

Cognitive Services

LM Informatica & Data Science - 2nd semester - 6 CFU

Visual Recognition and Machine Perception

Lamberto Ballan

Our class on Cognitive Services

- P1 (Sperduti): “Applied ML & Cognitive Services”
 - Introduction: from human cognition to smart cognitive services; brief intro to AI and ML paradigms.
 - Cognitive Services: major services and API (IBM Watson, Microsoft, Google Cloud); enabling technologies.
 - Machine Learning and Application Issues; Representation learning; Training and testing; Evaluation measures.
- P2 (Ballan): “Visual Recognition & Machine Perception”
 - Computer Vision: “Teaching computers to see”: extract rich information from visual data; designing visual features.
 - Representation learning in vision and image understanding.
 - What’s in the box? How to build a visual recognition pipeline; combining different services in a multi-modal scenario.

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

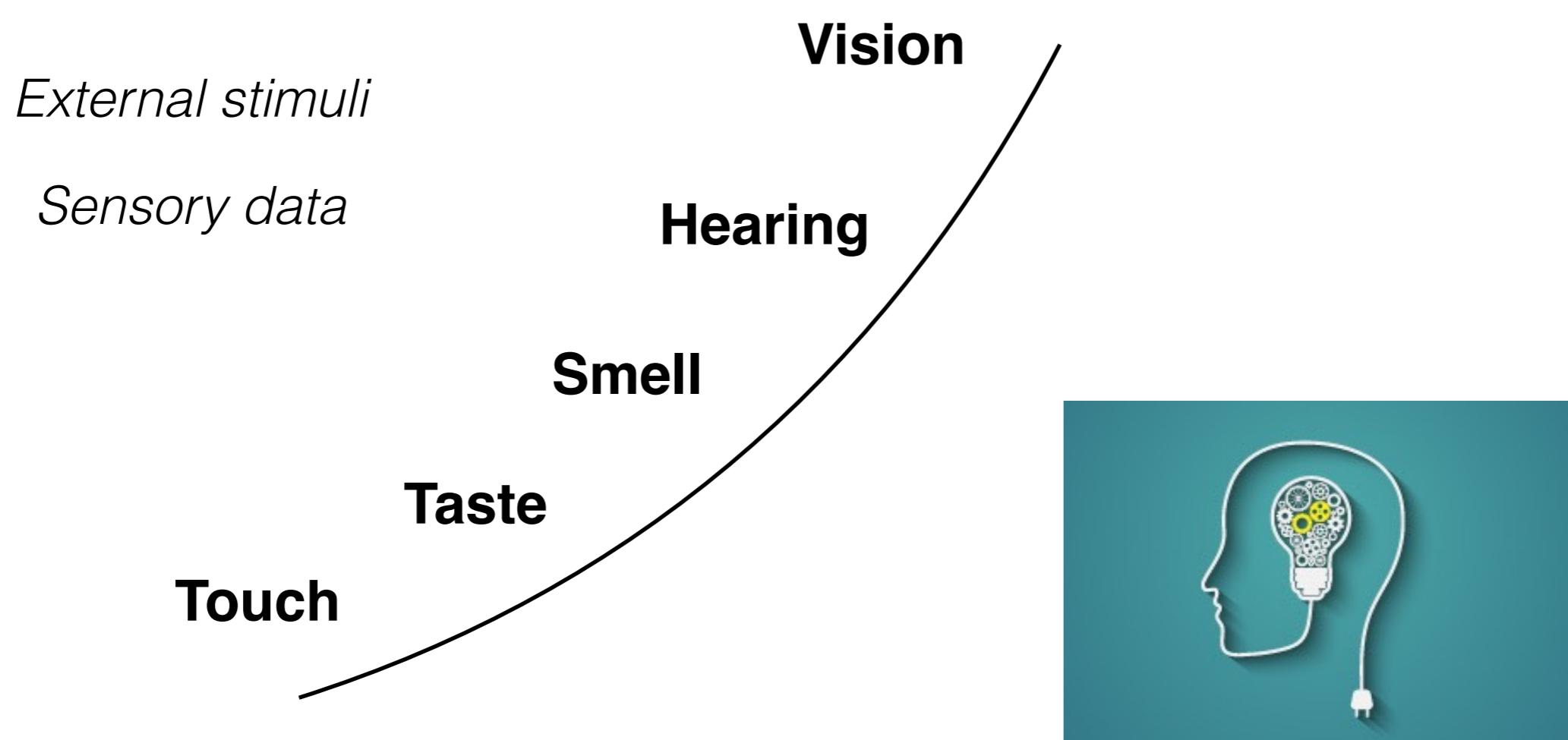
Deep Learning revolution

- March 27, 2018: “Fathers of the Deep Learning revolution” receive ACM Turing Award (*i.e. the “Nobel prize” in CS*)



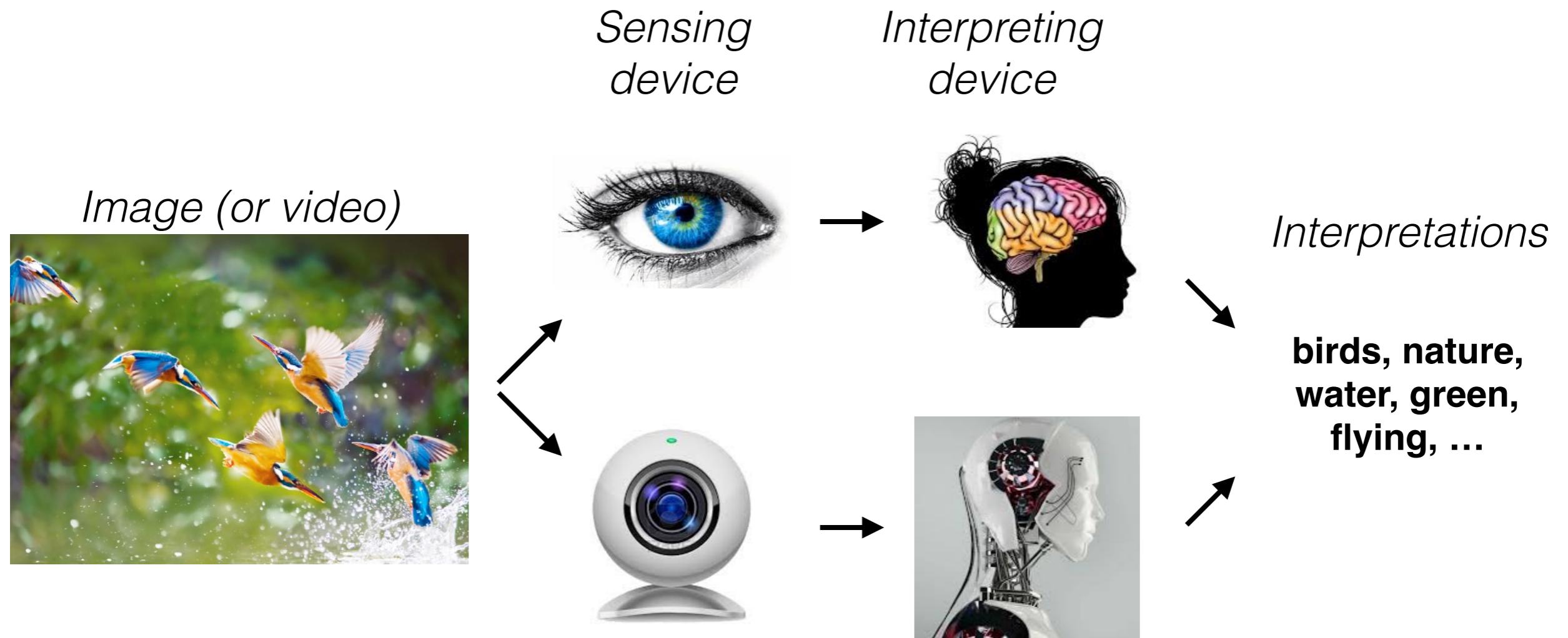
(Machine) Perception

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

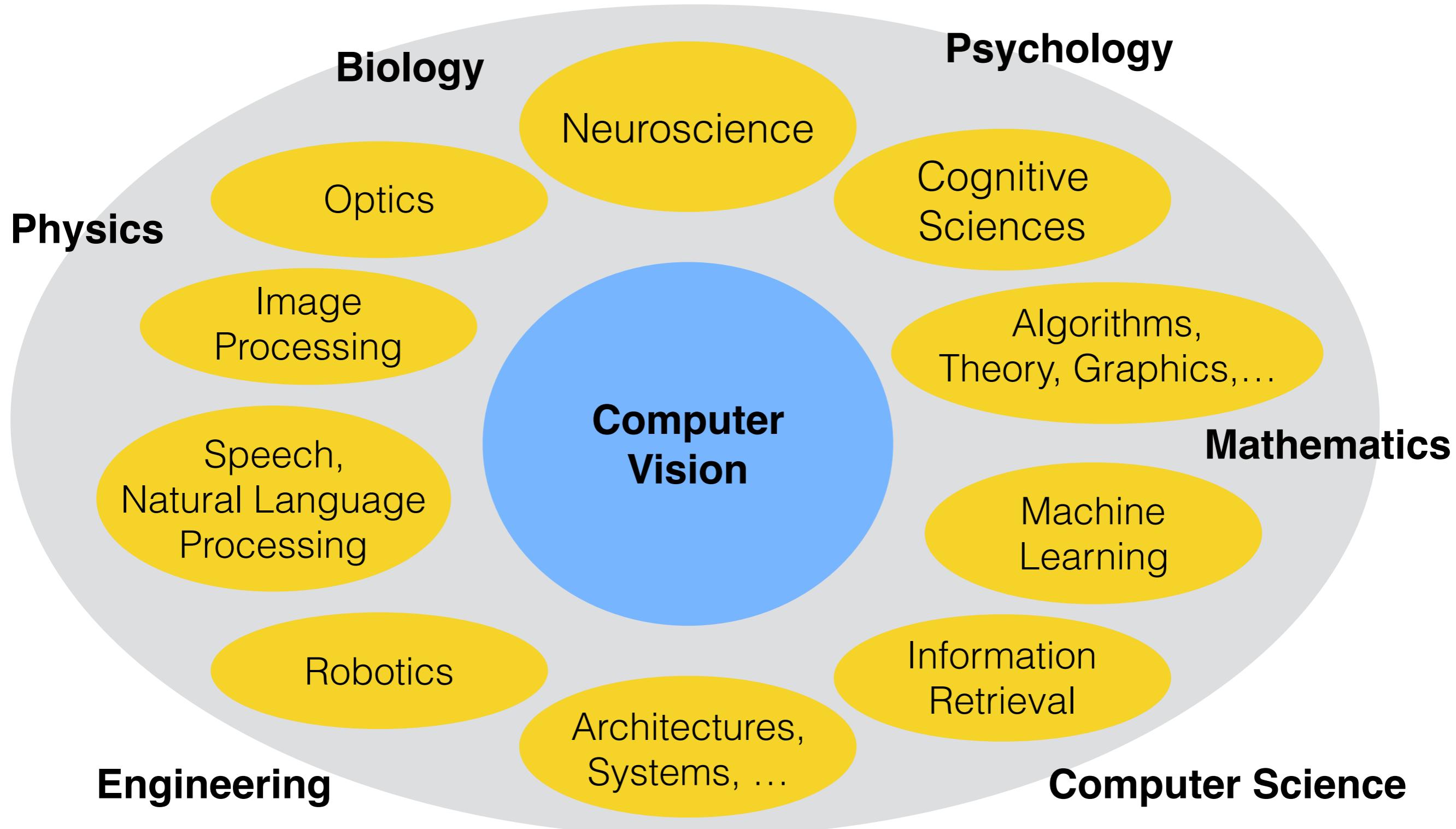


What is (computer) vision?

- Visual recognition is the holy grail of computer vision

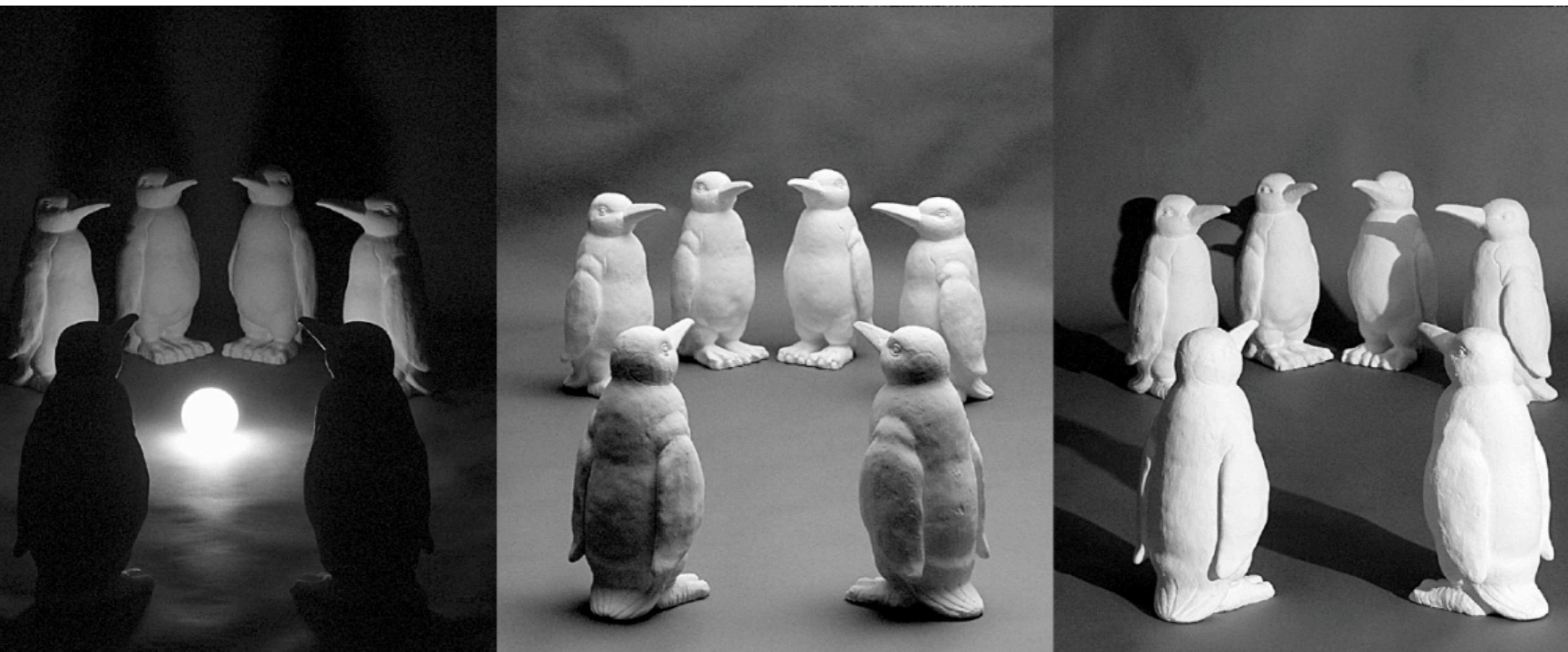


What is computer vision related to?



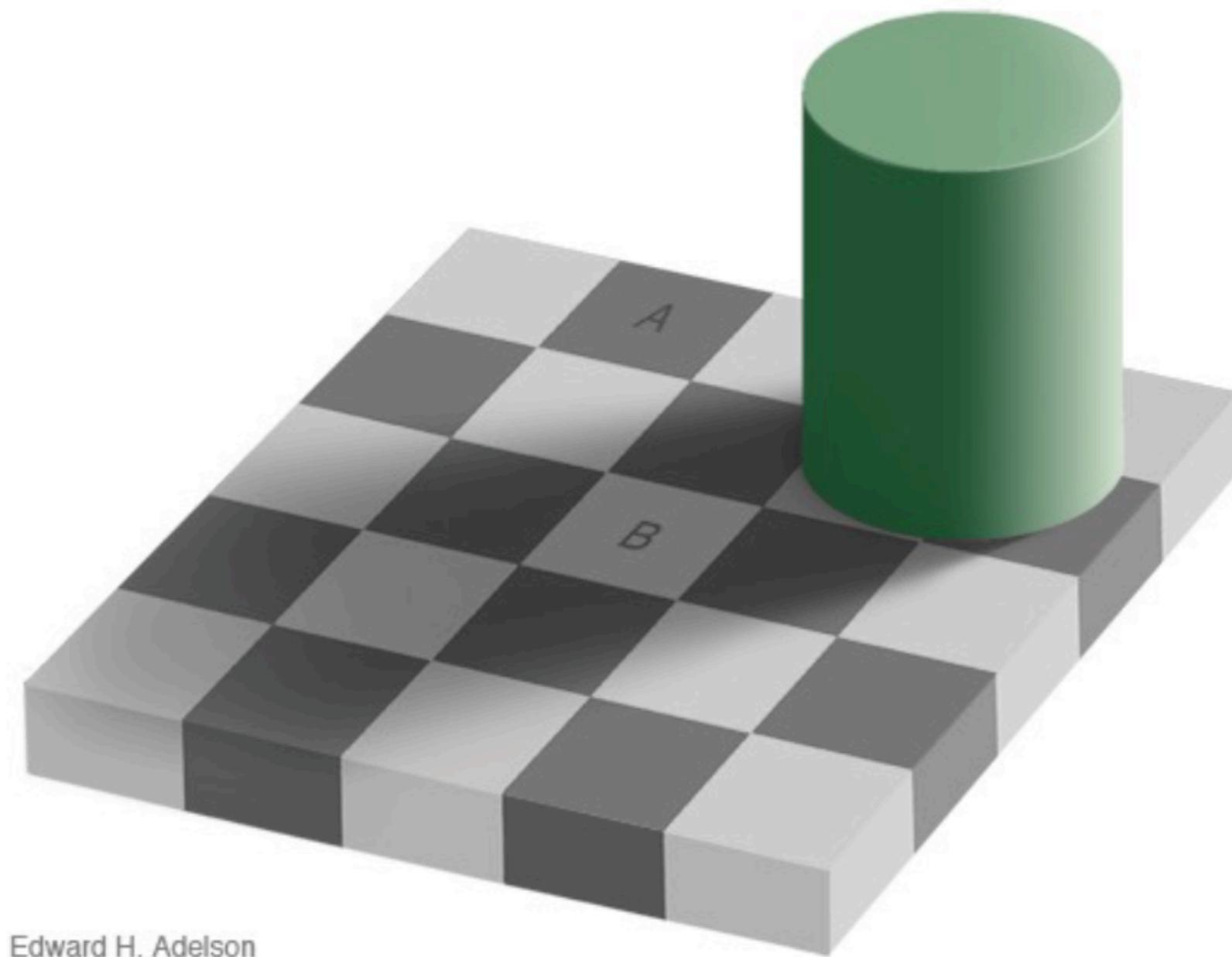
Visual perception: seeing is challenging

