16-899A

Visual World
as seen by
Neurons and Machines

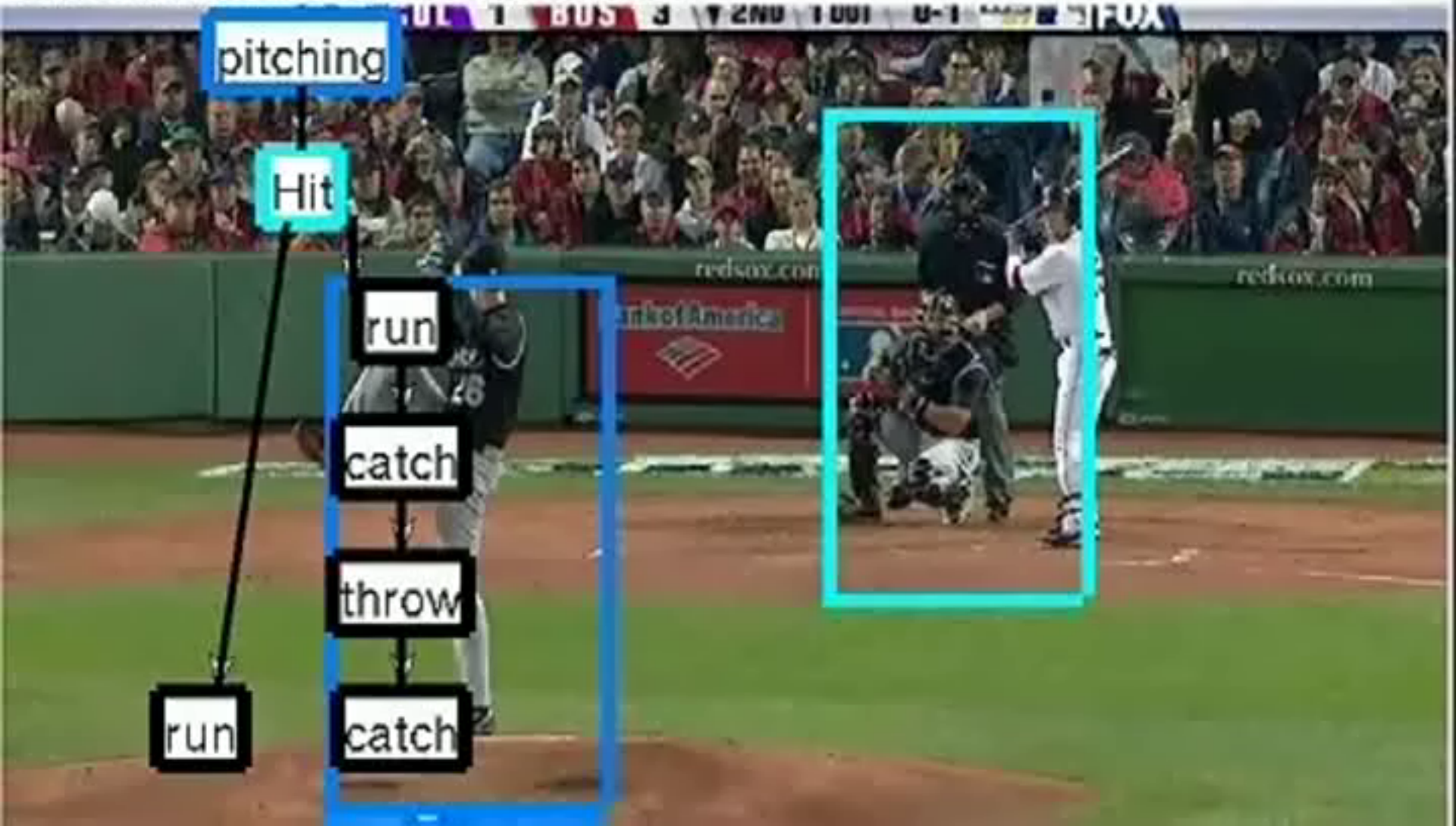
Course Information

- Time:
  - Monday, Wednesday 12:00-13:20
- Location:
  - NSH 3002
- Office Hours:
  - Email us for appointments
- Contact:
  - abhinavg@cs.cmu.edu , EDSH 213
  - elissa@cnbc.cmu.edu , Mellon Institute
- Website: http://graphics.cs.cmu.edu/courses/16-899A/2014_spring/

People

- Abhinav Gupta
- Ph.D. 2009, University of Maryland

Pitcher pitches the ball before Batter hits. Batter hits and then simultaneously Batter runs to base and Fielder runs towards the ball. Fielder runs towards the ball and then Fielder catches the ball. Fielder catches the ball and then Fielder throws to the base. Fielder at Base catches the ball at base after Fielder throws to the base.

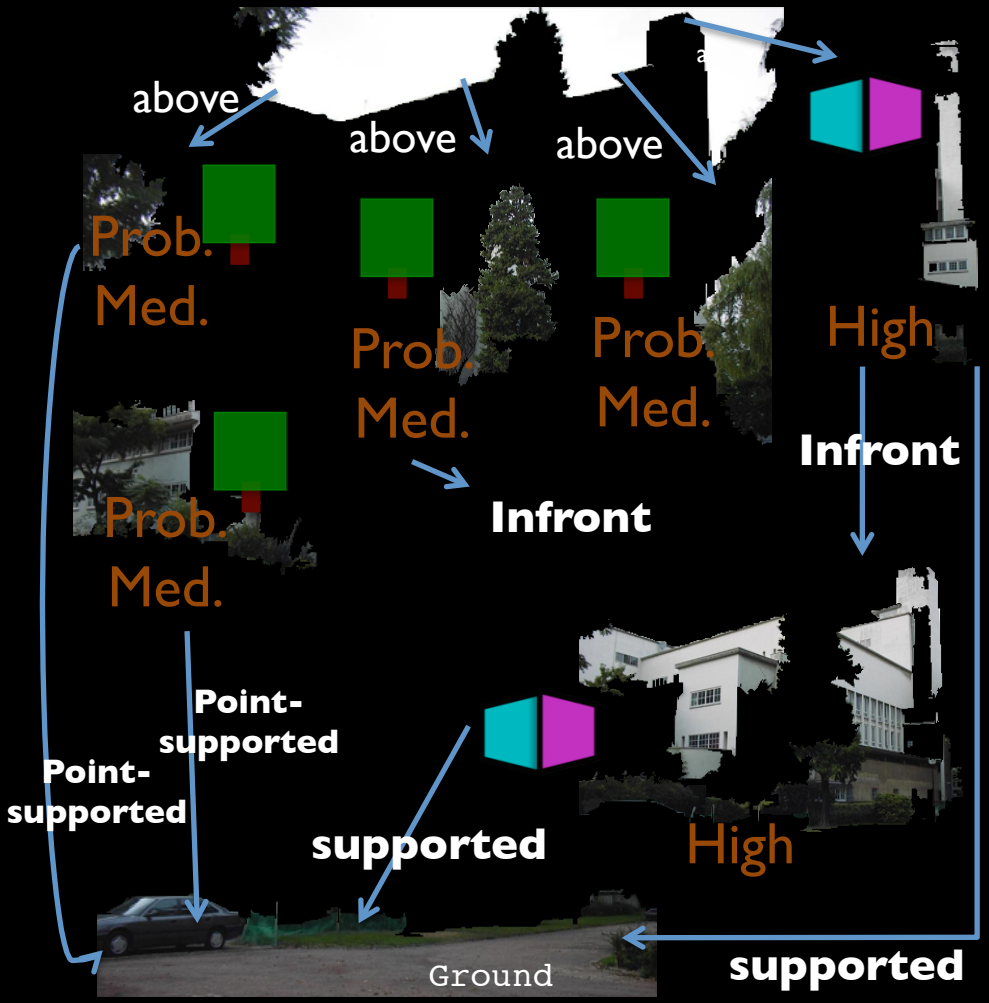


People

- Abhinav Gupta
- Ph.D. 2009, University of Maryland
- Postdoctoral Fellow, Carnegie Mellon University, 2009-11

blocks world revisited

Original Image



3D Parse Graph

All results and Code: http://www.cs.cmu.edu/~abhinavg/blocksworld

NEIL - Never Ending Image Learner

- Computer program that runs 24 hours per day, 7 days per week to download images, understand them.

