Computer Vision

P. Fua
(2-3 weeks taught by M. Salzmann)
IC-CVLab
EPFL

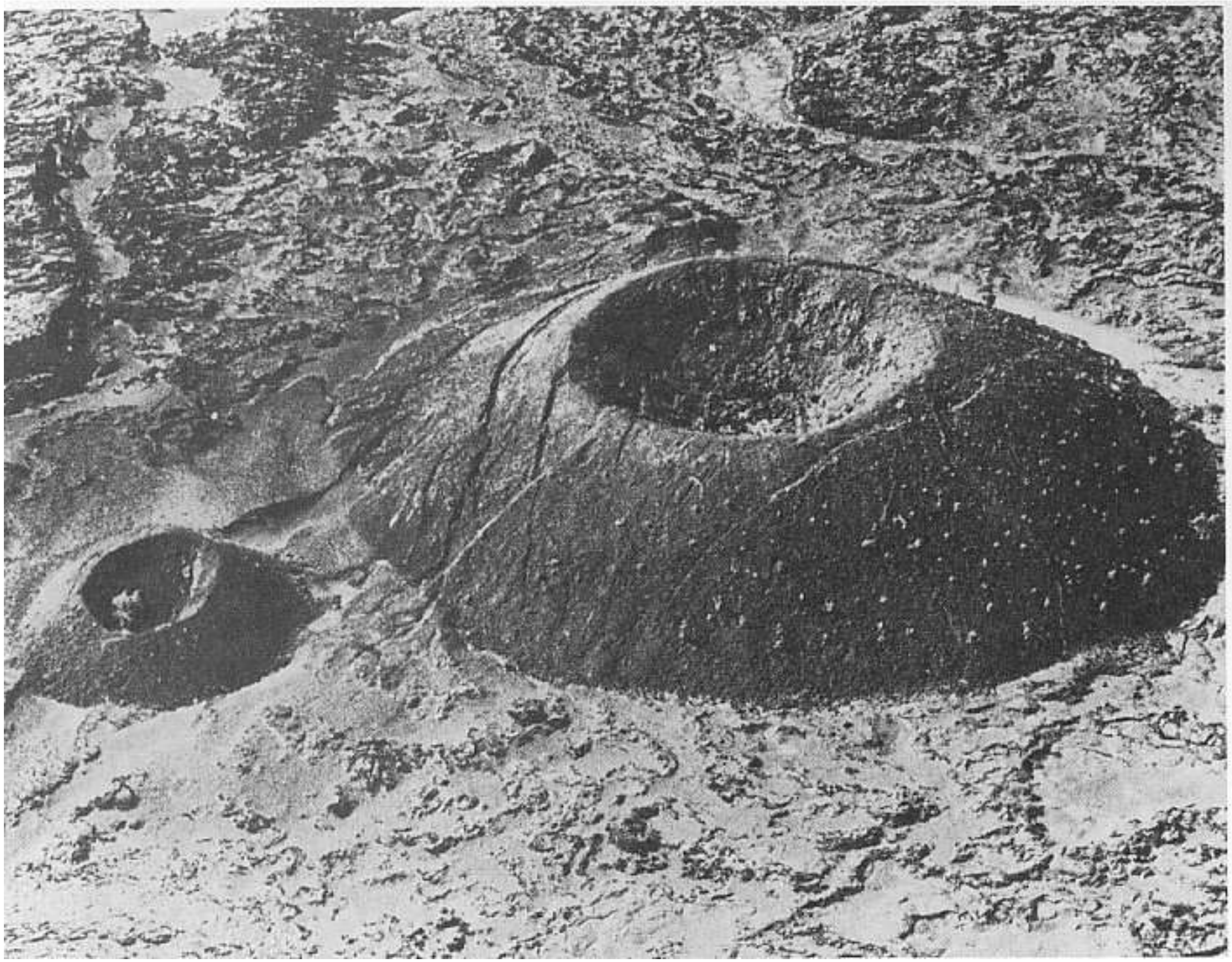
Computer Vision

Goal: Inferring the properties of the world from one or more images

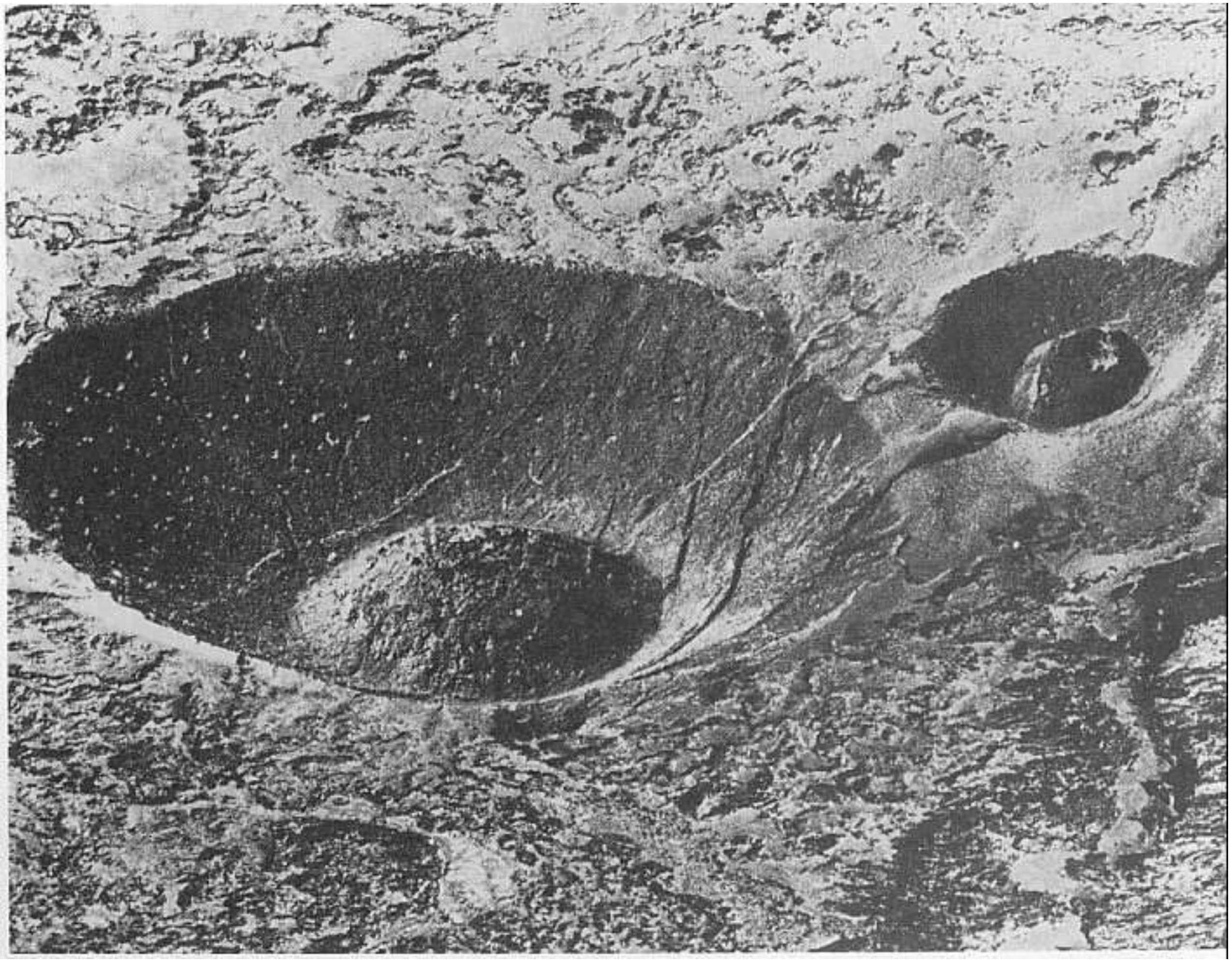
- Photographs
- Video Sequences
- Medical images
- Microscopy data

→ **Image Understanding**

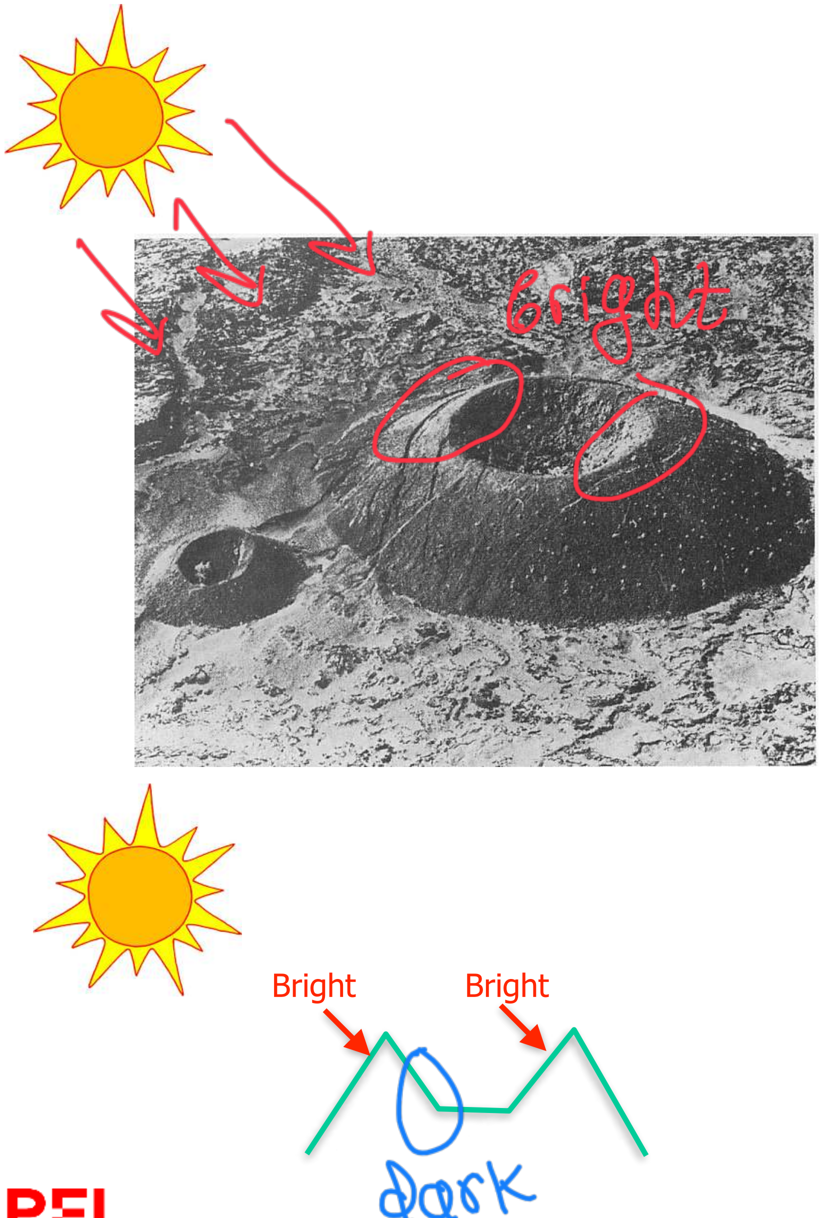
What do You See?

