CS5100 Foundations of Artificial Intelligence

Module 08 Lesson 13

Applications: Perception & Vision,   Autonomous Cars

Some images and slides are used from CS188 UC Berkeley/AIMA with permission
All materials available at http://ai.berkely.edu / http://aima.cs.berkeley.edu

Overview

Visual Perception

Image Formation

Image Processing

Face Detection & Object Recognition

Reconstructing 3D

Autonomous Driving

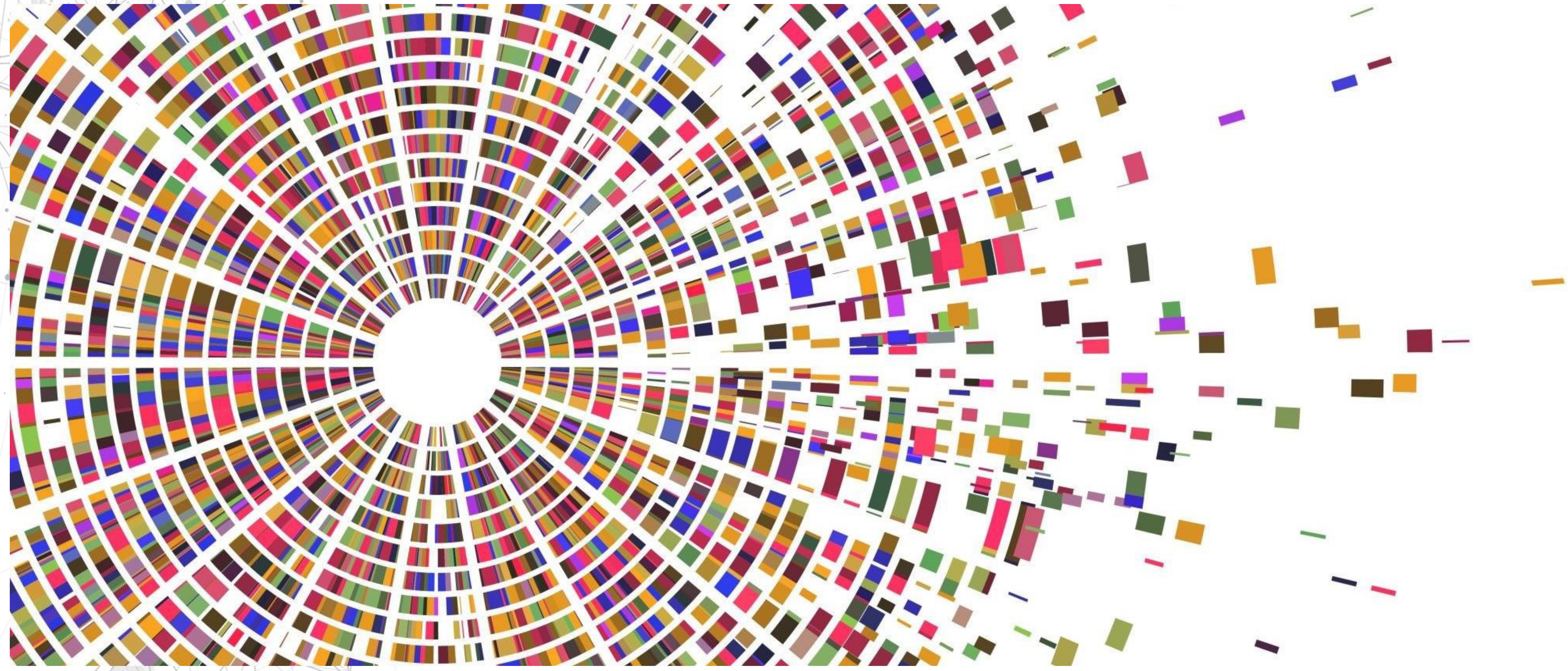
Applications
Vision

- Scene and object recognition
- Special Effects in movies
- Urban modeling
- Face/smile detection (security, better cameras)
- Biometrics
- Medical image interpretation
- Optical character recognition
- Visual search
- Automotive safety, Autonomous cars
- Augmented/Virtual Reality
(e.g. in Oculus Quest 2)
- Computational Photography
- Vision in shopping
- Vision in games (e.g. Kinect)
- Vision in robotics, space & underwater exploration …

David Hubel & Torsten Wiesel

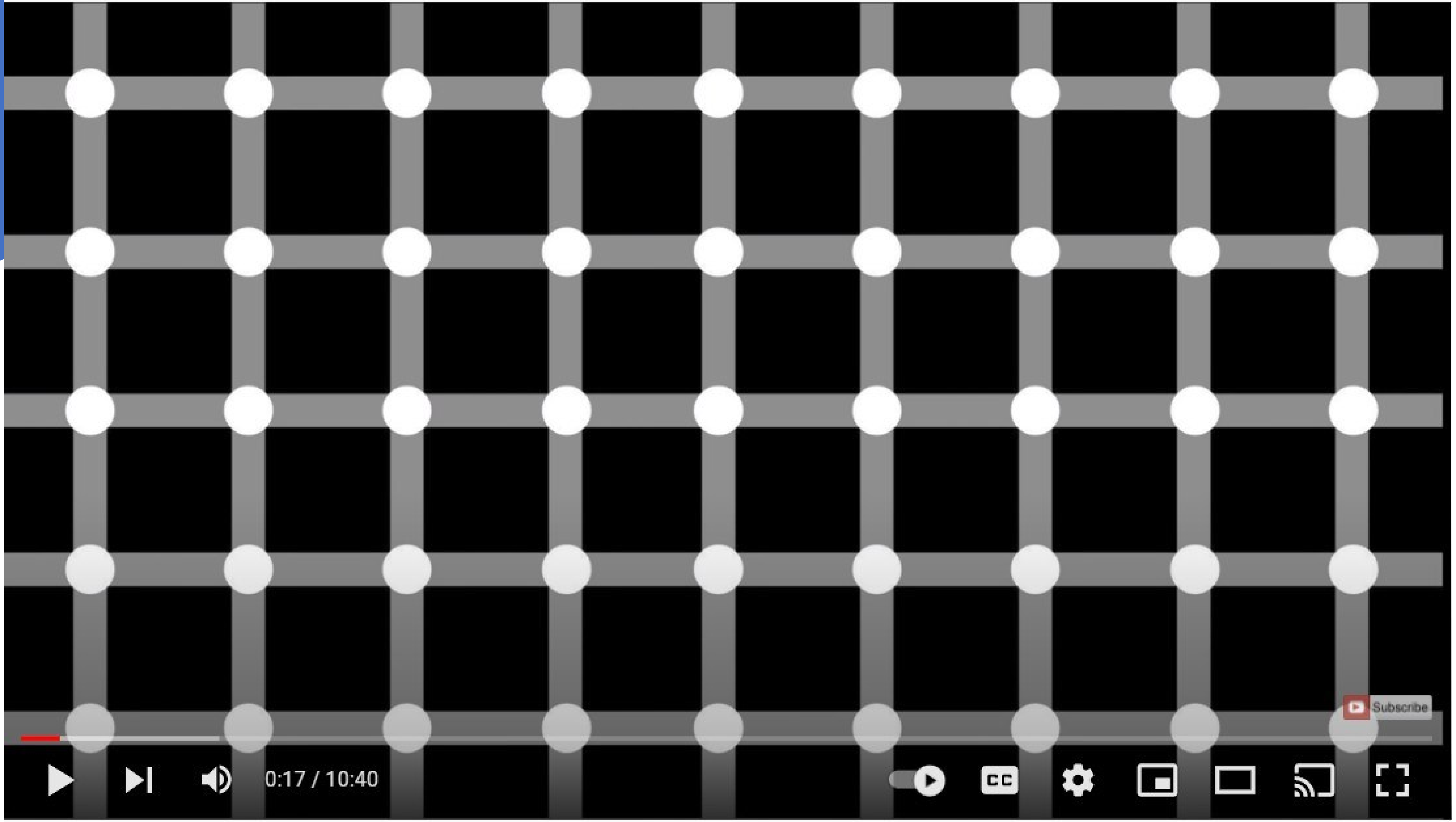
- Important impact on our understanding of visual perception
- Poked electrodes into the visual cortex of experimental animals
- Studied the activity in the animal's visual cortex while presenting various stimuli to its eyes.
- Development of the visual system,
  visual neurophysiology
  (including edge, orientation & motion detectors, color & stereoscopic depth detectors, complex representations)
- 1981 Nobel Prize in Medicine

