16-720
Computer Vision

Deva Ramanan

About me

Deva Ramanan

First class at CMU!

Moved from sunny California to Pittsburgh :)

Deva Ramanan
Associate Professor
Robotics Institute
Carnegie Mellon University
Elliot Dunlap Smith Hall (EDSH), Rm 221
deva@cs.cmu.edu
412-268-6966
Mailing address

Bio

I recently moved to CMU after spending 8 wonderful years as a faculty at UC-Irvine. A more formal bio is here.

Research

My research focuses on computer vision, often motivated by the task of understanding people from visual data. My work tends to make heavy use of machine learning techniques, often using the human visual system as inspiration. For example, temporal processing is a key component of human perception, but is still relatively unexploited in current visual recognition systems. Machine learning from big (visual) data allows systems to learn subtle statistical regularities of the visual world. But humans have the ability to learn from very few examples. Here's a recent talk (from 2015) that discusses some thoughts on these issues.

About you

Majors?

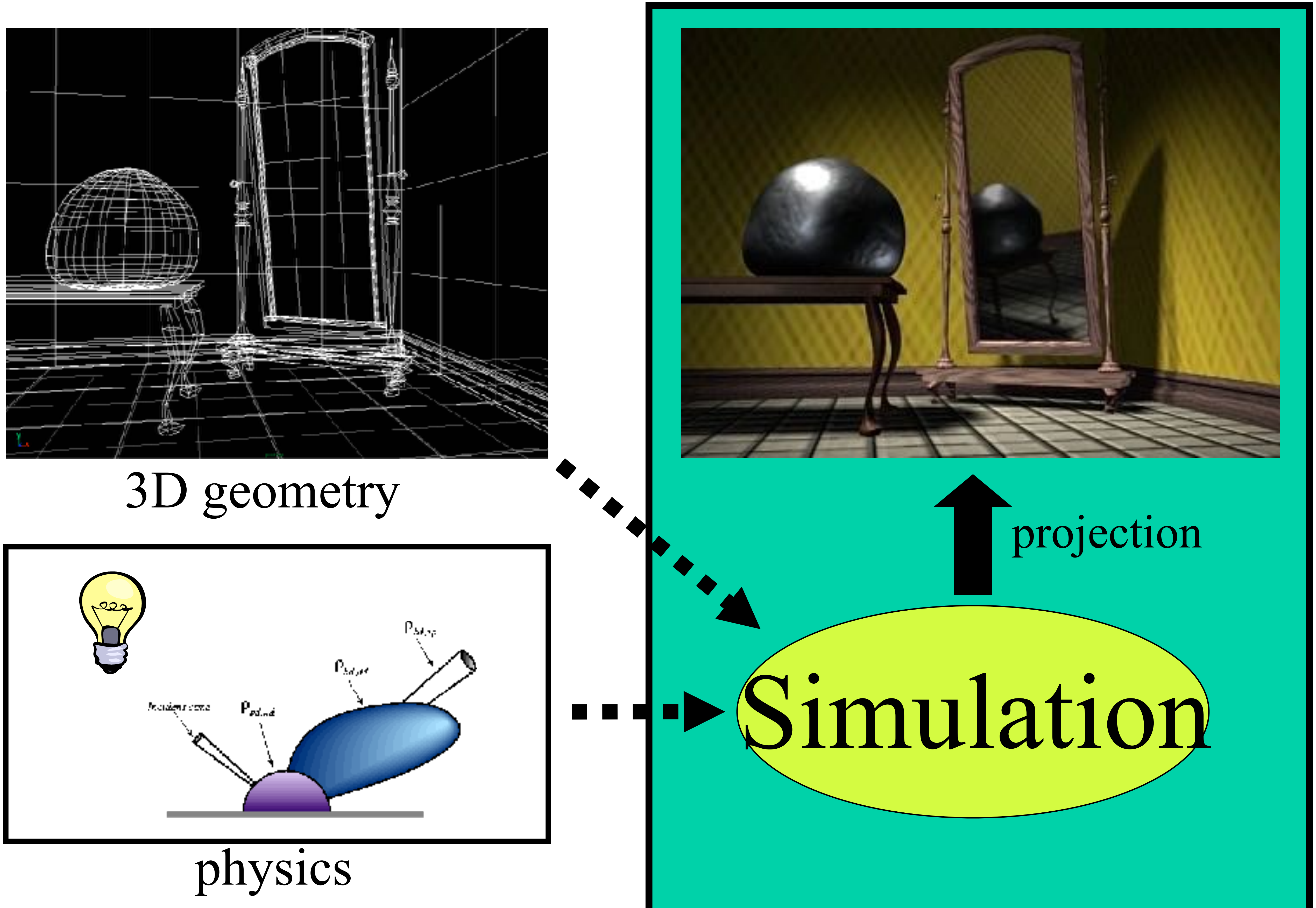
Years?

Related classes?
(computer graphics, machine learning,…)

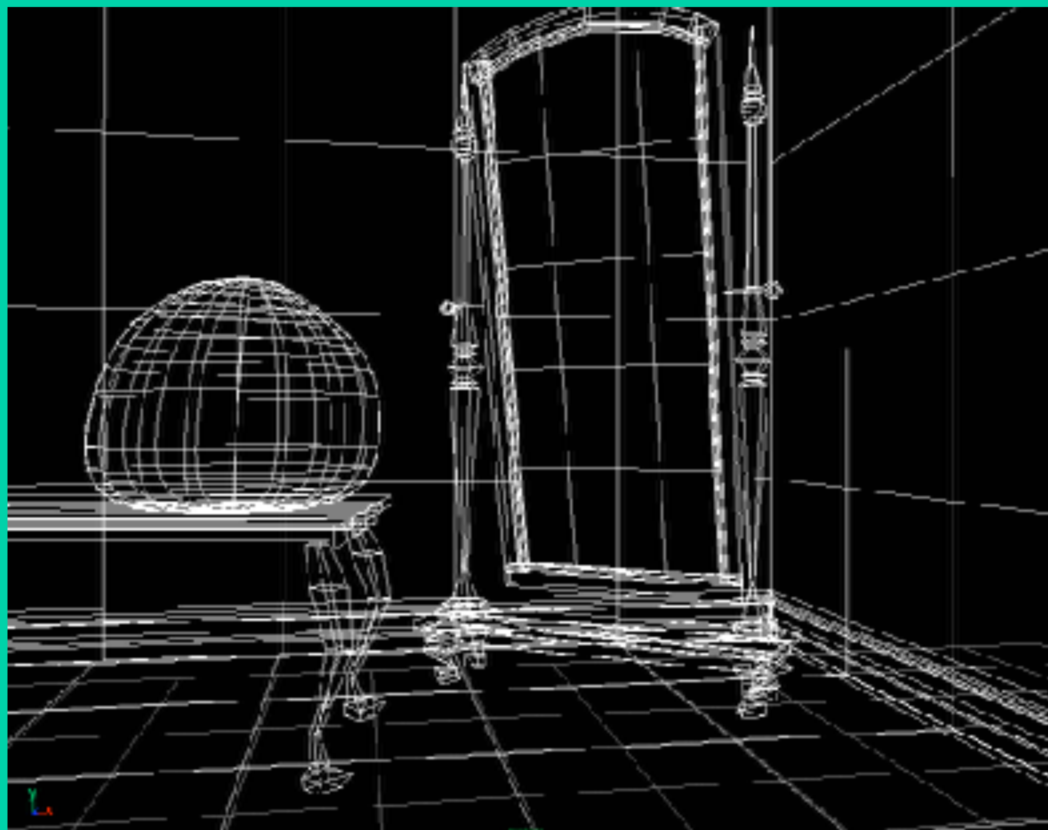
Outline

- What is this class about?
- Logistics for class (homeworks, grading)
- Historical perspective in computer vision
- HW0

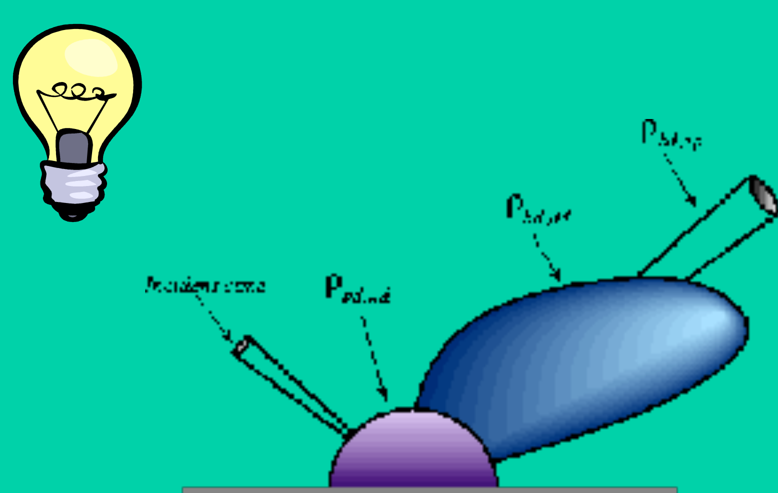
Related field: computer graphics



What is computer vision?



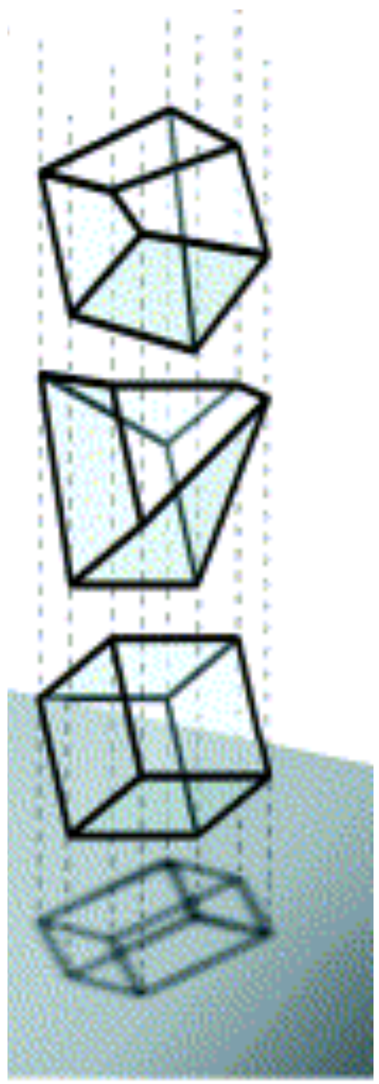
3D geometry

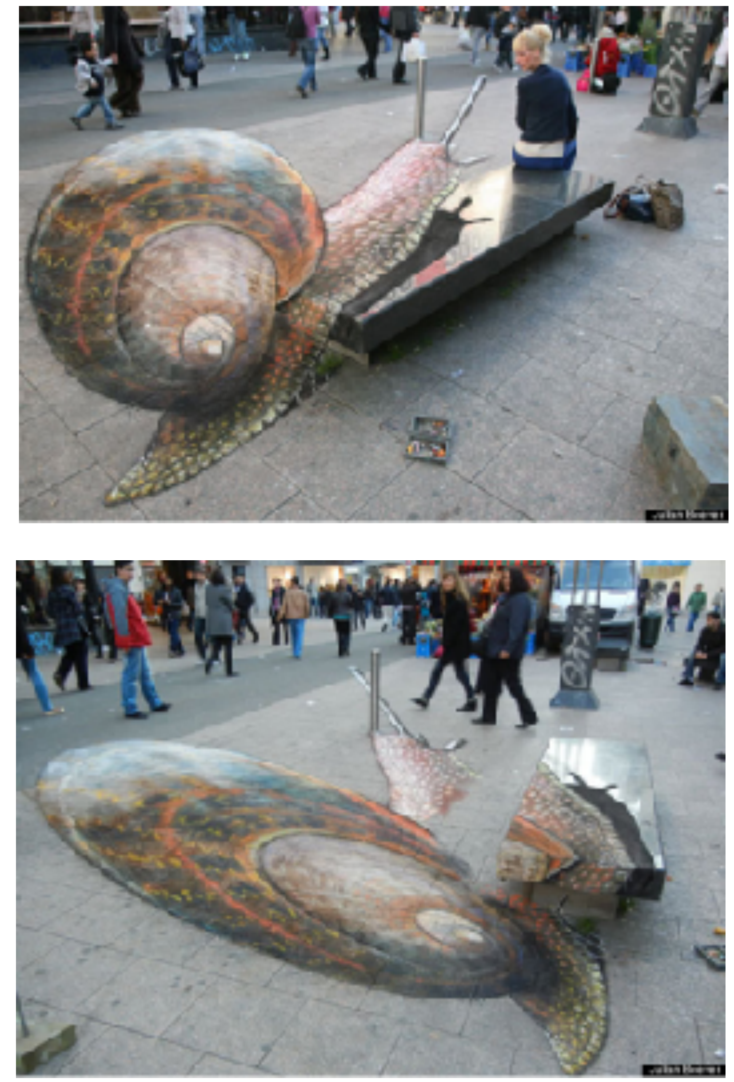


physics

Estimation

Challenge: ill-posed problem





Certain 3D interpretations are more *probable* than others

We want to build models that capture such prior knowledge about the word

Why do we want to understand the world from images (or videos)?

Visual media is everywhere

8 years worth of video is uploaded to YouTube
250 million photos are uploaded to FaceBook

... each day

By 2018, 80% of internet traffic will be video data

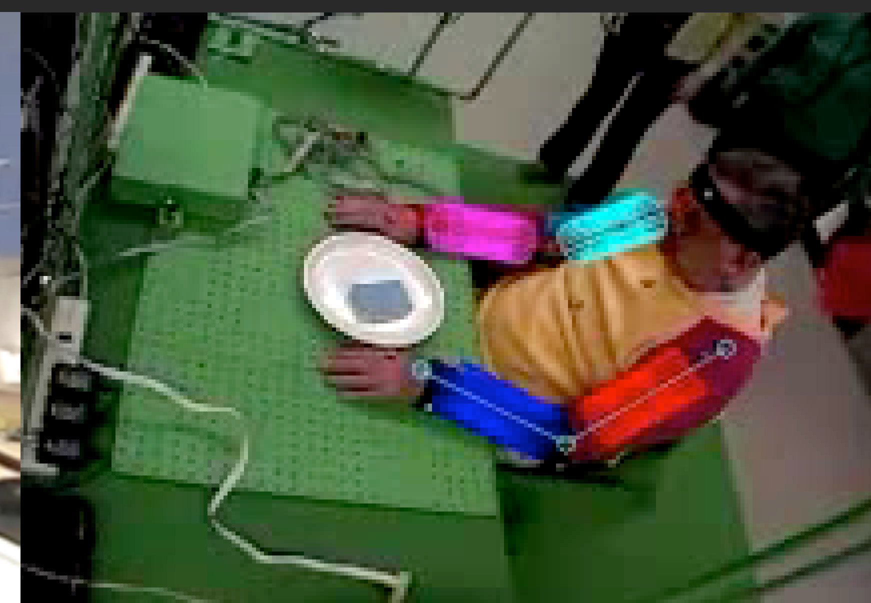
Distributed sensors

Everyone is carrying a camera in their back pocket

Reconstructing the (4D) world

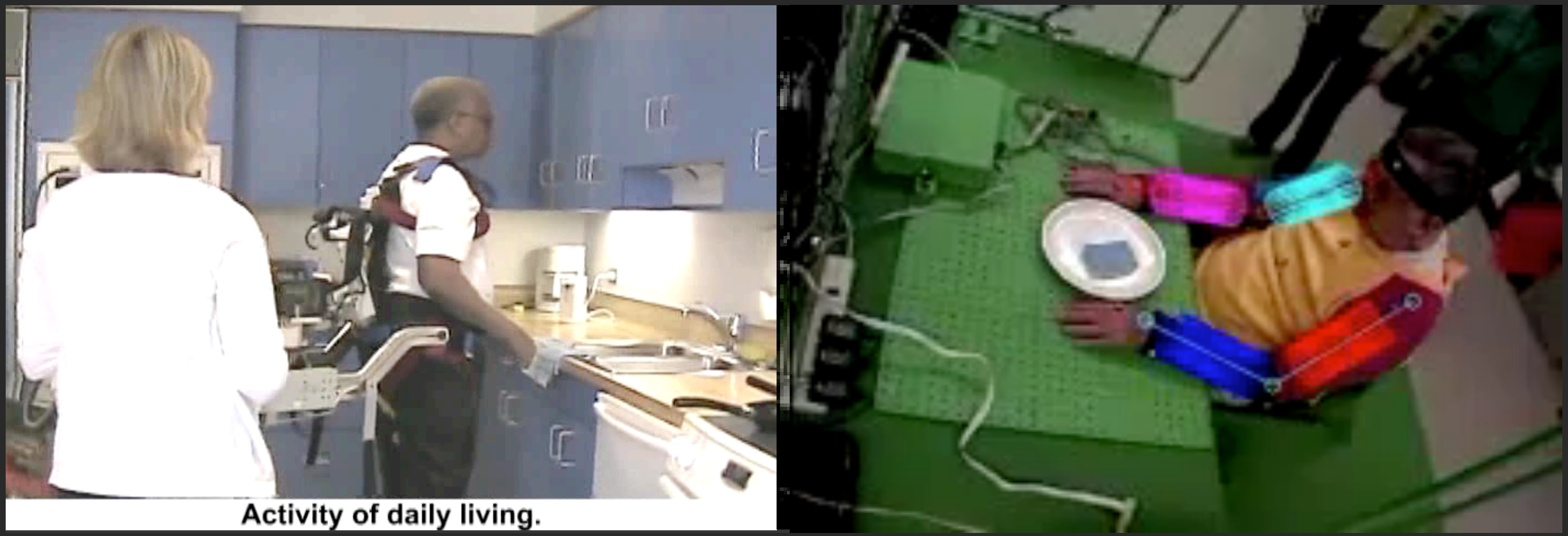
Photo Tourism (UWashington/Microsoft)

Applications: assistive technology



CMU Quality of Life Center

Applications: assistive technology



CMU Quality of Life Center

Application: field robotics

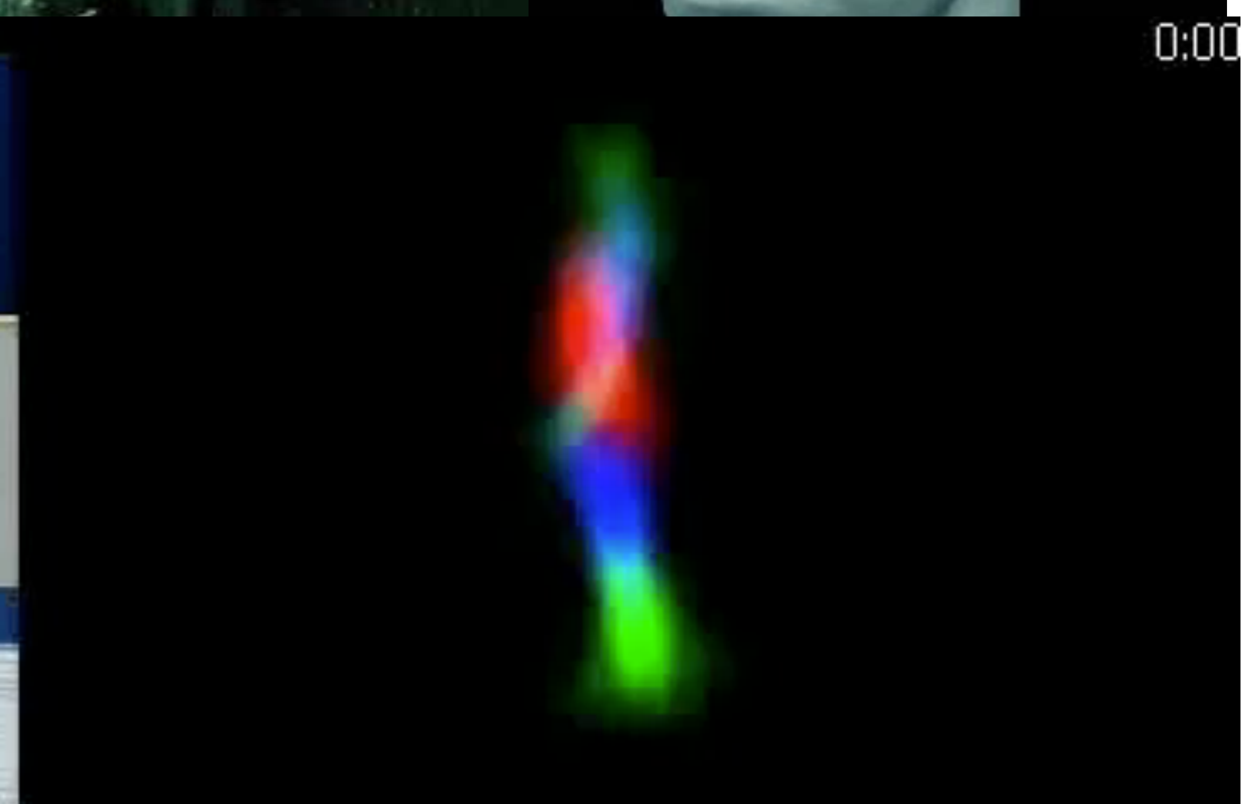




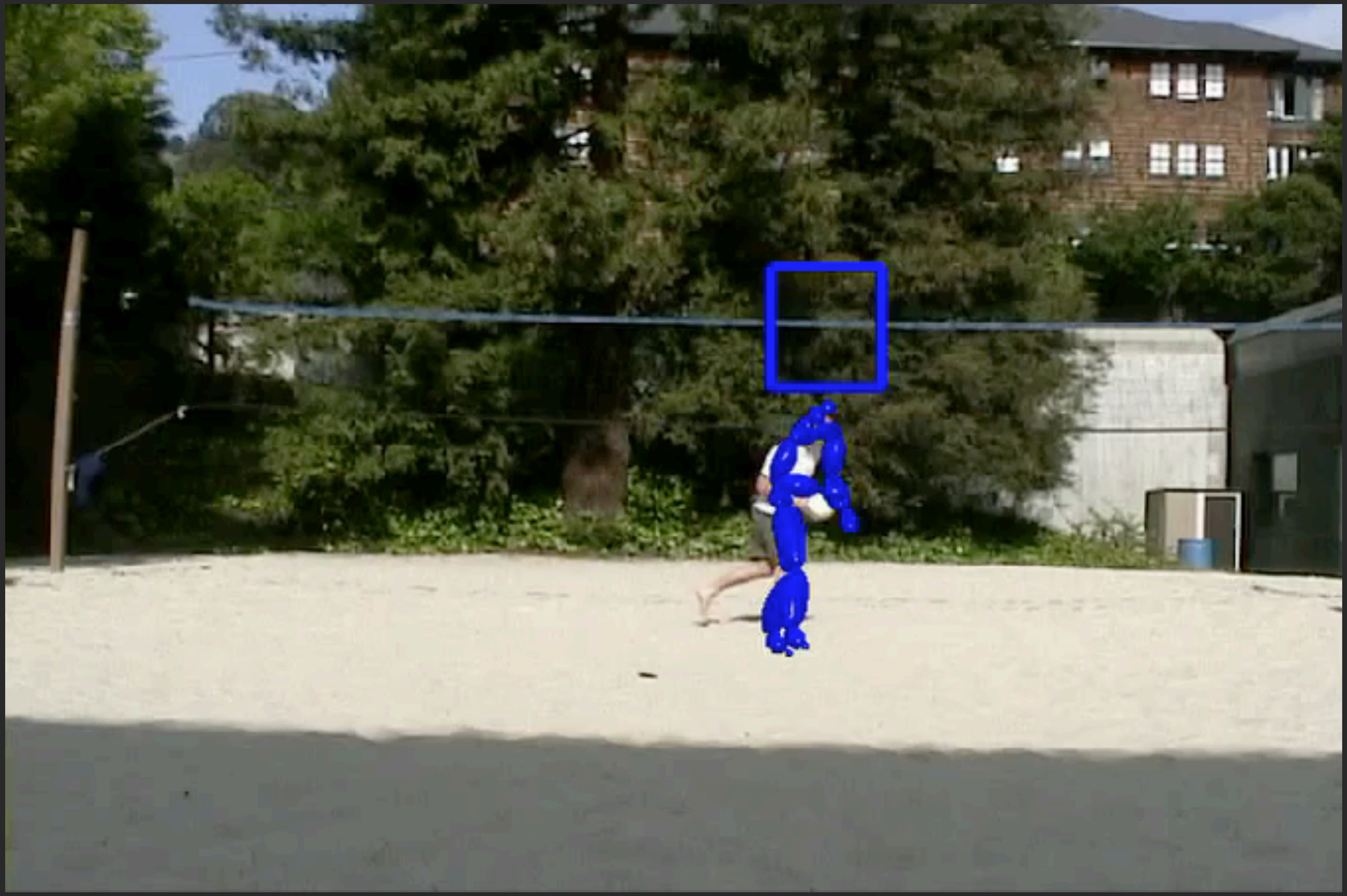
Applications: autonomous vehicle navigation

