Lecture 11: Computer vision

EMAT31530/Feb 2018/Raul Santos-Rodriguez

Have a look at ...

... Russell and Norvig (3rd edition) Chapter 24

... Computer Vision: Algorithms and Applications, Richard Szeliski, Springer, 2010.

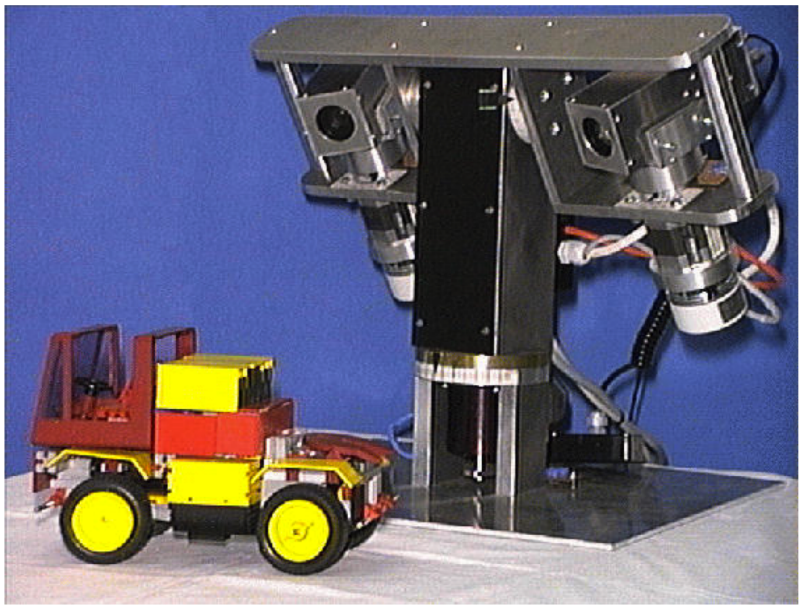
... Computer Vision: A Modern Approach, David A. Forsyth and Jean Ponce, Prentice-Hall, 2002.

... OpenCV

... `http://www.vision.caltech.edu/Image_Datasets/Caltech101/`

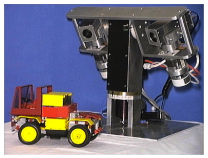

Outline

Although perception appears to be an effortless activity for humans, it requires a significant amount of sophisticated computation. The goal of vision is to extract information needed for tasks such as manipulation, navigation, and object recognition.



Outline

Although perception appears to be an effortless activity for humans, it requires a significant amount of sophisticated computation. The goal of vision is to extract information needed for tasks such as manipulation, navigation, and object recognition.

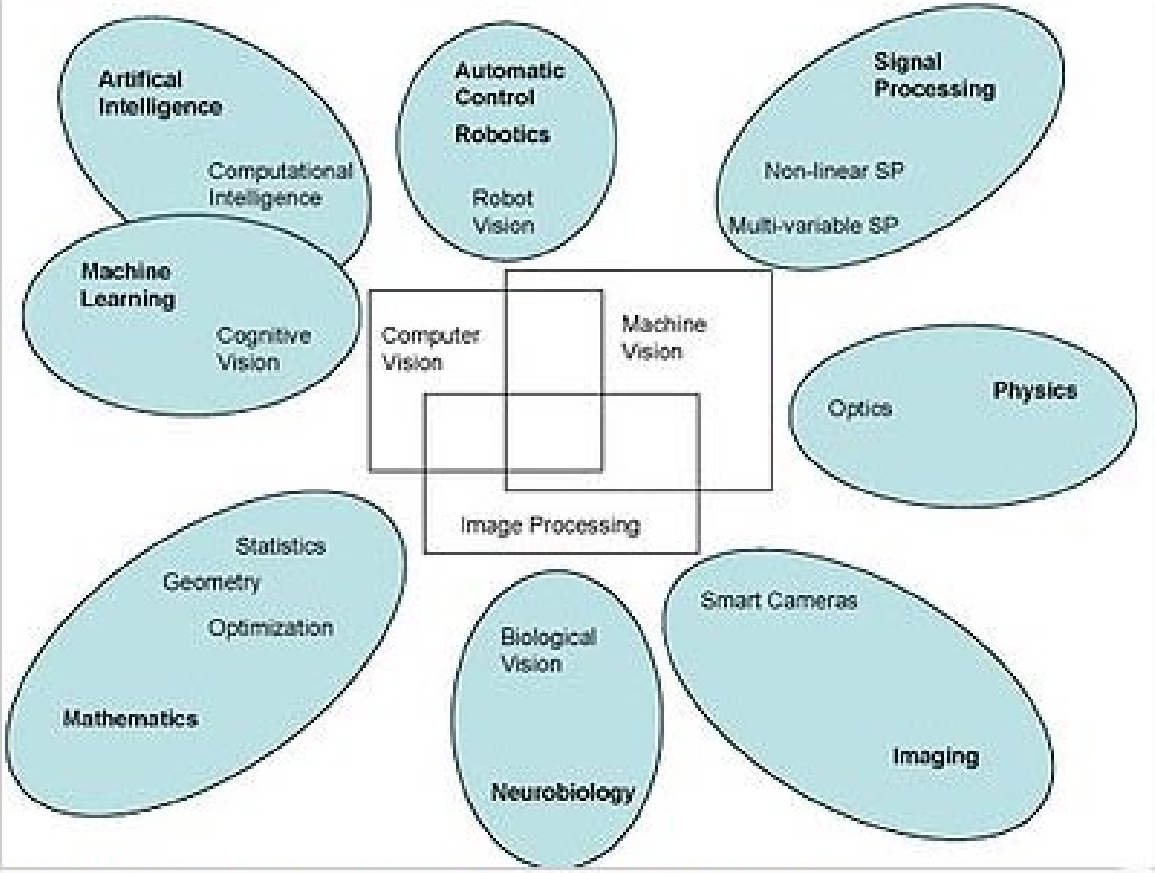


Computer Vision is the science of building systems that can extract certain task-relevant information from a visual scene. Such systems can be used for applications such as optical character recognition, analysis of satellite and microscopic images, magnetic resonance imaging, surveillance, identity verification, quality control in manufacturing, etc. The image data can take many forms, such as a video sequence, depth images, views from multiple cameras, or multi-dimensional data from a medical scanner.

Alternative definitions

- Trucco and Verri: computing properties of the 3D world from one or more digital images
- Stockman and Shapiro: to make useful decisions about real physical objects and scenes based on sensed images
- Ballard and Brown: the construction of explicit, meaningful description of physical objects from images
- Forsyth and Ponce: extracting descriptions of the world from pictures or sequences of pictures

Related fields



Relation between Computer vision and various other fields

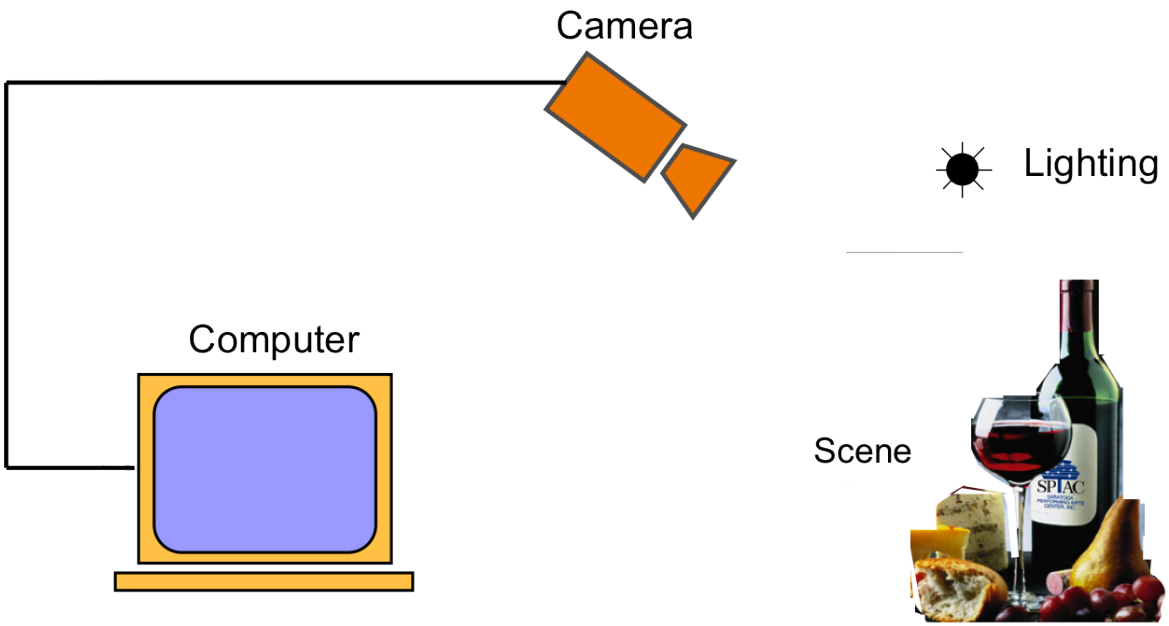
Every picture tells a story

Every picture tells a story

The goal of computer vision is to write computer programs that can interpret images

- What kind of scene?
- Where are the cars?
- How far are the buildings?
- ...

Components of a computer vision system



Computer vision vs Human vision

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

What a computer sees

Can computers match (or beat) human vision?

Yes and no (but mostly no!)

- Humans are much better at hard things
- Computers can be better at easy things

Can computers match (or beat) human vision?