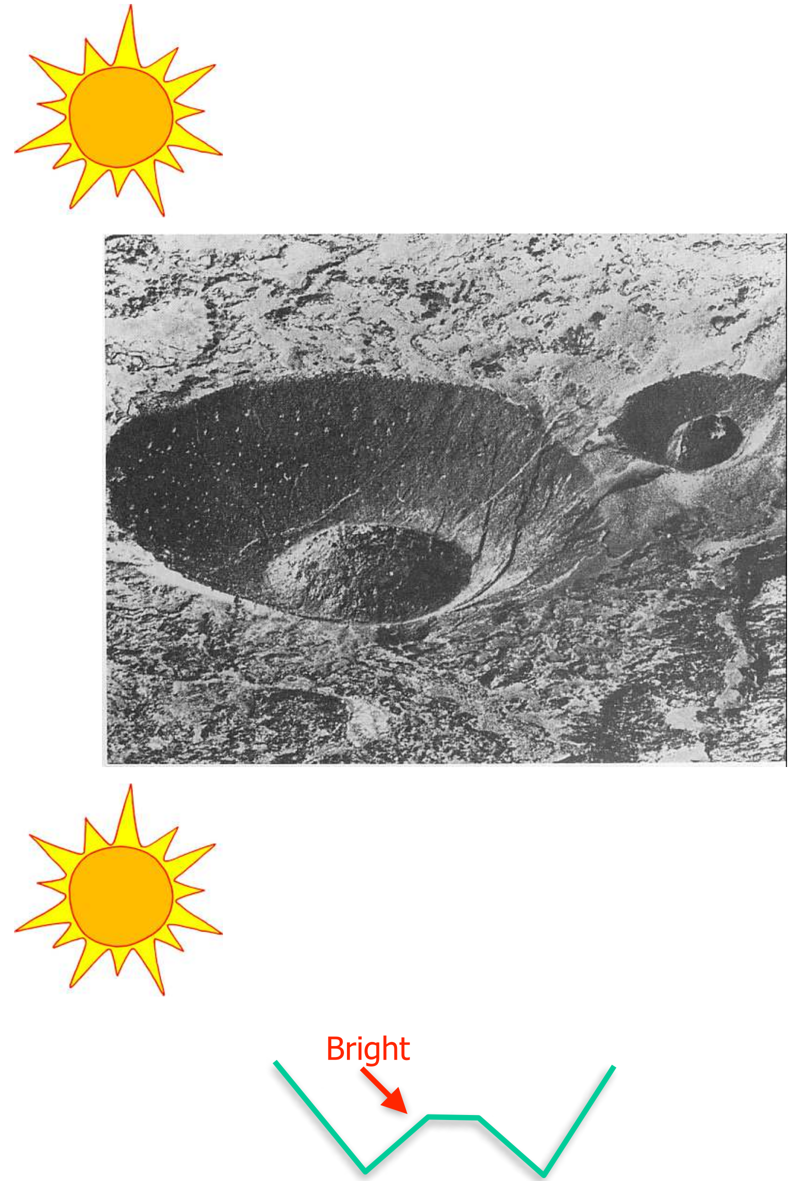


And Now?



Potential Interpretation





Shape from Contours

Shading

- Later in the class, we will formalize this in terms of a set of differential equations.
- This is probably not what we do in our heads when we look at images.

A Powerful Mechanism

differently represented

136	134	161	159	163	168	171	173	173	171	166	159	157	155
152	145	136	130	151	149	151	154	158	161	163	163	159	151
145	149	149	145	140	133	145	143	145	145	145	146	148	148
148	143	141	145	145	145	141	136	136	135	135	136	135	133
131	131	129	129	133	136	140	142	142	138	130	128	126	120
115	111	108	106	106	110	120	130	137	142	144	141	129	123
117	109	098	094	094	094	100	110	125	136	141	147	147	145
136	124	116	105	096	096	100	107	116	131	141	147	150	152
152	152	137	124	113	108	105	108	117	129	139	150	157	159
159	157	157	159	135	121	120	120	121	127	136	147	158	163
165	165	163	163	163	166	136	131	135	138	140	145	154	163
166	168	170	168	166	168	170	173	145	143	147	148	152	159
168	173	173	175	173	171	170	173	177	178	151	151	153	156
161	170	176	177	177	179	176	174	174	176	177	179	155	157
161	162	168	176	180	180	180	182	180	175	175	178	180	180

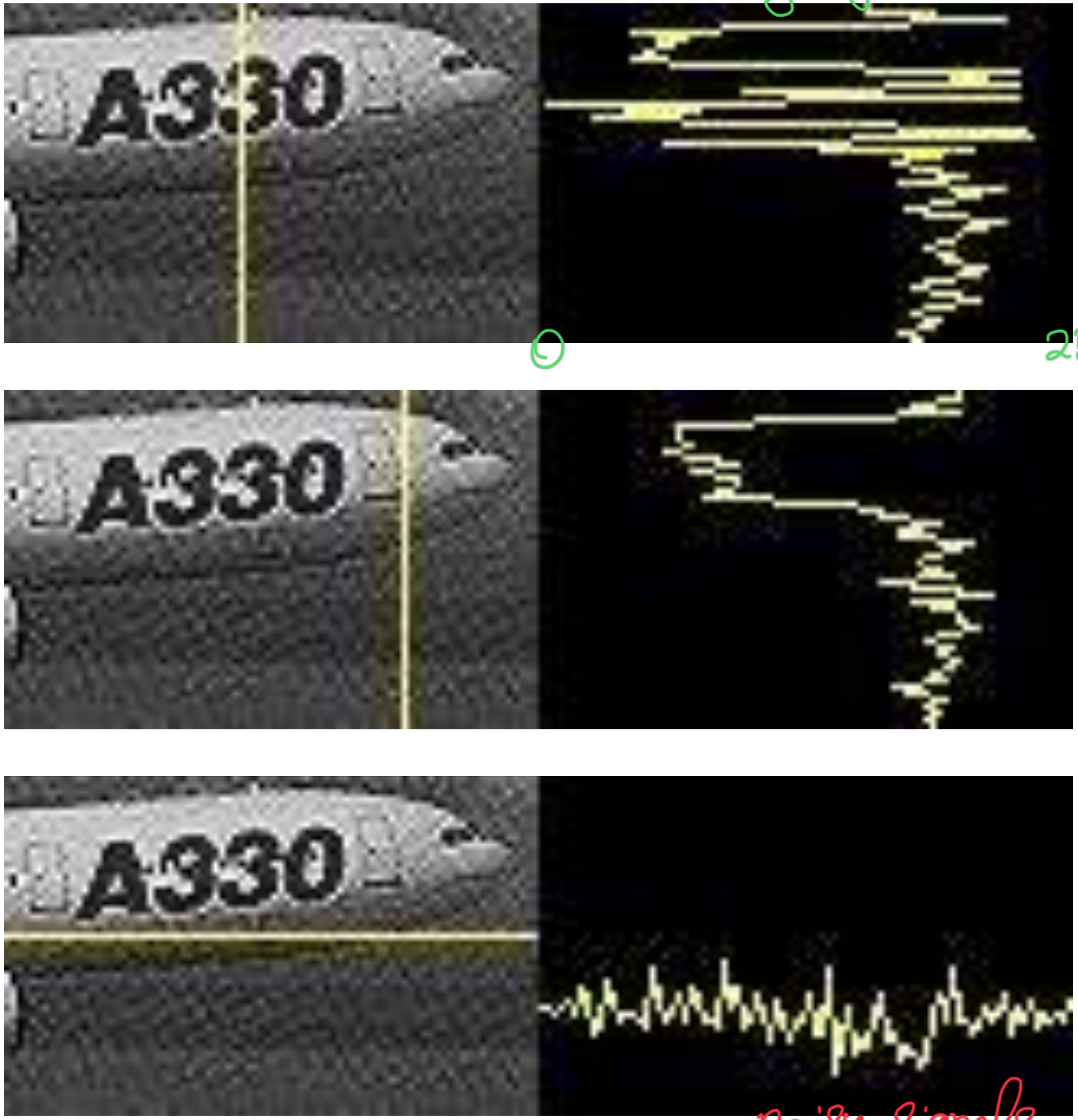
binarized image

255 → white
0 → black

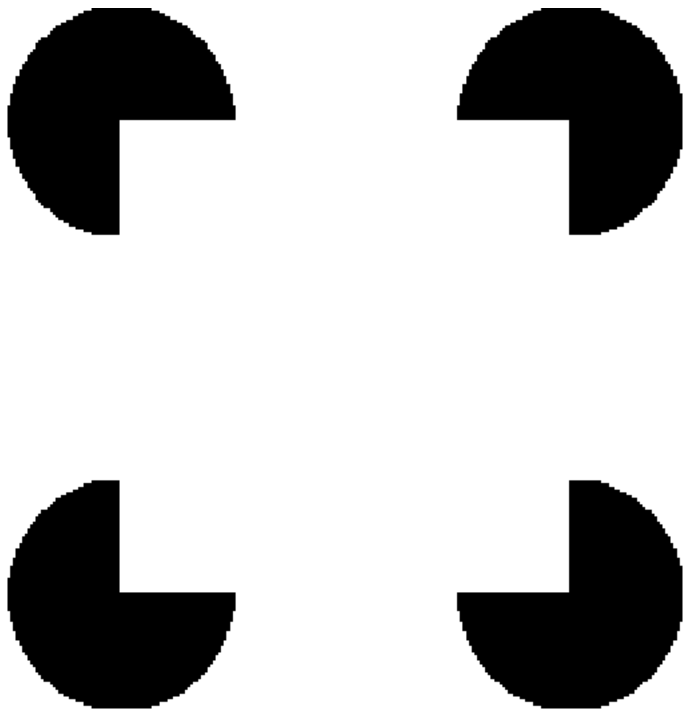
8-bit image
(0,255)

A Powerful Mechanism

gray levels along that line



Illusory Contours



Kanizsa's Triangle

Closure of Good Form Hypothesis: Illusory contours represent an example of the closure of good form (Osgood 1953), (Pastore 1971), (Kanizsa 1976, 1979).