Google,Uber,…

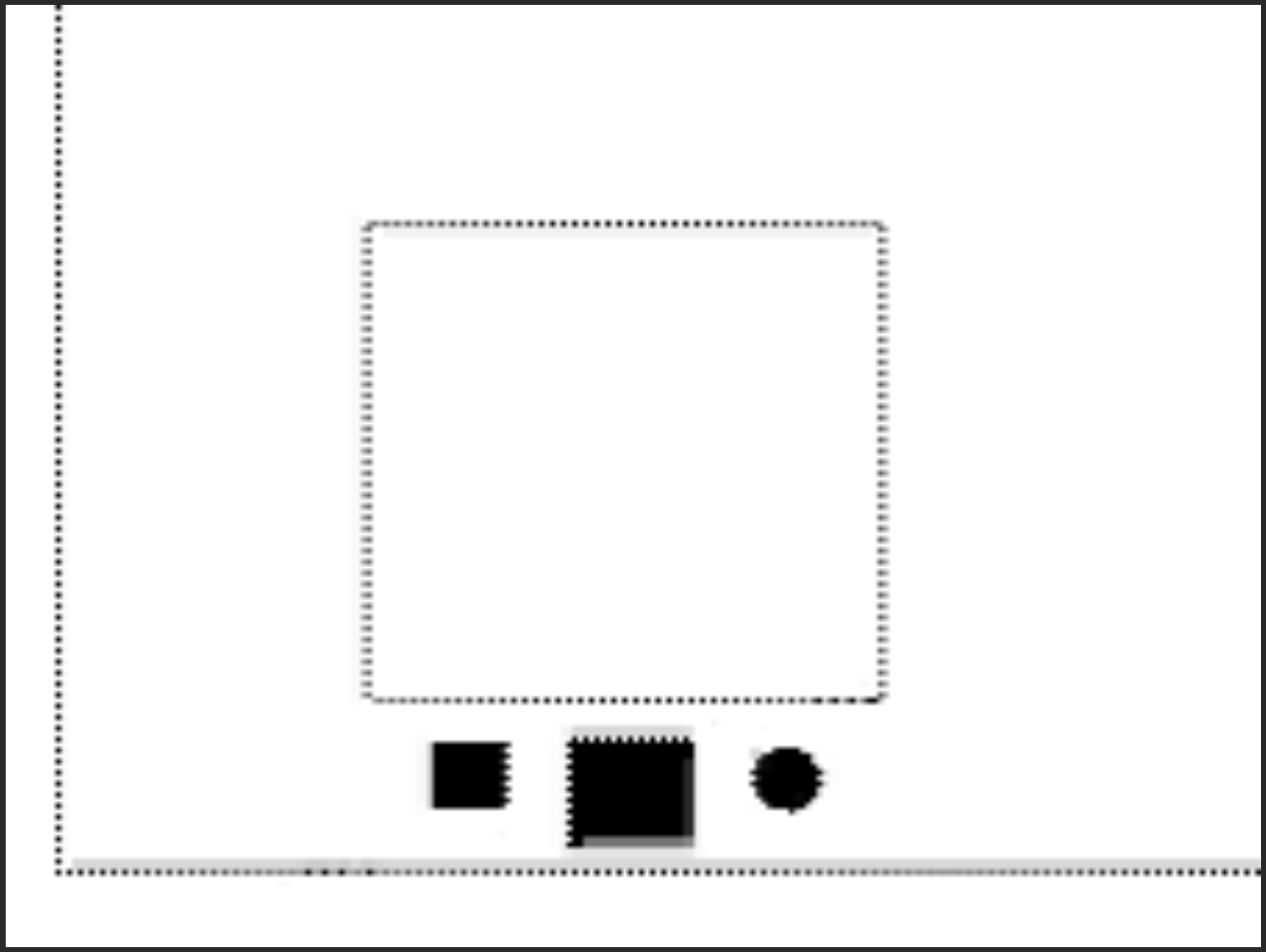
Applications:
motion/shape capture for graphics



Applications: surveillance



Aside: understanding intentions and goals is hard!

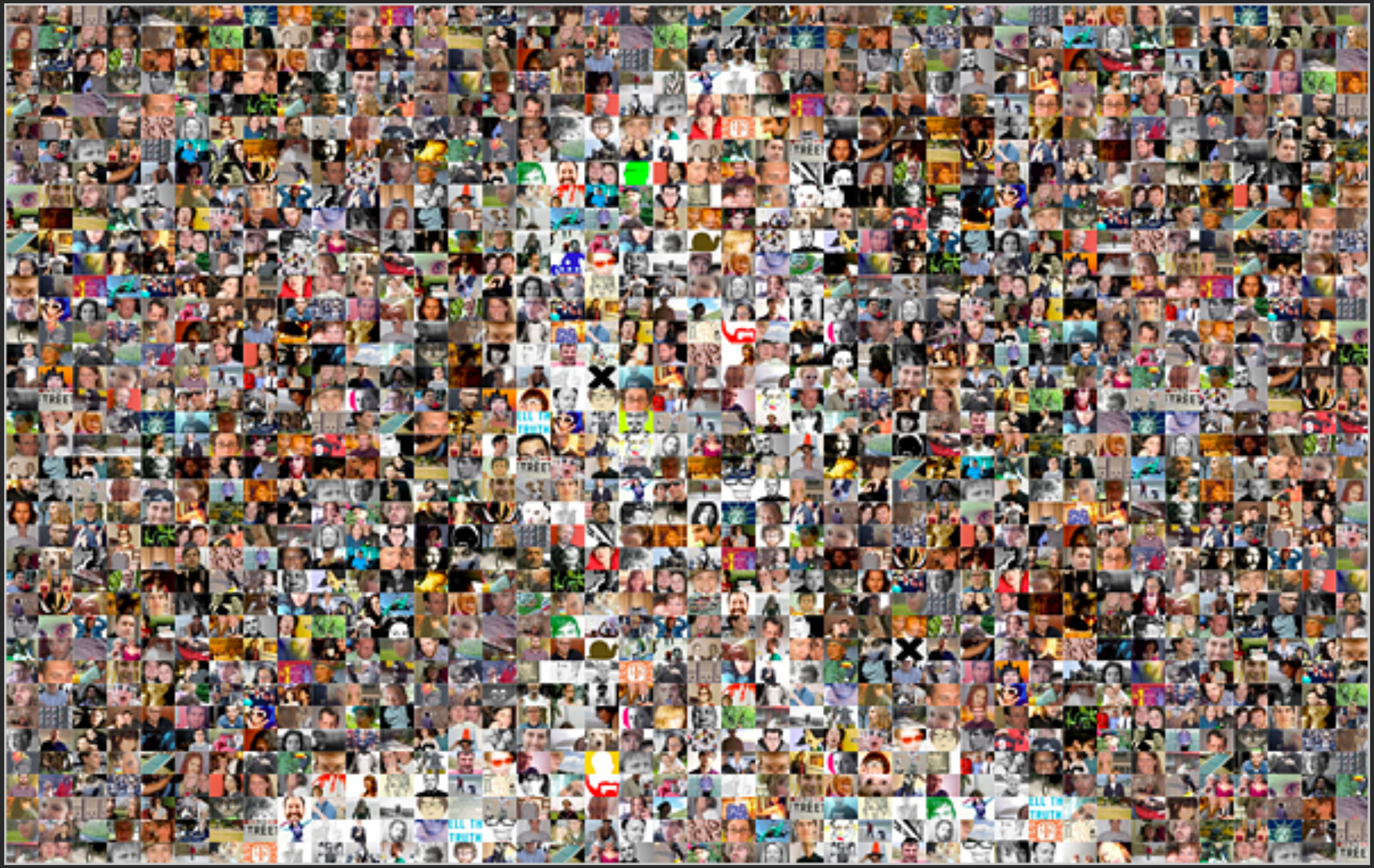


Applications: surveillance (cont'd)

"The work was painstaking and mind-numbing: One agent watched the same segment of video 400 times.The goal was to construct a timeline of images, following possible suspects as they moved along the sidewalks, building a narrative out of a random jumble of pictures from thousands of different phones and cameras. It took a couple of days, but analysts began to focus on two men in baseball caps who had brought heavy black bags into the crowd near the marathon's finish line but left without those bags."

Washington Post

Applications: image search



Applications: visual lifelogging

"Google Glass"

Applications: augmented reality

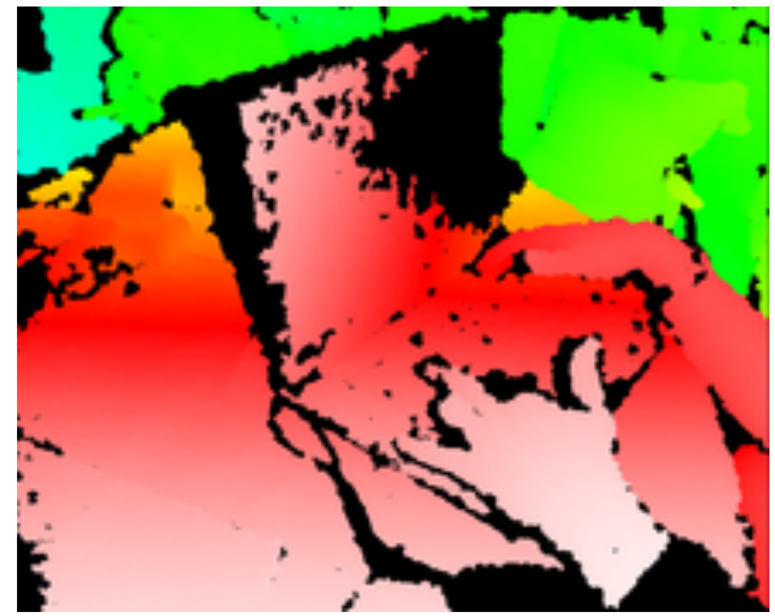
Occulus Rift: 2 billion

MagicLeap: .5 billion second stage funding (record)

Sports analytics

Nice [explanation](#) of first-down line on www.howstuffworks.com

Applications: gaming interfaces



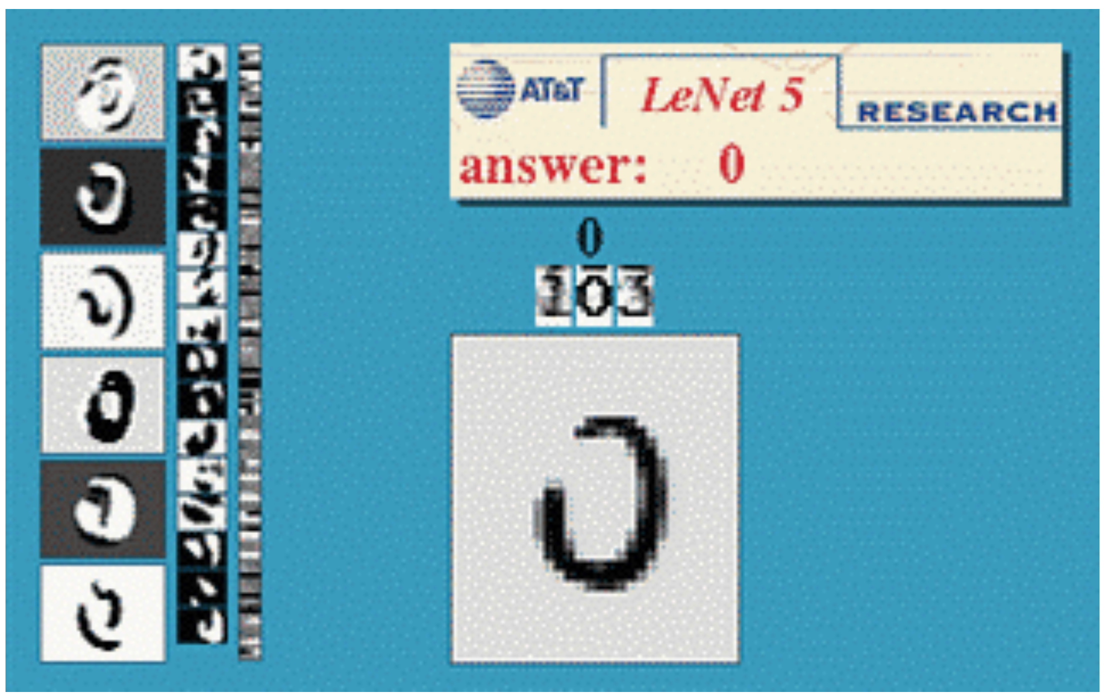
Kinect holds the Guinness World Record of being the "fastest selling consumer electronics device". It sold an average of 133,333 units per day with a total of 8 million units in its first 60 days.[18][19][20] 10 million units of the Kinect sensor have been shipped as of March 9, 2011.[1]

c.f. Wikipedia

Optical character recognition

Technology to convert scanned docs to text

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

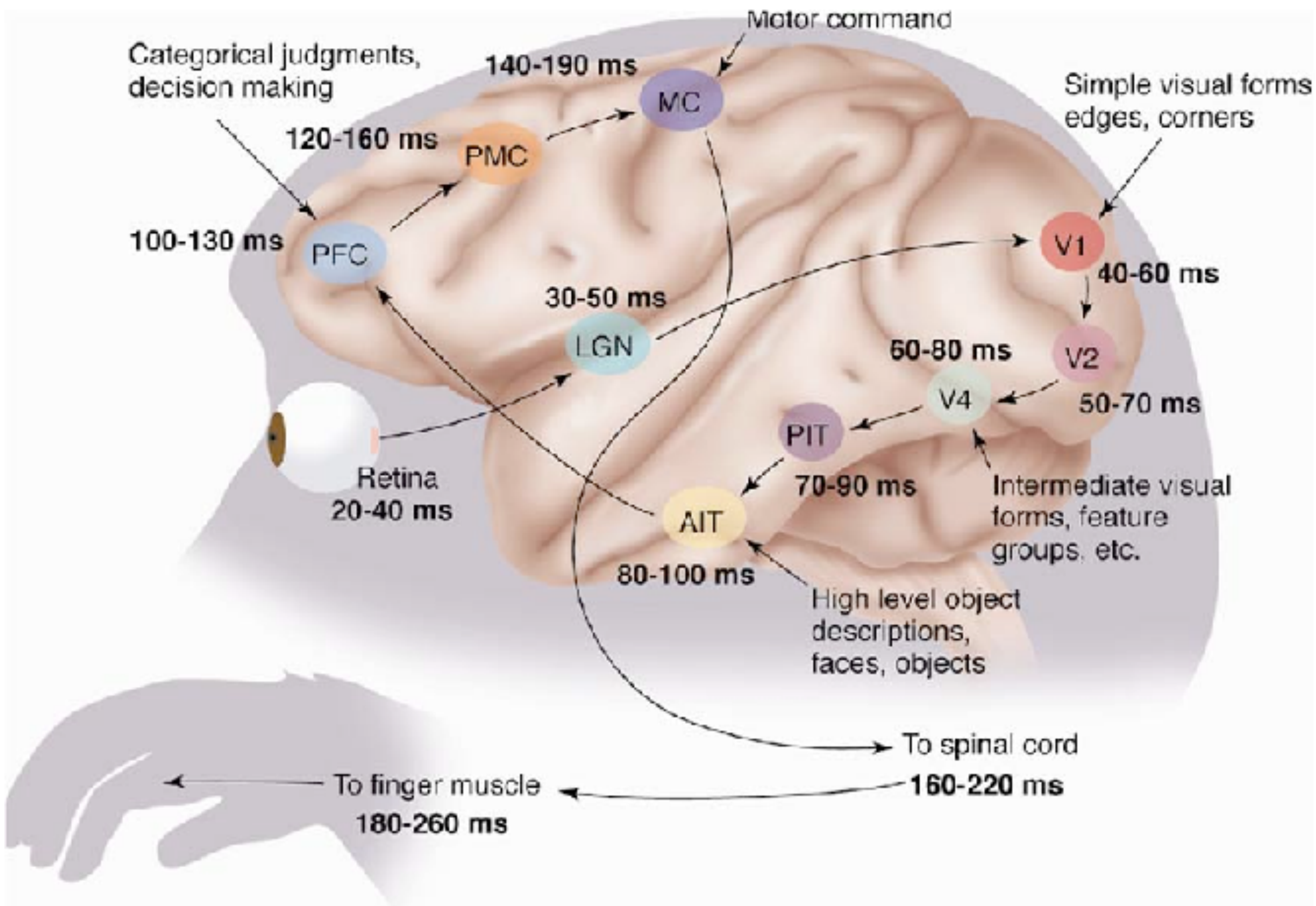


Digit recognition, AT&T labs
http://www.research.att.com/~yann/

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

Why is this interesting science?

Computational understanding of human perception



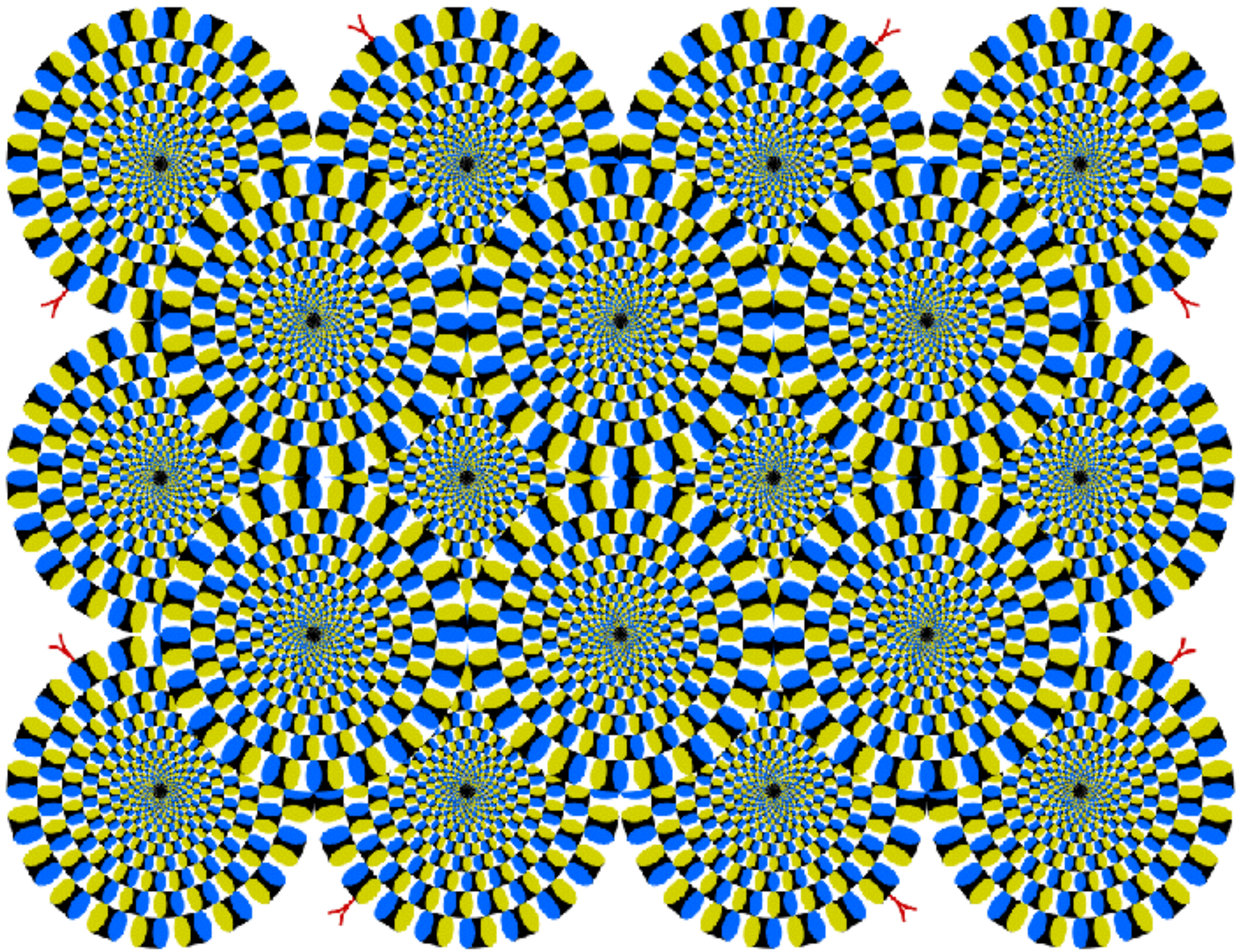
Vision is the most powerful of our senses
About 1/3 of brain is devoted to visual processing

Human perception has "shortcomings"

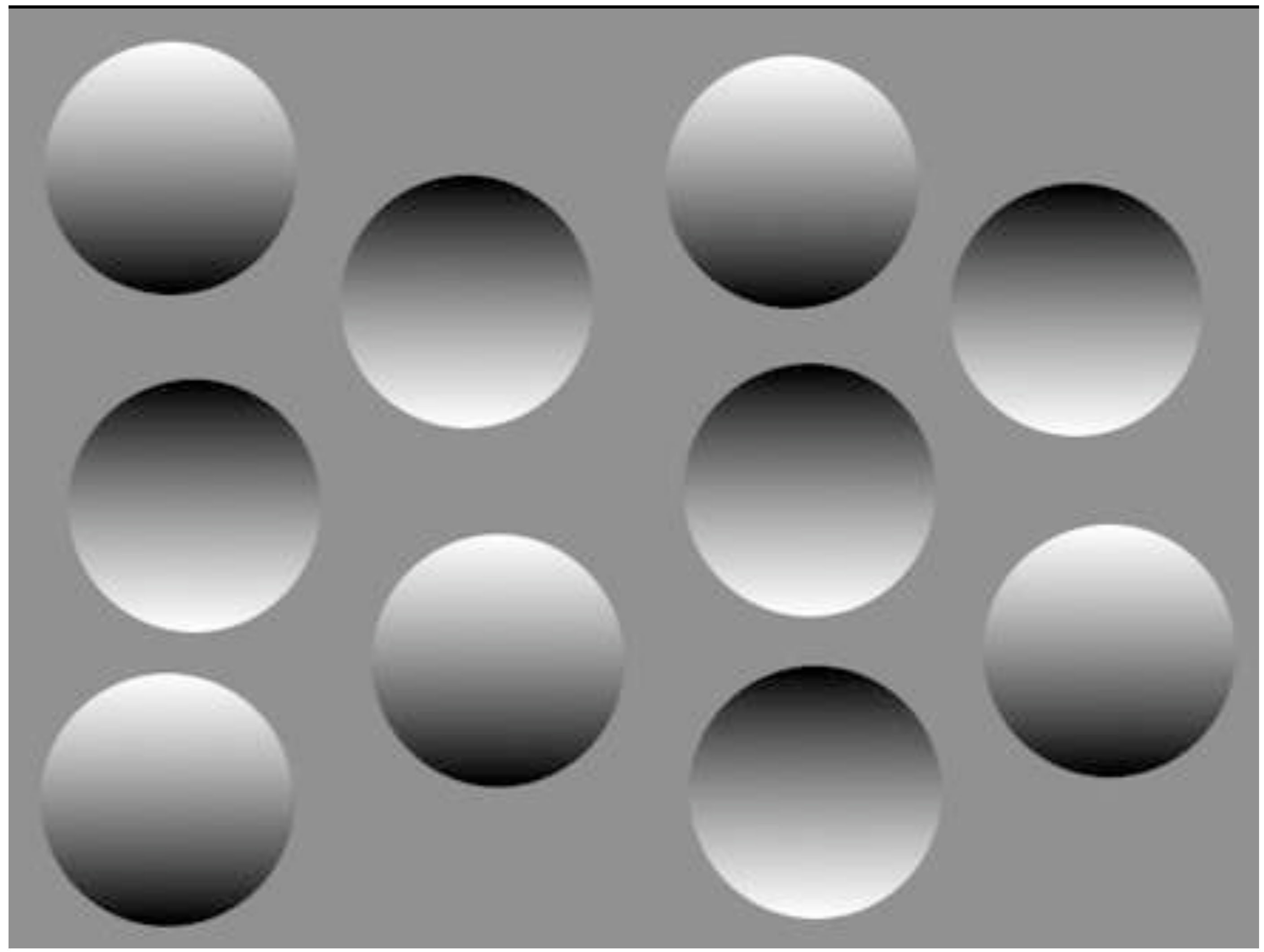
The human visual system exploits the ecological regularities of the world