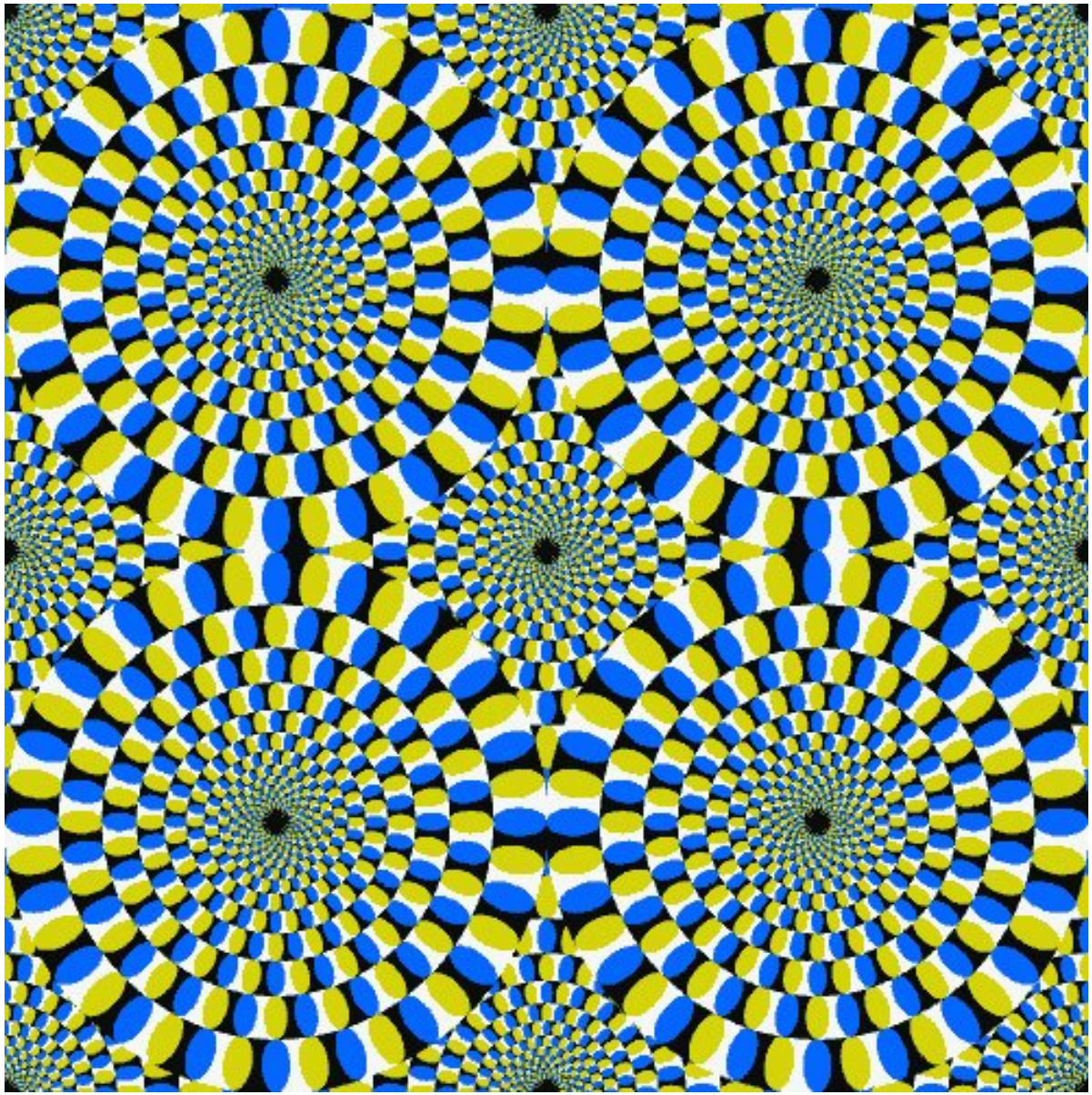
Can computers match (or beat) human vision?

Yes and no (but mostly no!)

- Humans are much better at hard things
- Computers can be better at easy things

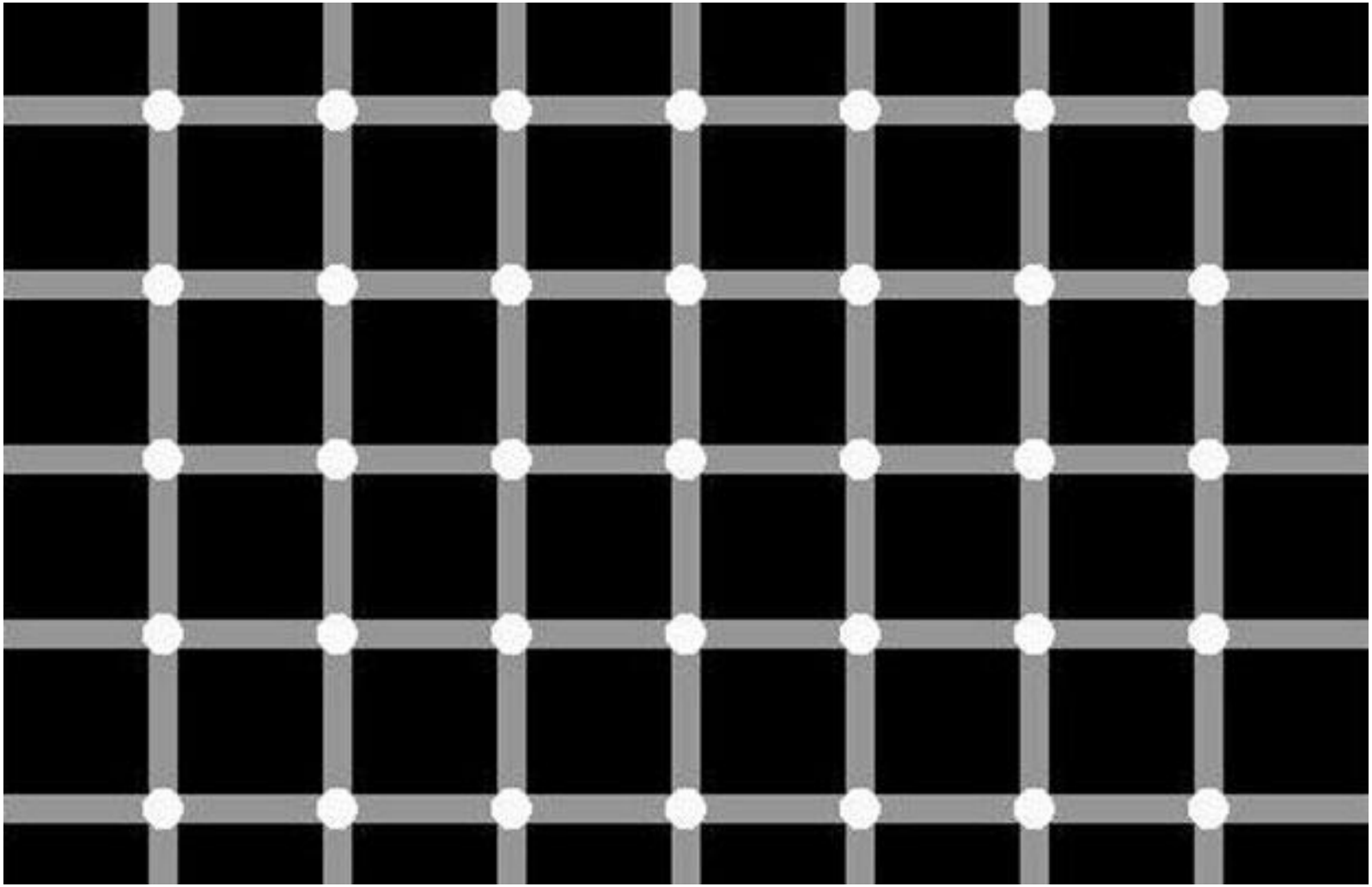
The generic "Vision Problem" is far from being solved. No existing system can come close to emulating the capabilities of a human. Systems such as the ones described above are fundamentally brittle: as soon as the input deviates ever so slightly from the intended format, the output becomes almost invariably meaningless.

Optical illusions



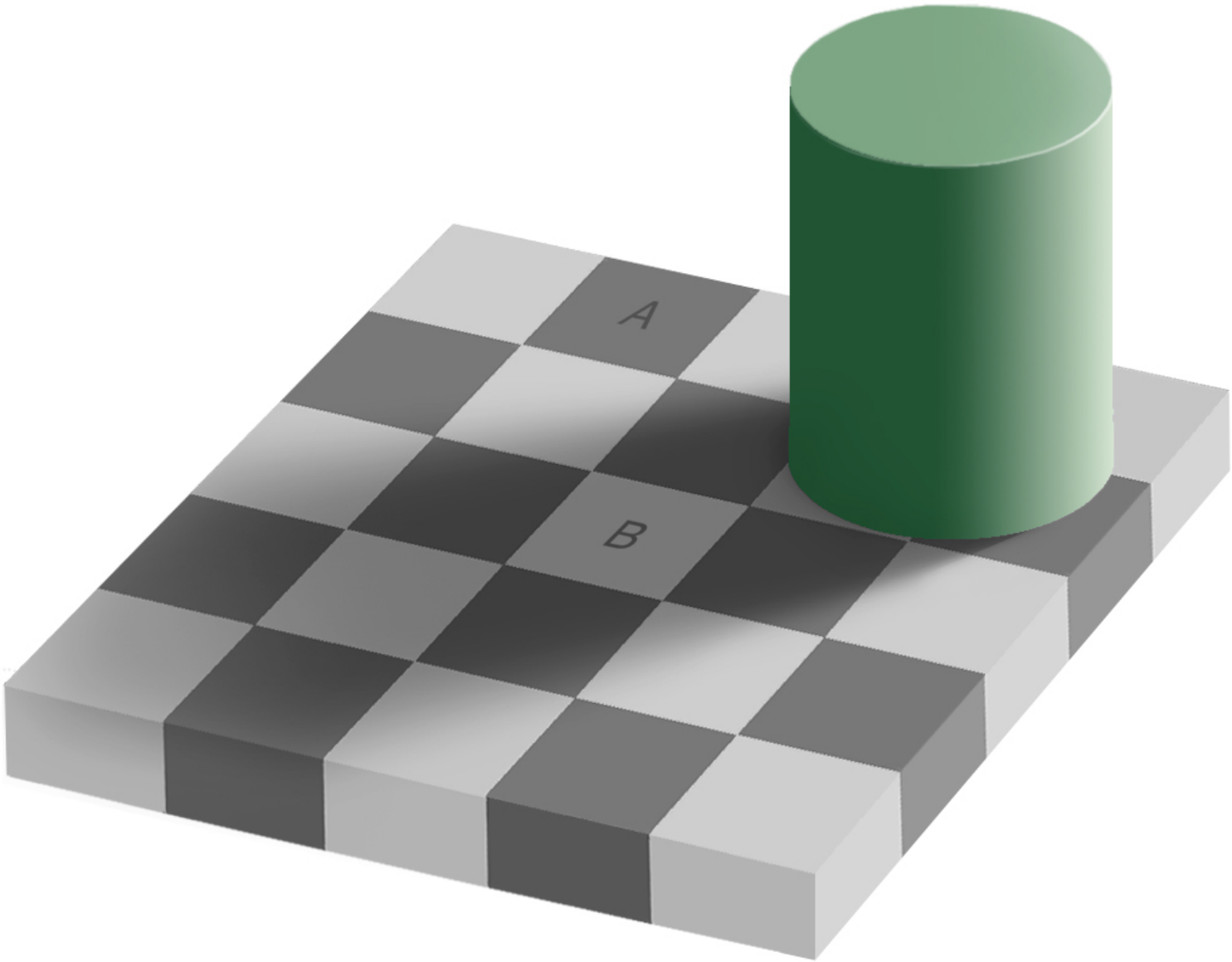
http://www.echalk.co.uk/amusements/opticalillusions/illusions.html