Figural-Cue Hypothesis: Illusory contours are responses to partial figural cues in the same way that meaning is abstracted from simple outline drawings or cartoons (Gregory 1972), (Piggins 1975), (Rock and Anson 1979).

Cues-to-Depth Hypothesis: Illusory contours are produced by the monocular depth cue of interposition to perceive a plane in depth (Coren 1972).

Organizational-Attentional Effects Hypothesis: Emphasizes that illusory contours are not totally stimulus-bound (Bradley and Dumais 1975), (Bradley and Petry 1977), (Kennedy 1976).

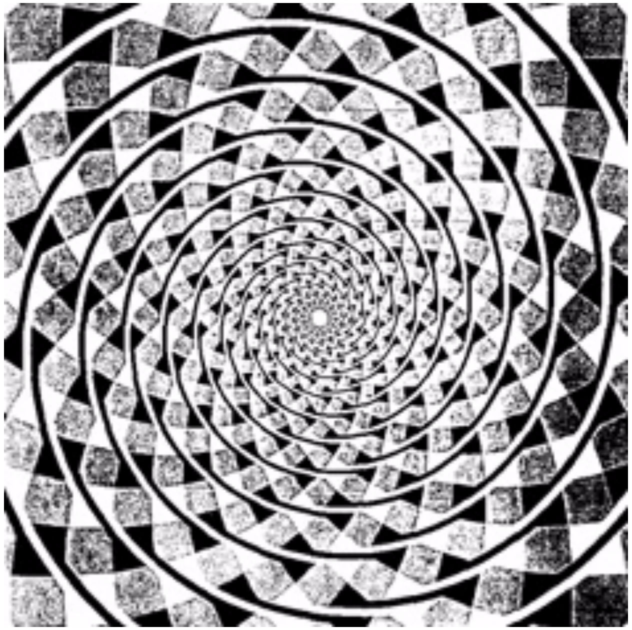
Retinal-Smearing Hypothesis: The edge effects created by the inducing areas are smeared over the retina during the course of normal eye movements to produce illusory contours (Kennedy and Chattaway 1975). This theory has been found to be untenable.

Brightness-Contrast Hypothesis: Illusory contour formation is secondary to the perception of brightness differences between the illusory figure and its background (Brigner and Gallagher 1974), (Frisby and Clatworthy 1975), (Day and Jory 1978,1979,1980).

Feature Analyzers Hypothesis: Illusory contours result from the partial triggering of contour-specific neural units by the physically present edge along the inducing areas (Stadler and Dieker 1972), (Smith and Over 1975,1979). Others have suggested neural networks that generate continuous contours (filling-in contours) from discontinuous stimulus (Ullman 1976), (Grossberg and Mingolla 1985).

Spatial-Frequency-Analysis Hypothesis: Existence of a stimulus correlate for illusory contours based on a Fourier analysis of the stimulus display (Ginsburg 1975) .

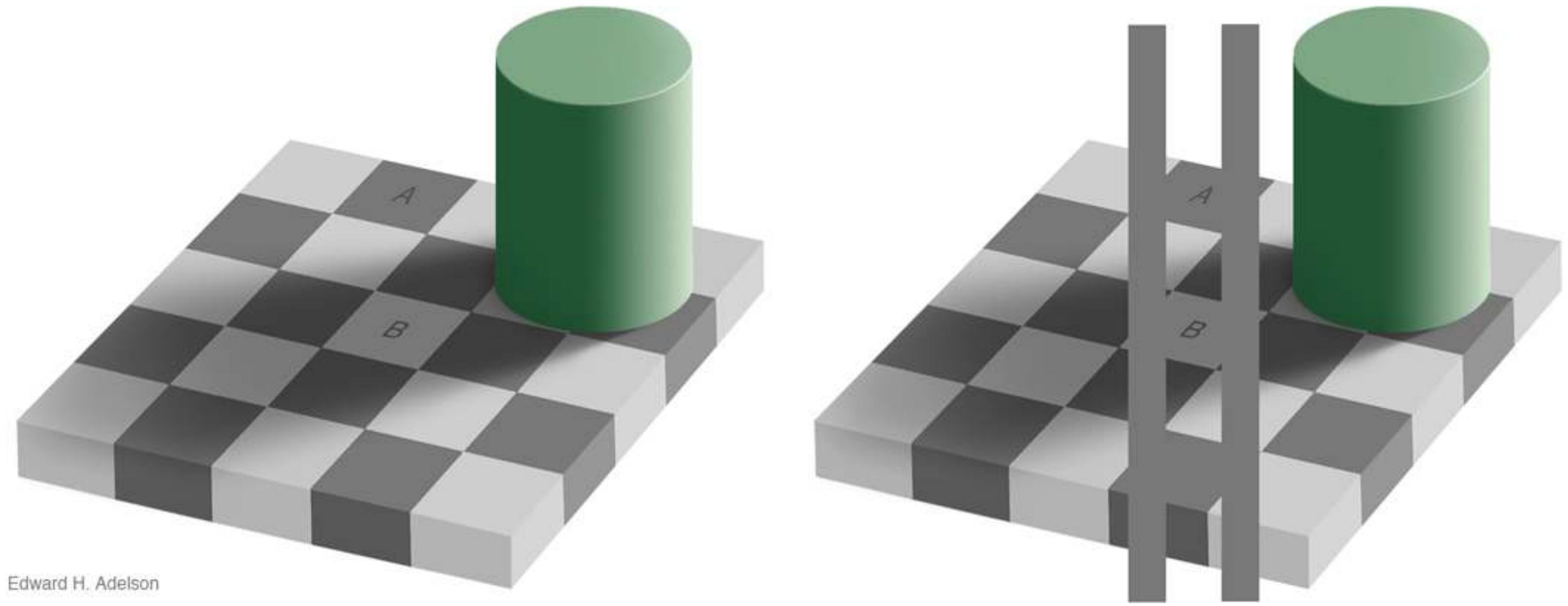
Optical Illusions



Every image is the image of thing merely to him who knows how to read it, and who is enabled by the aid of the image to form an idea of the thing.

Handbook of Physiological Optics
H. von Helmholtz

Photometric Illusion



The human eye measures relative rather than absolute intensity values.

Challenges

Vision involves dealing with:

- Noisy images
- Many-to-one mapping
- Aperture problem

→ Requires:

- Assumptions about the world   Priors
- Statistical and physics-based models
- Training data   Deep Networks

True image understanding seems to require a great deal of thinking. We are not quite there yet.

Applications

Duplicate on what brain can do

Cartography:
- Maps from aerial and satellite images

Imitation

Robotics:
- Autonomous navigation
- Visual servoing

Industrial inspection
- Quality control

Security applications
- Access control
- Surveillance

Databases
- Retrieval and Annotation

Medical Imagery
- Microscopy

Cartography on Mars

reconstruct 3D images ← views from above

Satellite view
(real)

Stereo
vision

Ground view
(synthetic)

Mars has been mapped

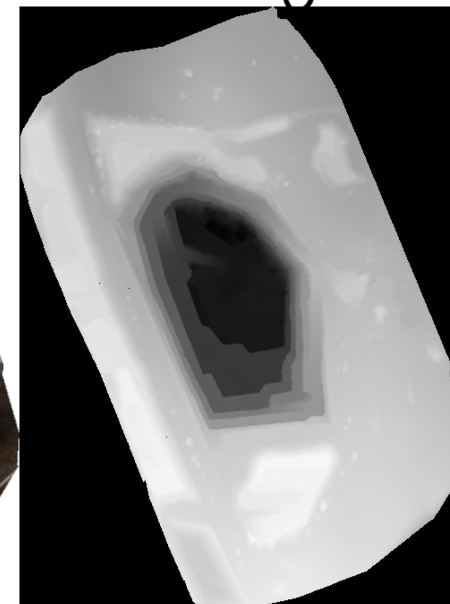
Cartography on Earth

visualize 3D through stereo vision

Virtual Matterhorn

Mining Site

depth image



Built using stereo vision

- Fully automated.
- Accurate.
- Inexpensive.

GCP statistics

	X[m]	Y[m]	Z[m]
RMS	0.086	0.074	0.053
σ	0.040	0.061	0.053

Applications

Cartography:
- Maps from aerial and satellite images

Robotics:
- Autonomous navigation
- Visual servoing

Industrial inspection
- Quality control

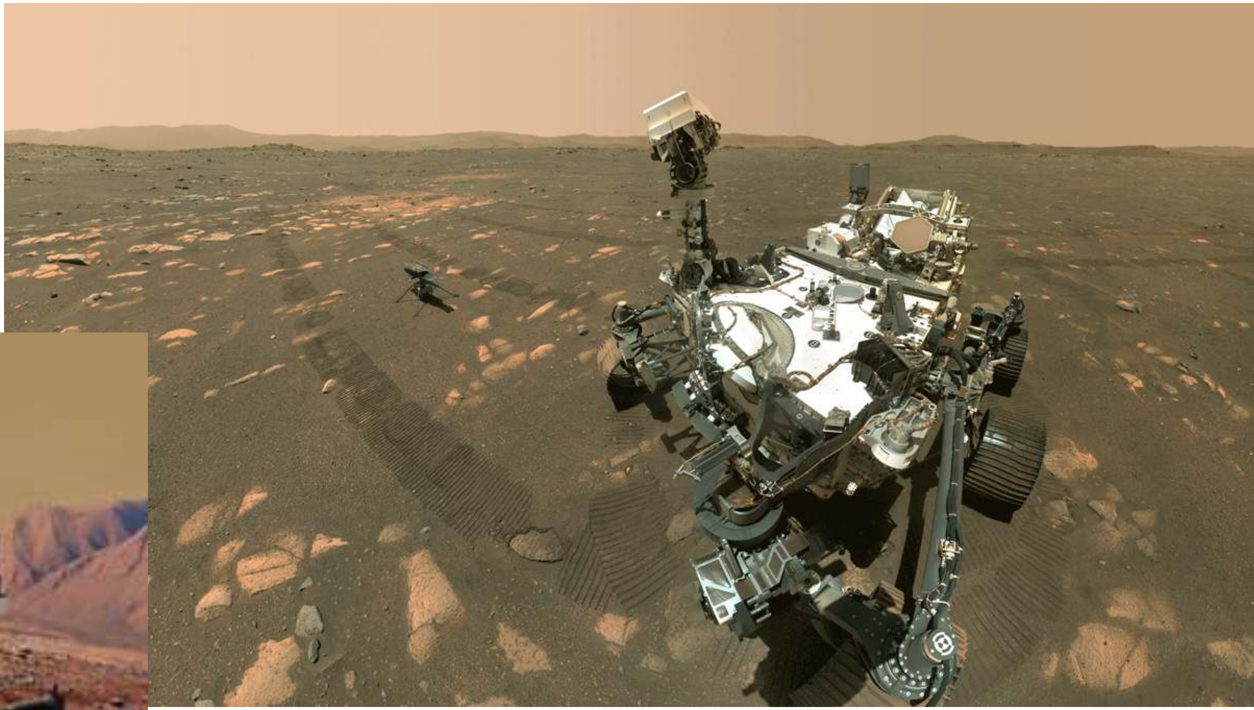
Security applications
- Access control
- Surveillance