Optical illusions are violations of this regularity

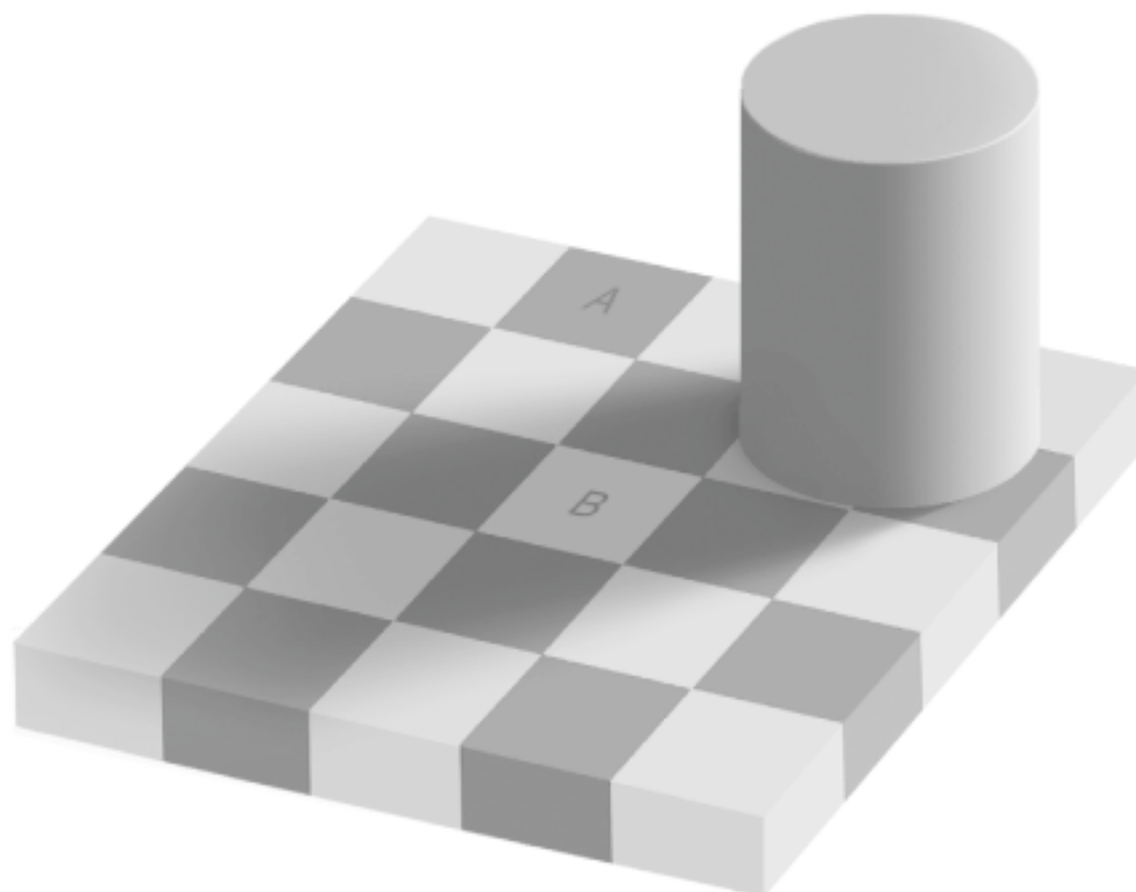
Illusory motion



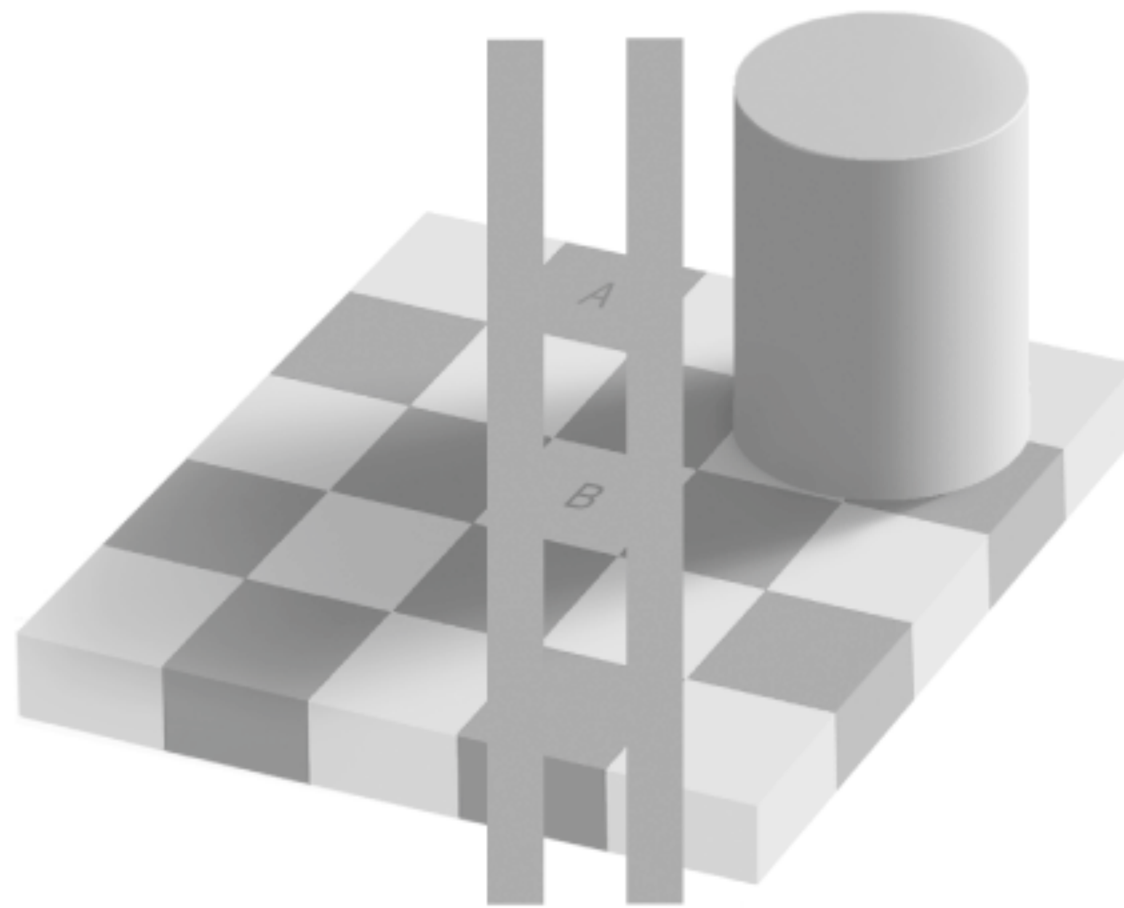
Shading



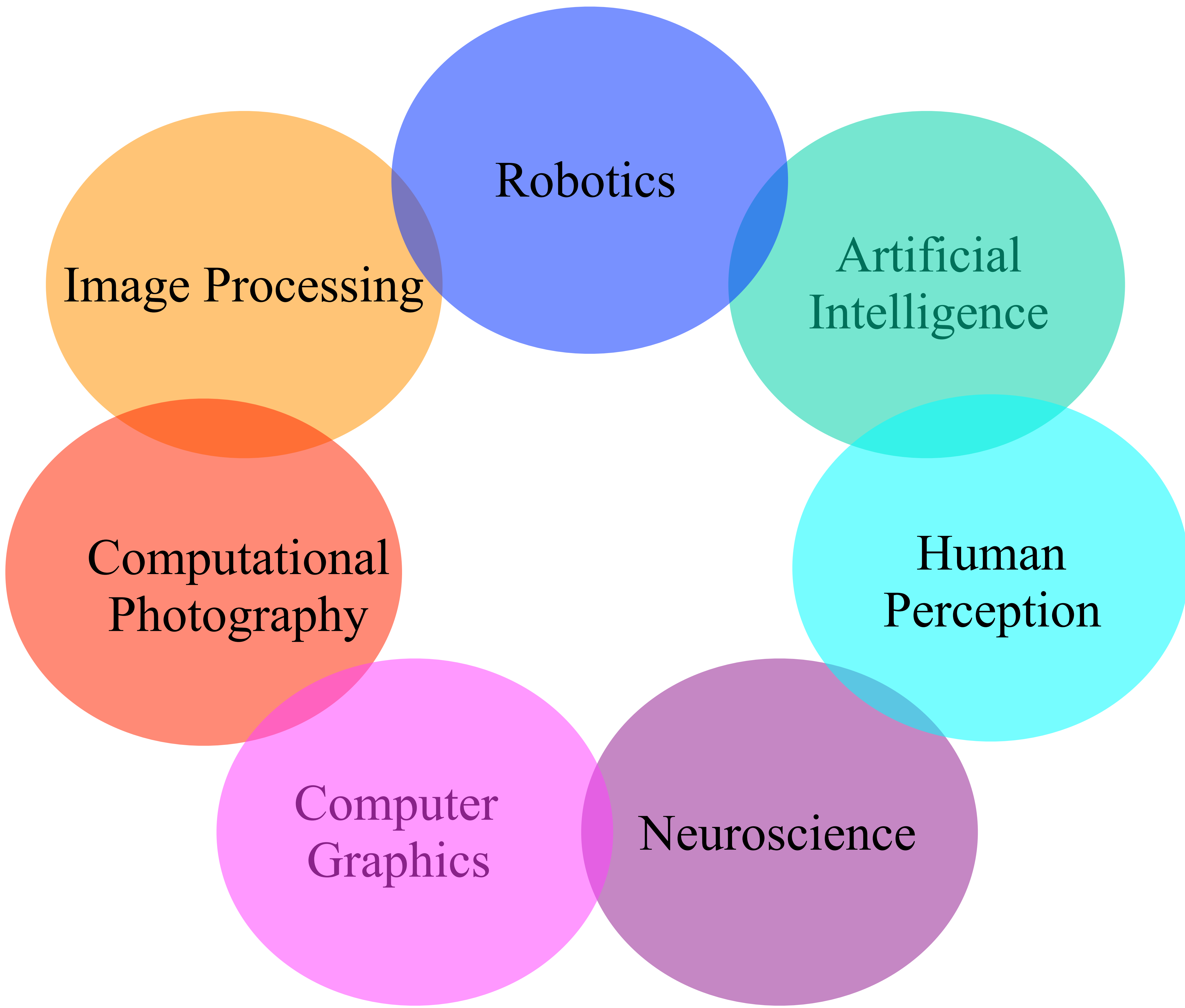
Cast shadows



Cast shadows



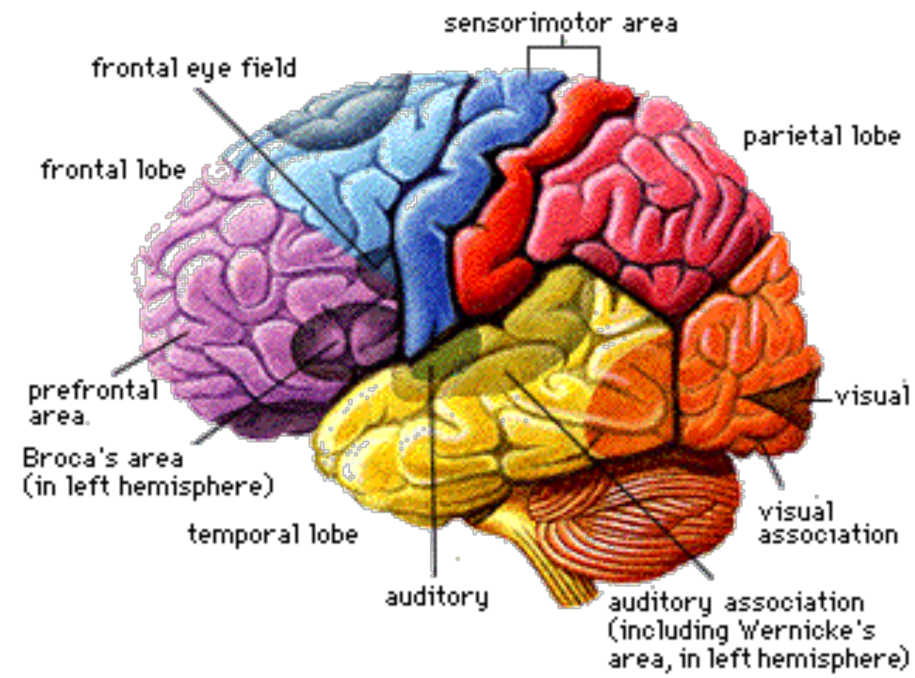
Other fields



Why is this interesting?

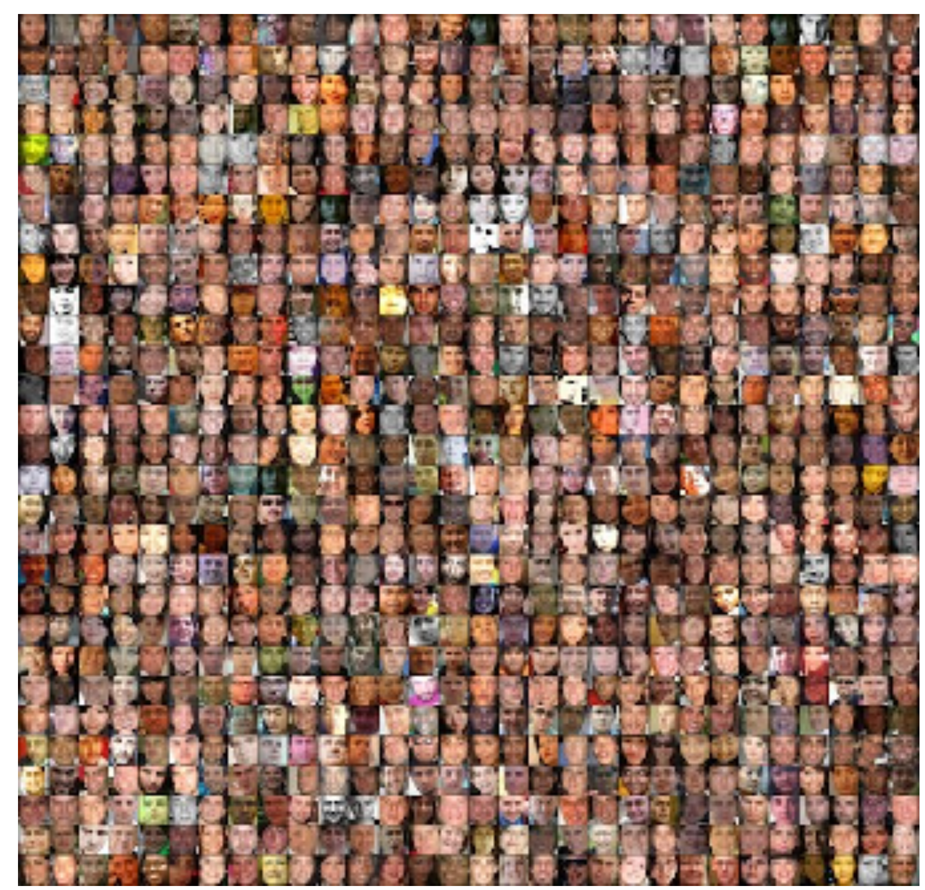
Science:

How does the human perceptual system work?

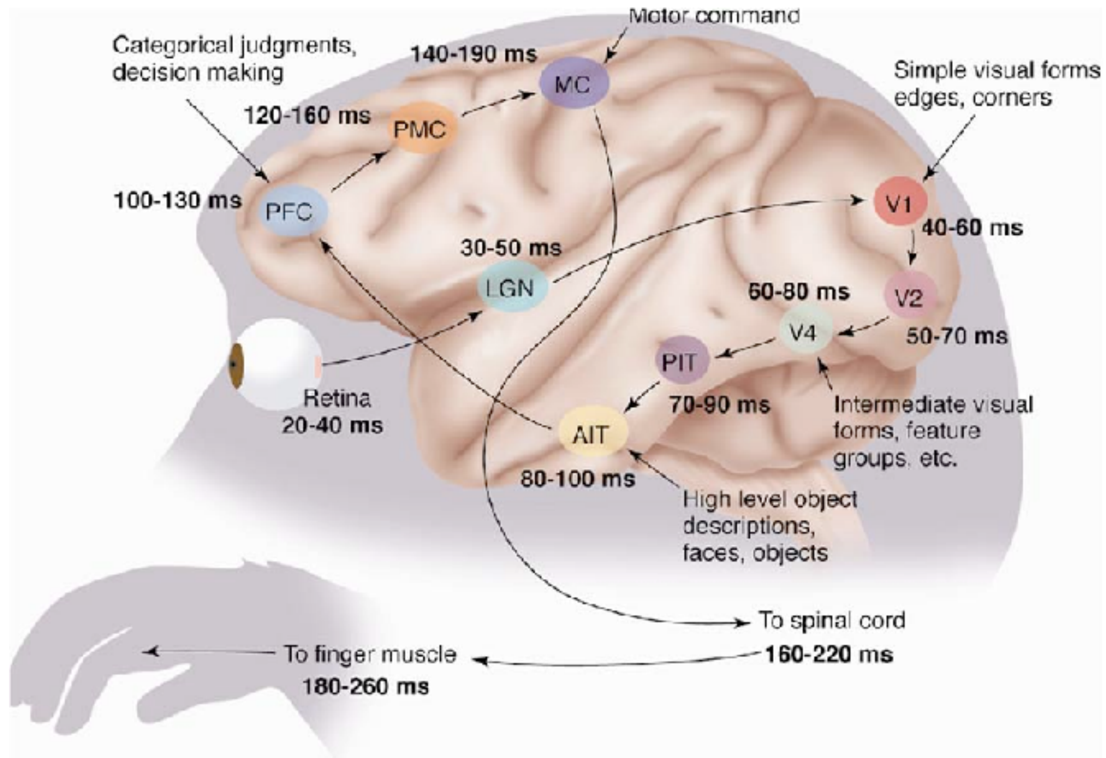


Engineering:

How do we process the abundance of visual media data around us?



Why not build a computer like a brain?



Visual cortex: 10^{11} neurons
parallel

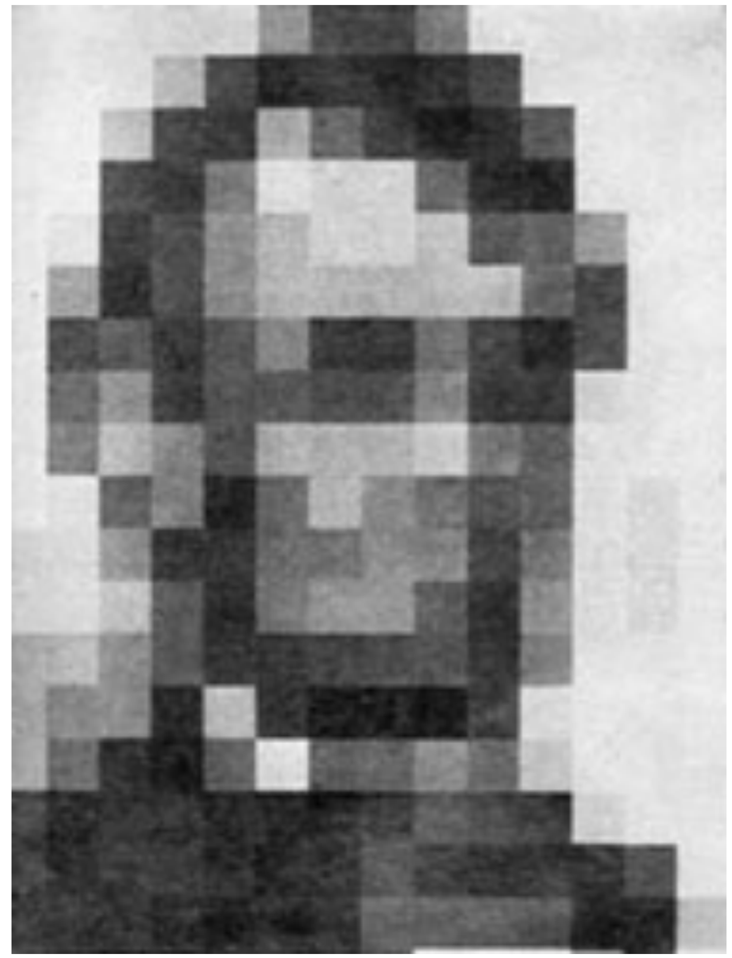
Computer: 10^{8} transistor
serial (but changing)

We still don't know the right software

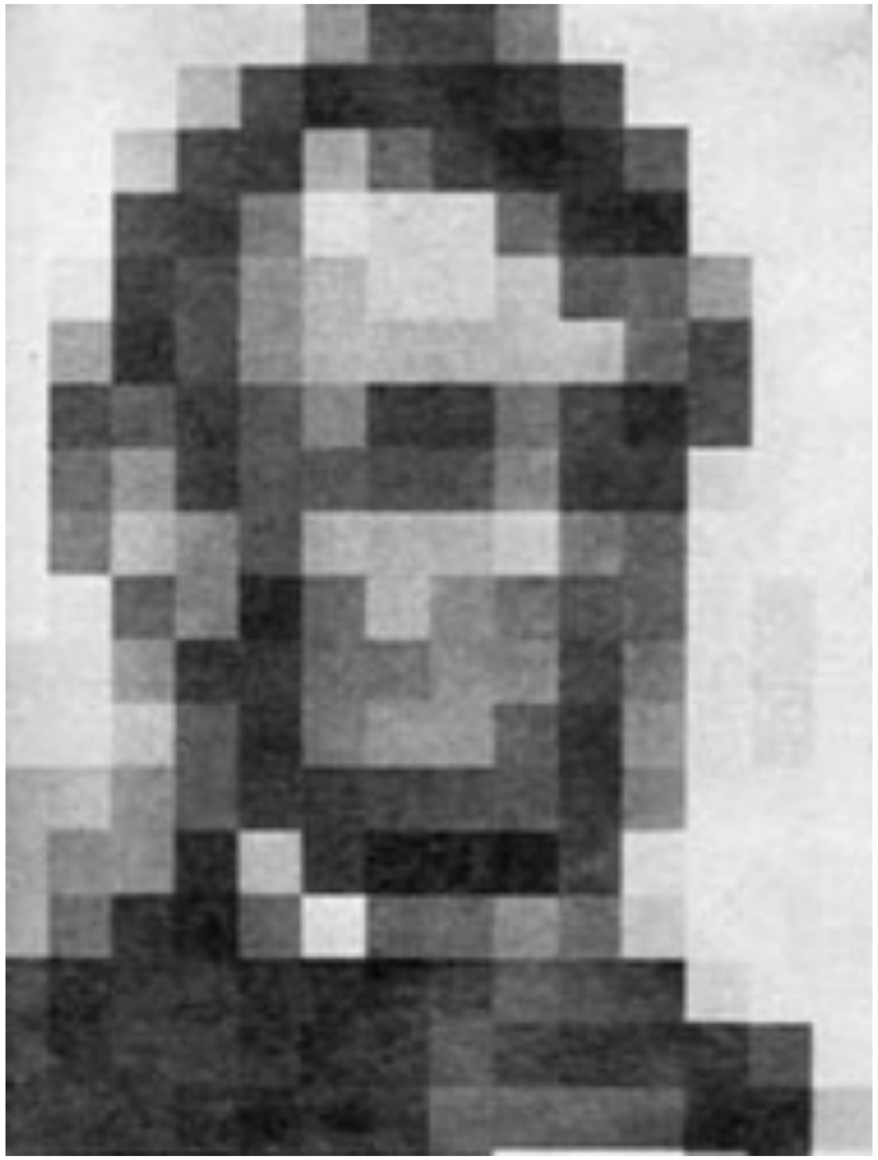
Many approaches (e.g., deep learning) are inspired by biology

Let’s discuss a simple computational approach to vision…

What is this a picture of?



