

Introduction à la vision artificielle



Jean Ponce and Josef Sivic

Email: Jean.Ponce@ens.fr and Josef.Sivic@ens.fr

Web: <http://www.di.ens.fr/willow/>

Planches après les cours sur :

<http://www.di.ens.fr/~ponce/introvis/lect1.pptx>

<http://www.di.ens.fr/~ponce/introvis/lect1.pdf>

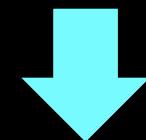
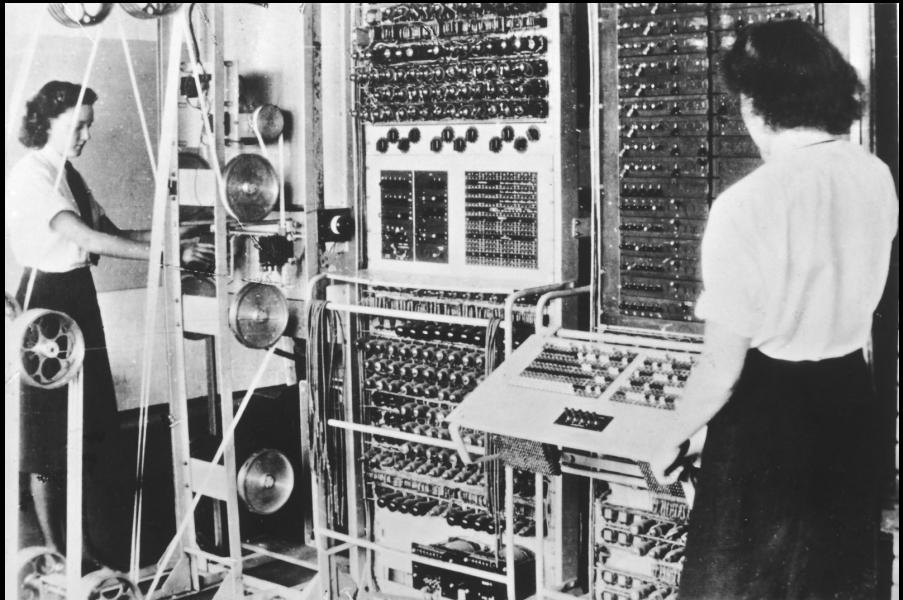
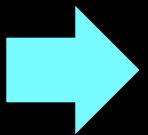
Today with some slides from J. Hays and S. Seitz

Change of time for the class

From next week the class is on:

Wednesdays 14:00-17:00 in the same room S16

Next class: Wednesday 21/9 14:00-17:00 in S16



Description:

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

Computer vision

... extracting information from images and video



Vision is really hard

- 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



Vision is really hard

081226352521213141332105071020232223263336424531222025262316141413203435241314131328231714091819212527212744423326363323191924333640424044352
081332422310102448372607081018272722283437495436312729282318171614406153521815151417261108071121242930253250473431383218192631263139394940322
09112942230507164739281118182026262630364853544234293027211816161533595448171415151414191409070815252529253754473536434024222531243135324540372
09092641394046495039292129282727272834406055573016202627221917161517272920141515151514152013080912252628223461463436403217202532223335334436311
09102343545752454839601412121528282835374752552915182527273226212020181818181918171817171814091013232628223654442735453622232532283332385341401
0709183945183122384032140408132828283133946523516182529263330262523232323282829313128241313120822272722364840273539291622243233404343748441
05081538421427223037311705091228303130343945487117142225243033302927252627262830323533292212171517263024203647382642453120252631243234454538280
08121433381714112636322105091027303129384042403110121924251713192624252930252627283229231613161823312923183240332237402716232530253331465448340
2225162831417011203631200506092628282443403630160810171917120417272422362410081311283514061217182330873960333725184344261926272924331394142270
23261923312528091531292205050922191318272929181413172221191603182425162517050912153538150506192321215771272936342938402217232828213131536042180
1819181928160307112320150605102111031823231916131416161819150411182309311400061512283110040423312922628232933293546412120252927172927424747190
0518241418160609122316141117202215031513151312120912121216120706112214291504021007233310040421193525242221223248343402617172733342933473833100
0513251316181417191816150713161918040613160709151518172120171071031132817020410092734110506111024272927252122203450393741261719364142504837100
09102615151211141516161506111413140710181918182318101728181921111332081410040709093429070508030112231907251826263148403638303228181736363630070
11082417141309121313141412151719161109101010121508041322201917101635110505070805083313040722230804131506211830303338303635252326313340453821050
57222117091417221622211625232020192014132019100500000307140602051313040304080605153614091114171904020205181215042824212632312624233635364022050
4824132009131935171919152427242526292218221203020103030609050508130602020303040405121412141004050802090804071111081711142330282929263739363219301
080401022091312352328210907101410060509121109040708090907060805060707090806050504070707030606070506041016100404076851030721343327221323738617363
0504041306070834221908030100020205050813110713191411081007120902040813121011110080407080825000508041519050211856703032546443311162631343116362
02050803030407100911040002010000060508172238161104020101020304050609111111101108020102071010010709051713030403201205043129233312122425272118372
0805050304050304050400000040400020307119030503010202030305060708090911101111030101021315030913041411020403072005033322131712142725241616262
05020103041514030608131208061114040101000104050804020203040506060707080913101117140500021507011220041410020203061804042912081511202522241416232
04040304041616132019131317251214150302030306050504040404050606060707101918121219261101010901011623070802030203071303052207050506162308151021242
050608101015170508020103091813131907030304060606060506070101626302014121538260603020000151702023051417121103051014120706202204101124262
14171714091405021401020305060503121201020406060606050605071223363823111219373316090503060905020406051619150702020304090205231603061424333
101412240915030410040405050302010403020203050606060606060606060718102034472917172335352010060607070905050507121706030001010100006210305121336623
130907070810060403030202030804000103030305050606050812050606070202233730181725475025090605061215070606091107000302010202010002060308160870996
1514090704030201010101010108040002030405040405060705050403060602031129102019172544431909060607121412080716170801010102020100000000108131162976
210903020101010100000000000010102030405060606060609111103040709030402030717192227231613110909081012110825351606000002020100000000005061525331
08021207000000000000000102030304050505060707080808070817322304082908000406081118253541342517111009060709092942191104000202010102020100000610050
01020503010101010203020304040606060708080808090907040708180104131100021634242148444437272015110907050707284424160802030100010102020100010101071
020602040403050812070503040506060708090910090909060403020202010516010517130910444945363124191311080506051843232614070401000000000010101000101010
000209110704071113090406100606070708080910101009050303020202061615130212100447574941352822161513080505133921282410090403020000000000000101000
000003050506040402030205160707070808080910101008060202010103180426350204001954635144372923181216090405092933262612090702020101000000010100000
0000010405060300000201614050606060607070809100906060302020101270903040101012753494847393123181210150705041533362208130704020101000000010201000
00000001020401000001001409020303030404040405060704040202010101130601000000010936373535292116130906080703040916301604130504030101000000000501000



WHY IS VISION DIFFICULT ?

Too much information:

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

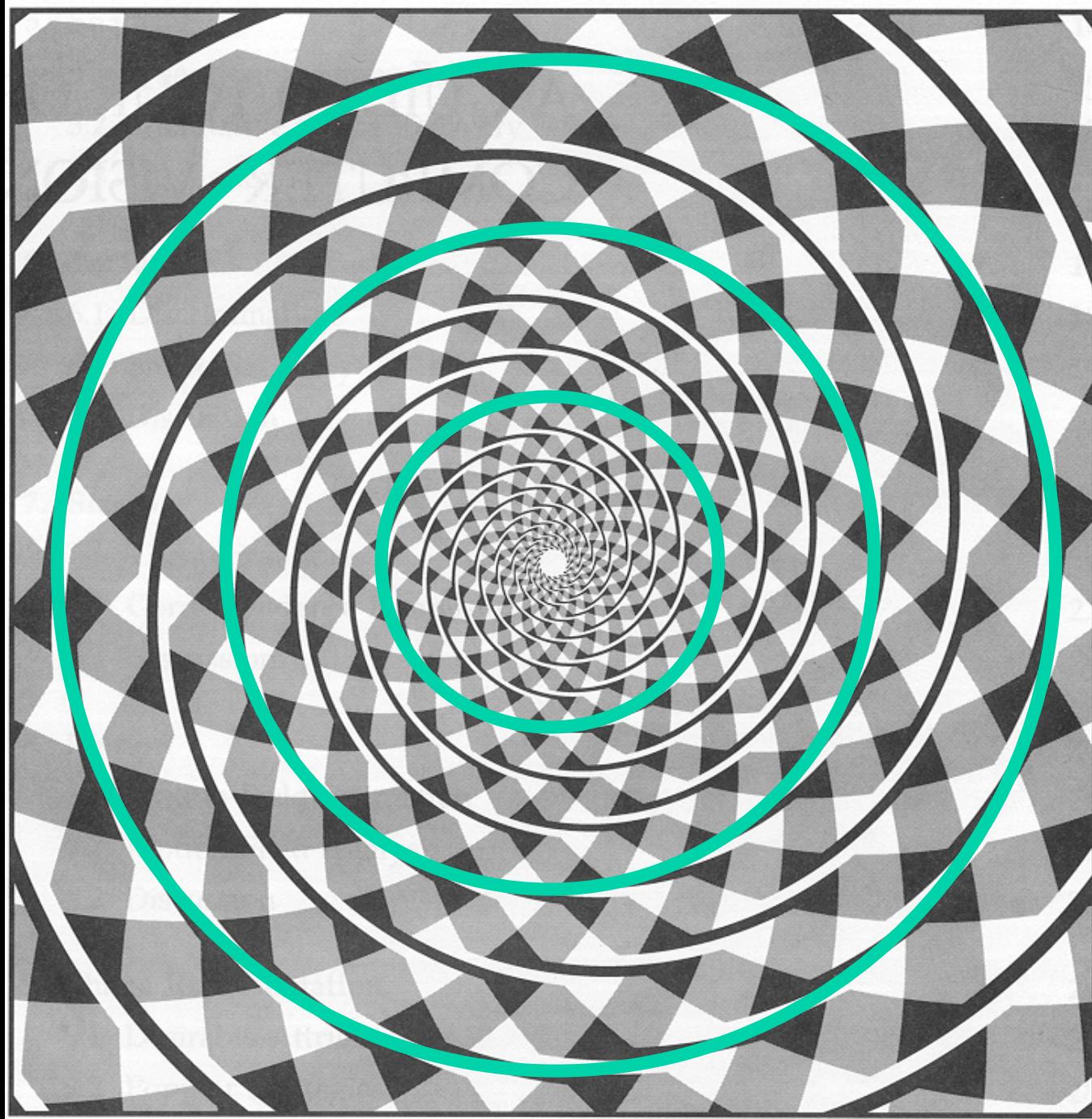
Too little information:

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

What are appropriate representations?

- for 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/clootcrans-press

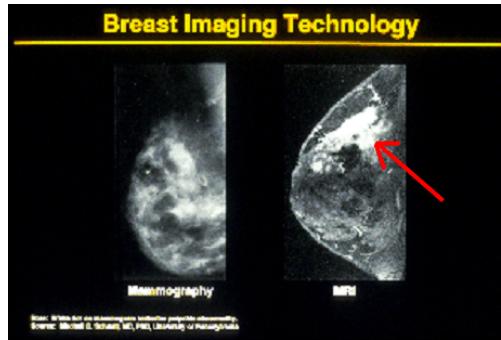
COMPUTER VISION IS **INTERESTING**.

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

Why computer vision matters



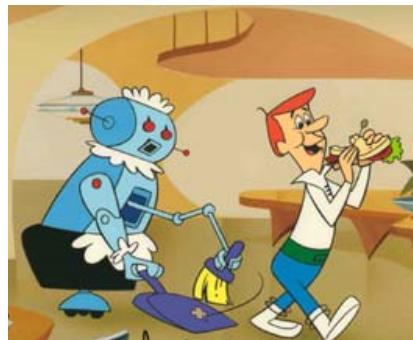
Safety



Health



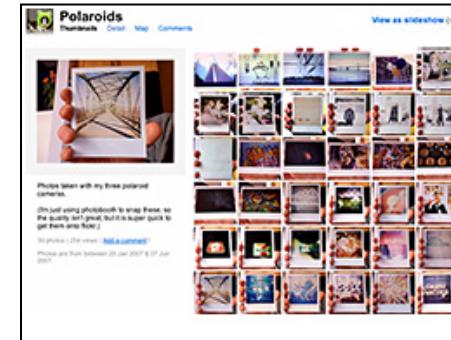
Security



Comfort



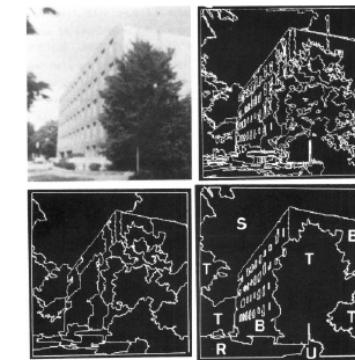
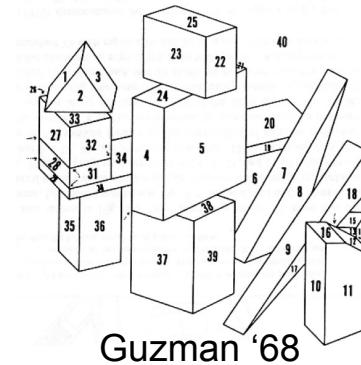
Fun



Access

Ridiculously brief history of computer vision

- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of 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: ...



Turk and Pentland '91

WHAT IS COMPUTER VISION GOOD FOR?

Traditionally:

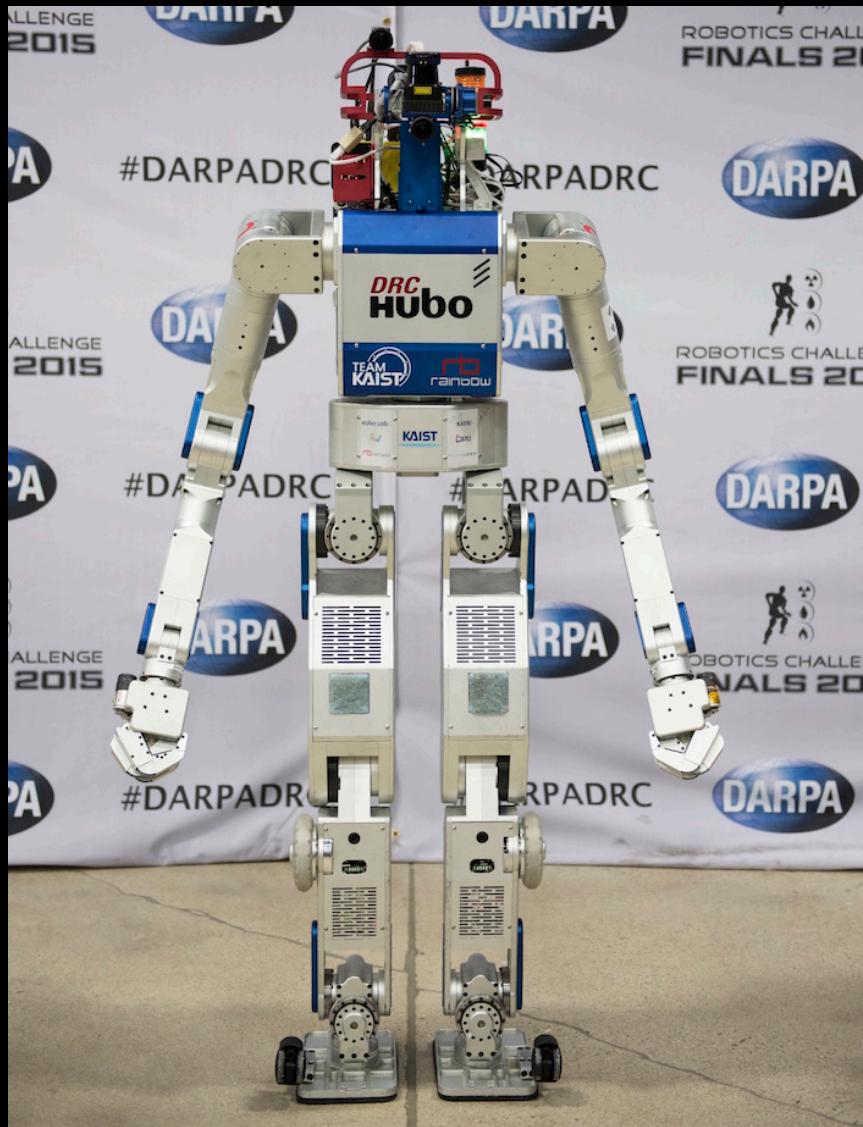
- Manufacturing: inspection, bin picking;
- Defense: ATR, photogrammetry, surveillance;
- Robotics: navigation, visual servoing.

Recently:

- Computer graphics, medical imaging, HCI
- 3D vision and recognition
- The Web, Internet, social networks.
- Robotics again.

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.



KAIST's Hubo



CMU's Chimp

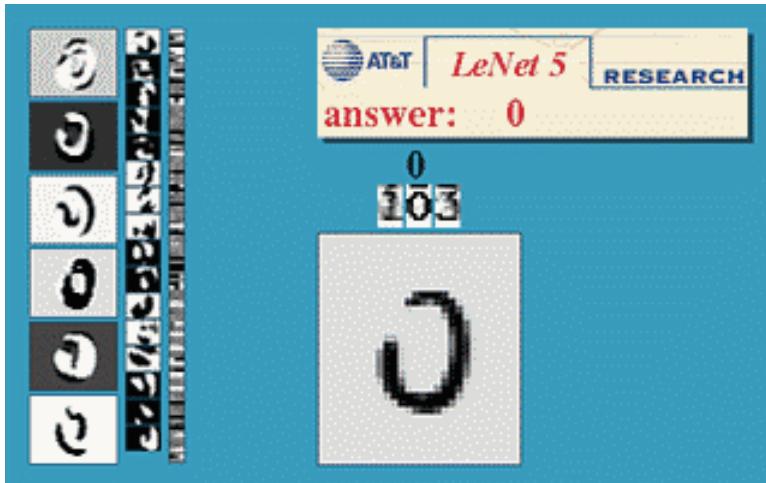
How vision is used now

- Examples of recent real world applications

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

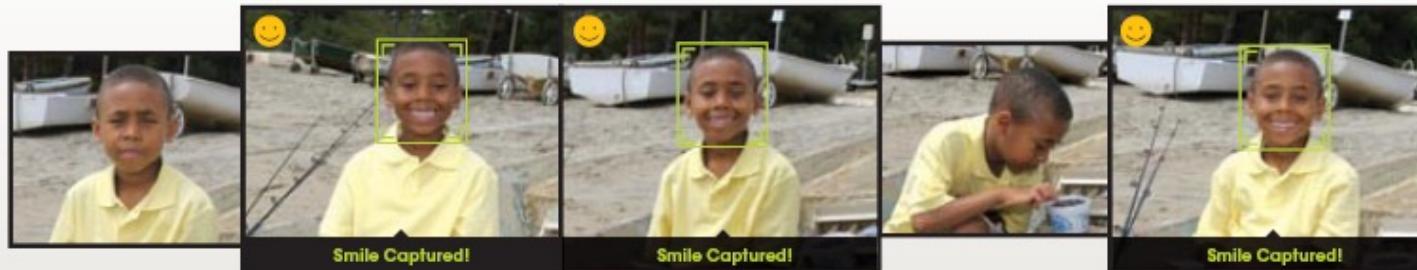


- Many new digital cameras now detect faces
 - Canon, Sony, Fuji, ...

Smile detection

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



[Sony Cyber-shot® T70 Digital Still Camera](#)

3D from thousands of images



Agarwal et al. 2009

Object recognition (in supermarkets)



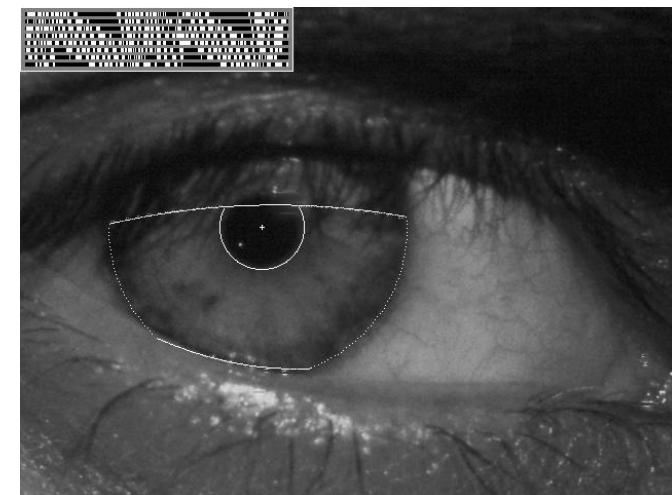
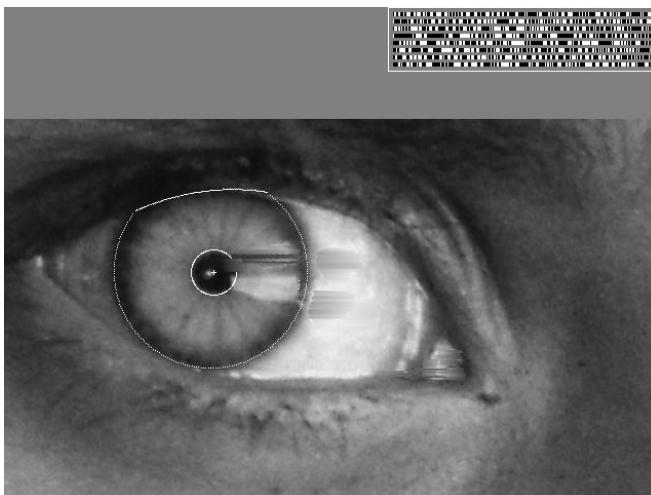
[LaneHawk by EvolutionRobotics](#)

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk, you are assured to get paid for it... “

Vision-based biometrics



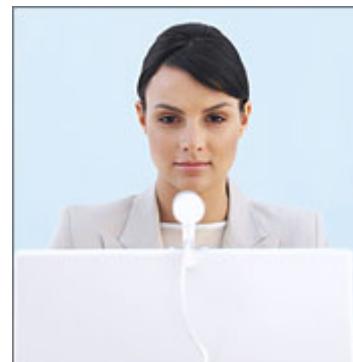
“How the Afghan Girl was Identified by Her Iris Patterns” Read the [story](#)
[wikipedia](#)



Login without a password...



