

# Introduction to computer vision

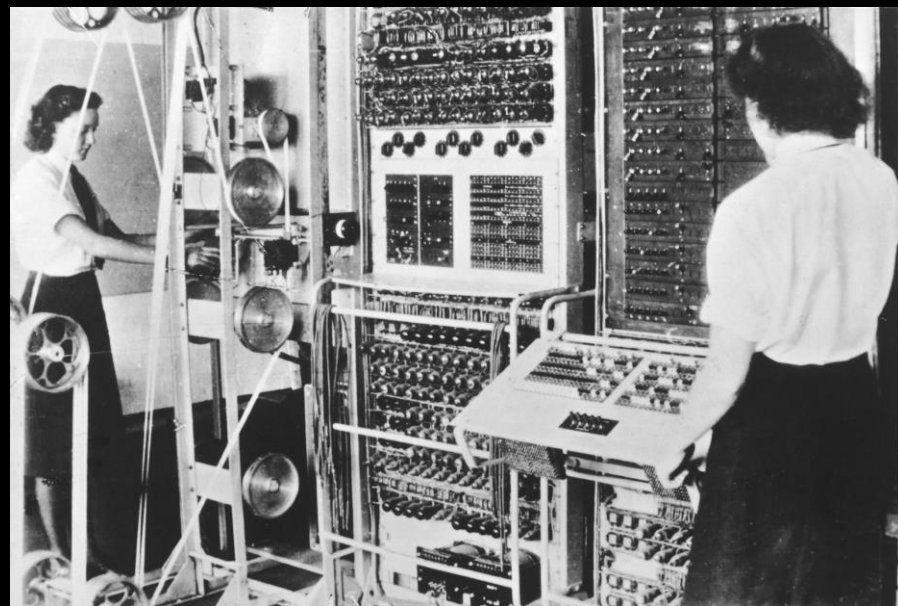
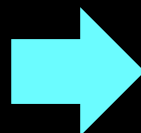
Jean Ponce and Elena Sizikova

[jean.ponce@inria.fr](mailto:jean.ponce@inria.fr), [es5223@nyu.edu](mailto:es5223@nyu.edu)

Manjusha Mishra and Aryaa Singh

[mam1974@nyu.edu](mailto:mam1974@nyu.edu), [as13538@nyu.edu](mailto: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”

08122635252121314133210507102023222326333642453122202526231614141320343524131413131328231714091819212527212744423326363323191924333640424044352  
0813324223101024483726070810182727222834374854363127292823181716144061535218151515141726110807112124293025325047343138321819263126313939494032  
09112942230507164739281118182026262630364853544234293027211816161533595448171415151414191409070815252529253754473536434024222531243135324540372  
0909264139404649503929212928272727283440505557301620262722191716151727292014151515151415201308091225262822345146343640321720253222335334436311  
09102343545752454839301412121528282835374752552915182527273226212020181818181918171817171814091013232628223654442735453622232532283332385341401  
0709183945183122384032140408132828283133394652351618252926333026252323232326282931312824131312082227272236484027353929162224323340434748441  
05081538421427223037311705091228303130343945487117142226243033302927252627262830323533292212171517263024203647382642453120252631243234454538280  
08121433381714112636322105091027303129384042403110121924251713192624252930262627283229231613161623312923183240332237402716232530253331465448340  
22251628314170112036312005060926282824434036301608101719171204172724223524100813112835140612171823303739303337251843442619262729243331394142270  
23261923312528091531292205050922191318272929181413172221191503182425162517050912153538150508192321215771272936342938402217232828213131535042180  
18191819281603071123201506051021110318232319161314161618191504111823093114000615122831100404233129222628232933293546412120252927172927424747190  
05182414181606091223161411170222150315131513121209121212161207061122142915040210072333100404211935252422212223243343402617172733342933473833100  
05132513161614171918161507131619180406131607091515161721201711071031132817020410092734110508111024272927252122203450393741261719364142504837100  
09102615151211141515161506111413140710181918182318101726181921111332081410040709093429070508030112231907251826263148403638303228181736363630070  
11082417141309121313141412151719161109101010121508041322201917101635110505070806083313040722230804131506211830303338303535252326313340463821050  
57222117091417221622211625232020192014132019100500000307140602051313040304060605153614091114171904020205181215042824212632312624233635364022050  
48241320091319351719191524272425262922182212030201030306090505081306020303040405121412141004050802090804071111081711142330282929263739363219301  
08041022091312352328210907101410060509121109040708090907060805060707090806050504070703060607050604101610040407685103072134332722213237383617363  
05040413060708342219080301000202050508131107131914110810071209020408131210111110080407080825000508041519050211856703032546443311162631343116362  
02050803030407100911040002010000060508172238161104020101020304050609111111101108020102071010010709051713030403201205043129233312122425272118372  
08050503040505030405040000000404000203071119030503010202030305060708090911101111030101021315030913041411020403072006033322131712142725241616262  
0502010304151403060813120806111404010100104050804020203040508060707080913101117140500021507011220041410020203061804042912081511202522241416232  
04040304041816132019131317251214150302030306050504040404050606060707101918121219261101010901011623070802030203071303052207050506162308151021242  
05060810101517050602010309181313190703030406060606050405050506071016263020141215382606030200001517020203051417121103051014120706202024101214262  
141717140914050214010203050605031212010204060606060605060605071223363823111219373318090503060905020406051619150702020304090205231603061424333  
10141224091503041004040505030201040302020305060606060606050607181020344729171723353520100606070709050506071217060300010101000006210305121335623  
13090707081006040303020203080400010303030305050606050812050606070202233730181725475025090606061215070606091107000302010202010002060308160870896  
151409070403020101010101080400020304050404050607050504030608020311291020191725444319090606071214120807161708010101020202010000000108131162976  
21090302010101010000000000010102030405060606060606091110304070903040203071719222723161311090908101211082535160600000202010000000000506152533  
080212070000000000000010203030405050506070708080807081732230408290800040608111826354134251711100906070909294219110400020201010202020100000610050  
0102050301010101020302030404060606070808080809090704070818010413110002163424214644443727201511090705070728442415080203010001010202010010101071  
02060204040305081207050304050606070809091009090906040302020201051601051713091044494536312419131108050605184323261407040100000000010101000101010  
000209110704071113090406100606070708080910101010090503030202020616151302121004475749413528221615130605051339212824100904030200000000000000101000  
00000306050604040203020516070707070808080910101008060202010103180426350204001954535144372923181216090405092933262612090702020101000000010100000  
00000104050603000002021814050606060607070809100906060302020101270903040101012753494847393123181210150705041533362208130704020101000000010201000  
00000001020401000001001409020303030404040405060704040202010101130601000000010936373535292118130906080703040916301604130504030101000000000501000