- Builds visual models and learn common-sense knowledge.

People

- Abhinav Gupta
- Ph.D. 2009, University of Maryland
- Postdoctoral Fellow, Carnegie Mellon University, 2009-11
- Research Interests:
  - Vision: Image and Video Understanding
  - Language and Vision
  - Learning from Big Data

People

- Elissa Aminoff
- PhD 2008, Harvard University, Psychology – Cognitive Neuroscience

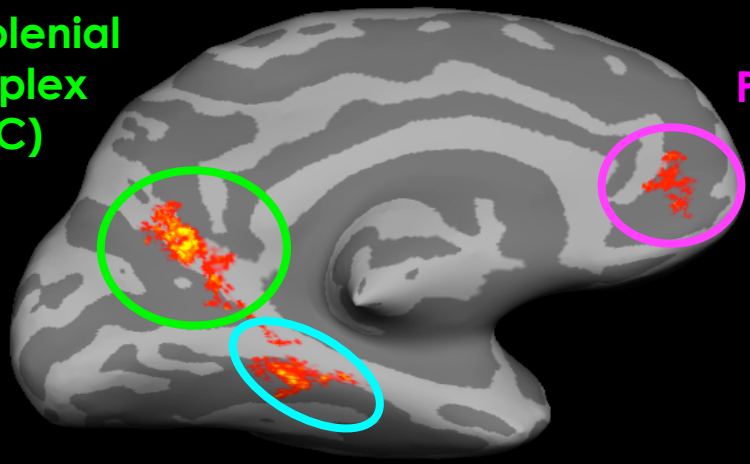
Strong Context

Weak Context

VS.

Retrosplenial Complex (RSC)

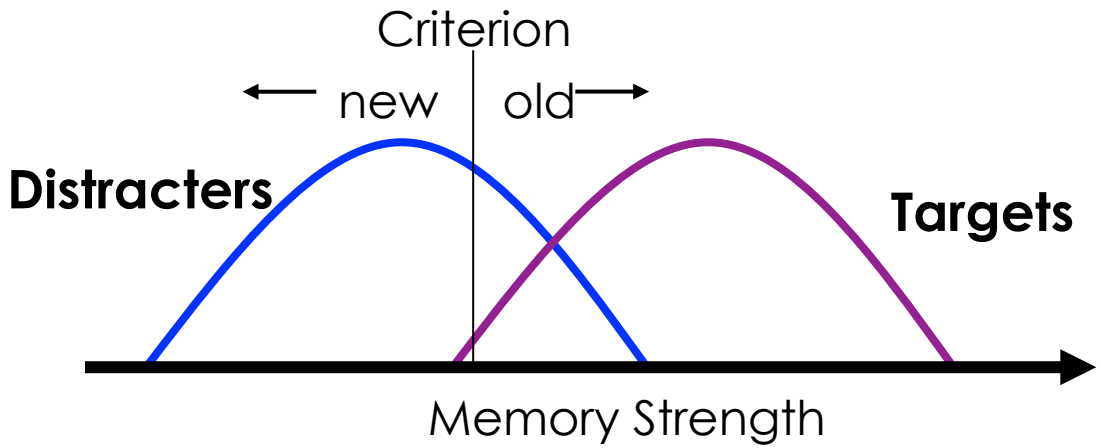
Medial Prefrontal Cortex (MPFC)

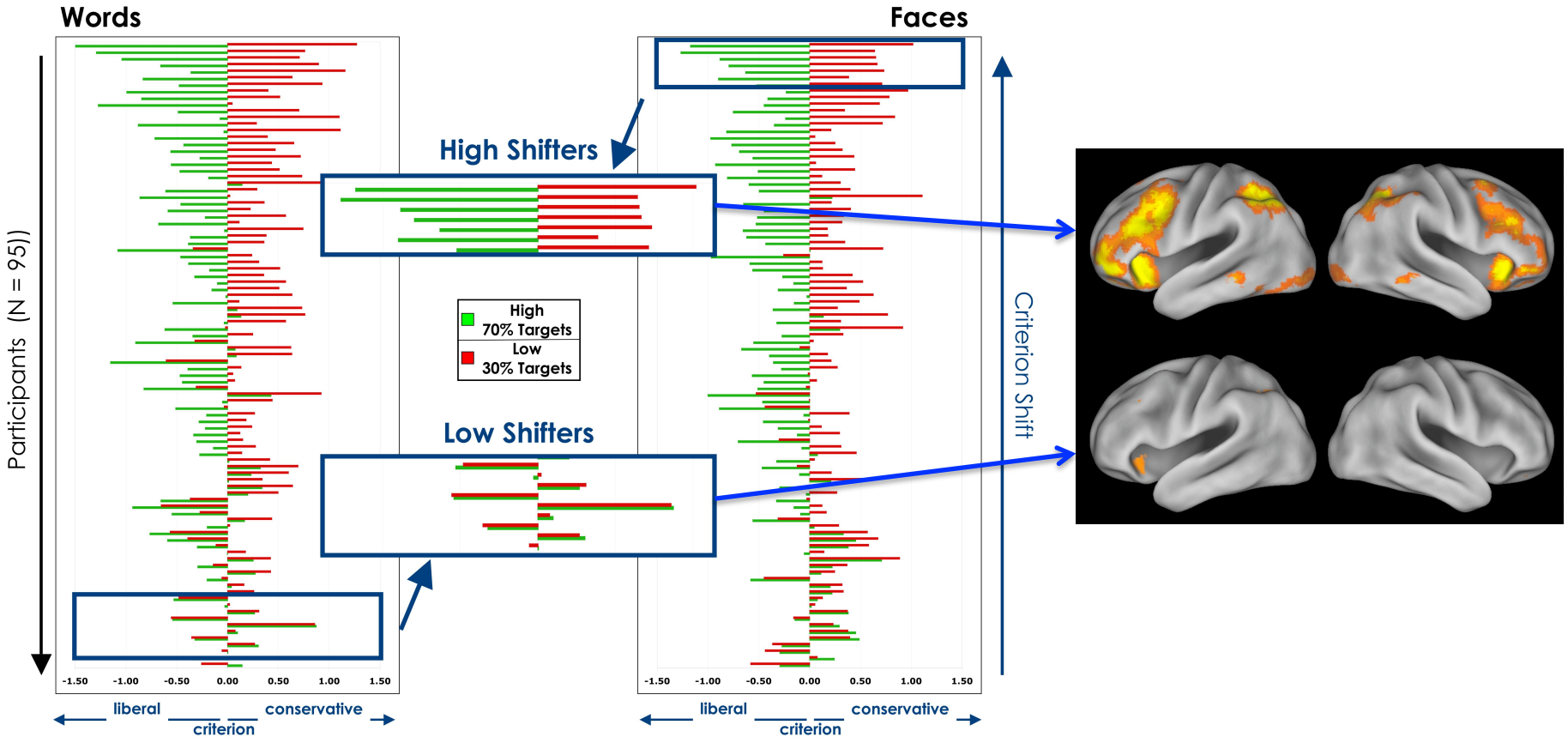


Parahippocampal Cortex (PHC)