- Challenge: illumination



Seeing is challenging

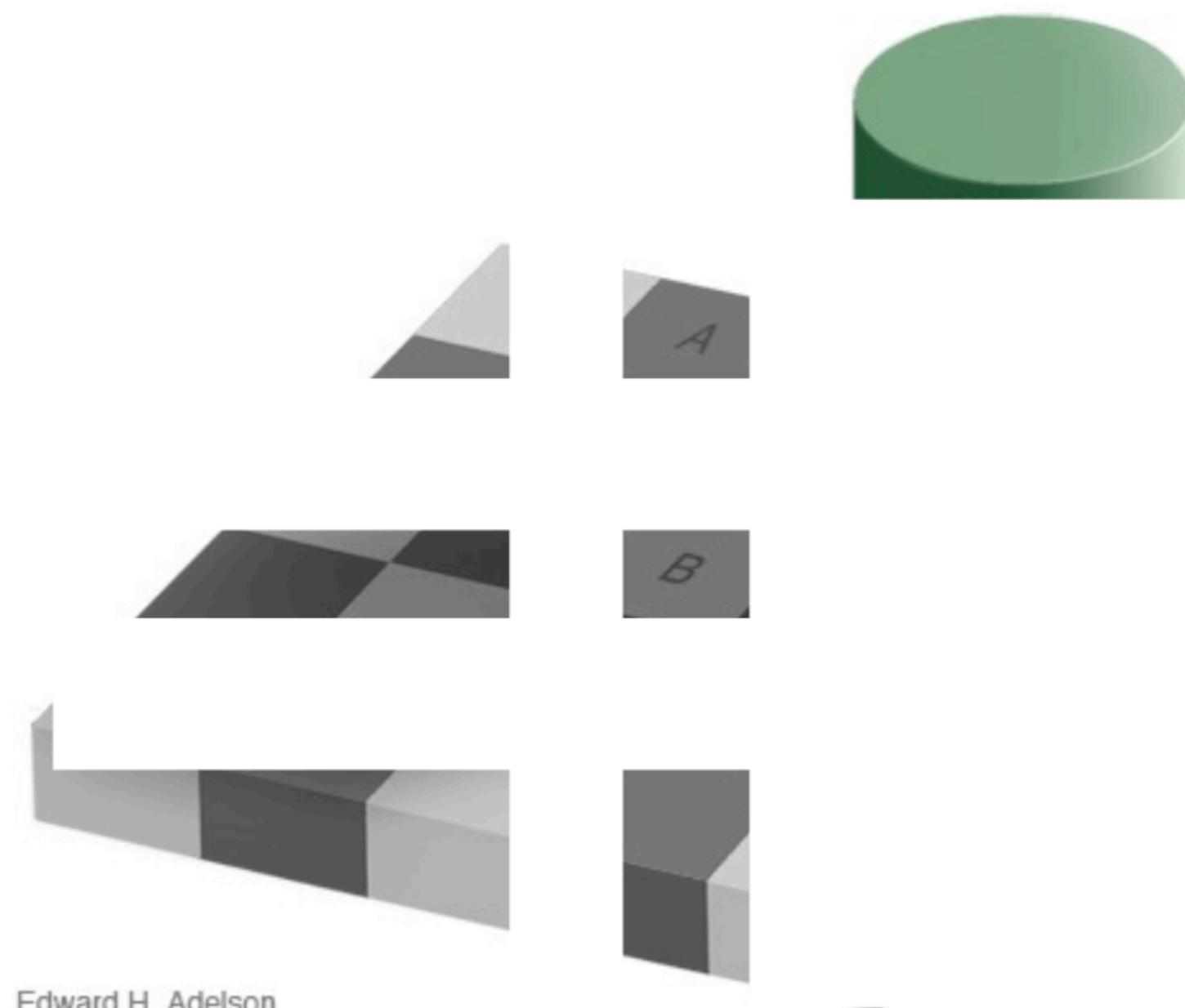
- Challenge: light and shadow



Edward H. Adelson

Seeing is challenging

- Challenge: light and shadow



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



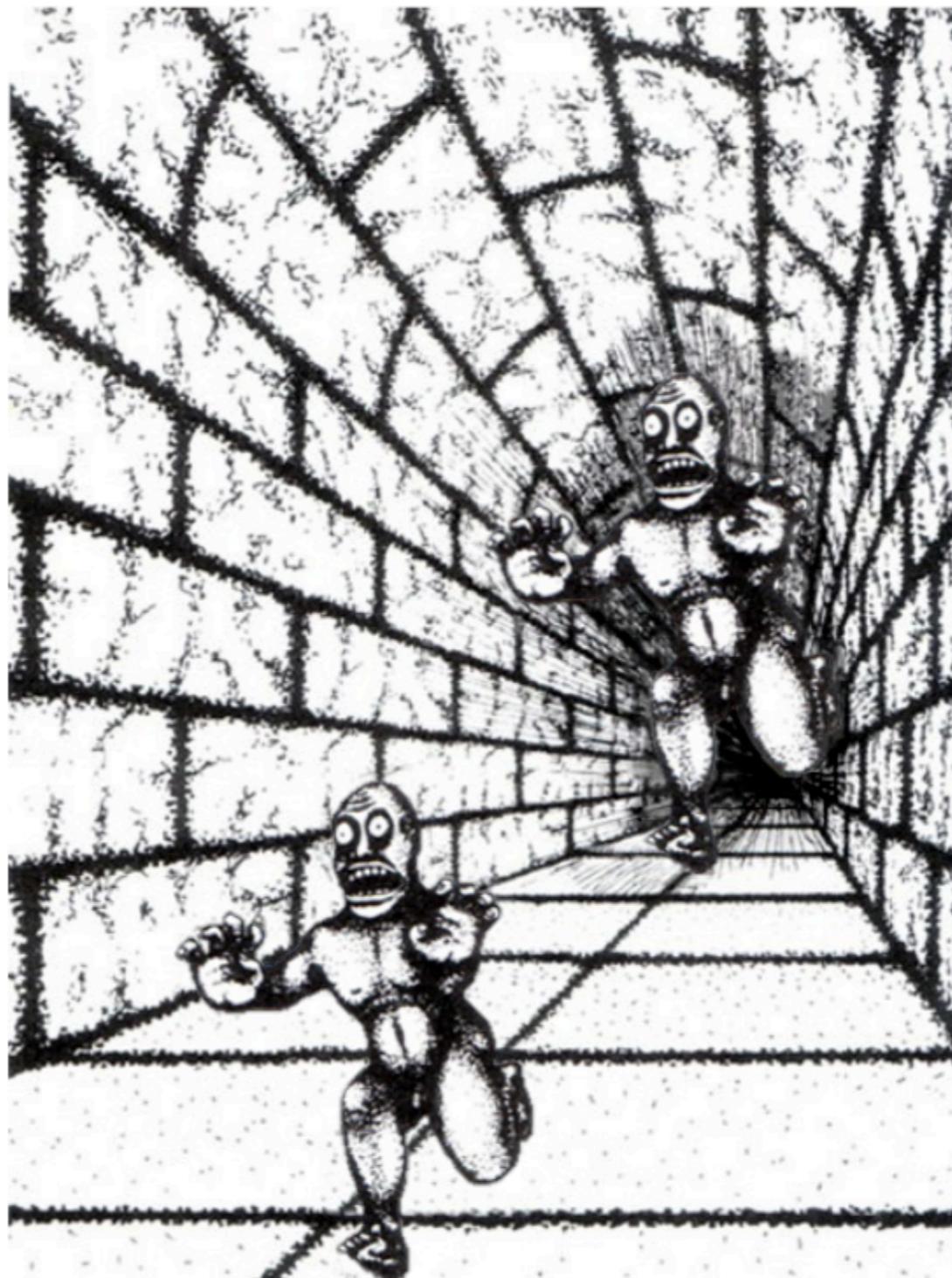
Seeing is challenging

- Challenge: scale



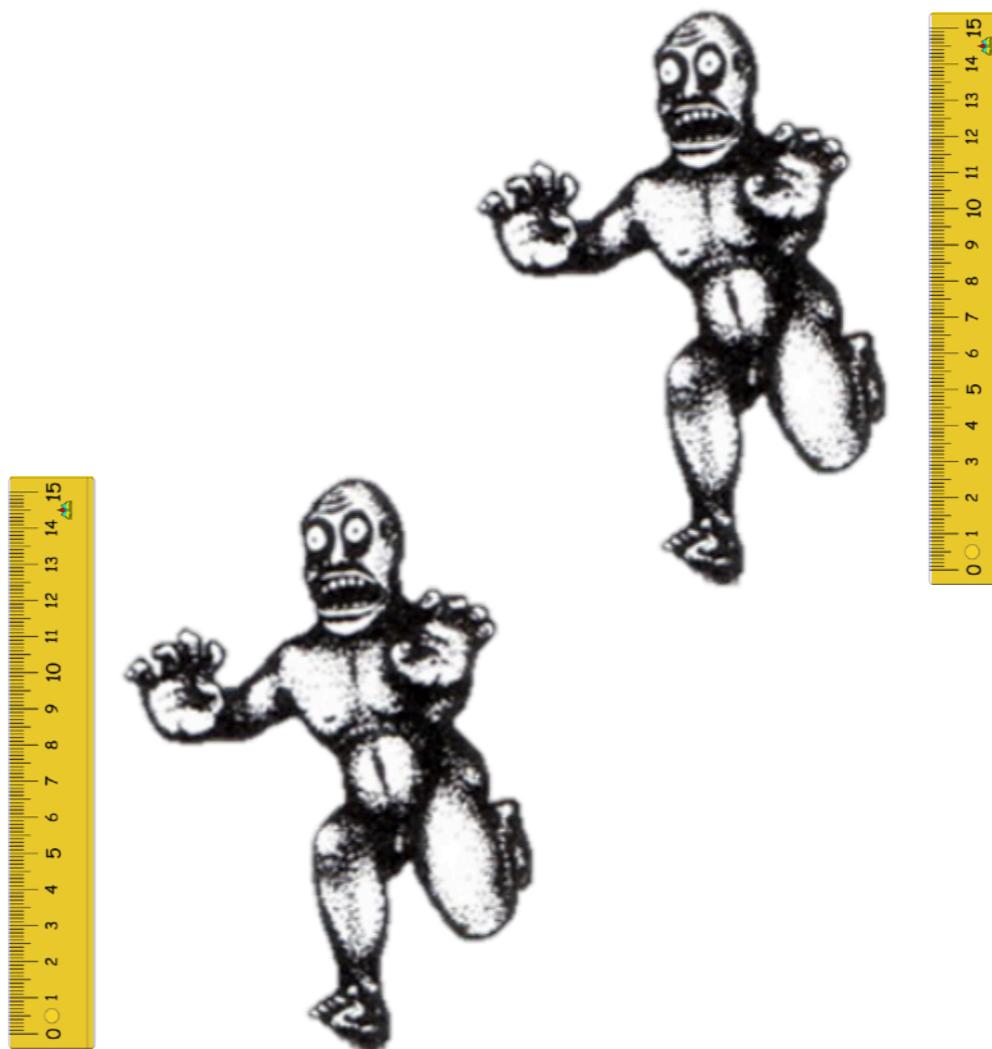
Seeing is challenging

- Challenge: scale and perspective view



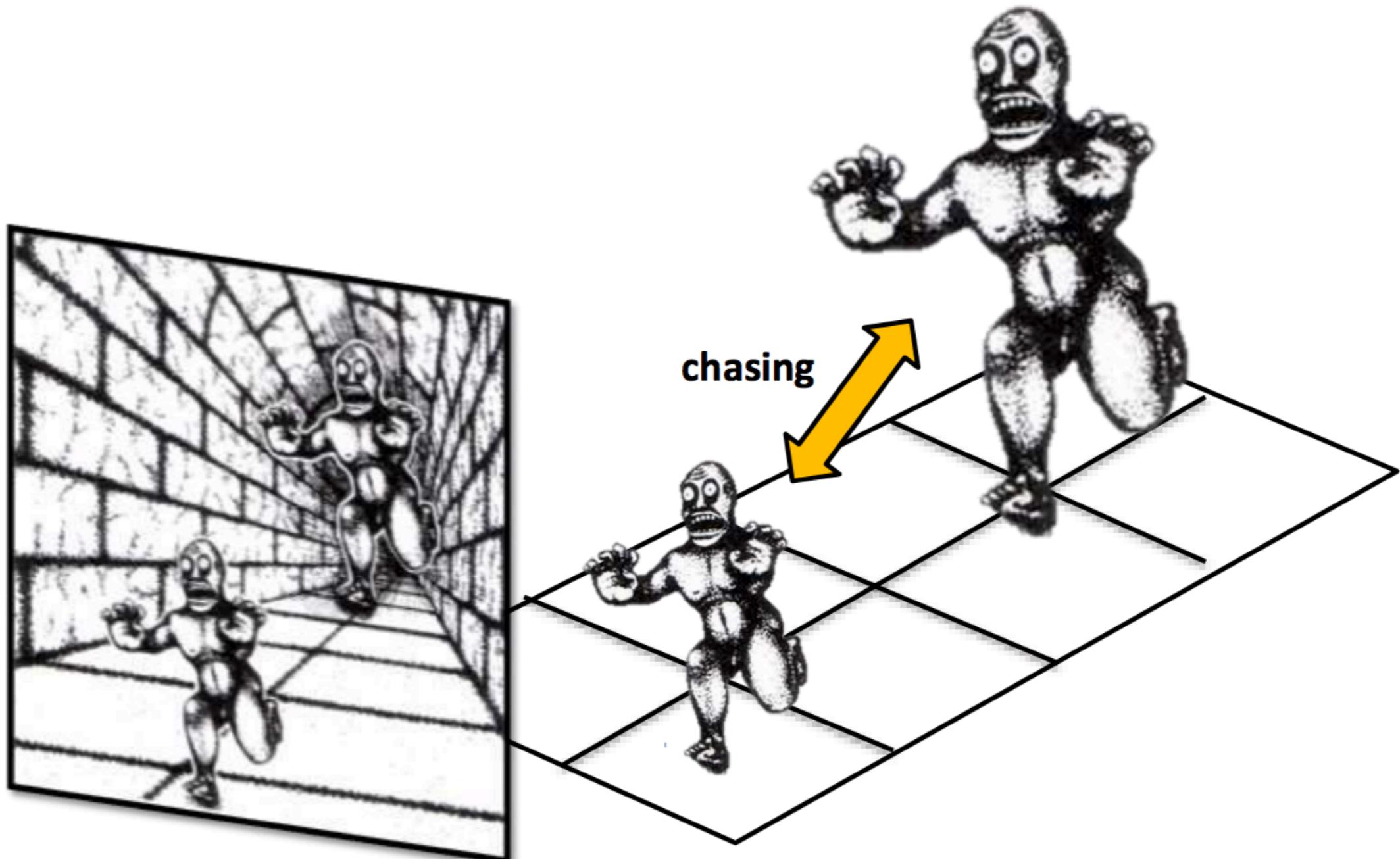
Seeing is challenging

- Challenge: scale and perspective view



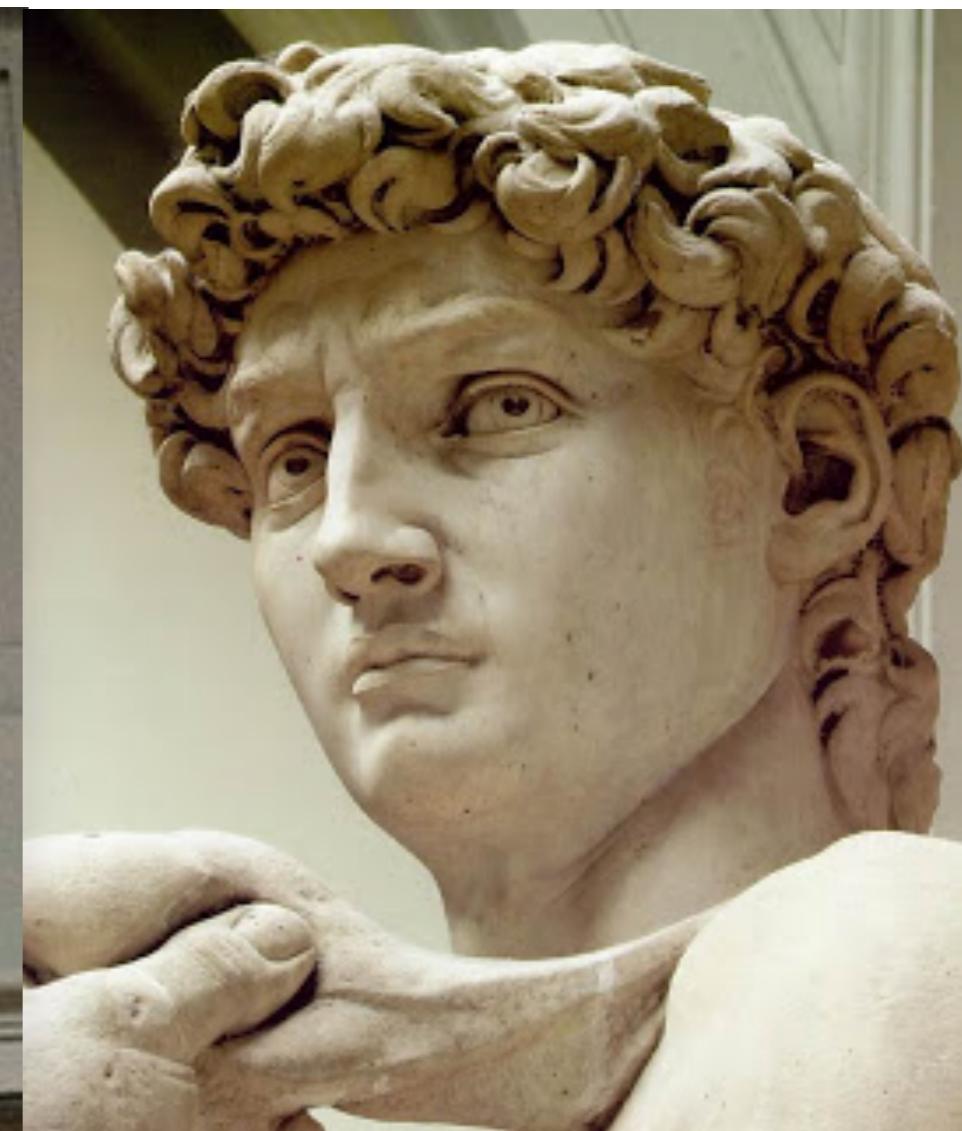
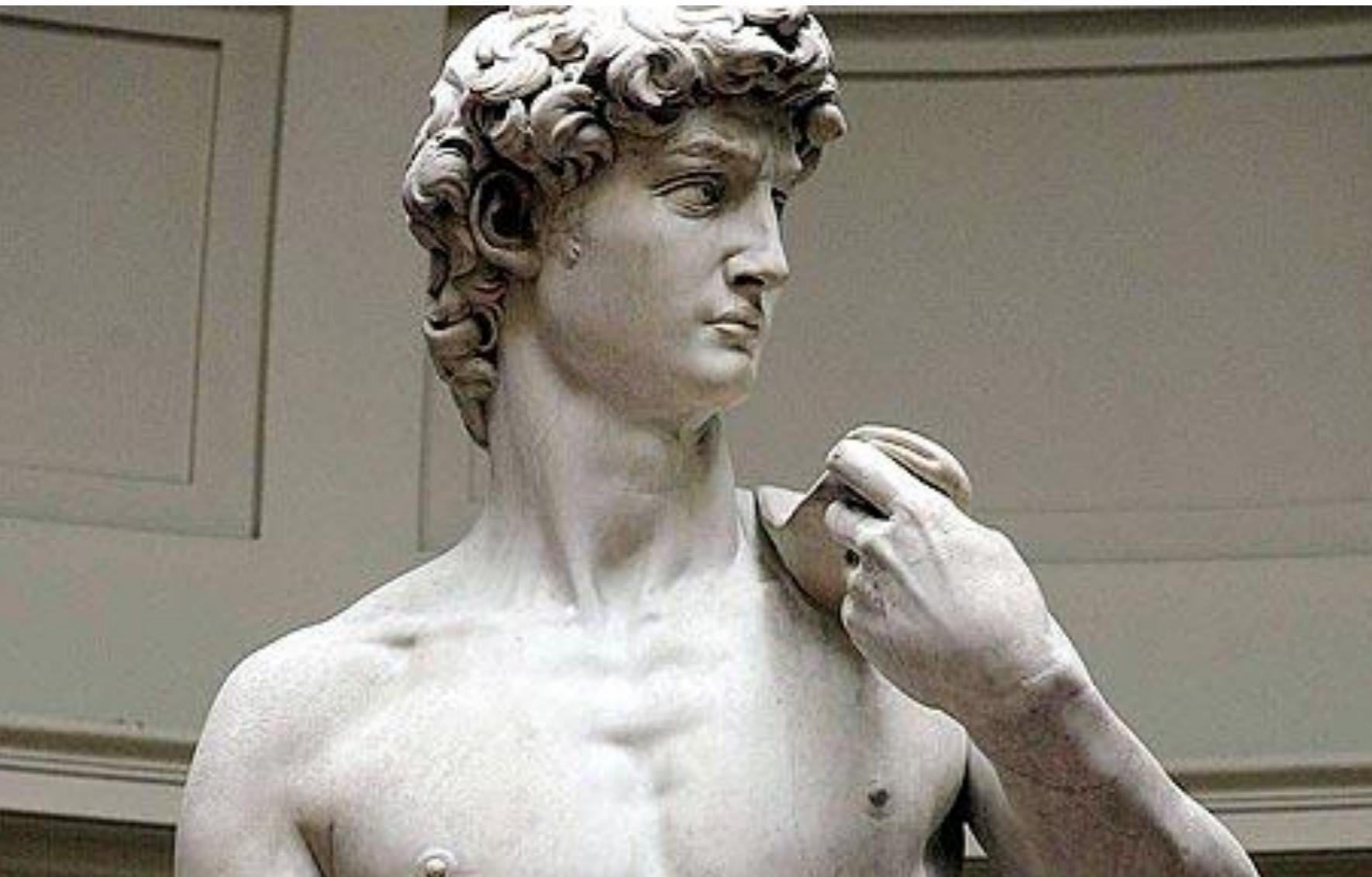