- 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:

- $1000 \times 1000 \times 24$  bits 30 times per second;
- matching  $n$  features against  $n$  features costs  $n!$ ;
- shadows, highlights, texture..

## Too little information:

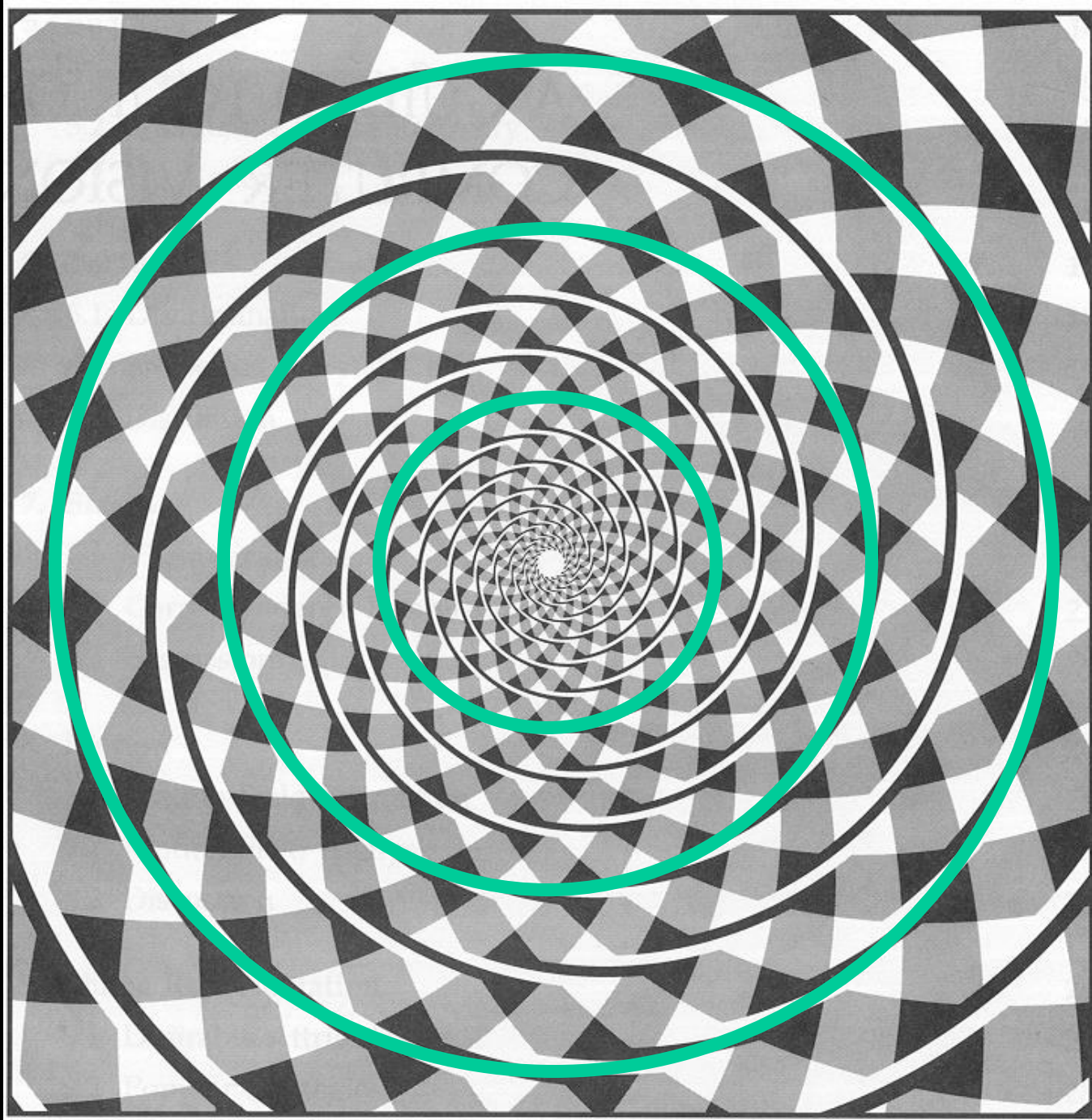
- Physical properties (depth, orientation, reflectance..) of the world are not directly observable.