People

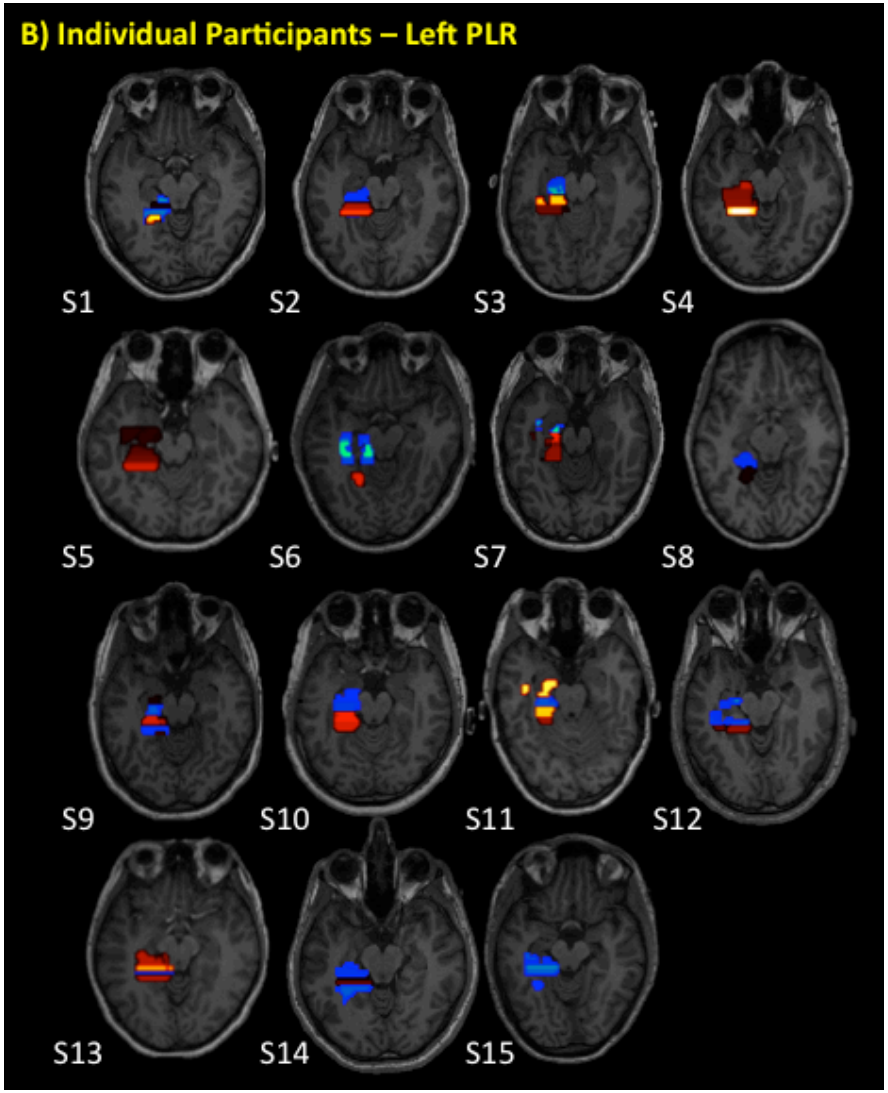
- Elissa Aminoff
- PhD 2008, Harvard University, Psychology – Cognitive Neuroscience
- Postdoctoral Fellow 2008-2011, University of California-Santa Barbara


Criterion
new
old
Distracters
Targets
Memory Strength


Words
Faces
High Shifters
Low Shifters
High
70% Targets
Low
30% Targets
Participants (N = 95))
Criterion Shift
-1.50 -1.00 -0.50 0.00 0.50 1.00 1.50
liberal
conservative
criterion

People

- Elissa Aminoff
- PhD 2008, Harvard University, Psychology – Cognitive Neuroscience
- Postdoctoral Fellow 2008-2011, University of California-Santa Barbara
- Research Scientist, Center for the Neural Basis of Cognition


B) Individual Participants – Left PLR
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15

People

- Elissa Aminoff
- PhD 2008, Harvard University, Psychology – Cognitive Neuroscience
- Postdoctoral Fellow 2008-2011, University of California-Santa Barbara
- Research Scientist, Center for the Neural Basis of Cognition

- Research Interests:
  - Scene Perception
  - Associative processing
  - High-level visual processing in the brain
  - Memory as a direct influence of perception
  - How to address these through computational models

16-899A: Visual World by Neurons and Machines

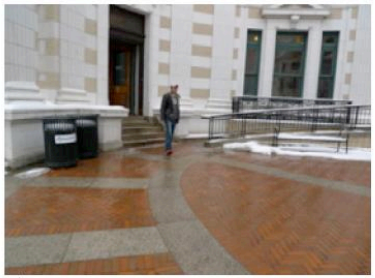
What is this course about?

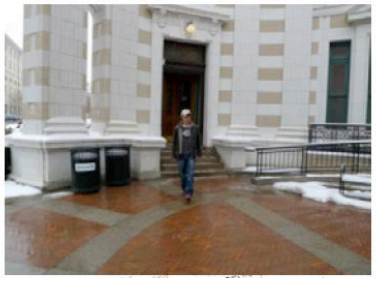
from a computer vision scientist perspective

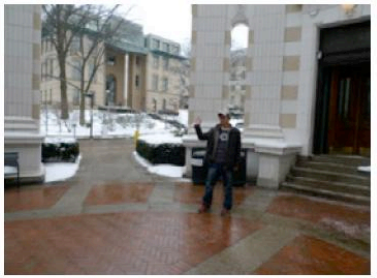
What is the goal of Computer Vision?

Systems that can “understand” Visual Data

understanding visual data

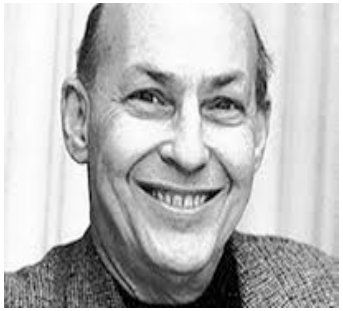






understanding visual data

understanding visual data



Minsky

"spend the summer linking a camera to a computer and getting the computer to describe what it saw"

Sussman


chair
table
car
table
road
keyboard
road

Why is understanding the visual world, while so effortless for humans, so excruciatingly difficult for computers?

Maybe solving the wrong problem?

Systems that can "understand" Visual Data

What does it mean to understand?

The Vision Story Begins…

“What does it mean, to see? The plain man's answer (and Aristotle's, too). would be, to know what is where by looking.”

-- David Marr, *Vision (1982)*

Slide Credit: Alyosha Efros

Vision: a split personality

"What does it mean, to see? The plain man's answer (and Aristotle's, too). would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is."

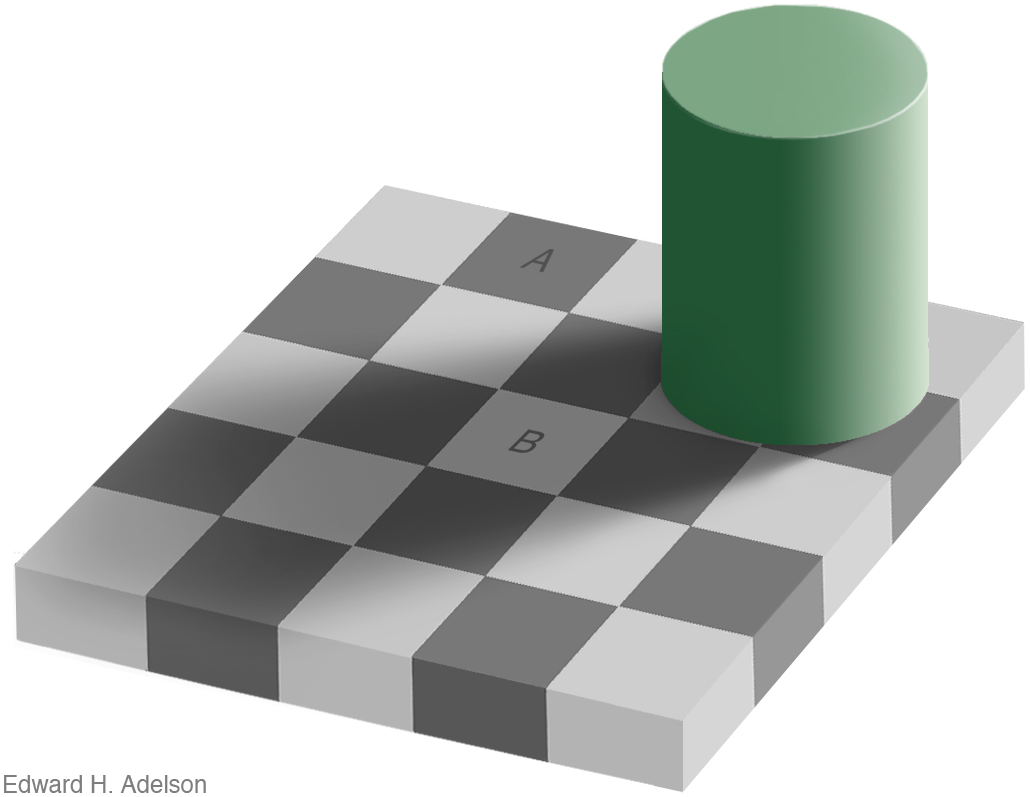
Answer #2: looks like flat sittable surface of the couch

Which do we want?
Is the difference just a matter of scale or is there some fundamental difference?