Visual Perception

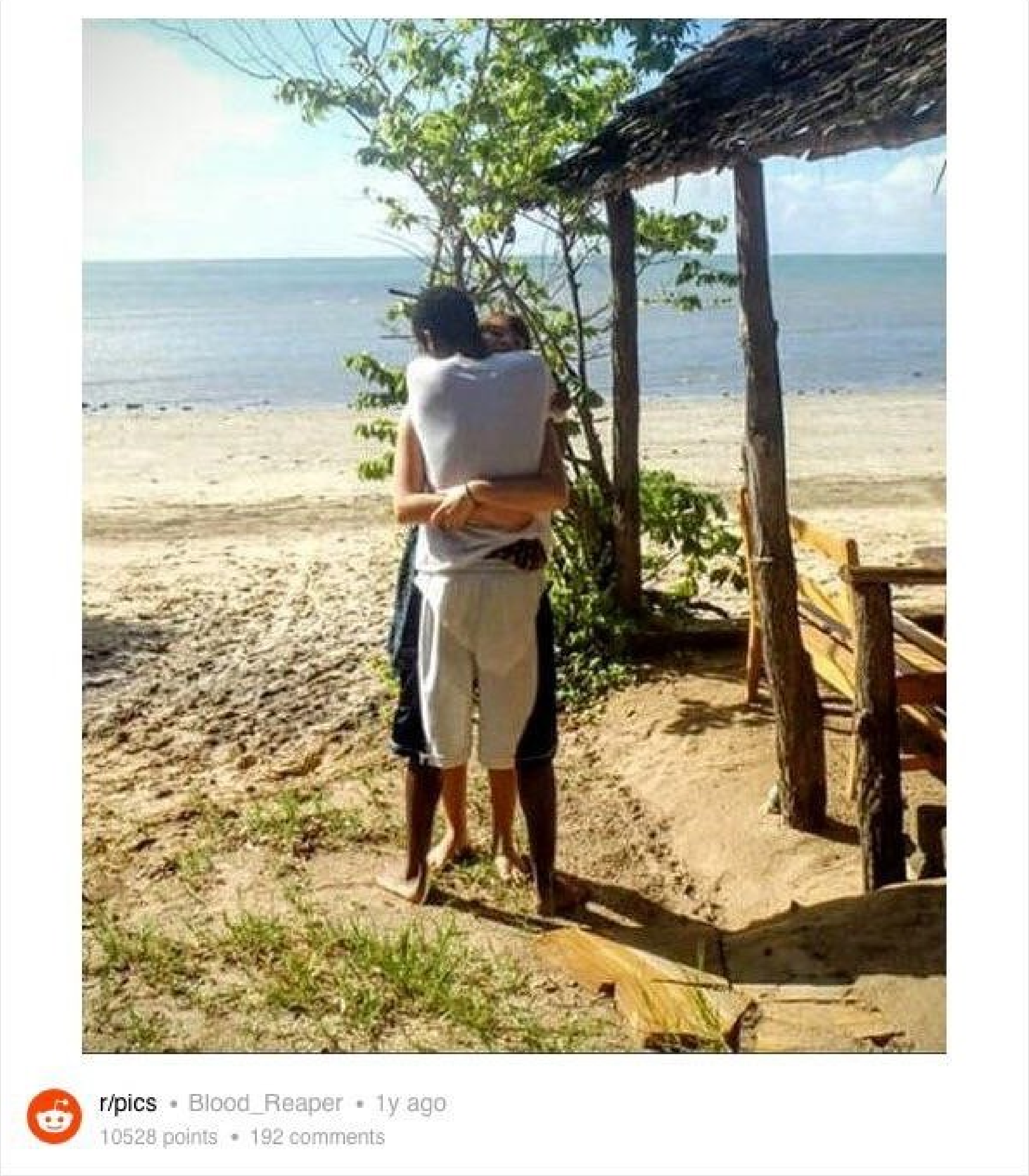


Visual Illusions Eye-brain link

https://www.youtube.com/watch?v=_t2ePlwTeBQ

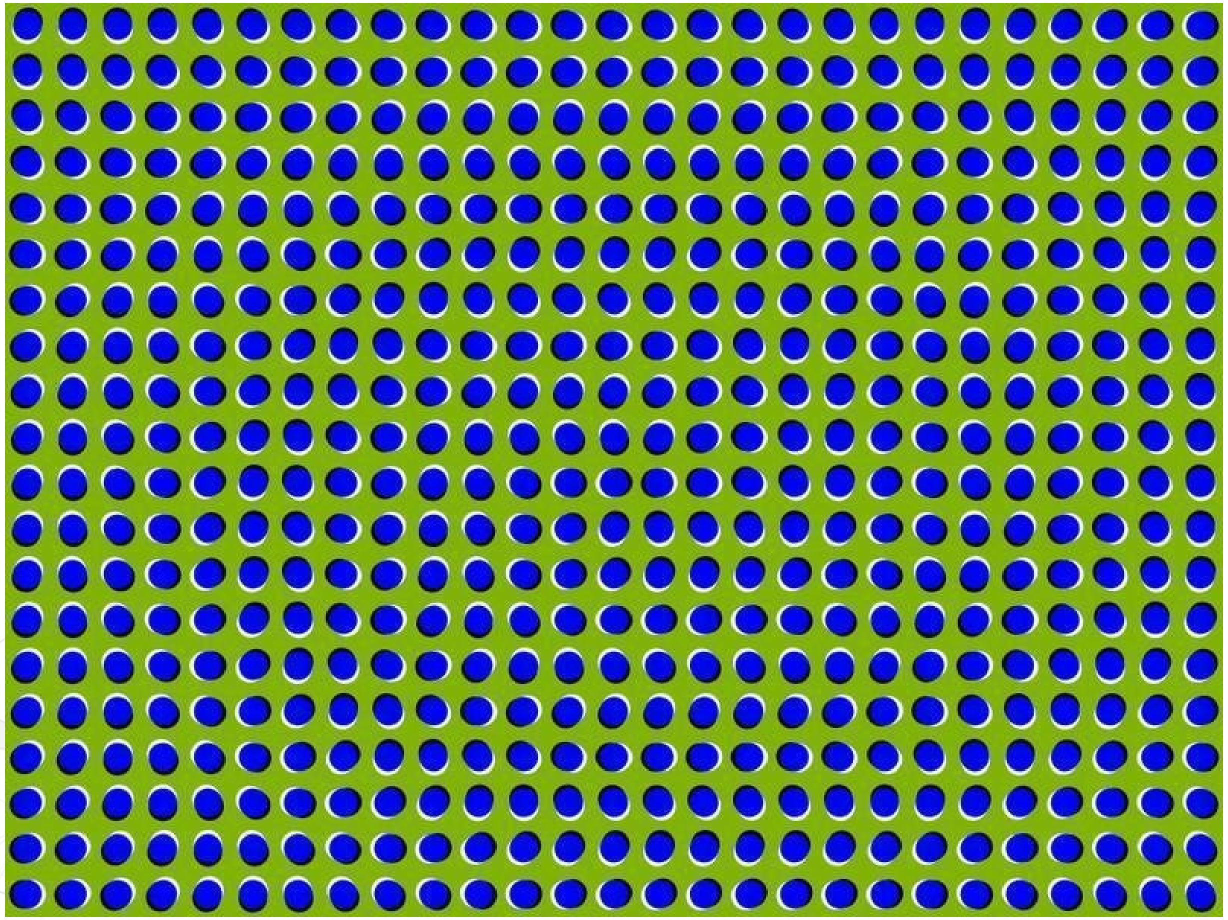


Optical Illusions



Optical Illusions

www.pcmag.com/news/21-optical-illusions-that-prove-your-brain-sucks



Perception

- Provides agents with info about the world by interpreting response of sensors
  - Sensors: switches, eyes, cameras…
  - Sensing may be passive (just get data) or
    active (send out signals, and sense response to them: e.g. radar, sonar)

- Focus on visual perception here, plus a little on other sensors for autonomous cars

Vision Models

Object Model

- Objects in the real world: what objects "are"
- May include constraints from real world or actual 3D models

Rendering Model

- Describes processes that produce stimulus from world
- What objects "look like"
  - E.g., nearer object looks bigger than object further away
  - May be ambiguous even when accurate

Approaches to Modeling Vision

Visual data very rich: which to consider, which to reject

Feature Extraction

- Simple computations on sensor observations e.g. birds & collision detection
- Like reflex actions

Recognition

- Distinguish between objects encountered, using visual and other cues
- Recognize food/not-food, recognize faces etc., and respond to these

Reconstruction

- Build a geometric model from one or more images

Image Formation

Images

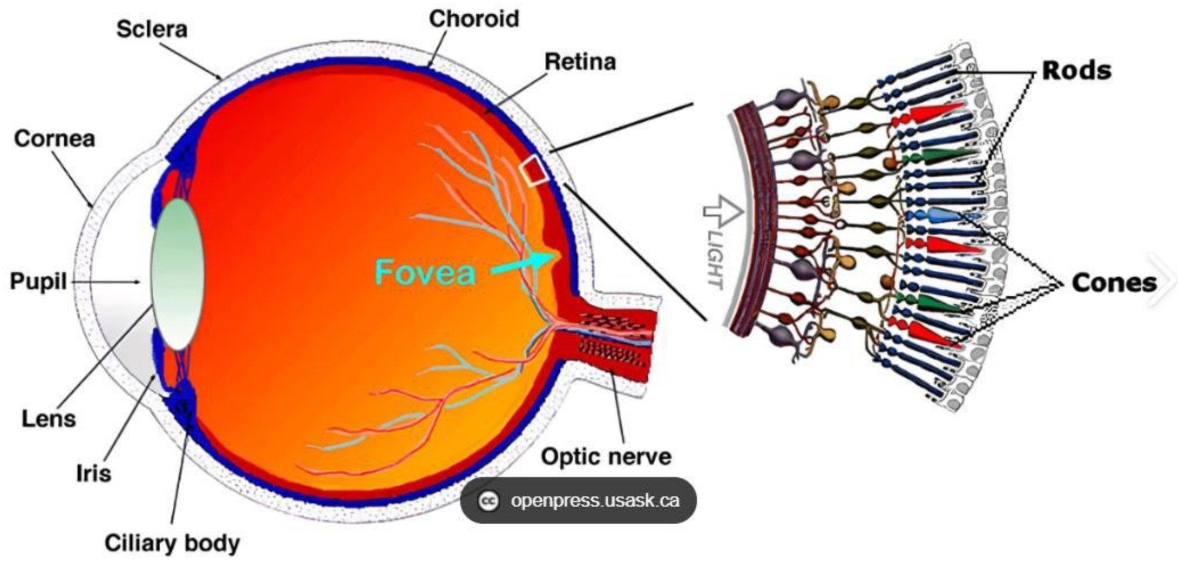
Image **sensors** gather light scattered from objects in a **scene** and create an **image**

Eye:

- About 100 million rods, sensitive to a range of wavelengths, *plus*
- About 5 million cones, for color vision, 3 main types each sensitive to a different set of wavelengths

Camera:

- Image formed on image plane
  - Photosensitive film OR
  - Photosensitive pixels (CMOS/CCD/…)



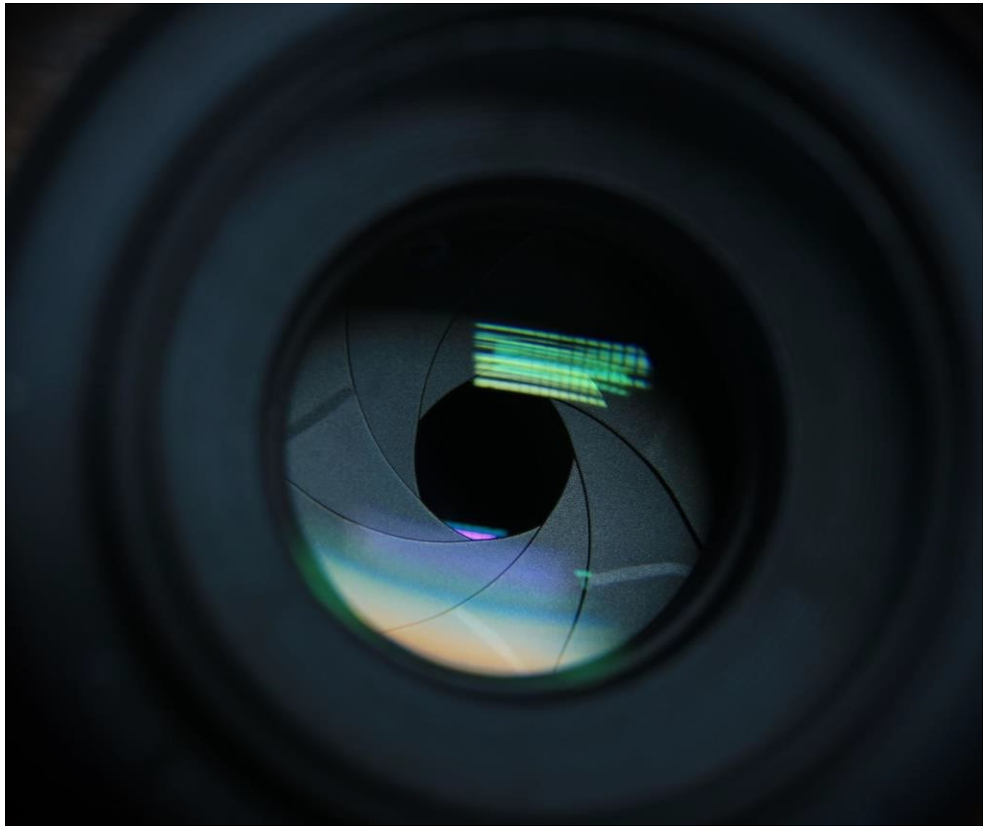
“Seeing”

How does a human eye “see”?