Optical illusions

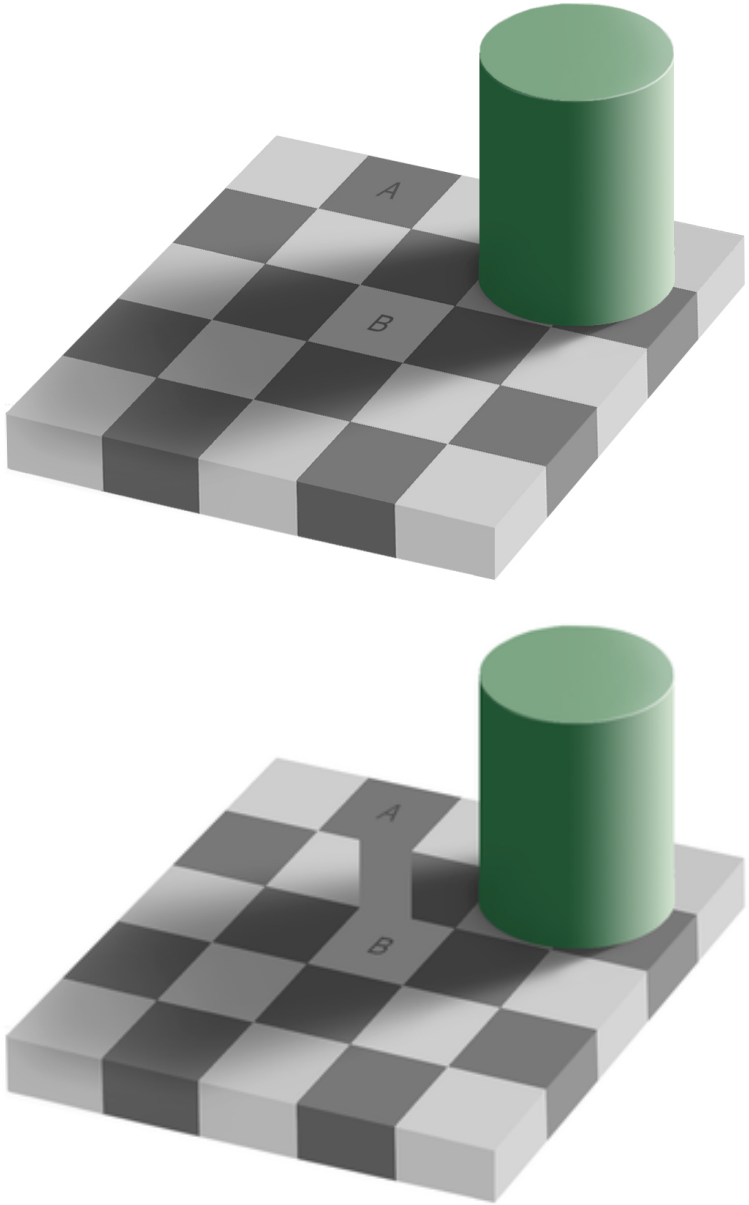


http://www.echalk.co.uk/amusements/opticalillusions/illusions.html

Optical illusions



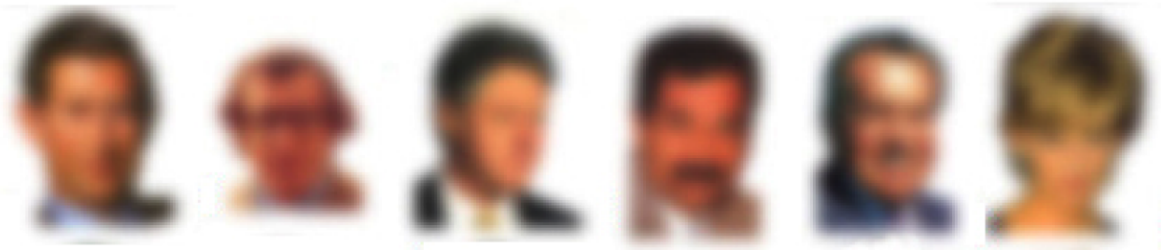
Optical illusions



Why is computer vision difficult?

- Ill-posed
- High-dimensional data
- Noise
- Variation

Why is computer vision difficult?



- Ill-posed
- High-dimensional data
- Noise
- Variation

Why is computer vision difficult?

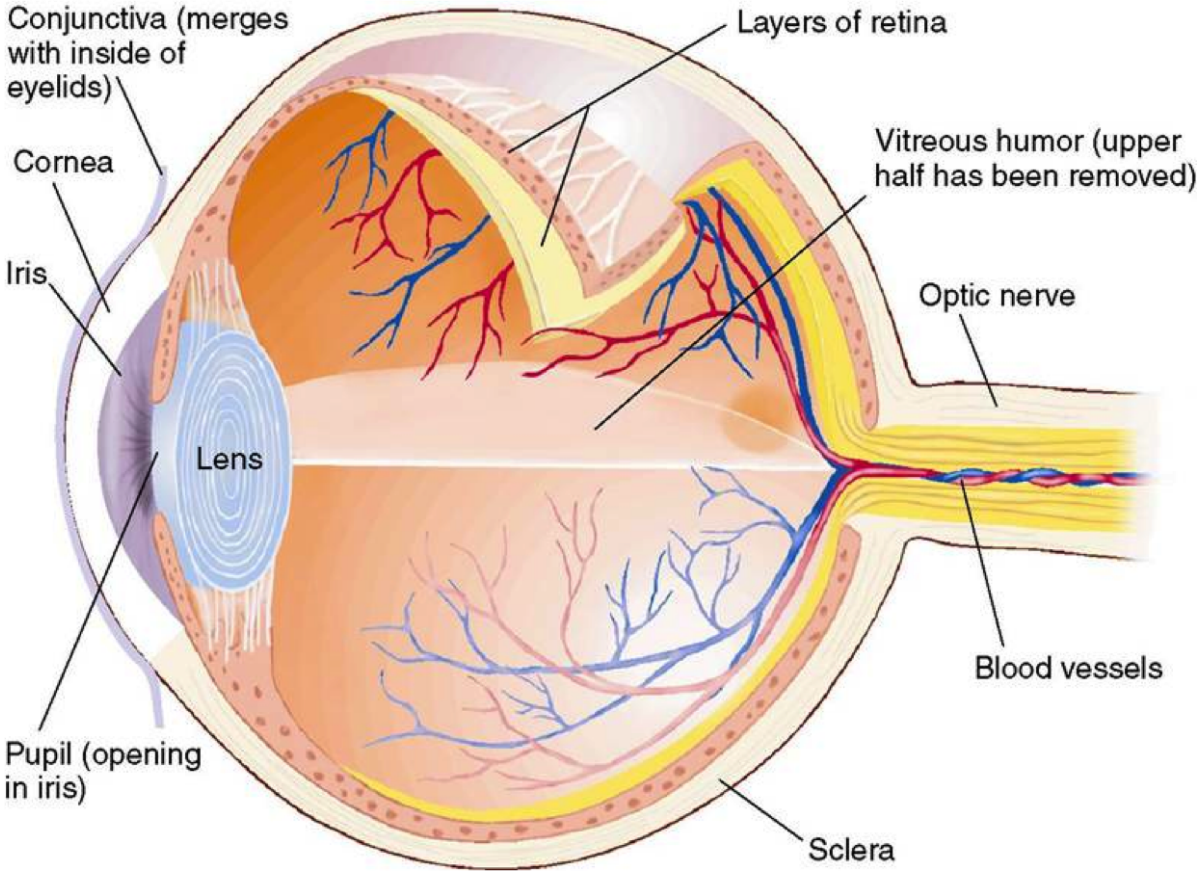
Why is computer vision difficult?

