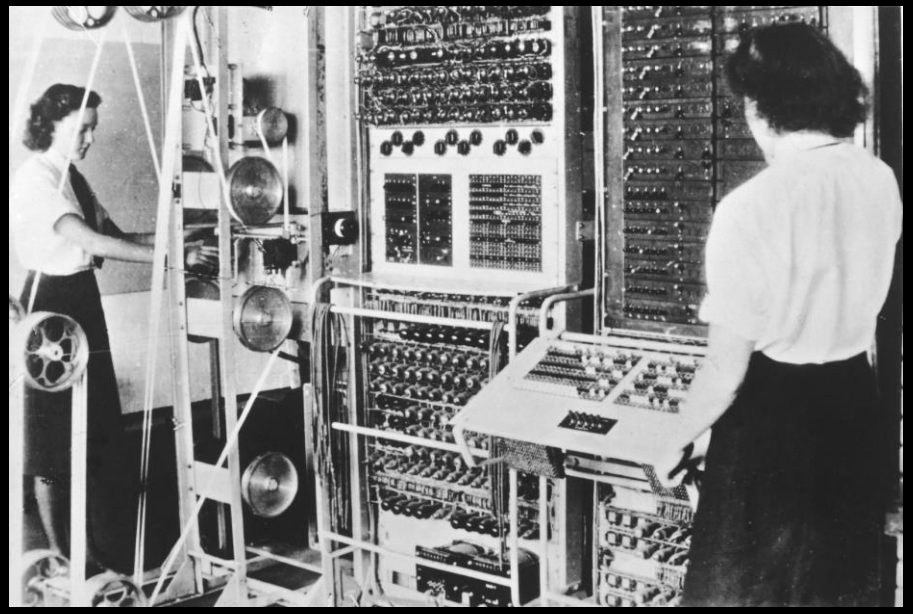
Introduction to computer vision

Jean Ponce and Elena Sizikova
jean.ponce@inria.fr, es5223@nyu.edu

Manjusha Mishra and Aryaa Singh
mam1974@nyu.edu, as13538@nyu.edu

- Slides will be available after classes
- See logistic details at the end



Description:

- Street scene
- Bar
- Chairs
- People drinking coffee
- Ashtray, etc.

Computer vision

… extracting information from images and video
(courtesy of I. Laptev)

Vision is hard—this is what the machine "sees"

- The visual cortex is about 50% of the macaque brain
- More human brain is dedicated to vision than anything else

WHY IS VISION DIFFICULT ?

Too much information:

- 1000x1000x24 bits 30 times per second;
- matching n features against n features costs n!;
- shadows, highlights, texture..