## What are appropriate models?

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

## What are appropriate algorithms and architectures?





(Nalwa, 1993)



<http://go.funpic.hu>

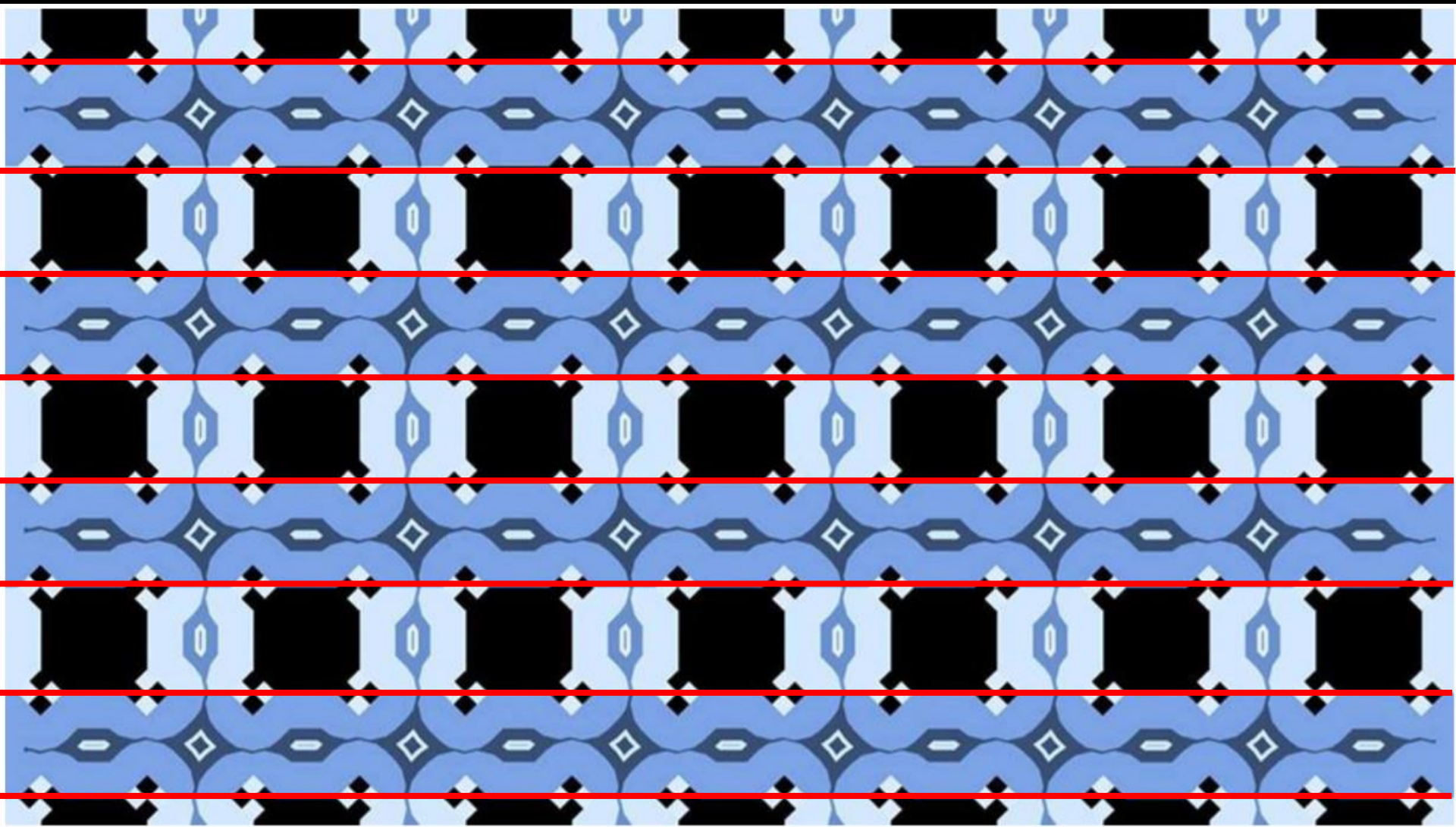
J.J. Koenderink, [www.gestaltrevision.be/en/resources/cloutcrans-press](http://www.gestaltrevision.be/en/resources/cloutcrans-press)