Answer #1: *pixel of brightness 243 at position (124,54) …and depth .7 meters*

Measurement vs. Perception

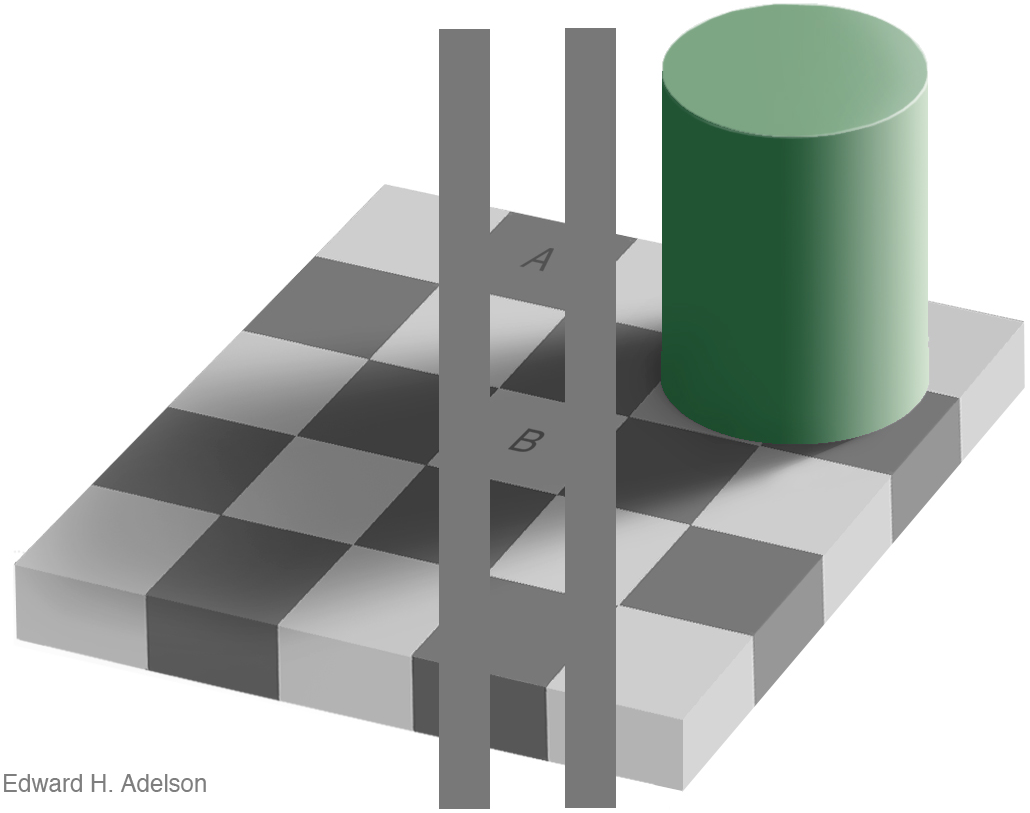
Brightness: Measurement vs. Perception



Edward H. Adelson

Slide Credit: Alyosha Efros

Brightness: Measurement vs. Perception



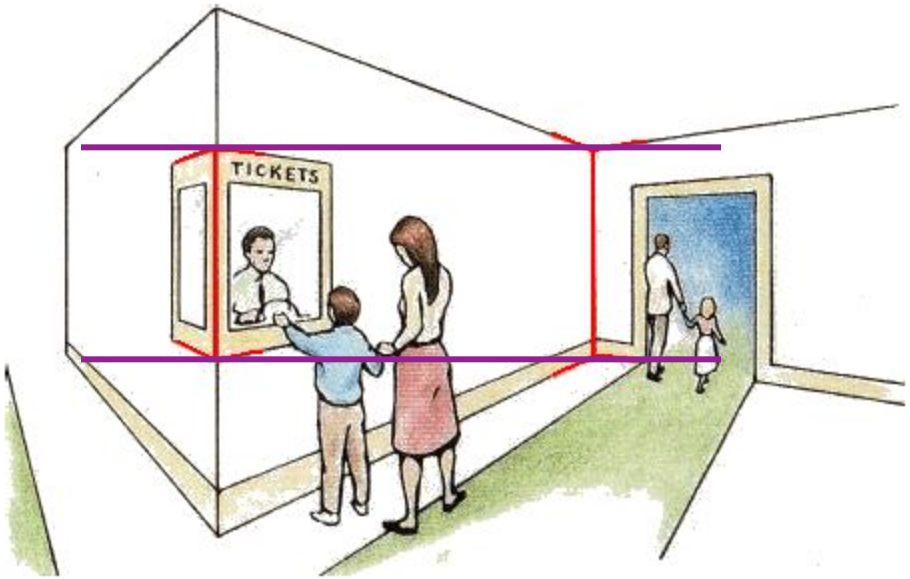
Edward H. Adelson

Proof!

Slide Credit: Alyosha Efros

Measurement

Length



Müller-Lyer Illusion

http://www.michaelbach.de/ot/sze_muelue/index.html

Slide Credit: Alyosha Efros

Measurement

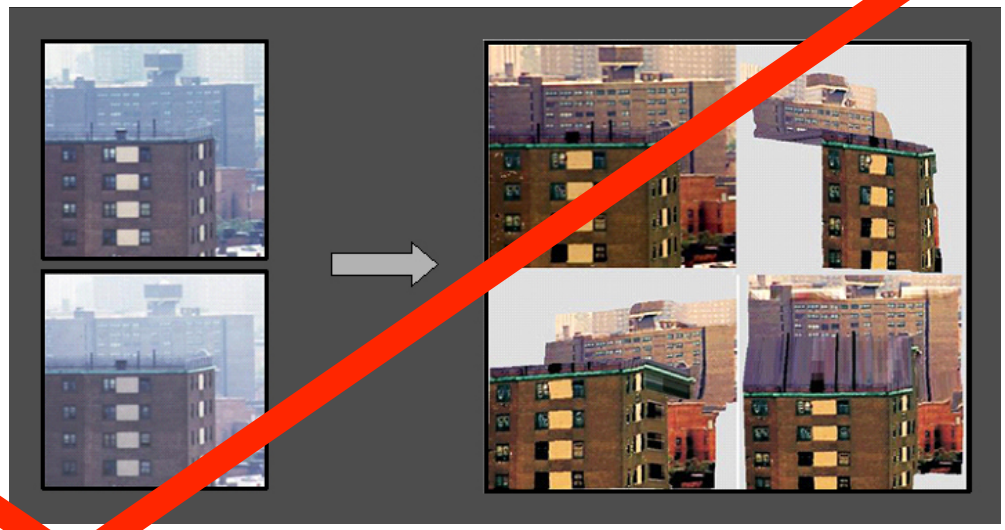
Capturing physical quantities like pixel brightness, depth, etc.