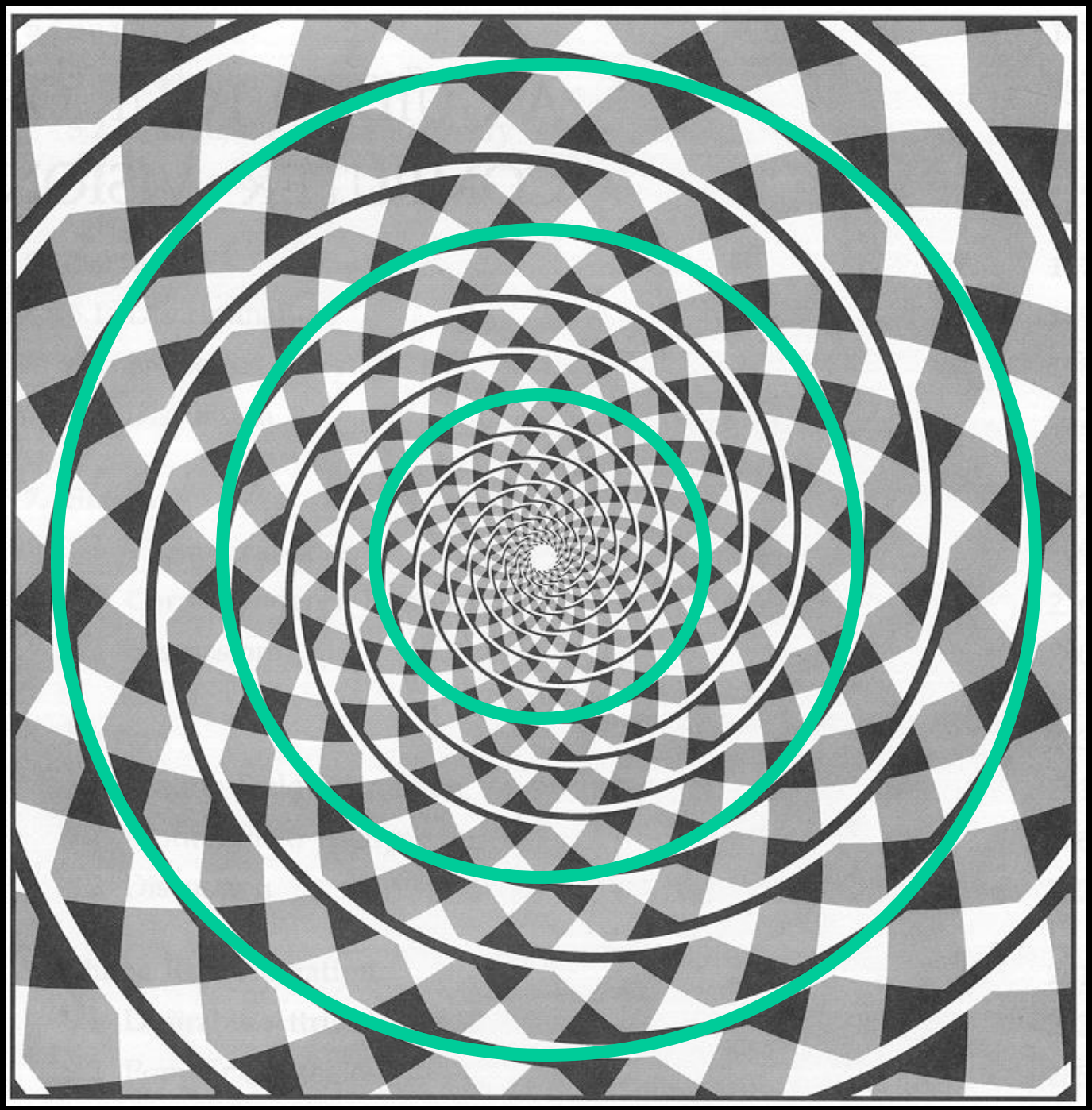
Too little information:

- Physical properties (depth, orientation, reflectance..) of the world are not directly observable.

What are appropriate models?

- of images, object instances, object classes, video content and the interpretation process..

What are appropriate algorithms and architectures?