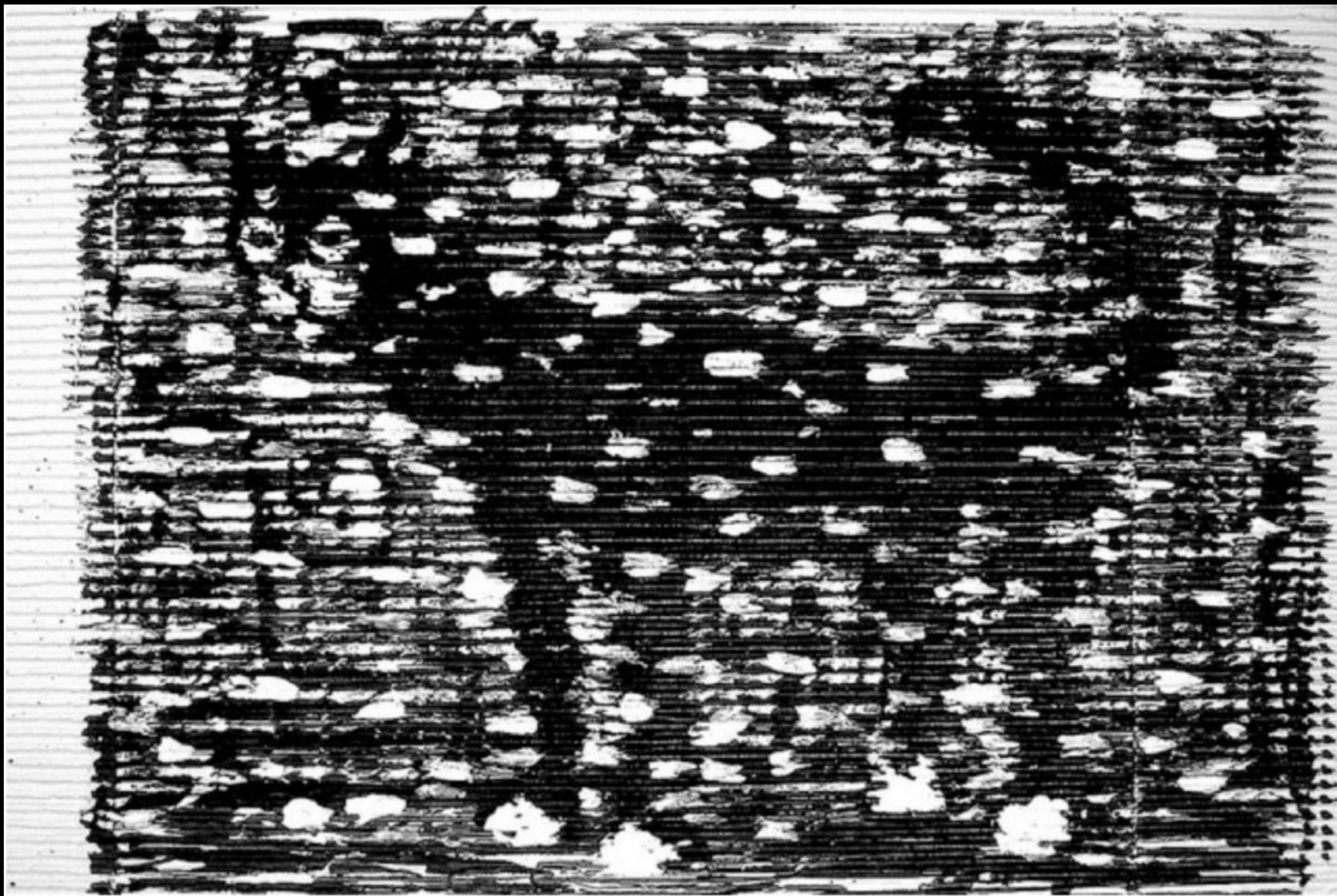
# COMPUTER VISION IS INTERESTING.

- 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)



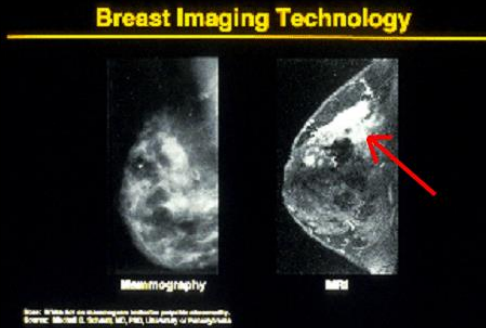


(Pau Buscato)