Seeing is challenging

- Challenge: scale and perspective view



Seeing is challenging

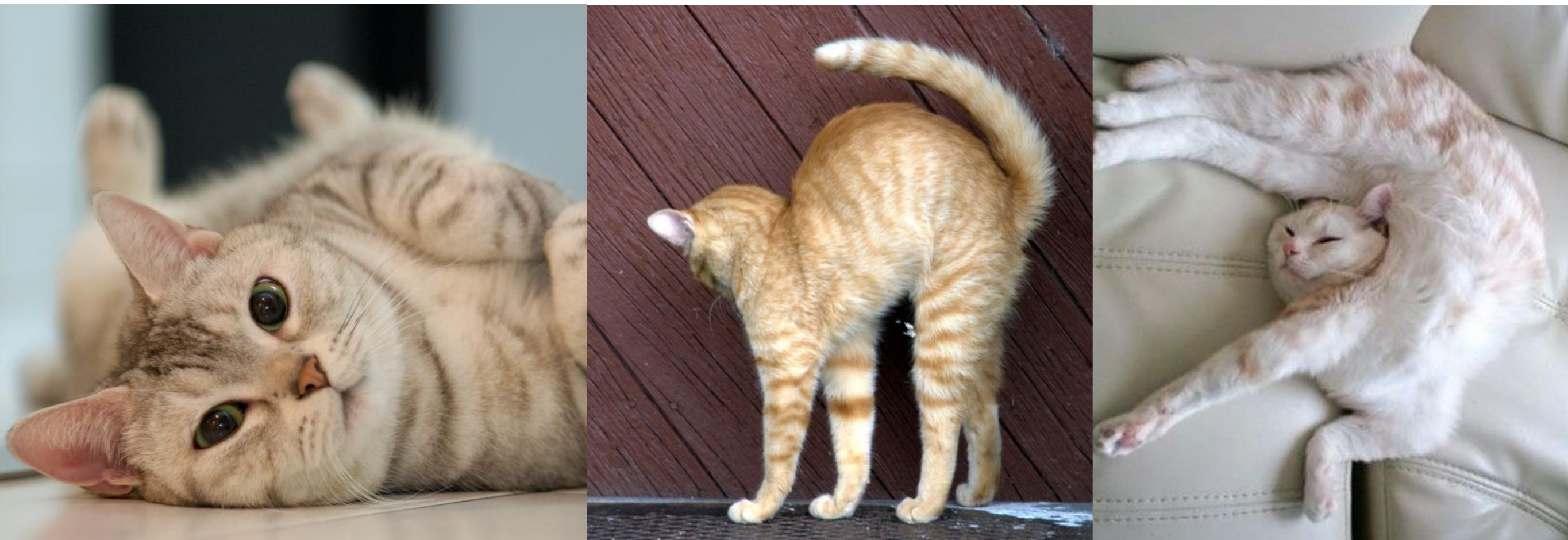
- Challenge: viewpoint



Michelangelo's David

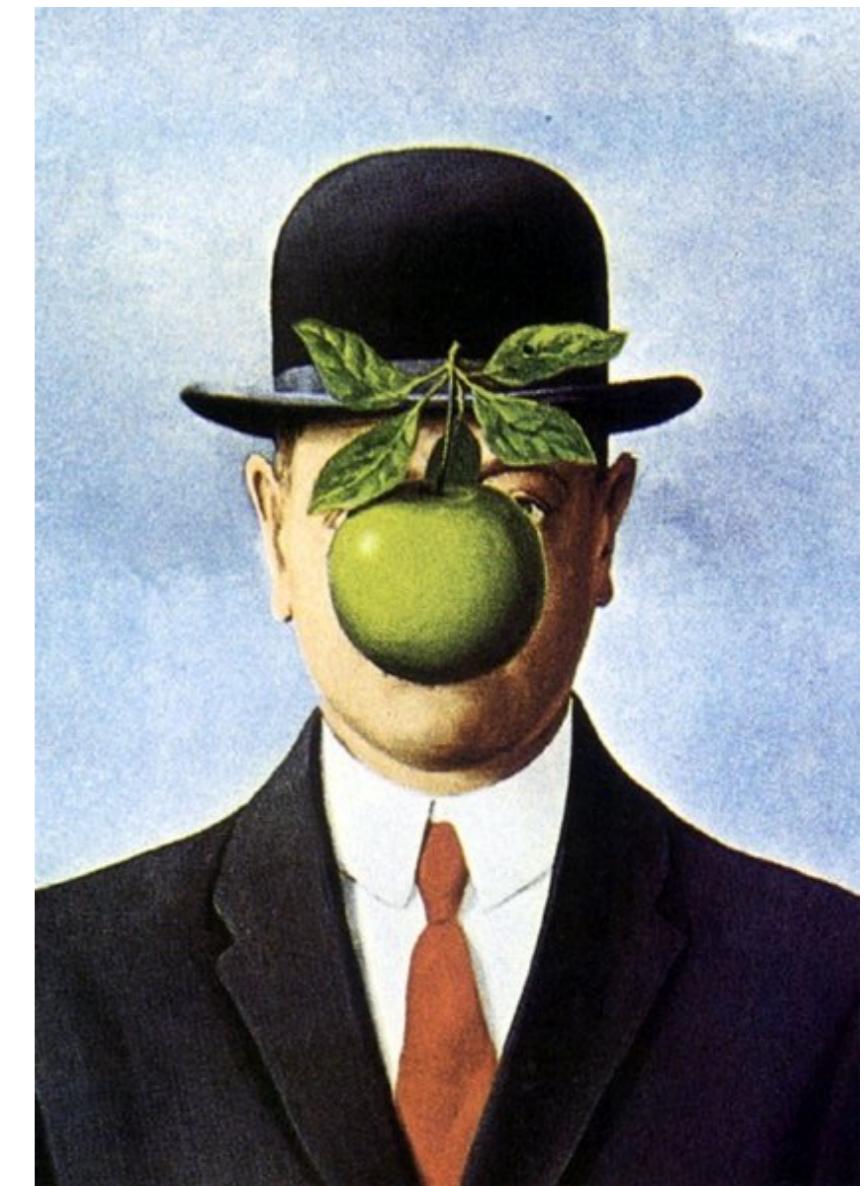
Seeing is challenging

- Challenge: deformation



Seeing is challenging

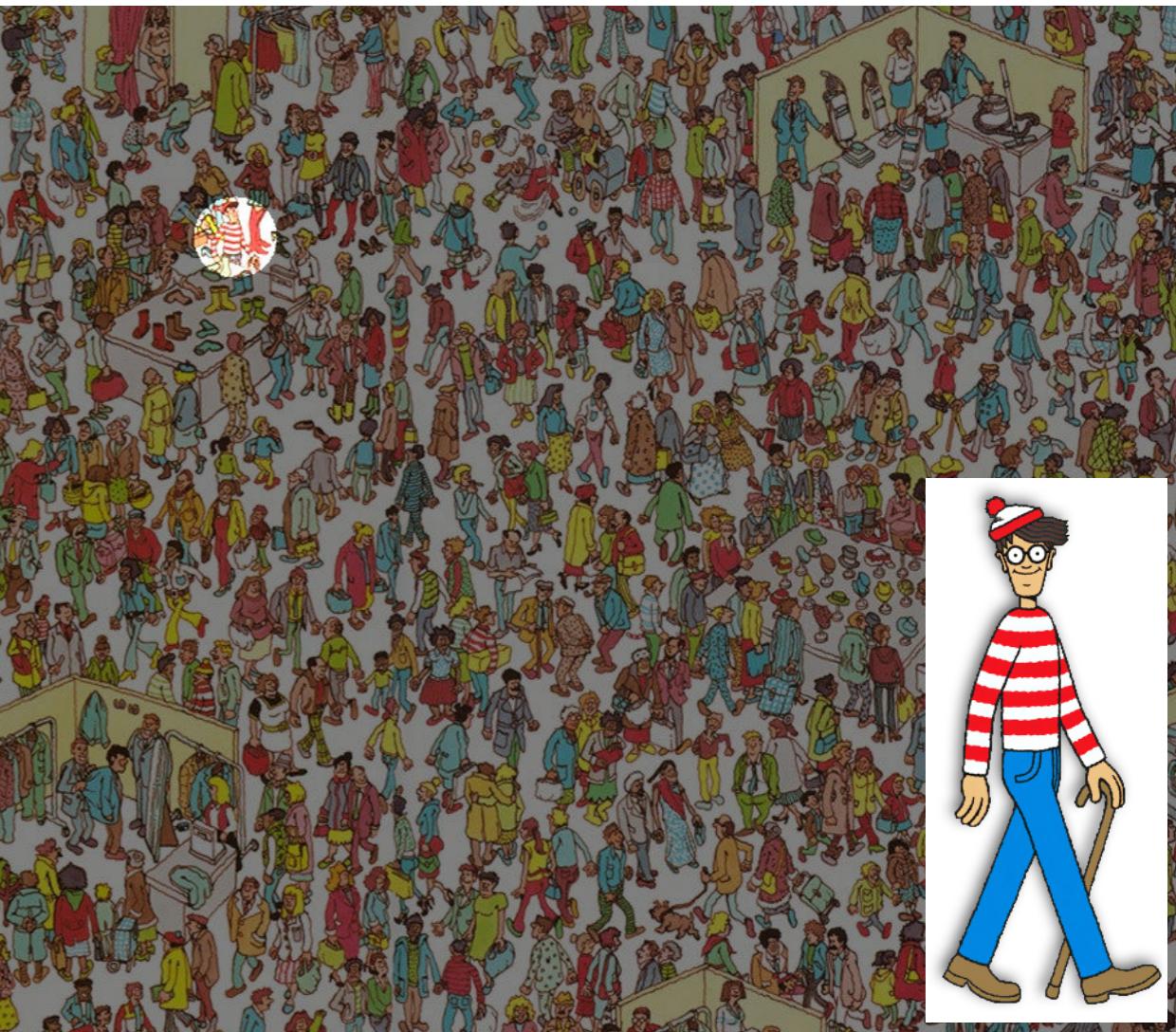
- Challenge: occlusion



Magritte's "The Son of Man"

Seeing is challenging

- Challenge: clutter



Seeing is challenging

- Challenge (categorization): intra-class variation



Seeing is challenging

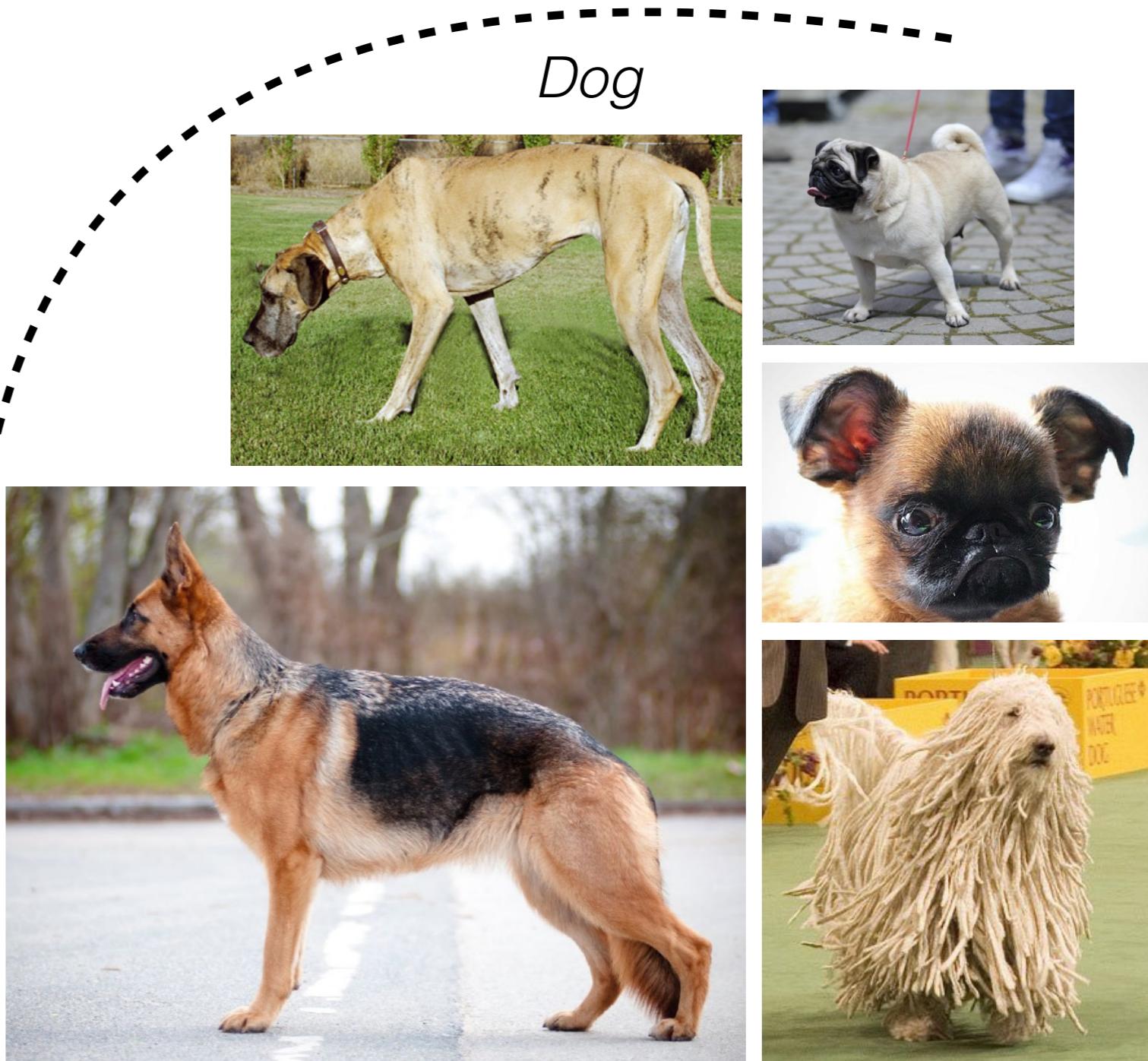
- Challenge (categorization): inter-class similarity