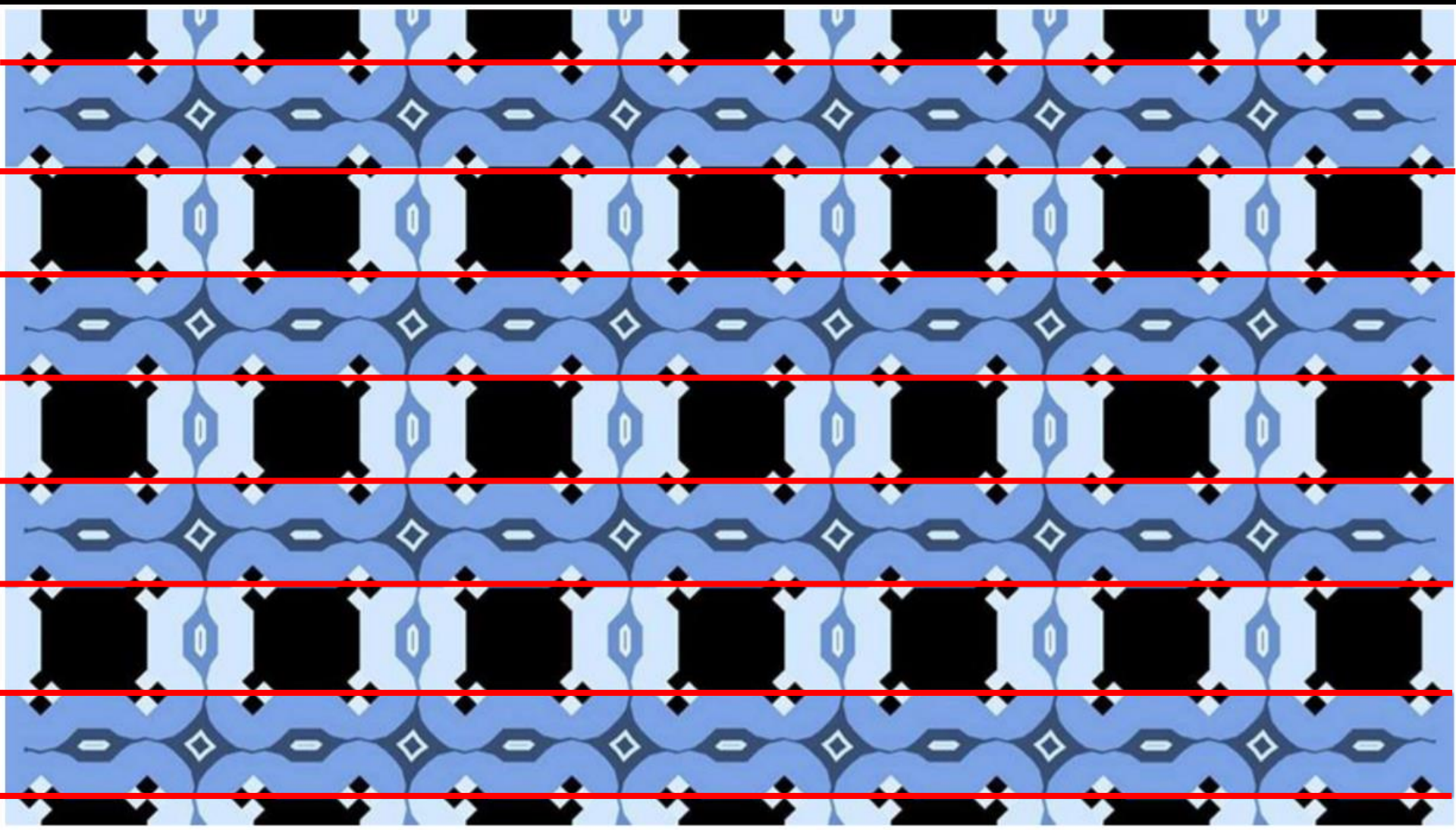
(Nalwa, 1993)

J.J. Koenderink, www.gestaltrevision.be/en/resources/clootcrans-press

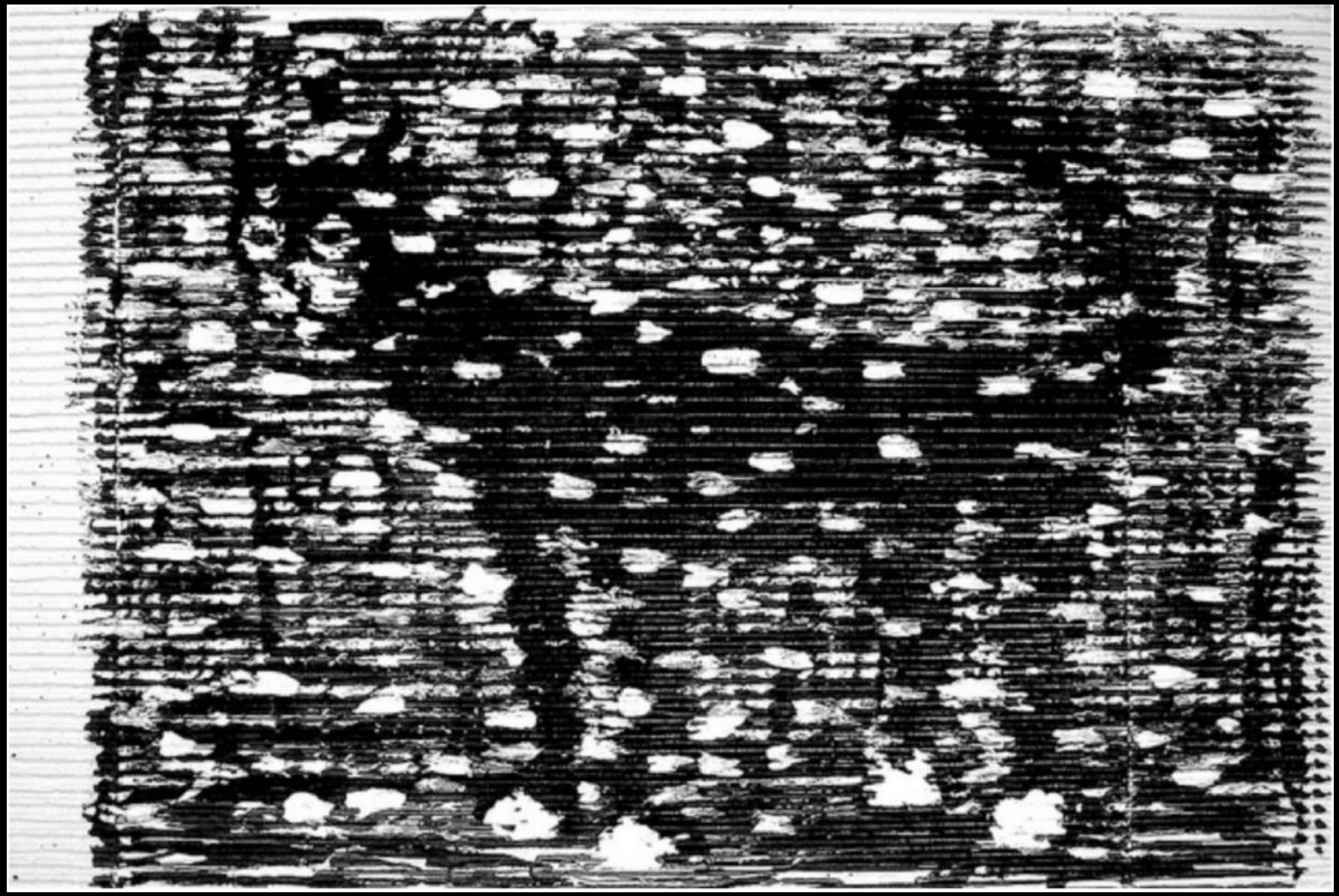
COMPUTER VISION IS INTERESTING.