# Why computer vision matters



Safety



Health



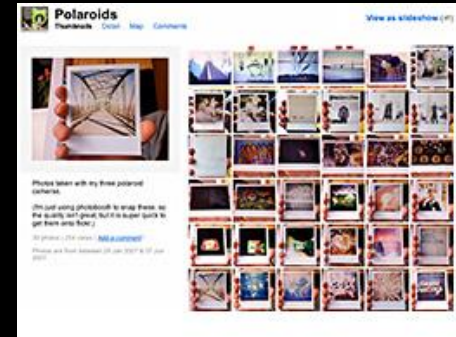
Security



Comfort

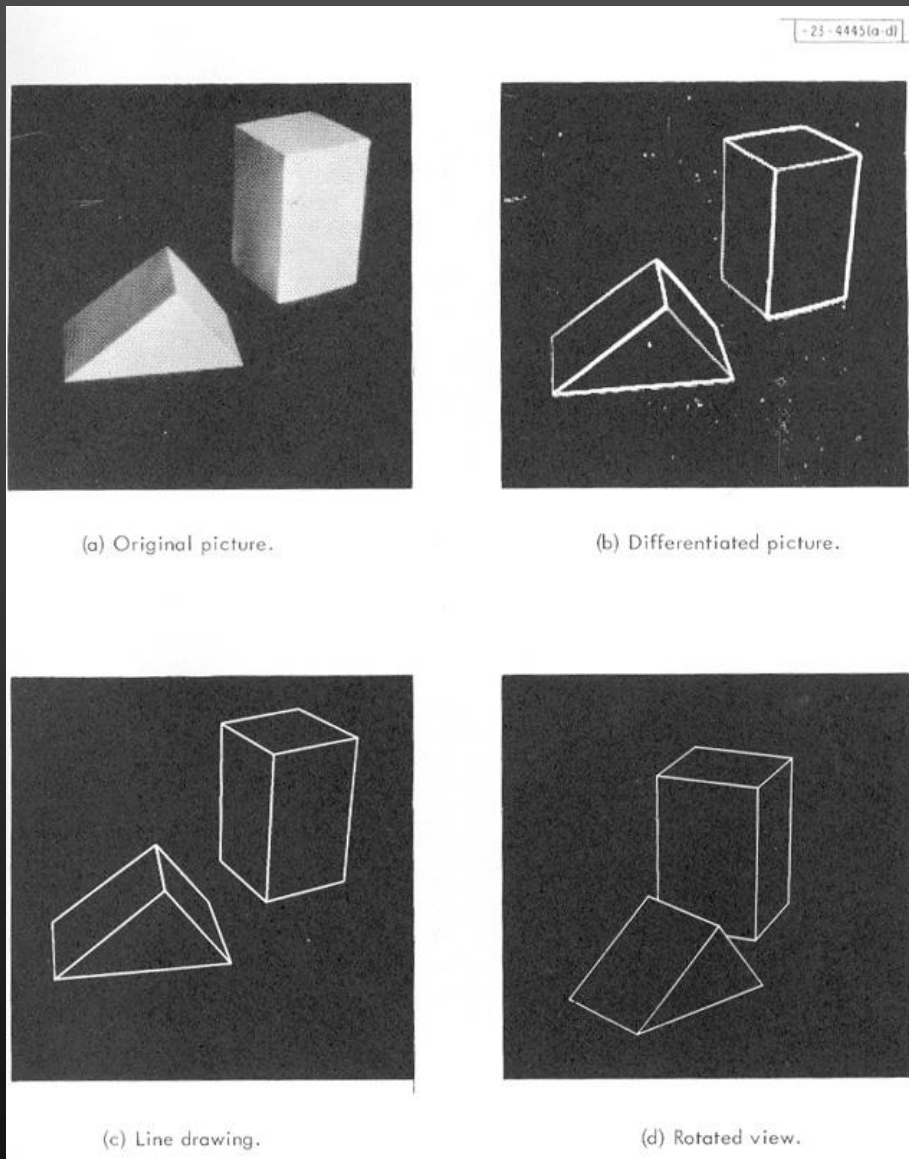


Fun



Access

# Origins of computer vision

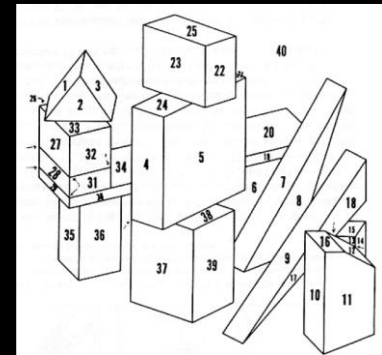


*L. G. Roberts, Machine Perception of Three Dimensional Solids, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.*

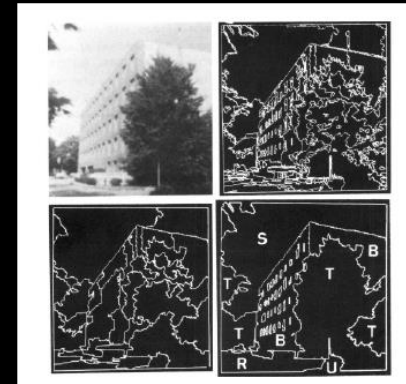


# After Roberts: a ridiculously brief history of computer vision

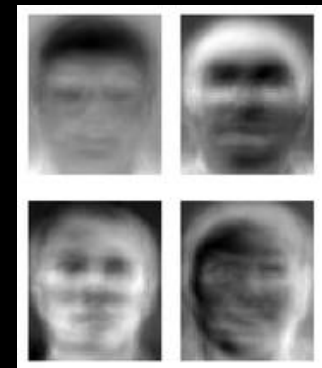
- 1966: Minsky assigns computer vision as an undergrad summer project (??)
- 1960's: interpretation of extremely simple images & synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; 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.