Wolf

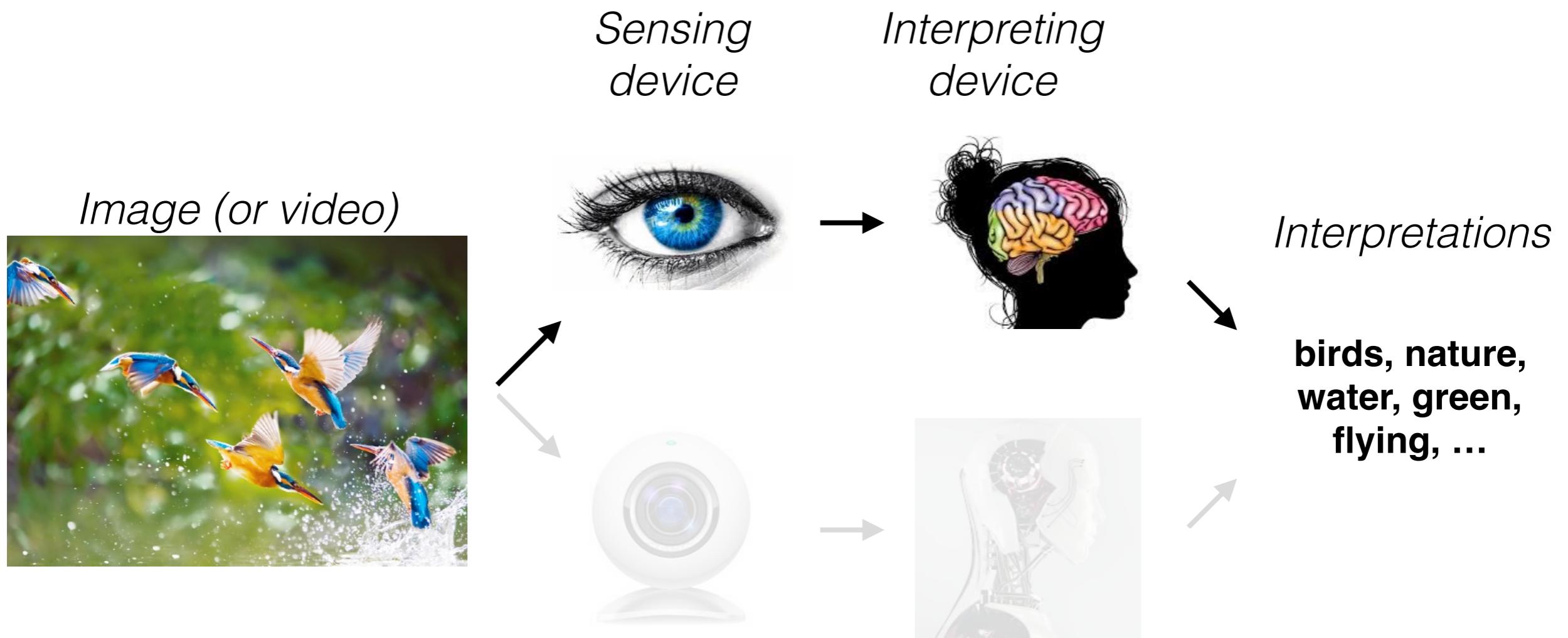


Copyright Jon Atkinson 2010

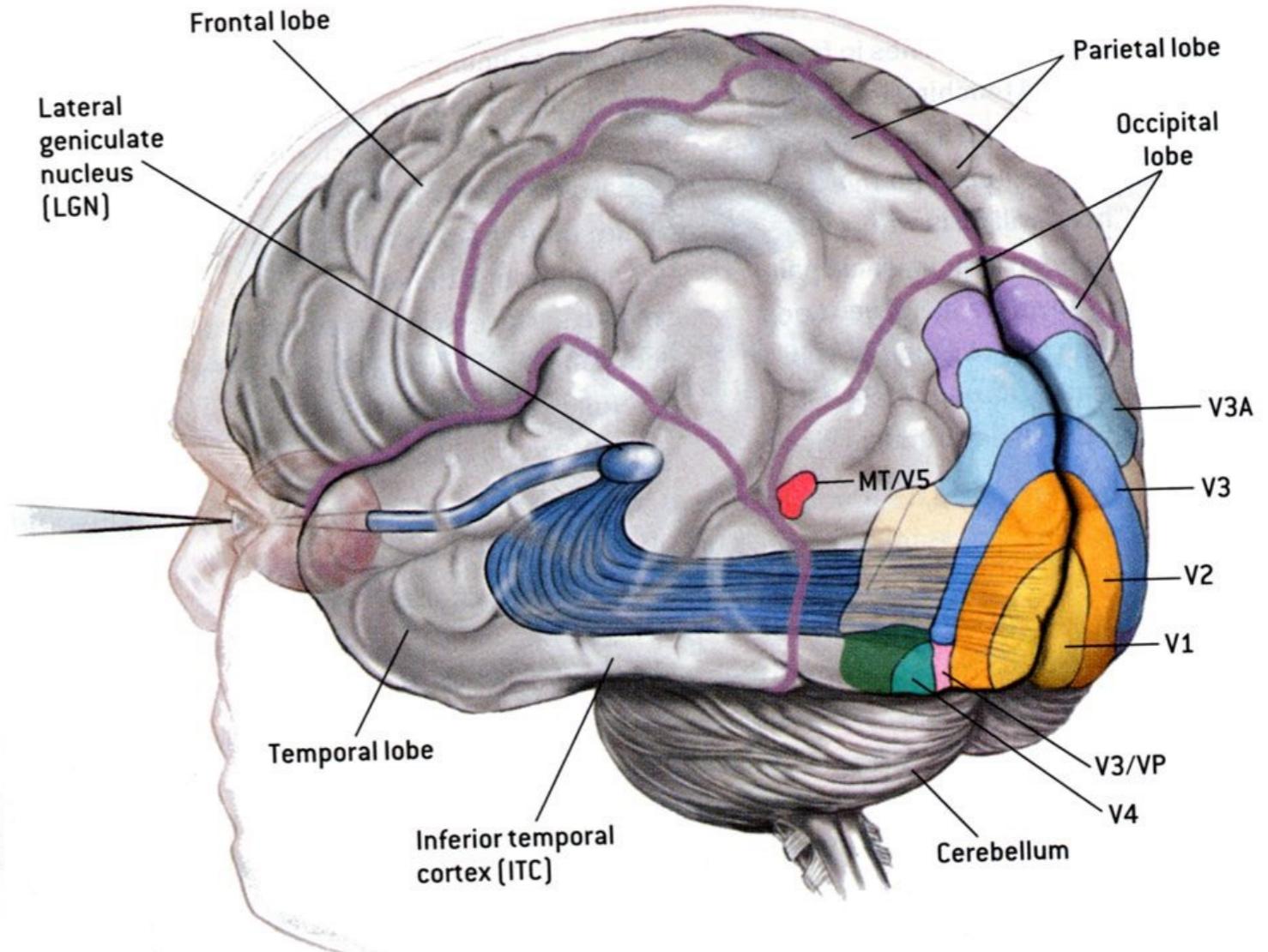
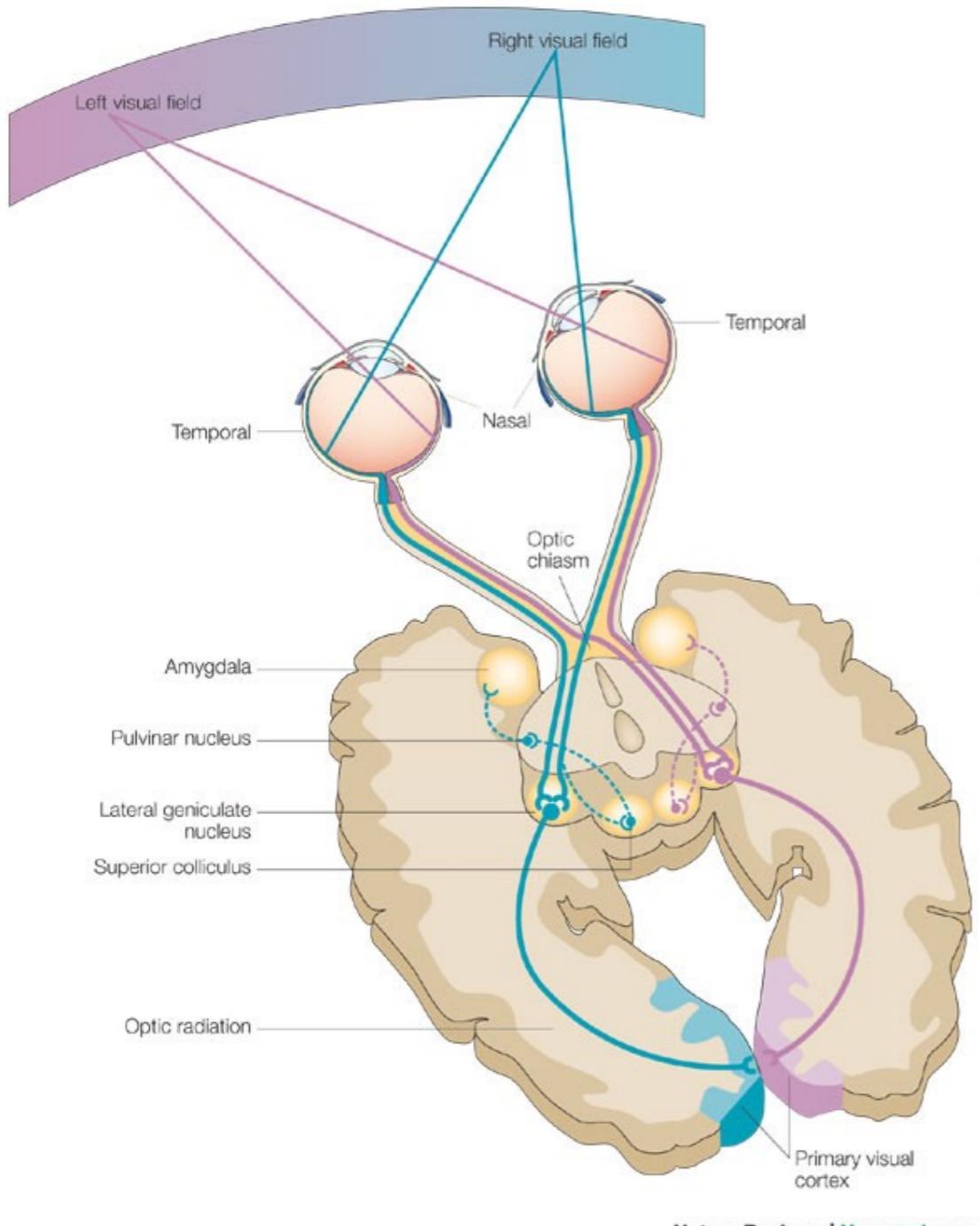
Dog



What is (computer) vision?



Human (mammalian) vision



Functional specialization:

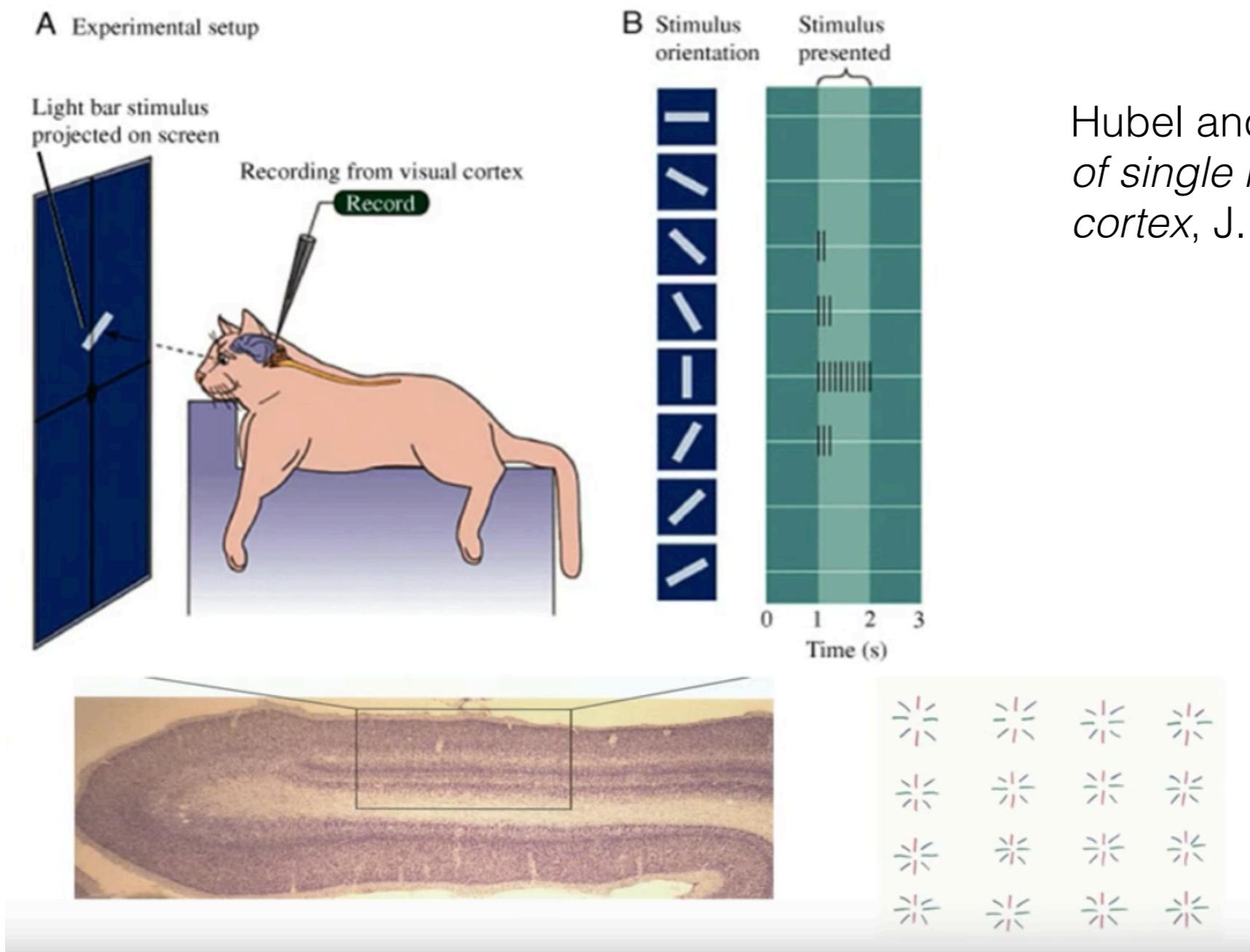
V1: primary visual cortex

V4: color, V3/VP: recognition

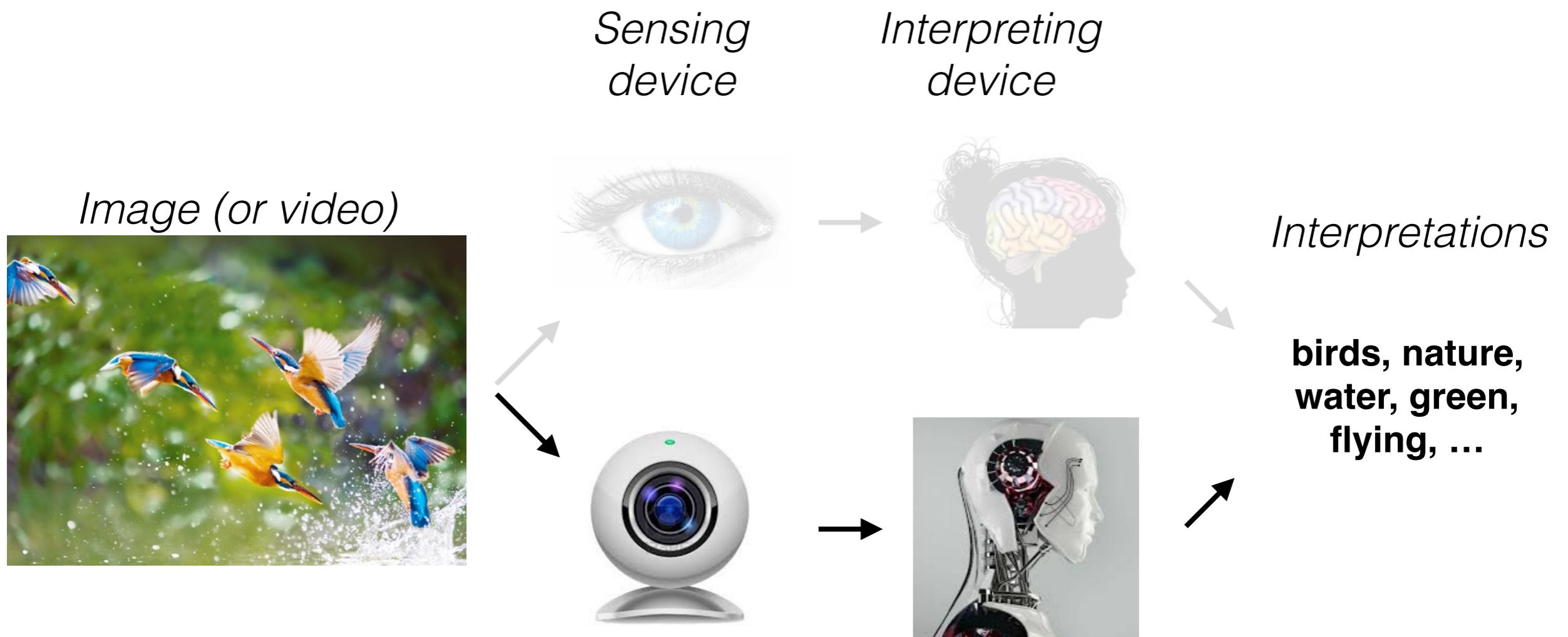
MT/V5: motion

Human (mammalian) vision

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



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

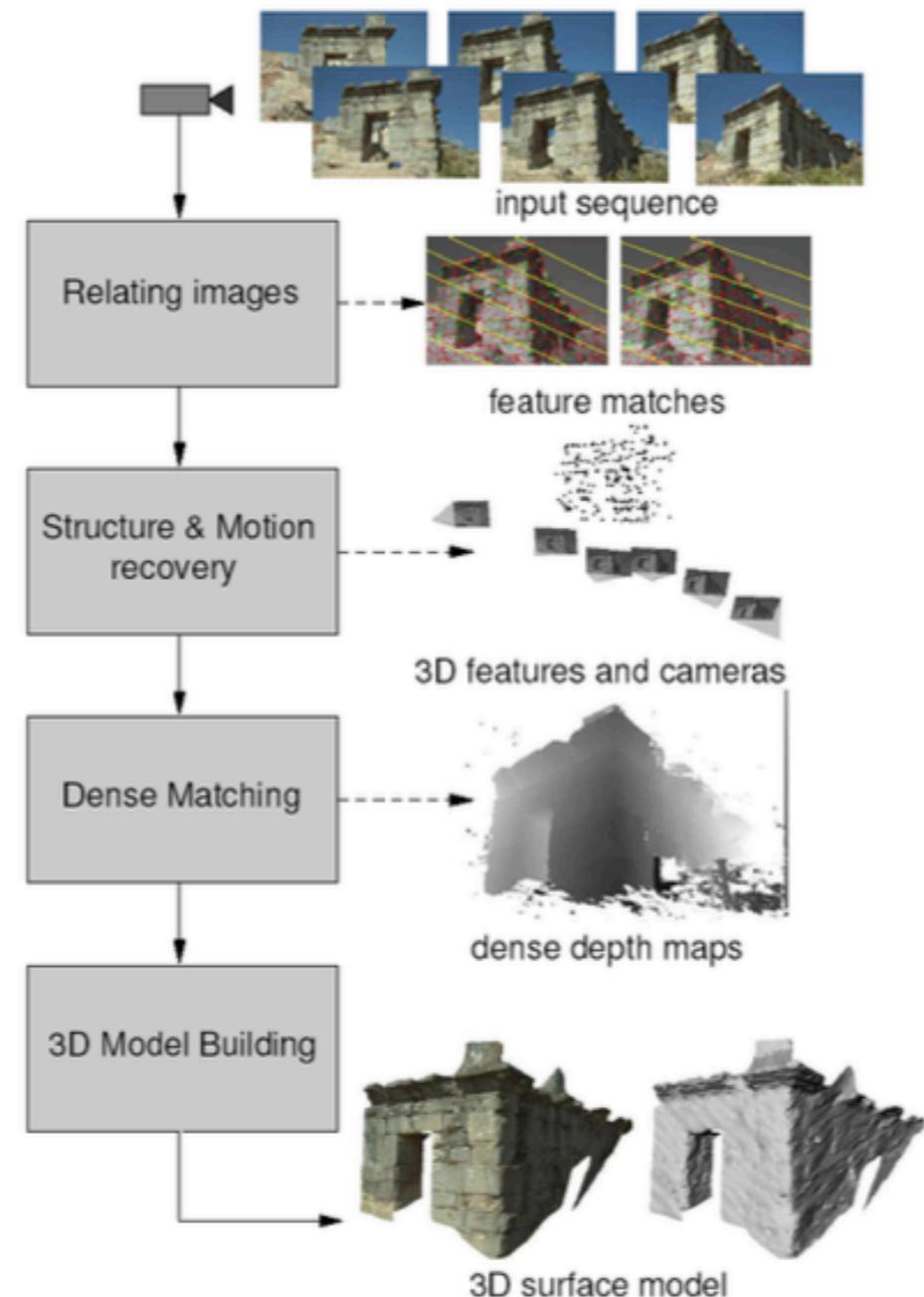
What a computer sees

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

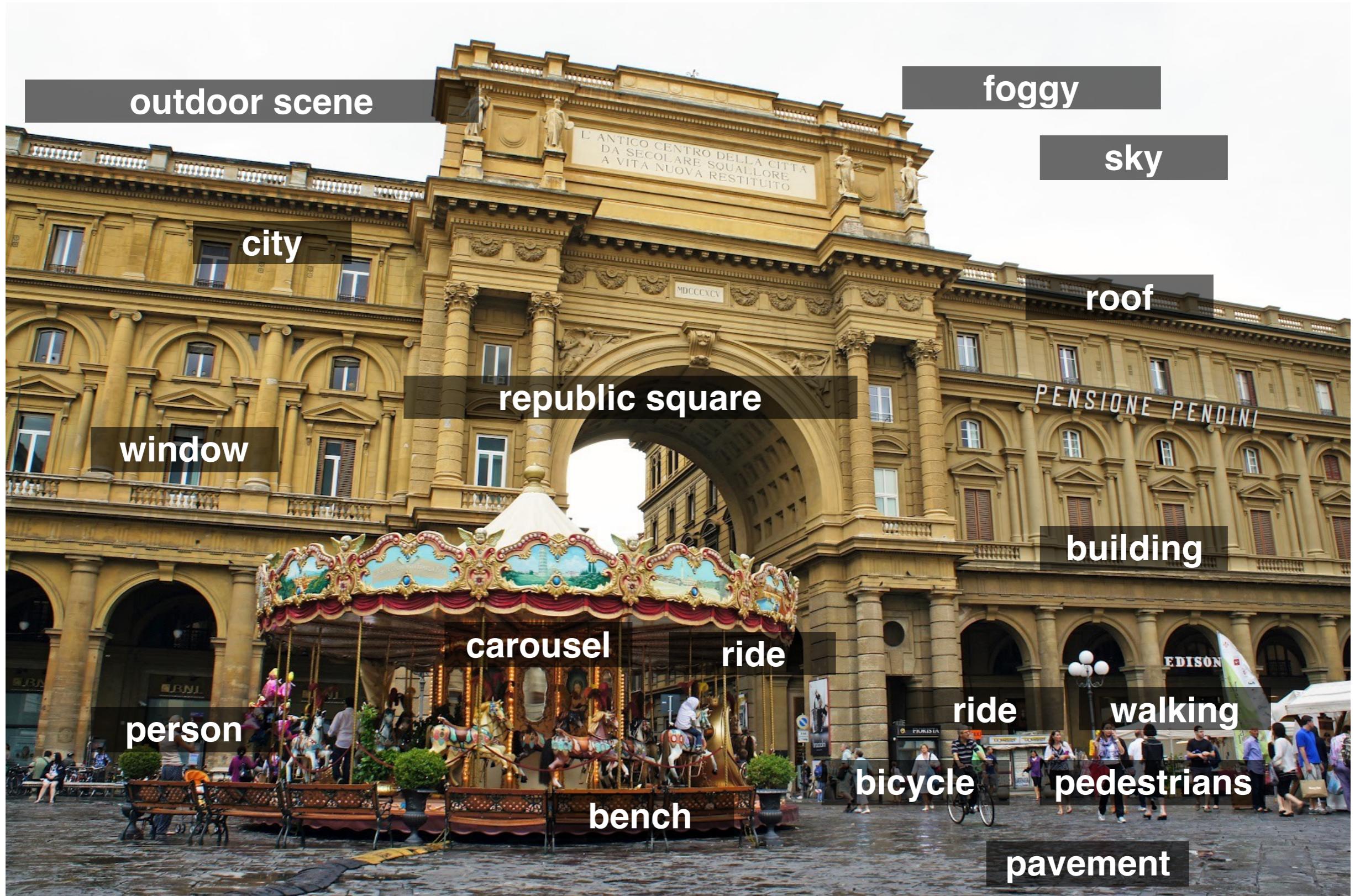


Pollefeys et al.



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

CV “started” as an undergraduate summer project at MIT

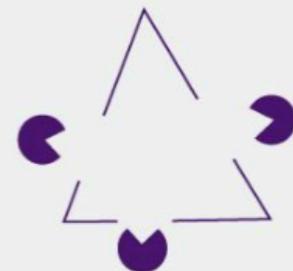
David Marr’s influential work in the 1970s

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".

VISION



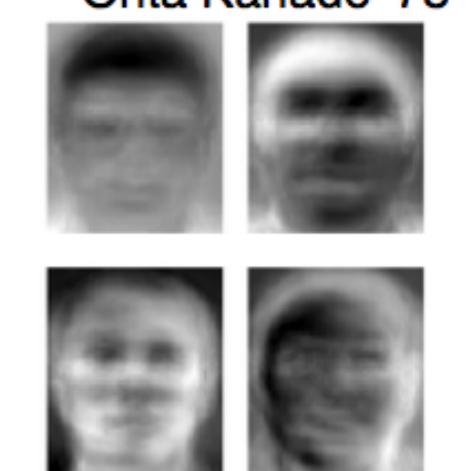
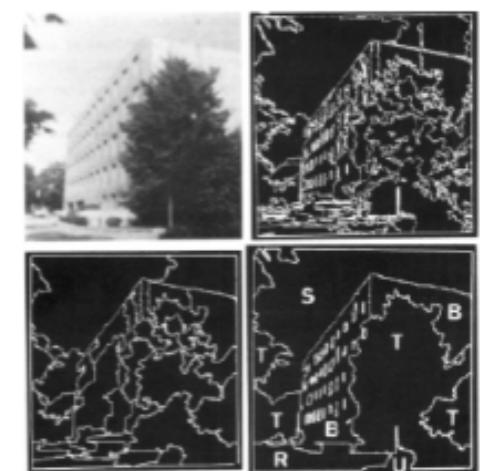
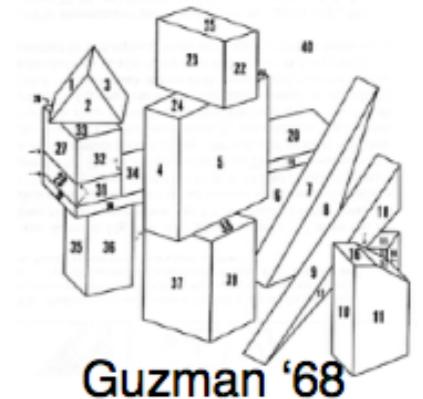
David Marr

FOREWORD BY
Shimon Ullman
AFTERWORD BY
Tomaso Poggio

Copyrighted Material

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

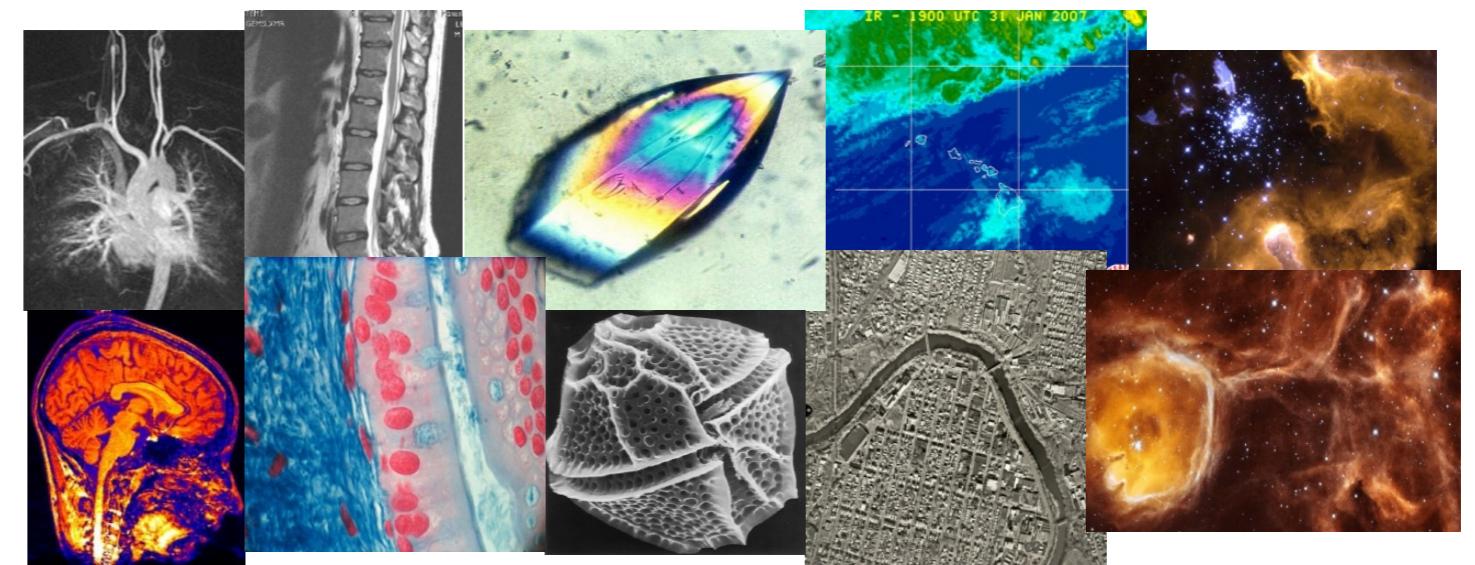


Why study computer vision?