Databases
- Retrieval and Annotation

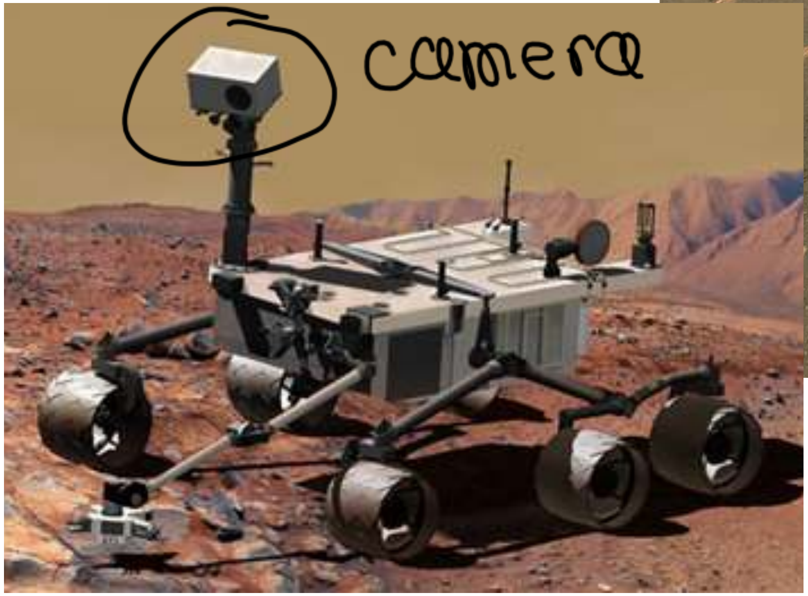
Medical Imagery
- Microscopy

Mars Rovers

Opportunity
Landed 2004

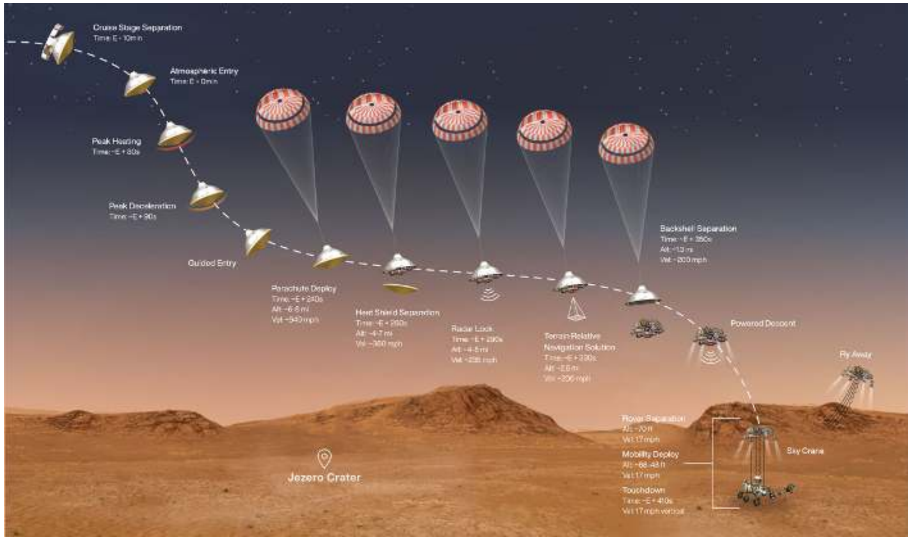


Perseverance
Landed 2021



Curiosity
Landed 2012

Landing System Valid



Andrew Johnson (NASA)

Computer vision will play an unprecedented role in the landing, ensuring that the rover avoids such obstacles as boulder fields, dunes and crater walls in the final seconds of its seven-month journey to Mars.

https://www.ri.cmu.edu/cmu-robotics-alum-leads-development-of-critical-landing-technology-computer-vision-system-will-enable-safe-martian-landing-for-nasas-perseverance-rover/

Earth Rovers

Autonomous vehicle

lidar

1985
DARPA ALV

2007
DARPA Urban Challenge

2014 / 2020
Google Car / Tesla FSD / Waymo Taxis

- Much more computing power.
- More reliable sensors.
- Detailed maps and models of the environment.

duplicating

Applications

Cartography:
- Maps from aerial and satellite images

Robotics:
- Autonomous navigation
- Visual servoing

Industrial inspection
- Quality control

Security applications
- Access control
- Surveillance

Databases
- Retrieval and Annotation

Medical Imagery
- Microscopy

Visual Inspection Of Assembled Devices

Software embedded in the camera to find and read serial numbers

- Accurate localization
- Robustness to Illumination changes
- Generality

APPLICATIONS

Cartography:
- Maps from aerial and satellite images

Robotics:
- Autonomous navigation
- Visual servoing

Industrial inspection
- Quality control

Security applications
- Access control
- Surveillance

Databases
- Retrieval and Annotation

Medical Imagery
- Microscopy

EPFL

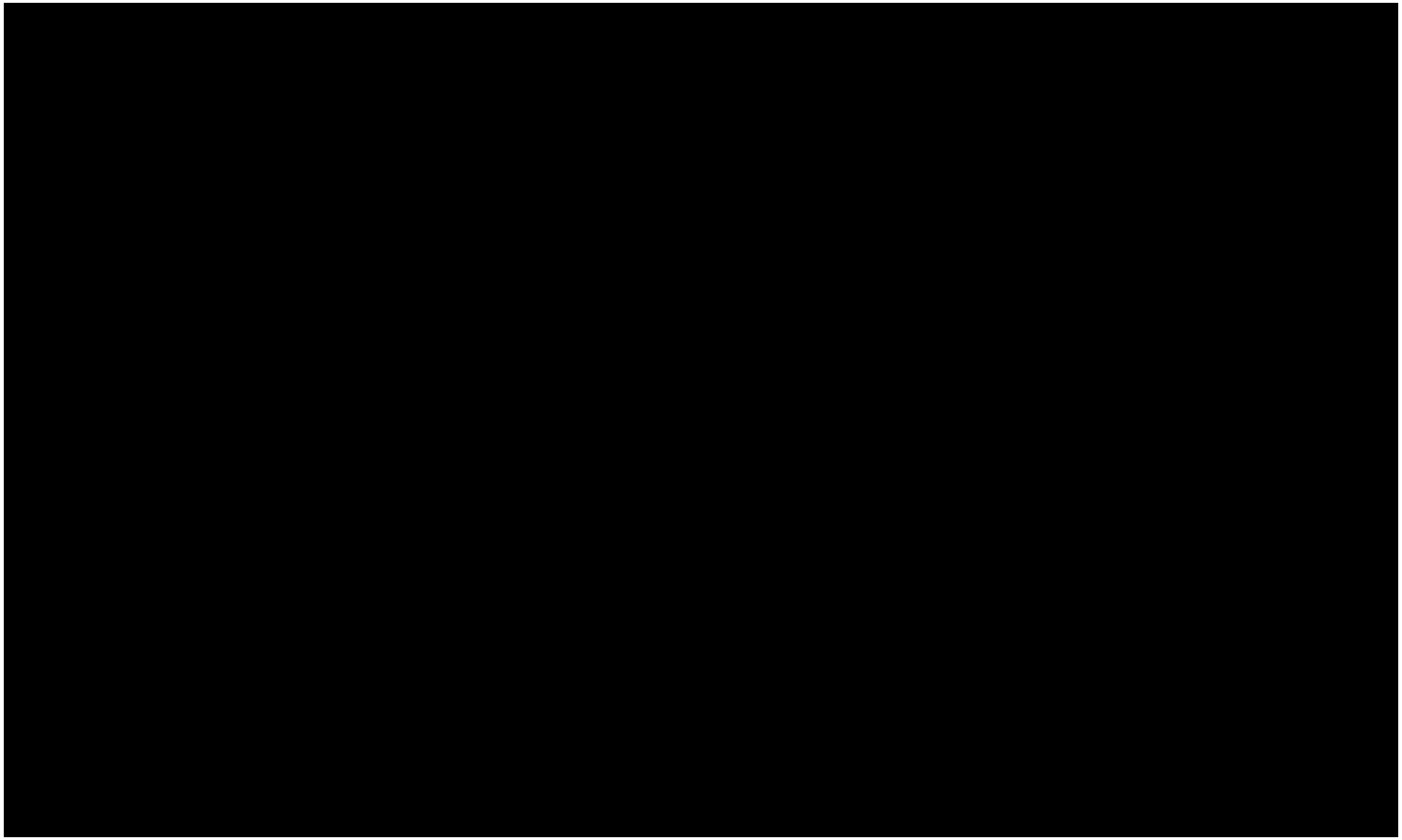
License Plates

http://www.appian-tech.com/products/anpr overview.html

Tracking People

… and the ball → Behavioral analysis.

Tech Transfer



- 2005: First ICCV paper published.
- 2017: System deployed in NBA arenas.
- 2019: Premier League Optical Tracking Provider.