- We know it is possible.
- We know it is difficult.
- We don't (really) know how to do it.

(Next three slides and many others courtesy of Jan Koenderink.)



(Victoria Skye)

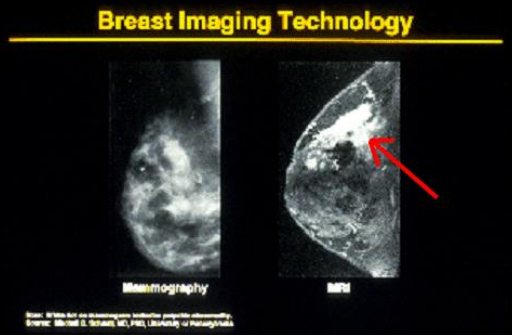


(Franco Mattichio)

(Pau Buscato)

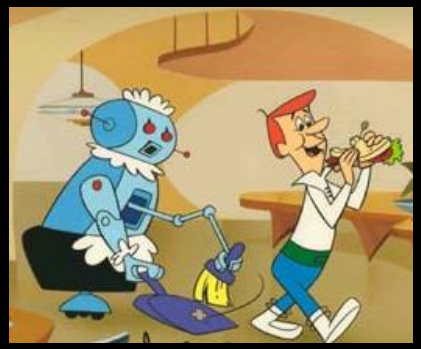
Why computer vision matters

Safety



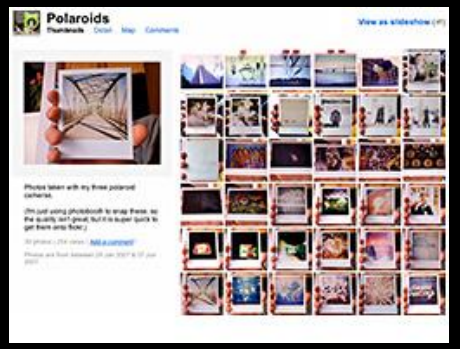
Health

Security



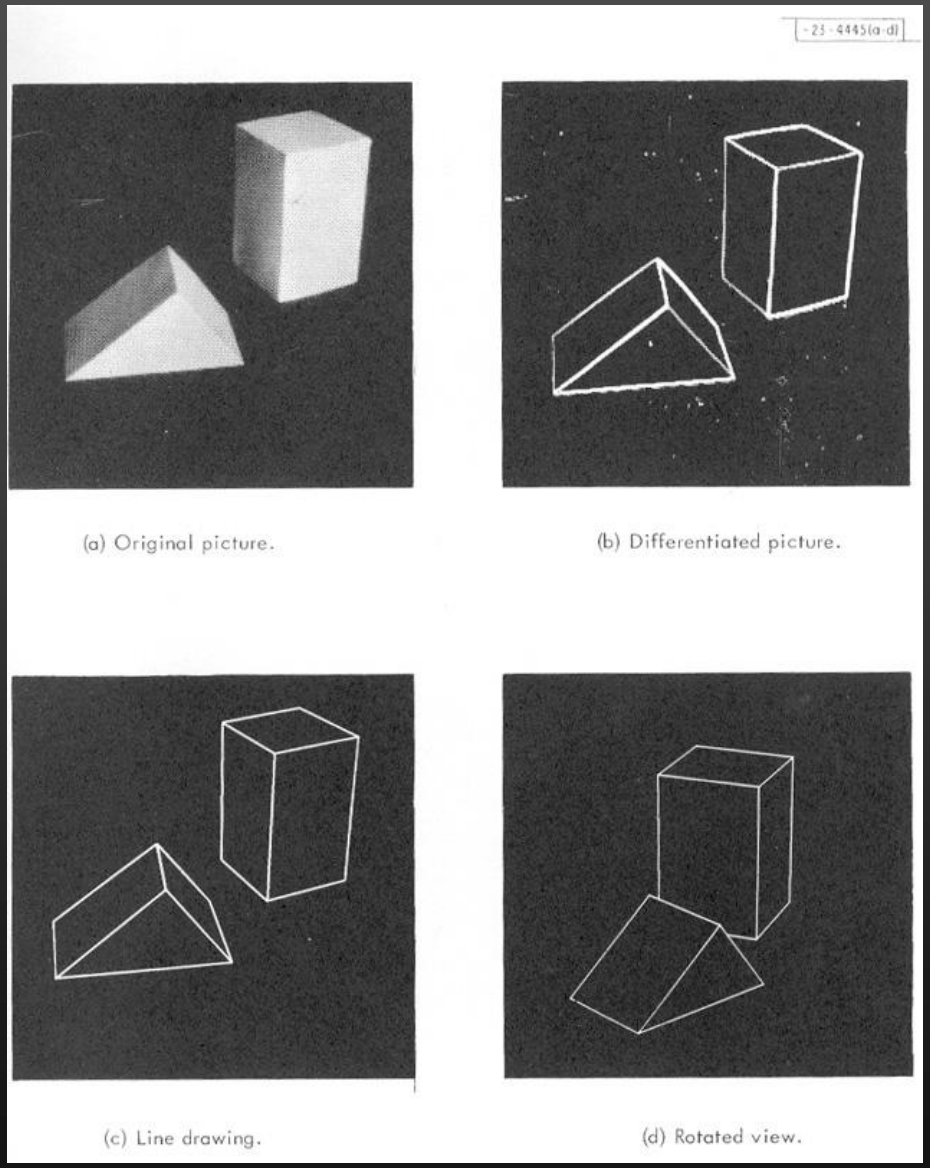
Comfort

Fun



Access

Origins of computer vision

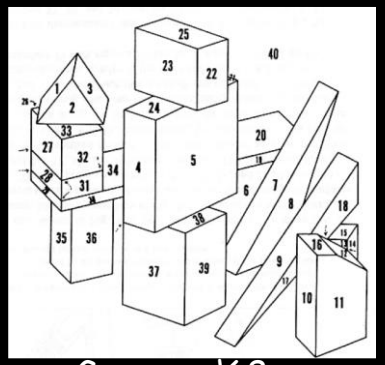


L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

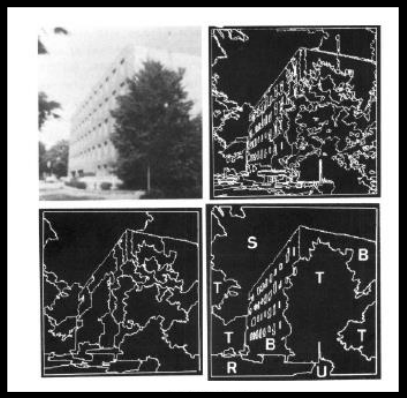
photo credit: Joe Mundy

After Roberts: a ridiculously brief history of computer vision

- 1966: Minsky assigns computer vision as an undergrad summer project (??)
- 1960's: interpretation of extremely simple images & synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2010's: Deep learning with ConvNets
- 2030's: …



Guzman '68



Ohta Kanade '78

Turk and Pentland '91

WHAT IS COMPUTER VISION GOOD FOR?