- Ill-posed
- High-dimensional data
- Noise
- Variation

Vision is an amazing feat of natural intelligence:

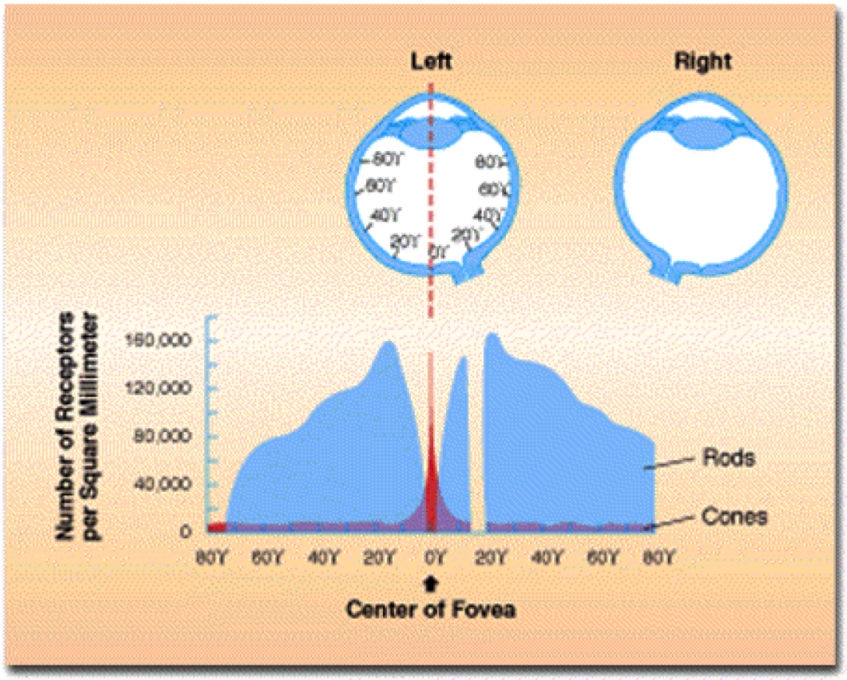
- Visual cortex occupies about 50% of Macaque brain
- More human brain devoted to vision than anything else

Cross Section of the Human Eye



Rods and Cones

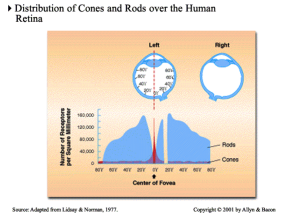
- Distribution of Cones and Rods over the Human Retina



Source: Adapted from Lidsay & Norman, 1977.

Rods and Cones

Rods and Cones



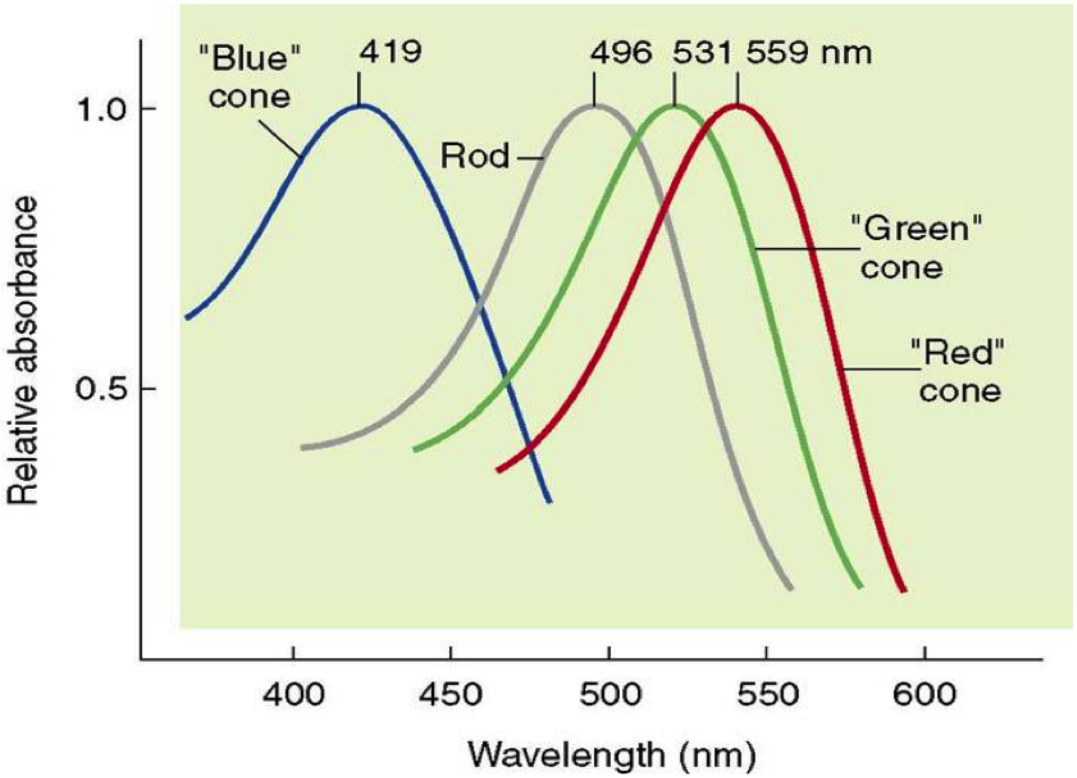
Rods

- Perceive brightness only
- Night vision

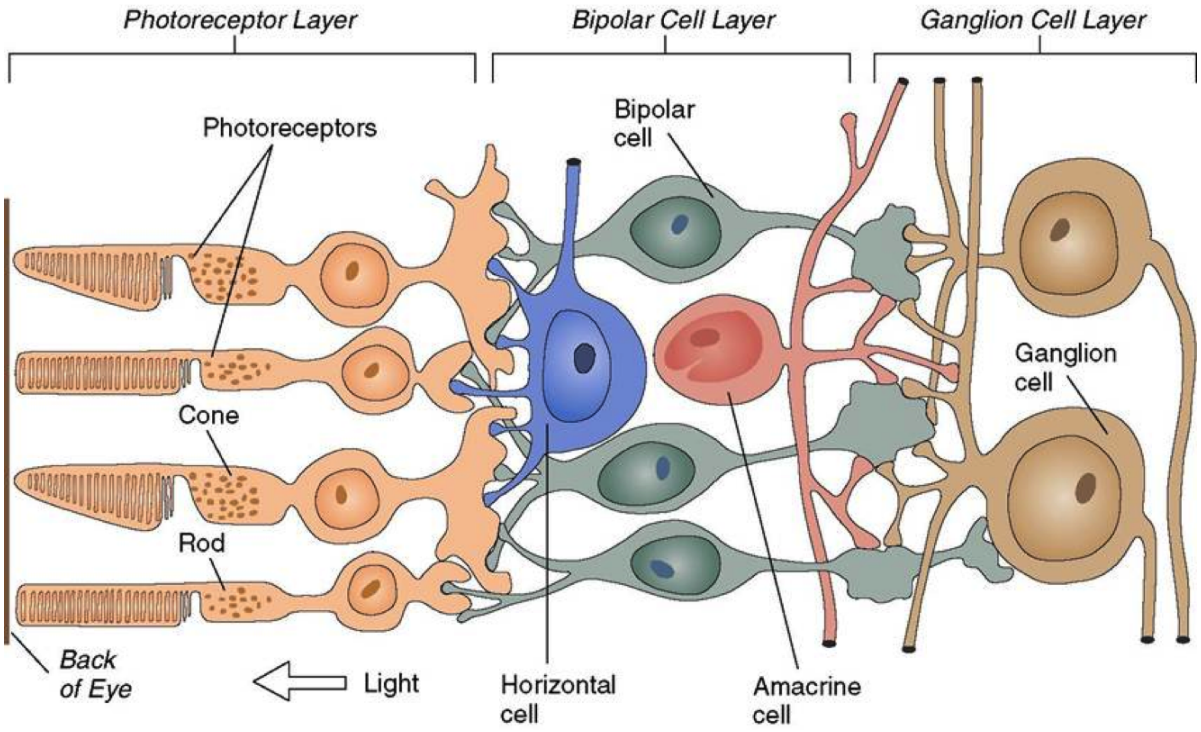
Cones

- Perceive color
- Day vision
- Red, green, and blue cones

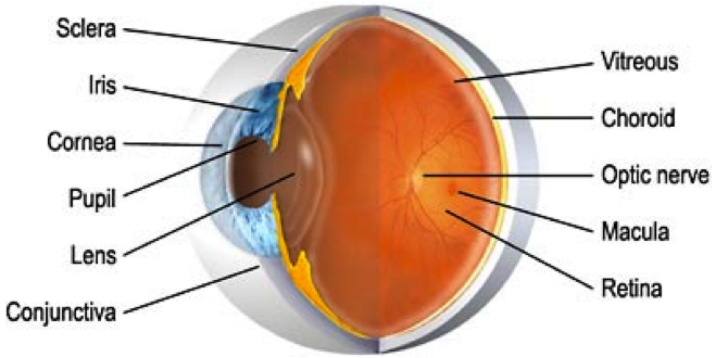
Retina


"Blue" cone
419
496
531
559 nm
Rod
"Green" cone
"Red" cone
1.0
0.5
Relative absorbance
400
450
500
550
600
Wavelength (nm)

Cell types of the retina



Camera vs Eye



Camera components	**Eye components**
Lens	Lens, cornea
Shutter	Iris, pupil
Film	Retina
Cable to transfer images	Optic nerve to send the incident light information to the brain

Digital image representation

(0,0)

126	127	126	128	127	124	158	
125	126	127	123	120	144	163	
123	126	125	121	128	155	160	
126	123	127	122	142	162	164	
120	122	124	130	157	161	166	
119	121	123	145	162	164	165	

0 → black, 255 → white

n

$0 \le s(m,n) \le 255$ } **quantization**

$0 \le m \le M-1$

$0 \le n \le N-1$ } **sampling**

Topics

- Image formation: camera model, camera calibration, radiometry, color, shading
- Low-level vision: stereopsis, structure from motion, illumination, reflectance, shape, texture
- Mid-level vision: segmentation, grouping, Kalman filter, particle filter, shape representation
- High-level vision: correspondence, matching, object detection, object recognition, visual tracking

Topics

- Image formation: camera model, camera calibration, radiometry, color, shading
- Low-level vision: stereopsis, structure from motion, illumination, reflectance, shape, texture
- Mid-level vision: segmentation, grouping, Kalman filter, particle filter, shape representation
- High-level vision: correspondence, matching, object detection, object recognition, visual tracking