—> It takes time!

https://www.secondspectrum.com/

APPLICATIONS

Cartography:
- Maps from aerial and satellite images

Robotics:
- Autonomous navigation
- Visual servoing

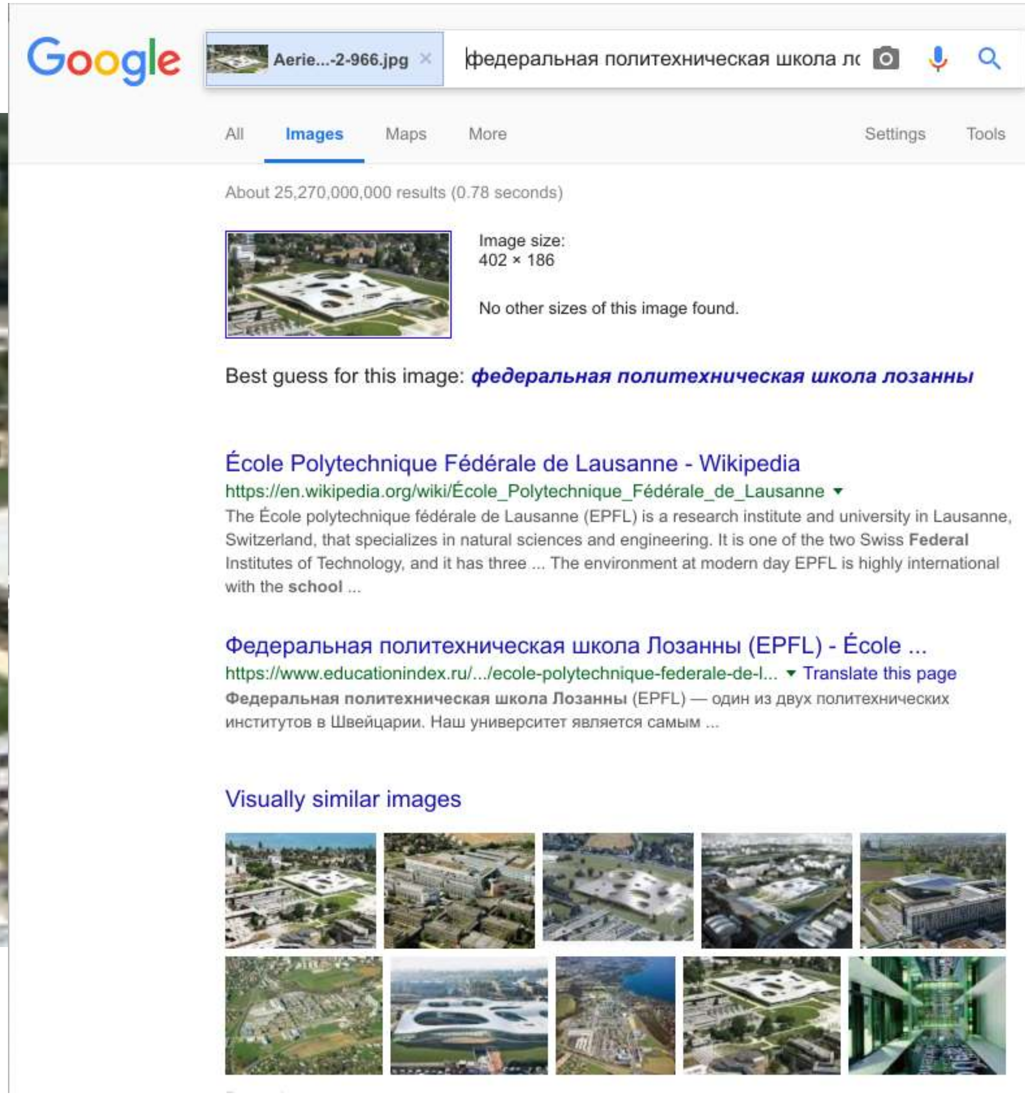
Industrial inspection
- Quality control

Security applications
- Access control
- Surveillance

Databases
- Retrieval and Annotation

Medical Imagery
- Microscopy

Image Retrieval



Identify a picture of EPFL in a large database or on the web **without** words.

Applications

Cartography:
- Maps from aerial and satellite images

Robotics:
- Autonomous navigation
- Visual servoing

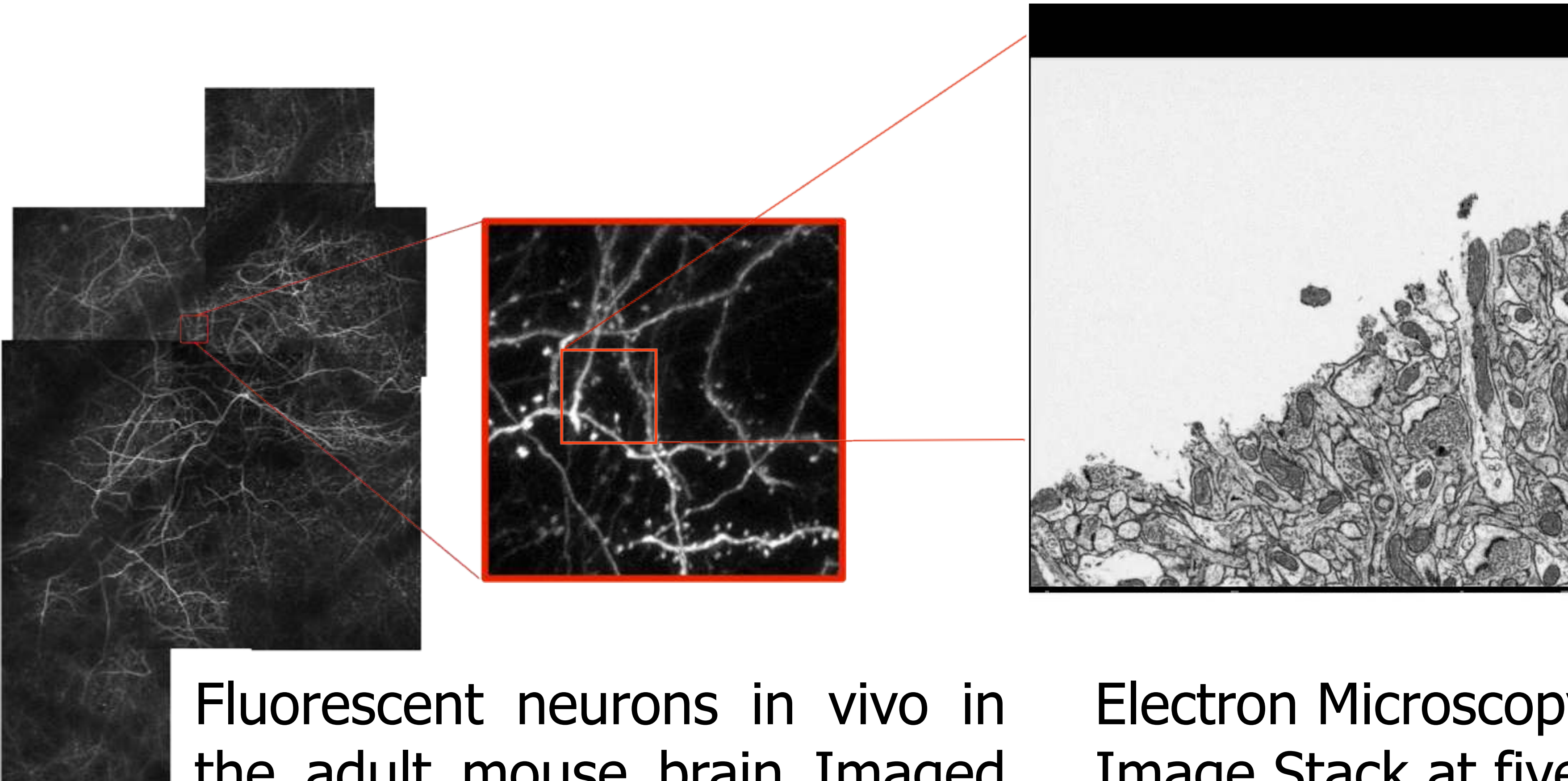
Industrial inspection
- Quality control

Security applications
- Access control
- Surveillance

Databases
- Retrieval and Annotation

Medical Imagery
- Microscopy

Microscopy

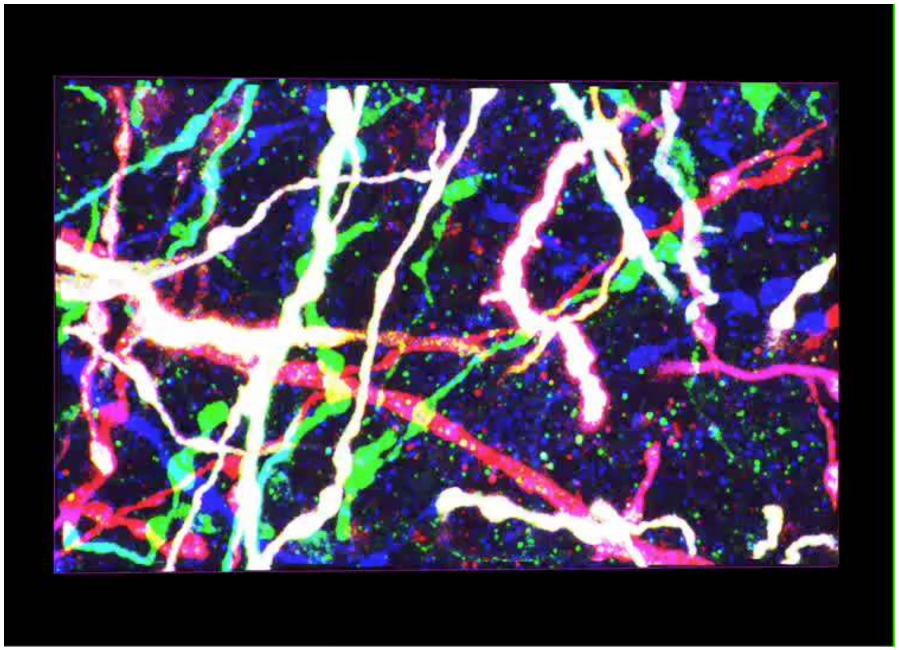


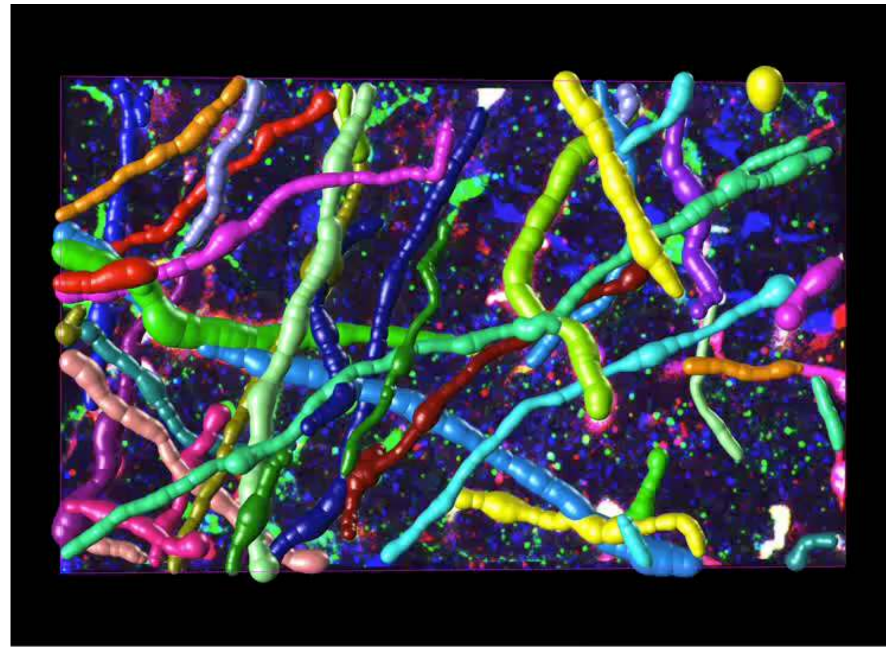
Fluorescent neurons in vivo in the adult mouse brain Imaged through a cranial window using a 2-photon microscope.