How is this different from how a camera sees?

What are Focus, and Depth of focus ?

Start with a pin-hole camera model



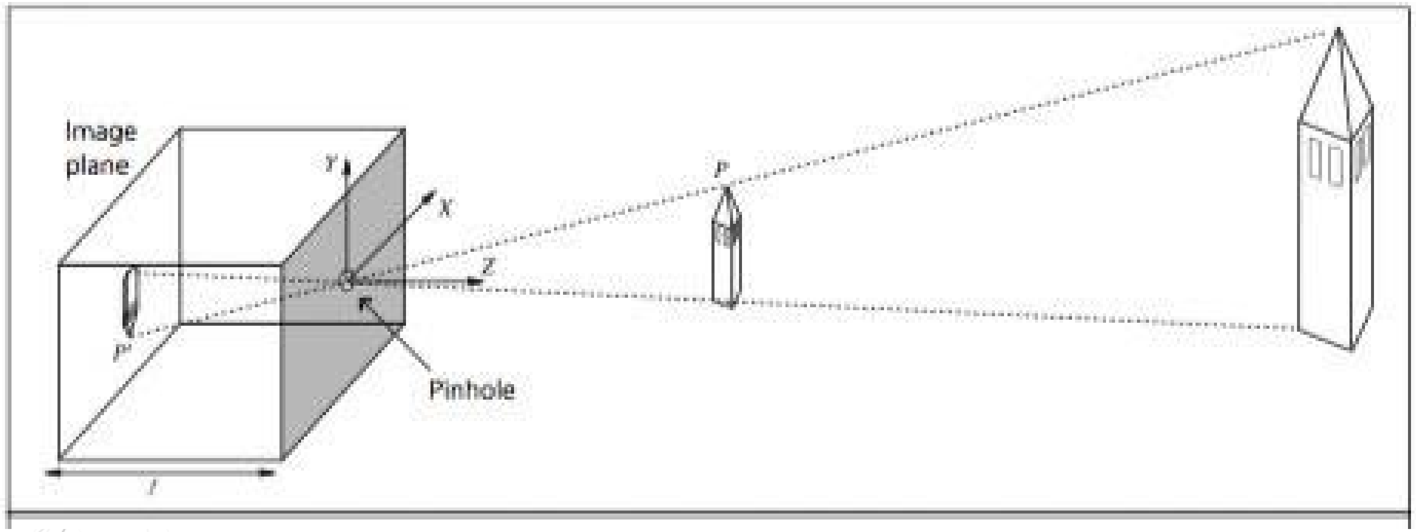
Pin-hole Camera Model

Image inverted (left-right and up-down)

$-x/f = X/Z$, $-y/f = Y/Z$ $\quad$ $x = -fX/Z$, $\quad y = -fY/Z$

Perspective projection

- Distant objects look small,
- Parallel lines meet (at the vanishing point)

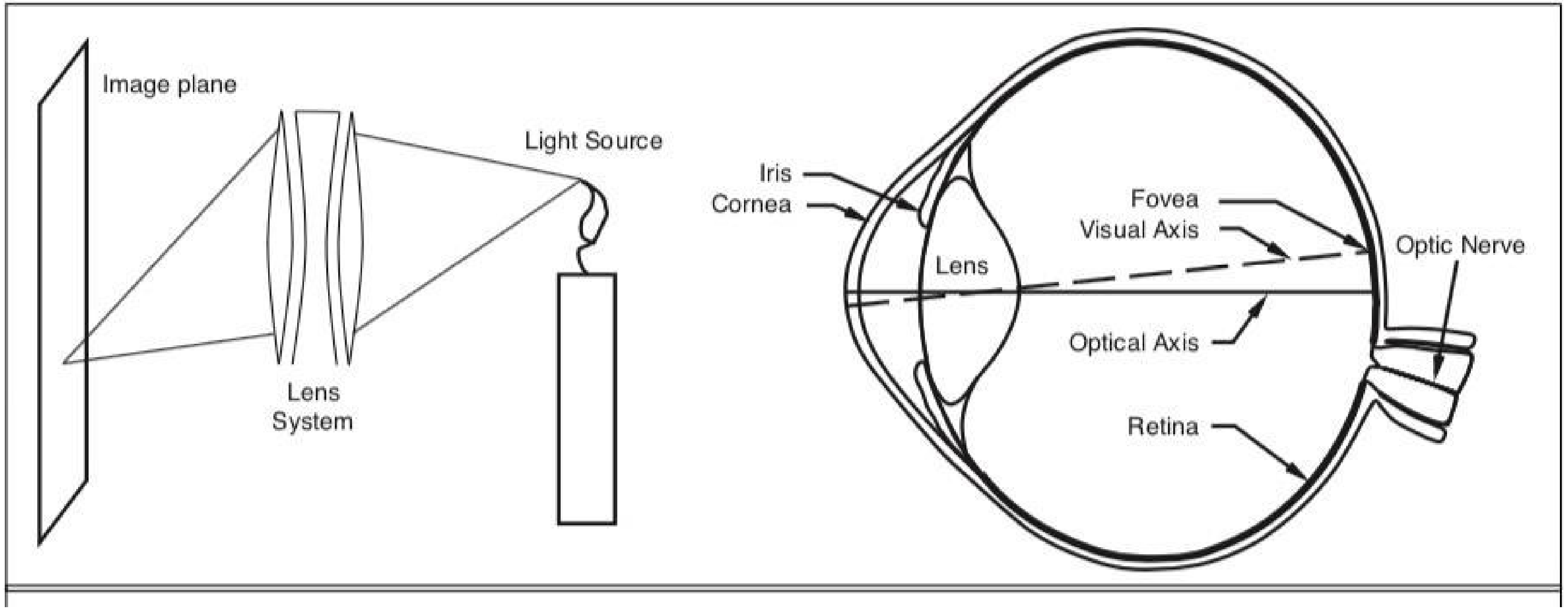


Lens Systems.. 1

Lens systems gather light and keep image in focus

- From points within a specific range around focal plane
  - Depth of Field

Image 'straightened' in brain



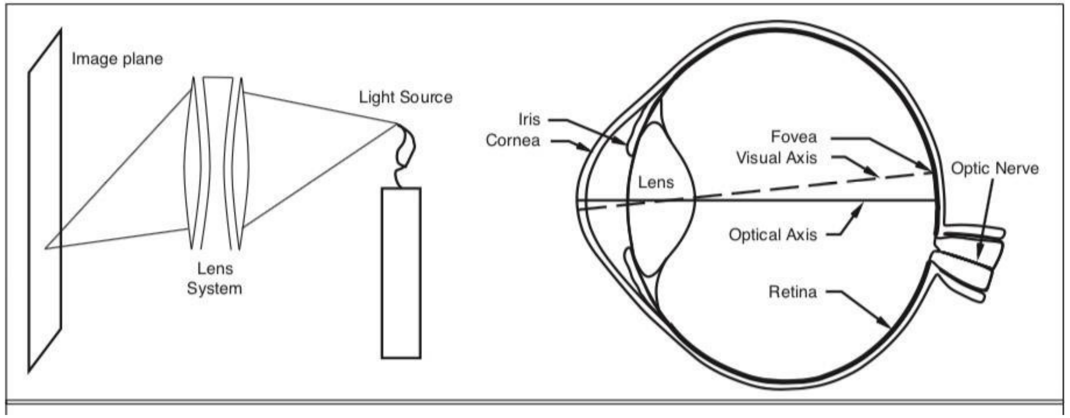
Lens Systems ..2

To move focal plane:

- Camera lenses can be moved back and forth, Eye lens can change shape, more difficult in old age

To change depth of field, adapt to brightness:

- Change camera aperture (opening)
- Change iris opening in eye

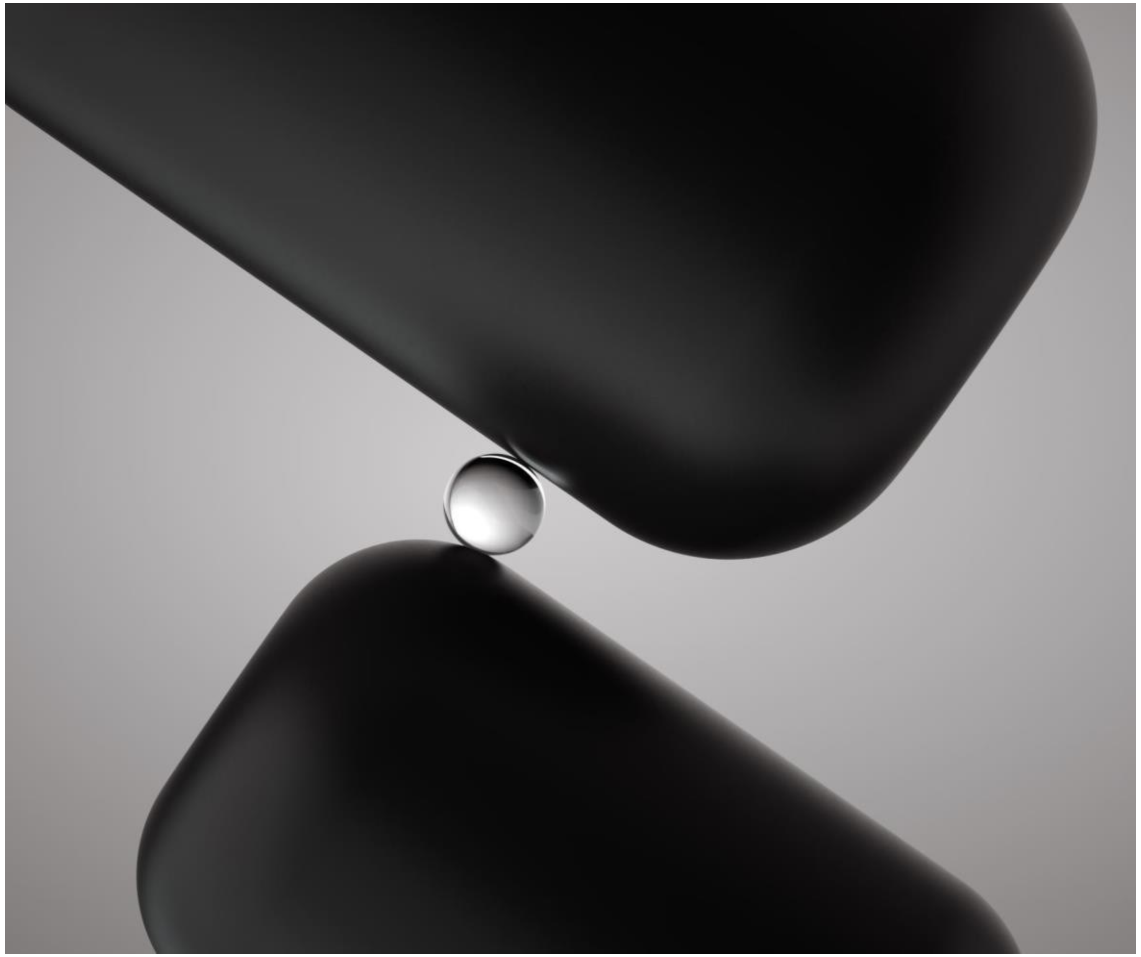


Light & Shading ..1

Image brightness is a strong cue to object shape & identity

Three main causes of varying brightness

- **Overall intensity** of light
- Different **reflectivity** of points on object (e.g. ring)
- **Shading**