Fingerprint scanners on
many new laptops,
other devices



Face recognition systems now
beginning to appear more widely
<http://www.sensiblevision.com/>

Object recognition (in mobile phones)



Point & Find, Nokia
Google Goggles

Special effects: shape capture



The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture



Pirates of the Caribbean, Industrial Light and Magic

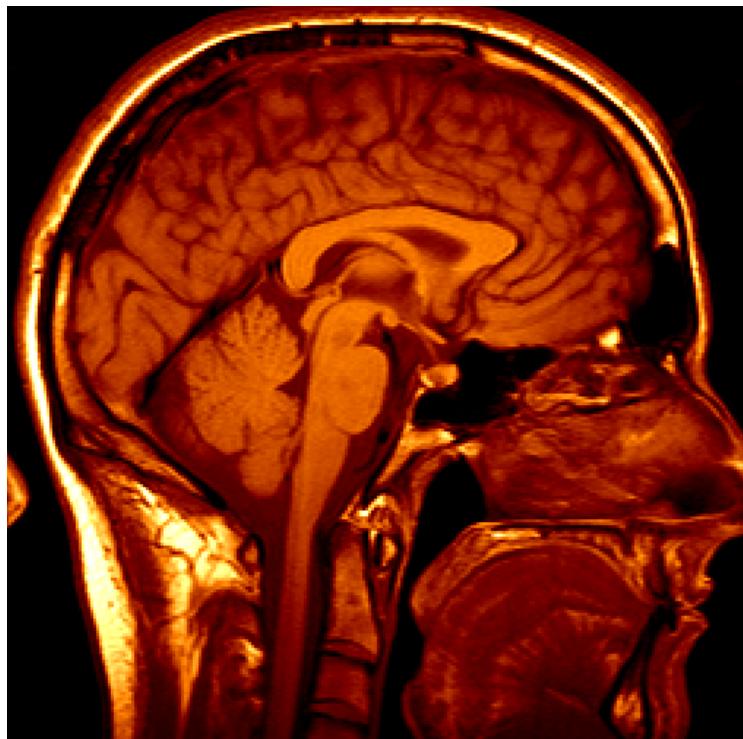
Sports



Sportvision first down line
Nice [explanation](http://www.howstuffworks.com) on www.howstuffworks.com

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

Medical imaging



3D imaging
MRI, CT



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

Smart cars

Slide content courtesy of Amnon Shashua

The image is a screenshot of the Mobileye website. At the top, there are two tabs: 'manufacturer products' (highlighted in blue) and 'consumer products' (highlighted in grey). Below the tabs, the slogan 'Our Vision. Your Safety.' is displayed. The main visual is a top-down view of a car with three camera systems highlighted: 'rear looking camera' (top left), 'forward looking camera' (top right), and 'side looking camera' (bottom). Below this, there are three main sections: 'EyeQ Vision on a Chip' (with an image of a chip), 'Vision Applications' (with an image of a pedestrian crossing a street), and 'AWS Advance Warning System' (with an image of a display screen). To the right, there are two columns: 'News' (listing 'Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System' and 'Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end') and 'Events' (listing 'Mobileye at Equip Auto, Paris, France' and 'Mobileye at SEMA, Las Vegas, NV').

- Mobileye
 - Market Capitalization: 11 Billion dollars
 - See also CVPR 2016 keynote

Google cars



Oct 9, 2010. ["Google Cars Drive Themselves, in Traffic"](#). [The New York Times](#). John Markoff

June 24, 2011. ["Nevada state law paves the way for driverless cars"](#). [Financial Post](#). Christine Dobby

Aug 9, 2011,
["Human error blamed after Google's driverless car sparks five-vehicle crash"](#). [The Star \(Toronto\)](#)

Ford acquires SAIPS for self-driving machine learning and computer vision tech

Posted Aug 16, 2016 by [Darrell Etherington \(@etherington\)](#)



Ford outlined a few of the ways it's aiming to [ship driverless cars by 2021](#), and part of the plan involves acquisitions. CEO Mark Fields revealed at a press event in Palo Alto today that the automaker [acquired SAIPS](#), an Israeli company focusing on machine learning and computer vision. It's also partnering exclusively with Nirenberg Neuroscience, to bring more "humanlike intelligence" to machine learning components of driverless car systems.

SAIPS' technology brings image and video processing algorithms, as well as deep learning tech focused on processing and classifying input signals, all key ingredients in the special sauce that makes up autonomous vehicle tech. This company's expertise should help with on-board interpretation of data captured by sensors on Ford's self-driving cars, and turning that data into usable info for the car's virtual driver system. SAIPS' offerings include detection of anomalies, persistent tracking of objects detected by sensors, and much more. The company's past clients include HP and Trax, but its partner group doesn't appear to have included much in the way of driving-specific applications.

CrunchBase

Ford Motor Company

FOUNDED
1903

OVERVIEW

Ford is an automotive company that develops, manufactures, distributes, and services vehicles, parts, and accessories worldwide. It operates through two sectors: automotive and financial services. The automotive sector offers vehicles primarily under the Ford and Lincoln brand names. This sector markets cars, trucks, parts, and accessories through retail dealers in North America and distributors ...

LOCATION
Dearborn, MI

CATEGORIES
Automotive

WEBSITE
<http://www.ford.com/>

[Full profile for Ford Motor Company](#)

TC NEWSLETTERS

The Daily Crunch

Our top headlines
Delivered daily

TC Week-in-Review

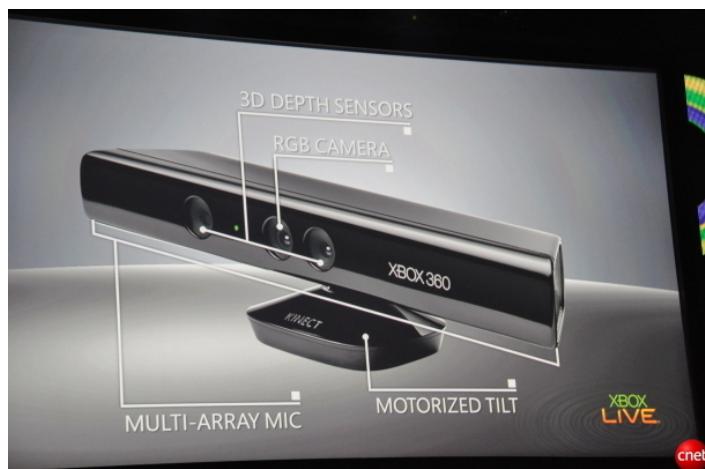
Top stories of the week
Delivered weekly

CrunchBase Daily

The latest

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>



Industrial robots



Vision-guided robots position nut runners on wheels

Vision in space

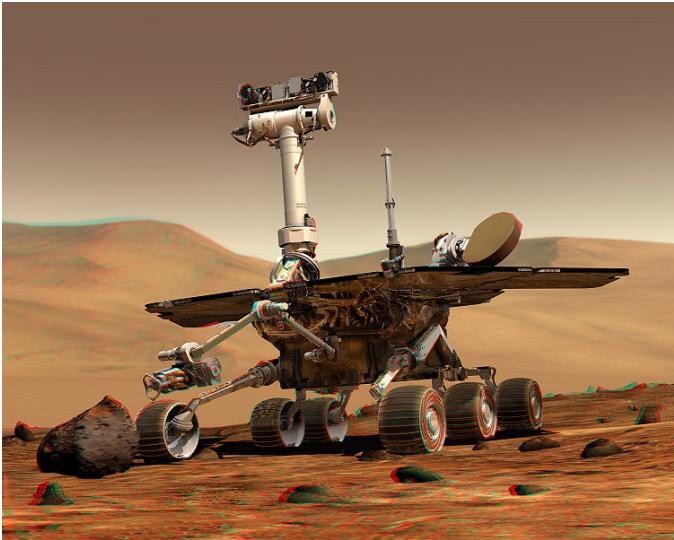


[NASA's Mars Exploration Rover Spirit](#) captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

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.

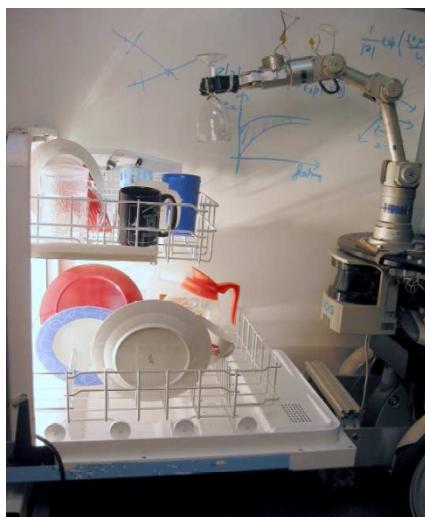
Mobile robots



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover



<http://www.robocup.org/>

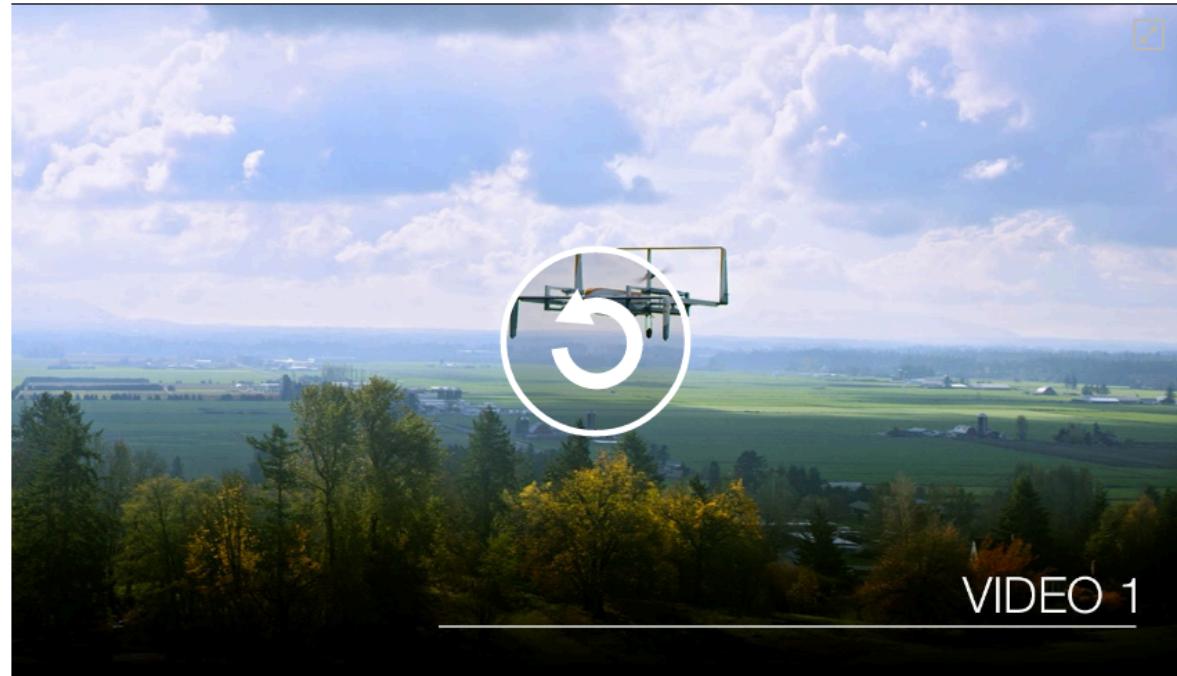
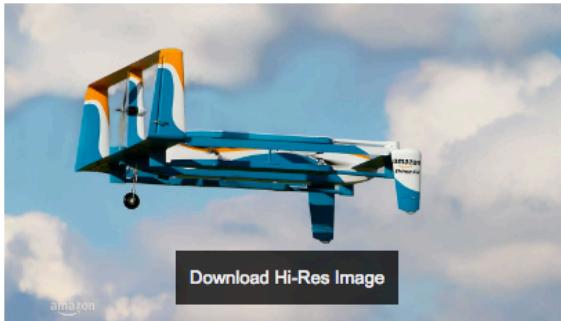


Saxena et al. 2008
[STAIR](#) at Stanford

Amazon Prime Air



We're excited about Prime Air — a future delivery system from Amazon designed to safely get packages to customers in 30 minutes or less using small unmanned aerial vehicles, also called drones. Prime Air has great potential to enhance the services we already provide to millions of customers by providing rapid parcel delivery that will also increase the overall safety and efficiency of the transportation system. Putting Prime Air into service will take some time, but we will deploy when we have the regulatory support needed to realize our vision.



<https://www.amazon.com/b?node=8037720011>

Skydio

<https://www.skydio.com/>

Augmented Reality and Virtual Reality



Magic Leap, Oculus, Hololens, etc.

State of the art today?

With enough training data, computer vision nearly matches human vision at some recognition tasks

Deep convolutional neural networks have been a disruption to the field. More and more techniques are being “deepified”.

Major research challenges, however, remain.

Computer Vision and Nearby Fields

- Computer Graphics: Models to Images
- Comp. Photography: Images to Images
- Computer Vision: Images to Models



Steve Sullivan

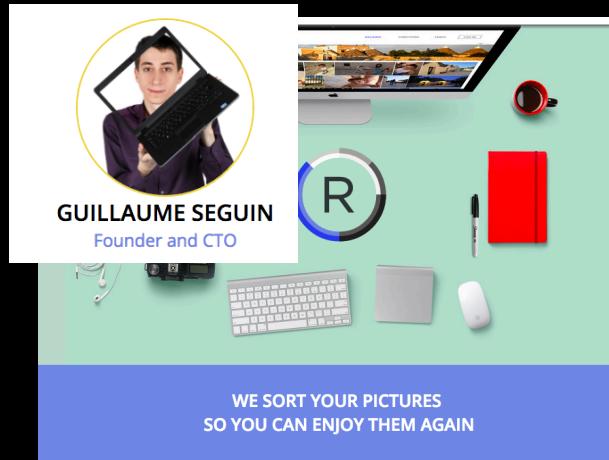
- Ph.D., UIUC, 1996
- Head of R&D, ILM, 2003
- Cover, IEEE Spectrum, 2004
- CSO, Lucasfilm, 2009-2012
- Microsoft (2013-)
- 3 Academy Awards

Computer vision as a job

Vincent Delaitre
Phd, 2015
Start-up Deepomatic.com



Guillaume Seguin
Phd, 2016
Start-up regaind.io



Piotr Bojanowski
Phd, 2016
Facebook AI Research



Oliver Whyte
Phd, 2012
Engineer at Microsoft



Mathieu Aubry
Phd, 2015
Faculty at ENPC



Relja Arandjelovic
Post-doc, 2016
Google DeepMind



Computer vision books

- D.A. Forsyth and J. Ponce, "Computer Vision: A Modern Approach", Prentice-Hall, 2003, 2nd edition, 2011.
- 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.

Other relevant books

- J.J. Koenderink, "Solid Shape", MIT Press, 1990.
- J.J. Koenderink, <http://www.gestaltrevision.be/en/resources/clootcrans-press>
- M. Berger, "Géométrie", Nathan, 1992.
- D. Hilbert and S. Cohn-Vossen, "Geometry and the Imagination", Chelsea, 1952.

Course outline:

1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

Programming assignments + final presentation

Reconnaissance d'objets et vision artificielle

<http://www.di.ens.fr/willow/teaching/recvis16/>



Ivan Laptev



Josef Sivic



Cordelia Schmid



Jean Ponce

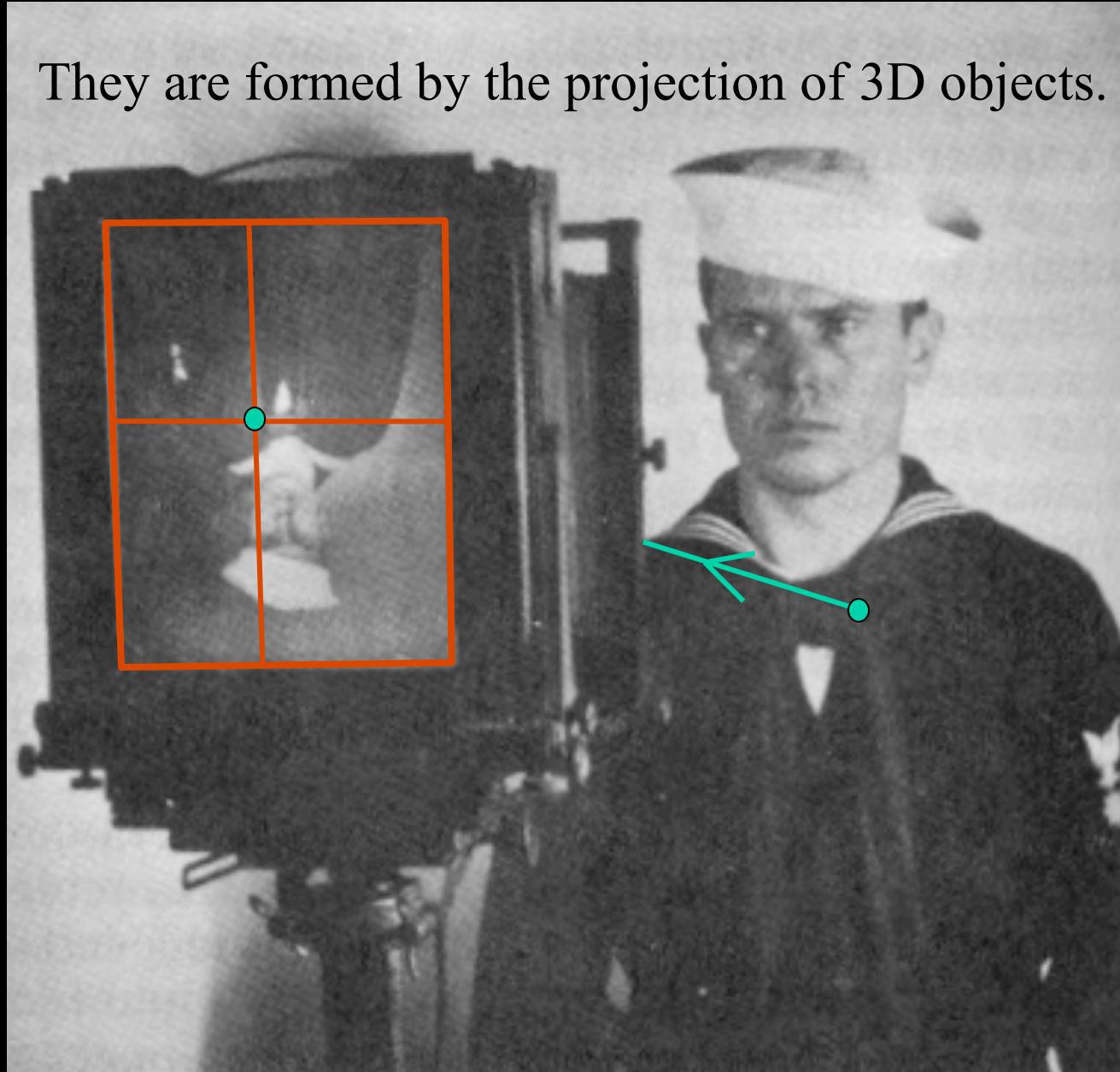
Tuesdays 16h15 salle Conference, at 46 rue d'Ulm
First class: Tuesday Oct 4th

Course outline:

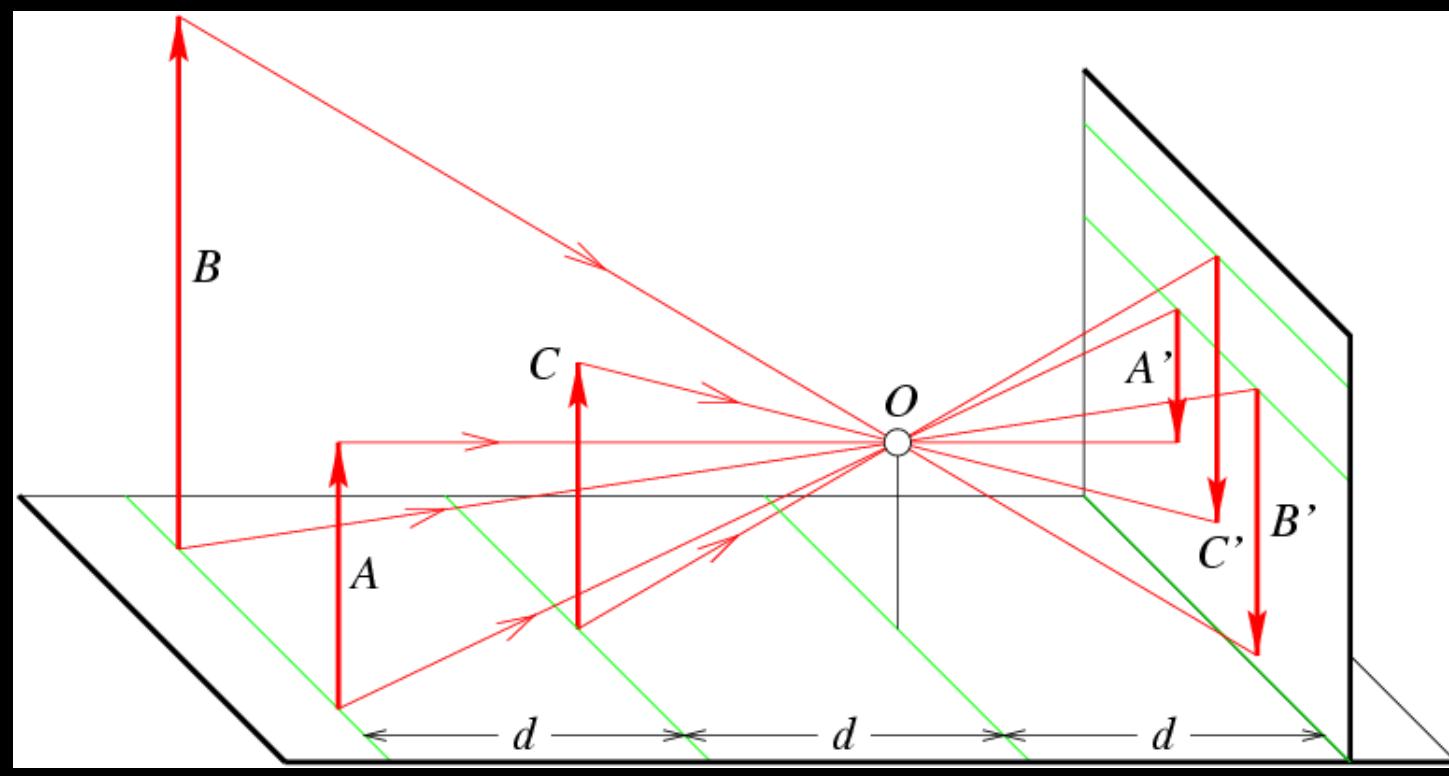
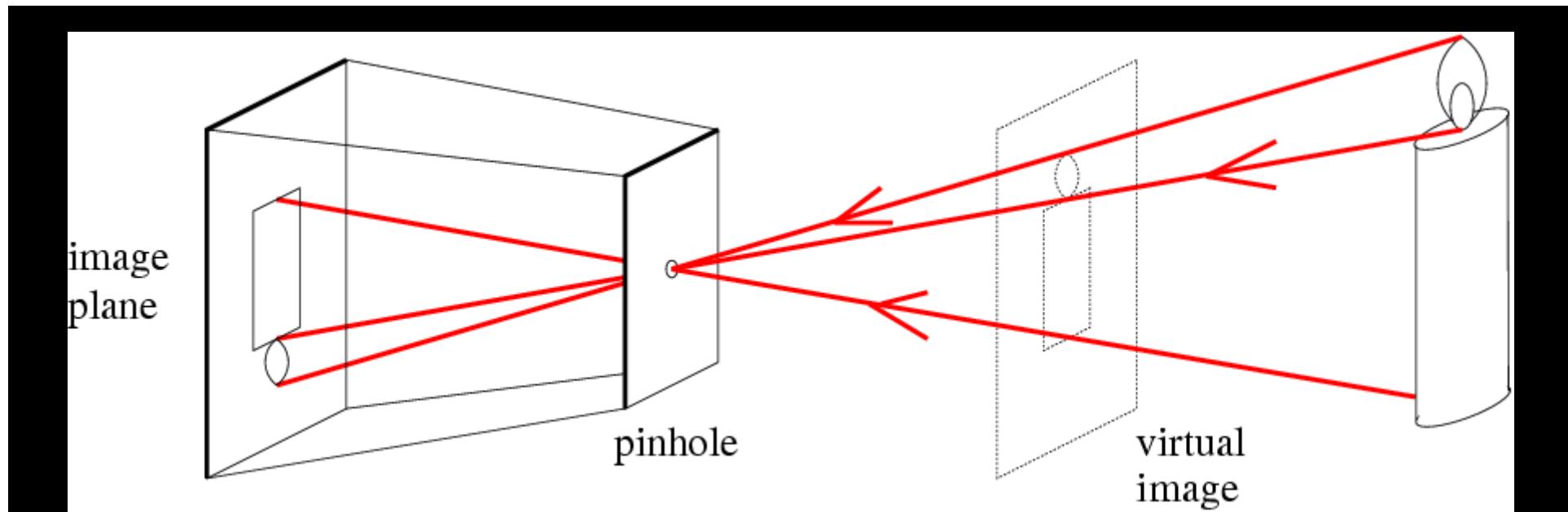
1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

Programming assignments + final presentation

They are formed by the projection of 3D objects.



Images are two-dimensional patterns of brightness values.

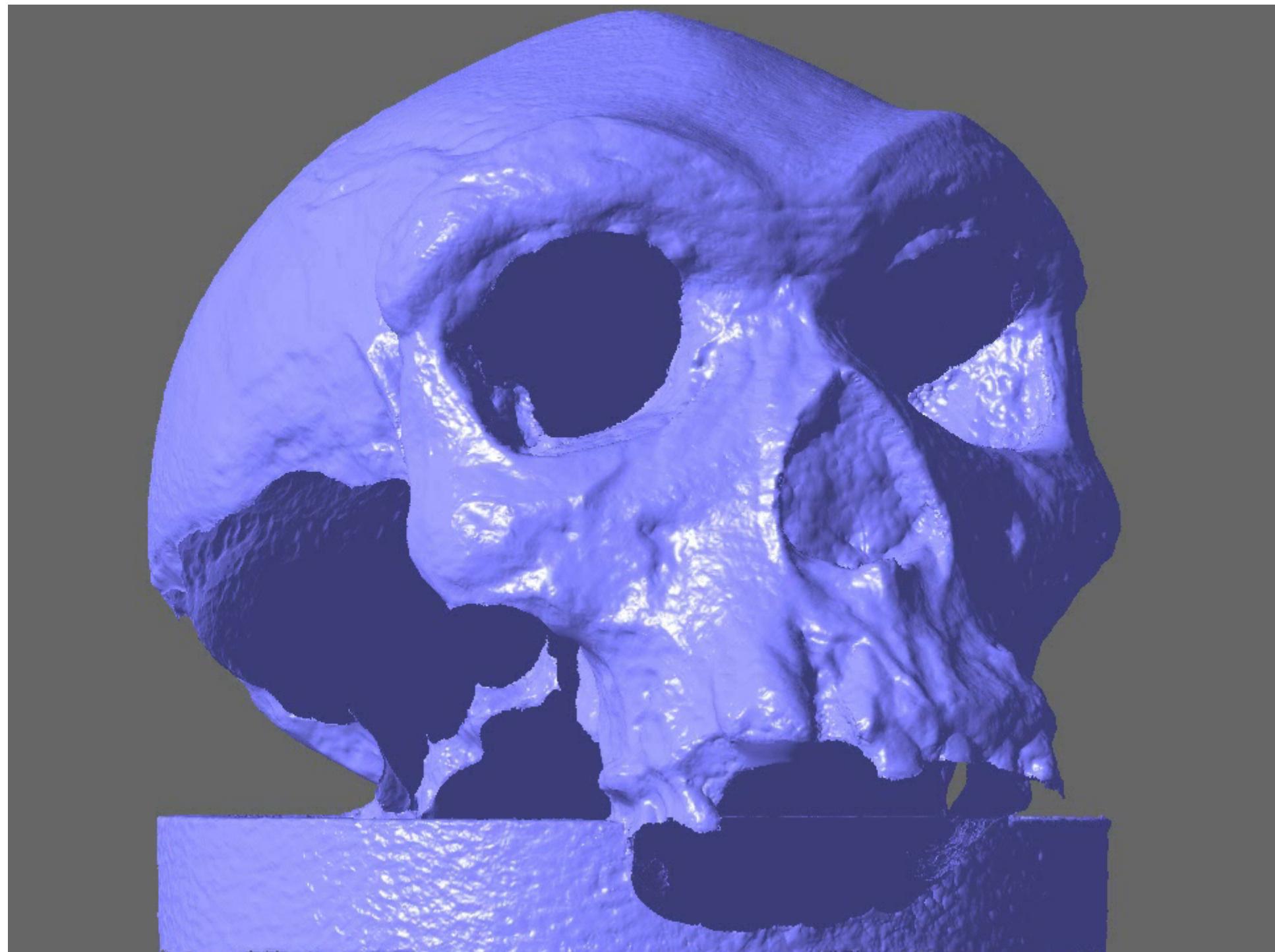


High-fidelity multi-view stereopsis (Furukawa and Ponce, CVPR'07, PAMI'10)

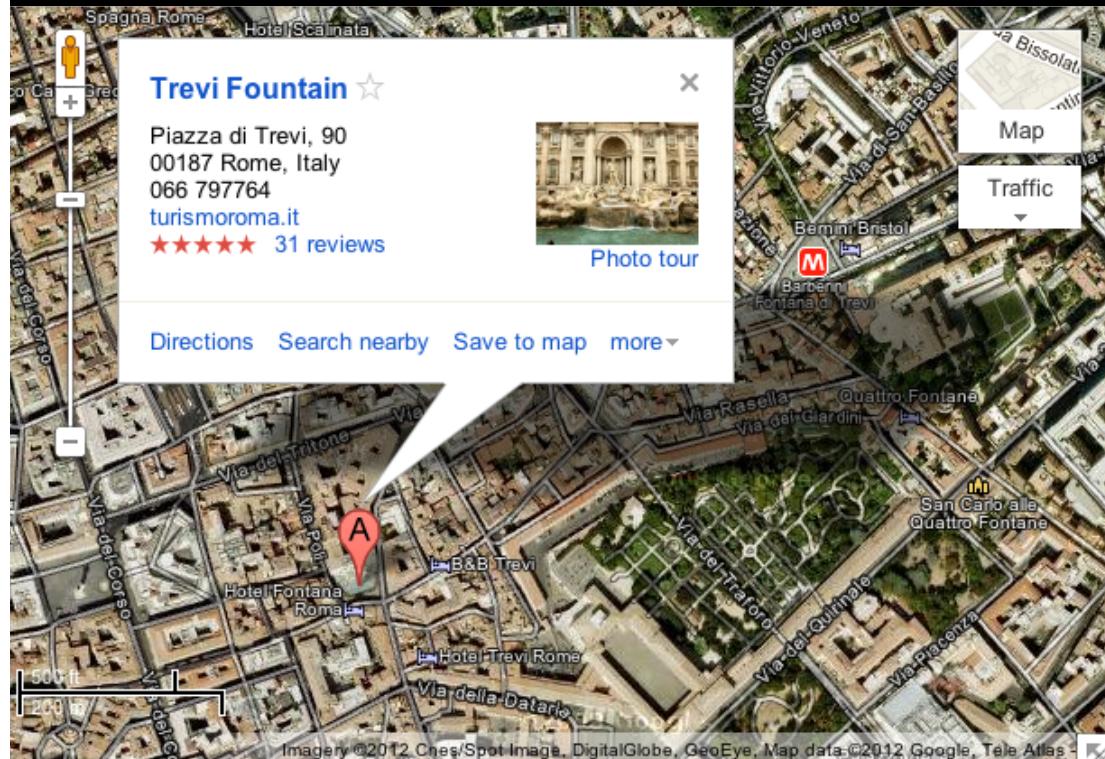
<http://www.cs.washington.edu/homes/furukawa/research/pmvs/index.html>



Data courtesy of S. Leigh, UIUC Anthropology Department. See for example (Hernandez and Schmitt, 2004; Strecha et al., 2006) for related work.



PMVS (<http://www.di.ens.fr/pmvs>)



(© Bath & Burke, Weta Digital, Siggraph'11)



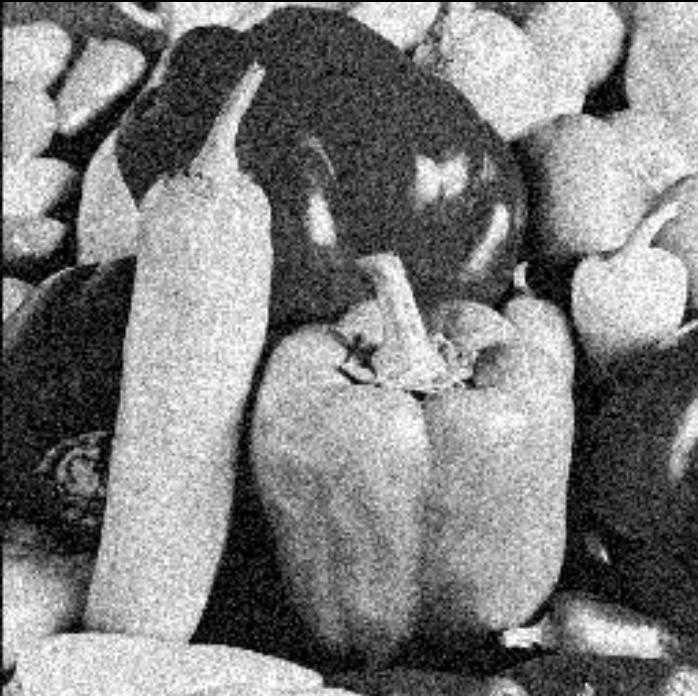
- Google Maps Photo Tour
- Lucasfilm
- Weta Digital

Course outline:

1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

Programming assignments + final presentation

Filtering



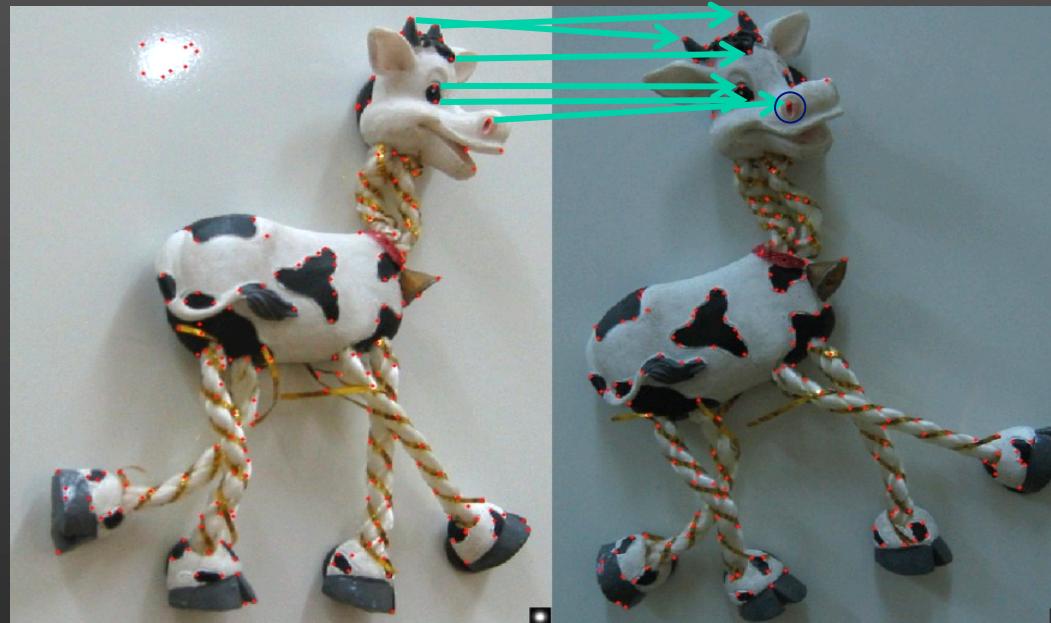
Edge Detection



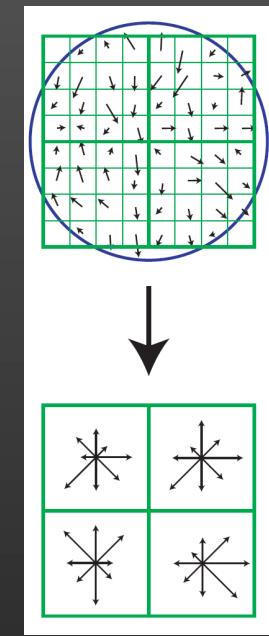
Edge Detection



Interest points and local appearance models



(Image courtesy of C. Schmid)



(Lowe 2004)

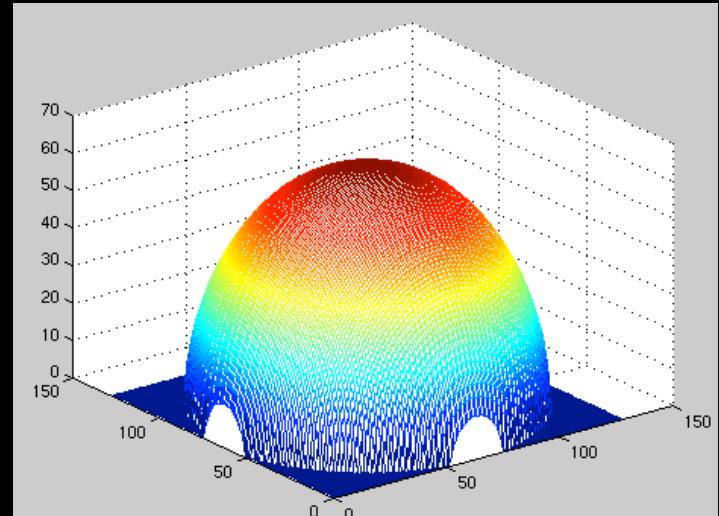
- Find features (interest points)
- Match them using local invariant descriptors (jets, SIFT)

Course outline:

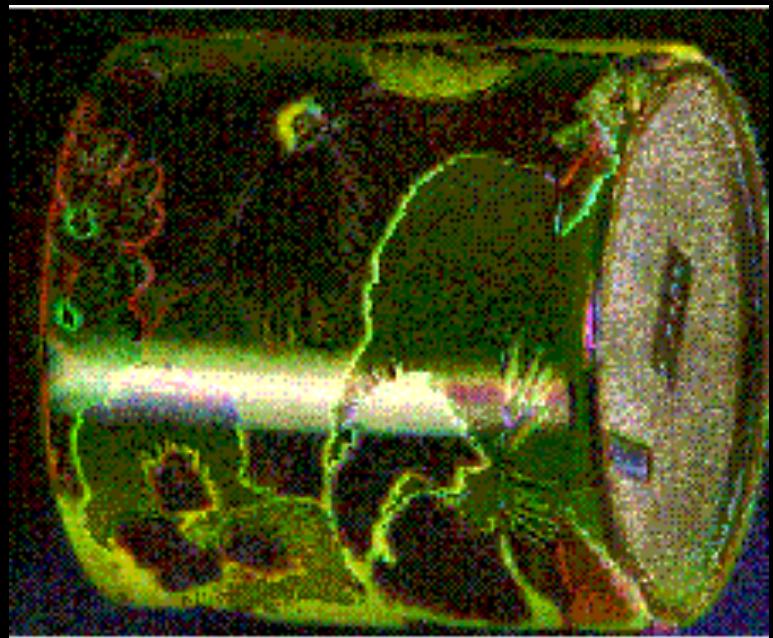
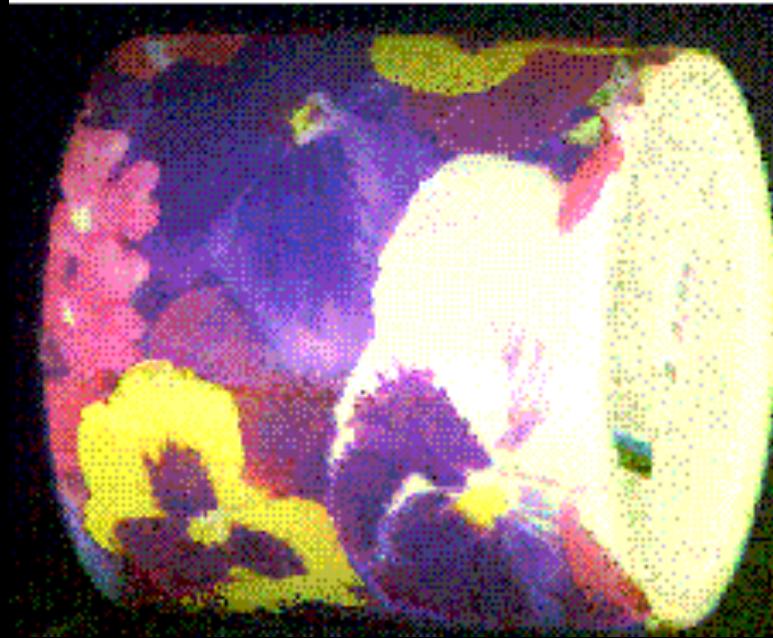
1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