Traditionally:

- Manufacturing: inspection, bin picking;
- Defense: ATR, photogrammetry, surveillance;
- Robotics: navigation, visual servoing.

Recently:

- Computer graphics, medical imaging, HCI;
- 3D vision and recognition;
- The Web, Internet, social networks;
- Robotics again;
- And zillions of other industries.

Really:

- Understanding the principles of object recognition;
- Building the robots of tomorrow, for home and space;
- Understanding how people tick;
- It is just difficult, fun, and interesting.

Optical character recognition (OCR)

Technology to convert scanned docs to text

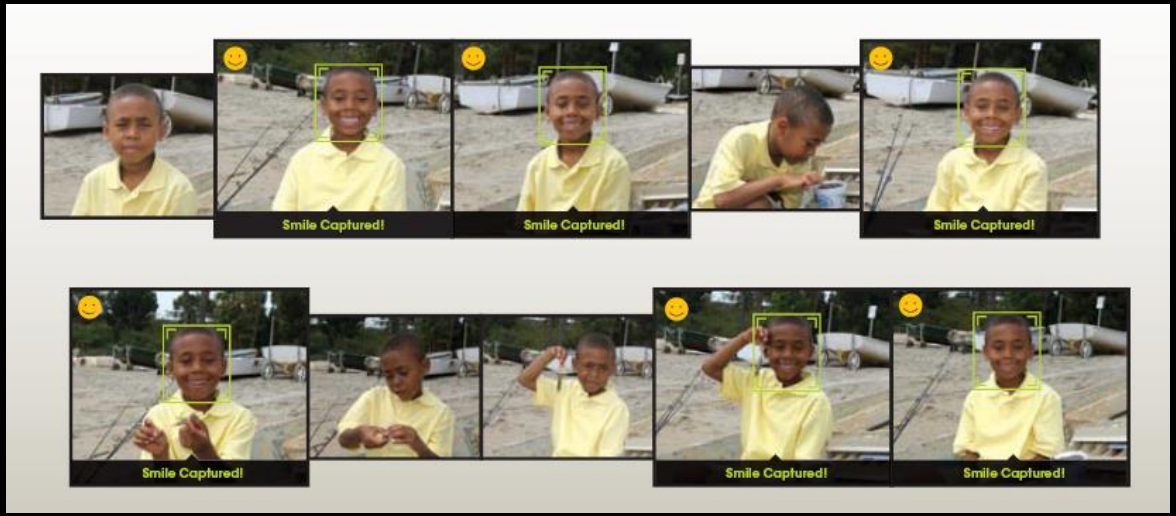
- If you have a scanner, it probably came with OCR software

Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face detection

Even smile detection

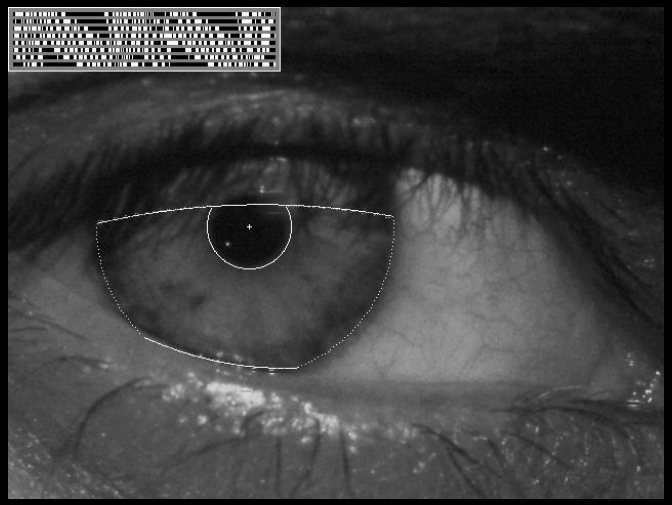


Structure from motion from busloads of images

(Agarwal et al. 2009)