Light & Shading ..2

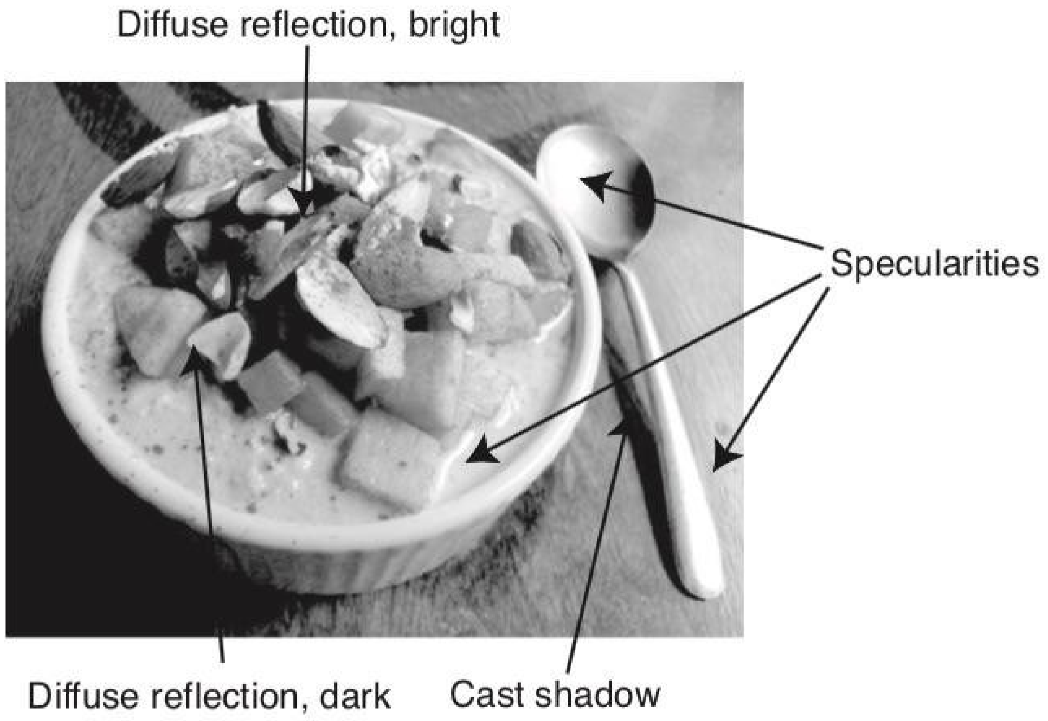
Diffuse Reflection

- From cloth, rough wood, plants, rough stone…

Specular Reflection

- Polished metal, polished stone, plastic, oily or wet surfaces

Most surfaces:
diffuse with speculiarities (reflecting regions)



Light & Shading ..3

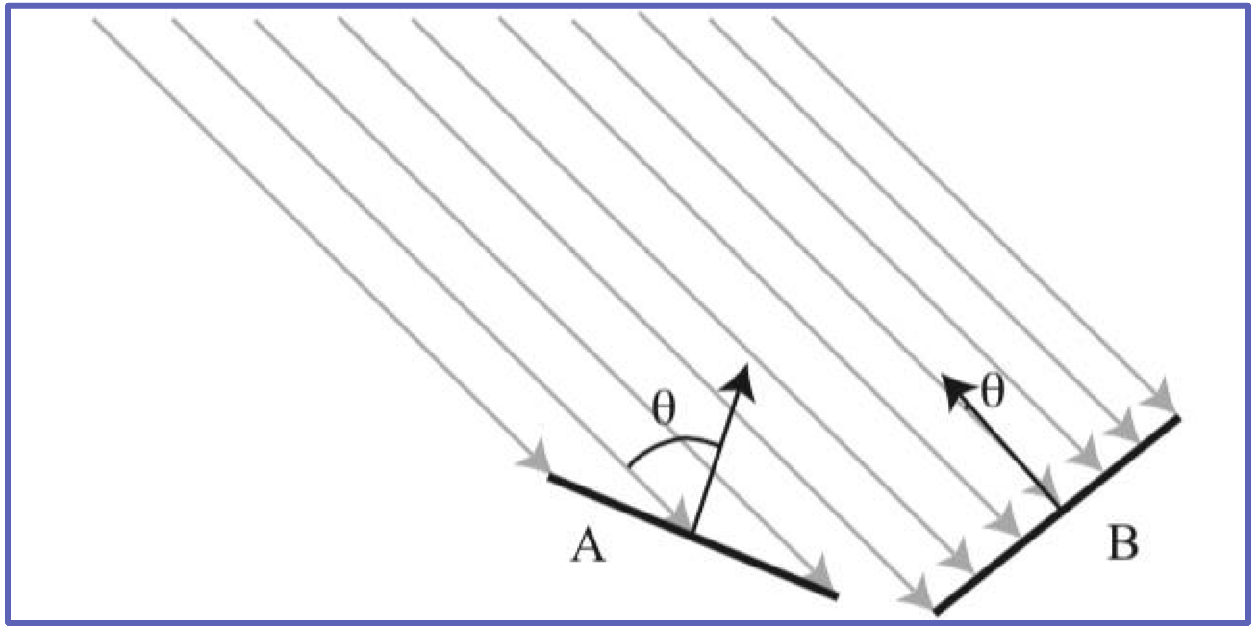
Sun: main source of illumination outdoors

- **Distant point light source**

Usually, multiple sources indoors

Amount of light collected by surface: depends on angle between illumination direction & normal to surface

Shadows give clues to object shapes



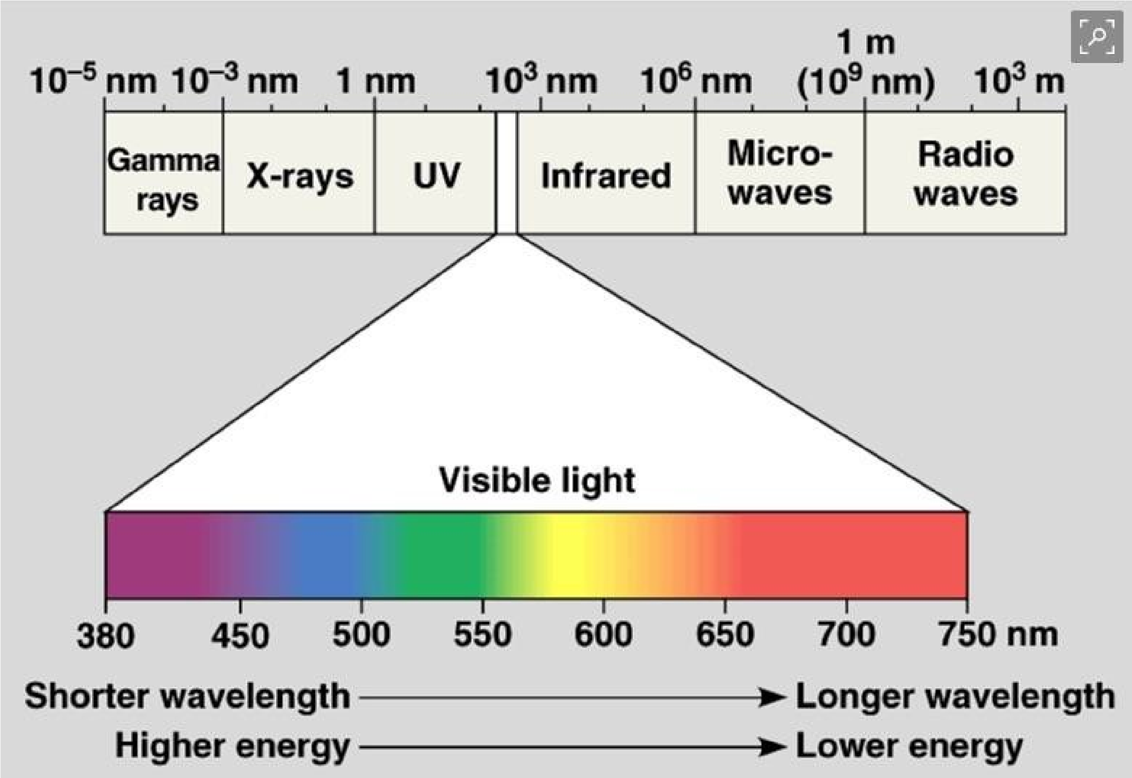
Color

Human eyes best for wavelengths 380-750nm

- 3 different types of Cones with peak sensitivities at: 420nm, 540nm, 570 nm

Principle of trichromacy

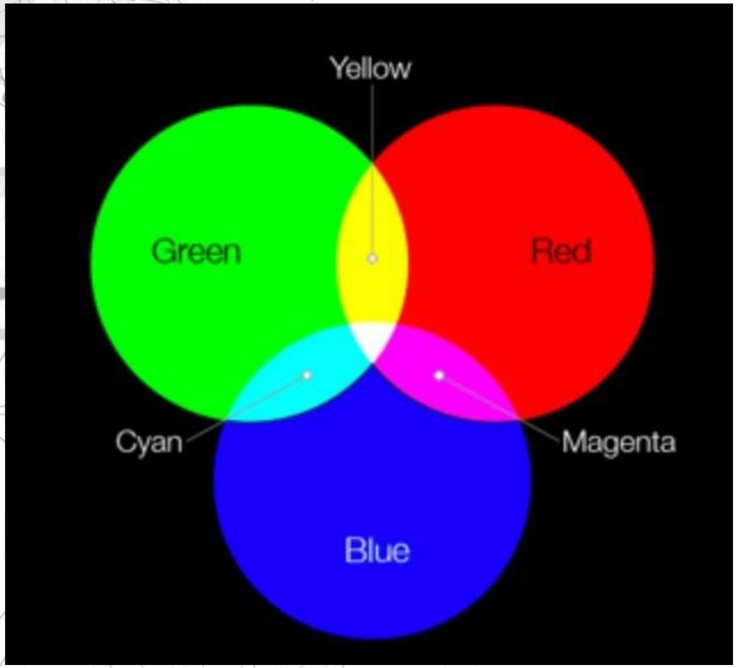
- For any ‘spectral energy density’, can construct another spectral energy density from a mix of 3 colors (usually R, G & B) that is indistinguishable from the original color