Programming assignments + final presentation

Radiometry/Shading



Color

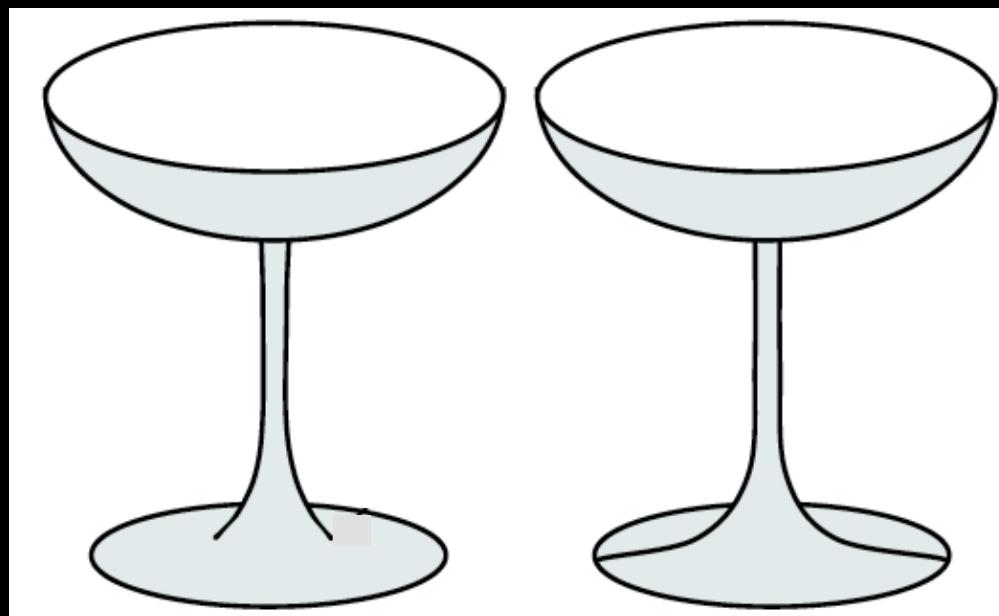
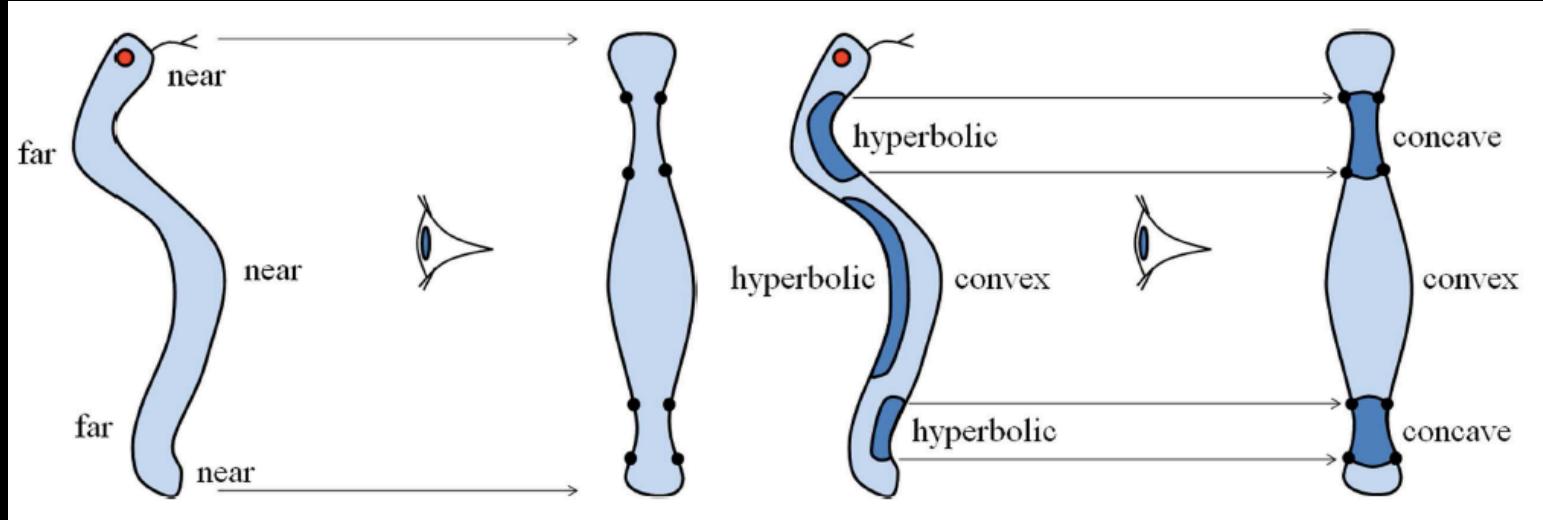


Course outline:

1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

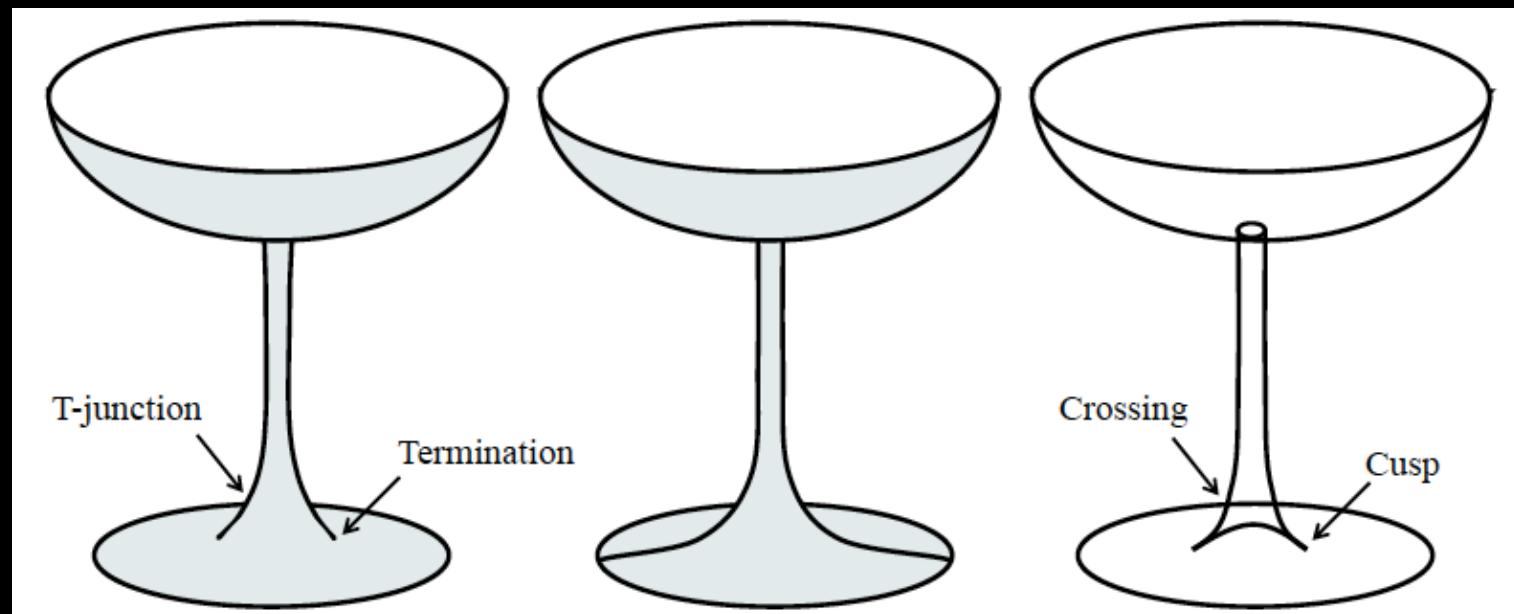
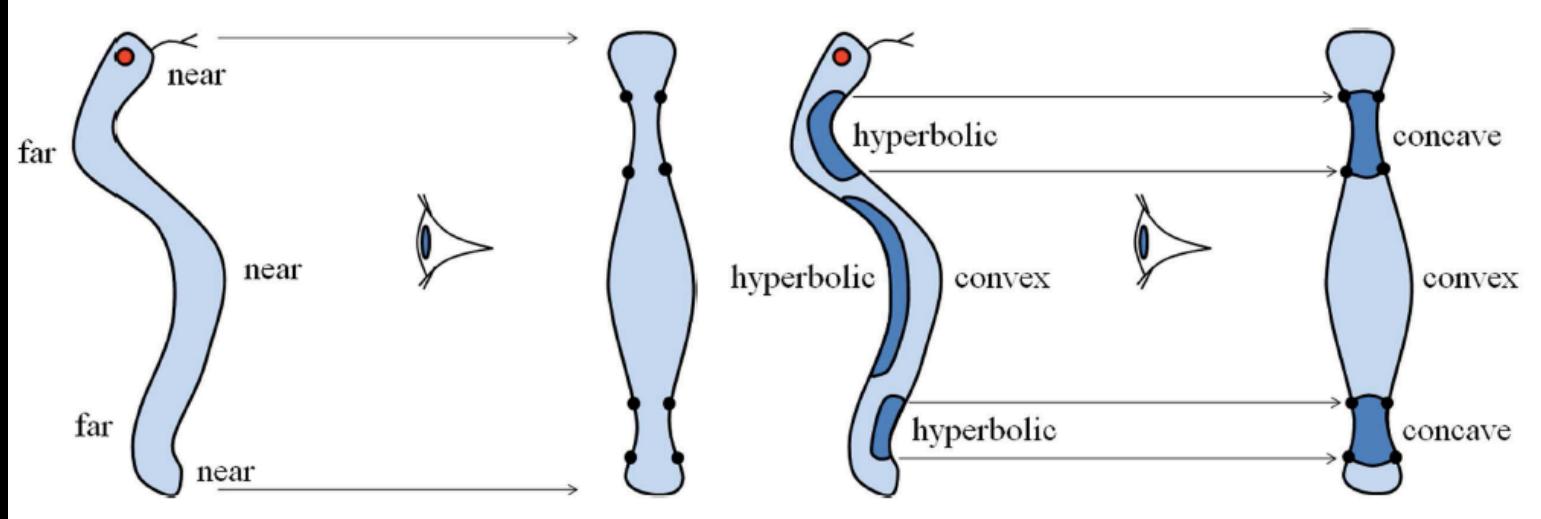
Programming assignments + final presentation

One-view (differential) geometry



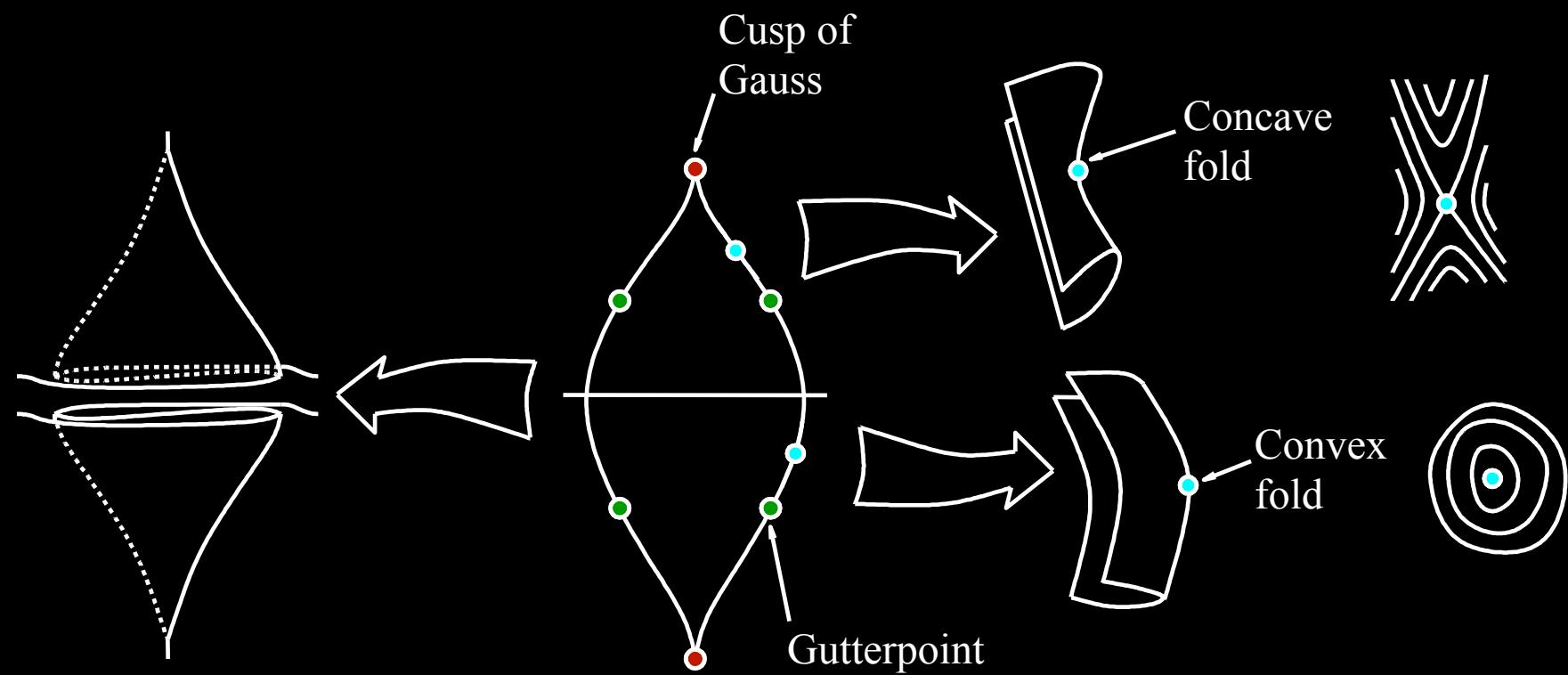
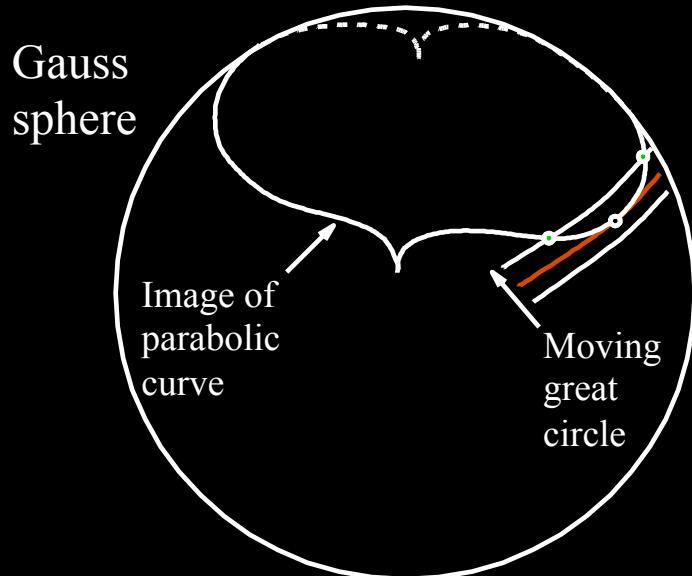
(Marr & Nishihara, 1978; Koenderink, 1984)

One-view (differential) geometry



(Marr & Nishihara, 1978; Koenderink, 1984)

The Geometry of the Gauss Map

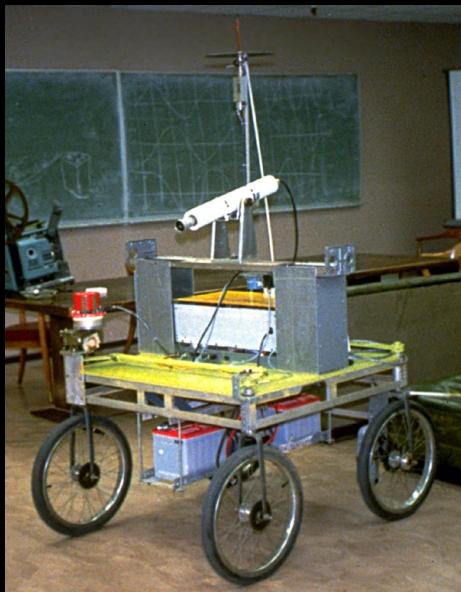
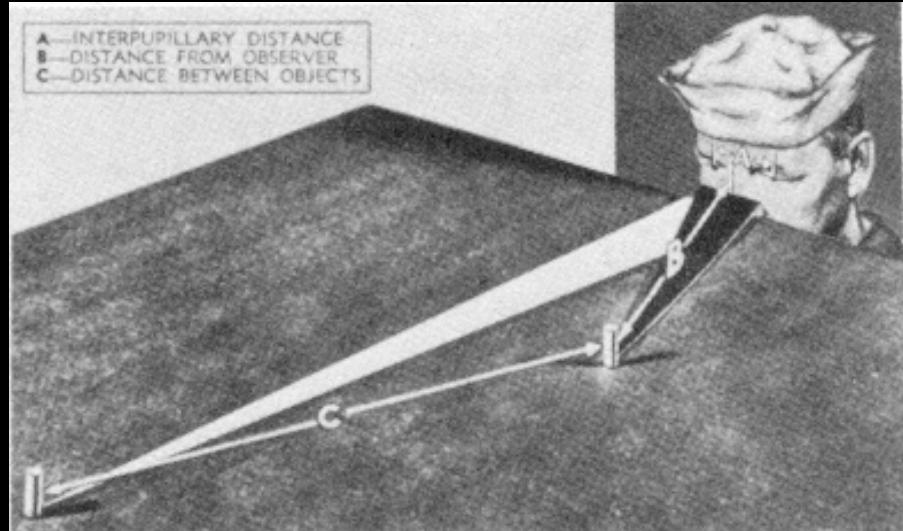


Course outline:

1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

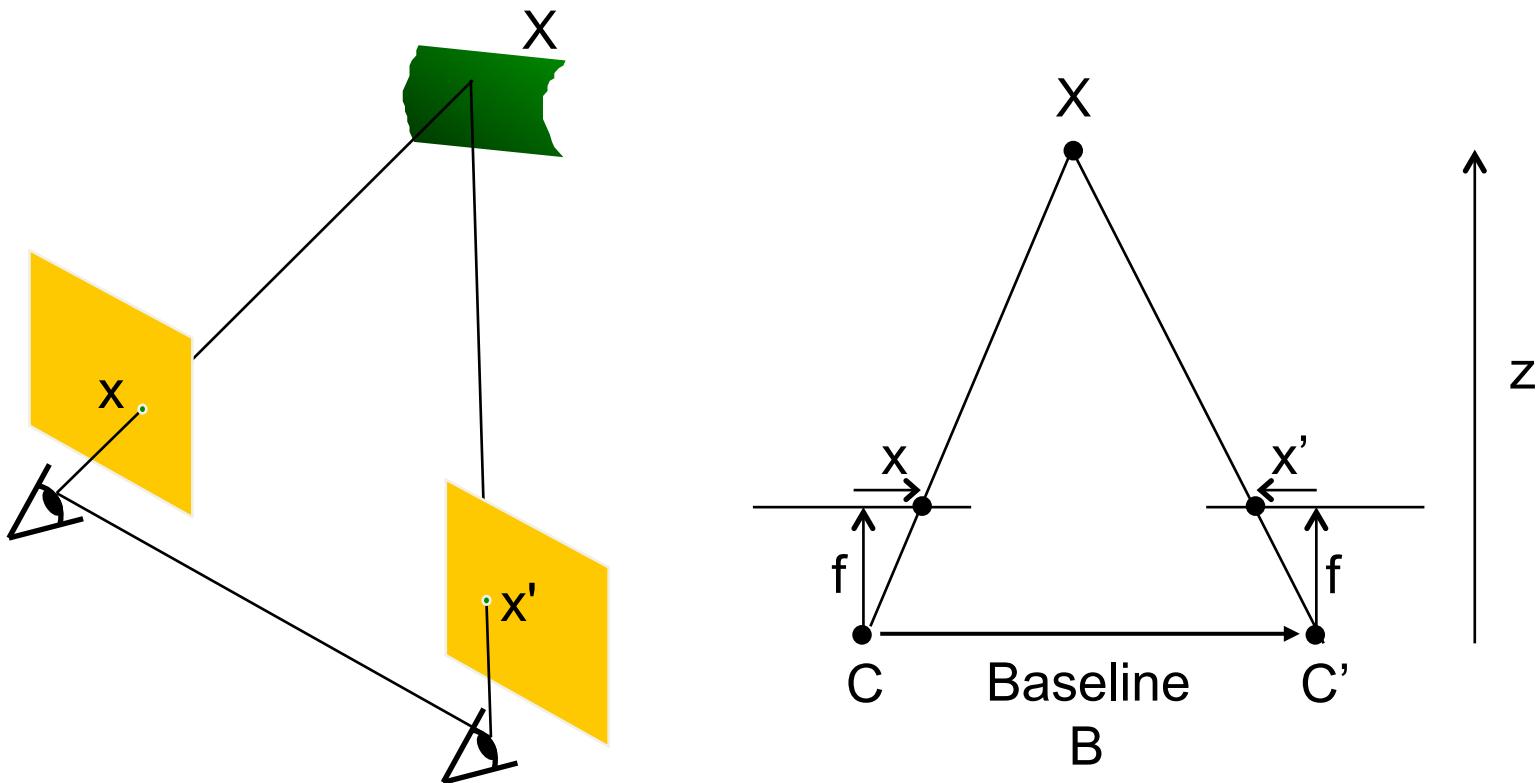
Programming assignments + final presentation

How do we perceive depth?