Perception/Understanding

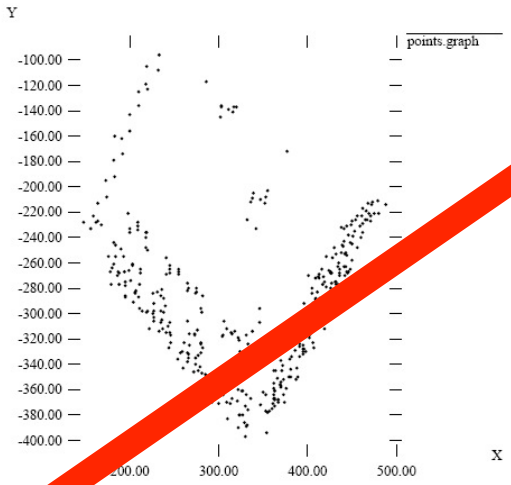
- a high-level representation that captures the semantic structure of the scene and its constituent objects.
- Subjective – Depends on Task and Agent
- Intersection of what you see and what you believe (prior knowledge)

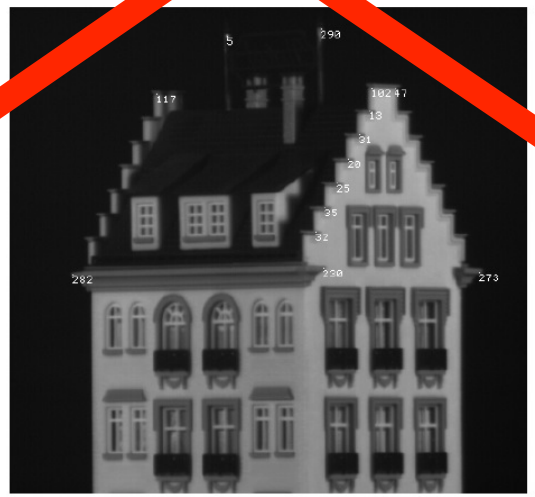
Vision as Measurement Device

Real-time stereo on Mars



Physics-based Vision





Structure from Motion

Virtualized Reality

Slide Credit: Alyosha Efros

…but why do we care about perception?

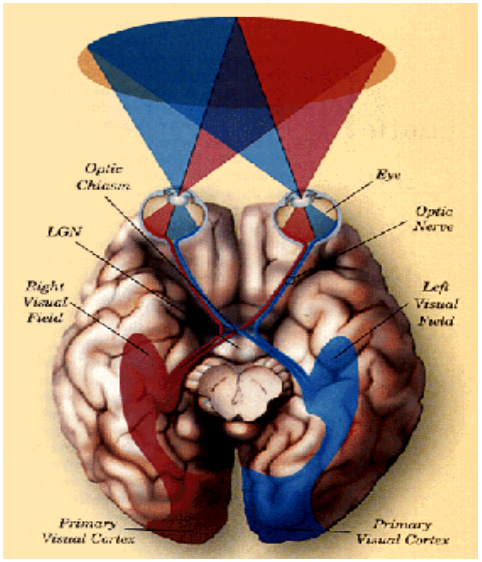
The goals of computer vision (**what + where**) are in terms of what humans care about.

Our Problem

- Input is world designed by “humans”..
- Output is defined in terms of what “humans” want..
- Only machine that solves this task is “humans”..

Yet

None of us care to read what information does human brain extract and how does it do it.



Airplane Reasoning!

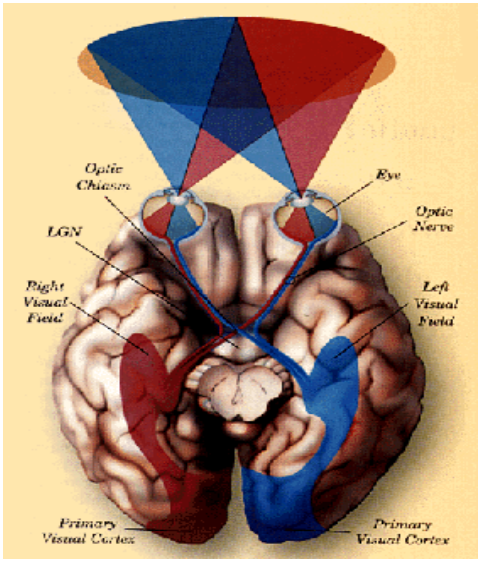
- Computer vision does not need to solve it exactly the way humans do it…
- Do we really need computer vision?? Start putting RFID's on every object in the world….sensing it becomes quite easy….
- **Hidden Goal of Computer Vision:** We are scientist and we want to understand human cognition and reasoning…

Our Problem

- Input is world designed by “humans”..
- Output is defined in terms of what “humans” want..
- Only machine that solves this task is “humans”..

Yet

None of us care to read what information does human brain extract and how does it do it.



This Class!

an opportunity to correct that …. learn more about how does human vision system work

in context of problems we face in computer vision everyday !

Goals of this Class

Goal 1: What insights from the human visual system can we use in computer vision?

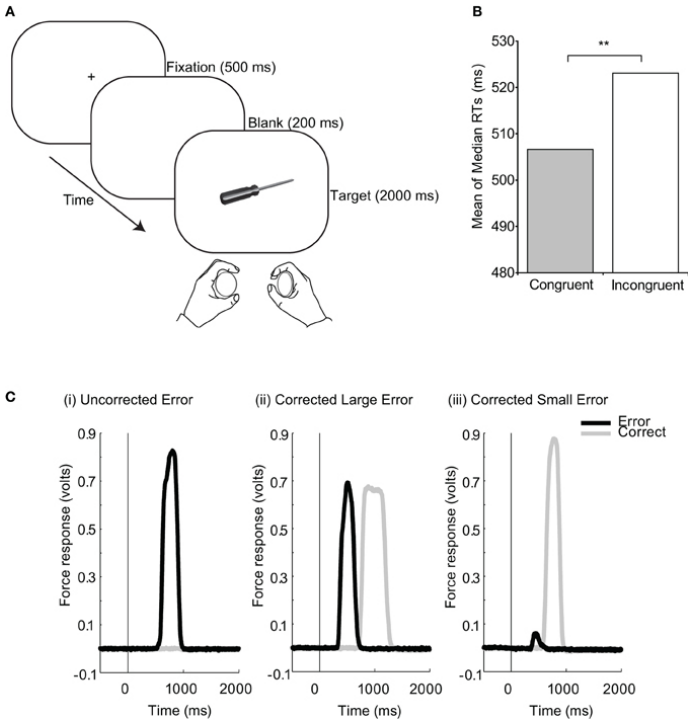
an awesome opportunity to get ideas to publish in the field of computer vision

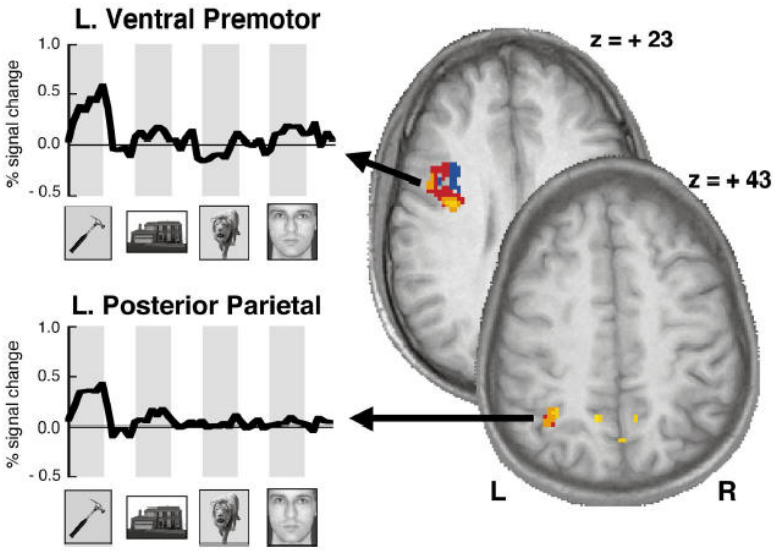
Personal Story!

Personal Story!

Personal Story!