Pigment colors vs. Light colors

Additive (Light) Color Model – RGB

Color transmitted through transparent media
Use: In TV/Computer monitors,
digital presentations, …
All colors together: white; absence of light = black



Subtractive (Pigment) Color Model – CYMK

Color absorbed by/reflected off of media
Use: paints, printers, …
All colors together: near black; K: black

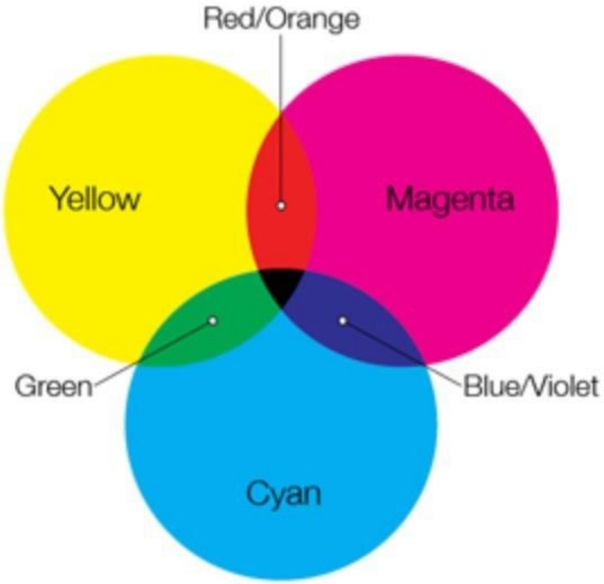


Image Processing

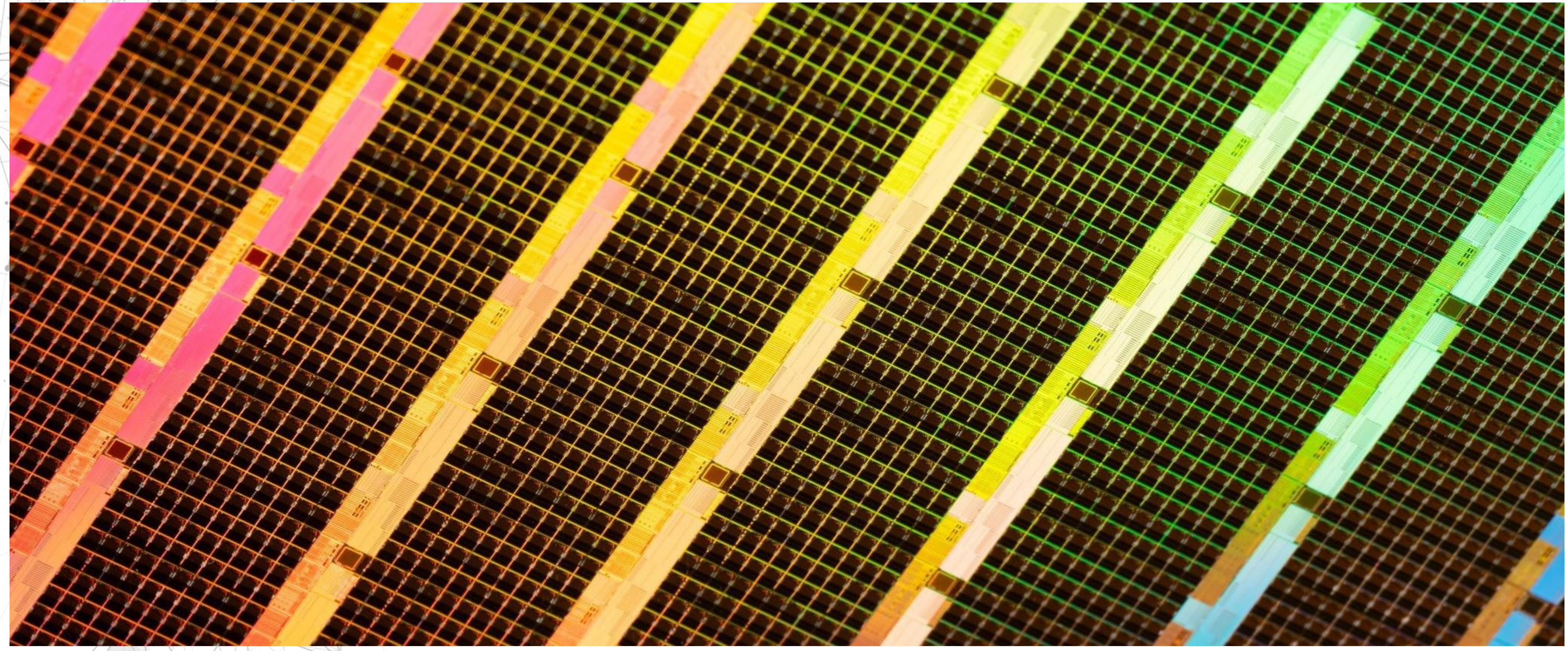
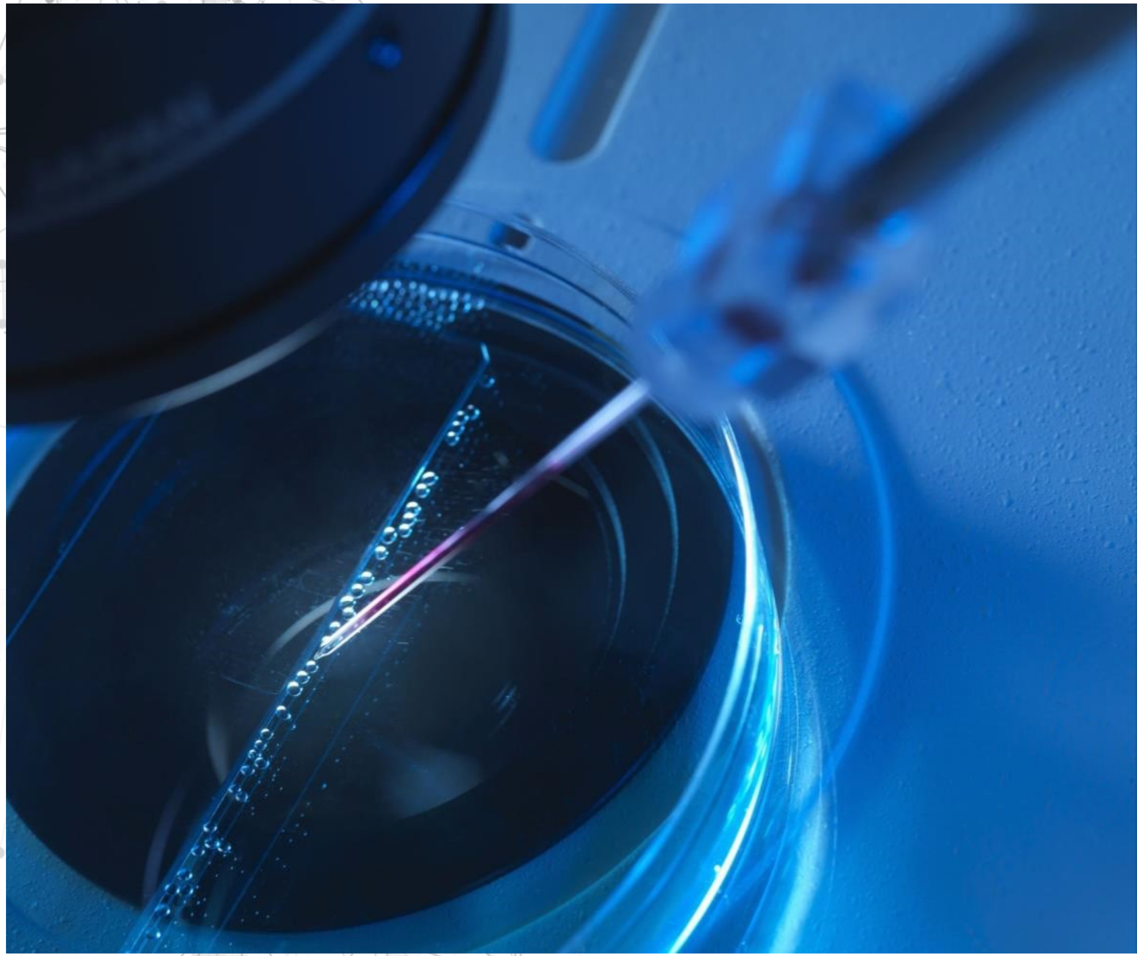


Image Processing Operations



Early/Low-level Operations

- Edge Detection
- Texture Analysis
- Optical Flow Computation

'Local': can operate on one part of image

➔ Parallelizable, e.g. in eyes or a GPU

Mid-level Operation

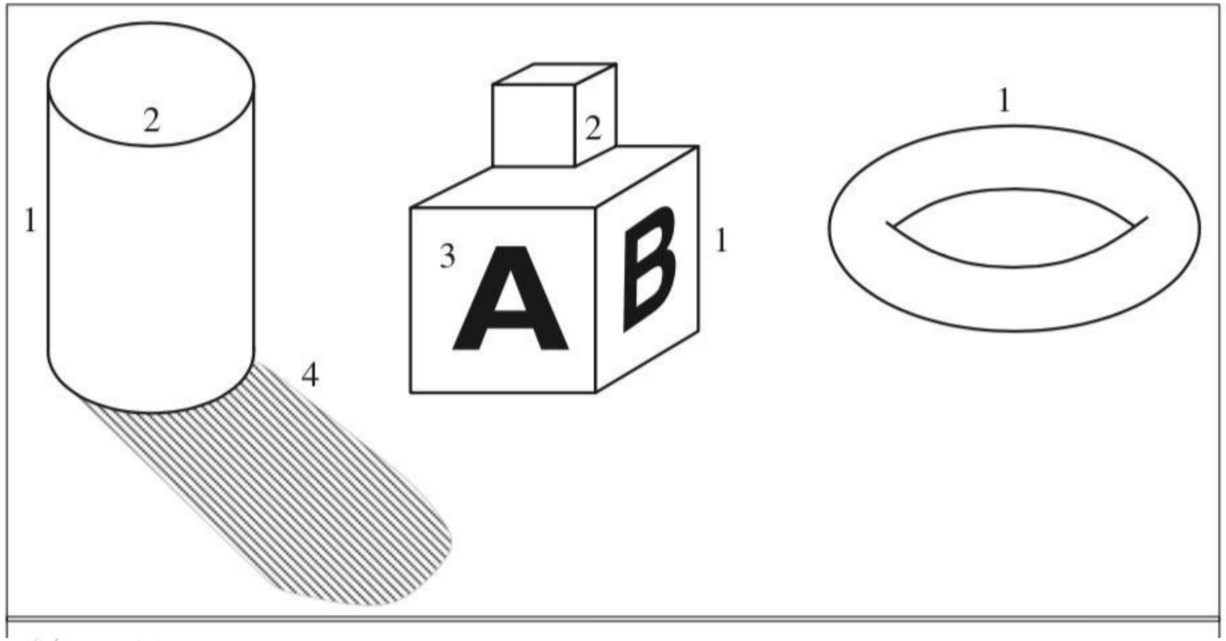
- Image Segmentation

Edge Detection ..1

Lines along significant change of brightness

Abstraction:

- detailed image ➔ more compact representation



Discontinuities:
1. depth
2. Surface orientation
3. Reflectance
4. illumination

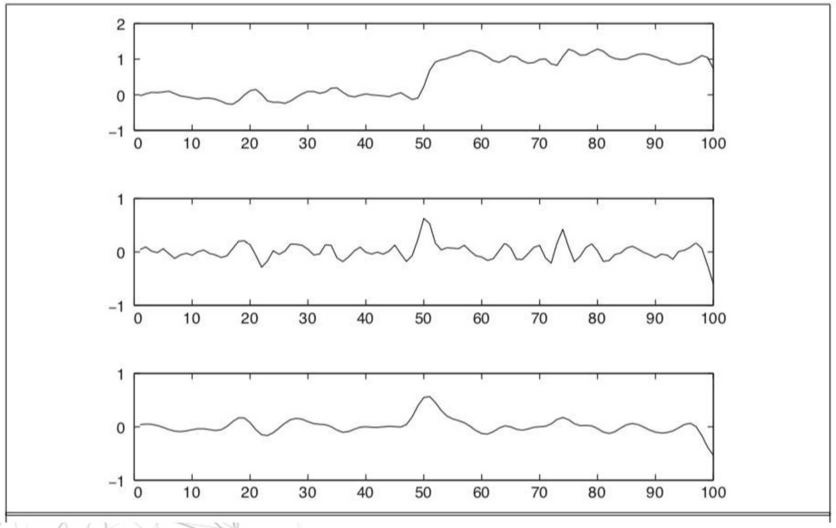
Edge Detection ..2

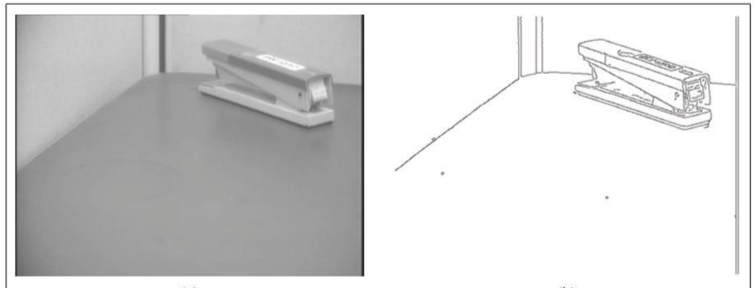
Look for brightness change

Differentiate brightness curve, look for large magnitude changes

Even better after smoothing brightness graph

Deal with noise




(a) (b)

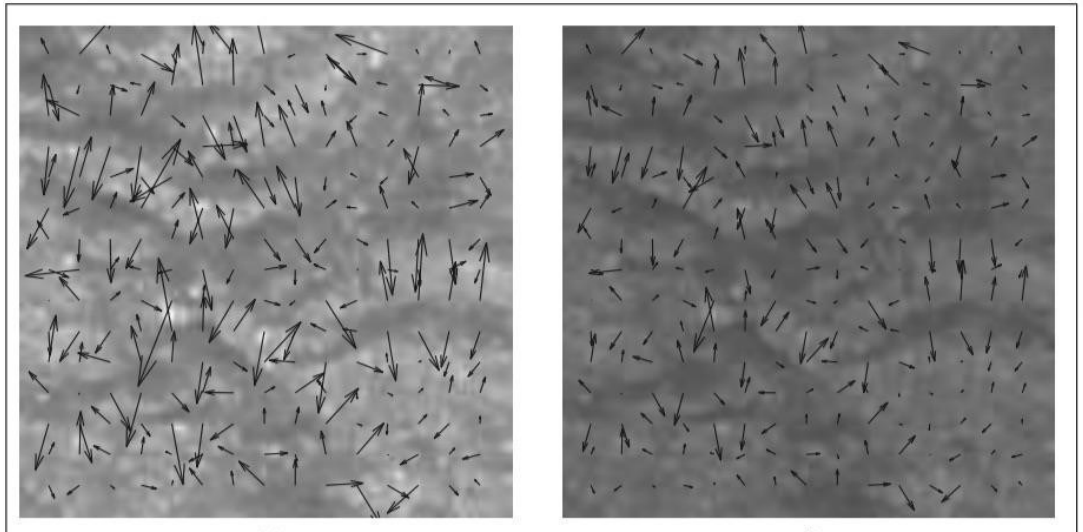
Texture

Visual feel of surface

- Spatially repeating pattern on surface, sensed visually
- e.g. Stiches on sweater, people in a stadium, grass patch

Largely invariant to lighting

Can look for differences in texture



(a) (b)

Optical Flow

Apparent motion in video, when an object is moving, or if camera is moving

Shows speed & direction of motion of image features

- Provides info about scene, e.g., because further points move slower etc.
- Lets us recognize actions