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

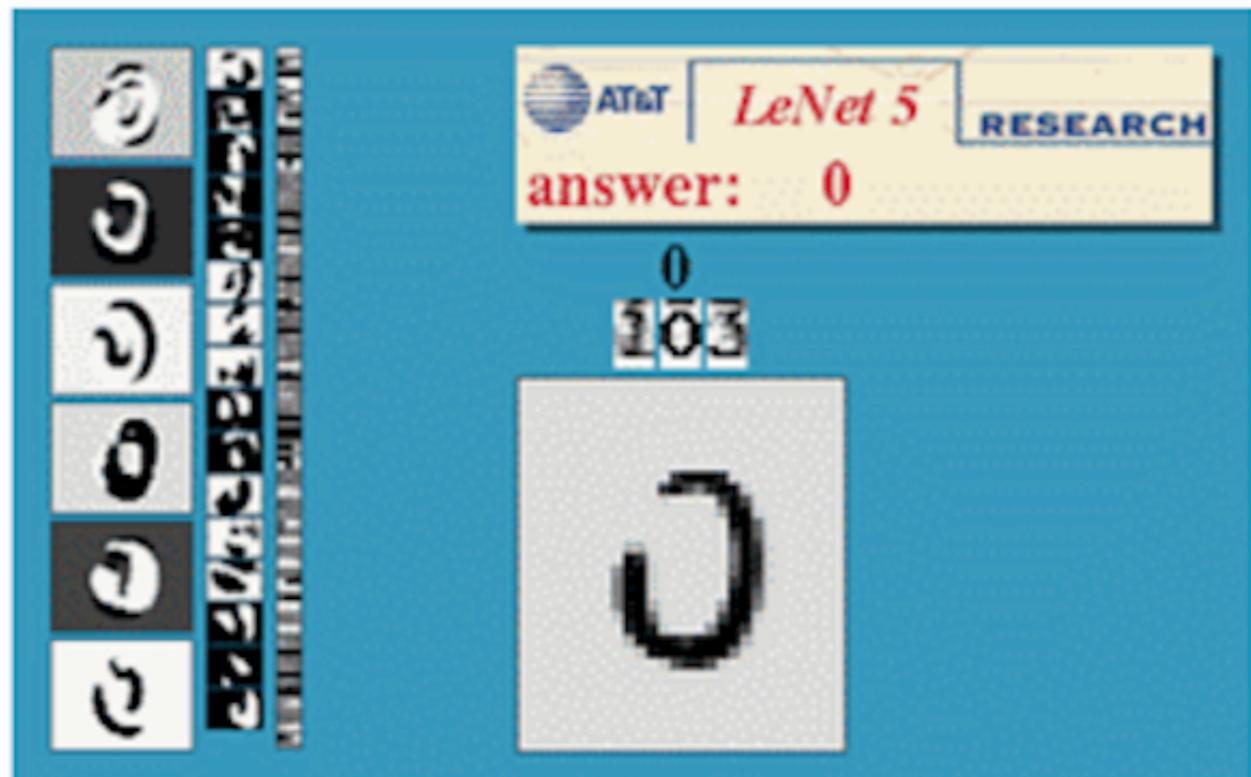


Surveillance and security



Medical and scientific images

Applications: optical character recognition

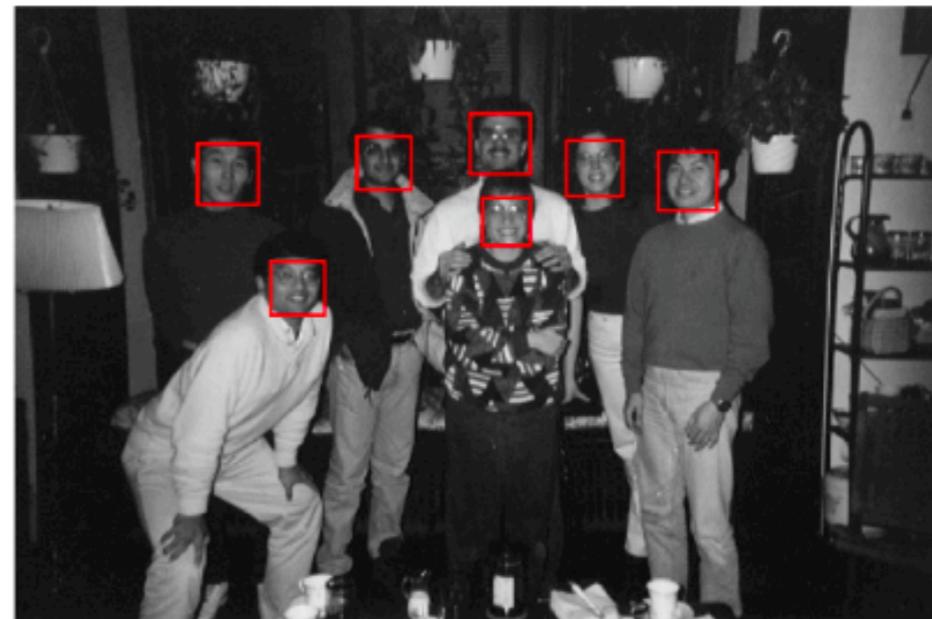
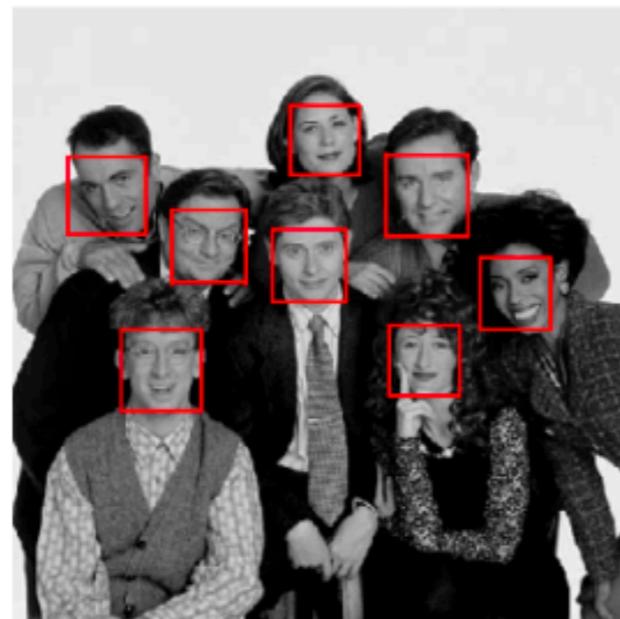
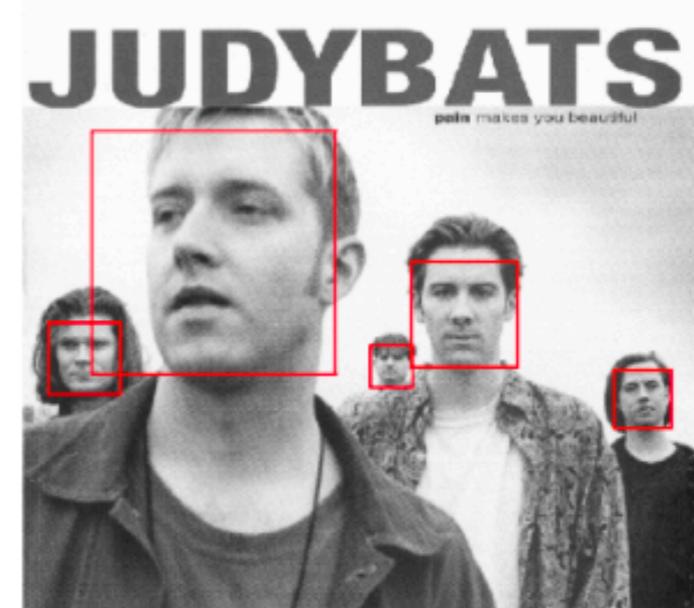
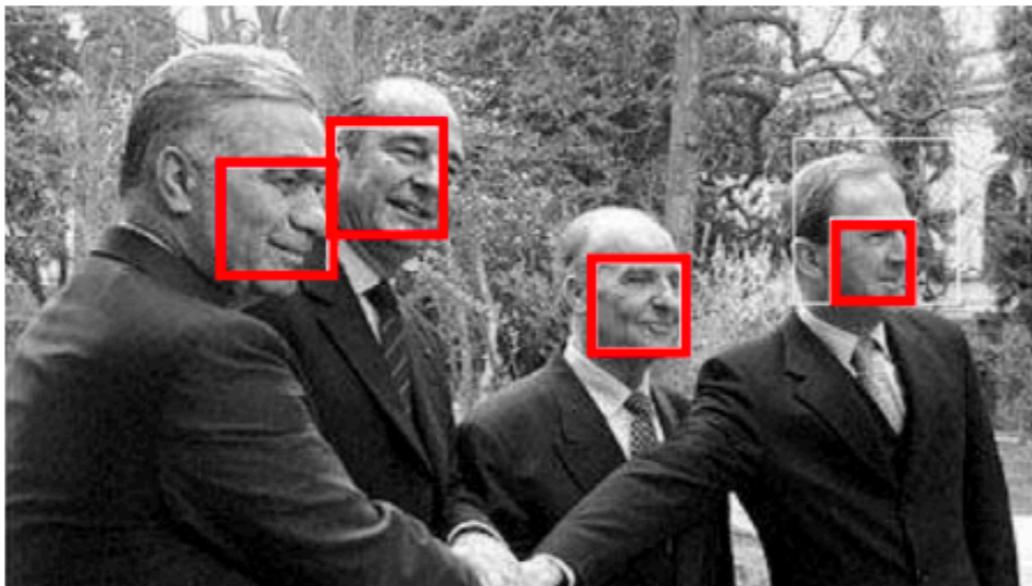


Digit recognition, AT&T labs



License plate readers

Applications: face detection



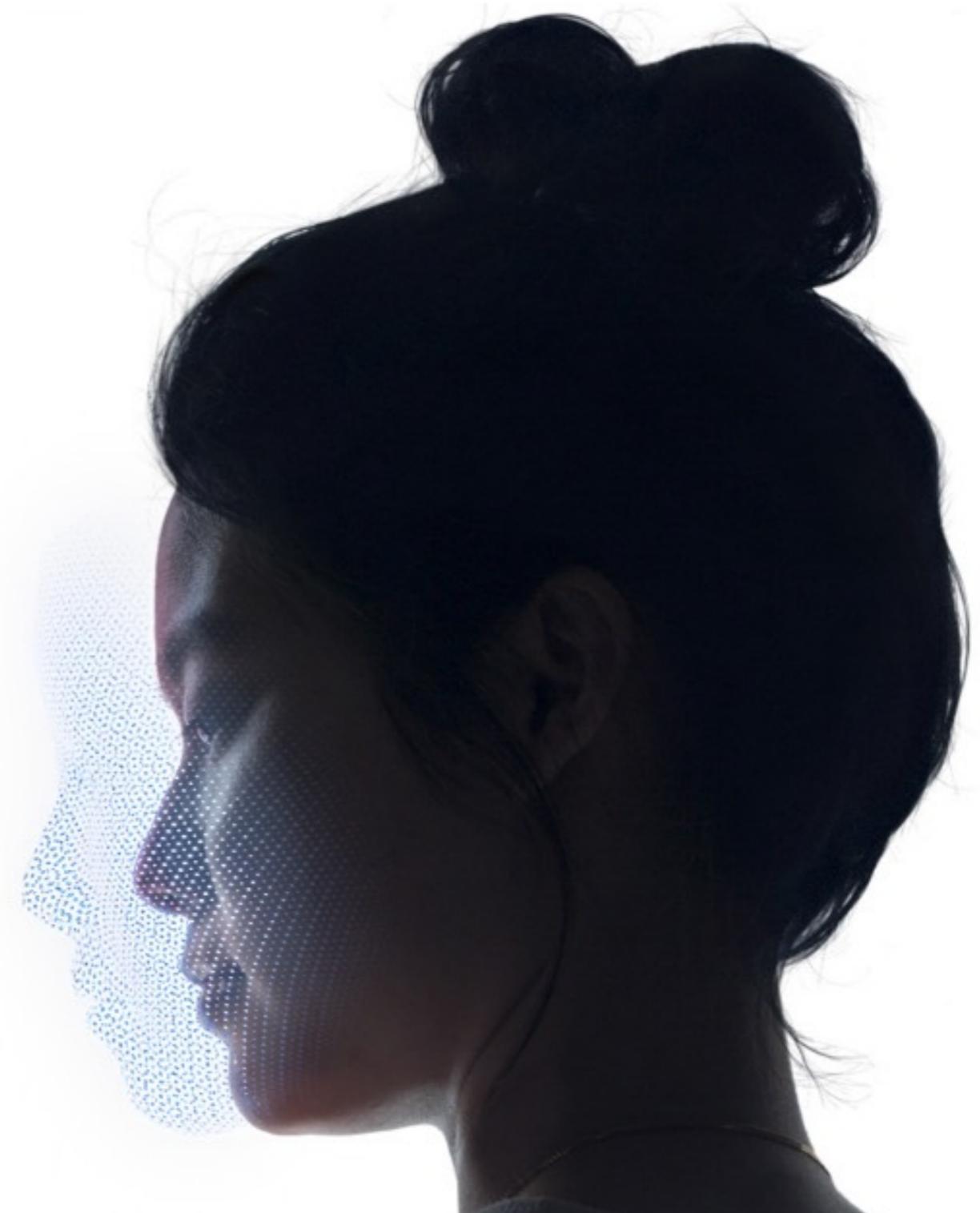
FinePix S6000fd, by Fujifilm, 2006

Viola & Jones, 2001

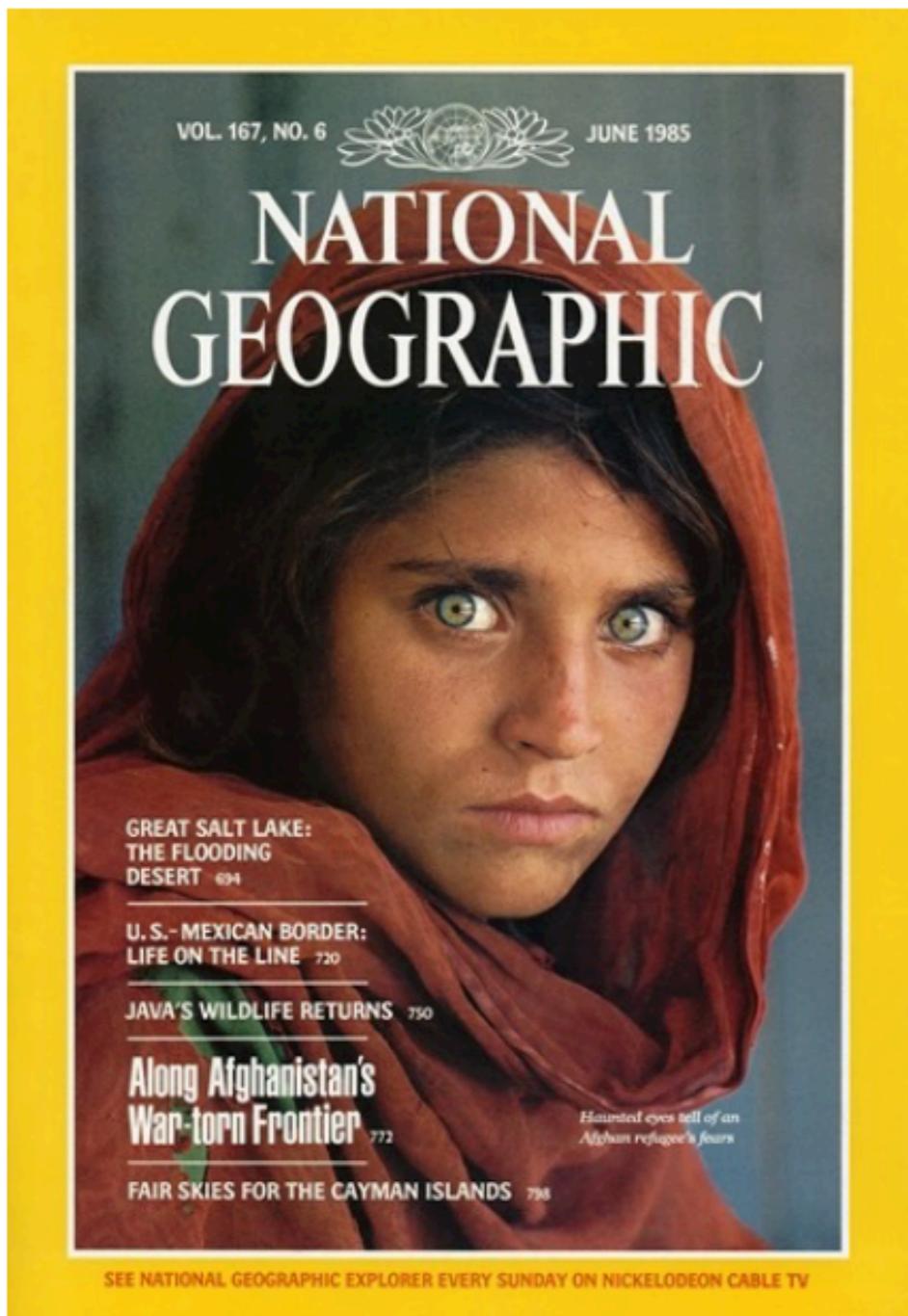
Applications: face recognition

Face ID

Your face is your secure password.



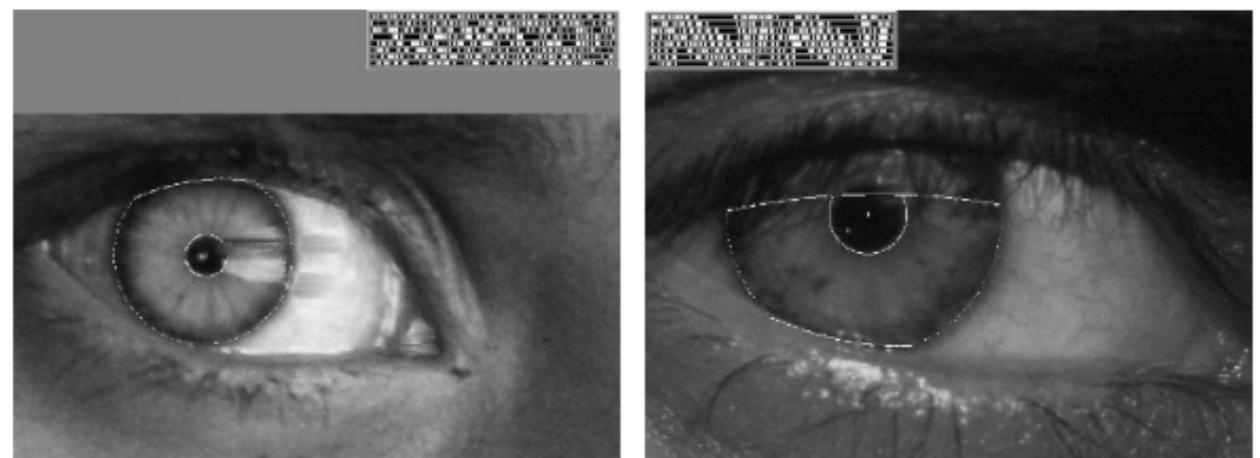
Applications: biometrics



Who is she?



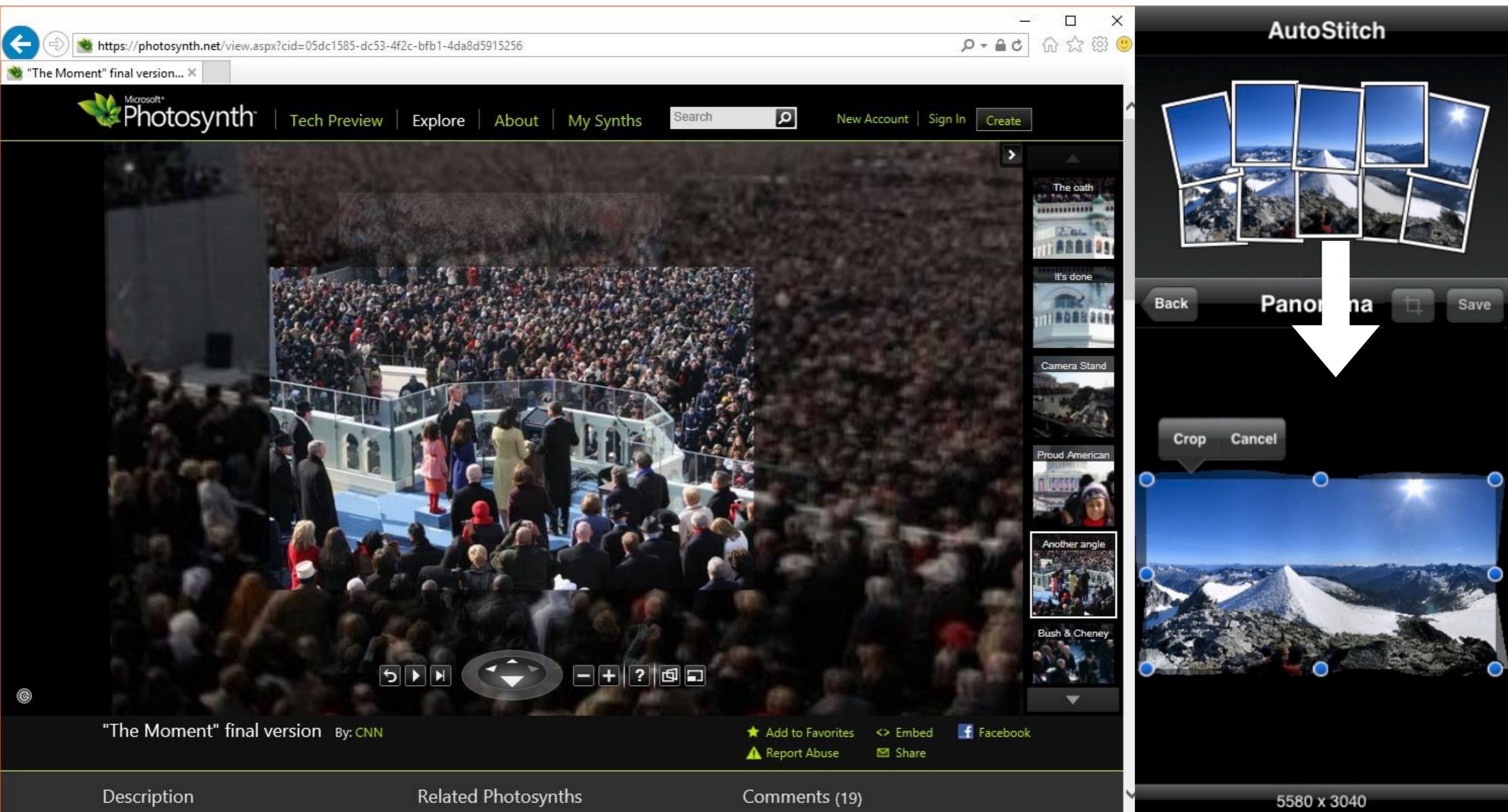
"How the Afghan Girl was Identified by Her Iris Patterns" [[read the story](#)]