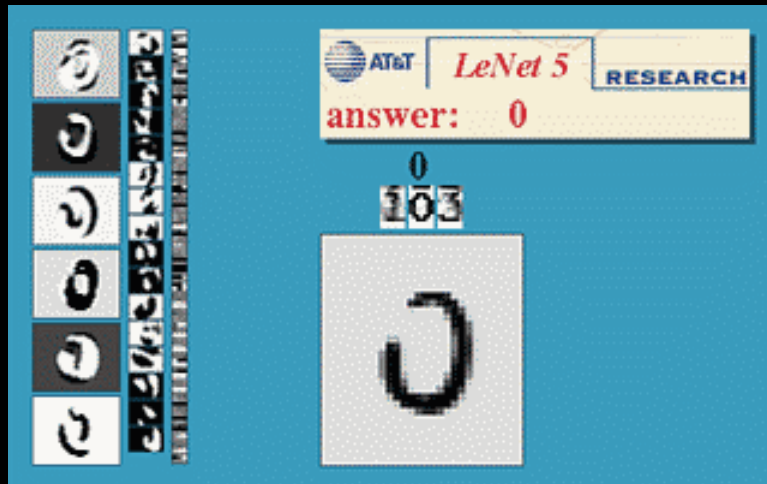
## Really:

- 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

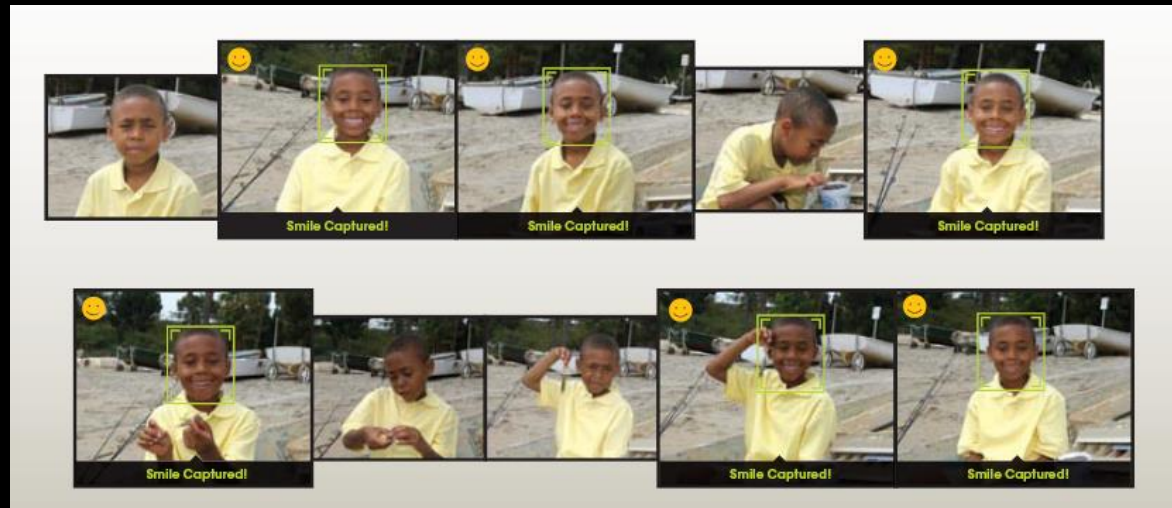
[http://en.wikipedia.org/wiki/Automatic\\_number\\_plate\\_recognition](http://en.wikipedia.org/wiki/Automatic_number_plate_recognition)