Electron Microscopy Image Stack at five nanometer resolution.

Courtesy of G. Knott

Delineating Dendritic Trees

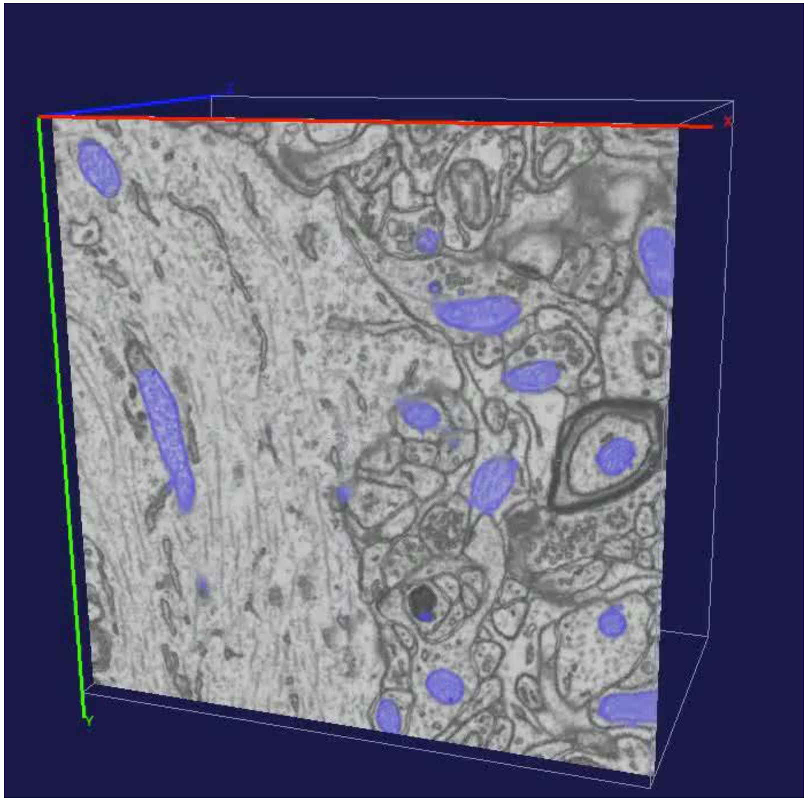


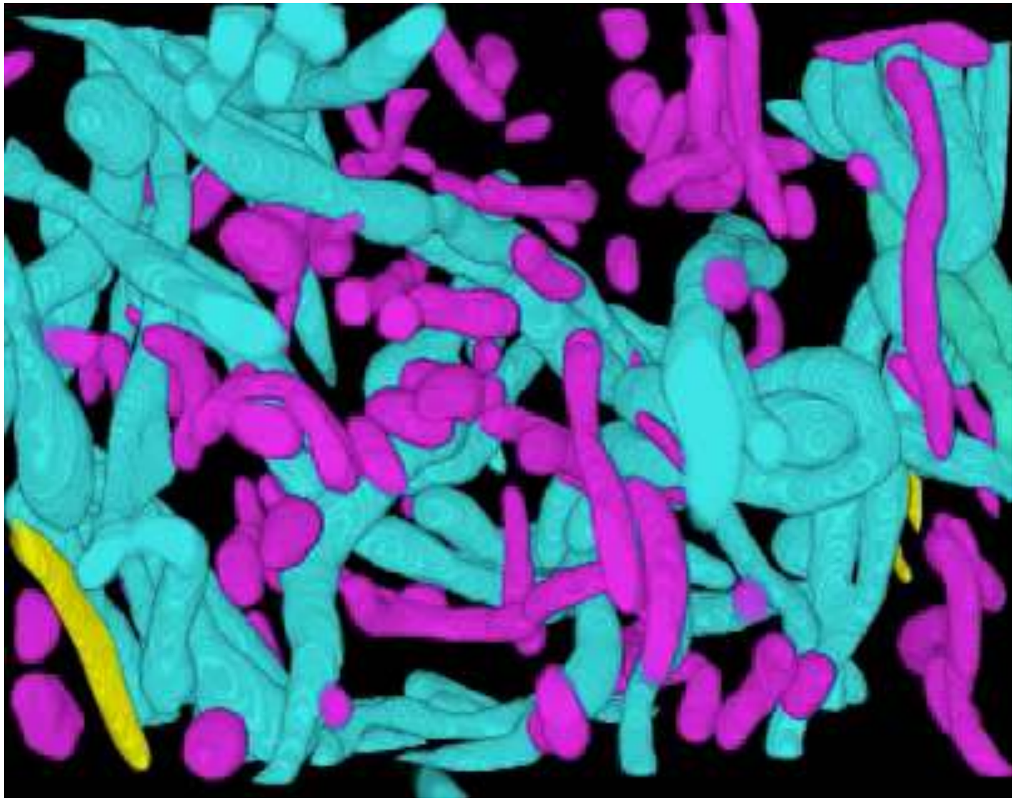


pictures → graphs

Turetken et al., PAMI'16

Finding Mitochondria





Lucchi et al. TMI'11

Google Earth For The Brain

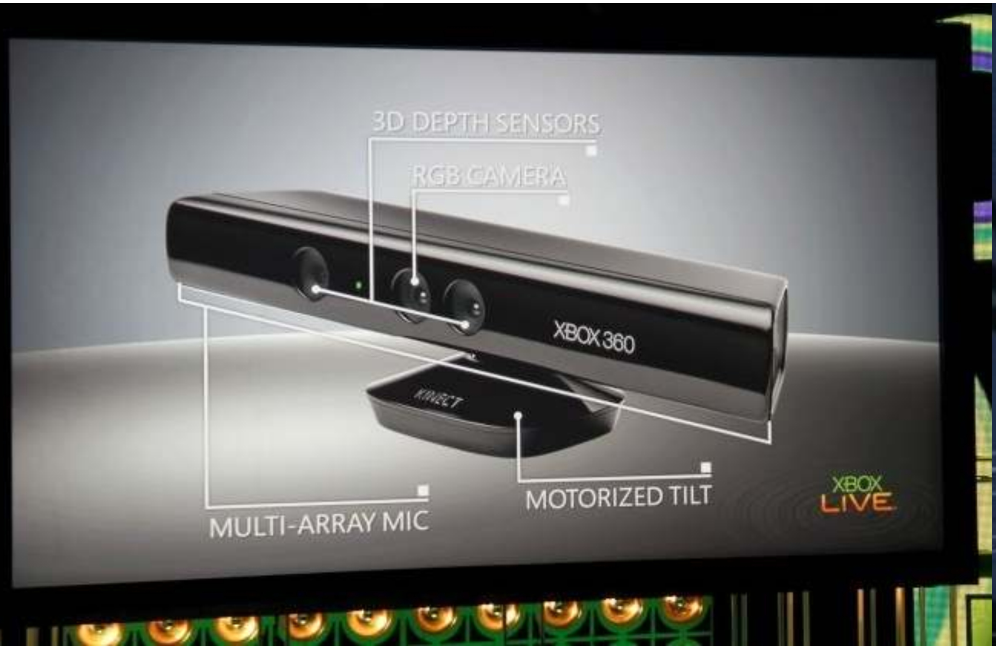
• A human brain contains approximately 100 billion neurons and 100 trillion synapses.
• It would take 1000 Exabytes to store an uncompressed digitization at 5nm resolution.

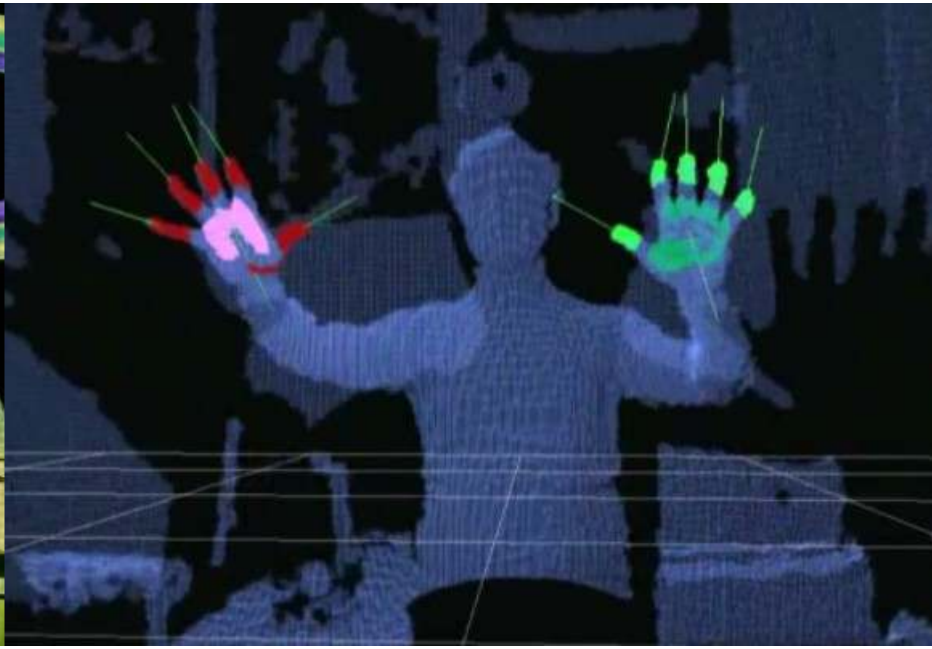
amazon x 500!