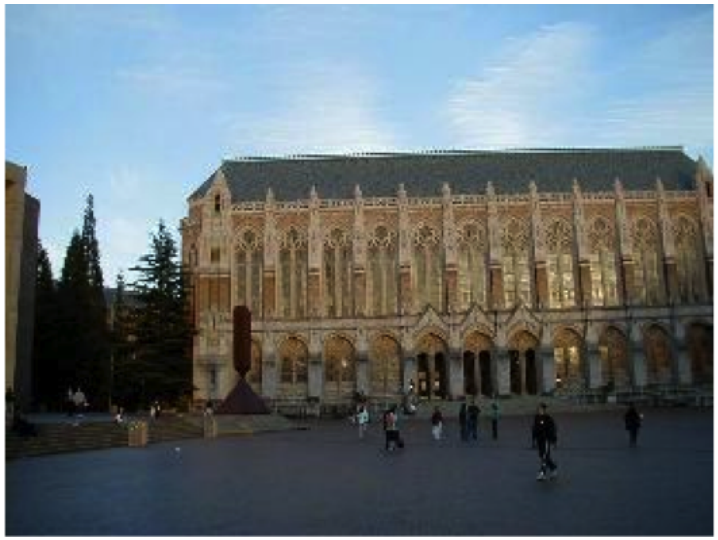
Recent topics: image and video retrieval, internet vision

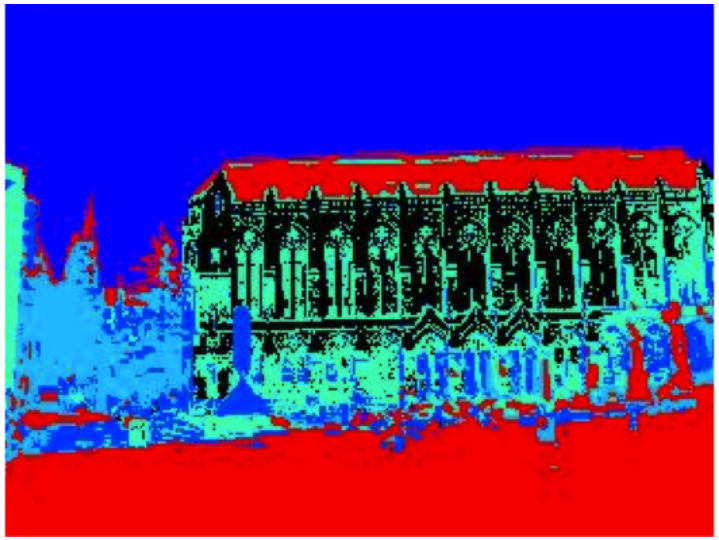
- Segment an image into useful regions
- Perform measurements on certain areas
- Determine what object(s) are in the scene
- Calculate the precise location(s) of objects
- Visually inspect a manufactured object
- Construct a 3D model of the imaged object
- Find interesting events in a video