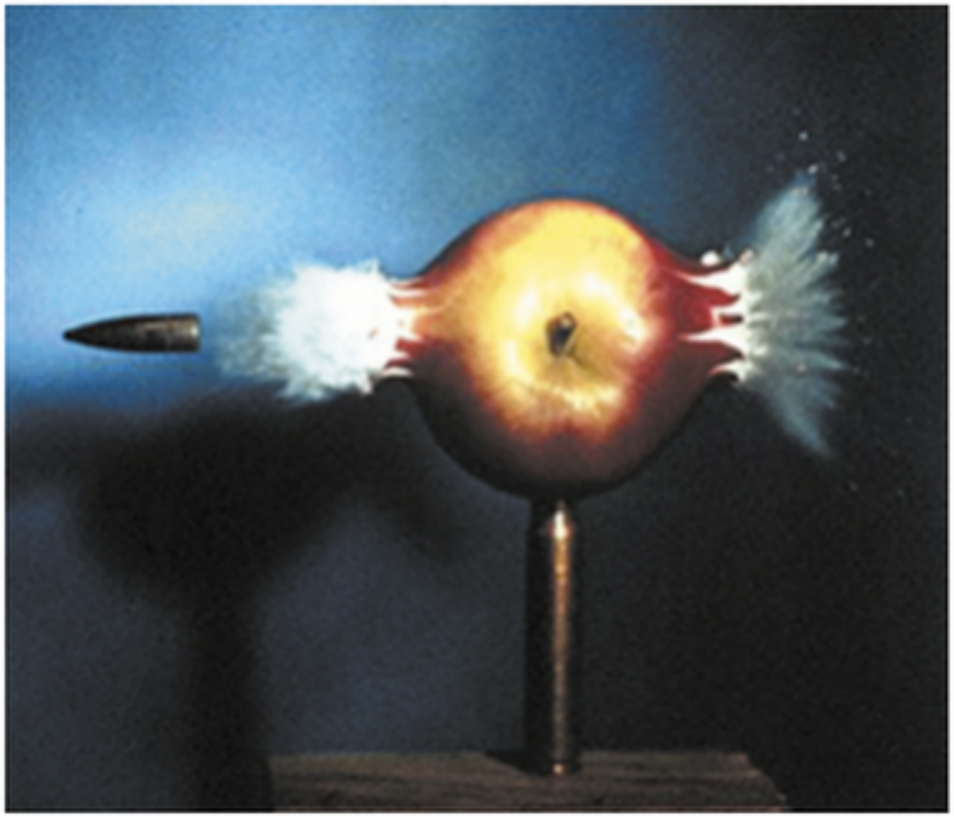
- Neuroscience section was close to the place she worked!
- Spend time reading neuroscience and behavioral science journals…





My First Paper: Objects in Action

My Second Paper: But Still it Moves



My Second Paper: But Still it Moves

phoning

running

walking

ridinghorse

PASCAL Action Recognition Challenge

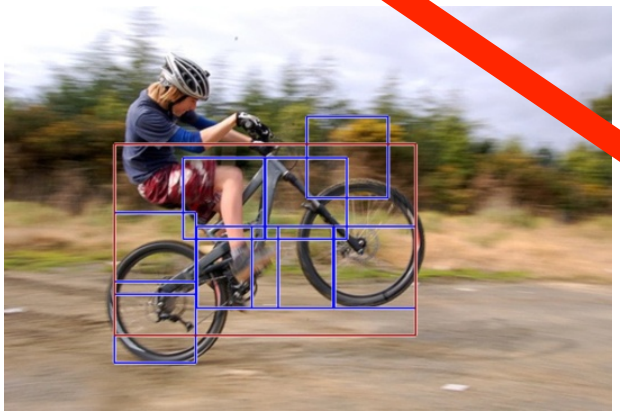
My Latest Paper: Jacob Walker

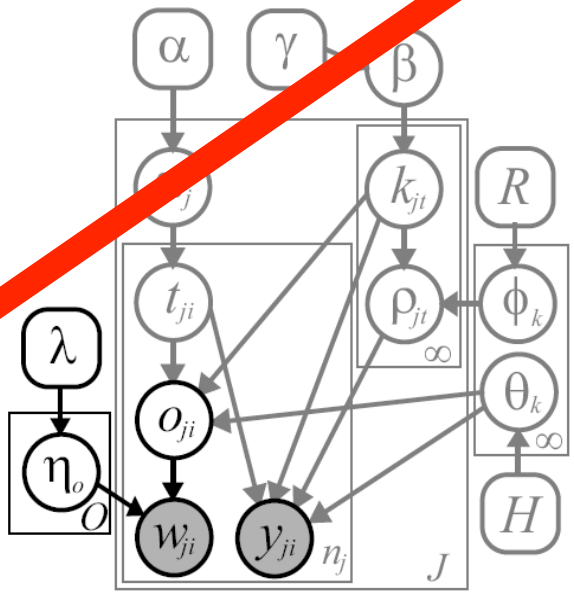
Goal 2: Broaden your perspective, think more than technical issues and solving problems…

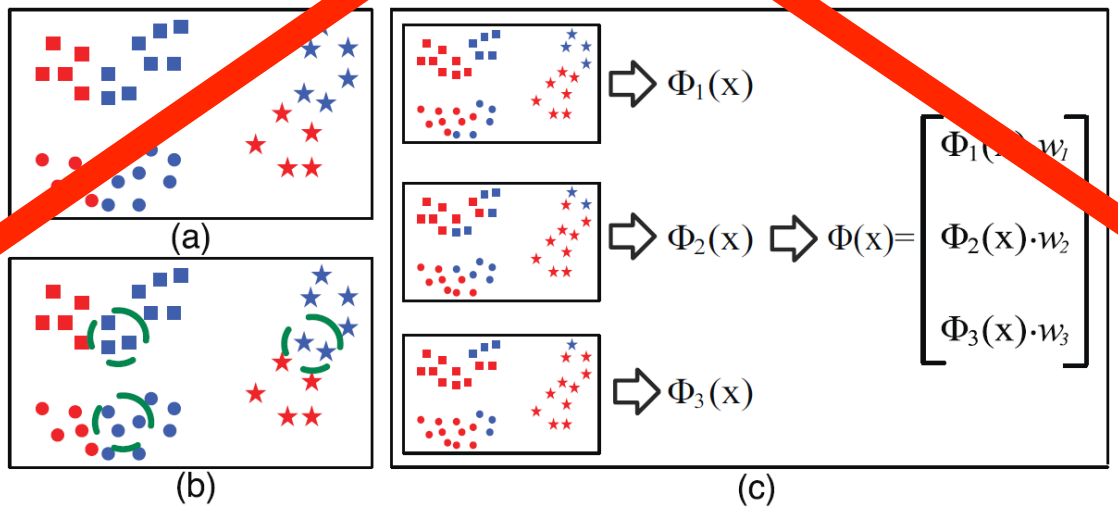
- *Stand in-front of a renowned computer vision expert and discuss/ fight for your ideas…*
- *an awesome opportunity to discuss what are the right ideas to work on and find the right problems*