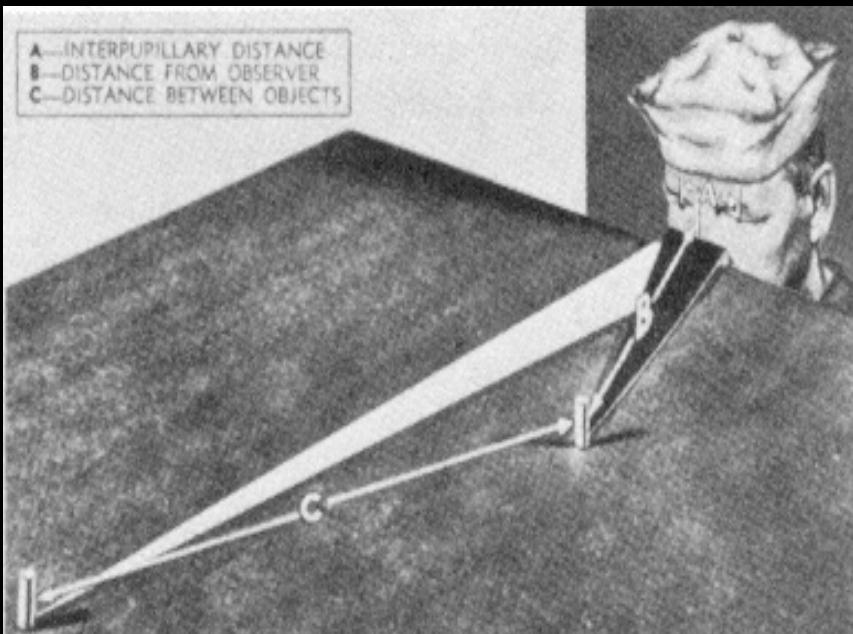


Depth from Stereo

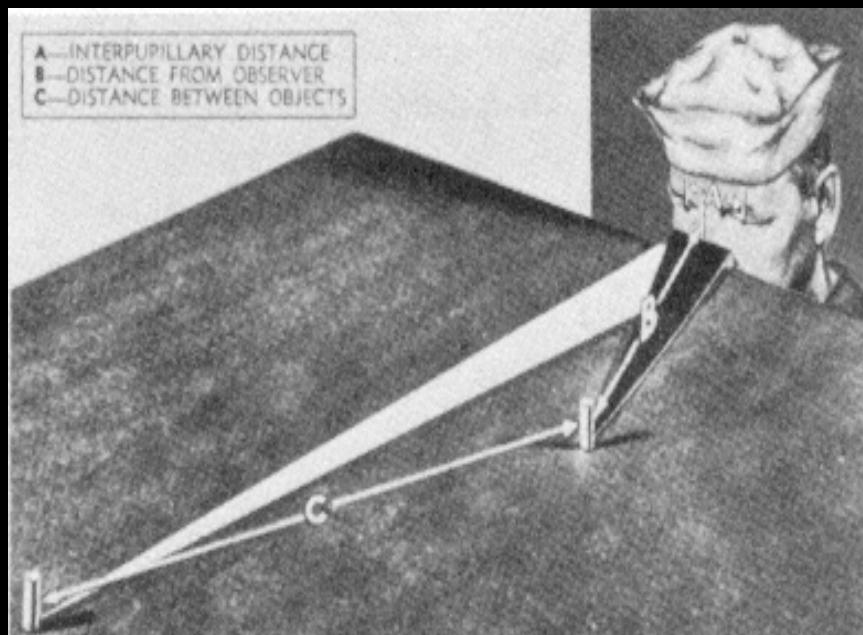
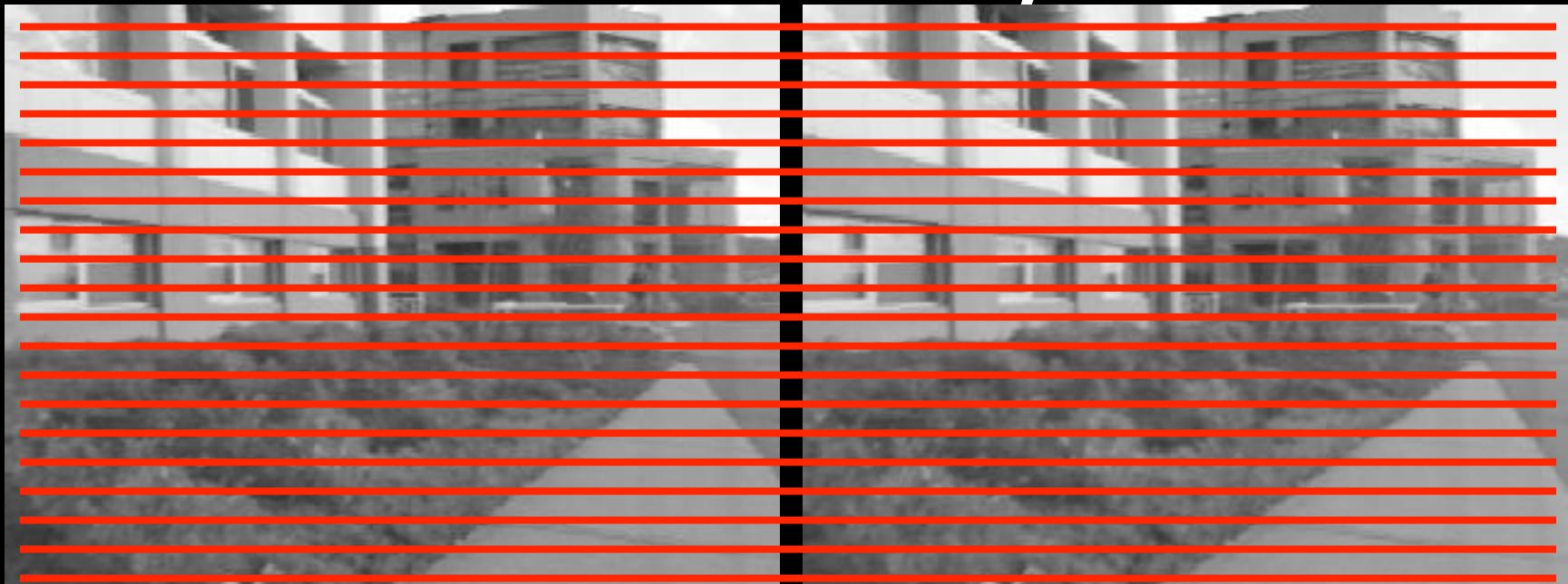
- Goal: recover depth by finding image coordinate x' that corresponds to x



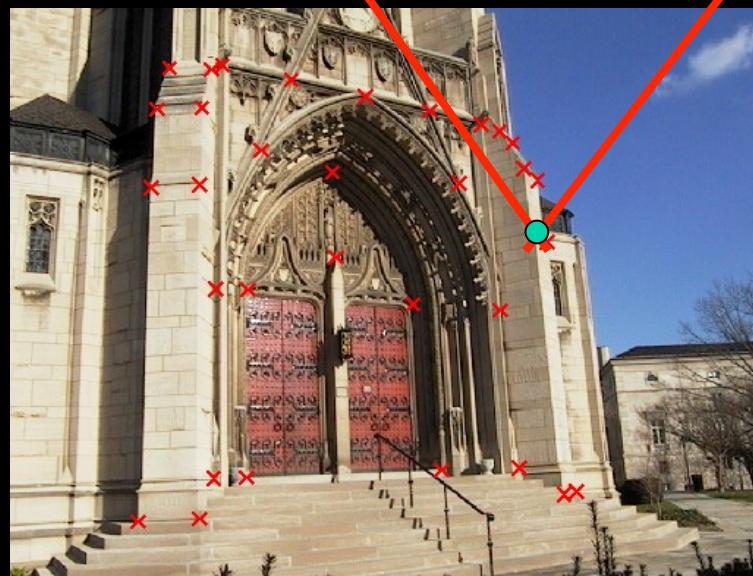
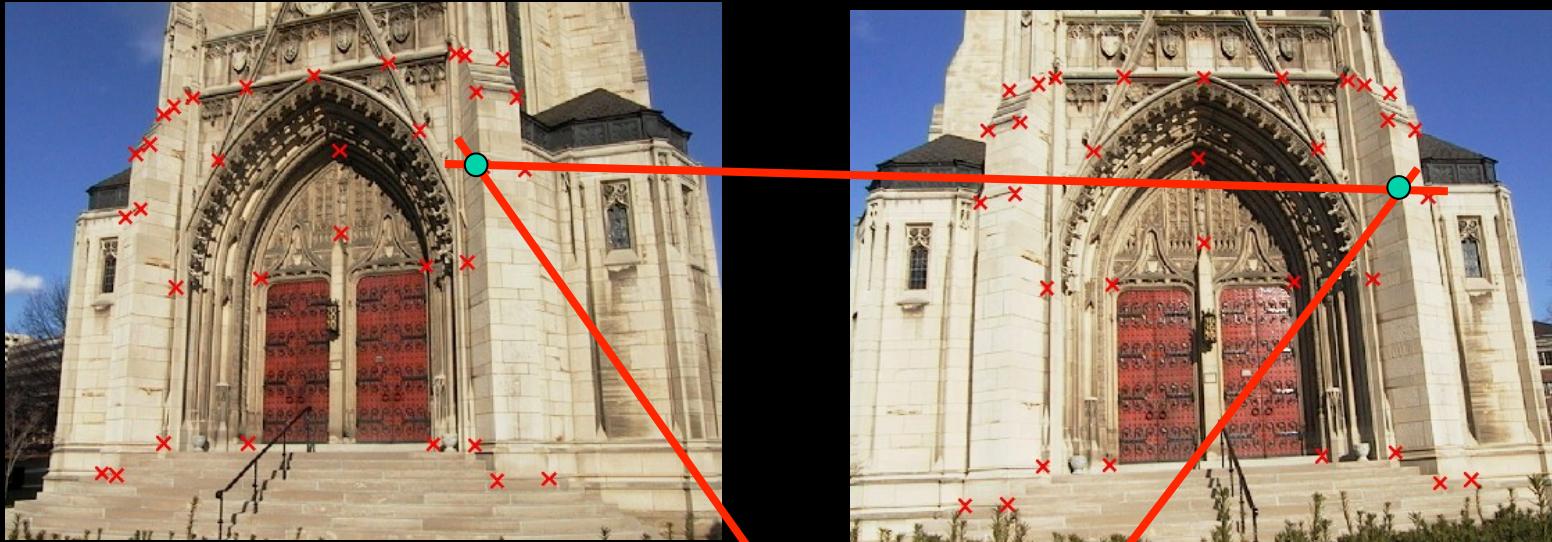
Two-View Geometry: Stereo



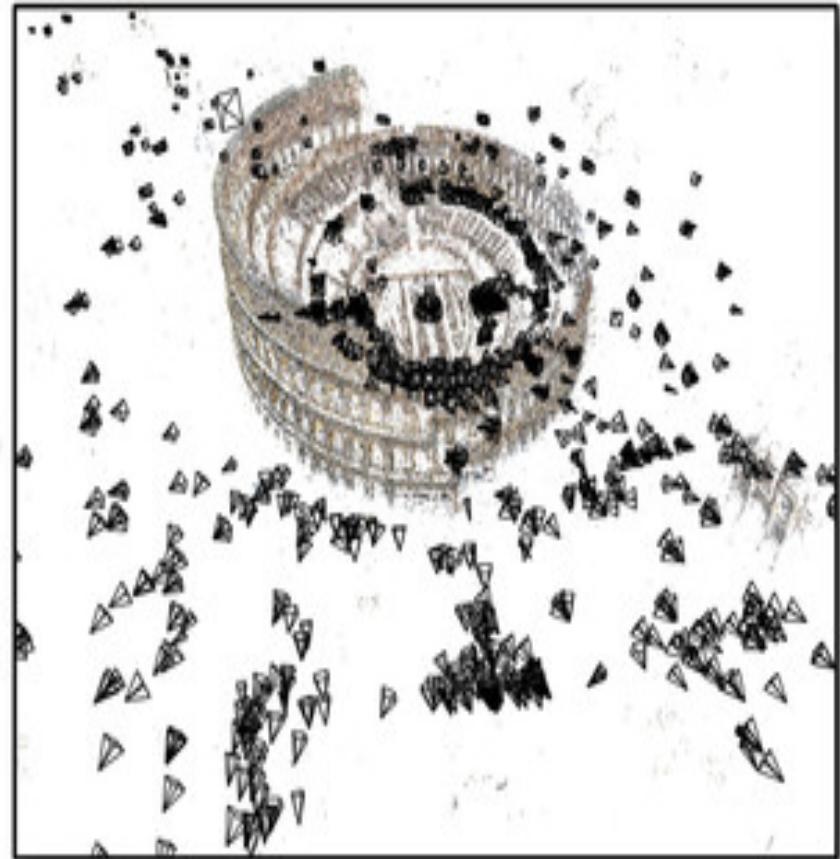
Two-View Geometry: Stereo



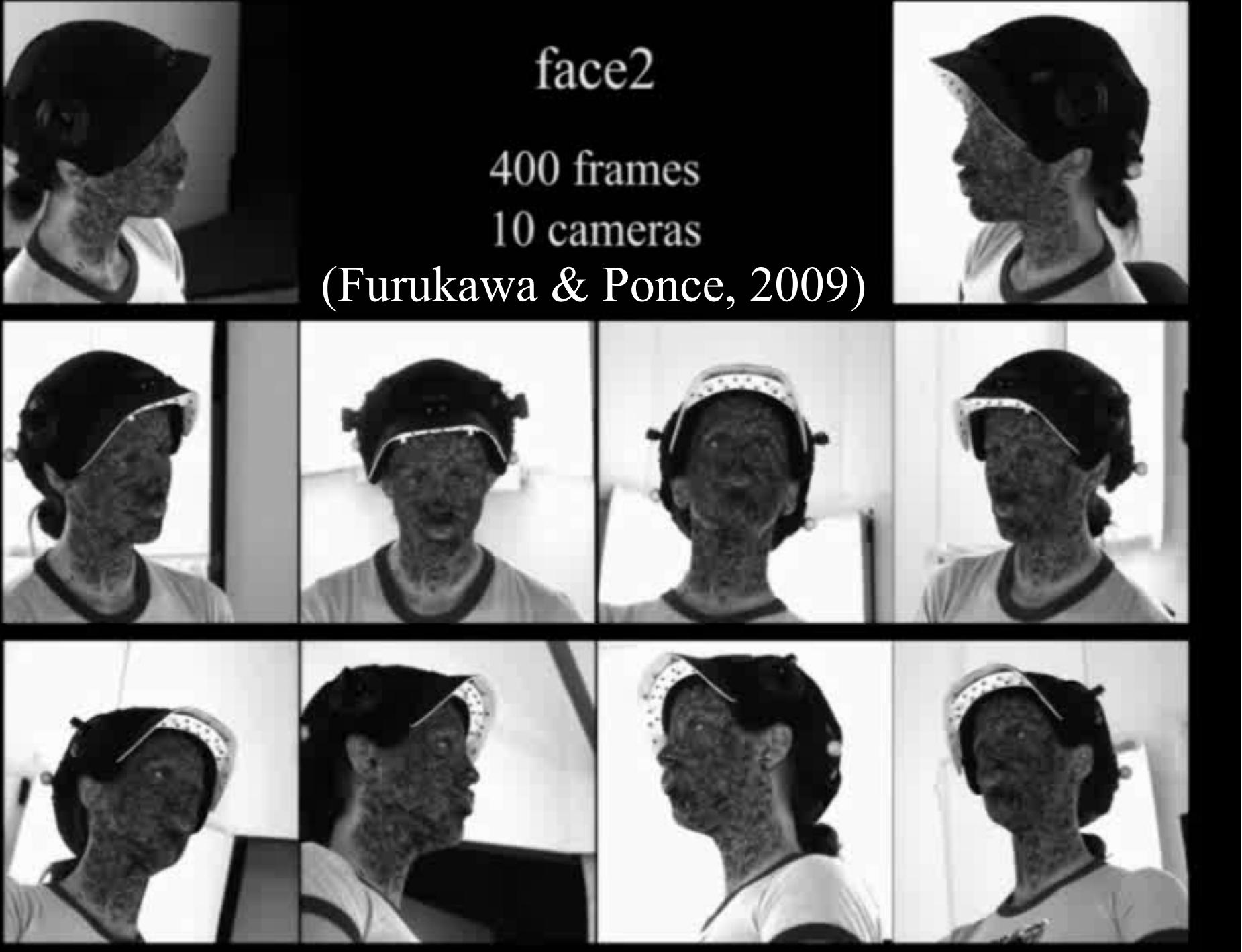
Multi-Camera Geometry



Phototourism



(Snavely, Seitz, Szeliski, 2006)
<http://phototour.cs.washington.edu/>



face2

400 frames

10 cameras

(Furukawa & Ponce, 2009)

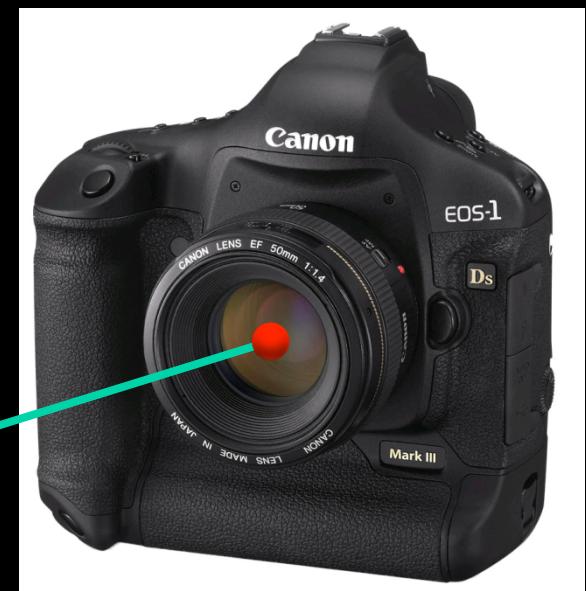
Course outline:

1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

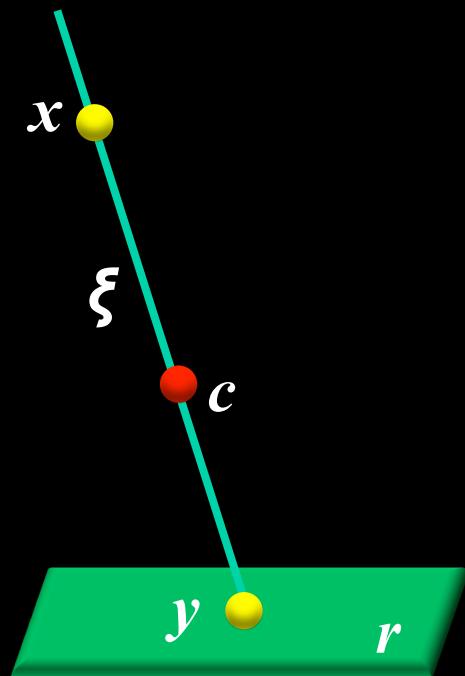
Programming assignments + final presentation

What is a camera?

x

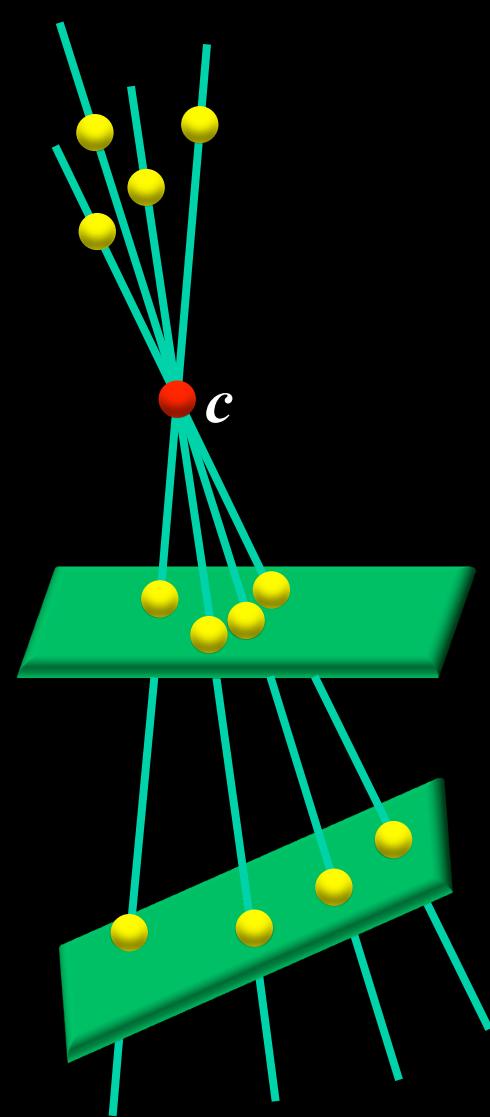
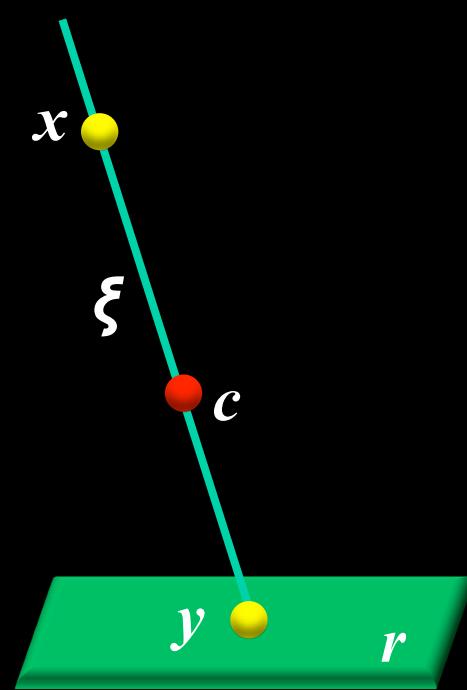


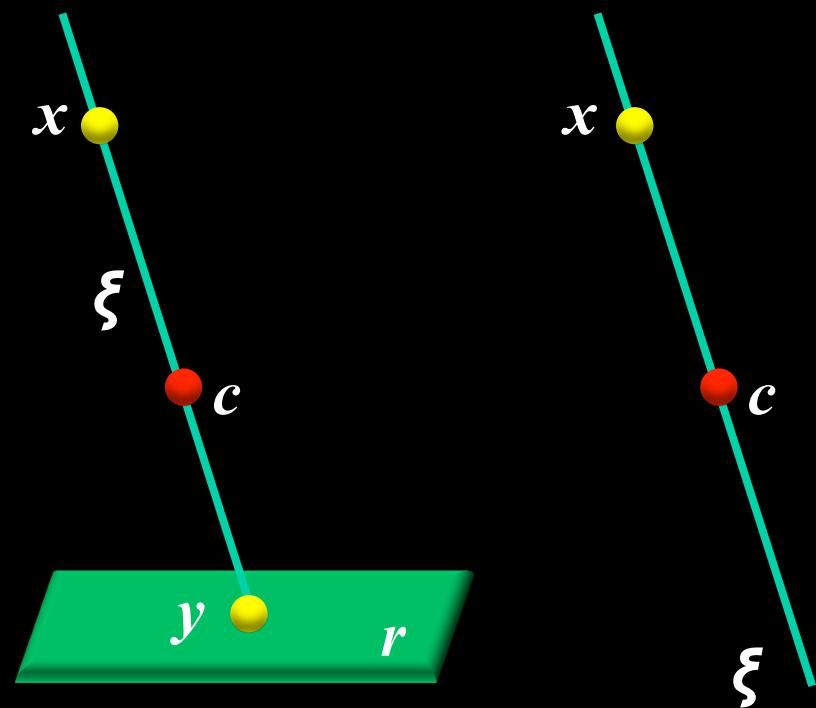
What is a camera?