Low-level

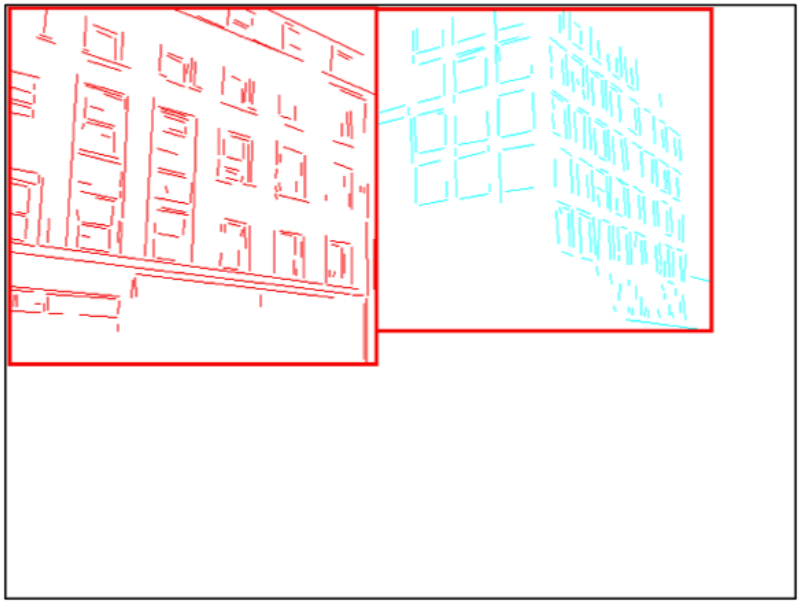
Low-level

Mid-level

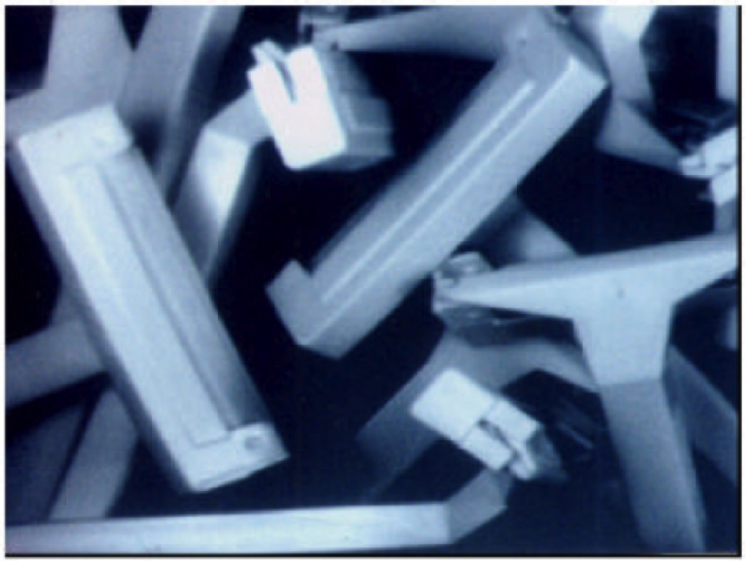


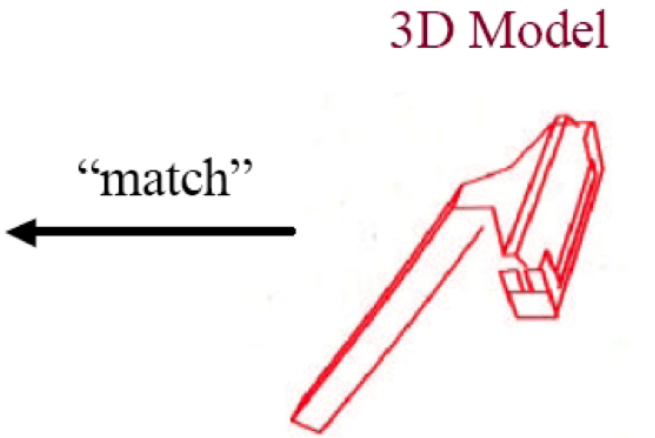


From Low-level to high-level



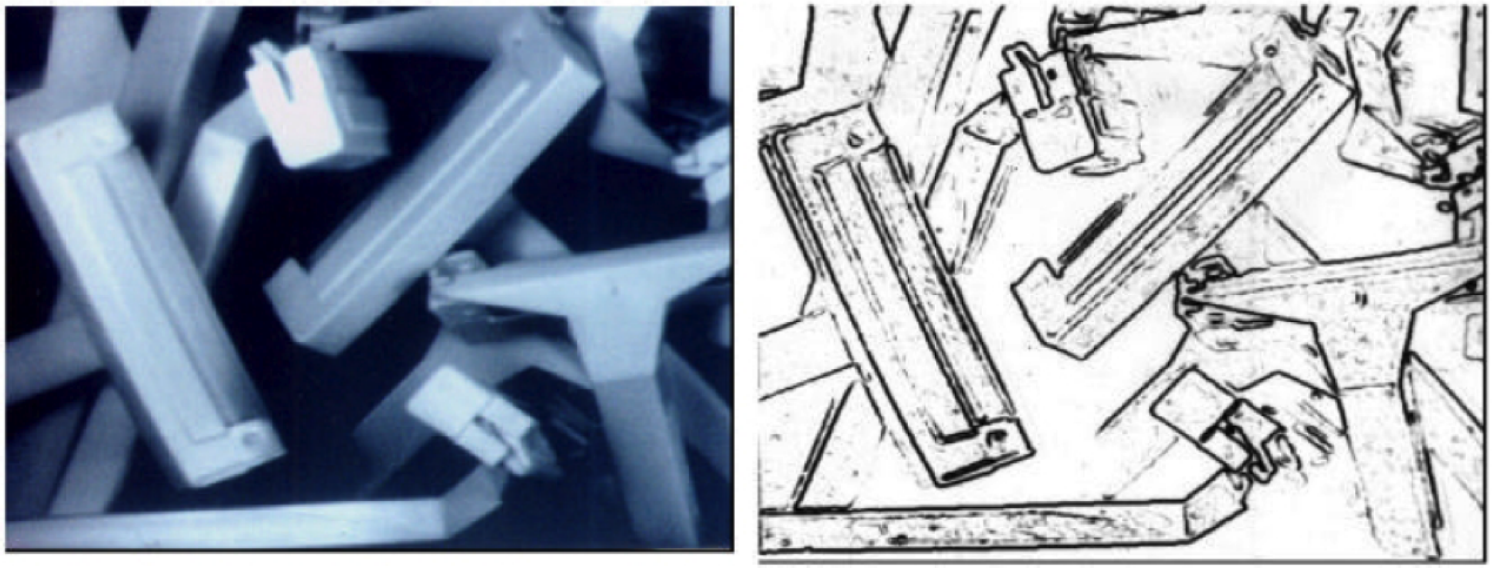
Example: Recognition





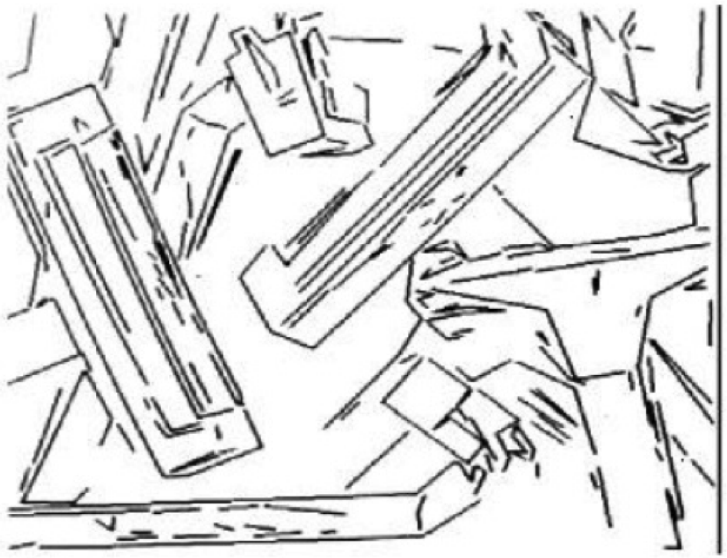
Parameters: 3D position and orientation

Example: Recognition



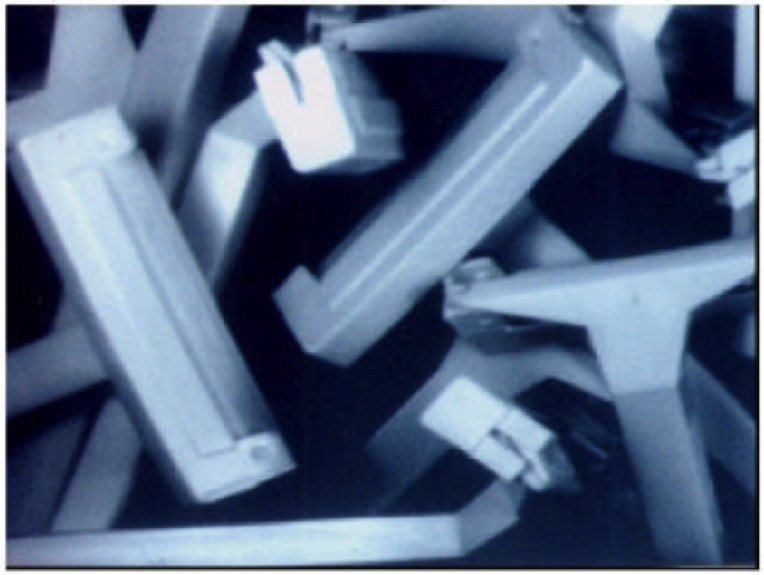
"Filter" image to find brightness changes.

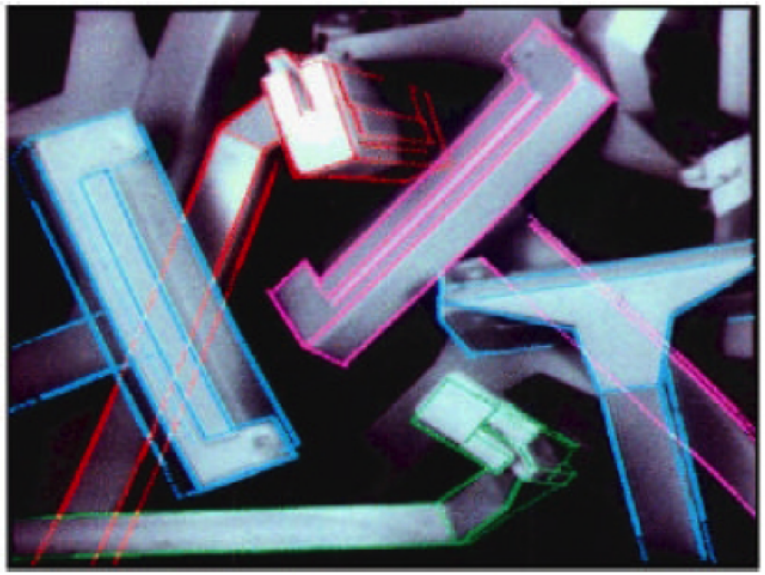
Example: Recognition



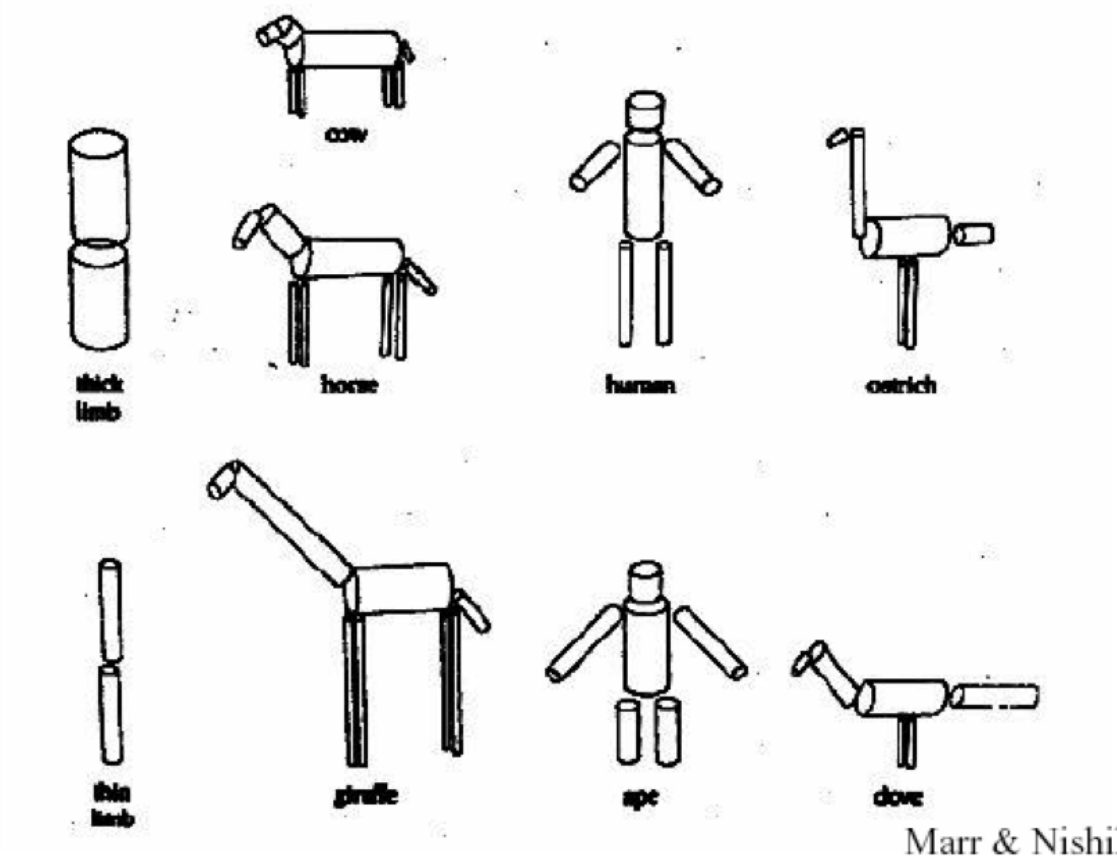
"Fit" lines to the raw measurements.