Therefore,

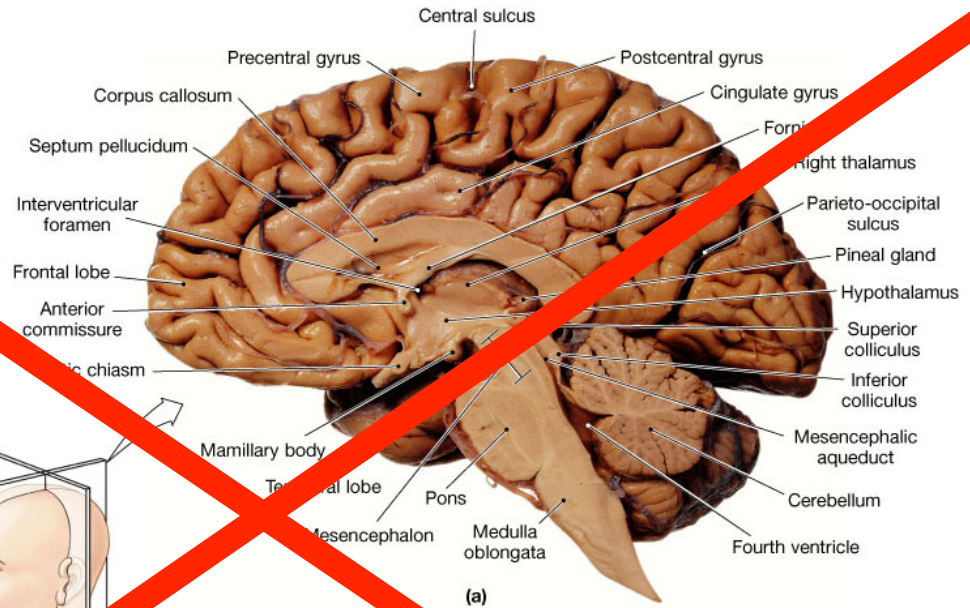
- Latent SVM

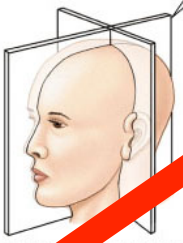






Therefore,

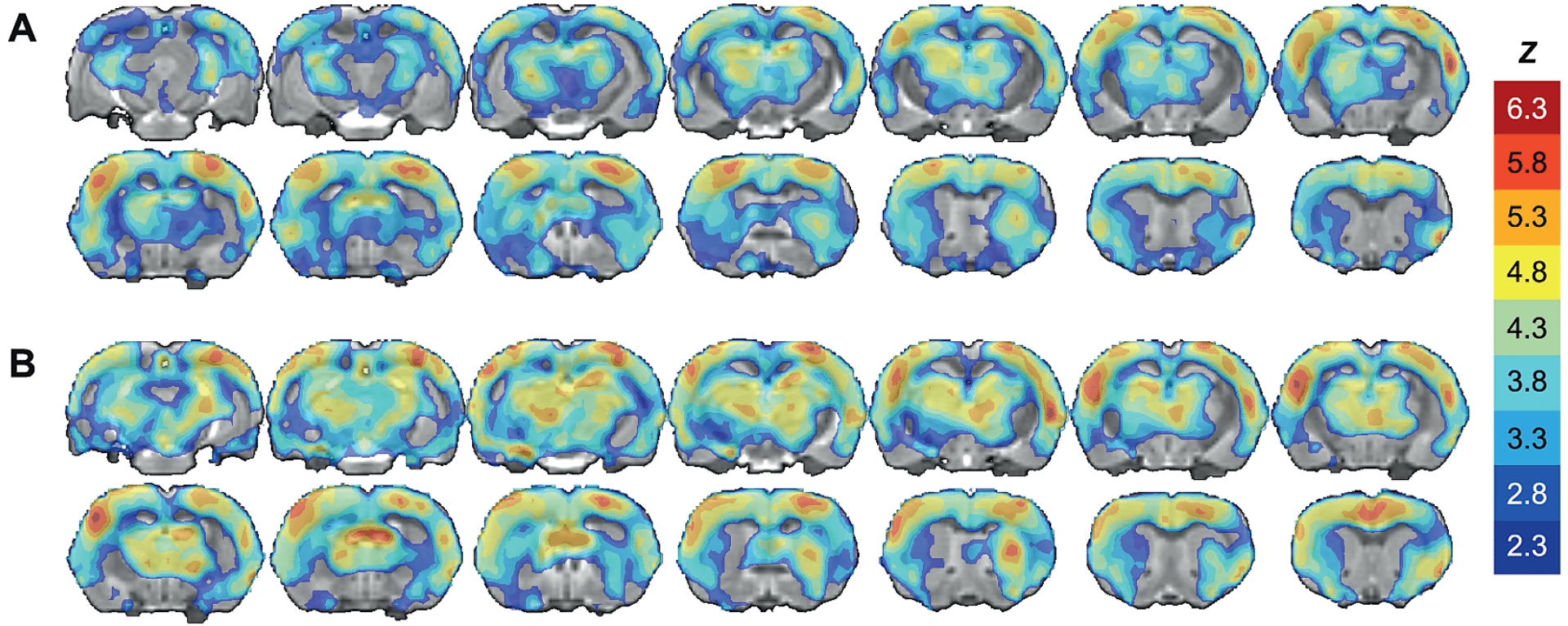




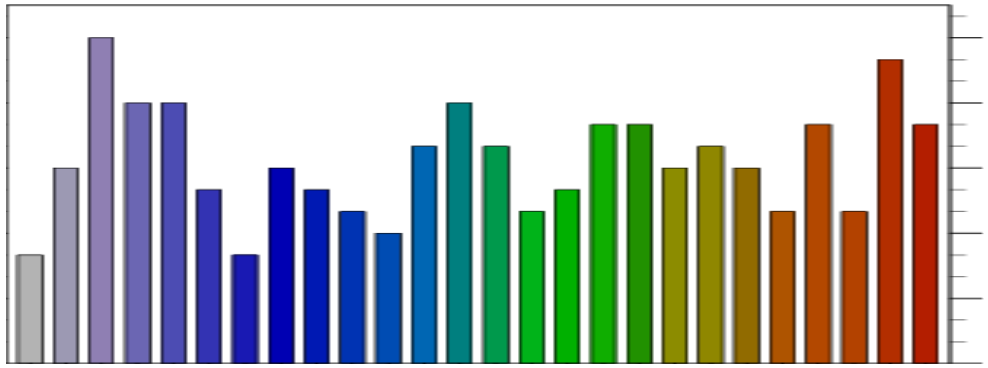
Information Processing

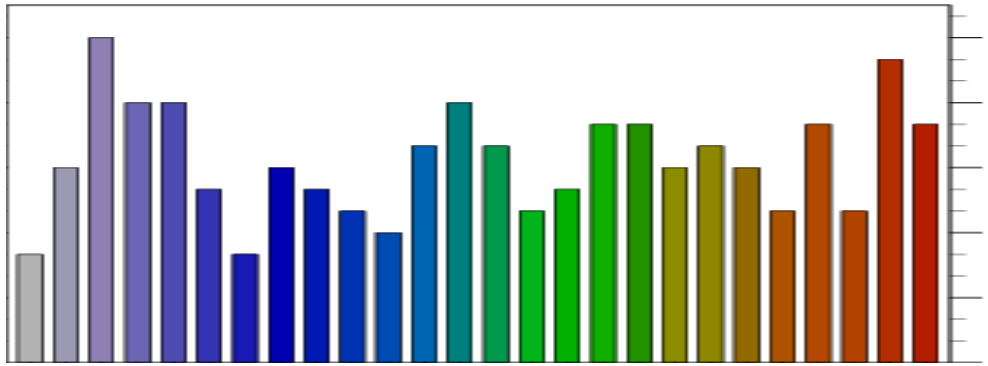
- Analyze how human brain processes visual information.
- Use modern techniques such as FMRI, MEG etc.