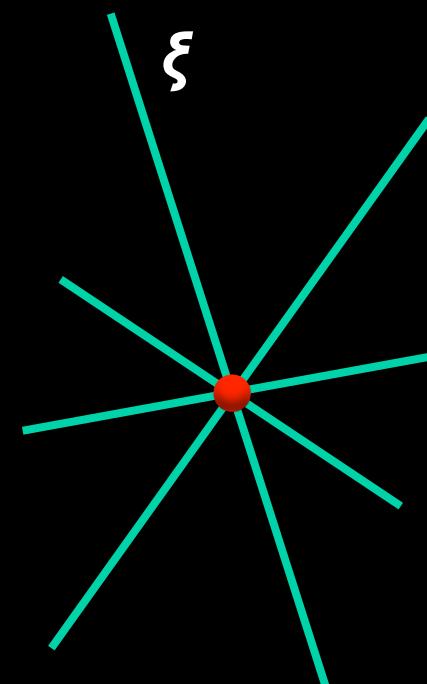
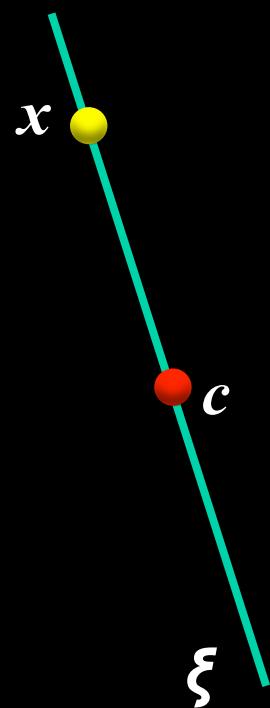
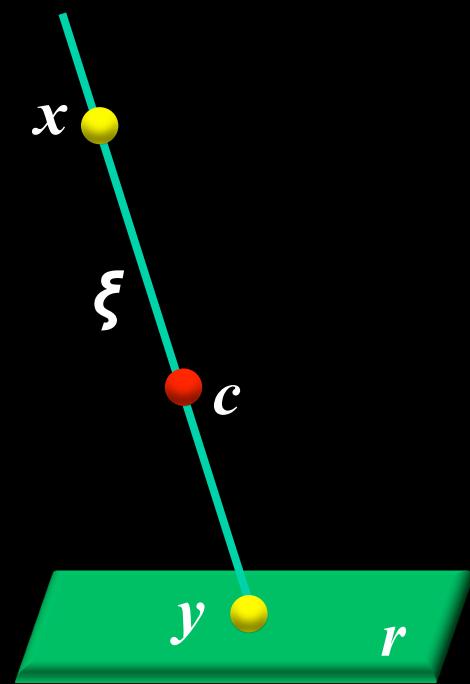


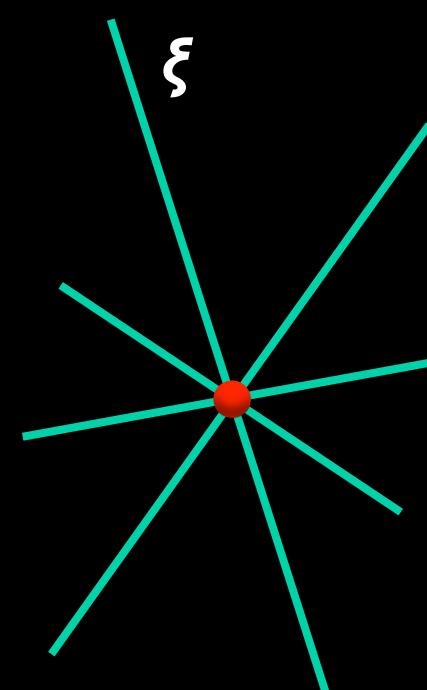
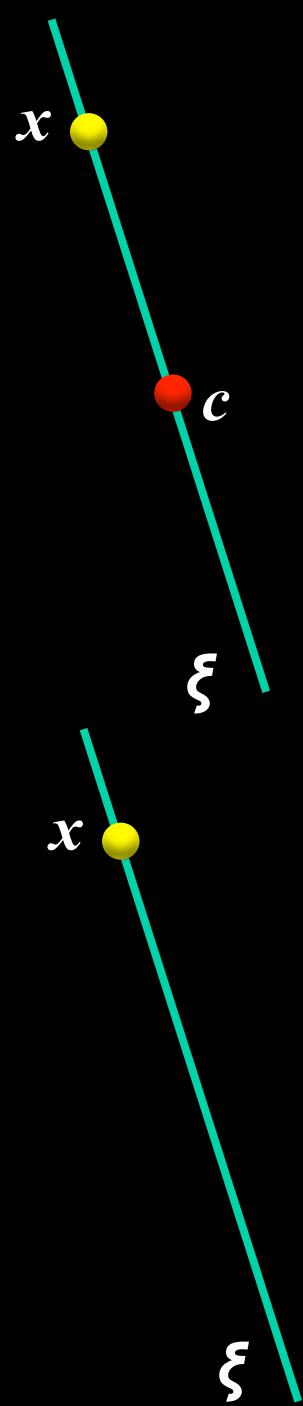
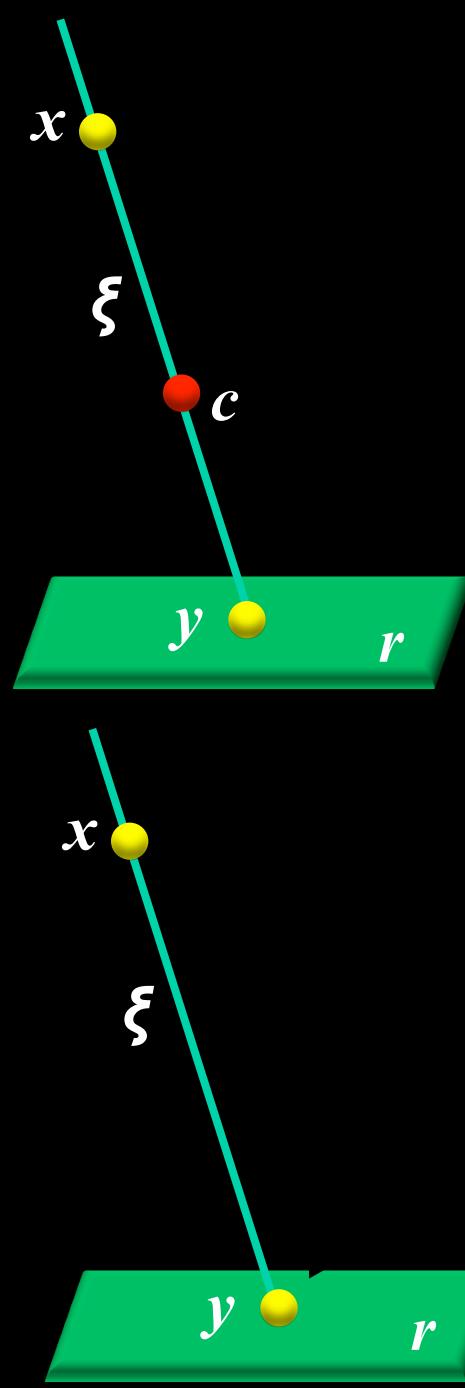
x









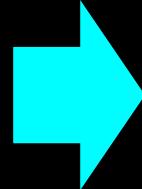


Course outline:

1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

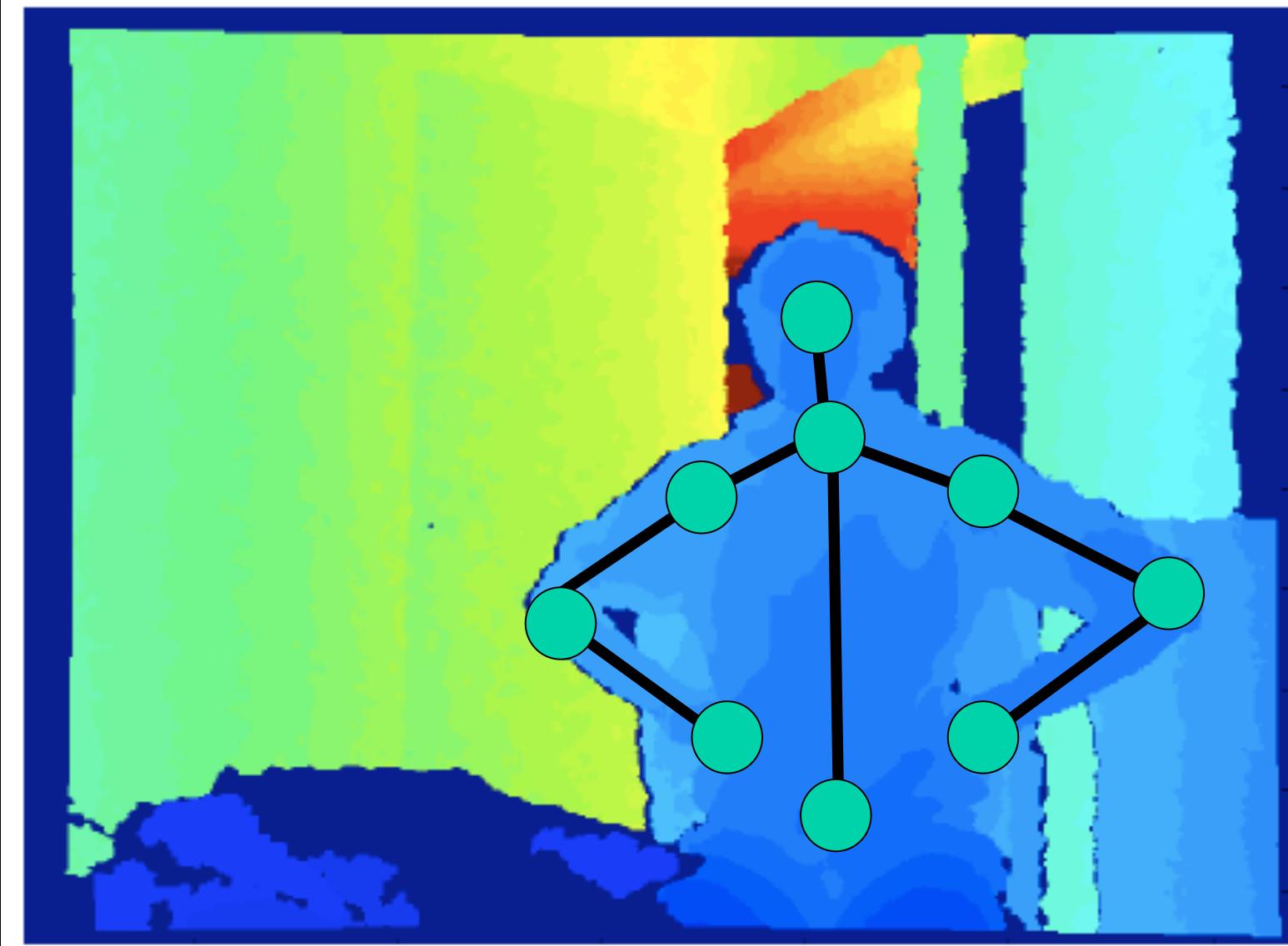
Programming assignments + final presentation

New sensors



Problem: find the 3D skeleton of people

Solution: Use random forest to classify pixels as belonging to some body part



(Shotton et al., 2011)

Course outline:

1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

Programming assignments + final presentation

Segmentation



(Joulin, Bach, Ponce, CVPR'12)

Unsupervised object discovery

aeroplane-0004-029

- Object colocalization per class
- Unsupervised object discovery



(Suha et al., 2015)

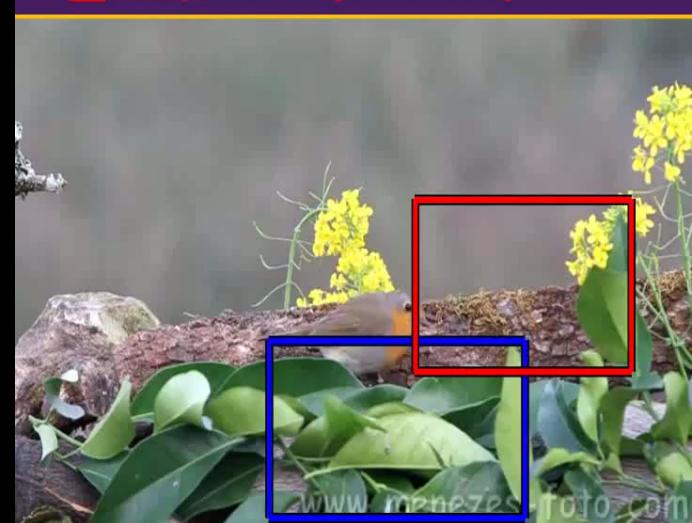
aeroplane-0013-140

- Object colocalization per class
- Unsupervised object discovery



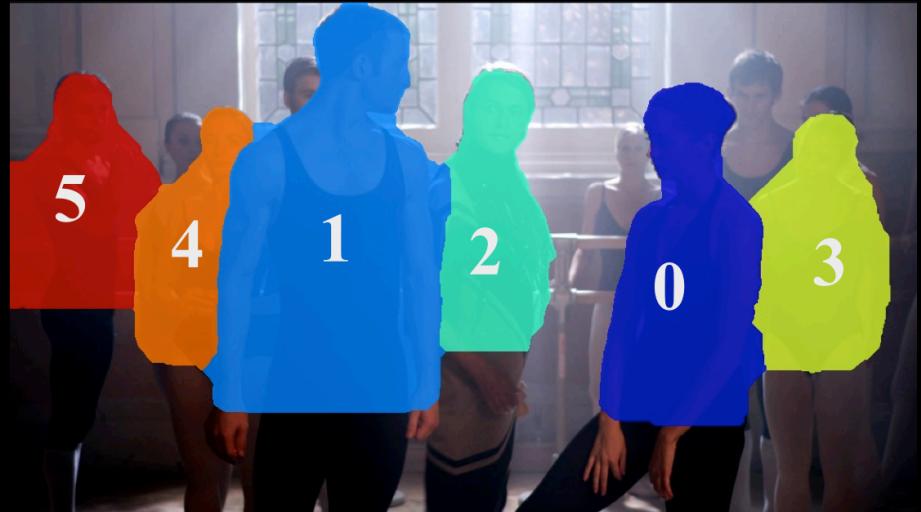
bird-0004-016

- Object colocalization per class
- Unsupervised object discovery



Layered person segmentation

[Seguin et al., 2015]



Course outline:

1. Camera geometry and calibration
2. Filtering, edge and feature detection
3. Radiometry, shading and color
4. One-view (differential) geometry
5. Two-view geometry and stereo
6. Multi-view geometry and stereo, SFM
7. Projective cameras
8. Range data
9. Segmentation
10. Recognition

Programming assignments + final presentation

How to make sense of “pixel-chaos”?

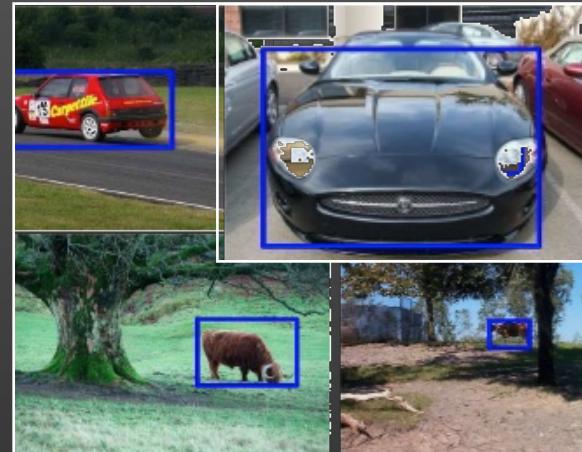
Face recognition



Action recognition



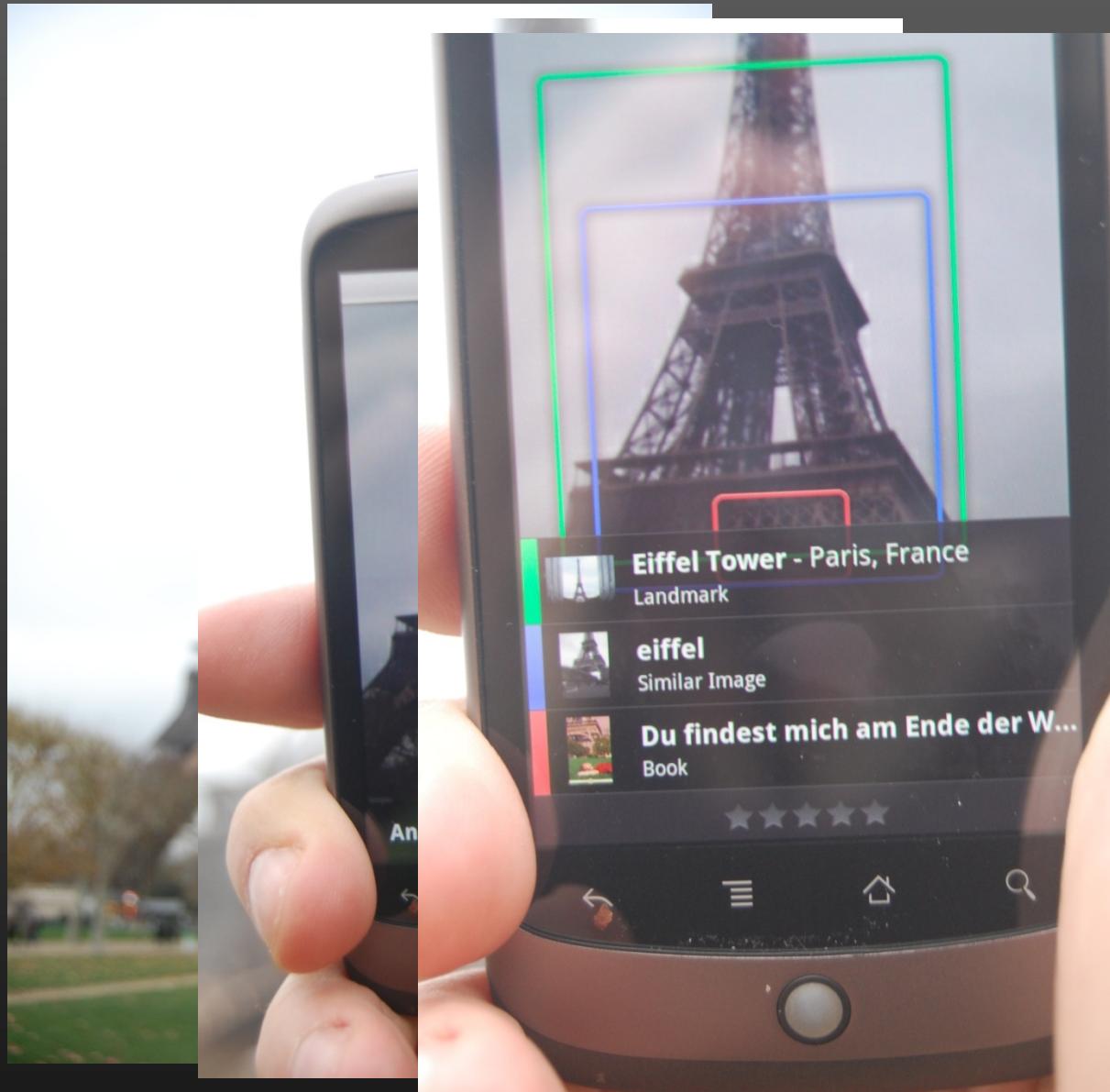
Object class recognition



3D Scene reconstruction



Object **instance** recognition

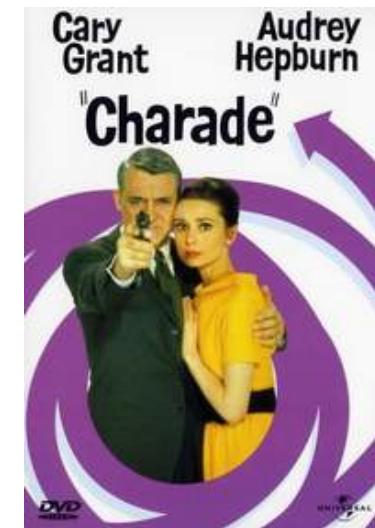


Example: Visual search in an entire feature length movie

Visually defined query



“Find this bag”