Example: Recognition



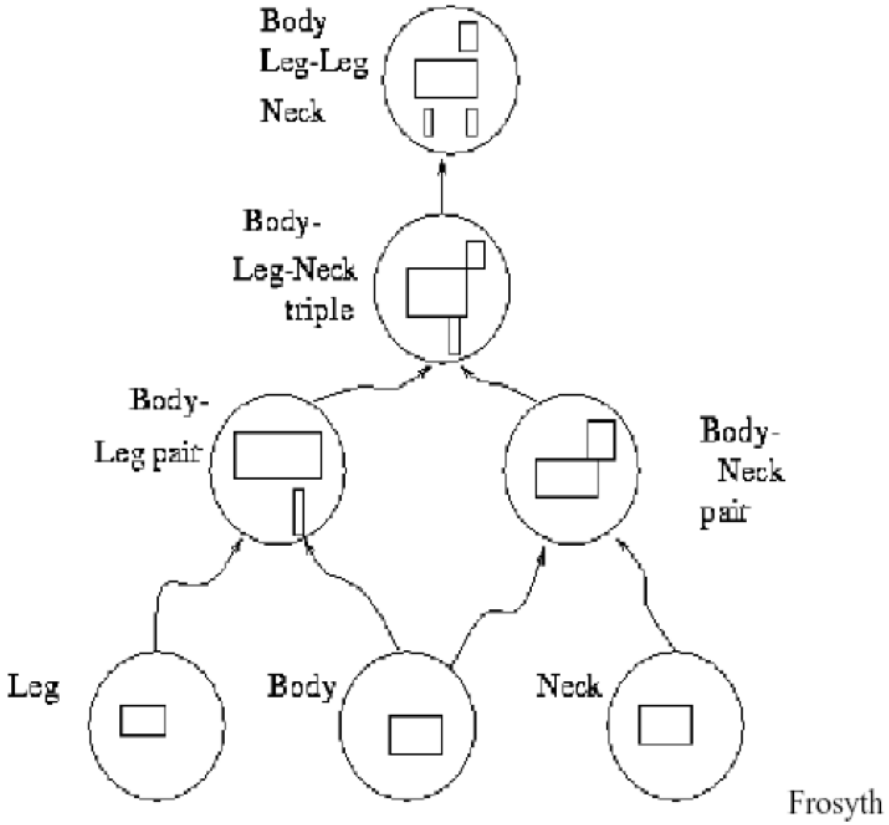


Example: Recognition



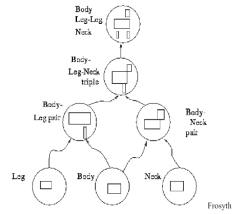
Marr & Nishihara

Example: Recognition



Example: Recognition

Example: Recognition

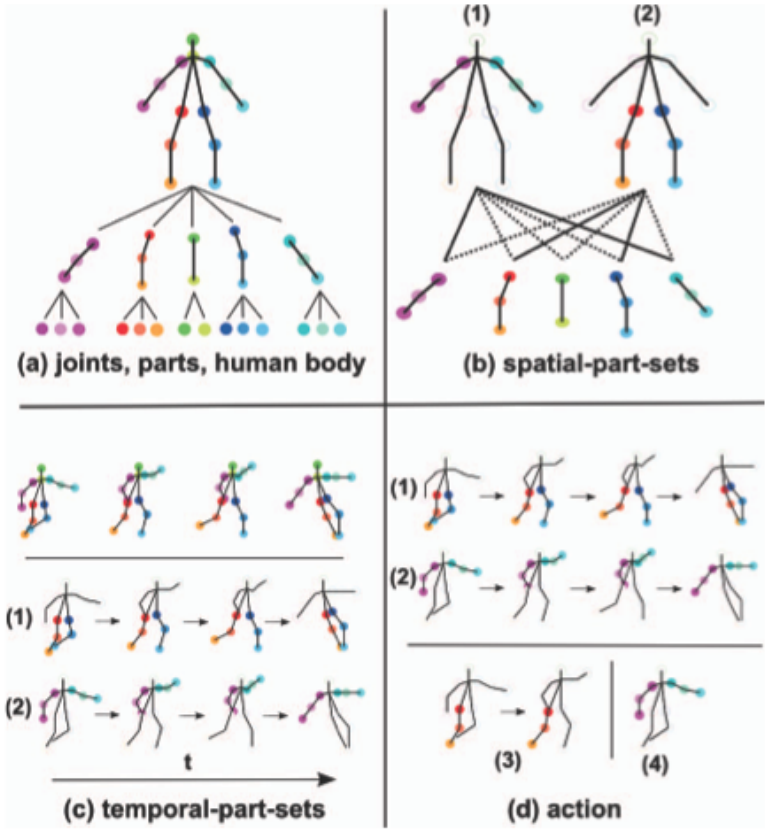


We need some sort of perceptual organization process that tells us what low-level measurements might group together. There are many different cues, including multiple views, texture, shading, etc.
Several problems:

- Which bits of image should be recognised together? Segmentation.
- How can objects be recognised without focusing on detail? Abstraction.
- How can objects with many free parameters be recognised? No popular name, but its a crucial problem anyhow.
- How do we structure very large model bases? again, no popular name; abstraction and learning come into this.

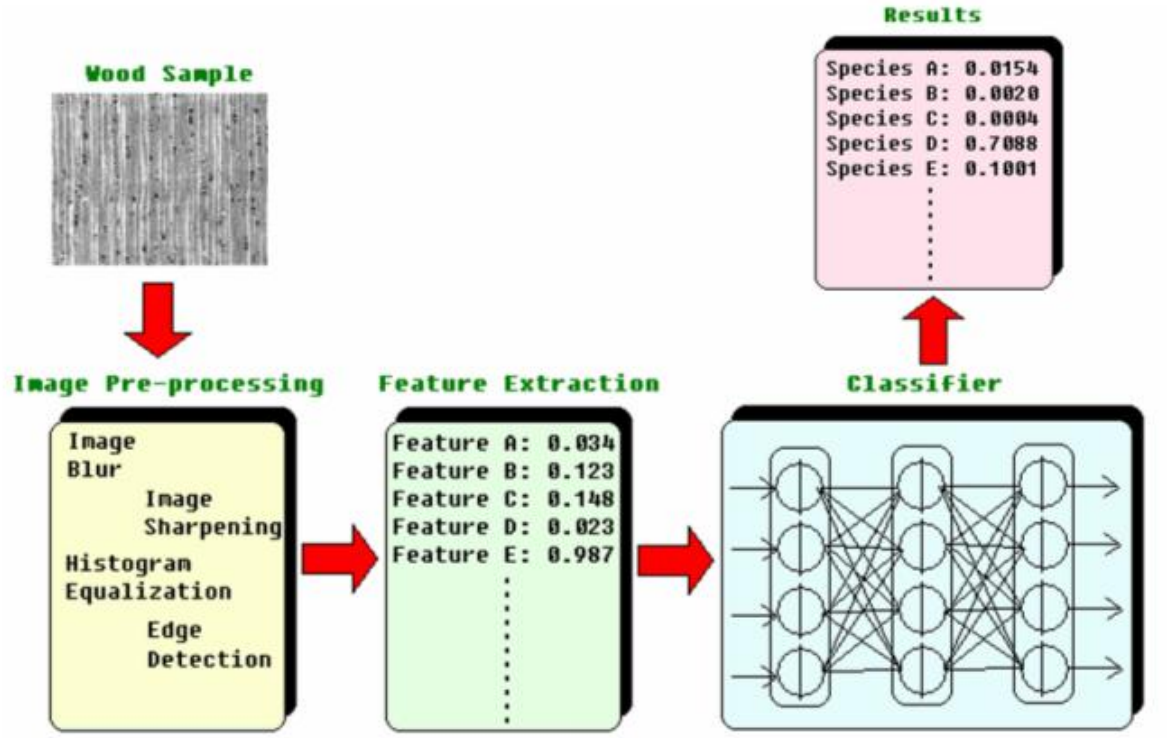
Example: Kinnect

Remember: Random forests



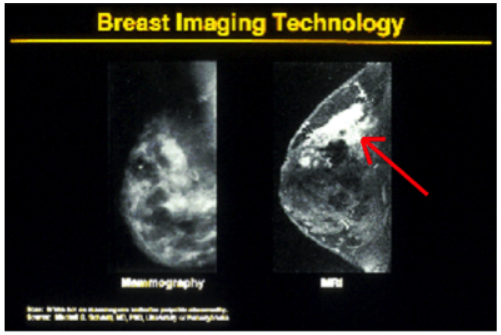
https://www.youtube.com/watch?v=ifEVIsdBQr4
http://research.microsoft.com/en-us/projects/vrkinect/