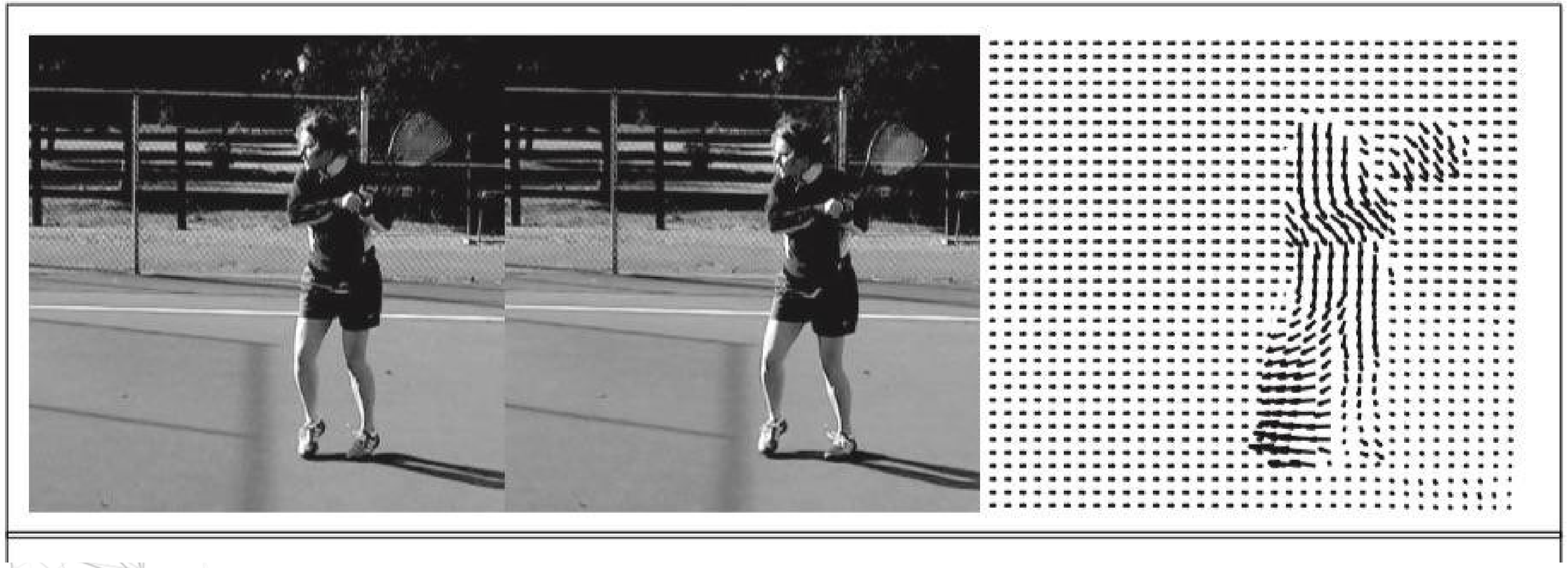


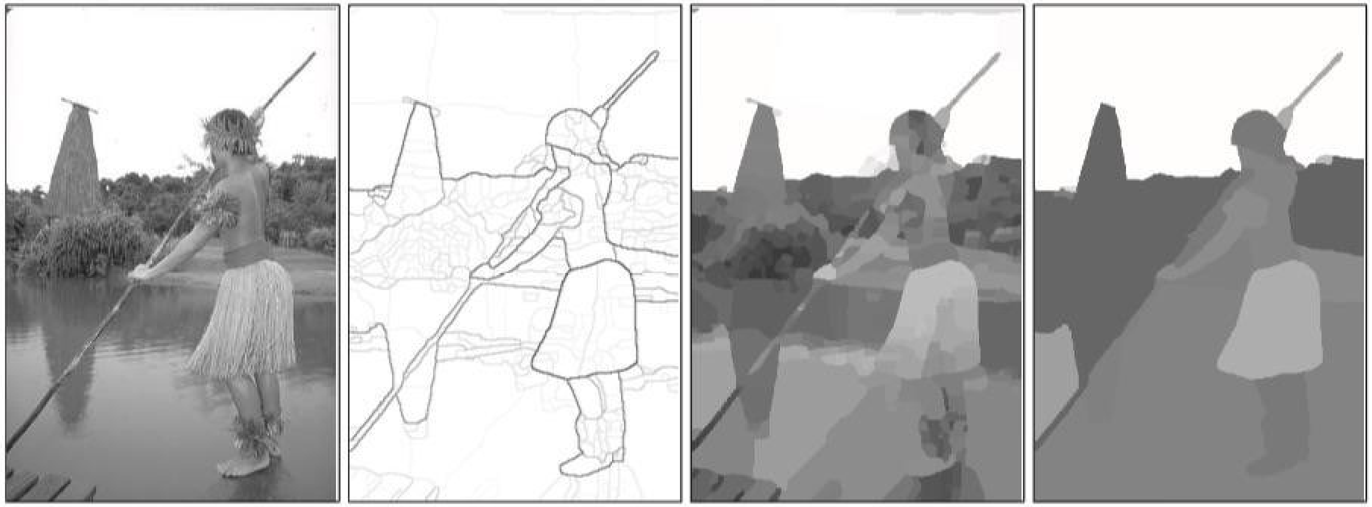
Image Segmentation

Splitting image into regions of similar pixels

- Similar in brightness, color, texture

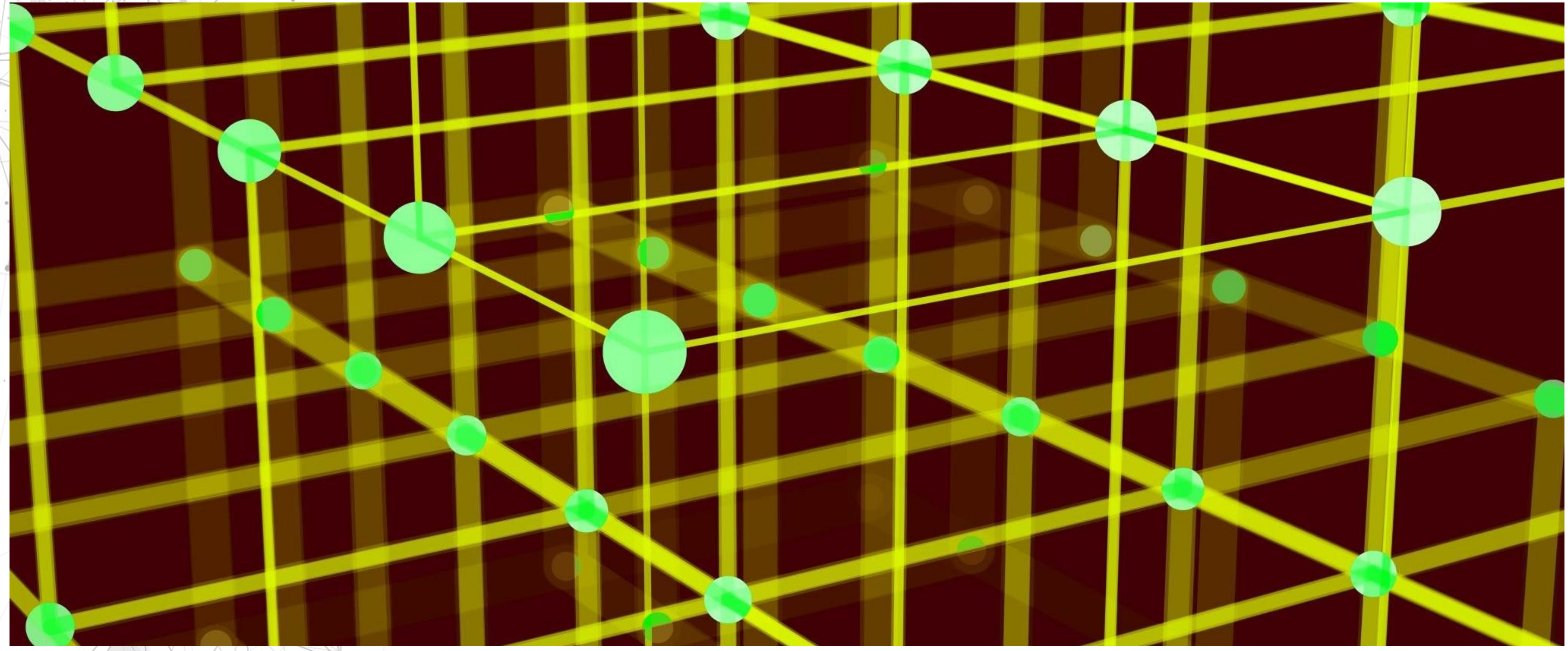
Two approaches:

- Detect boundaries
- Detect different regions directly



(a) (b) (c) (d)

Object Recognition

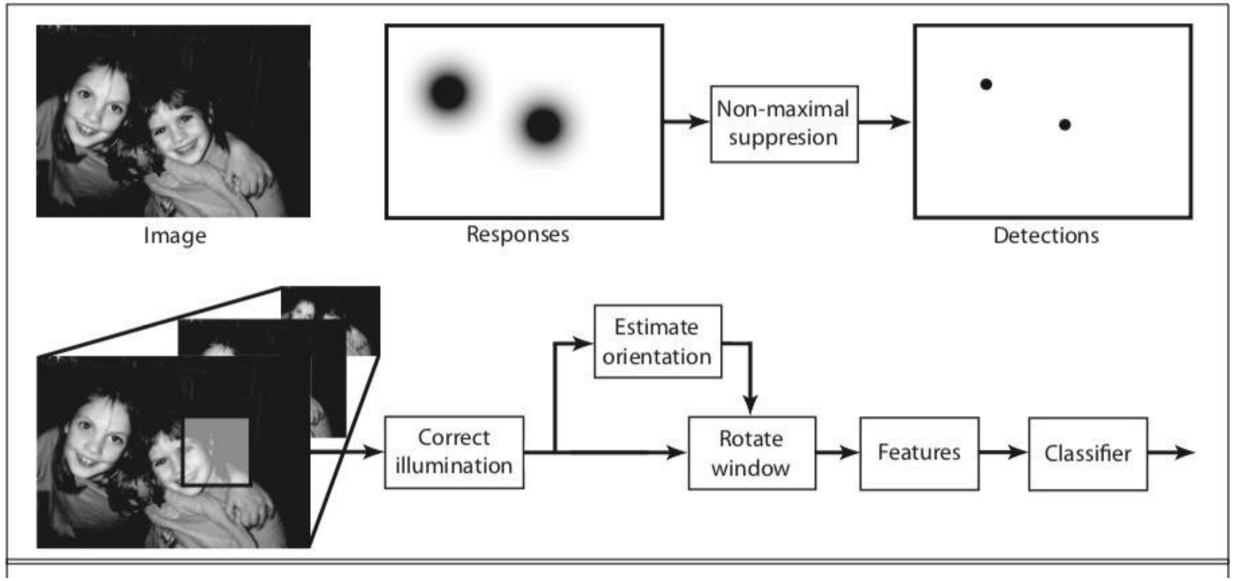


Face Detection

Look for face-like segments, correct illumination, rotate images, extract features, send through classifier

Look for faces of different sizes

Look for multiple faces

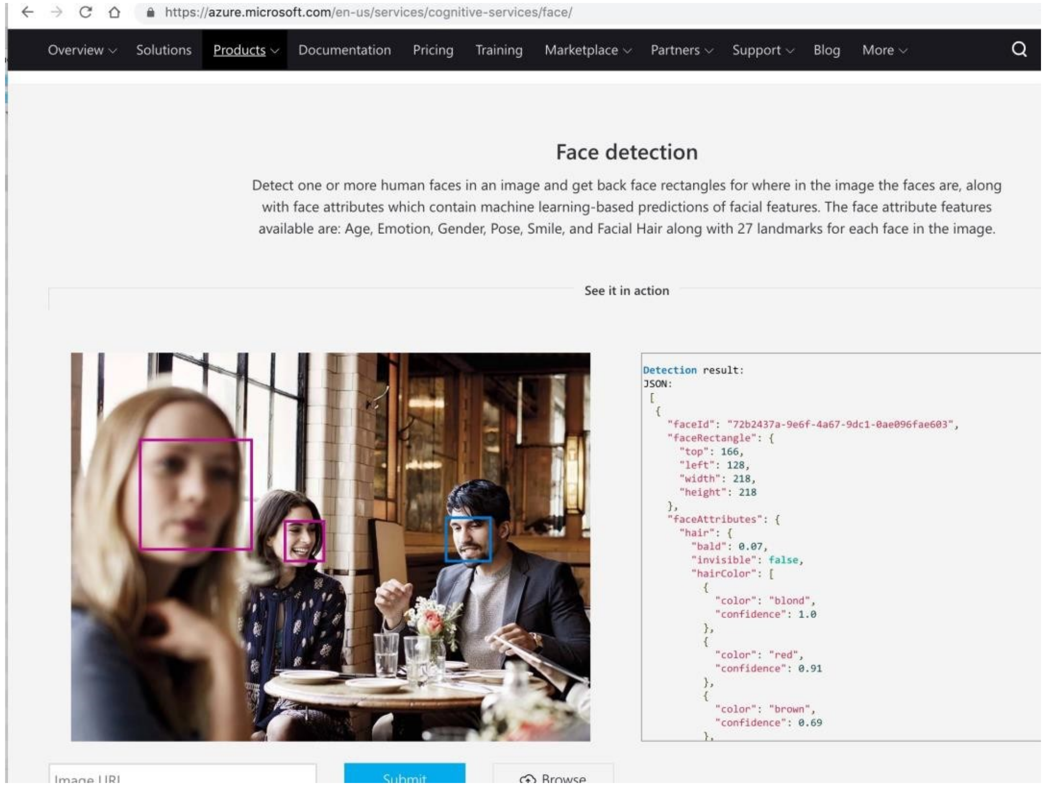


Face + Emotions Demos .. 1

Face API

https://azure.microsoft.com/en-us/services/cognitive-services/face/

- Identify/Verify Faces
- Detect Faces
- Recognize emotions
- Find similar faces, cluster similar faces