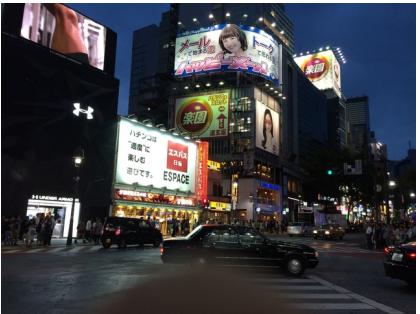


“Charade” [Donen, 1963]

Demo:

<http://www.robots.ox.ac.uk/~vgg/research/vgoogle/index.html>

Instance level recognition : still difficult



See also: this workshop, e.g. Perdoch et al.'15, Fernando et al.'14, Schindler et al.'06, Martin-Buralla'15, Matzen&Snavely'14

Example I.: Localize non-photographic depictions

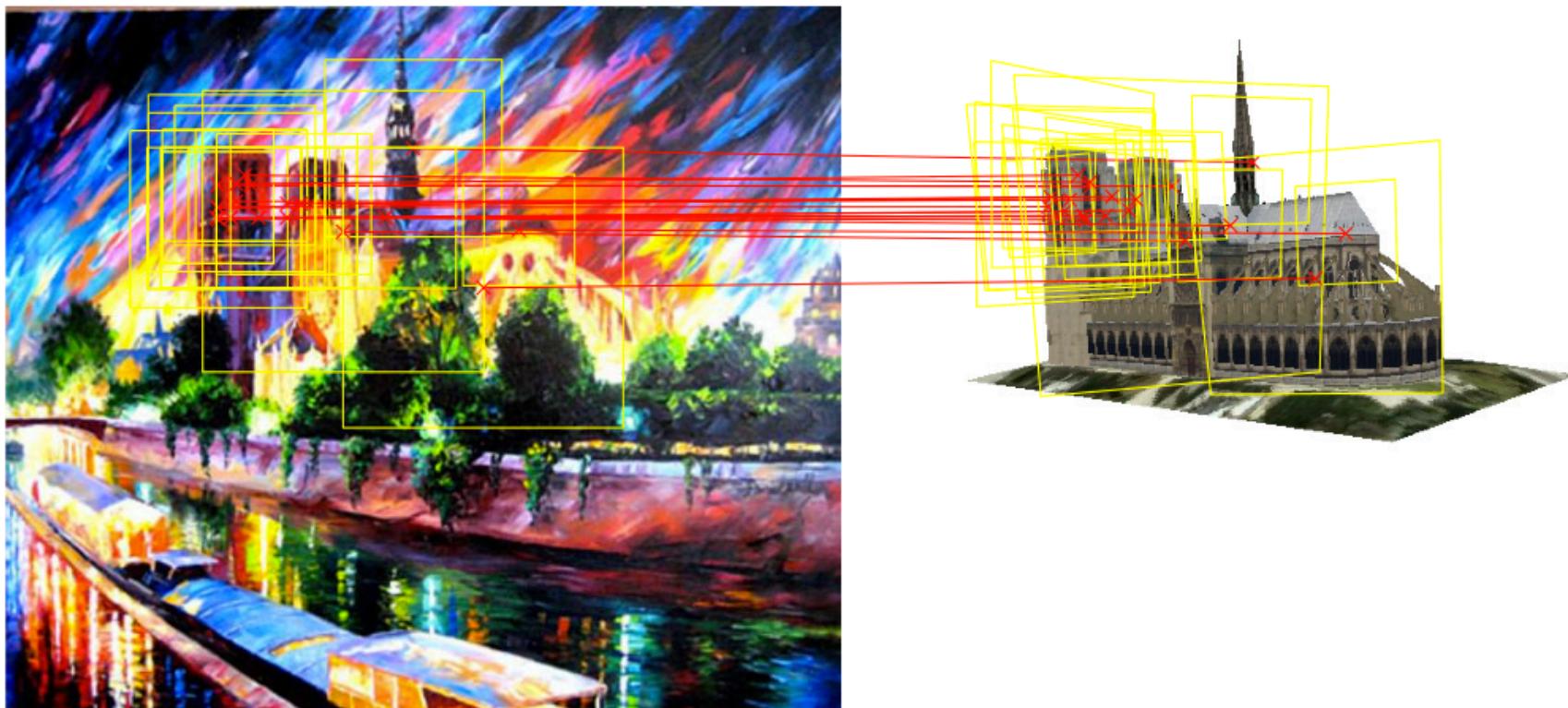


Inputs: paintings, drawings,
historical photographs,
reference 3D model



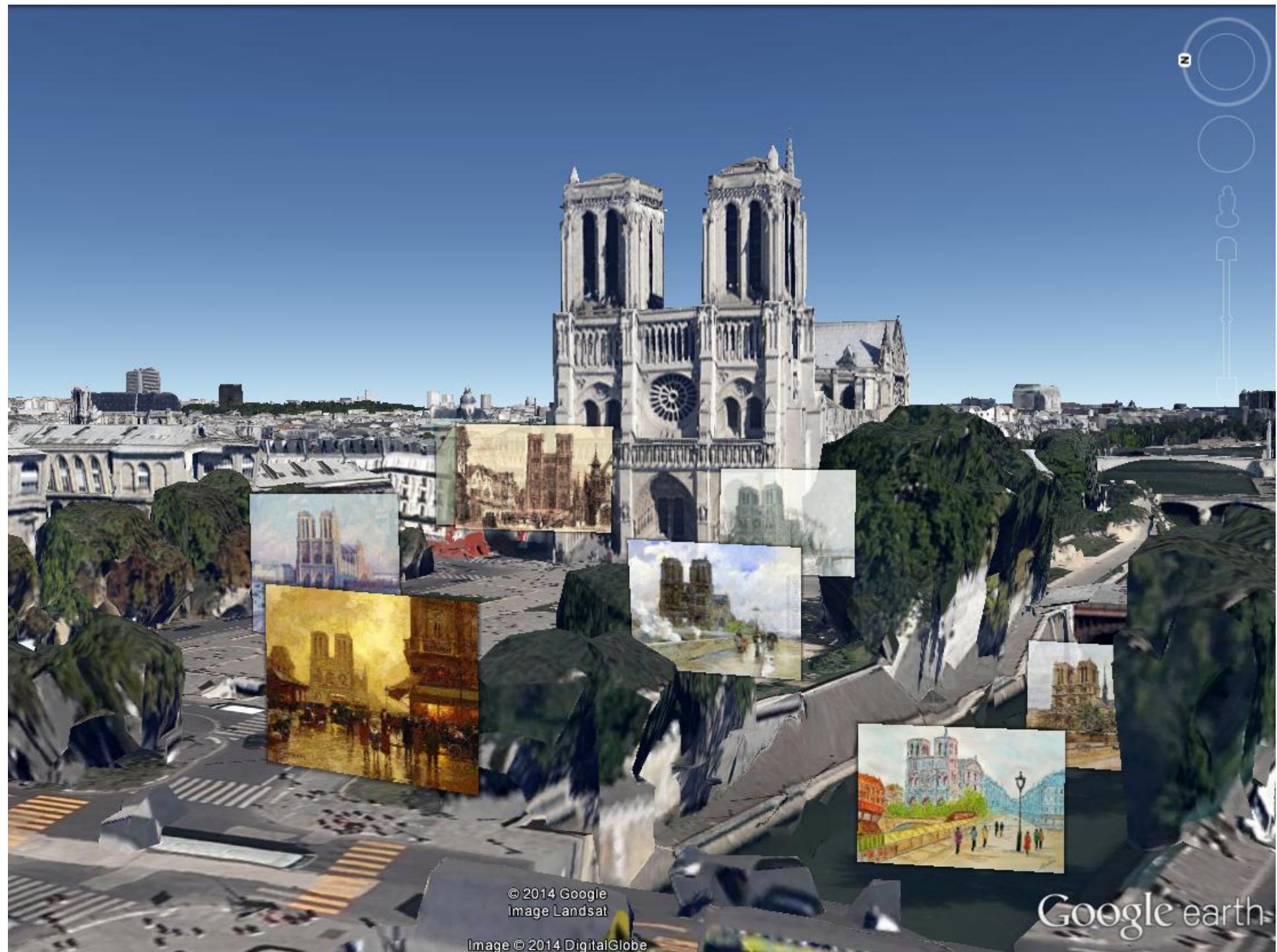
Output: recovered artist/camera
viewpoints

Matching non-photographic depictions



Geo-localization of historical and non-photographic depictions





© 2014 Google
Image Landsat

Image © 2014 DigitalGlobe

Google earth

Recognizing people

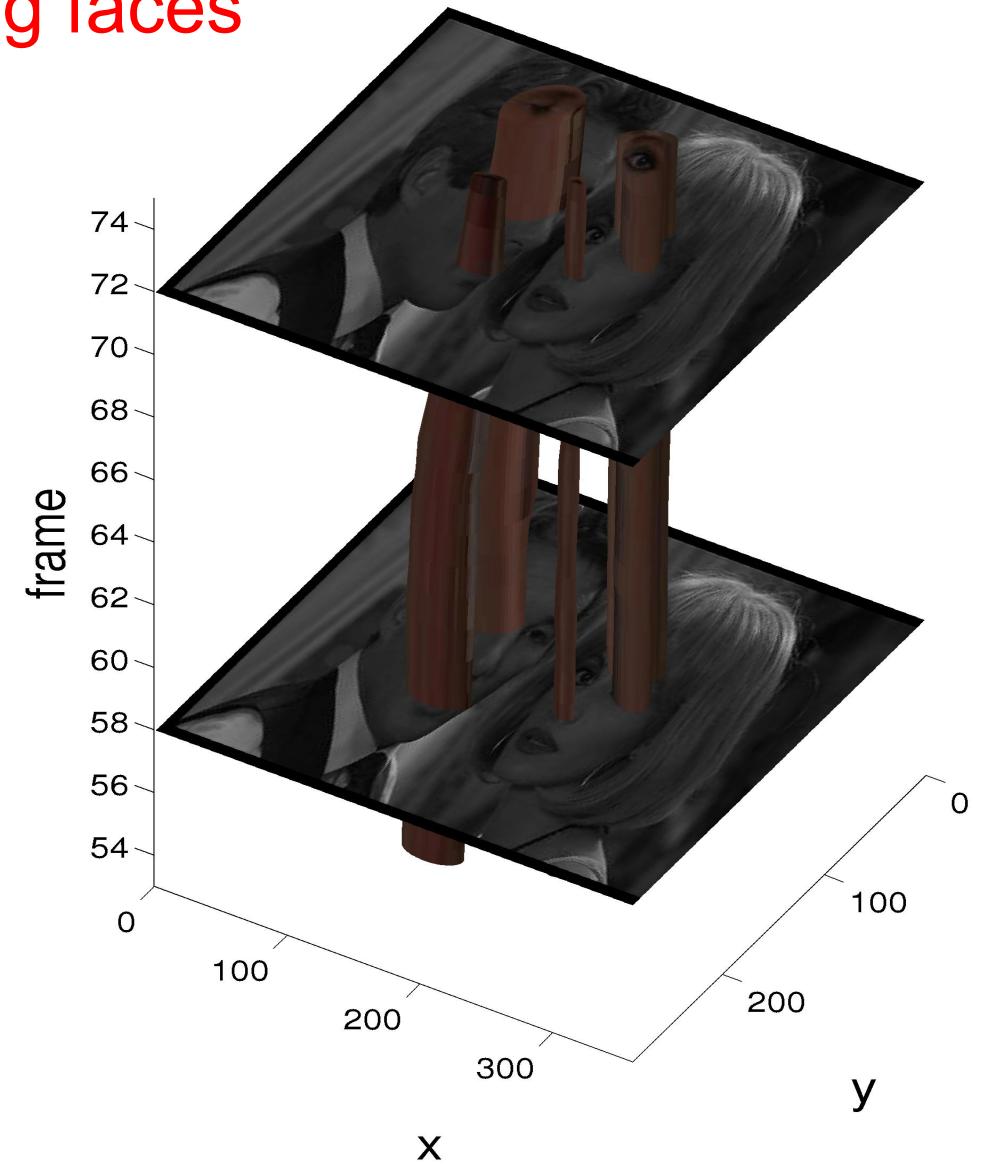


Edward Lewis
(Richard Gere)

Vivian
(Julia Roberts)

(Sivic, Everingham, Zisserman, 2005)

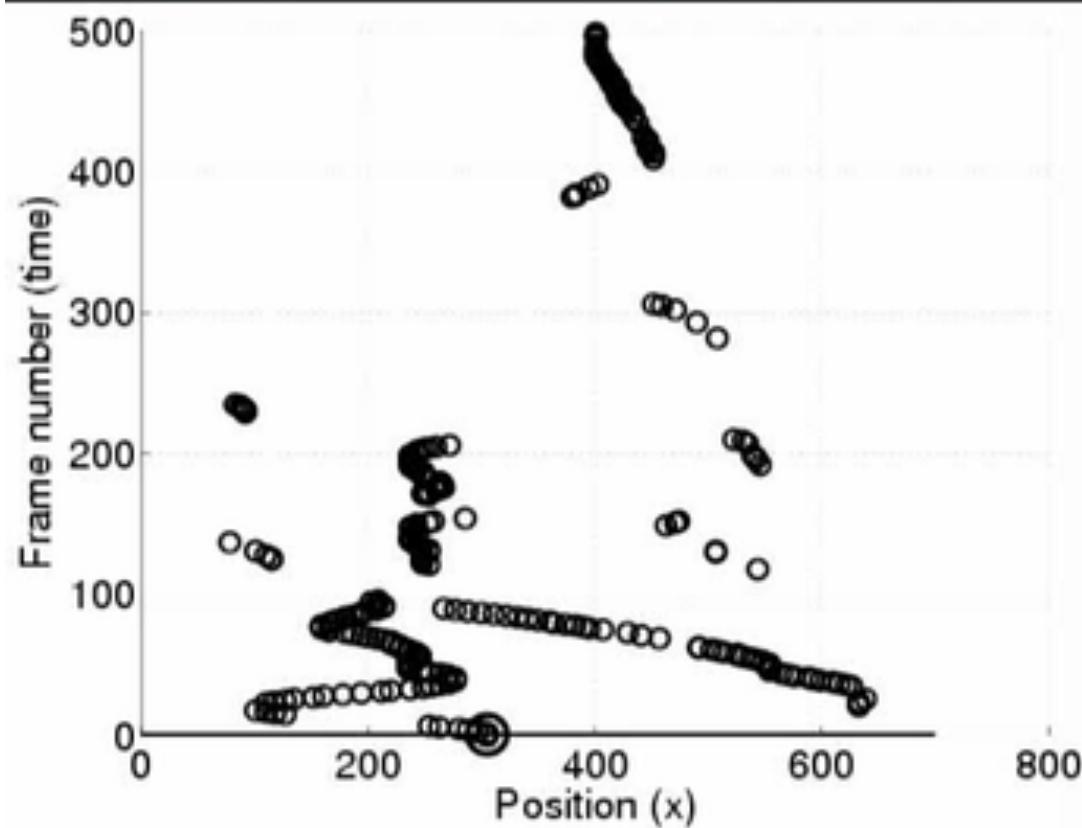
Faces: Region tubes for tracking faces



[Sivic, Everingham and Zisserman, 2005]

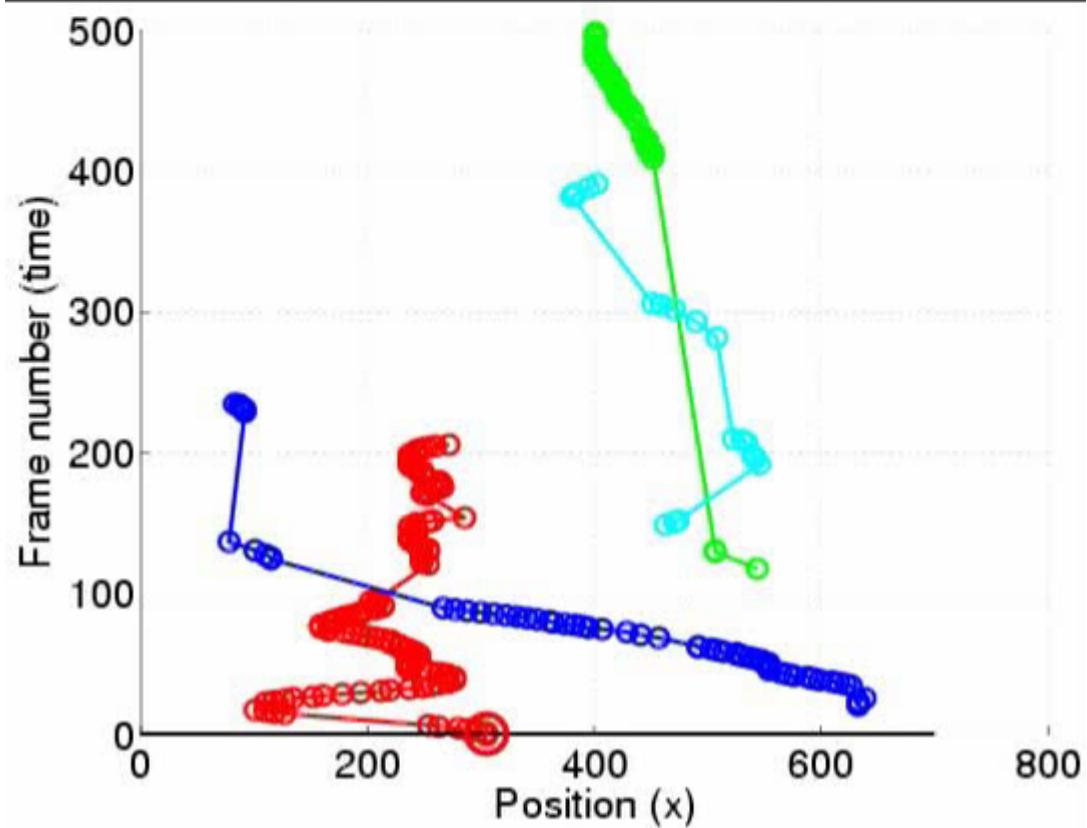


Raw face
detections



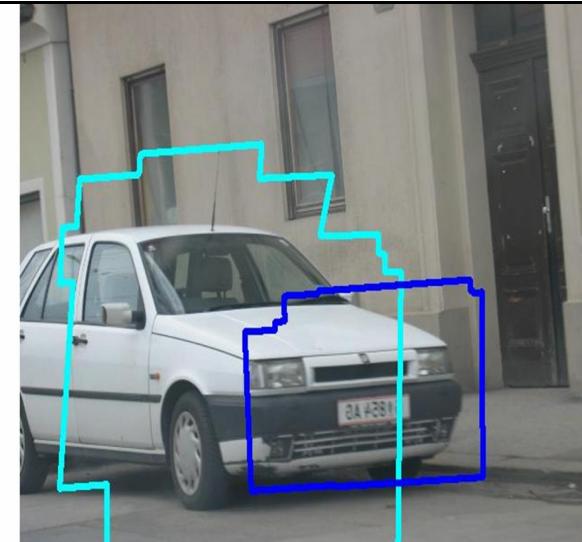
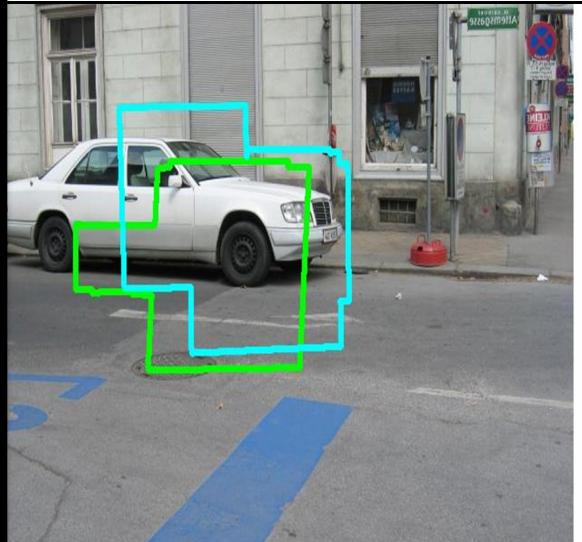
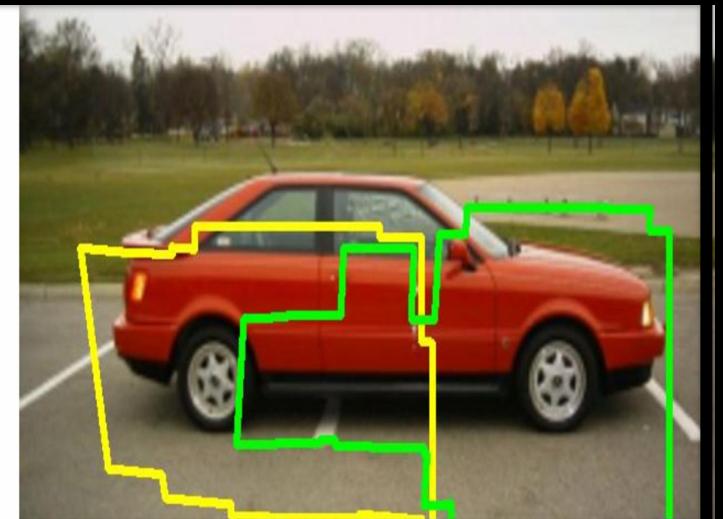
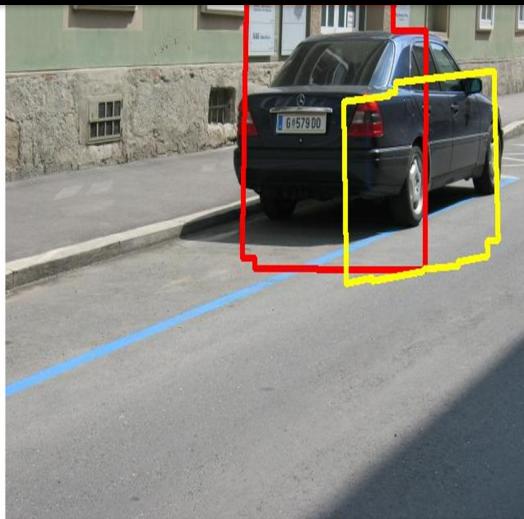