A general approach



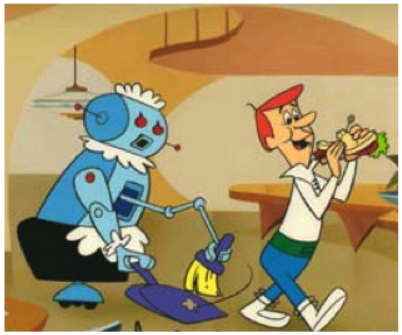
Why computer vision matters

Safety



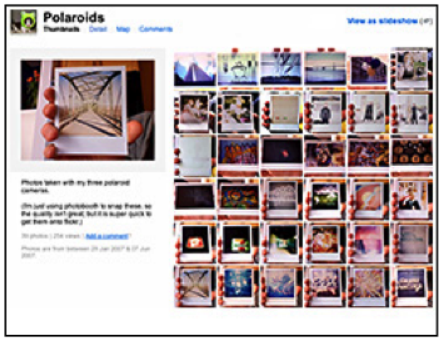
Health

Security



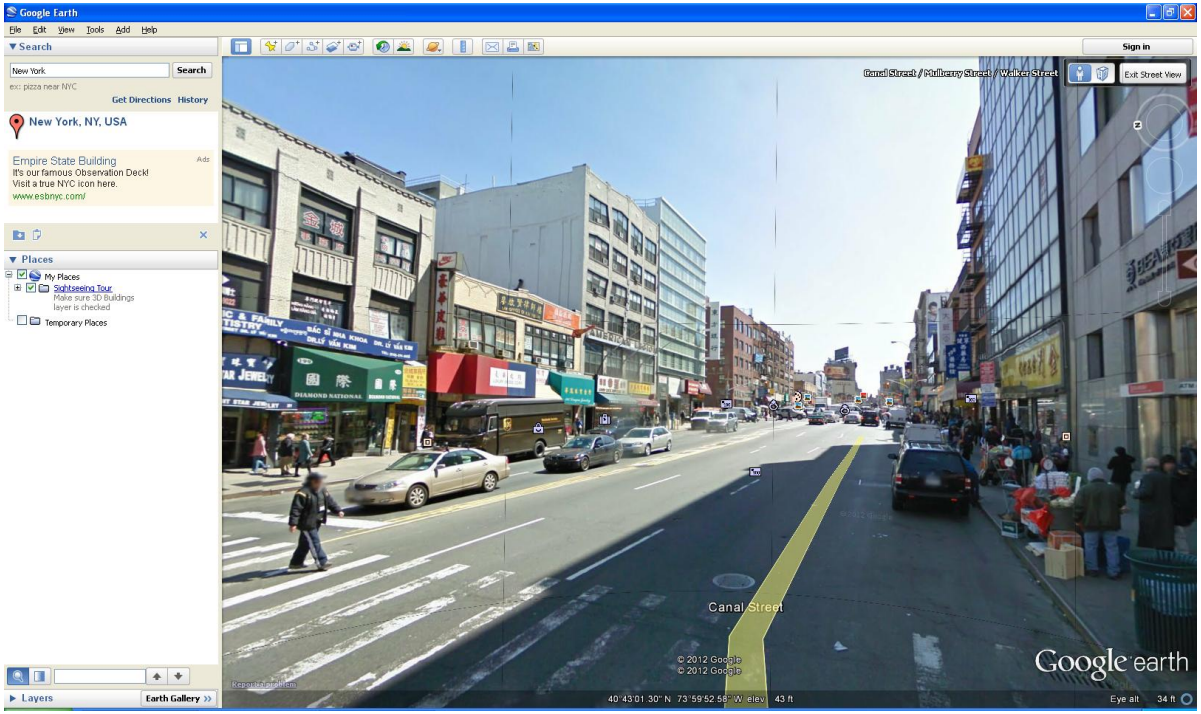
Comfort

Fun

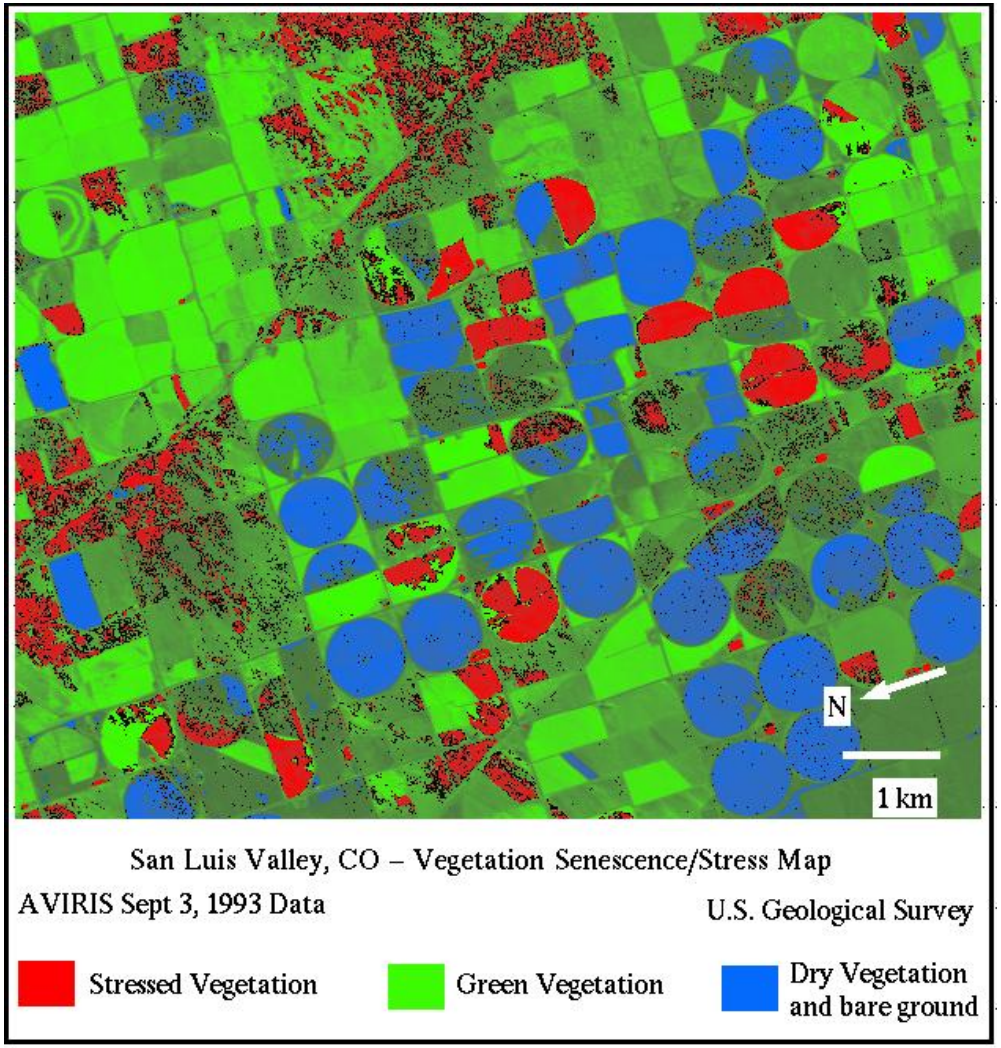


Access

Applications: Earth viewers (3D modeling)

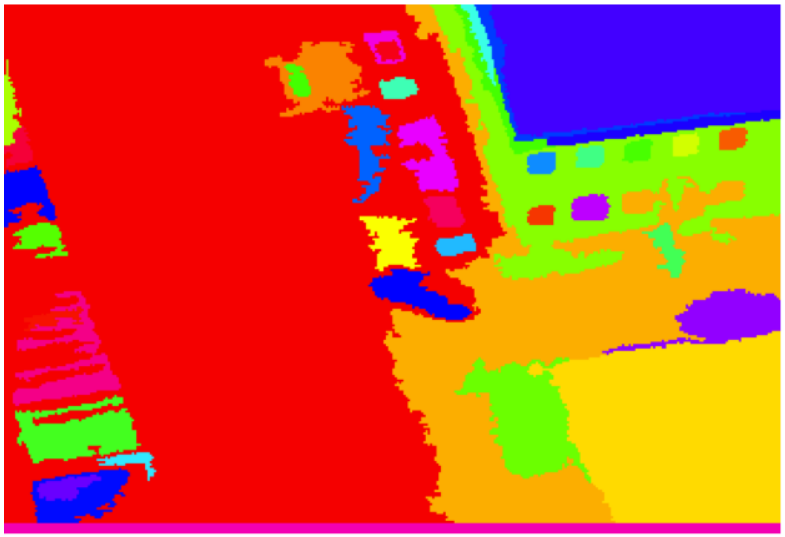


Applications: Remote sensing



Applications: Surveillance

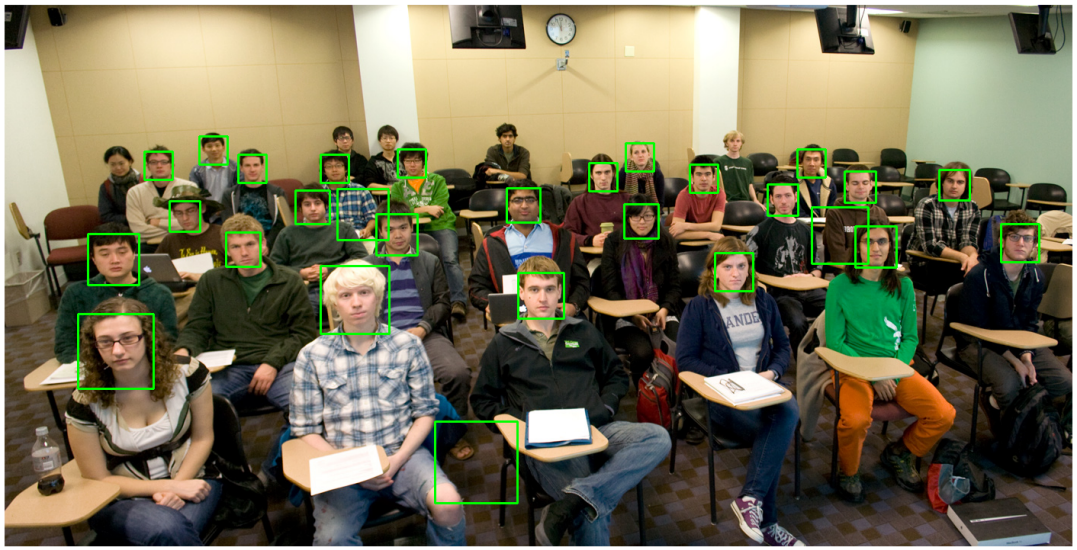
Original Video Frame



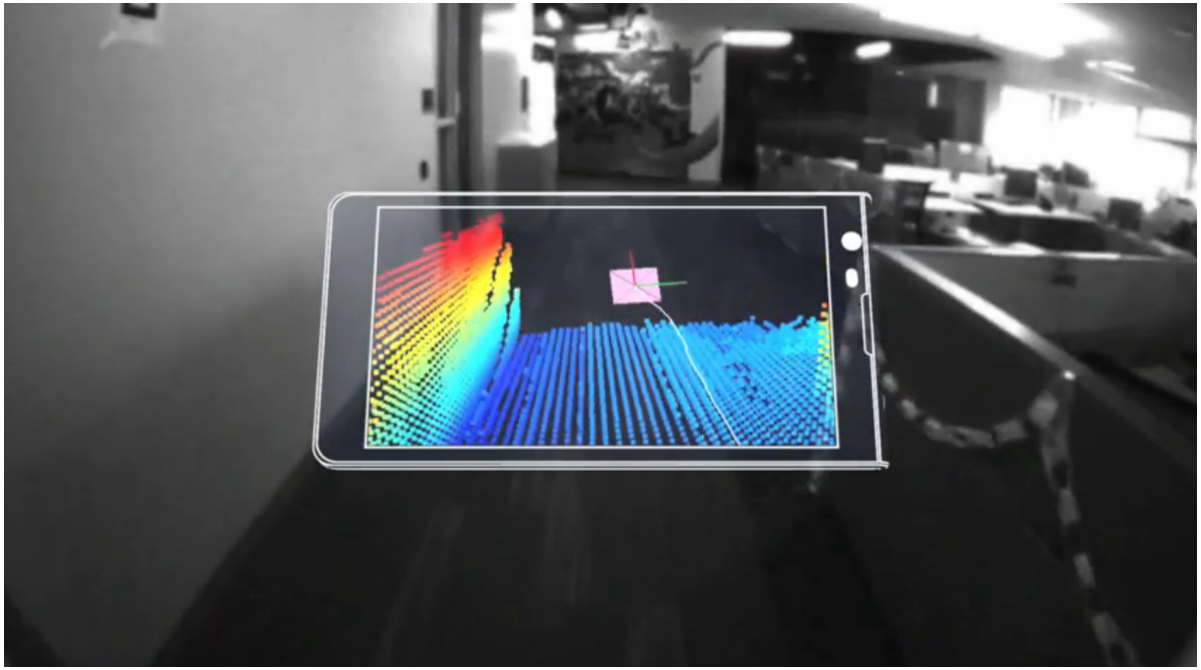
Color Regions

Structure Regions

Applications: Face detection



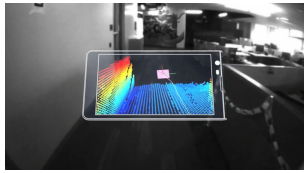
Applications: Google Tango / ARCore



https://developers.google.com/ar/

Applications: Google Tango / ARCore

Applications: Google Tango / ARCore



https://developers.google.com/ar/

Project Tango aims to merge the physical and digital worlds. Humans use visual cues to interact with their environments, Google is building a prototype smartphone that aims at replicating the same thing.

Applications: Special effects

Applications: Smart cars

https://www.youtube.com/watch?v=5DtdeOg3piY

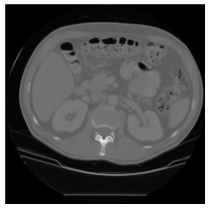
Applications: Gaming

Applications: Gaming

Techniques used in computer and video games to produce the illusion of intelligence in the behavior of non-player characters. A game must feel natural:

- Obey laws of the game
- Characters aware of the environment
- Path finding (search)
- Decision making
- Planning

Applications: Medical imaging



CT image of a patient's abdomen