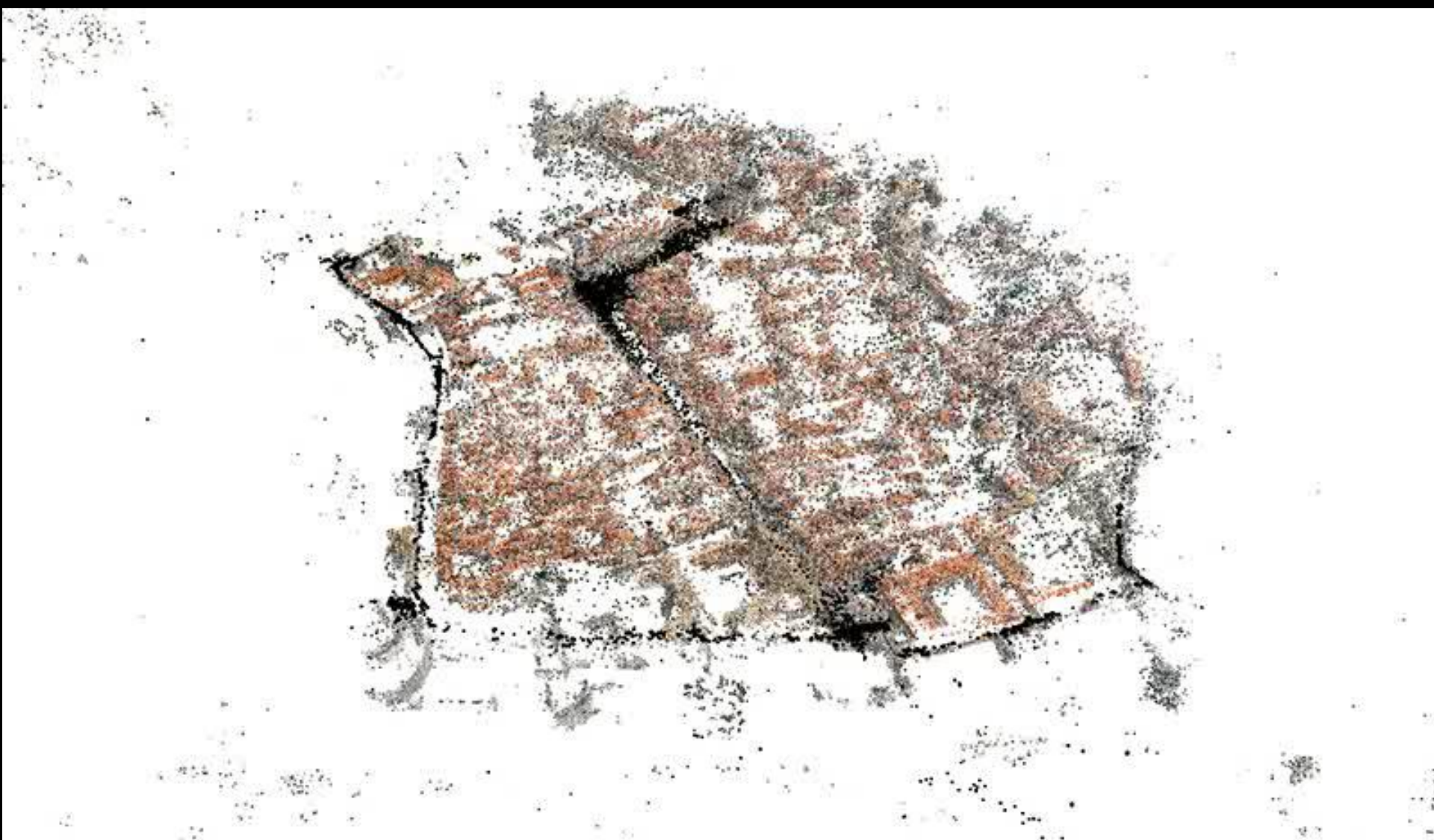
# Face detection



## Even smile detection



# Structure from motion from busloads of images

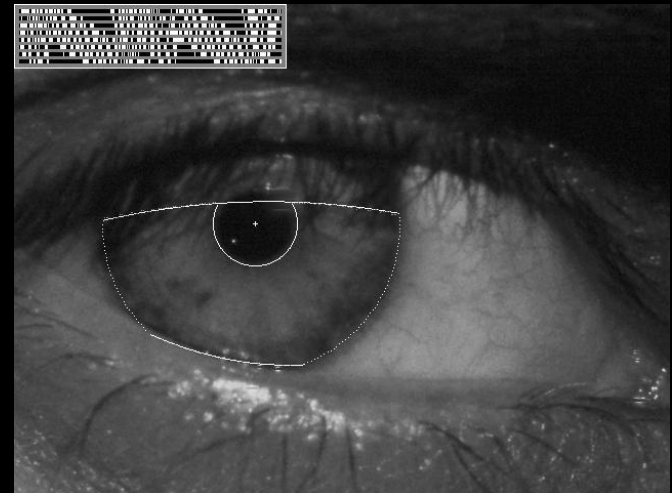


(Agarwal et al. 2009)

# 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 Caribbean, 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

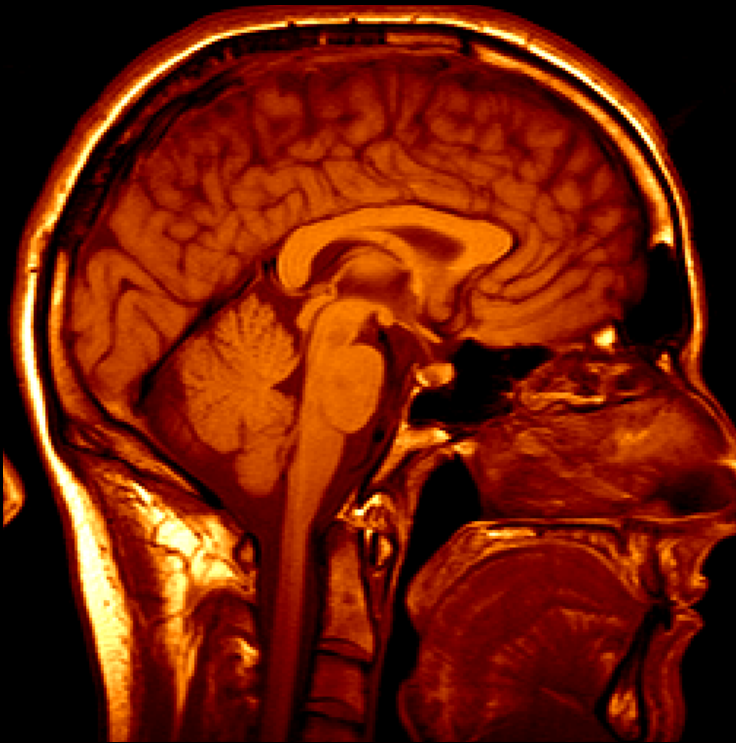


*Sportvision first down line*

Nice [explanation](http://www.howstuffworks.com) on [www.howstuffworks.com](http://www.howstuffworks.com)

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

# Medical imaging



3D imaging  
MRI, CT



Image guided surgery  
[Grimson et al., MIT](#)