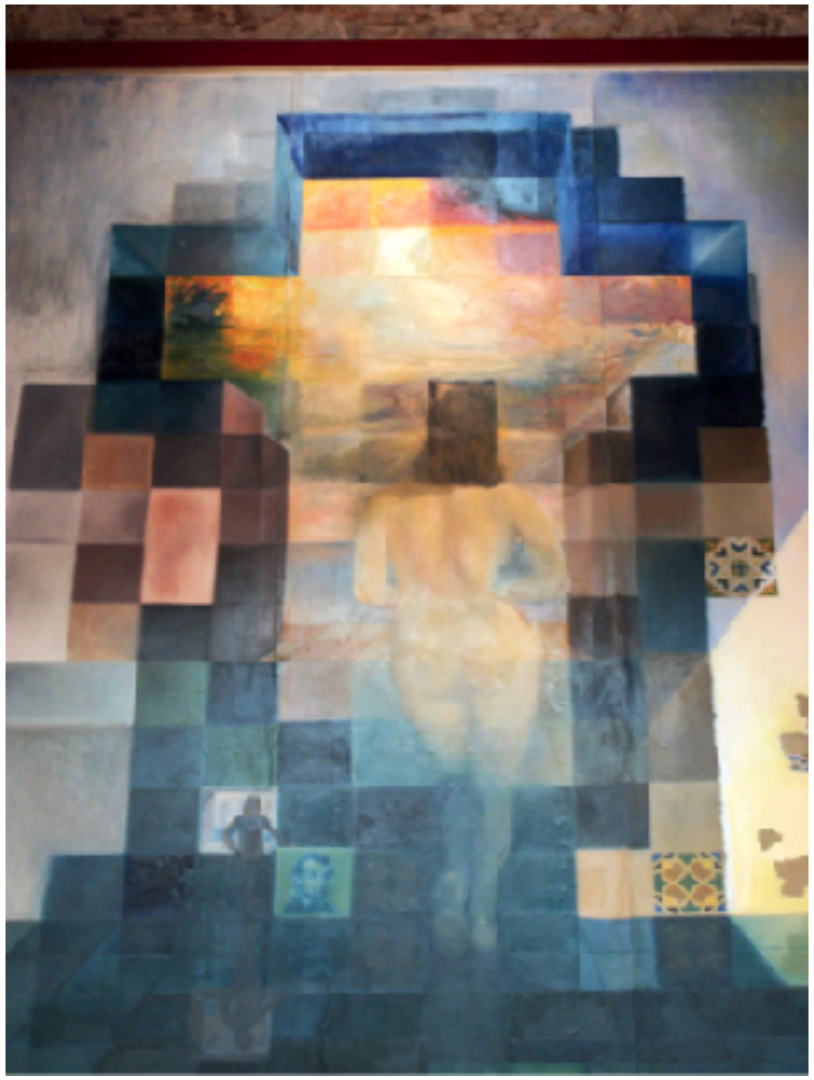
Digital images



234	7	89	7	98	98	7	9	7	5
43	7	0	123	4	13	454	23	5	87
67	5	76	4	3	56	67	87	65	45
97	0	6	3	6	25	7	3	587	8
78	5	54	7	876	71	54	76	9	75
45	81	67	78	78	5	4	75	86	8
5	4	3	35	8	256	6	4	3	36
7	6	64	3	4	7	77	76	4	54
64	35	46	46	64	56	7	56	4	7
75	464	576	75	75	75	57	64	75	75

A digital image = table of numbers

Digital Color images



S. Dali “Gala Contemplating the Mediterranean Sea”

234	7	89	7	98	98	7	9	7	5
									87
									45
									8
									75
									8
									36
									54
									7
									75

234	7	89	7	98	98	7	9	7	5
									87
									45
									8
									75
									8
									36
									54
									7
									75

234	7	89	7	98	98	7	9	7	5
43	7	0	123	12	13	454	23	5	87
67	5	76	4	3	56	67	87	65	45
97	0	6	3	6	25	7	3	587	8
78	5	54	7	876	19	54	76	9	75
45	81	67	78	78	5	4	75	86	8
5	4	3	35	8	256	6	4	3	36
7	6	64	3	4	7	77	76	4	54
64	35	46	46	64	56	7	56	4	7
75	464	576	75	75	75	57	64	75	75

3 tables of numbers (Red, Green Blue)

Example problem: face recognition

Query

Obama

G.W.Bush

Clinton

G.H. Bush

…

Library of known faces

Example problem: face recognition

Query

Obama

G.W.Bush

Clinton

G.H. Bush

…

Library of known faces

23	7	89	7	98	98	7	9	7	5
43	7	0	12	12	13	45	23	5	87
67	5	76	4	3	56	67	87	65	45
97	0	6	3	6	25	7	3	58	8
78	5	54	7	87	19	54	76	9	75
45	81	67	78	78	5	4	75	86	8
5	4	3	35	8	25	6	4	3	36
7	6	64	3	4	7	77	76	4	54
64	35	46	46	64	56	7	56	4	7
75	46	57	75	75	75	57	64	75	75

23	7	89	7	98	98	7	9	7	5
43	7	0	12	12	13	45	23	5	87
67	5	76	4	3	56	67	87	65	45
97	0	6	3	6	25	7	3	58	8
78	5	54	7	87	19	54	76	9	75
45	81	67	78	78	5	4	75	86	8
5	4	3	35	8	25	6	4	3	36
7	6	64	3	4	7	77	76	4	54
64	35	46	46	64	56	7	56	4	7
75	46	57	75	75	75	57	64	75	75

23	7	89	7	98	98	7	9	7	5
43	7	0	12	12	13	45	23	5	87
67	5	76	4	3	56	67	87	65	45
97	0	6	3	6	25	7	3	58	8
78	5	54	7	87	19	54	76	9	75
45	81	67	78	78	5	4	75	86	8
5	4	3	35	8	25	6	4	3	36
7	6	64	3	4	7	77	76	4	54
64	35	46	46	64	56	7	56	4	7
75	46	57	75	75	75	57	64	75	75

23	7	89	7	98	98	7	9	7	5
43	7	0	12	12	13	45	23	5	87
67	5	76	4	3	56	67	87	65	45
97	0	6	3	6	25	7	3	58	8
78	5	54	7	87	19	54	76	9	75
45	81	67	78	78	5	4	75	86	8
5	4	3	35	8	25	6	4	3	36
7	6	64	3	4	7	77	76	4	54
64	35	46	46	64	56	7	56	4	7
75	46	57	75	75	75	57	64	75	75

23	7	89	7	98	98	7	9	7	5
43	7	0	12	12	13	45	23	5	87
67	5	76	4	3	56	67	87	65	45
97	0	6	3	6	25	7	3	58	8
78	5	54	7	87	19	54	76	9	75
45	81	67	78	78	5	4	75	86	8
5	4	3	35	8	25	6	4	3	36
7	6	64	3	4	7	77	76	4	54
64	35	46	46	64	56	7	56	4	7
75	46	57	75	75	75	57	64	75	75

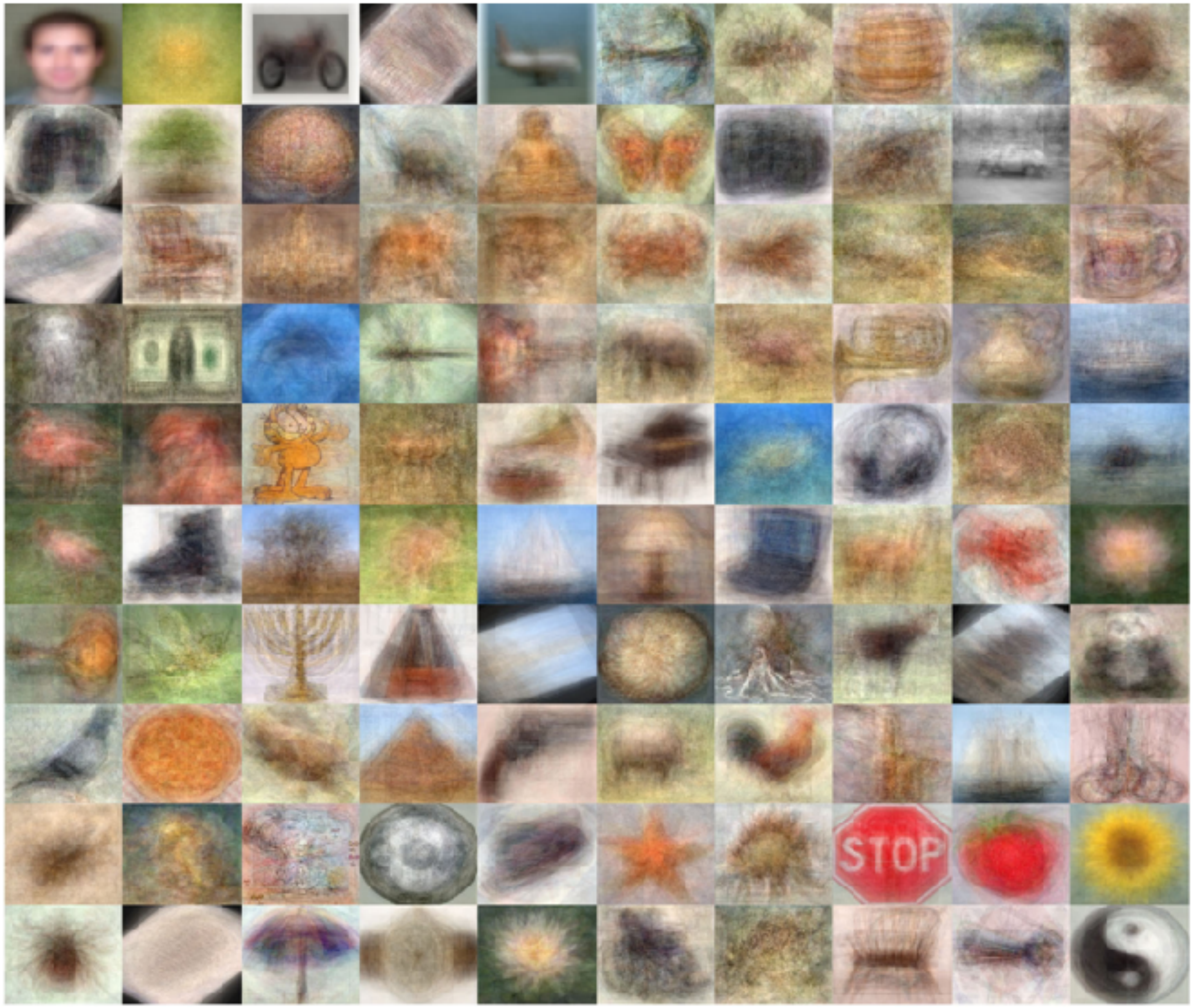
Can you think of an algorithm for matching?

“Nearest-Neighbor” Face Recognition

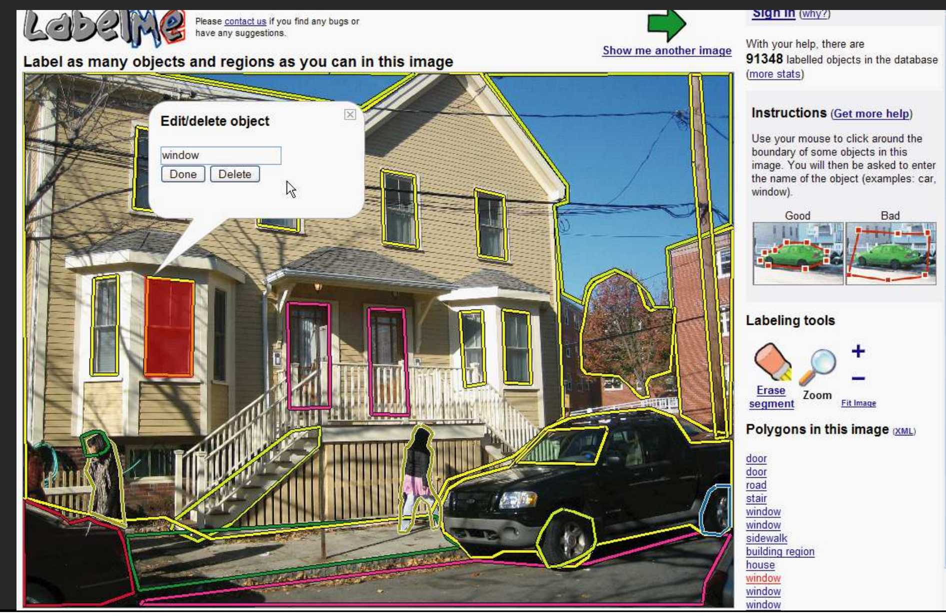
...

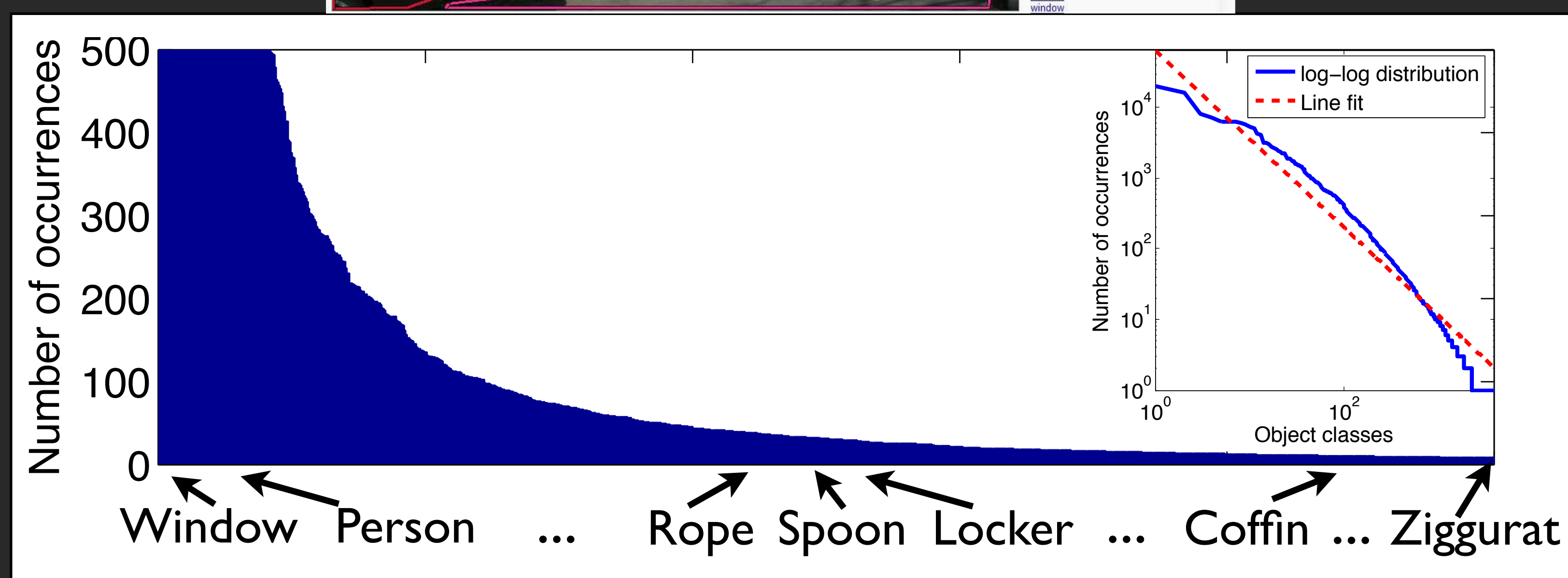
Variant of this implemented in Picasa, iPhoto, FaceBook

Google approach to intelligence



Why doesn't big-data solve vision?



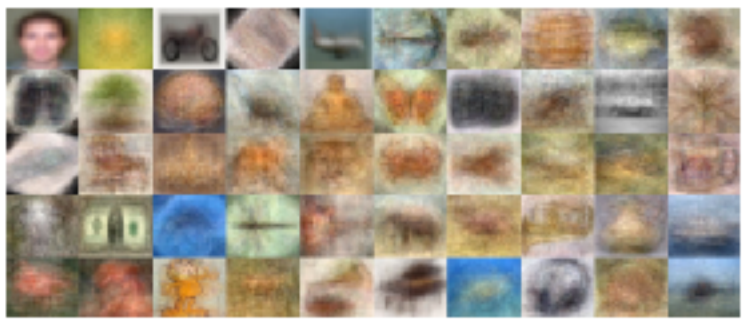


$$y = x^a$$
$$\log y = a \log x$$

Long-tail statistics: we still care about small-data problems in big-data world

Interlude: Logistics

http://16720.courses.cs.cmu.edu/



16-720
Computer Vision
Spring 2016

Administrivia

Instructor:	Deva Ramanan (deva@cs.cmu.edu)			
TAs:	Achal Dave (achald@cmu.edu>)			
	Shashank Jujjavarapu (bjujjava@andrew.cmu.edu)			
	Siddarth Malreddy (smalredd@andrew.cmu.edu)			
	Brian Pugh (bpugh@andrew.cmu.edu)			
Piazza link:	piazza.com/cmu/spring2017/16720/home			
Gradescope link:	https://gradescope.com/login (use entry code MNNZ8M)			
Lectures:	Tues,Thur	12:00-1:20pm	Doherty Hall	1212
TA office hours:	Mon-Wed	5:00-6:30pm	Smith Hall	200
	Thurs	3:00-4:30pm	Smith Hall	200
Deva's office hours:	Tues	1:30-3:00pm	Smith Hall	221

Waitlist

We are at the capacity of this room (can’t add much more).

If you are considering dropping, please do so as a courtesy to your fellow classmates

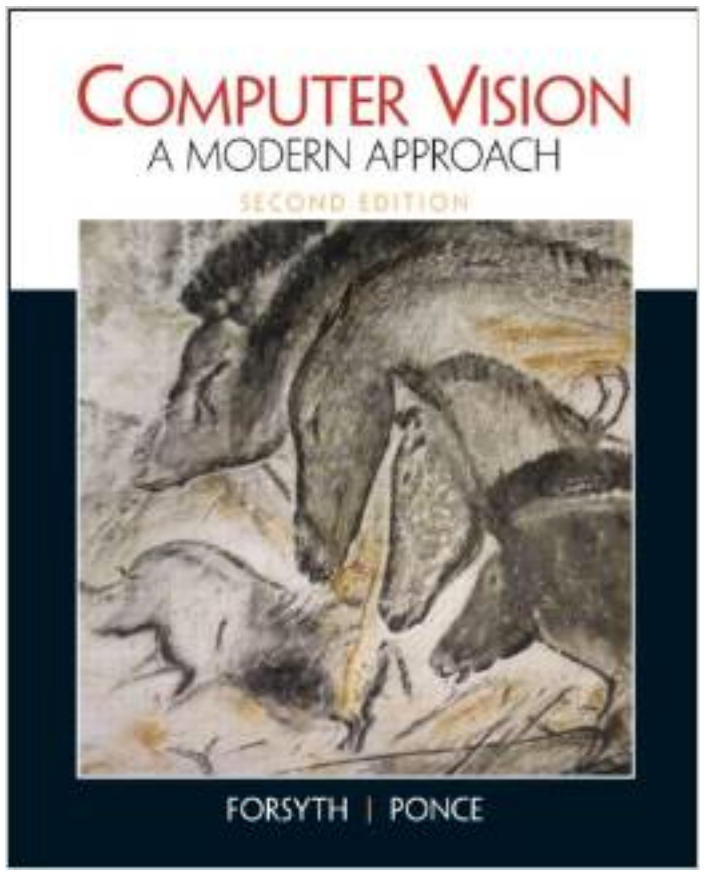
Because class is oversubscribed by nearly 200 people, even handling special cases is overwhelming

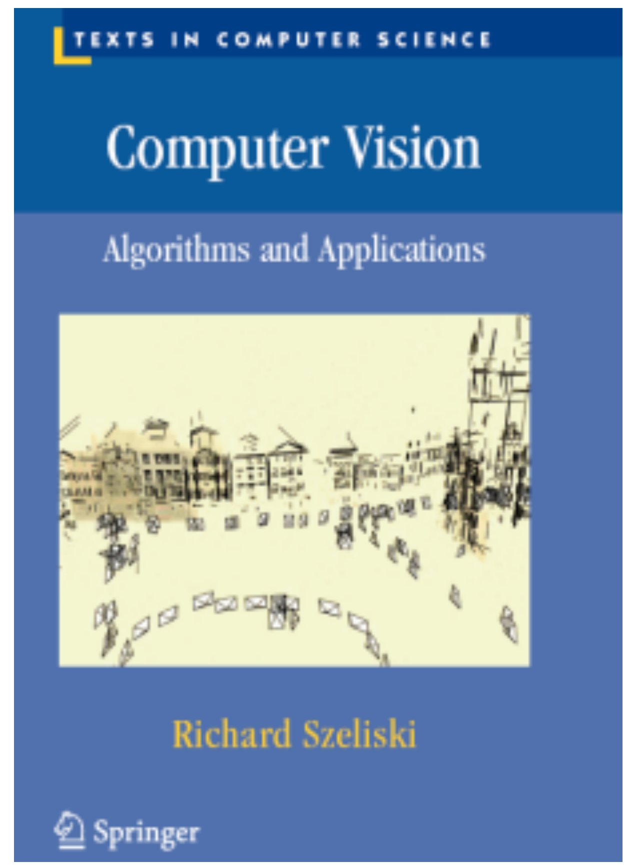
Default order for pulling off waitlist (prioritized by seniority and departments where class is required)

1. Phd Robotics
2. Masters Robotics
3. Phd SCS
4. Masters SCS
5. Phd Engineering
6. Masters Engineering
7. PhD - remaining
8. Masters - remaining
9. Undergrads - remaining

You can try e-mail me any special circumstances, but if I don’t respond, assume that I’ll use the default above

Textbooks (recommended)

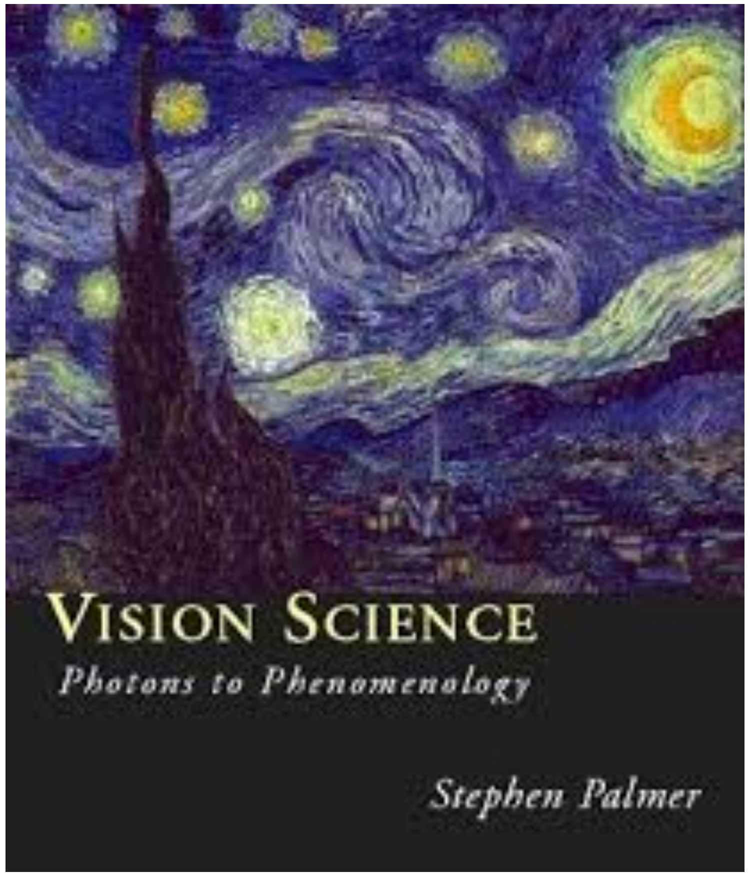




http://szeliski.org/Book/

Other texts

(classic and nice presentation)



Gives the “human” side of things

Pre-reqs

Knowledge of linear algebra, vector calculus, and basic probability are required. MATLAB programming experience and previous exposure to image processing are desirable, but not required.

Grading

5 homeworks (with considerable MATLAB implementation) worth 17% each, a class project worth 12%, and class participation (measured by instructor-approved answers to Piazza questions) worth 3%.

Accounts used by class

Blackboard: submission of code

Gradescope: submission of PDF writeup and returning of grades

Piazza: asking and answering questions about course and homeworks

HW submission

Homework Code Submission

Submit a zipfile named YourAndrewId.zip to Blackboard. This zip file should contain a single folder called YourAndrewId that contains a matlab/ folder containing all the .m and .mat files you were asked to write and generate. Make sure the matlab directory includes everything that your code needs to run. If we hand out code as part of the assignment, please include that also. However, to prevent large zip files, do not include image or data files that we prodivid in the data directory of skeleton code.

Homework Writeup Submission

You must submit your writeups as as a single PDF file titled YourAndrewId.pdf to Gradescope. We will not accept writeups in any word processing format such as MS Word or Open Office. You are welcome to create the writeup in whichever word processor you like but in the end you need to save it as a pdf file.

Late policy

Assignments and due dates can be found on the lecture webpage. Homeworks must be submitted on both Blackboard (code) and Gradescope (writeup) by 11:59 pm on the given due date. You will be allowed a total of 3 late days througout the semester - use them wisely! Each additional day late will result in an penalty of 50% of the homework grade, and no homework will be accepted after 3 days past its due date.

Piazza discussions

Questions are encouraged as a way to get help from other students when you are stuck, or feel like you're going down the wrong path. **Questions are not meant as a way to solve a problem before you've struggled with it**. Most learning comes from a little bit of struggling, and the assigments are meant to make you think a little about how to implement things. Before posting a question, ask yourself whether you're truly stuck or are just avoiding spending the time to figure it out. Struggling and debugging is a big part of learning in this class!

Piazza discussions

Questions are encouraged as a way to get help from other students when you are stuck, or feel like you're going down the wrong path. **Questions are not meant as a way to solve a problem before you've struggled with it**. Most learning comes from a little bit of struggling, and the assigments are meant to make you think a little about how to implement things. Before posting a question, ask yourself whether you're truly stuck or are just avoiding spending the time to figure it out. Struggling and debugging is a big part of learning in this class!

Questions about homework wording, lecture slides, or class logistics are welcome and encouraged; however, please keep complaints about the course off of Piazza. If you have a concern or a suggestion about anything to do with the course, the best way to deal with it and get results is to take it to the TA or Instructor directly . **Piazza is not a complaint forum**.

Piazza discussions

Questions are encouraged as a way to get help from other students when you are stuck, or feel like you're going down the wrong path. **Questions are not meant as a way to solve a problem before you've struggled with it**. Most learning comes from a little bit of struggling, and the assigments are meant to make you think a little about how to implement things. Before posting a question, ask yourself whether you're truly stuck or are just avoiding spending the time to figure it out. Struggling and debugging is a big part of learning in this class!

Questions about homework wording, lecture slides, or class logistics are welcome and encouraged; however, please keep complaints about the course off of Piazza. If you have a concern or a suggestion about anything to do with the course, the best way to deal with it and get results is to take it to the TA or Instructor directly . **Piazza is not a complaint forum**.

We encourage you to be an active participant and help your fellow students out by answering questions on Piazza. Avoid becoming overzealous and giving the answer though. Full lines of code should never be posted on Piazza, and **hints rather than solutions should be given**. For example:

Question: "I tried subtracting the vectur mu from the matrix A and MATLAB keeps breaking with the error message "...". I've tried several things and can't figure out how to make MATLAB subtract mu from each column."

Poor Answer: "bsxfun(@minus, A, mu)"

Good Answer: "Look into MATLAB's documentation/help on "bsxfun". MATLAB's minus function only works on matrices that are the same dimension; you actually want to perform several subtraction operations on the matrix A."

Instructor-approved answers will contribute toward class participation grade!

Academic honesty

Homeworks can be discussed, but each student must independently write up their own solutions. In particular, no sharing of code. Please see the university policy on academic honesty. It is fine to use reference materials found online, but do not search for homework solutions. Rather, students are strongly encouraged to ask questions at both office hours and on the class discussion group.

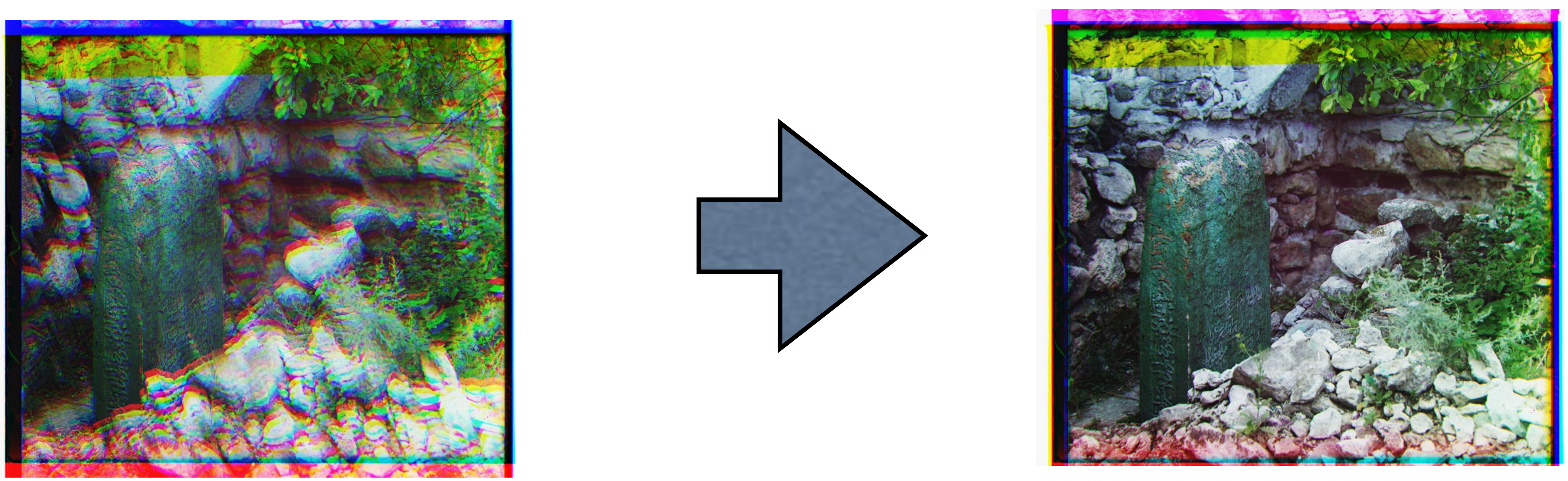
Syllabus

Date	Topic	Reference	HW Due
1/12	Introduction	Ballard & Brown Chap 1	
1/14	Matlab		
Image Processing			
1/19	Convolutions		HW0: Intro (pdf, code)
1/21	Filtering + Bag-of-words		
1/26	Fourier		
1/28	Interest Points		
Image Formation			
2/2	Descriptors		HW1 (Matching)
2/4	Camera Projection + Homographies		
2/9	Optical flow		
2/11	Image Alignment		
Geometry			
2/16	Motion Segmentation		
2/18	2-View Geometry		HW2 (Mosaics)
2/23	SFM		
2/25	Structured light		
Recognition			
3/1	Machine learning		
3/3	Matching and Retrieval		HW3 (Tracking)
3/15	Template Matching		
3/17	Parts		Project Proposal
3/22	Deep learning 1		
3/24	Deep learning 2		HW4 (Reconstruction)
Grouping			
3/29	Binary Morphology		
3/31	Textures		
4/5	Graphcuts		
4/7	Catchup		
Physics			
4/12	Light Transport		HW5 (Recognition)
4/14	NO CLASS (spring carnival)		
4/19	Color		
4/21	Computational Imaging		
4/26	Project presentations		Proj. Presentation + Writeup
4/28	Project presentations		

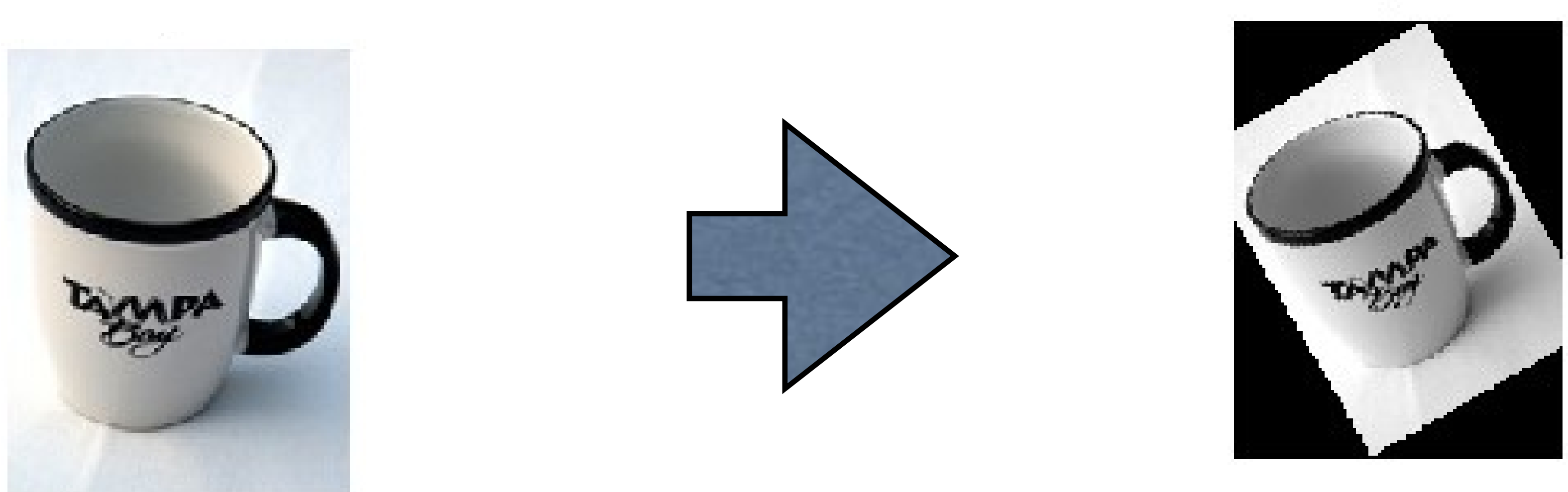
HW0

Required to submit, but not graded (those enrolled in class must submit something to keep their spot)

Warning: If you find that HW0 is difficult, class will be quite challenging



automatic alignment
of R,G,B planes



affine warp of an image

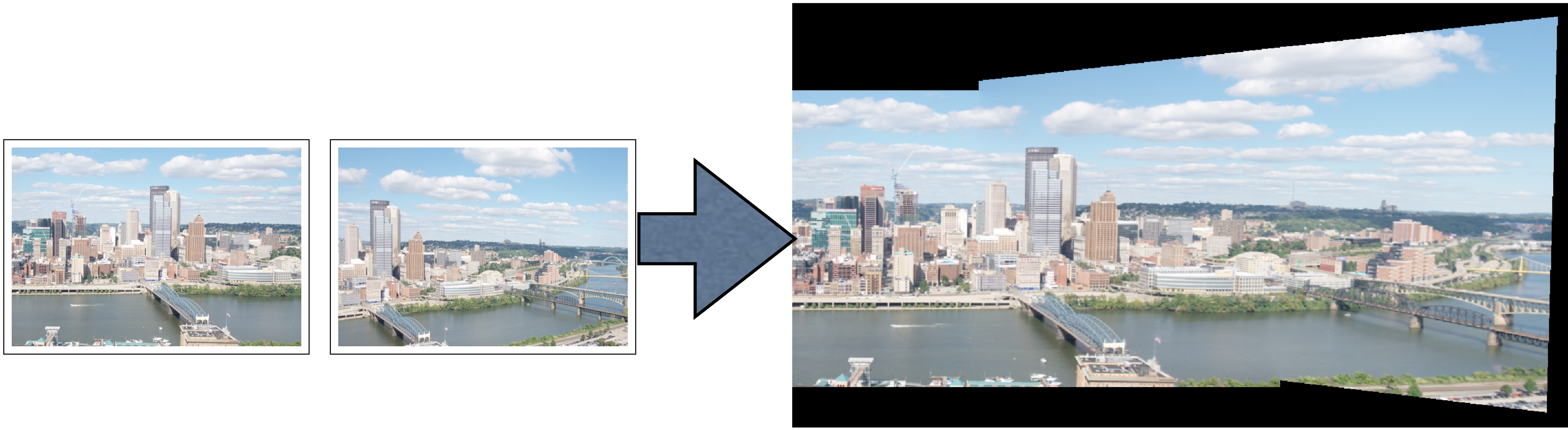
HW1

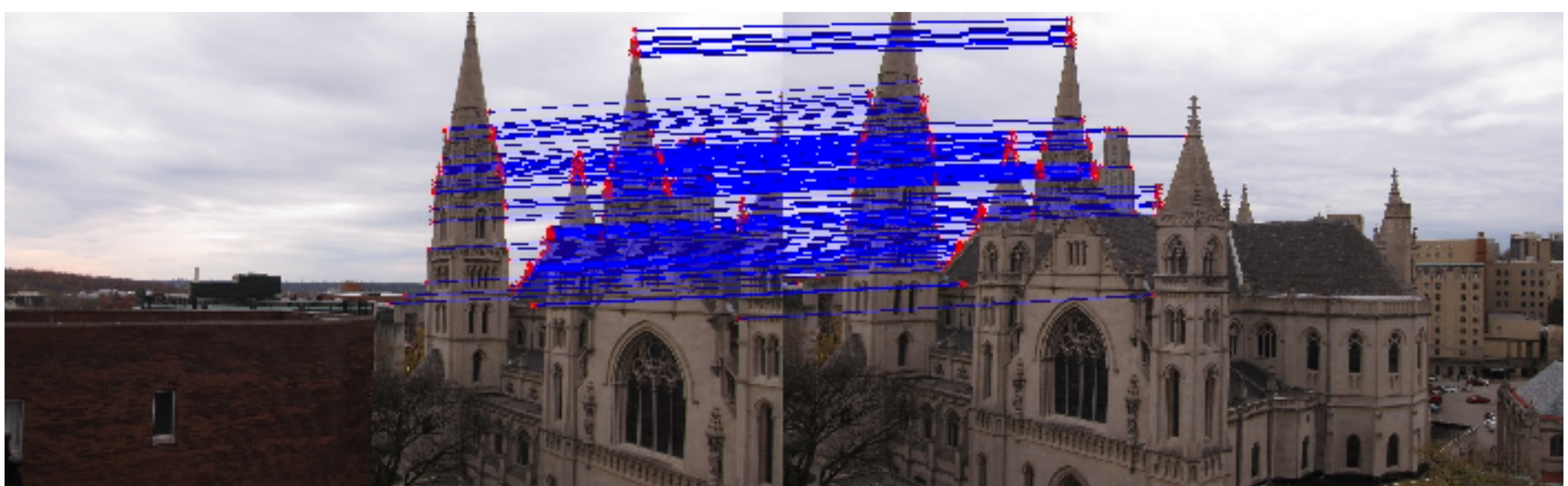
Scene recognition using bag-of-words models

Concepts: spatial correspondence, filter banks, clustering

HW2

Panoramic Mosaics





Concepts: interest points, descriptors, RANSAC, homographies

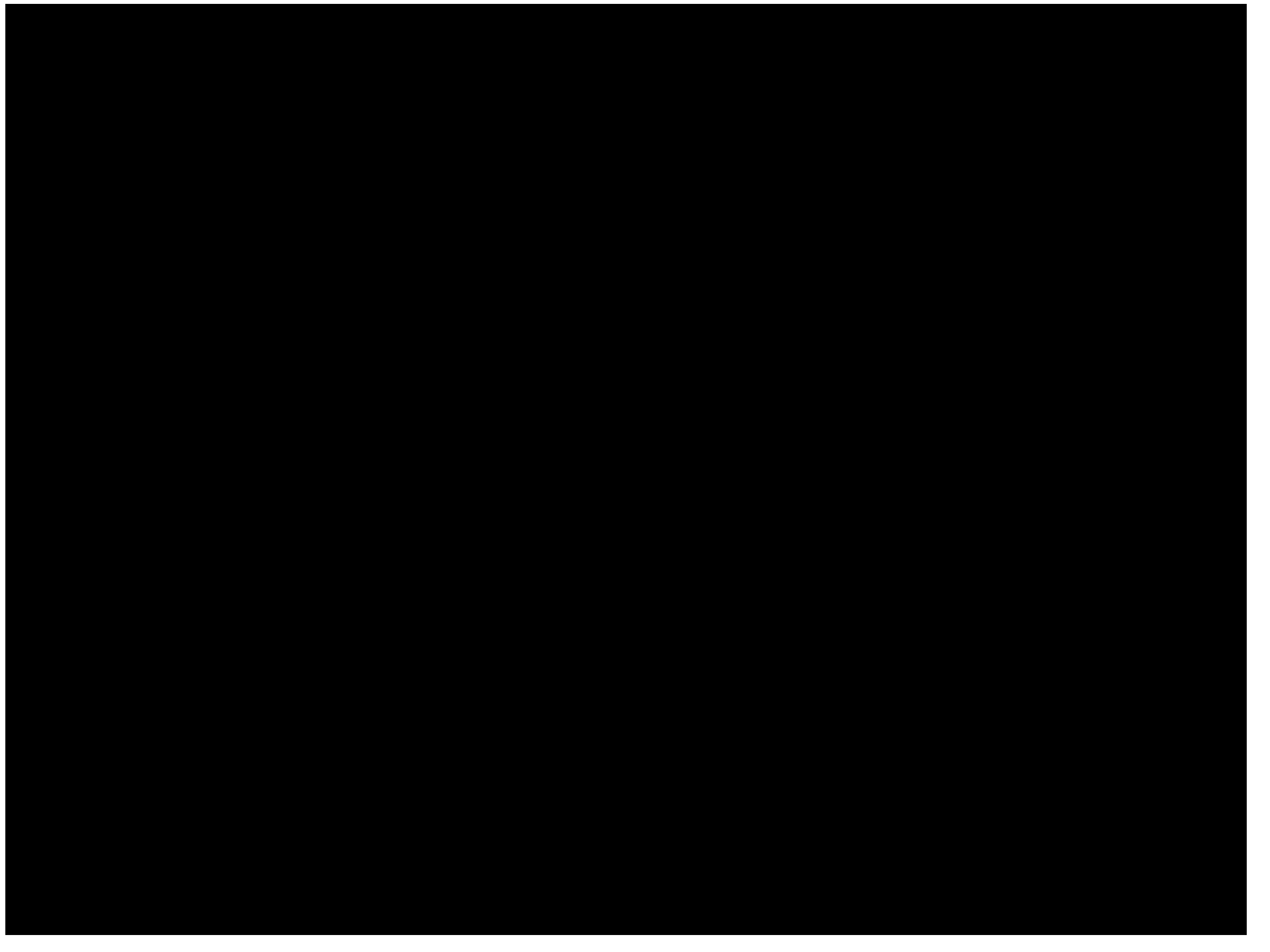
HW3

Tracking

Concepts: Templates, Lucas-Kanade alignment

HW4

3D Reconstruction



Concepts: Epipolar geometry, triangulation

HW5

Deep learning

Concepts: Pattern classification

Projects

A typical class project might be implementing and evaluating an algorithm from a research paper. The choice of projects can be very open-ended; ideally you can incorporate their own research. But your class project must be about new things you have done this semester; you can't use results you have developed in previous semesters. People can work in groups of size 1 to 3, though we encourage students to work in groups because the projects tend to be more substantial.

Previous course projects have resulted in papers - you are encouraged to explore wacky ideas!

Suggestions for this class

Its all about the assignments

A big factor in success will be your comfort and debugging capability in Matlab

(I probably spent 80 percent of my time as a grad student “debugging”)

Previous evaluations of this class

Difficulty: 7/10 (10 = impossible to get through)

“I really appreciate the use of blackboards in addition to slides”

“Lectures go a bit fast”

“This is a demanding course, but is well worth it. I really appreciate how the assignments require reading of academic papers. This is sometimes frustrating, but a skill that is very useful to develop.”

Related courses

16-822: Geometry-based Methods in Vision

The course focuses on the geometric aspects of computer vision: the geometry of image formation and its use for 3D reconstruction and calibration. The objective of the course is to introduce the formal tools and results that are necessary for developing multi-view reconstruction algorithms. The fundamental tools introduced study affine and projective geometry, which are essential to the development of image formation models. Additional algebraic tools, such as exterior algebras are also introduced at the beginning of the course. These tools are then used to develop formal models of geometric image formation for a single view (camera model), two views (fundamental matrix), and three views (trifocal tensor); 3D reconstruction from multiple images; and auto-calibration.

16-421: Vision Sensors

This course covers the fundamentals of vision cameras and other sensors - how they function, how they are built, and how to use them effectively. The course presents a journey through the fascinating five-hundered-year history of "camera-making" from the early 1500's "camera obscura" through the advent of film and lenses, to today's mirror-based and solid-state devices. The course includes a significant hands-on component where students learn how to use the sensors and understand, model and deal with the uncertainty (noise) in their measurements. While the first half of the course deals with conventional "single viewpoint" or "perspective" cameras, the second half of the course covers much more recent "multi-viewpoint" or "multi-perspective" cameras that include an array of lenses and mirrors. These sensors provide unusual and compelling forms of visualizations of the world around us that also drive new display technologies.

16-824: Visual Learning and Recognition

graduate course in Computer Vision with emphasis on representation and reasoning for large amounts of data (images, videos and ssociated tags, text, gps-locations etc) toward the ultimate goal of Image Understanding. We will be reading an eclectic mix of classic and cent papers on topics including: Theories of Perception, Mid-level Vision (Grouping, Segmentation, Poselets), Object and Scene ecognition, 3D Scene Understanding, Action Recognition, Contextual Reasoning, Image Parsing, Joint Language and Vision Models, etc. 'e will be covering a wide range of supervised, semi-supervised and unsupervised approaches for each of the topics above.