Vision-based biometrics

"How the Afghan Girl was Identified by Her Iris Patterns" Read the story wikipedia



Object recognition (in mobile phones)

Point & Find, Nokia

Google Goggles

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Steve Sullivan

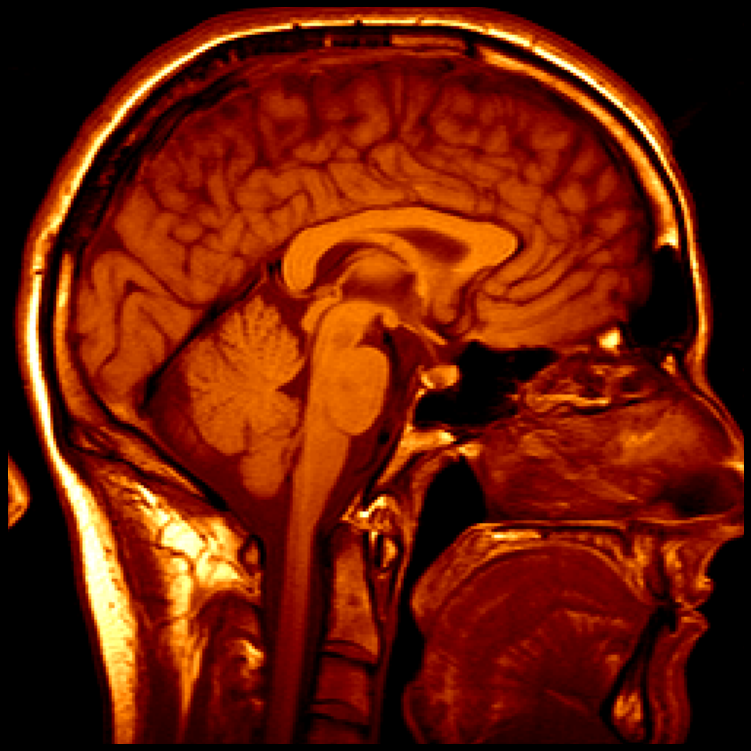
- Ph.D., UIUC, 1996
- Head of R&D, ILM, 2003
- Cover, IEEE Spectrum, 2004
- CSO, Lucasfilm, 2009-2012
- Microsoft (2013-)
- 3 Academy Awards