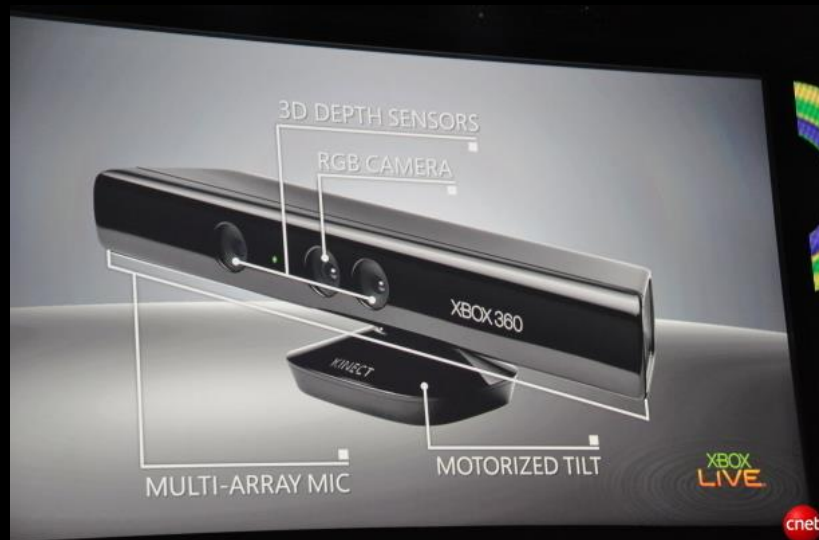


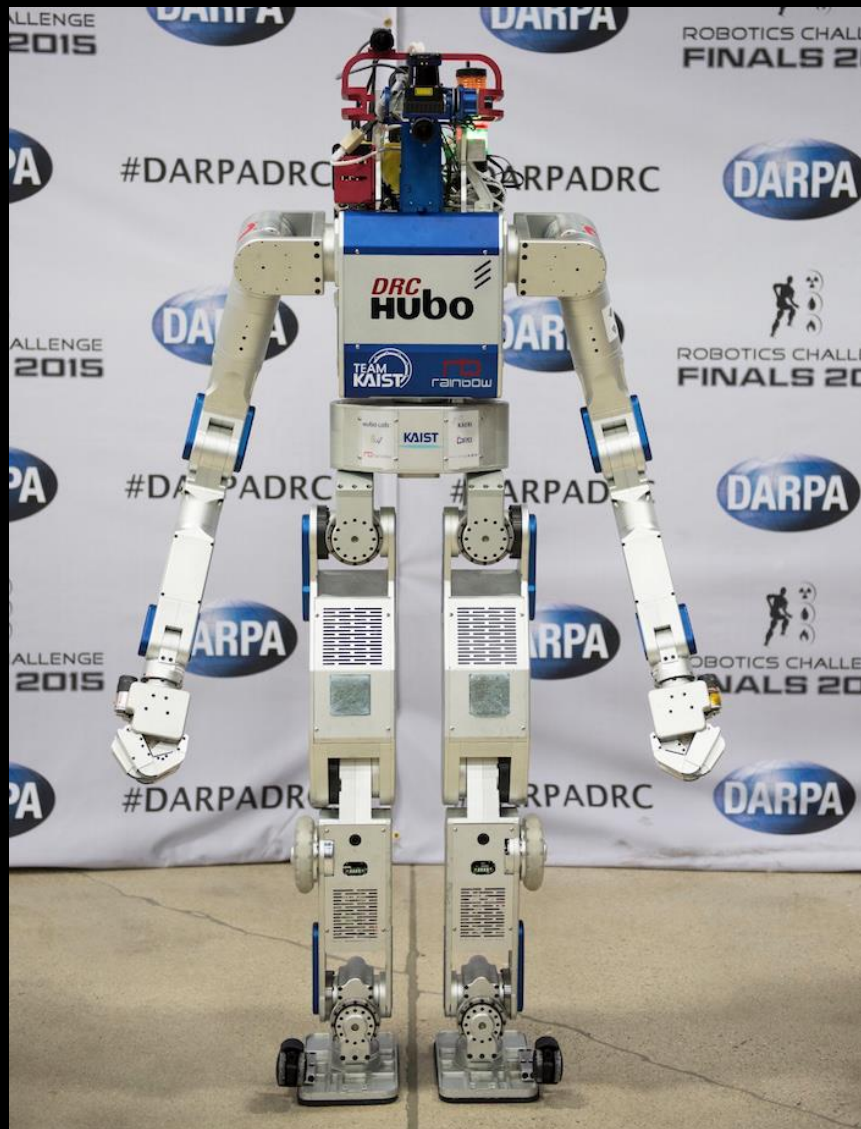


The Waymo autonomous car

# Interactive Games: Kinect

- Object Recognition:  
<http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o>
- Mario: <http://www.youtube.com/watch?v=8CTJL5IUjHg>
- 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, 2<sup>nd</sup> edition, 2011.
- J.J. Koenderink,  
<http://www.gestaltrevision.be/en/resources/cloutcrans-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. Papadopoulos, "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 is about computer vision, **not** about
- deep learning, pytorch, etc.