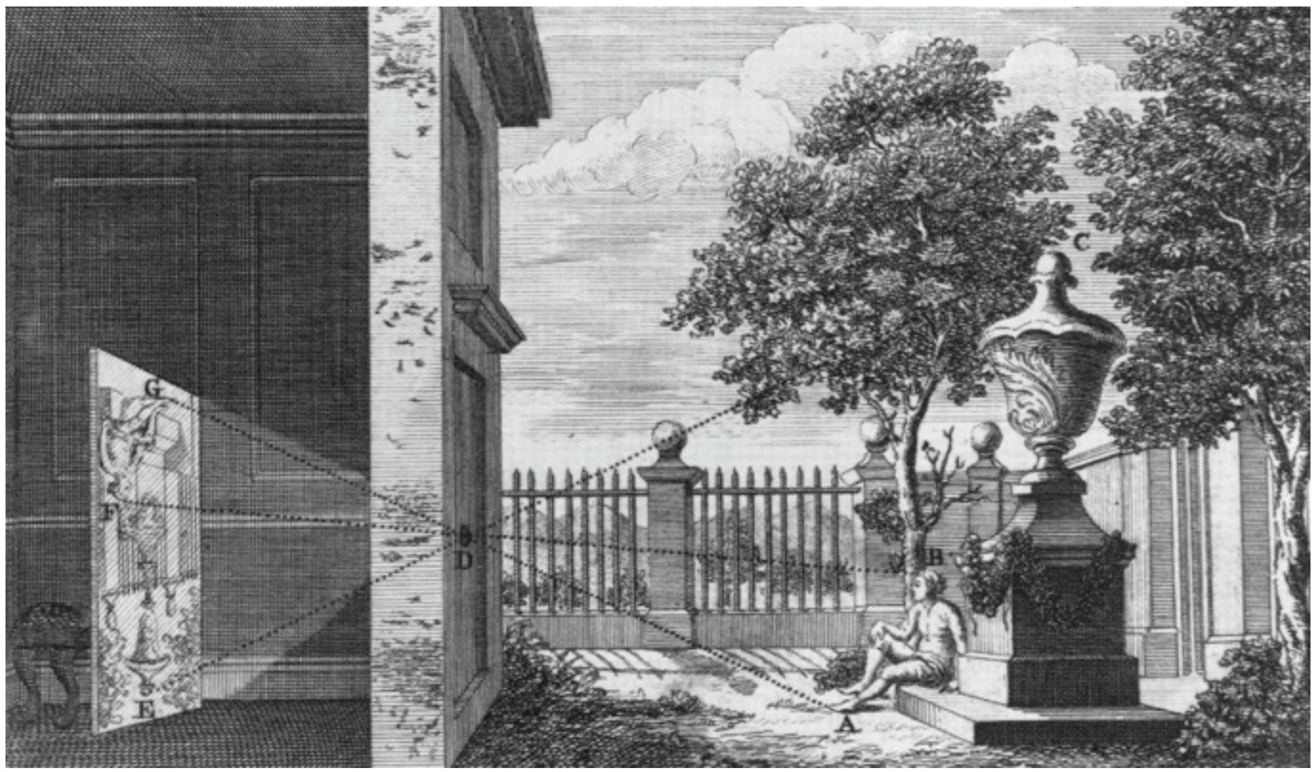
CMU Computer Vision

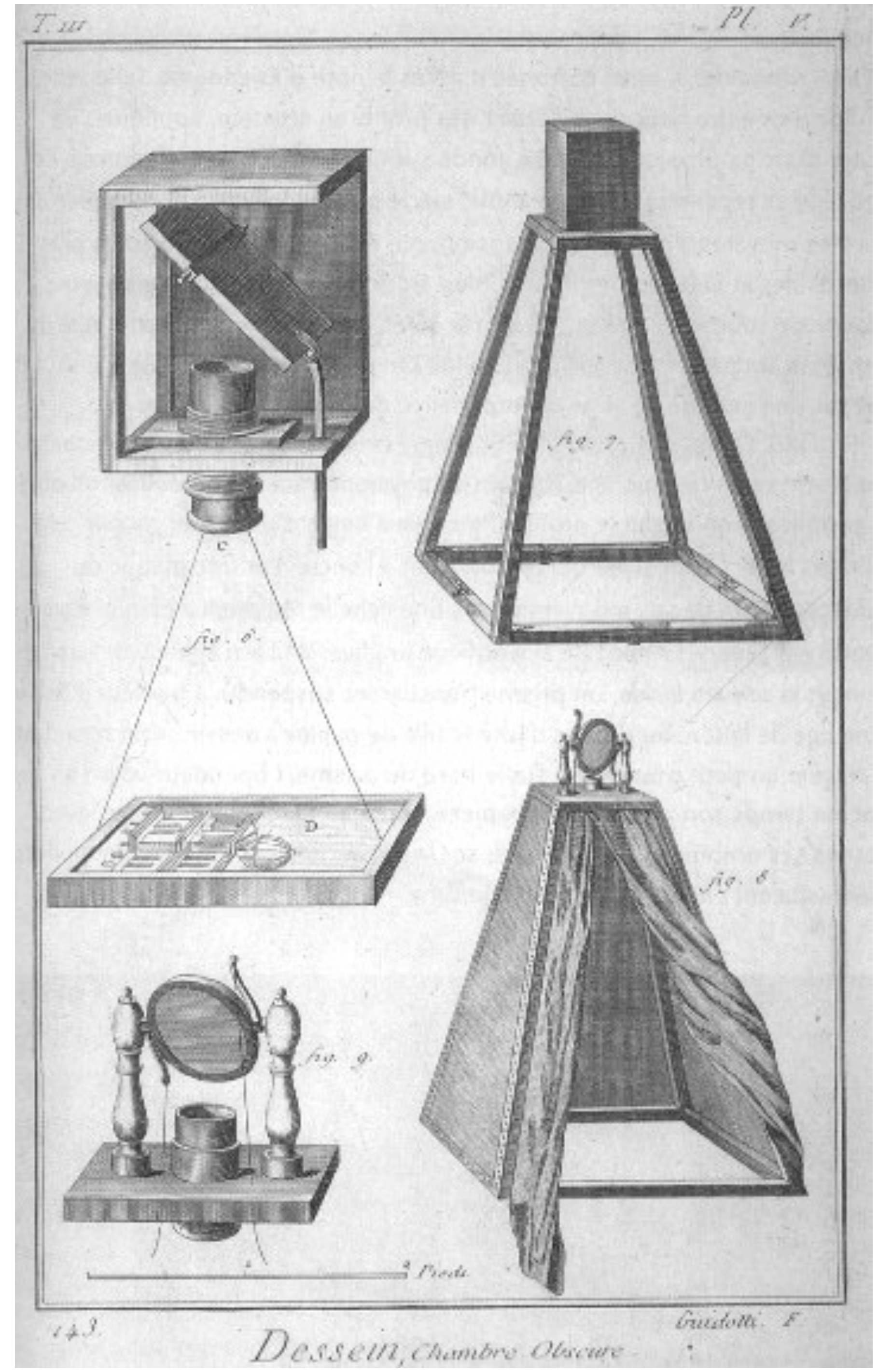
- Ten faculty covering all areas of Computer Vision research.
- State of the art laboratory facilities with strong staff researchers.
- Large class of graduate/undergraduate students.
- Over 10 courses in subareas of computer vision.
- Collaborations with Robotics, Machine Learning, AI, Systems, Computer Science, Medicine, Transportation, Arts, etc.

Outline

- What is this class about?
- Logistics for class (homeworks, grading)
- Historical perspective in computer vision
- HW0

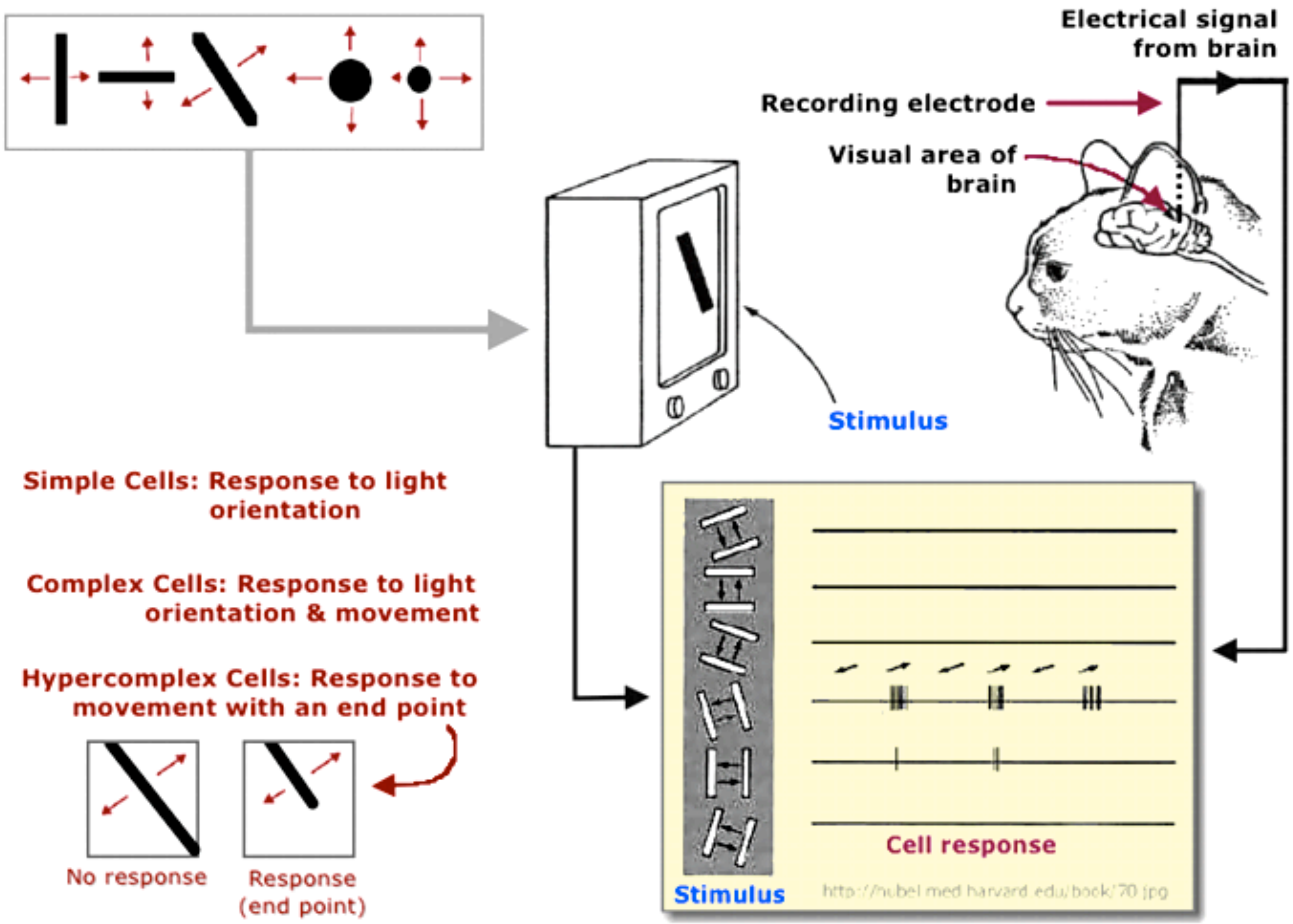
Camera Obscura





4th century BC (Aristotle)

Neuroscience



Huber & Wiesel, 1959

The start of modern computer vision

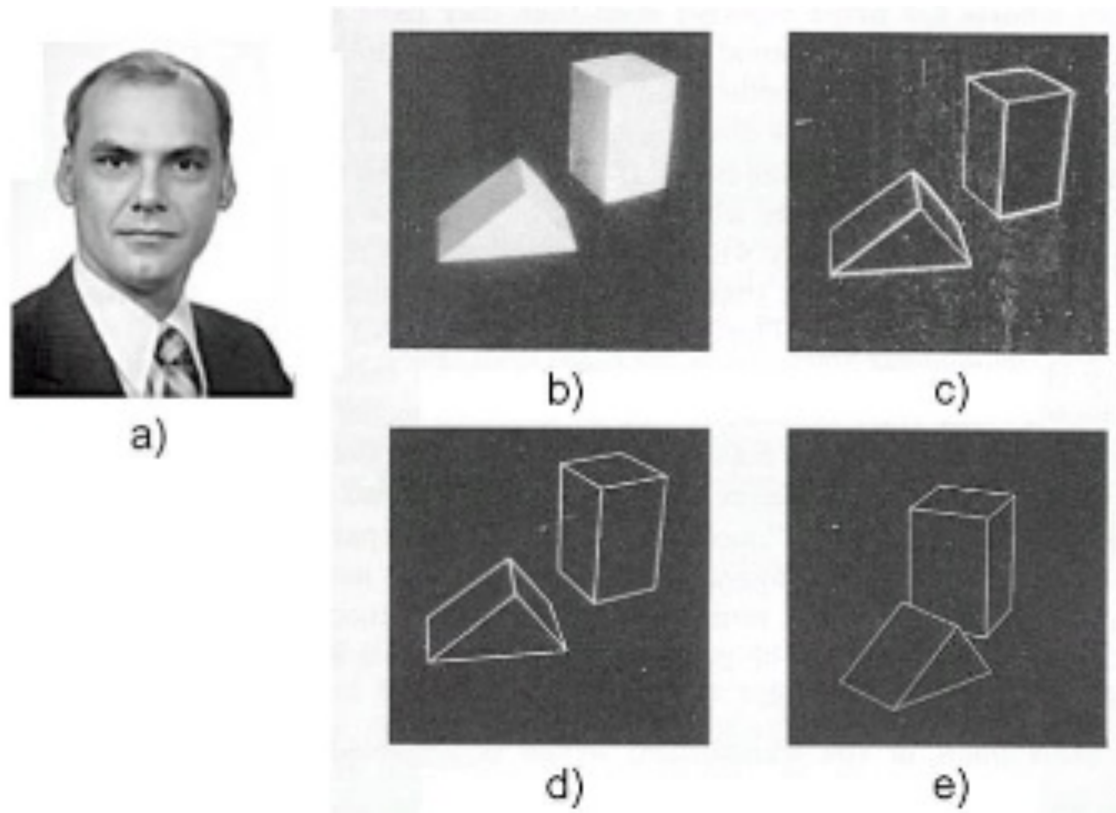


Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

Object Recognition in the Geometric Era: a Retrospective. Joseph L. Mundy. 2006

1963

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966

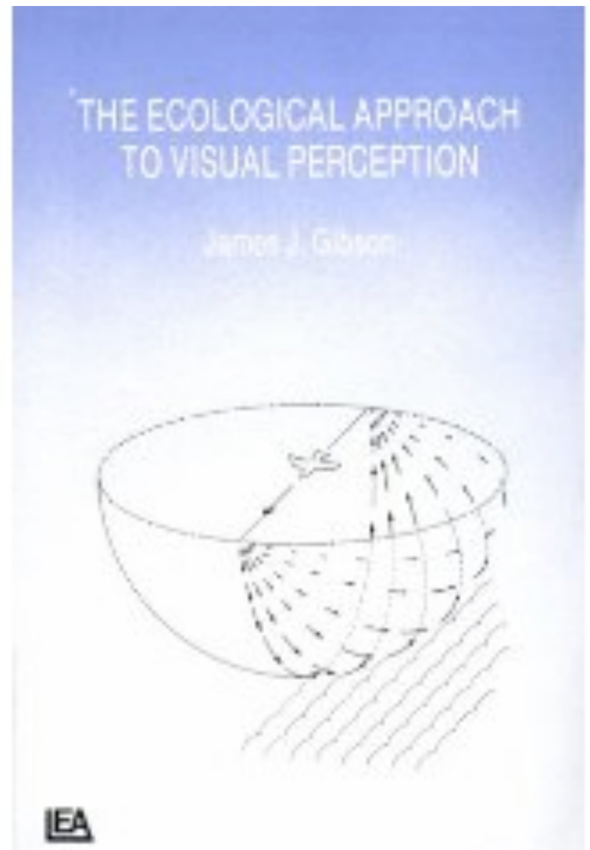
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Influential approach: ecological perspective



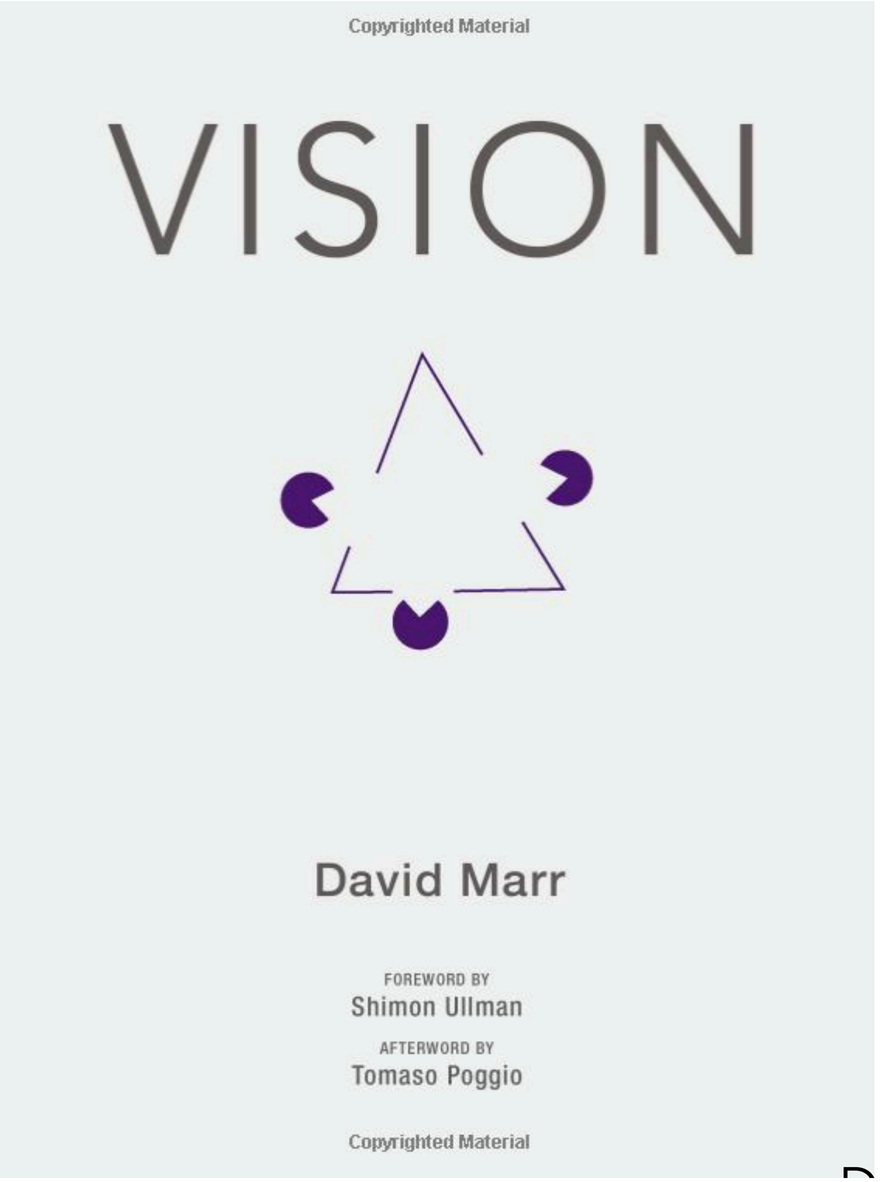
J. J. Gibson 79

Philosophical approach

Number of interesting arguments we'll look at towards the end of class

Excellent review in Palmer's Vision Science textbook

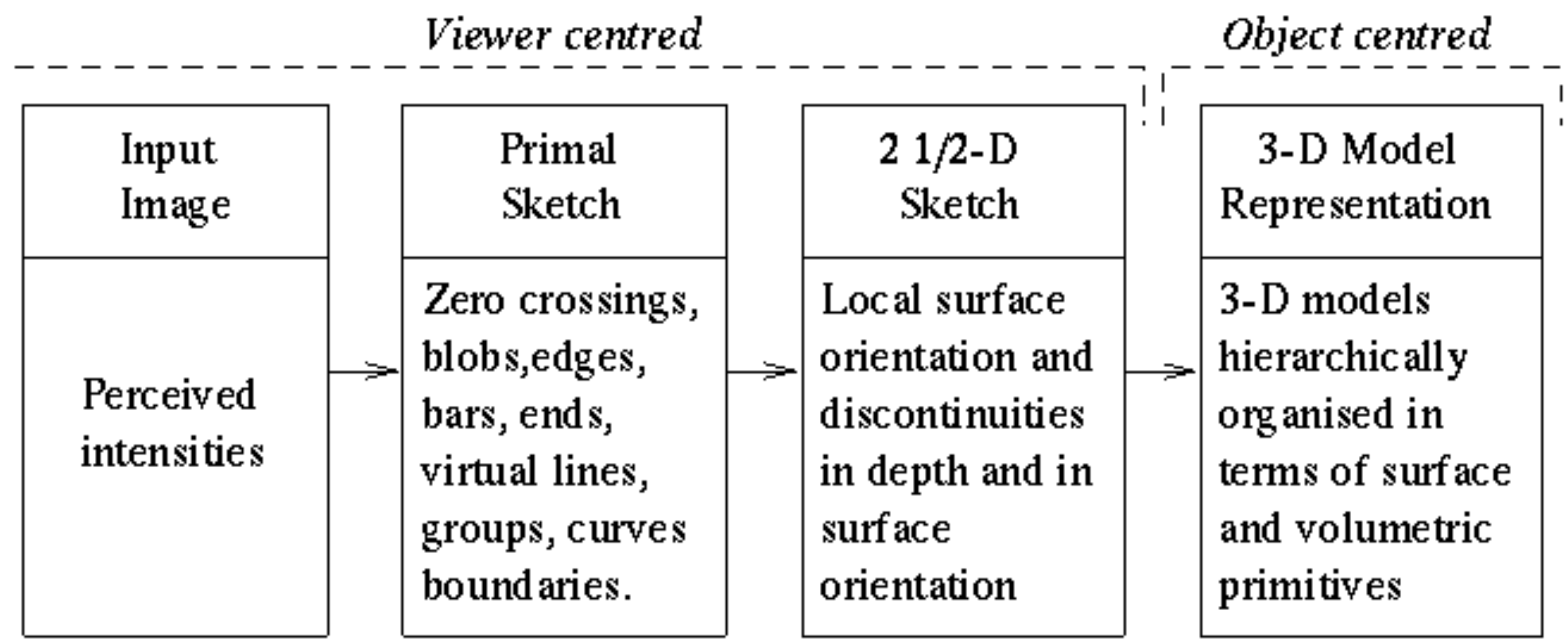
Influential approach: computational perspective

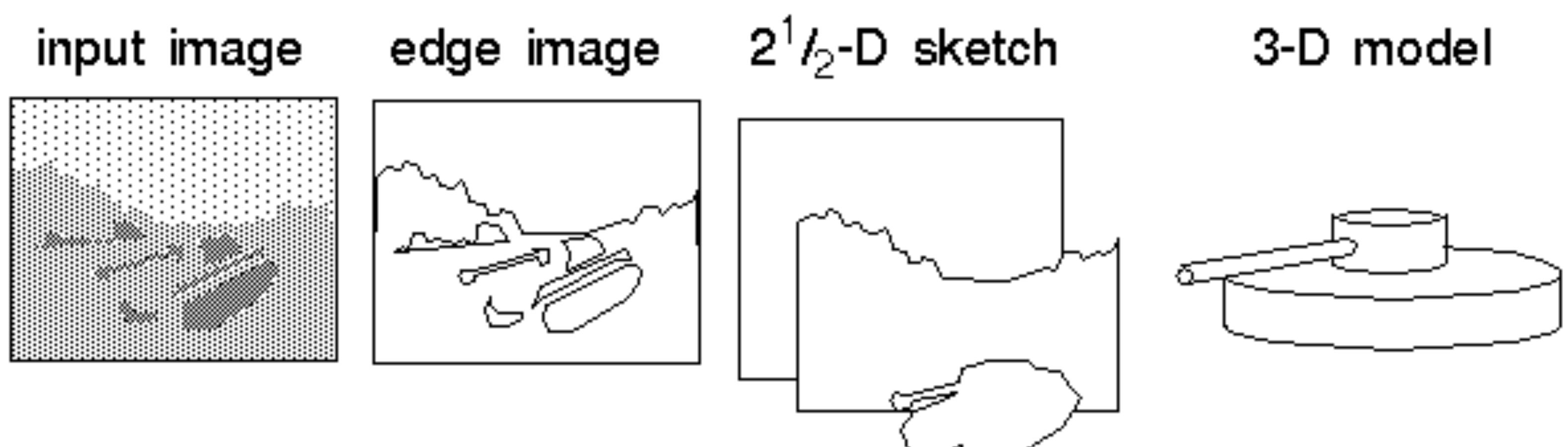


David Marr, 1982

David Marr

Credited with early computational approach for vision



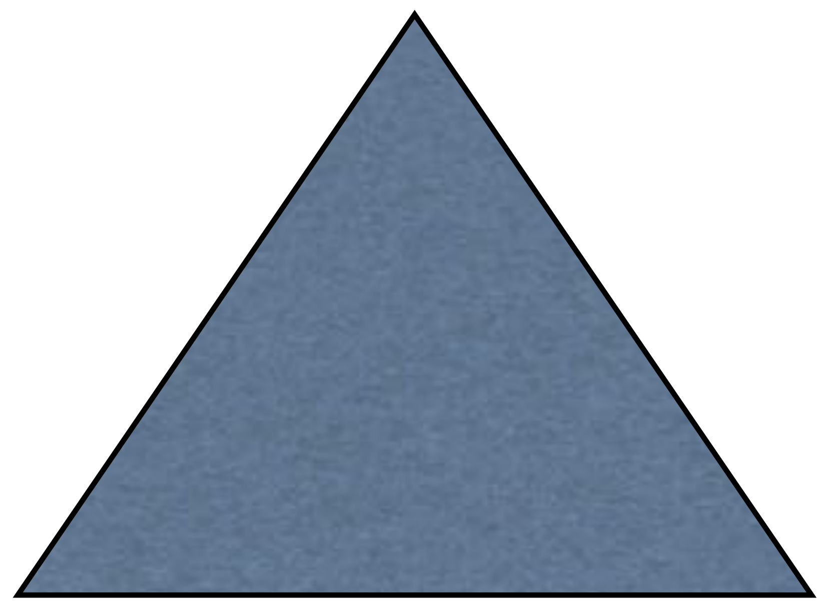


Low-level Mid-level High-level

(Aside) 3 R's: Another taxonomy of vision

(from Jitendra Malik)