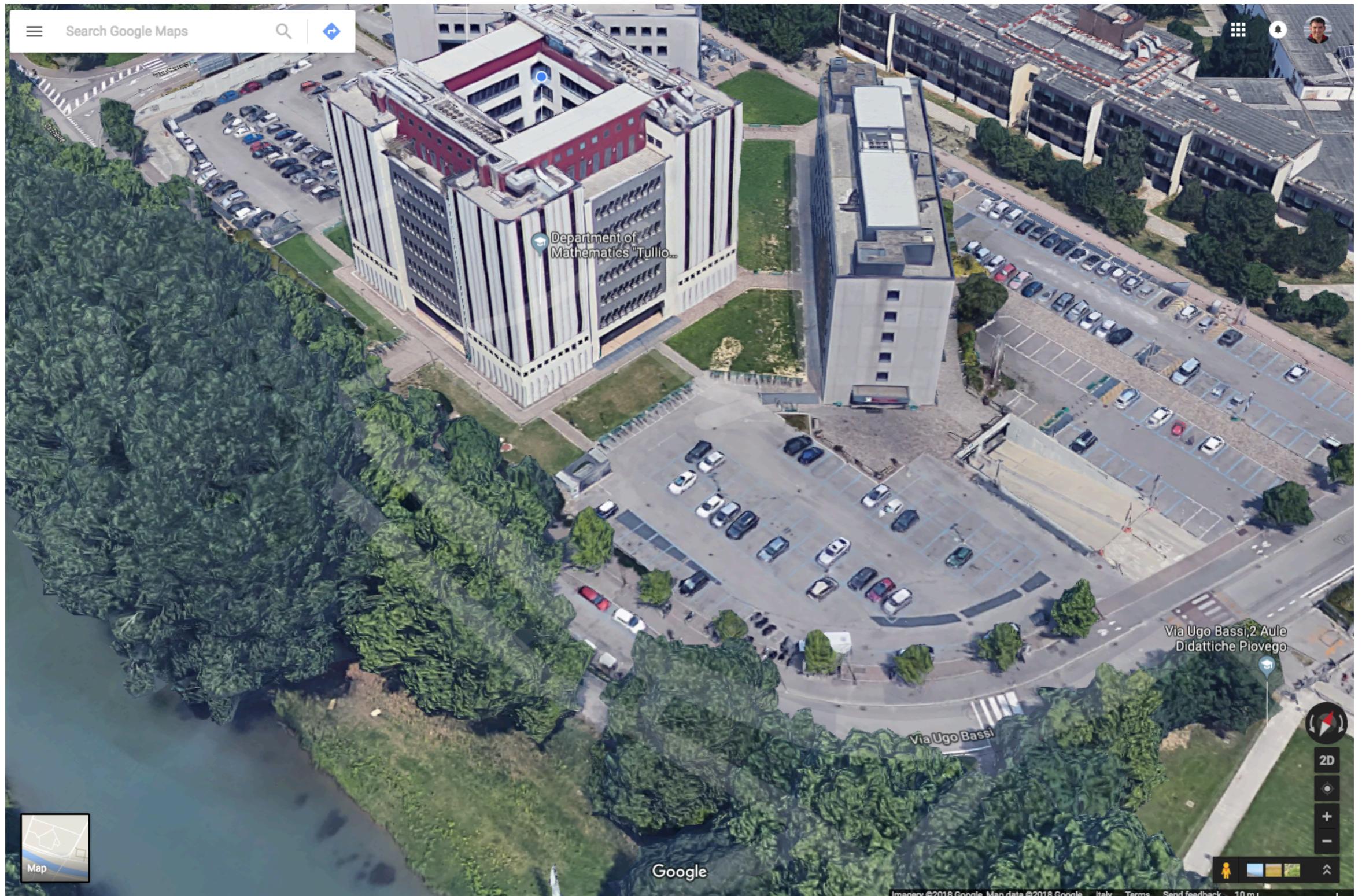
Applications: motion capture



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

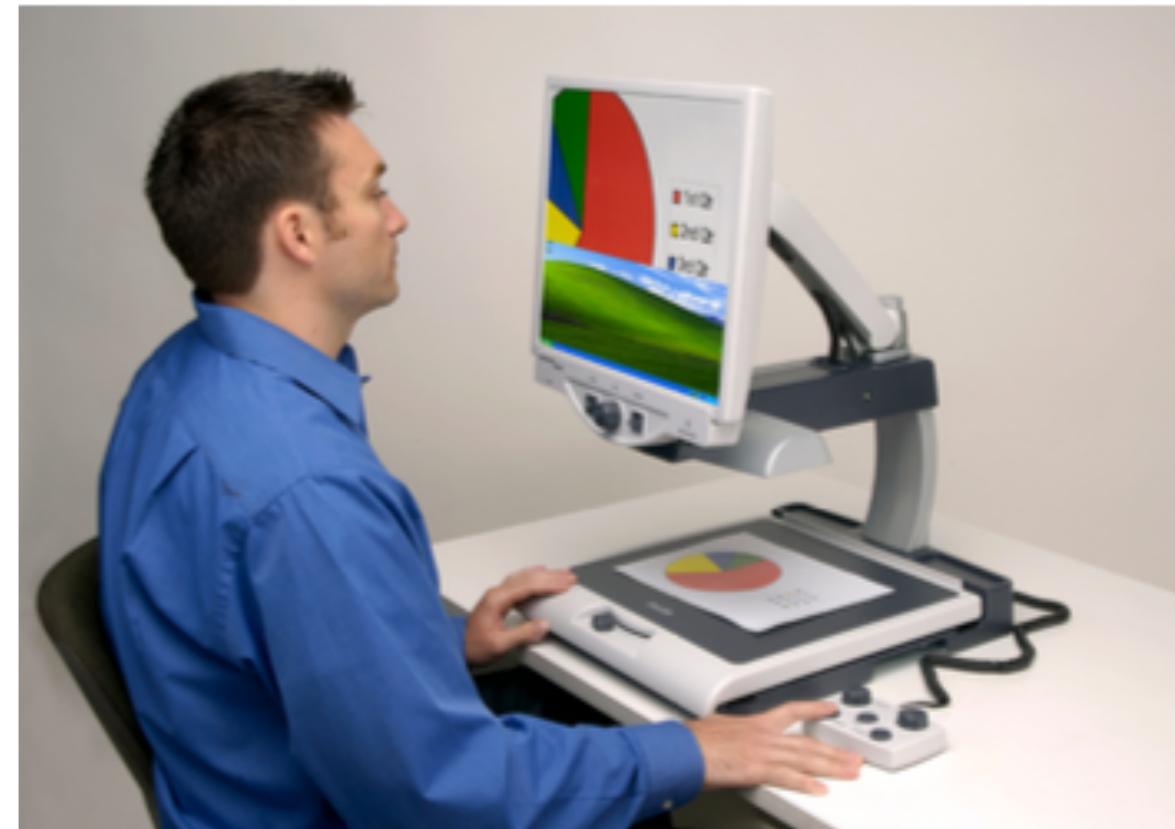
For more info read “computer vision on Mars”, Matthies et al.

Applications: interaction

Motion sensing for videogames
(Microsoft's Kinect)



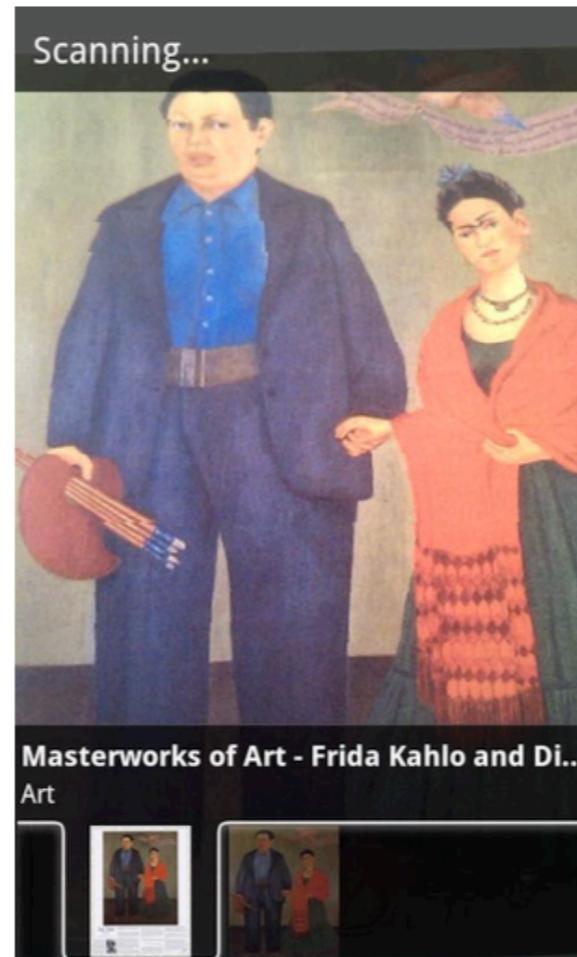
Assistive technologies



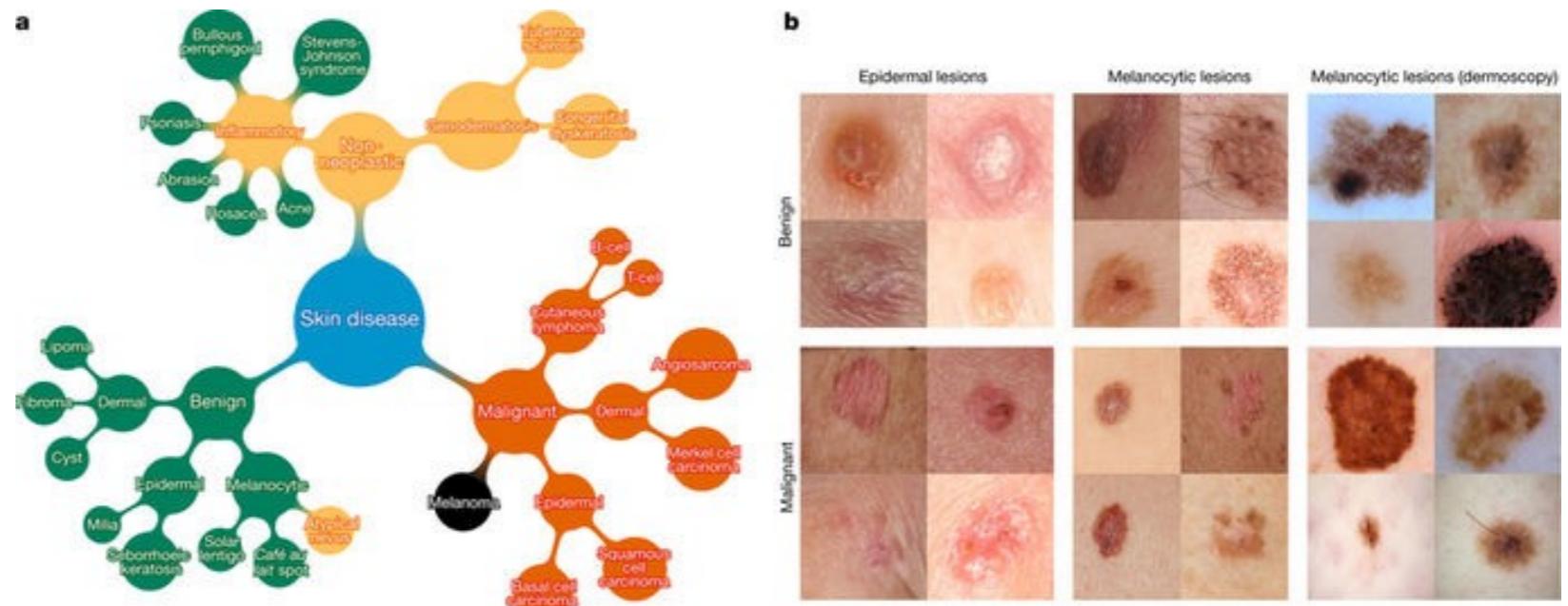
Applications: visual search



Google Goggles



Applications: healthcare

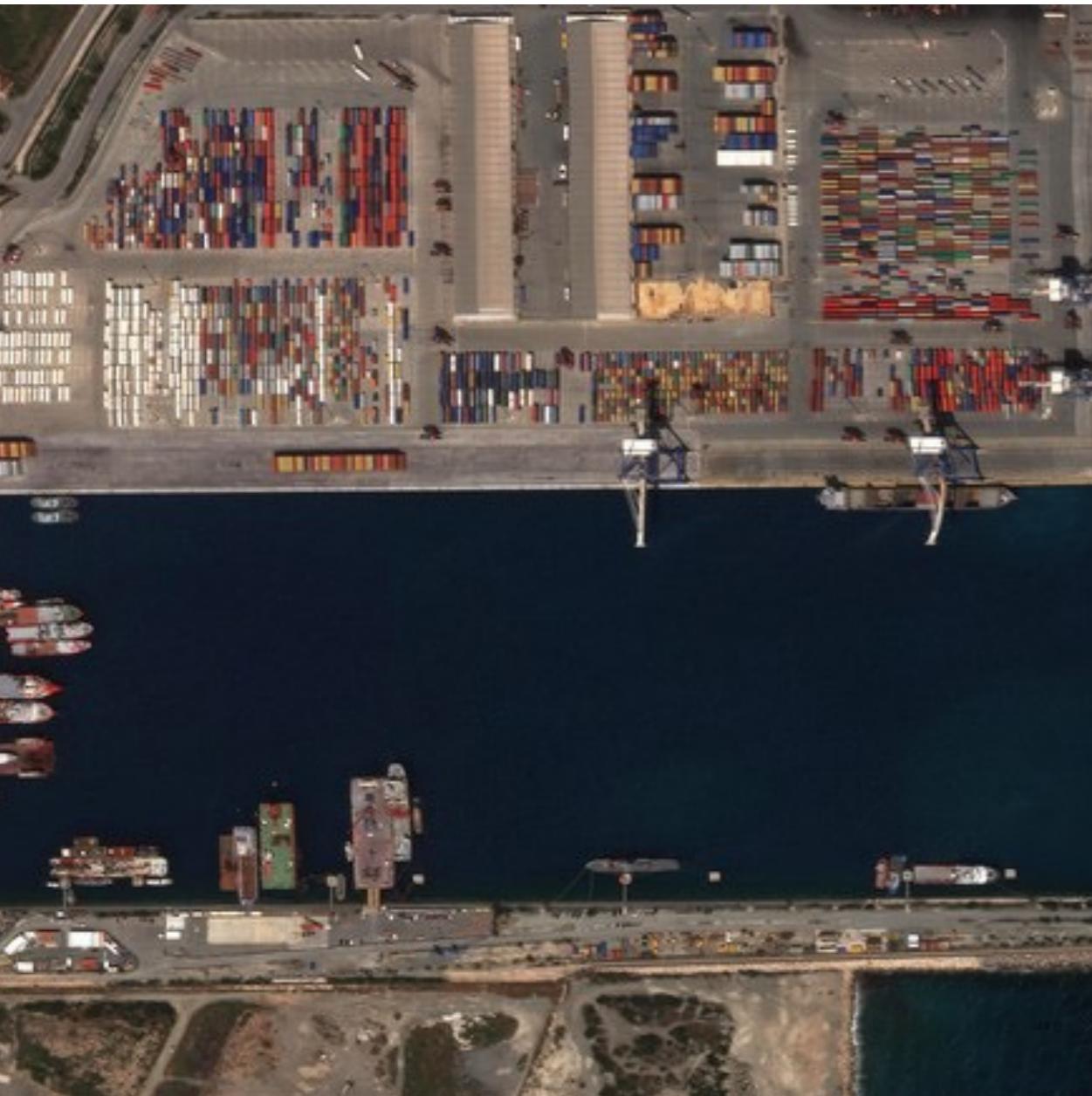


Dermatologist-level Classification of Skin Cancer with Deep Neural Networks

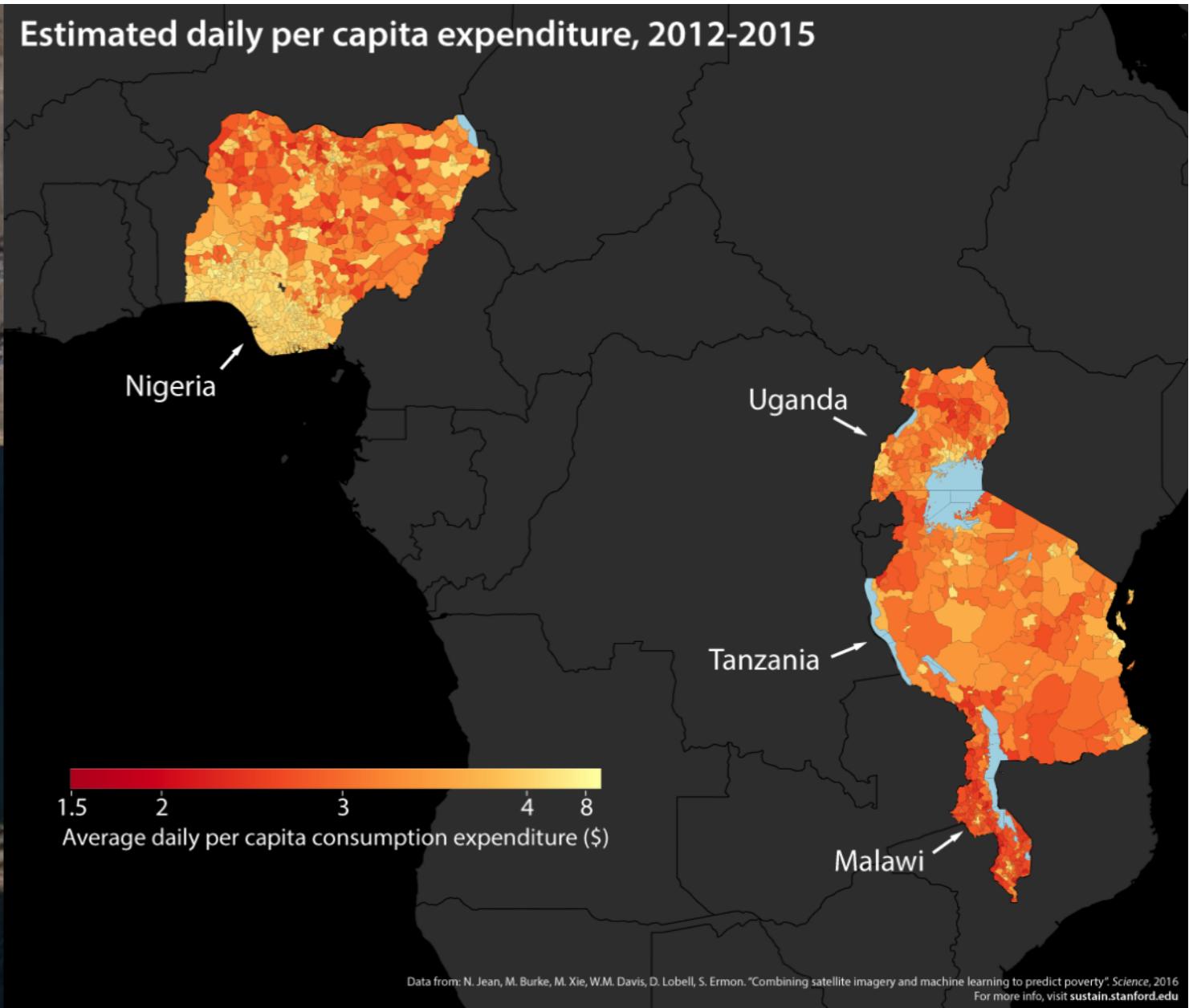
[more info: [watch the video](#)]