—> Seriously big data!

clear, automated CV solutions needed!

KINECT





→ Image whose pixel values are distances

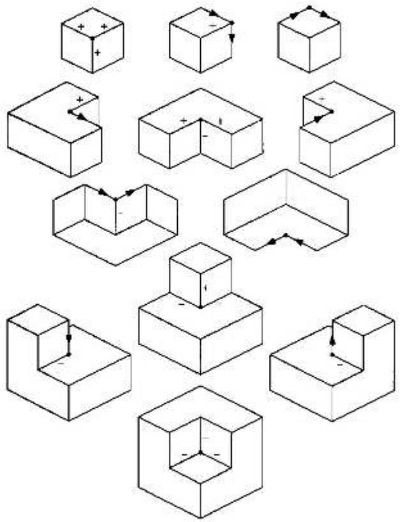
→ Analyze what person doing!

How it Began

- Computer Vision started in 1965 at MIT as a short term project.

- A world of perfect blocks and strong assumptions.

→ The real world is not like that!



Roberts, PhD 65

Historical Perspective

- 1960s: Beginnings in artificial intelligence, image processing and pattern recognition.
- 1970s: Foundational work on image formation.
- 1980s: Vision as applied mathematics, geometry, multi-scale analysis, control theory, optimization.
- 1990s: Physics-based models, Probabilistic reasoning.
- 2000s: Machine learning. → AdaBoost, DTs, SVMs
- 2010s: Deep Learning.
- 2020s: ?????

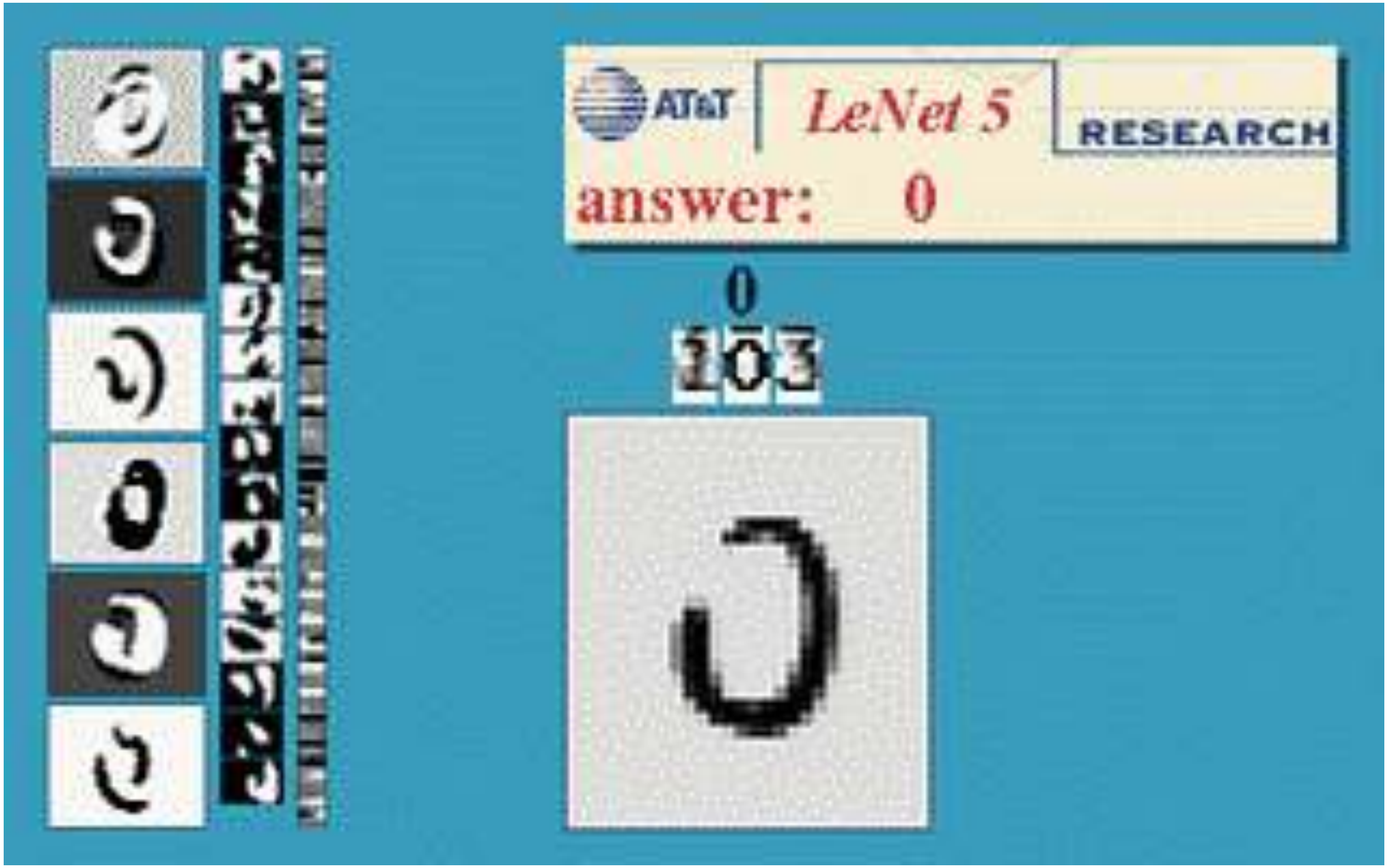
--> Improved understanding and successful applications in graphics, mapping, biometrics, and others but still far from human performance.

Human vs Machine Learning

EPFL

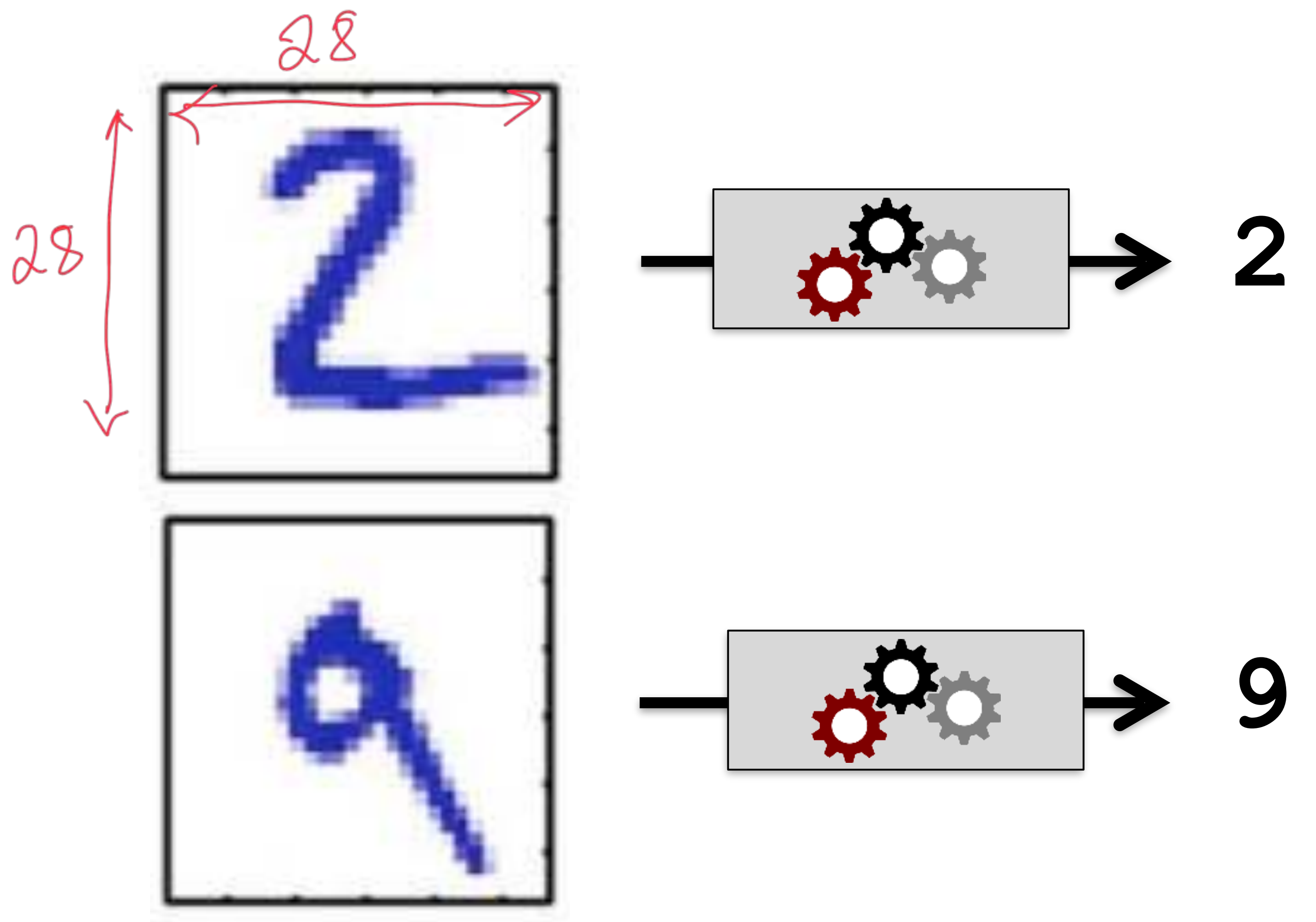
Recognizing Hand-Written Digits

Early DeepNets



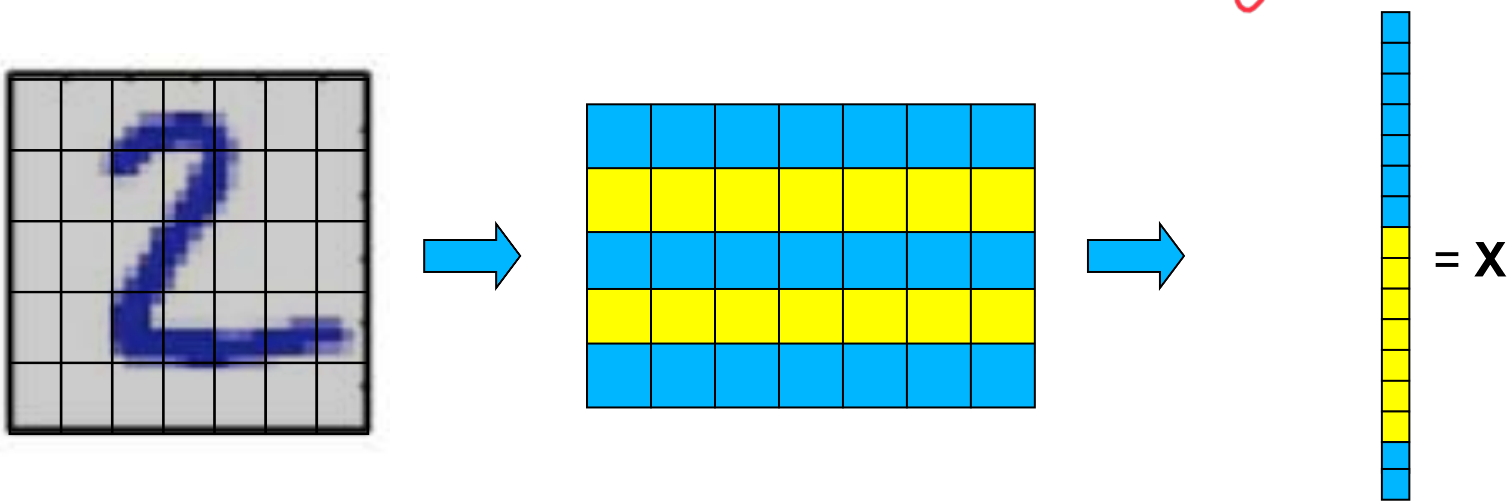
LeNet (1989-1999)