Recognition



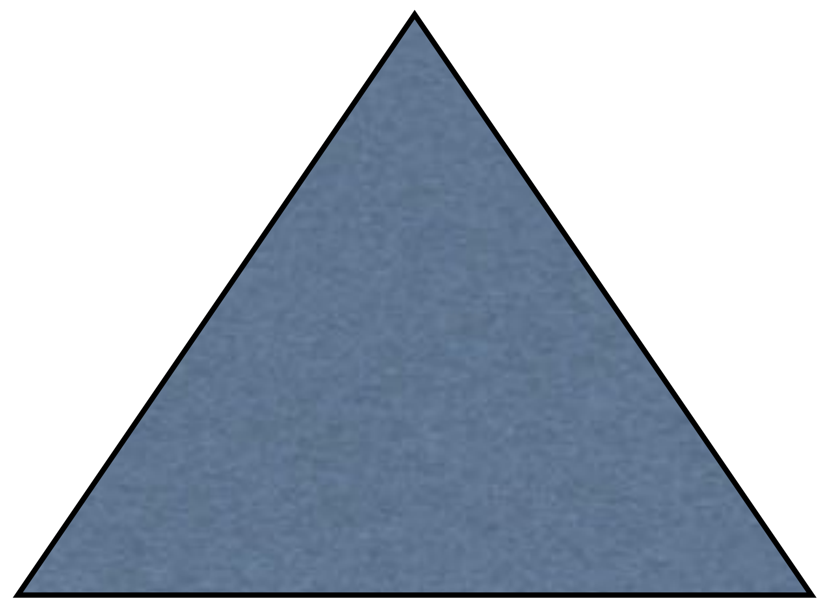
Reconstruction Reorganization

Future research directions seem to lie "in the middle"

1. Joint recognition and reconstruction (e.g., learning to infer 3D)
2. Joint recognition and reorganization (e.g., learning to group pixels)
3. Joint reconstruction and reorganization (e.g., grouping point clouds into surfaces)

(Aside) 4 R’s: 3R’s + rudiments

(from Serge Belongie)

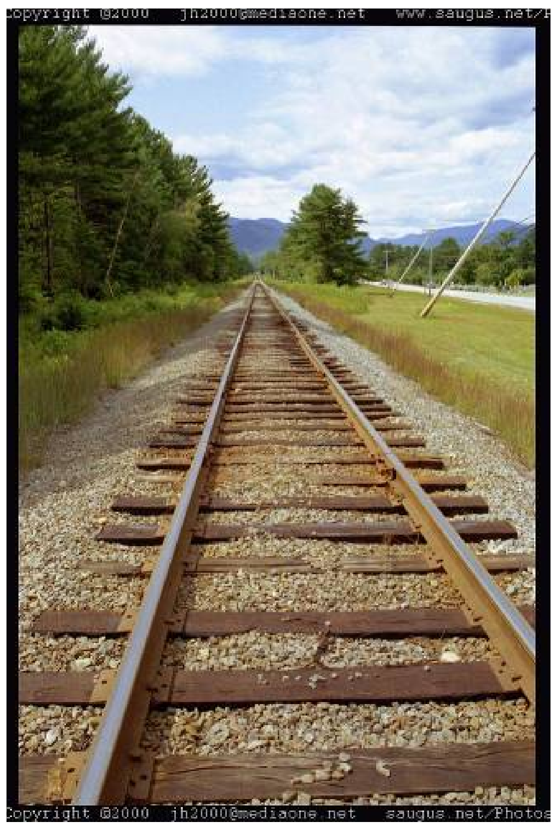


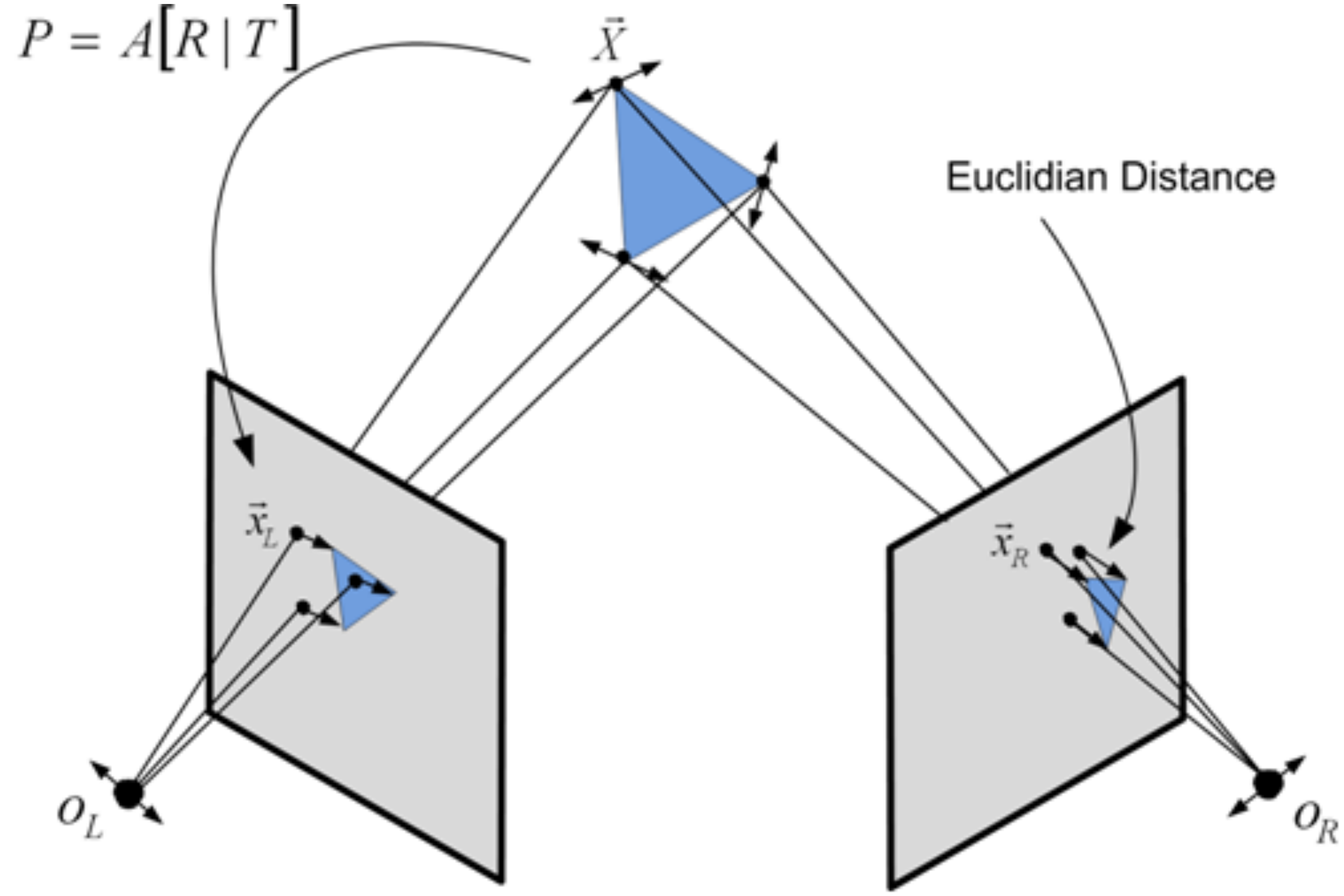
Toolboxes: linear algebra, probability, physics, statistics, machine learning

Syllabus

Date	Topic	Reference	HW Due
1/12	Introduction	Ballard & Brown Chap 1	
1/14	Matlab		
	Image Processing		
1/19	Convolutions		HW0: Intro (pdf, code)
1/21	Filtering + Bag-of-words		
1/26	Fourier		
1/28	Interest Points		
	Image Formation		
2/2	Descriptors		HW1 (Matching)
2/4	Camera Projection + Homographies		
2/9	Optical flow		
2/11	Image Alignment		
	Geometry		
2/16	Motion Segmentation		
2/18	2-View Geometry		HW2 (Mosaics)
2/23	SFM		
2/25	Structured light		
	Recognition		
3/1	Machine learning		
3/3	Matching and Retrieval		HW3 (Tracking)
3/15	Template Matching		
3/17	Parts		Project Proposal
3/22	Deep learning 1		
3/24	Deep learning 2		HW4 (Reconstruction)
	Grouping		
3/29	Binary Morphology		
3/31	Textures		
4/5	Graphcuts		
4/7	Catchup		
	Physics		
4/12	Light Transport		HW5 (Recognition)
4/14	NO CLASS (spring carnival)		
4/19	Color		
4/21	Computational Imaging		
4/26	Project presentations		Proj. Presentation + Writeup
4/28	Project presentations		

Geometric models

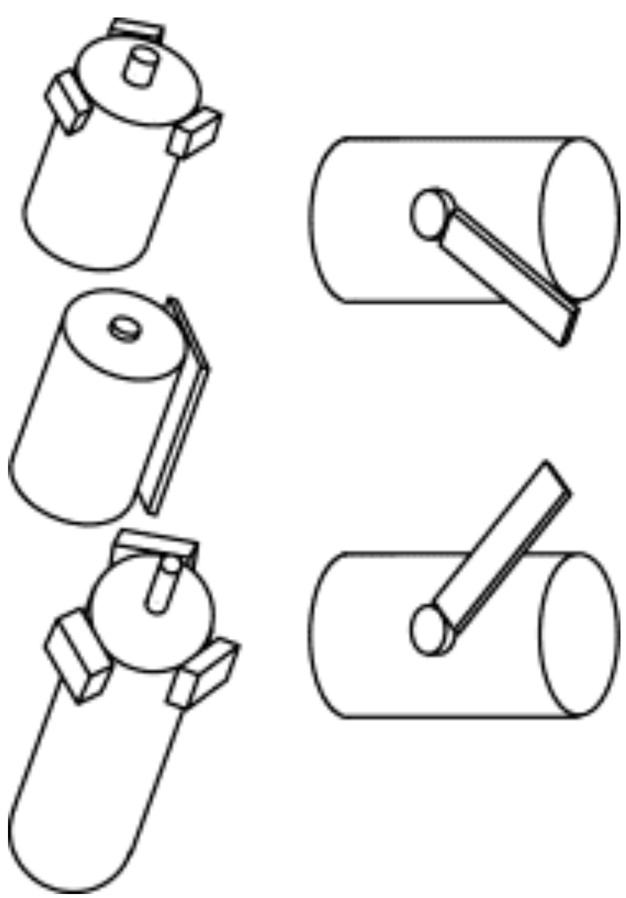




Geometric models (cont'd)

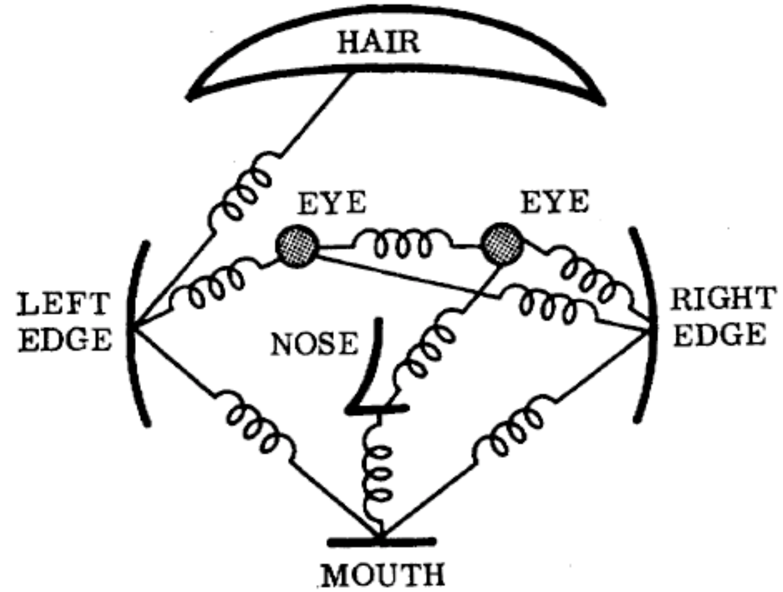
Generalized Cylinder

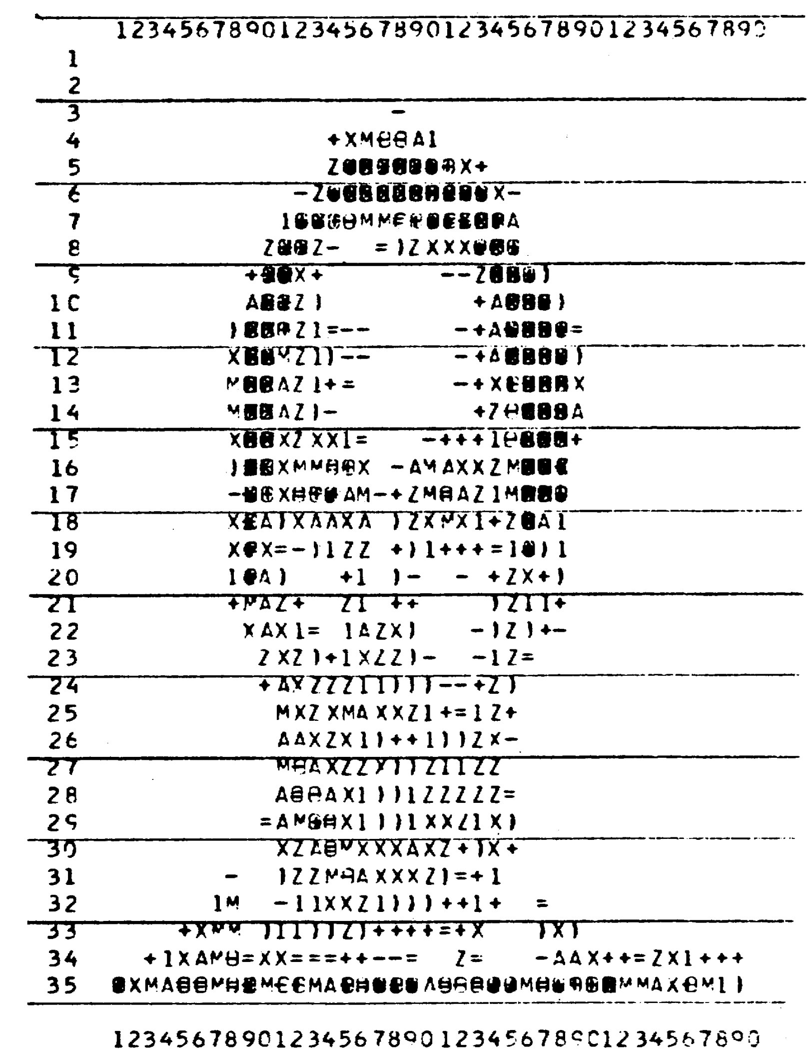
Brooks & Binford, 1979



- Pictorial Structure

Fischler and Elschlager, 1973

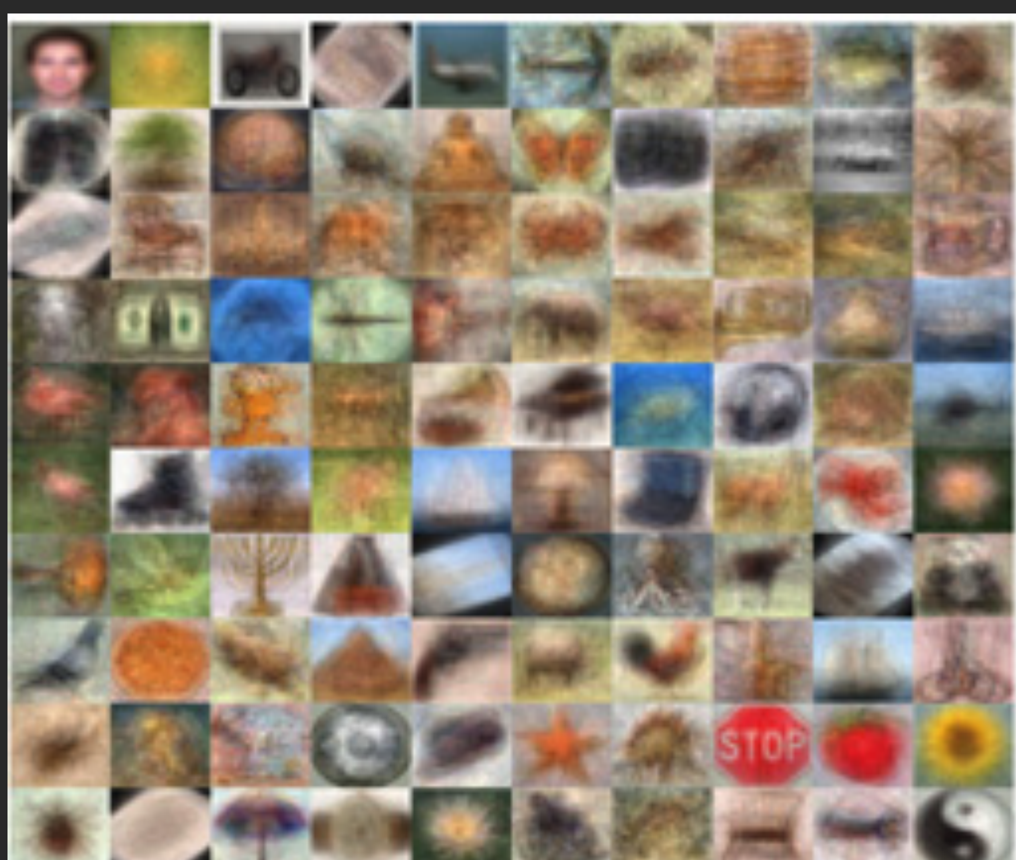




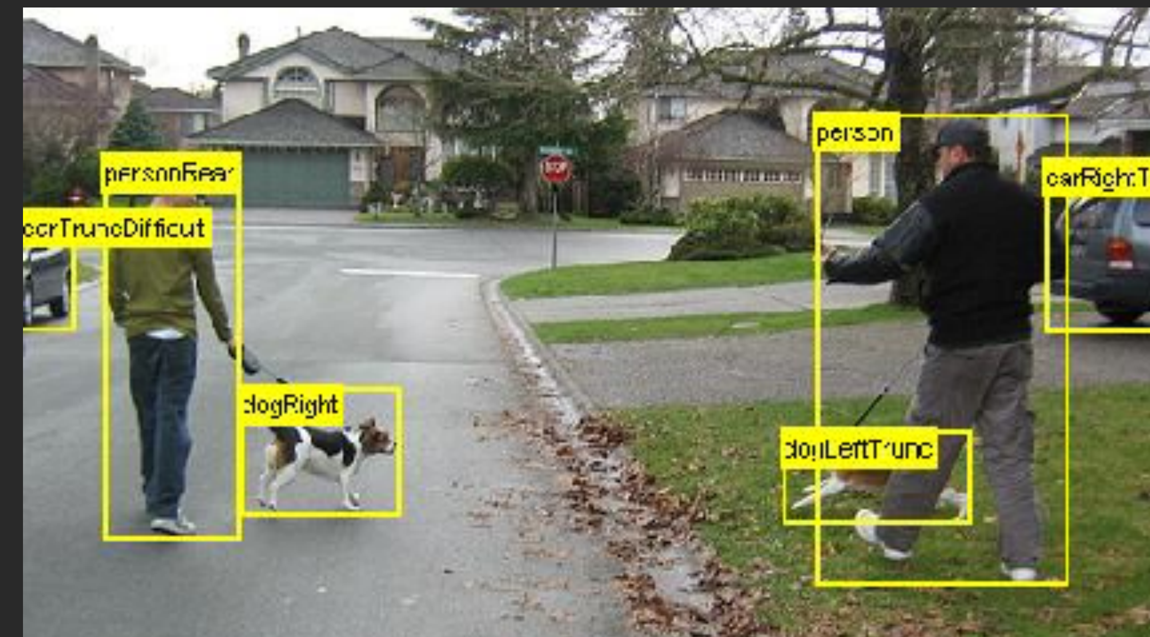
HAIR WAS LOCATED AT (6, 18)
L/EDGE WAS LOCATED AT (18, 10)
R/EDGE WAS LOCATED AT (18, 25)
L/EYE WAS LOCATED AT (17, 13)
R/EYE WAS LOCATED AT (17, 21)
NOSE WAS LOCATED AT (22, 18)
MOUTH WAS LOCATED AT (24, 17)

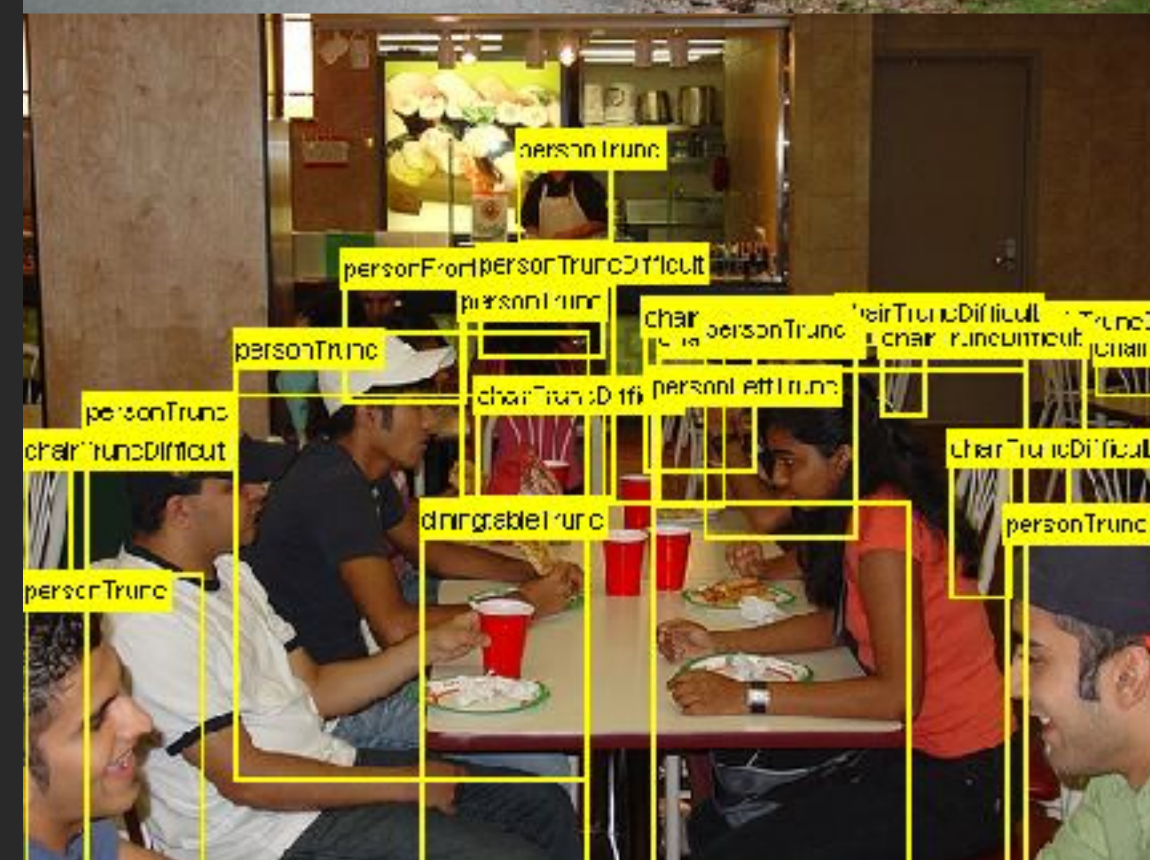
Contemporary vision: big data

Columbia Dataset (1996)