Tracking by
detection and
recognition



Connected face
tracks

Recognition



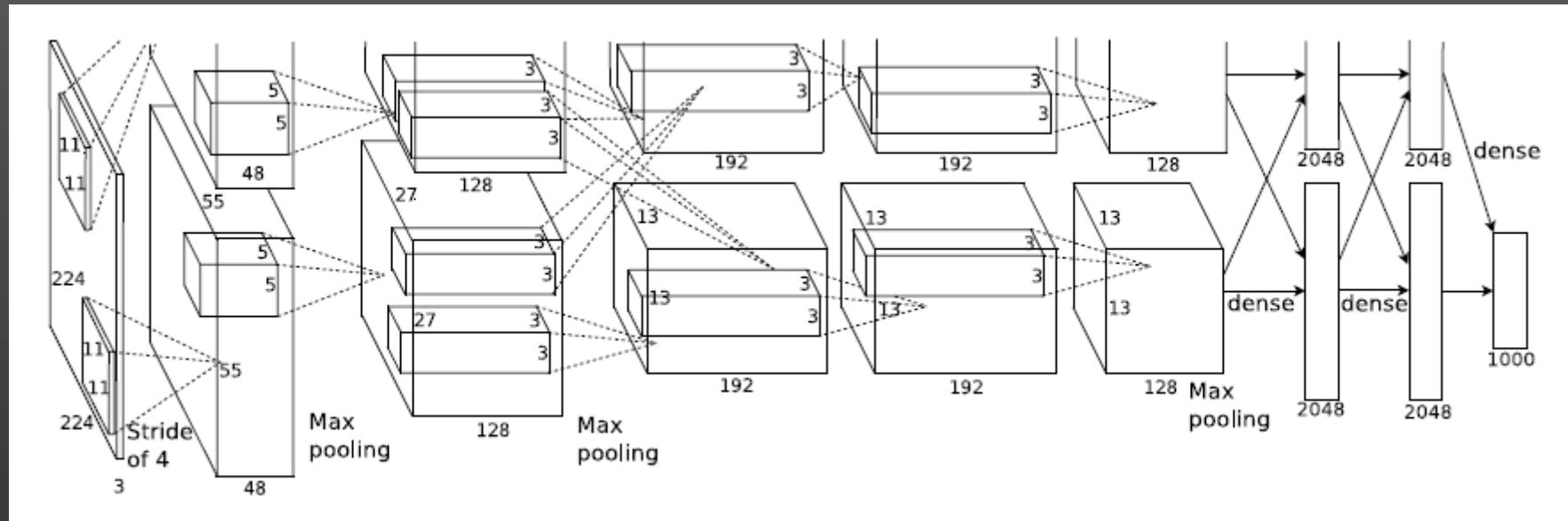
(Kushal et al., 2007)



(Aubry et al., 2014)

Convolutional neural networks

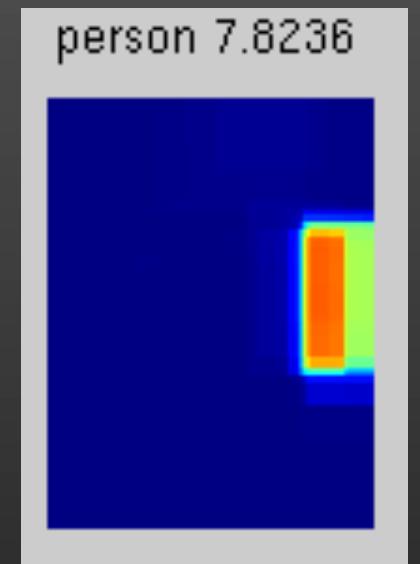
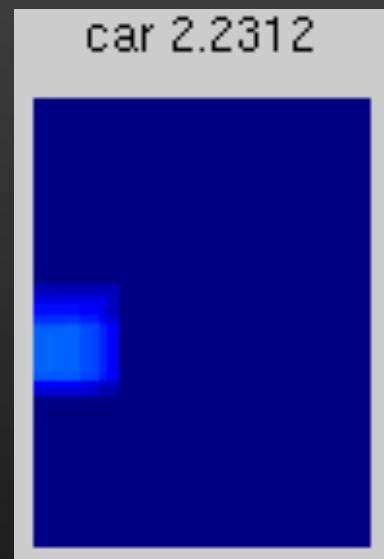
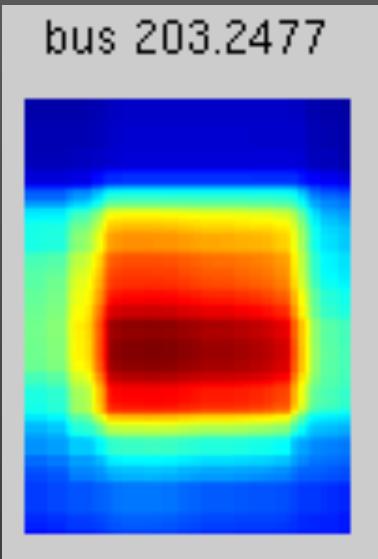
[Krizhevsky et al. NIPS'12]



Convolutional Neural Networks:

- The main principles are known since LeCun'88
- Has 60M parameters and 650K neurons.
- Success is determined by (a) lots of labeled images and (b) fast GPU implementation. **Both (a) and (b) have not been available until very recently.**

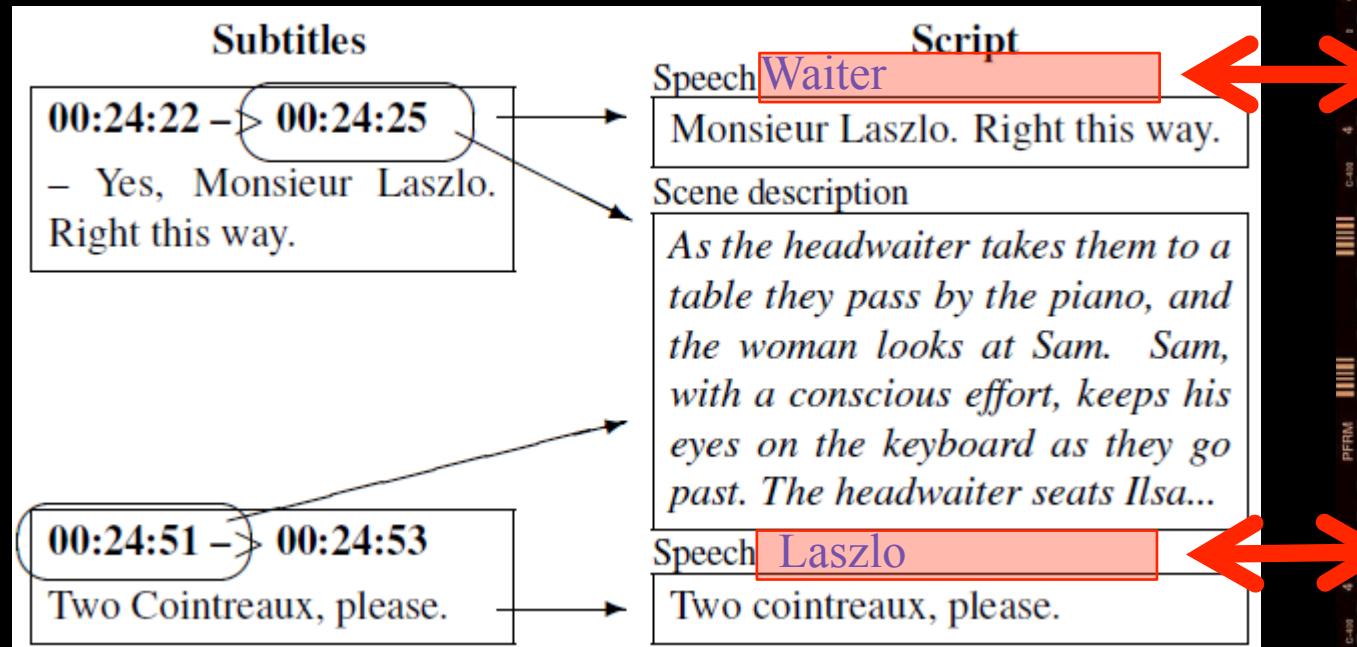
Some results



[Oquab, Bottou, Laptev, Sivic, CVPR 2014]

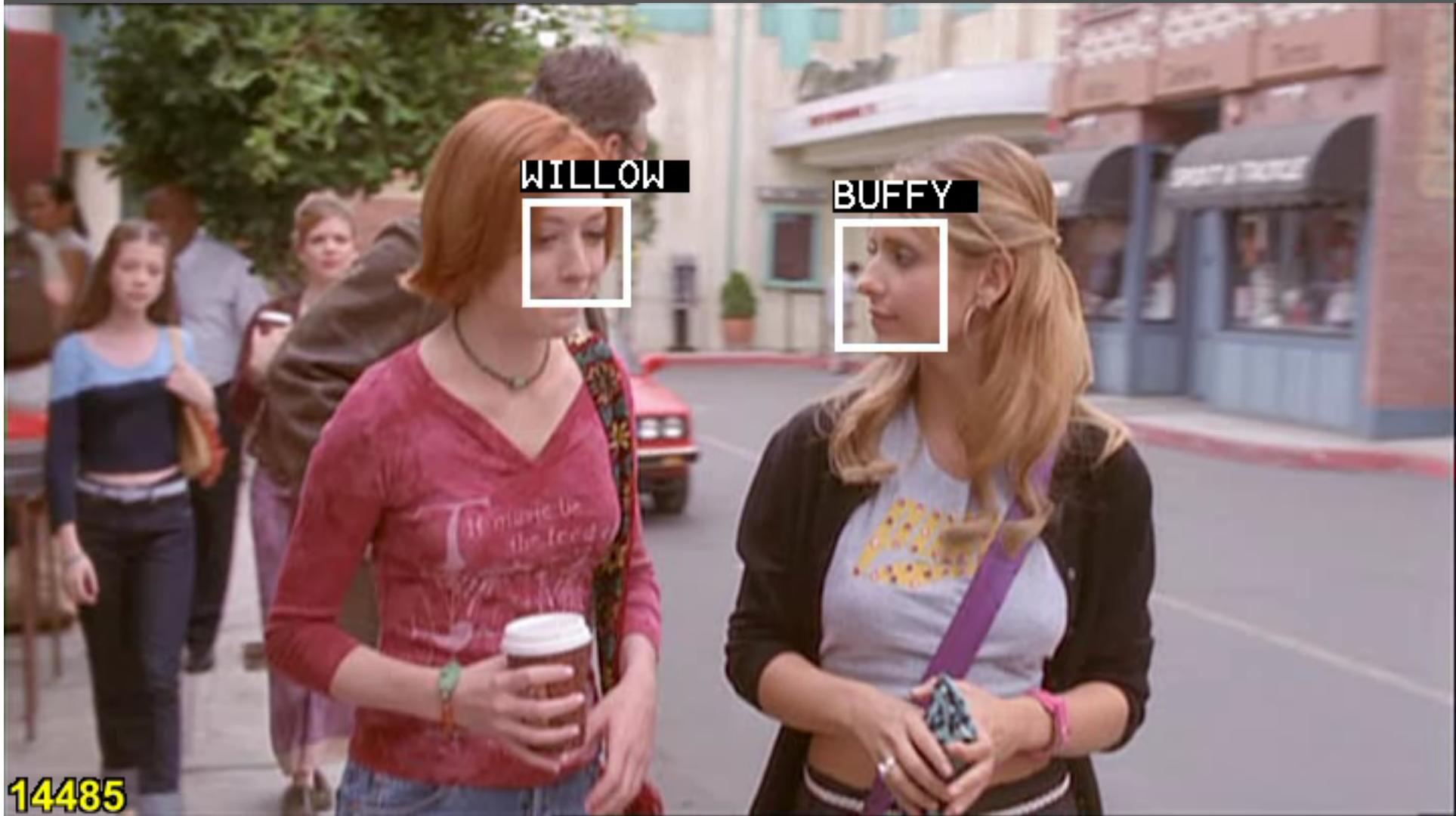
Automatic learning from video scripts

Input: Videos with aligned shooting scripts.



Output: Recognizer for each character in the video

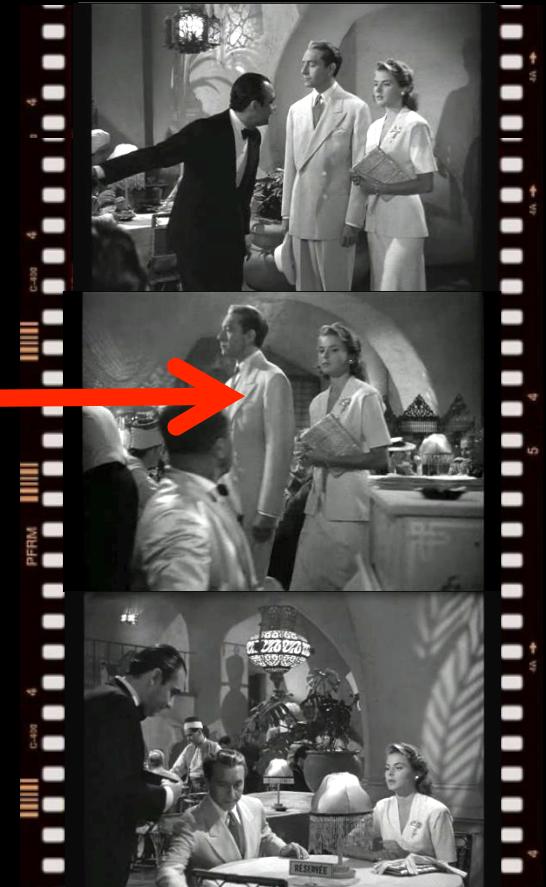
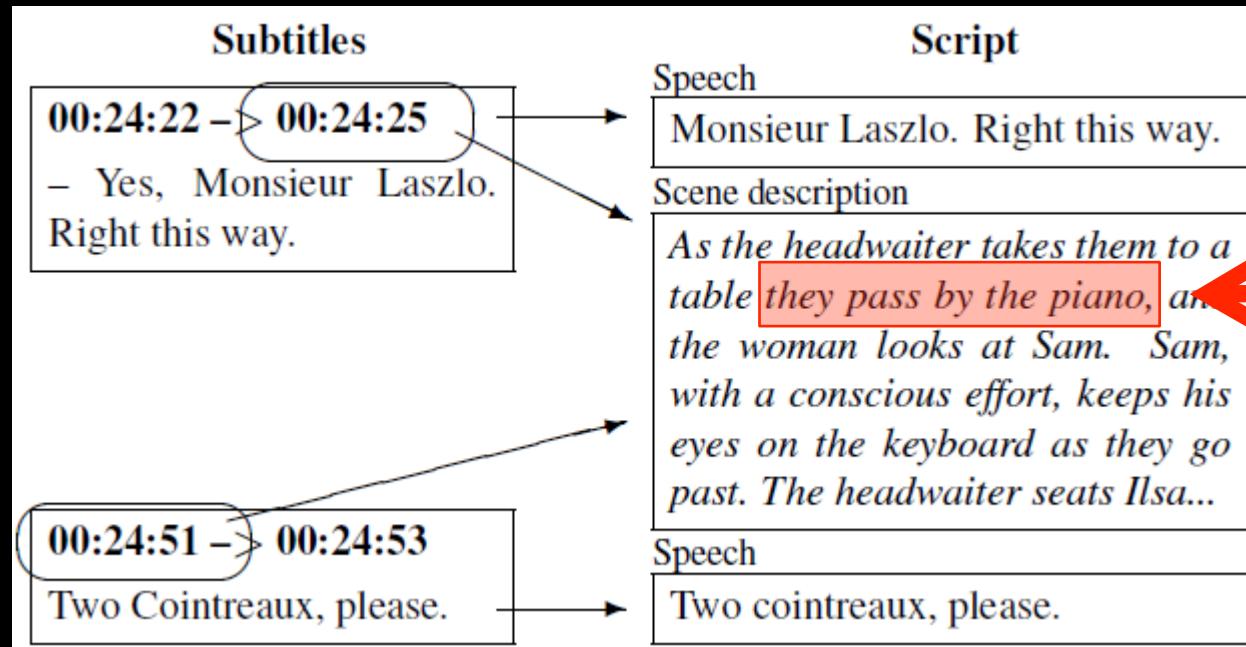
Recognizing people



(Everingham, Sivic, Zisserman, 2009)

Automatic learning from video scripts

Input: Videos with aligned shooting scripts.



Output: detector of human actions.

See also [Laptev, Marszałek, Schmid, Rozenfeld 2008]

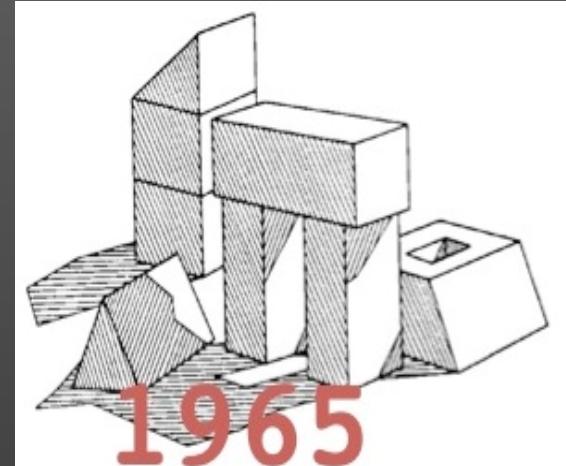
Weakly-supervised video interpretation

Clip number 0101

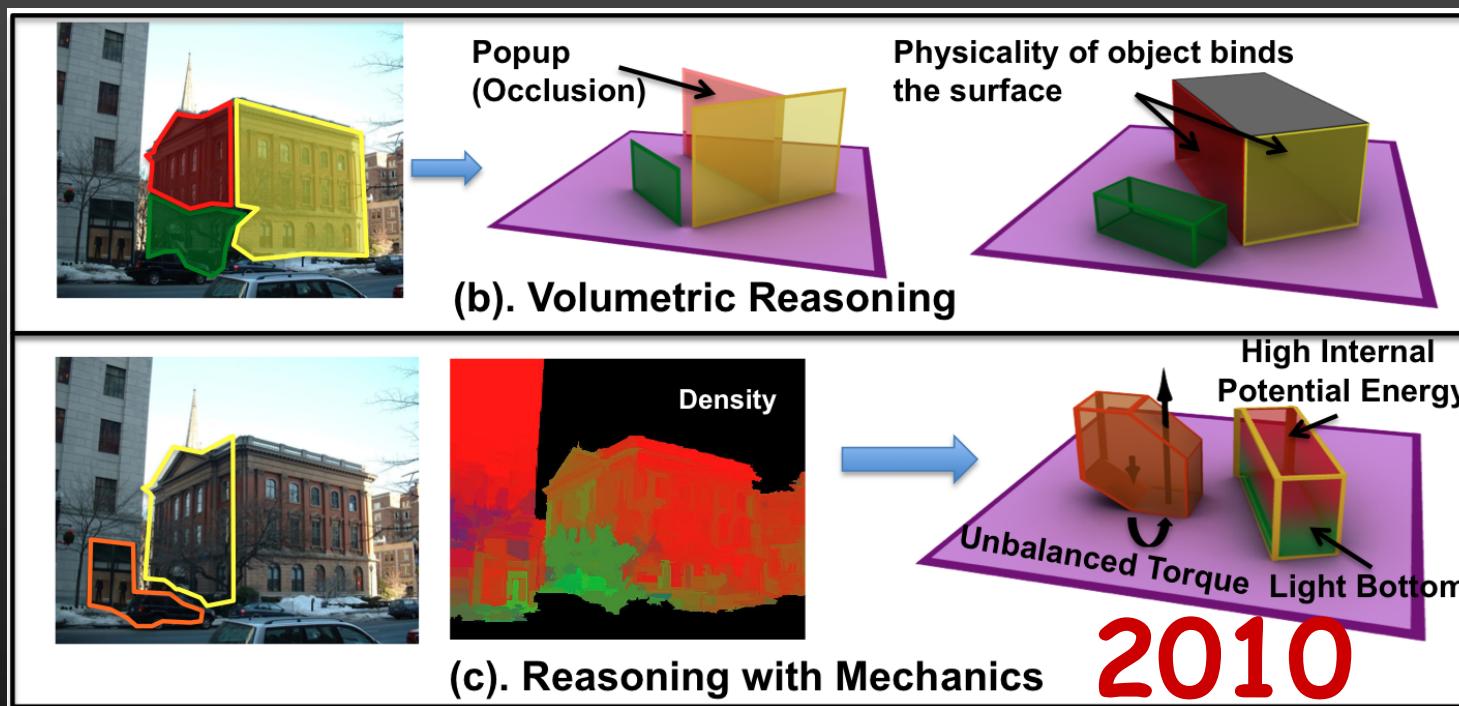
(Bojanowski et al., 2014)

What about scene understanding?

The blocks world revisited



1965



(Gupta, Efros, Hebert, ECCV'10)