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

http://www.sportvision.com/video.html

Medical imaging



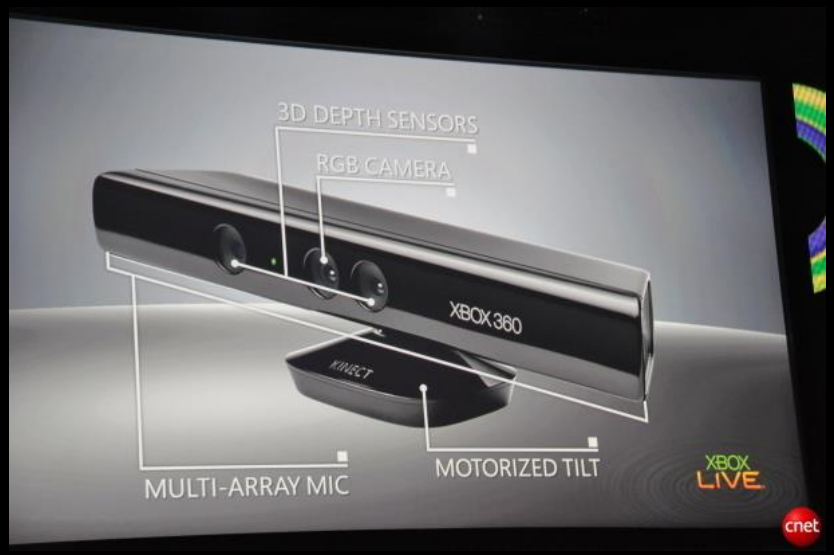
3D imaging
MRI, CT

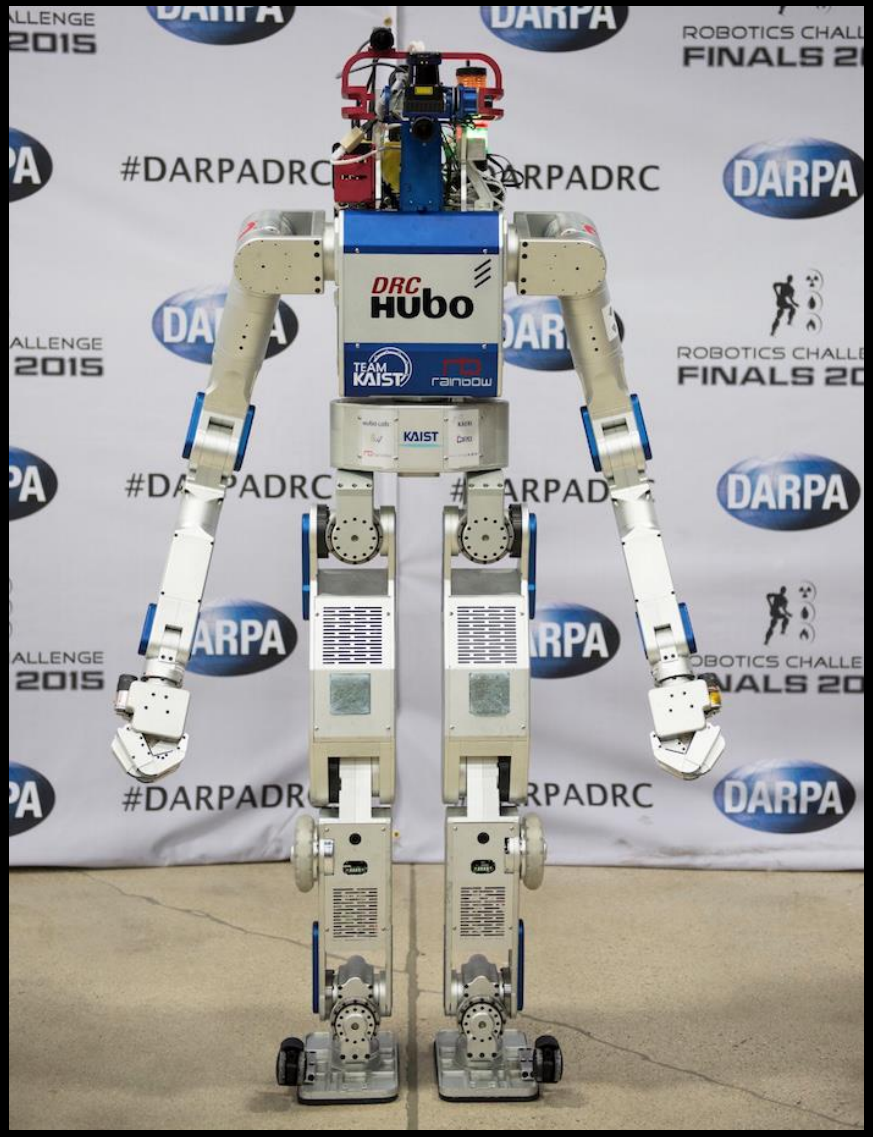
Image guided surgery
Grimson et al., MIT

The Waymo autonomous car

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY





KAIST's Hubo

And robots of course

CMU's Chimp

The SpotMini robot from Boston Dynamics

..and MetalHead from Black Mirror

The Atlas robot from Boston Dynamics (1m80, 150kg)

Vision in space

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

Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Computer vision books

- D.A. Forsyth and J. Ponce, "Computer Vision: A Modern Approach", Prentice-Hall, 2003, 2nd edition, 2011.
- J.J. Koenderink, http://www.gestaltrevision.be/en/resources/clootcrans-press
- M.F. Land and D.E. Nilsson, "Animal Eyes", Oxford University Press, 2012.
- R. Szeliski, "Computer Vision: Algorithms and Applications", Springer, 2010.
- O. Faugeras, Q.T. Luong, and T. Papadopoulo, "Geometry of Multiple Images," MIT Press, 2001.
- R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2004.

Course outline:

0. Introduction
1. Camera geometry and calibration
2. Filtering and feature detection
3. Radiometry and color
4. Texture and image segmentation
5. Stereopsis
6. Structure from motion and 3D models from images
7. Object recognition - historical perspective
8. CNNs for object classification and detection
9. 3D CNNs and applications to medical imaging
10. Weakly-supervised and unsupervised approaches to image and video interpretation

Important note:

- This class if about computer vision, not about
- deep learning, pytorch, etc.