Applications: analysing satellite images

Security and defense



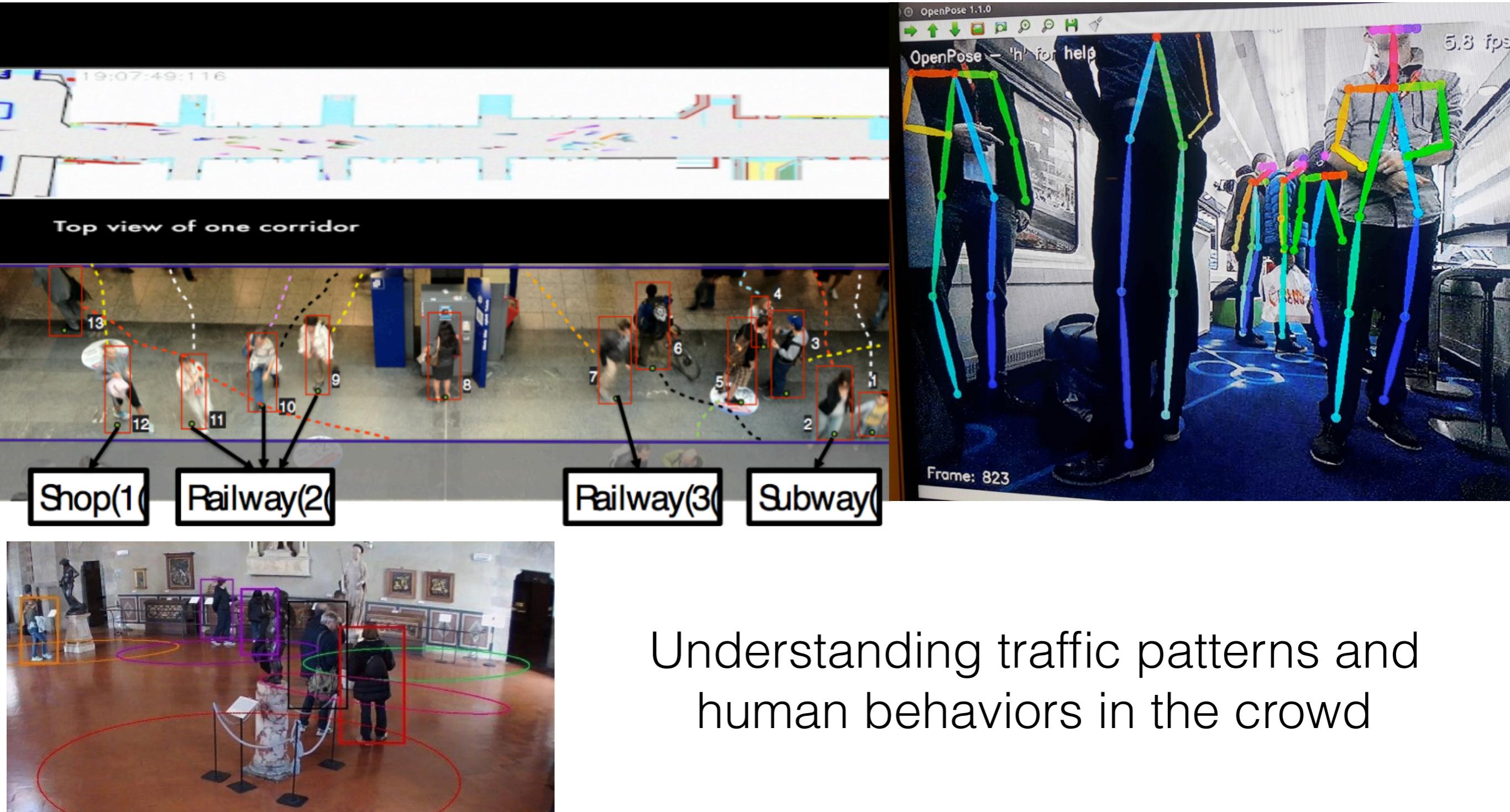
Environmental challenges



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

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

The diagram illustrates a car's multi-camera system. It shows a silver car from a rear three-quarter perspective. Three yellow dashed arcs indicate the field of view for different cameras: a 'rear looking camera' at the back, a 'forward looking camera' at the front, and a 'side looking camera' on the side. The text 'Our Vision. Your Safety.' is centered above the car.

►► manufacturer products consumer products ◀◀

Our Vision. Your Safety.

rear looking camera

forward looking camera

side looking camera

> EyeQ Vision on a Chip

A photograph of the EyeQ chip, which is a green square package with a black lid featuring the Mobileye logo and the text 'ST EyeQ1'. A small blue link button below it says '> read more'.

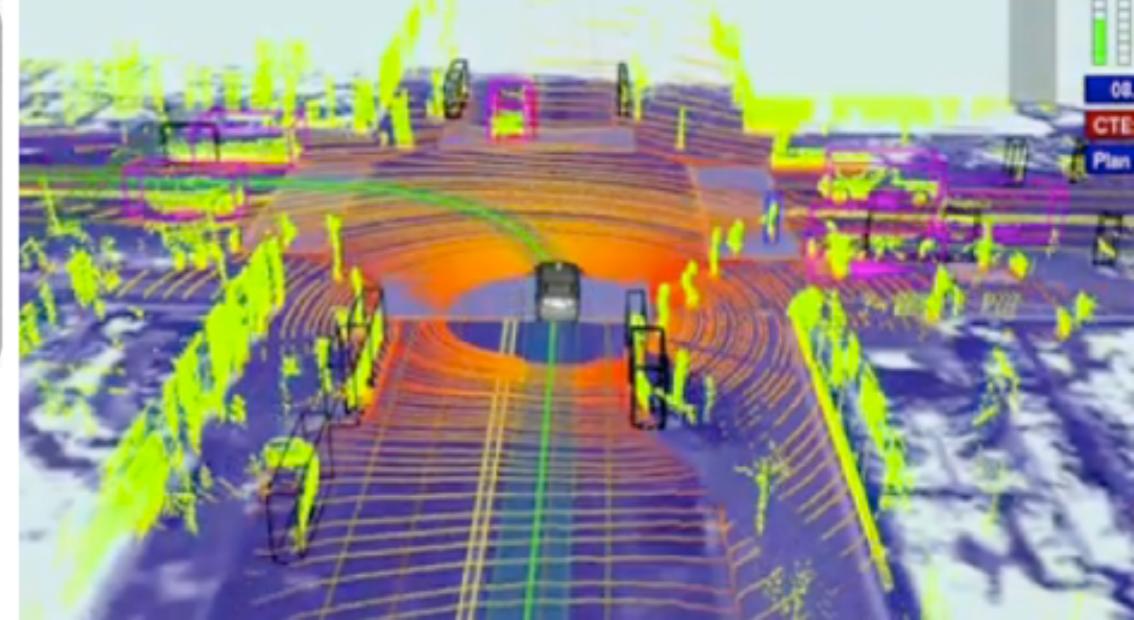
> Vision Applications

An illustration of a woman walking across a crosswalk. A red rectangular box highlights her, and a smaller red box highlights a car in the background. A blue link button below it says '> read more'.

Road, Vehicle, Pedestrian Protection and more

> AWS Advance Warning System

A photograph of a circular AWS display unit mounted on a stand. It shows a small orange car icon and the number '0.8'. A blue link button below it says '> read more'.



- 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 underlined), Solutions, Launcher, Pricing, Security, Customers, Documentation, Support, Partners, TRY IT FREE, and CONTACT SALES. The main content area has several sections:

- Powerful Image Analysis:** Describes the API's ability to understand image content using machine learning models. It includes a diagram of a sailboat and sun with text overlays "SAILBOAT" and "SUN".
- Try the API:** A section with a placeholder image and a button to "Drag image file here or Browse from your computer".
- Insight From Your Images:** Shows a thumbnail of a person and a list of detected objects: CARS, FLOWERS, RABBITS, MOUNTAINS, and BIRDS.
- Detect Inappropriate Content:** Powered by Google SafeSearch, it helps moderate content from crowd-sourced images. It includes a small icon of a person with a red 'X' and a green checkmark.
- Power of the Web:** Describes using Google Image Search to find topical entities like celebrities, logos, or news events. It includes a thumbnail of a person and a logo.
- Extract Text:** Describes Optical Character Recognition (OCR) for detecting text in images. It includes a thumbnail of a document page.

What are we going to learn?

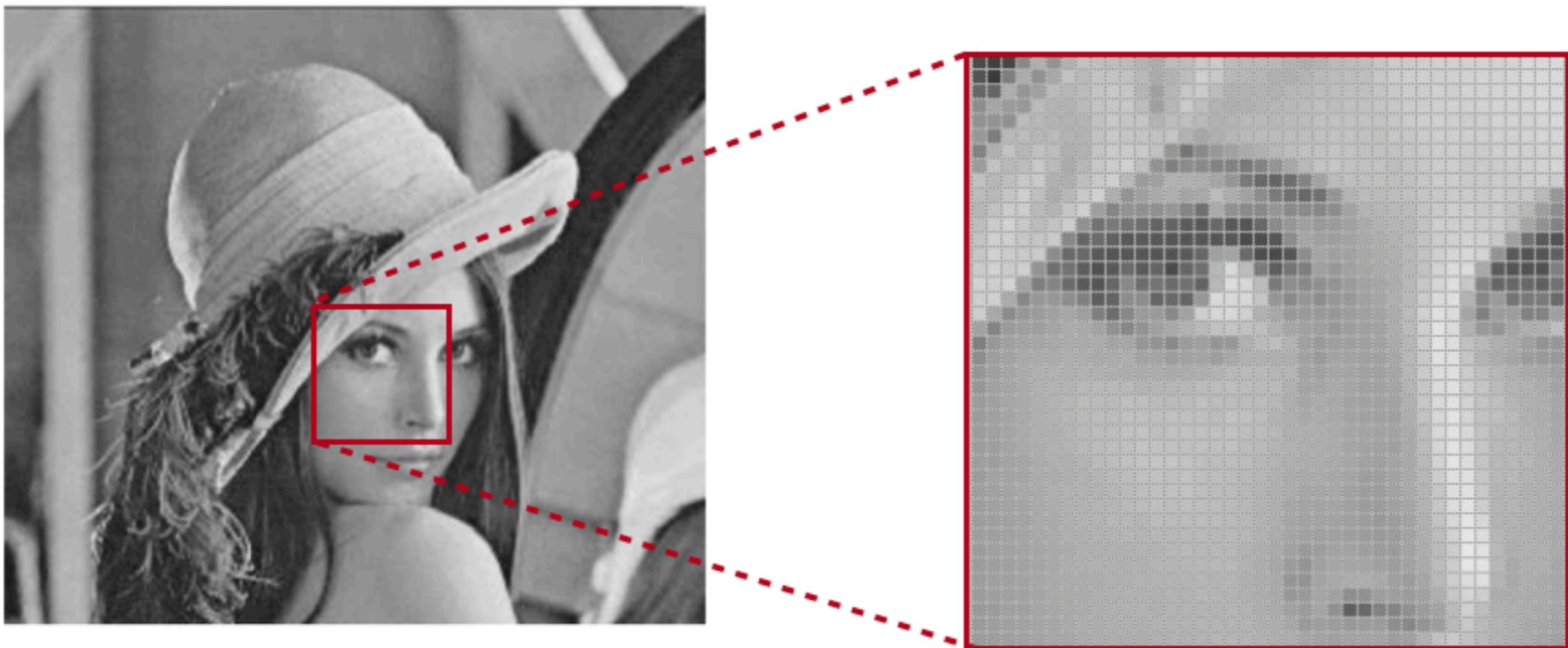
- Focus on high-level visual recognition tasks
- Course (2nd half) outline:
 - ▶ Foundations: images, convolution and filtering
 - ▶ Foundations: feature detectors and descriptors
 - ▶ Intro to visual recognition, classification, detection
 - ▶ Representation learning in vision, convolutional neural networks, *vision and language (multi-modal scenario)*
 - ▶ Hands-on visual recognition: features, BoW, CNN, cognitive vision services

Course schedule

#week	Date	Lecture	Hours	Note	Instructor
W1	lunedì 25 febbraio 2019	Introduction to the course (1st half)	2		A. Sperduti
W1	martedì 26 febbraio 2019	Intro to ML basics - part 1	2		A. Sperduti
W2	lunedì 4 marzo 2019	Intro to ML basics - part 2	2		A. Sperduti
W2	martedì 5 marzo 2019	Intro to ML - part 3	2		A. Sperduti
W3	lunedì 11 marzo 2019	ANN - part 1	2		A. Sperduti
L1	W3 martedì 12 marzo 2019	Lab GCloud	2		L. Ballan
	W4 lunedì 18 marzo 2019	ANN - part 2	2		A. Sperduti
	W4 martedì 19 marzo 2019	ANN - part 3	2		A. Sperduti
W5	lunedì 25 marzo 2019	Intro to Deep Learning	2		A. Sperduti
W5	martedì 26 marzo 2019	TF and Deep Learning Tools	2		A. Sperduti
W6	lunedì 1 aprile 2019	Intro to learning in sequential domains	2		A. Sperduti
L2	W6 martedì 2 aprile 2019	Lab2 - TensorFlow	2		A. Sperduti
	W7 lunedì 8 aprile 2019	Computer vision (2nd half) overview	2		L. Ballan
	W7 martedì 9 aprile 2019	Foundations: images, convolution, filters	2		L. Ballan
W8	lunedì 15 aprile 2019	Foundations: feature detectors, descriptors	2		L. Ballan
L3	W8 martedì 16 aprile 2019	Lab3 - Image filtering	2		L. Ballan
	W9 lunedì 22 aprile 2019				<i>Easter break</i>
	W9 martedì 23 aprile 2019				
W10	lunedì 29 aprile 2019	Visual recognition, classification, detection	2		L. Ballan
L4	W10 martedì 30 aprile 2019	Lab4	2		L. Ballan
	:				

Coming up

- **Tuesday, April 9** (next lecture):
 - ▶ Foundations: images, pixels, convolution, filtering



Seminar series on Big Data & AI

The Law of Big Data

Project
Team
Guest Speakers
Syllabus
News

Search

Books



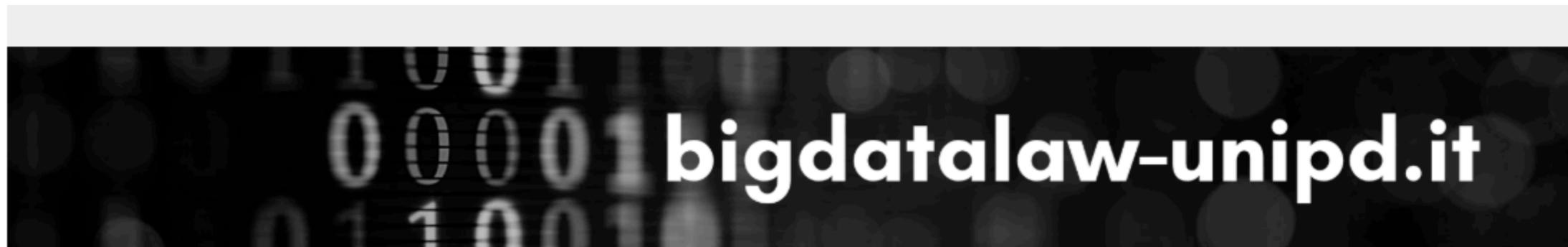
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DiPIC

Dipartimento di Diritto Pubblico
Internazionale e Comunitario



DIPARTIMENTO DI FISICA E
ASTRONOMIA "GALILEO GALILEI"



Project syllabus



12.10 2018 Artificial Intelligence and Big Data



09.11 2018 Under the Hood of Big Data and A.I.



07.12 2018 The Unconstitutional Algorithm



28.02 2019 Big Data Governance Beyond Personal Data



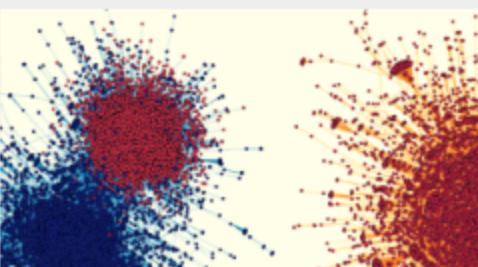
22.03 2019 The Era of Mass Data Litigation



08.05 2019 The Rise of Machines and the Disruption of Law



23.05 2019 Just Machine Learning in Unjust World?



06.06 2019 5 Reasons Why Social Networks Make Us Vulnerable to Misinformation

Contact

- **Office:** Torre Archimede 3CD, room 320
- **Homepage:** www.math.unipd.it/~ballan
- **My research group (VIMP):** vimp.math.unipd.it
- **E-mail:** lamberto.ballan@unipd.it
- **Phone:** +39 049 827 1332
- **Office hours (ricevimento):** Monday 10:30-12:30