Recognizing Hand-Written Digits



Predictor and Labels

concatenate



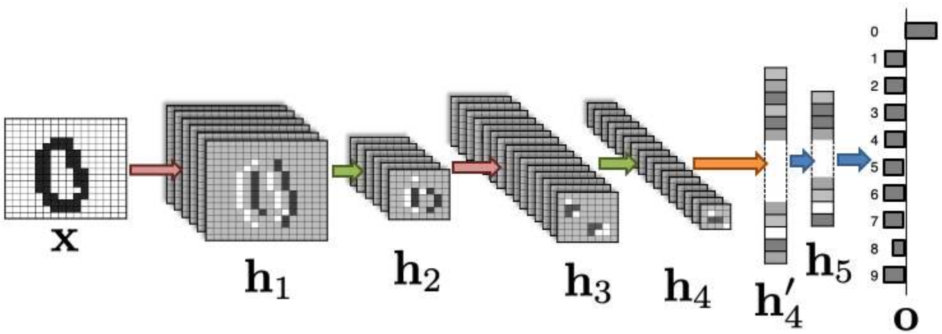
28x28 pixels

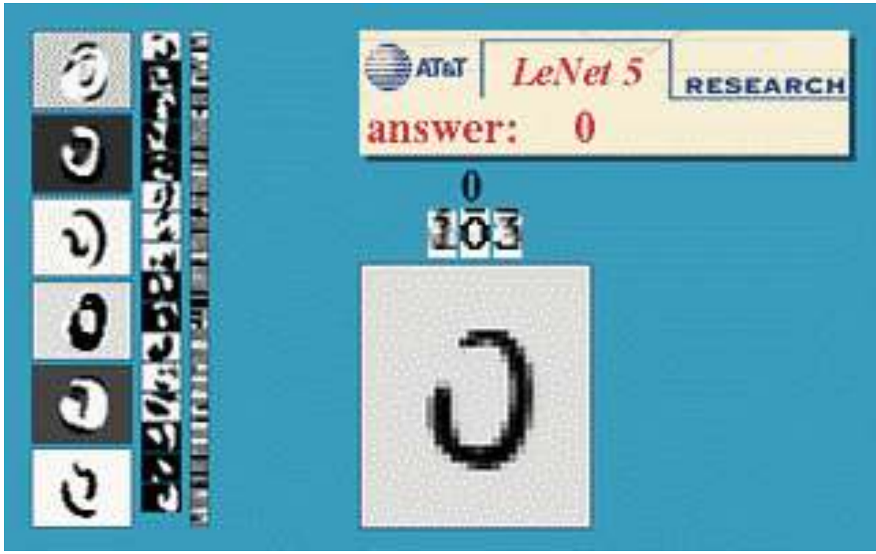
X is a 784-D Vector

$$\boxed{y} : \mathbf{x} \in \mathbb{R}^{784} \rightarrow \boxed{\{0, 1, 2, \cdots, 9\}} \quad ?$$

Predictor **Labels**

Convolutional Neural Networks





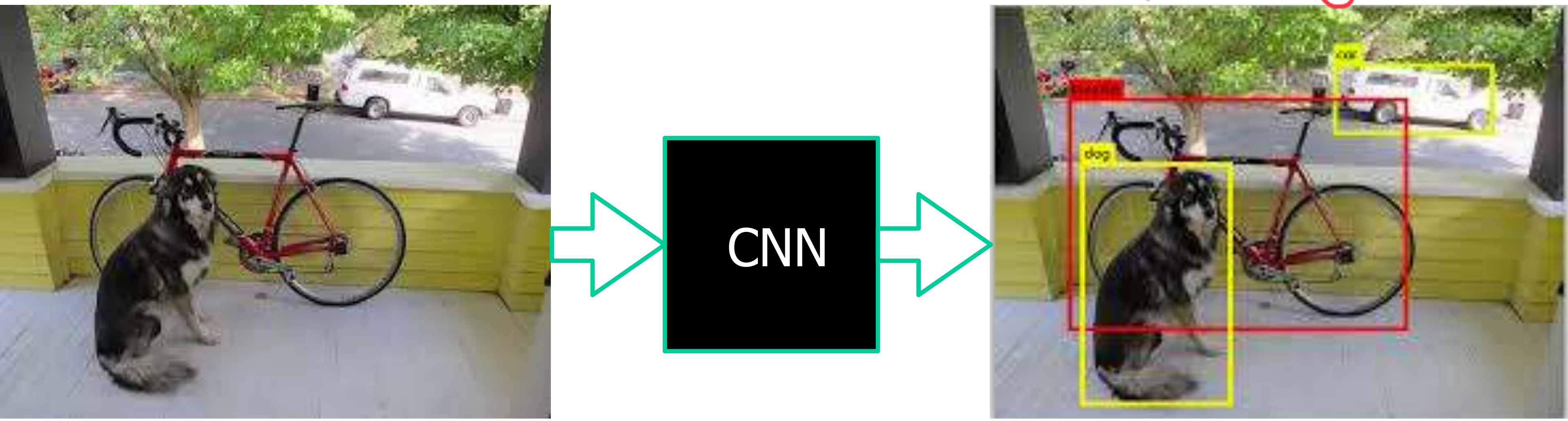
- Powerful way to encode the function y.
- But require much training data.

—> We will discuss them later in the class.

EPFL

Computer Vision in The Era of Deep Nets

Bounding boxes

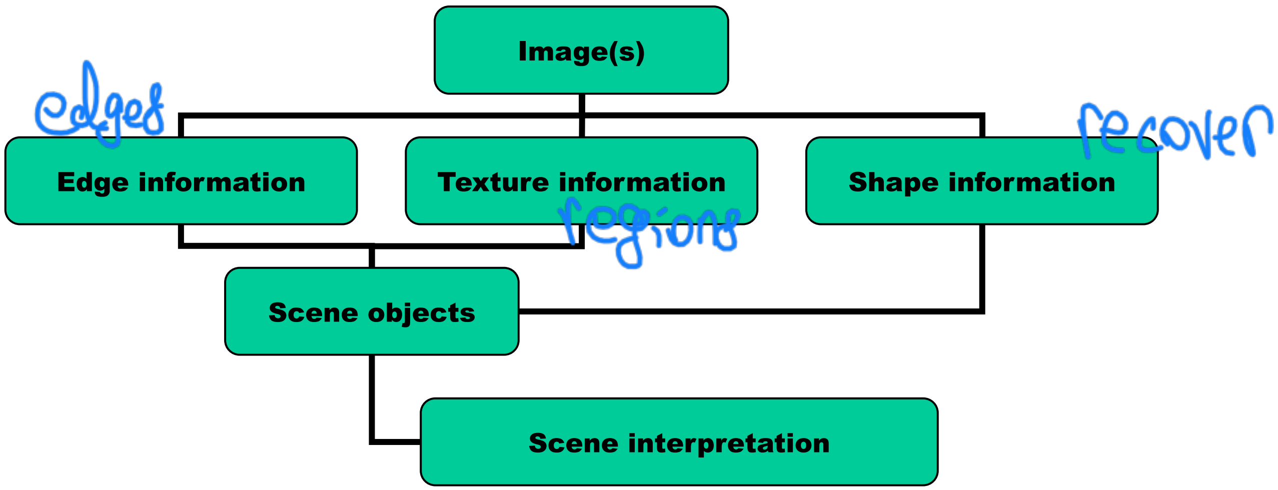


- Extremely effective in many cases but does not shed much light on the vision process.
- The best algorithms combine Deep Nets with more traditional techniques.

↳ geometry, physical properties of the world

Redmon et al. , CVPR'16

A Teachable Scheme



Decomposition of the vision process into smaller manageable and implementable steps.

--> Paradigm followed in this course

--> May not be the one humans use

Course Outline

Introduction:

- Definition
- Human vision
- Image formation

what we (don't) understand existence ✓

3D → (sequences) of images, 2D images

process 2D images

Extracting features:

- Contours
- Texture
- Regions

3D interpretation of images (e.g. volcanos)

Shape recovery:

- From one image
- Using additional images

recovering 3D images

Course Organization

- Formal lectures every week (Monday)
- Exercises every other week (Tuesday)
  - Two of them will be graded (10% of the grade, each).
  - Dates are on the moodle page.
  - Bring your laptops for the exercises.
- Written exam (80% of the grade).

Course Material

Textbooks:

- R. Szeliski, Computer Vision: Computer Vision: Algorithms and Applications, 2021.
- A. Zisserman and R. Hartley, Multiple View Geometry in Computer Vision, Cambridge University Press, 2003.

Web pages:

- moodle.epfl.ch (Computer Vision, CS-442)
- cvlab.epfl.ch/projects (Projects)
- cvlab.epfl.ch/research (Research)

Slide Codes

Training vs Testing

Normal slide: It is part of the course and I may ask exam questions about it.

Training vs Testing

Reminder

Reminder slide: We have already covered this earlier in the class. Go back to the appropriate lecture if you do not remember.

Training vs Testing

Optional Bishop, xxx

Optional slide: This is additional material for people interested in more details. I will not ask direct exam questions on this.

Reference to book or paper for even more details.