Caltech 101/256
Image Net

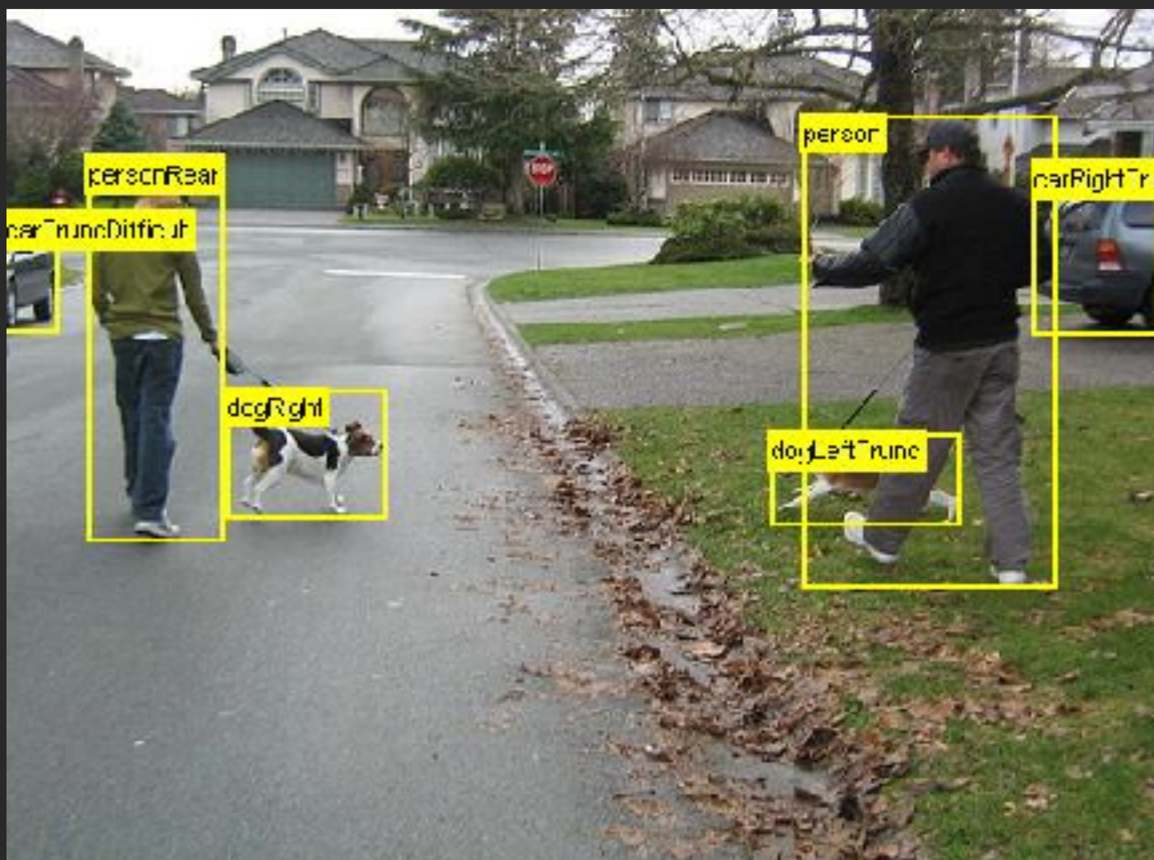




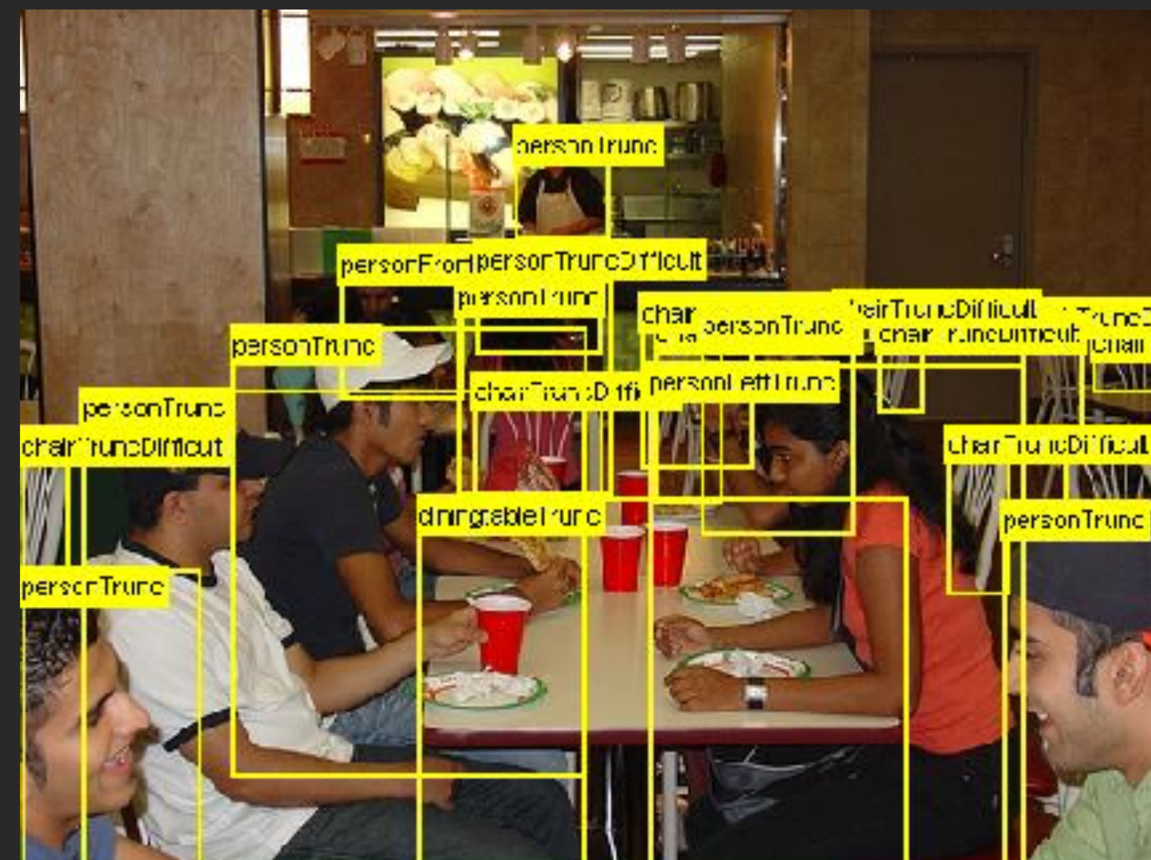
Flickr dataset (05-12)

“In-the-wild”

PASCAL Visual Object Challenge



5000 training images

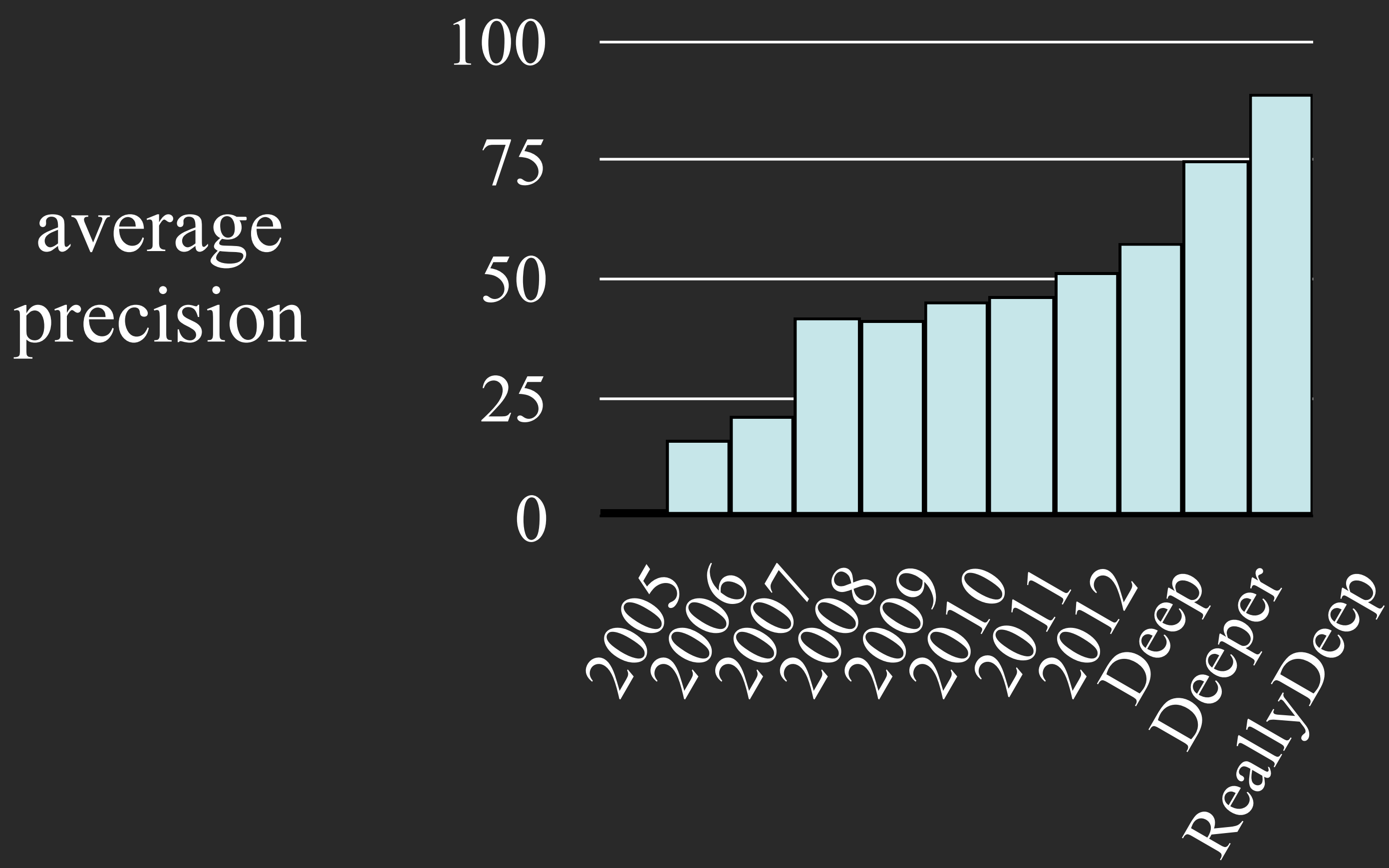


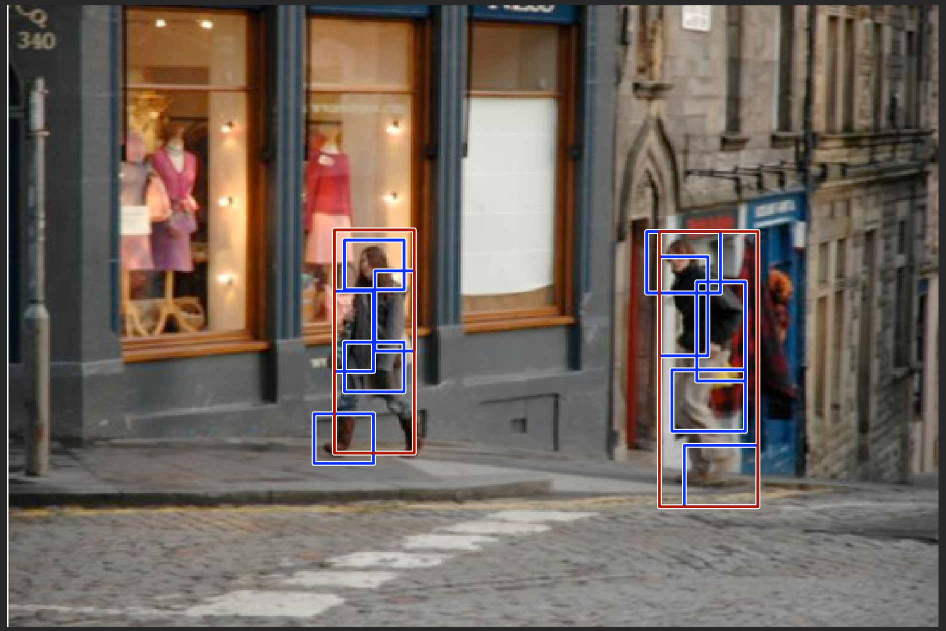
5000 testing images

20 everyday object categories from Flickr

airplane bike bird boat bottle bus car cat chair cow table dog
horse motorbike person plant sheep sofa train tv

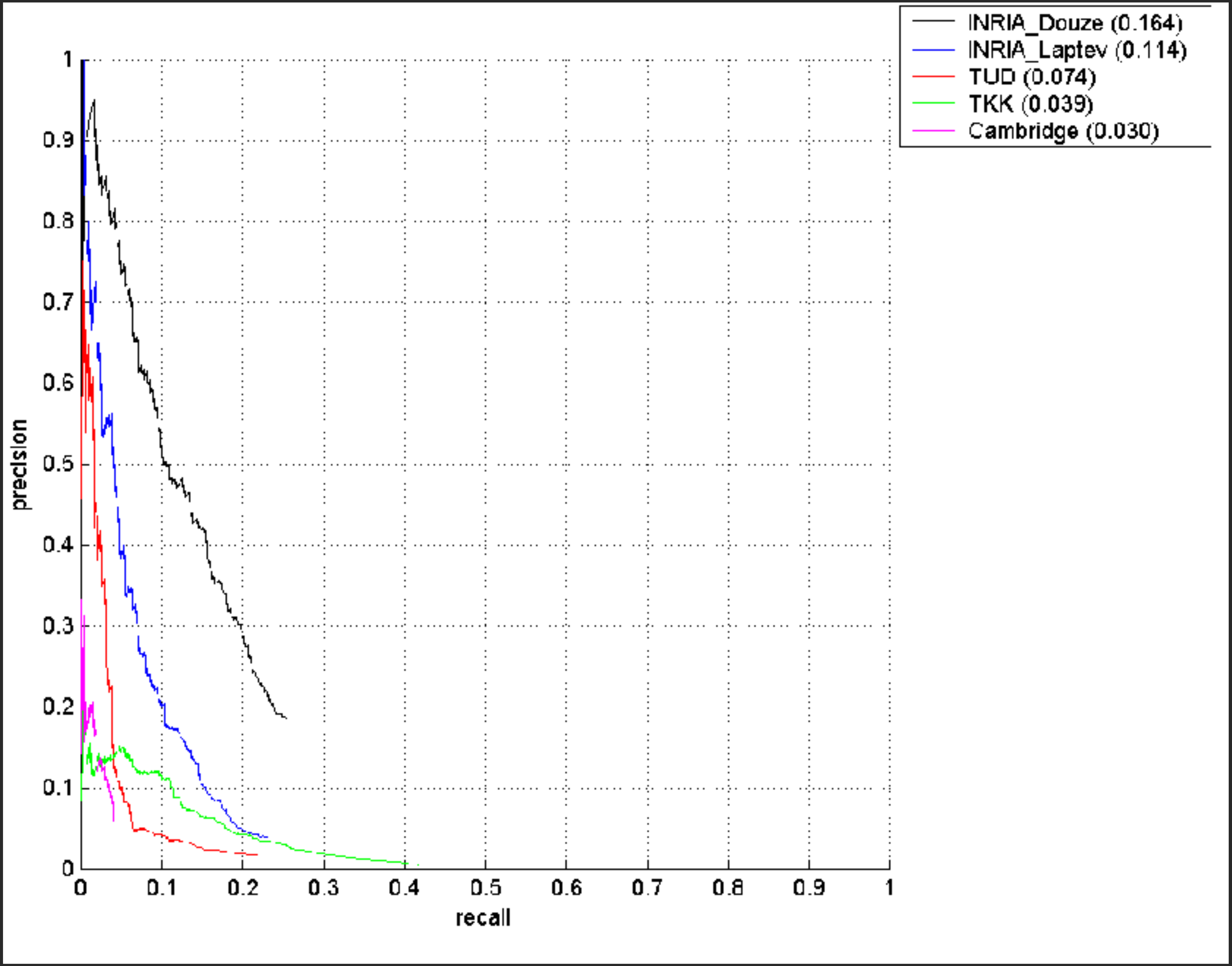
6 years of PASCAL people detection





1% to 90% in 10 years

Precision-recall (2006)

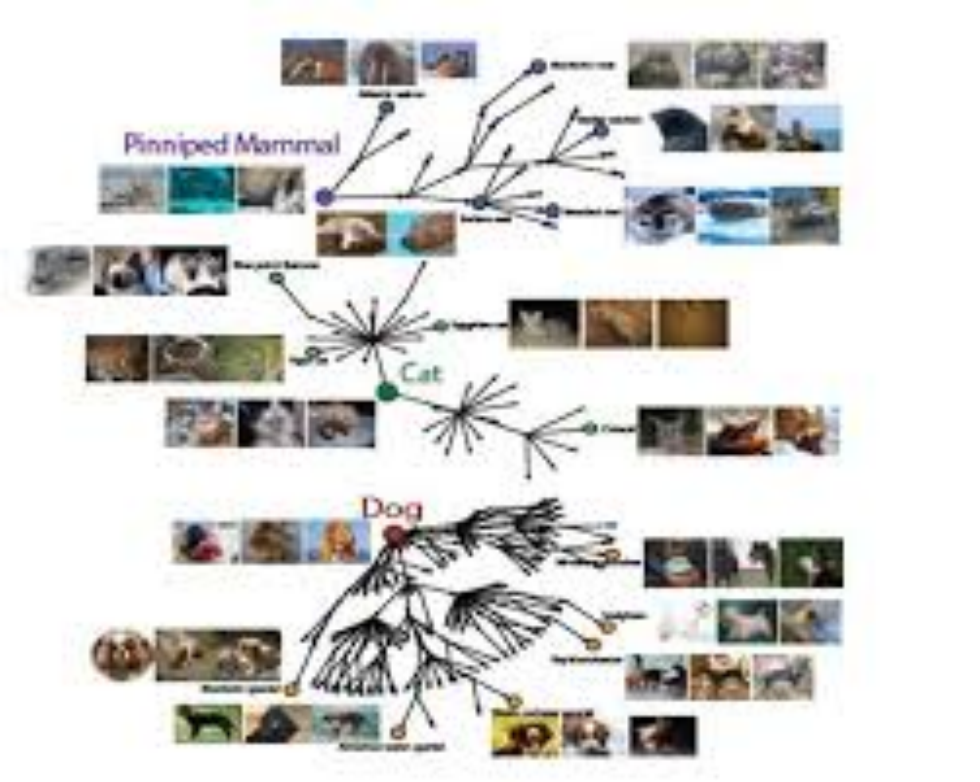


A glimpse into the current future…

Imagenet large-scale visual recognition challenge

1000 classes, ~1000 examples per class

Imagenet 2014 image classification results

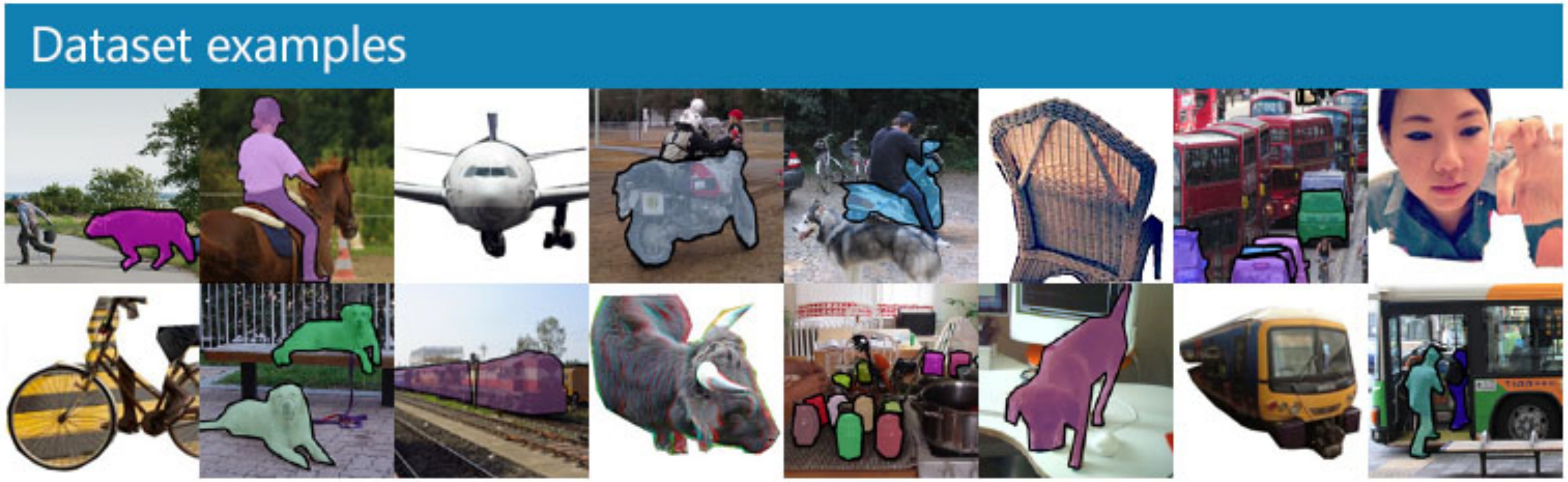


Model	Resolution	Crops	Models	Top-1 error	Top-5 error
GoogLeNet ensemble	224	144	7	-	6.67%
Deep Image low-res	256	-	1	-	7.96%
Deep Image high-res	512	-	1	24.88	7.42%
Deep Image ensemble	variable	-	-	-	5.98%
BN-Inception single crop	224	1	1	25.2%	7.82%
BN-Inception multicrop	224	144	1	21.99%	5.82%
BN-Inception ensemble	224	144	6	20.1%	**4.9%***

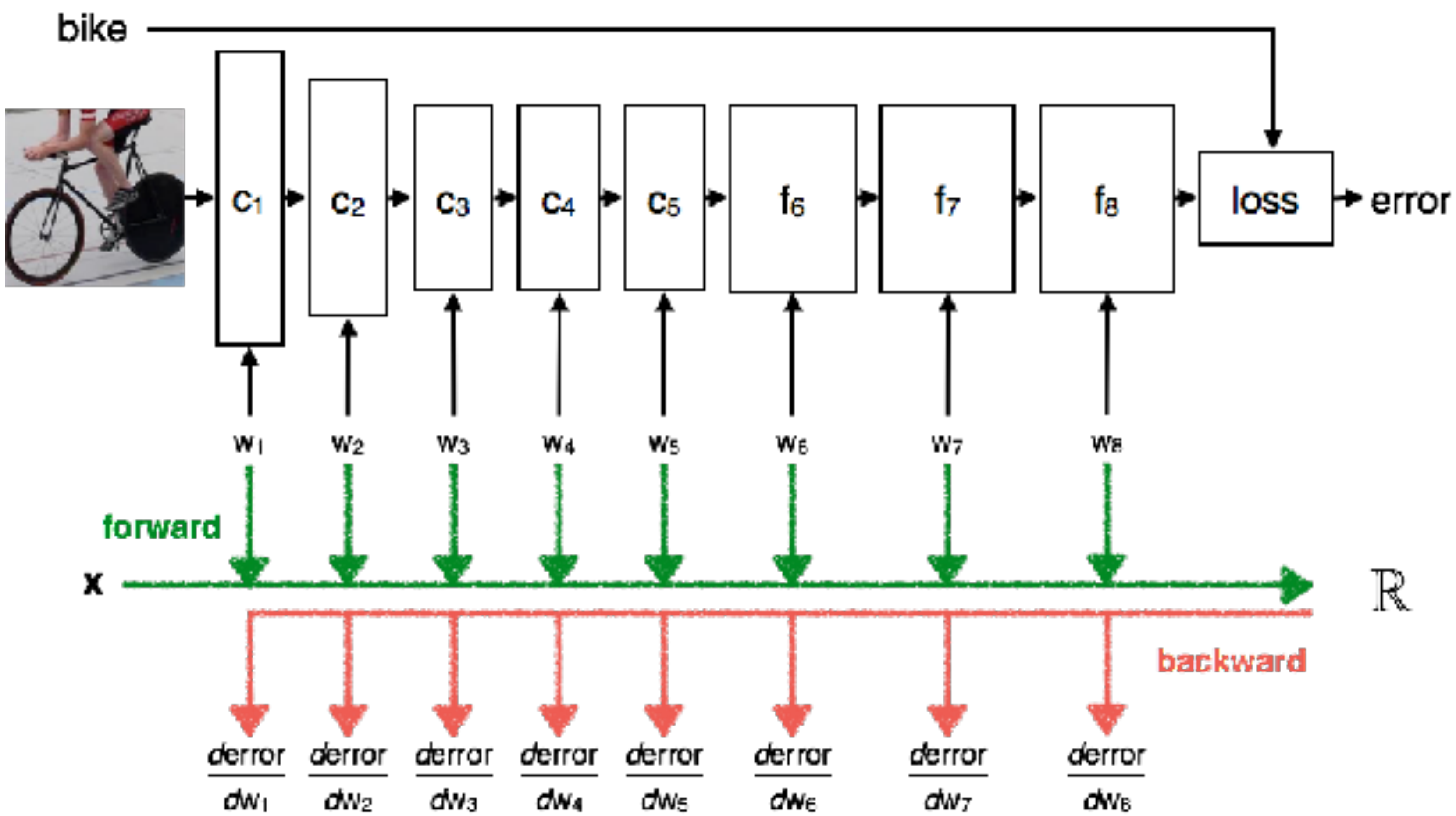
Human top-5 error: 5.1 %

A recent dataset

http://mscoco.org/



A recent toolbox…



http://www.vlfeat.org/matconvnet/

Next class

Basic image processing