Human brain is a feature space

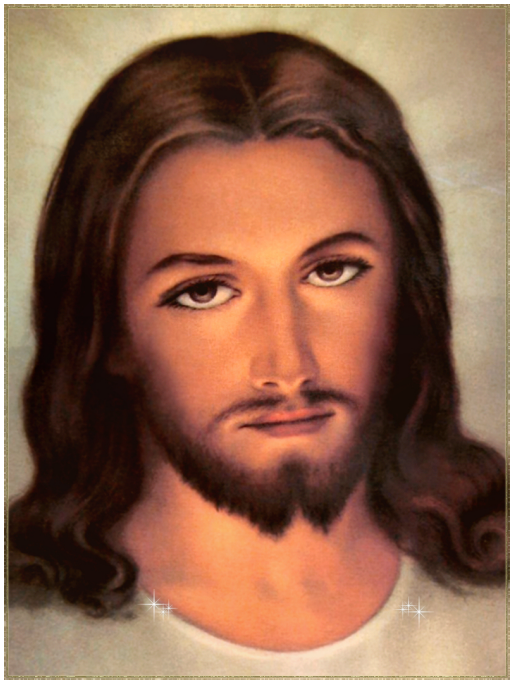


Analyze the feature space



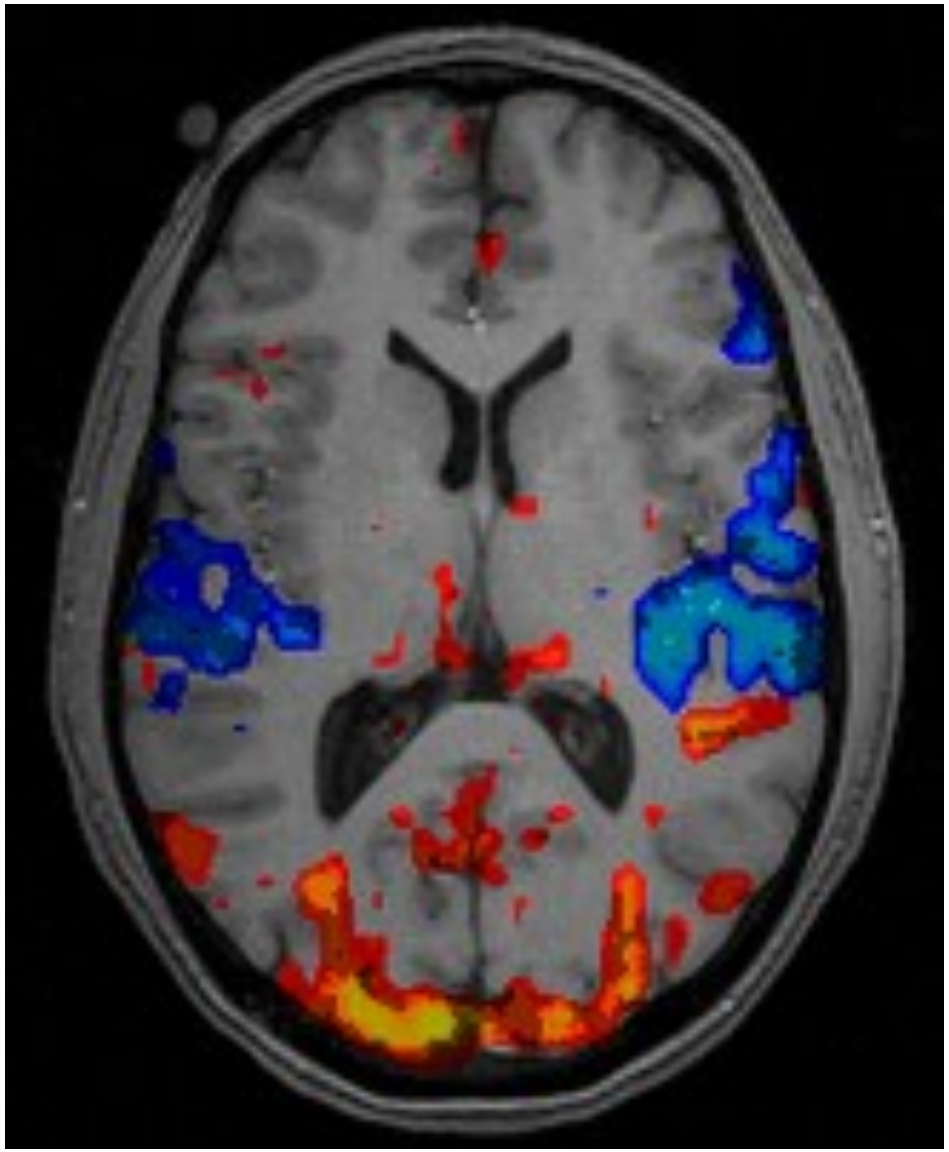


Goal 3: Explore opportunities with the new type of data….joint neuroscience + computer vision papers



Same Concept

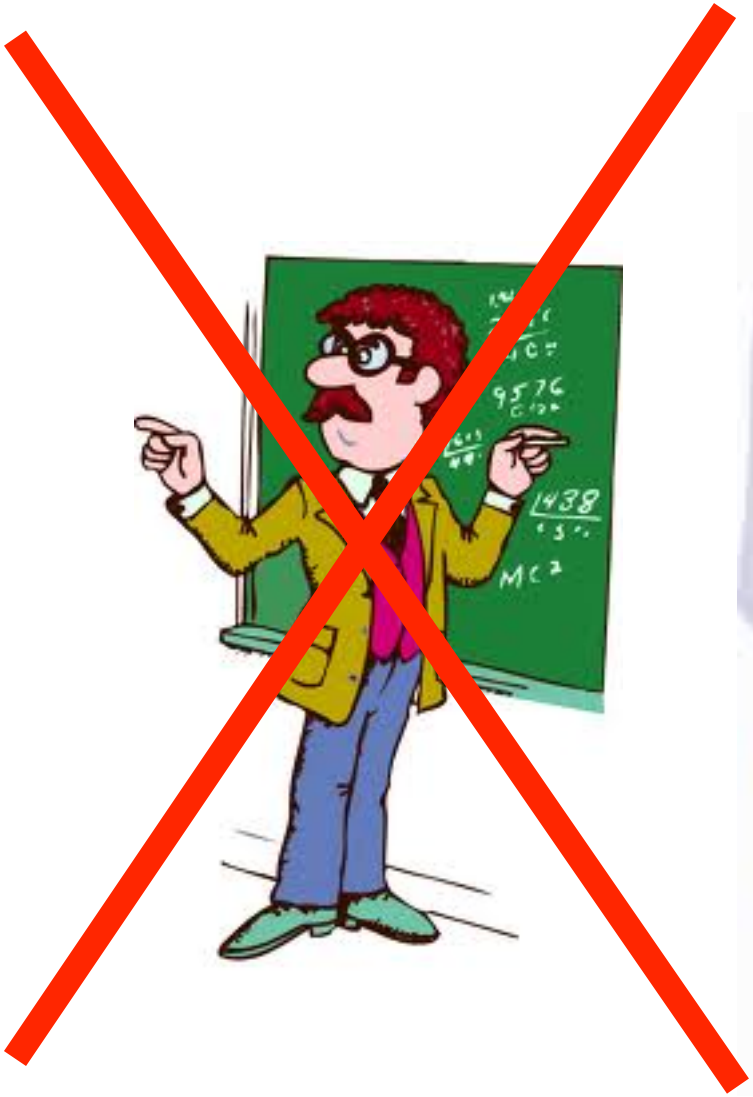
Human Labeling – Mechanical Turk



Use the Brain Signals as labels!

Hidden Goal: I want to learn more about neuroscience and how cool it is!

You guys are the group that will help me learn about it.


1438
$E=mc^2$

Course Outline

Design your Own Course

- First time being offered in this avatar – will be taught once in 3-4 years…
- Opportunity to design the course…and affect future versions.
- What problems plague computer vision..
  - Should I use semantic or functional categories etc.
- Feedback throughout the course appreciated. No feedback.

Lower-Level Goals

- Read some interesting papers together
  - Learn something new: both you and us!
- Get insight as to what does human brain say about some critical computer vision questions
  - ability to understand lots of neuroscience papers!
- Learn how to speak/discuss
- Learn to develop relation between human vision and computer vision
- Learn how think critically about papers

Next Gen Philosophers. Maybe?

Course Organization

- Requirements:
  1. Class Participation (35%)
     - Post on the Class Blog before **each class**
     - Ask questions / debate / flight / be involved!
  2. Presentation + Critic (25 %)
  3. Project (40%)

Class Participation

- Keep annotated bibliography of papers you read (always a good idea!). The format is up to you. At least, it needs to have:
  - Summary of key points
  - A few Interesting insights, "aha moments", keen observations, etc.
  - Weaknesses of approach. Unanswered questions. Areas of further investigation, improvement.
- Before **each class**:

  Submit a comment on the Class Blog
  - ask a question, answer a question, post your thoughts,praise, criticism, start a discussion, etc.

Presentation + Critic

1. Pick a topic from the list

2. Understand it as if you were the author

3. Prepare an amazing 45min presentation
   - Discuss with me twice – 2-weeks before the presentation, 5 days before the presentation

4. For every topic, there will be critics as well. These students will review and post the critical review on the blog.
   - Not the starting point of discussion though

Class Project

An Interesting 2-step Class Project !

- Release a 50-category scene dataset. Images from categories like beach, office etc..

- Each category has 15 images: 10 training, 2 validation, 3 test images.

Class Project

- 500 training, 100 validation and 150 test images.
- 100 validation images also have FMRI signals associated with them.
- Goal 1: Find a feature space used in computer vision which is similar to human brain.

Project

- Evaluation based on 100x100 similarity matrix.
- Closest Matrix…gets an AWARD!
- Goal 2: use 100 validation images to improve your classifier (sharing, active-learning etc.)
- Test on 150 test images.

End of Semester Awards

- We will vote for:
  - Best 2nd Project – CVPR paper submission
  - Best Presentation
  - Best Discussions

Logistics

- People not sure – talk to us…
- Sign-up to begin in 2 weeks…
- Mid-term Class Presentations and End-term Class Presentations…


FUN
AHEAD