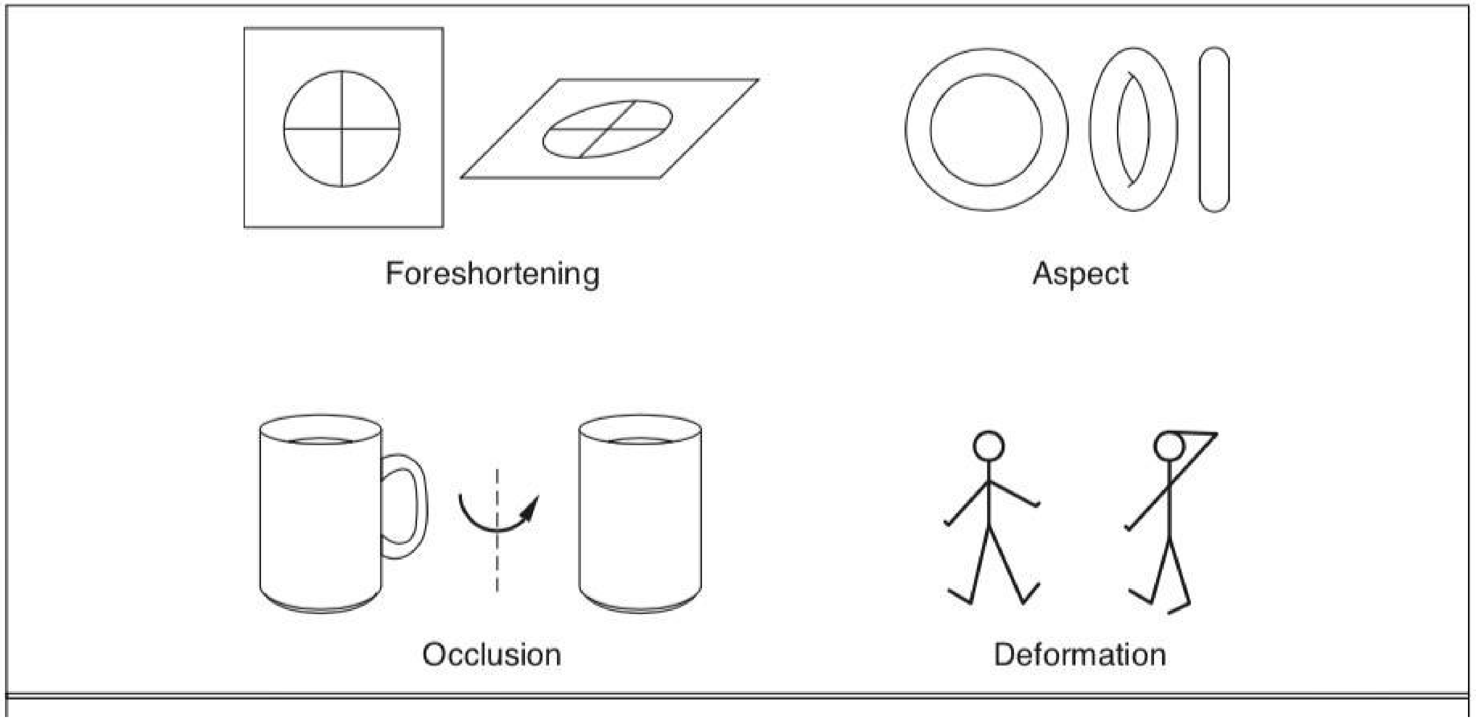


Face + Emotions Demos .. 2

Appearance Variation

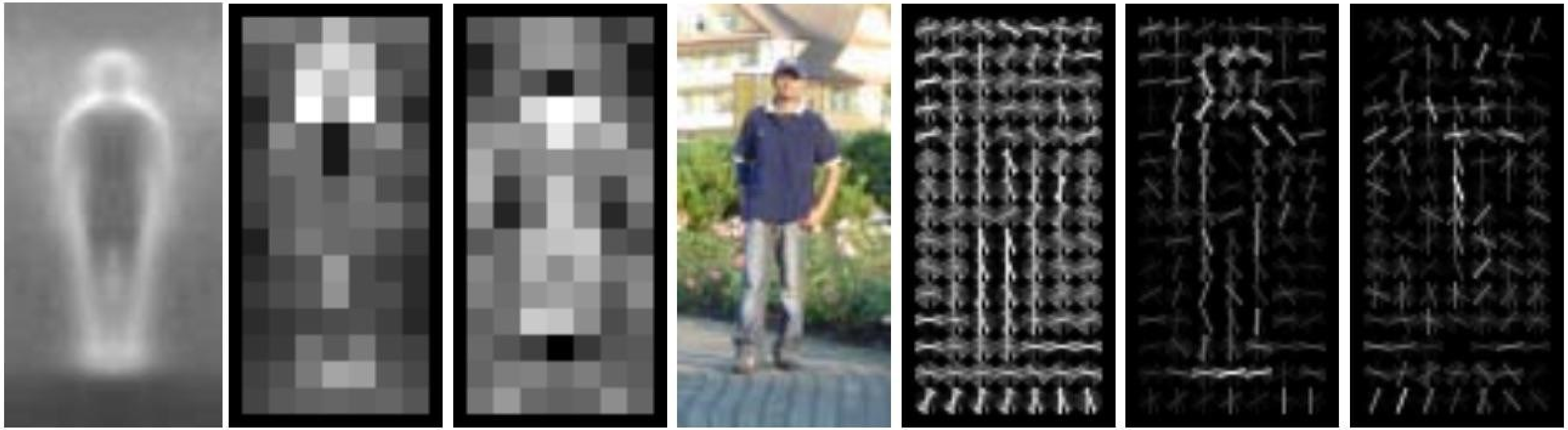
Many effects move features around in an image:

- Foreshortening
- Aspect
- Occlusion
- Deformation



Pedestrian Detection using HOG

Get features using the Histogram of Oriented Gradients (HOG) method, then present to classifier (SVM or similar)

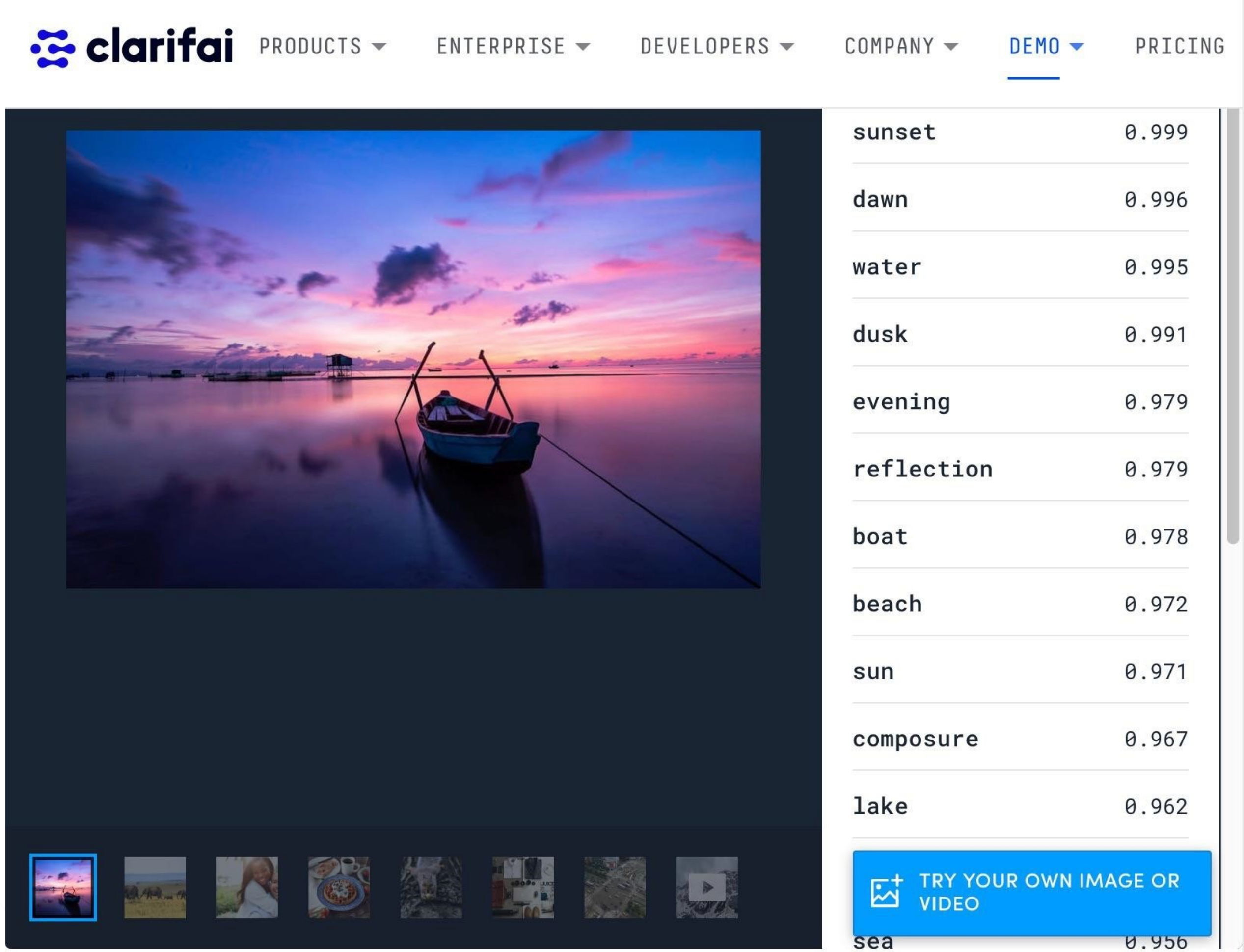


(a) (b) (c) (d) (e) (f) (g)

Figure 6. Our HOG detectors cue mainly on silhouette contours (especially the head, shoulders and feet). The most active blocks are centred on the image background just *outside* the contour. (a) The average gradient image over the training examples. (b) Each "pixel" shows the maximum positive SVM weight in the block centred on the pixel. (c) Likewise for the negative SVM weights. (d) A test image. (e) It's computed R-HOG descriptor. (f,g) The R-HOG descriptor weighted by respectively the positive and the negative SVM weights.

Object Recognition: Clarifai Demo

https://www.clarifai.com/demo



Reconstructing The World

Reconstructing 3D

Going from 2D to 3D

- From one or more images

With 2 or more images, can triangulate

Use background knowledge about scene

Several visual cues

- Brief overview in the next few slides

Motion Parallax

Camera moving relative to the scene

→ optical flow

→ clues about depth etc.

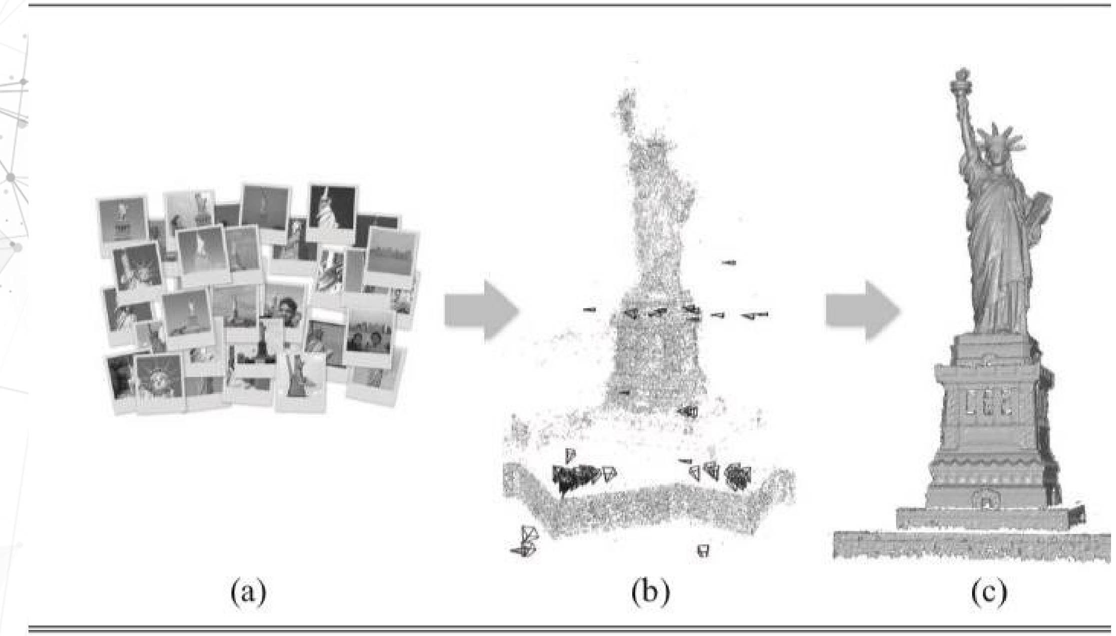
Binocular Stereopsis

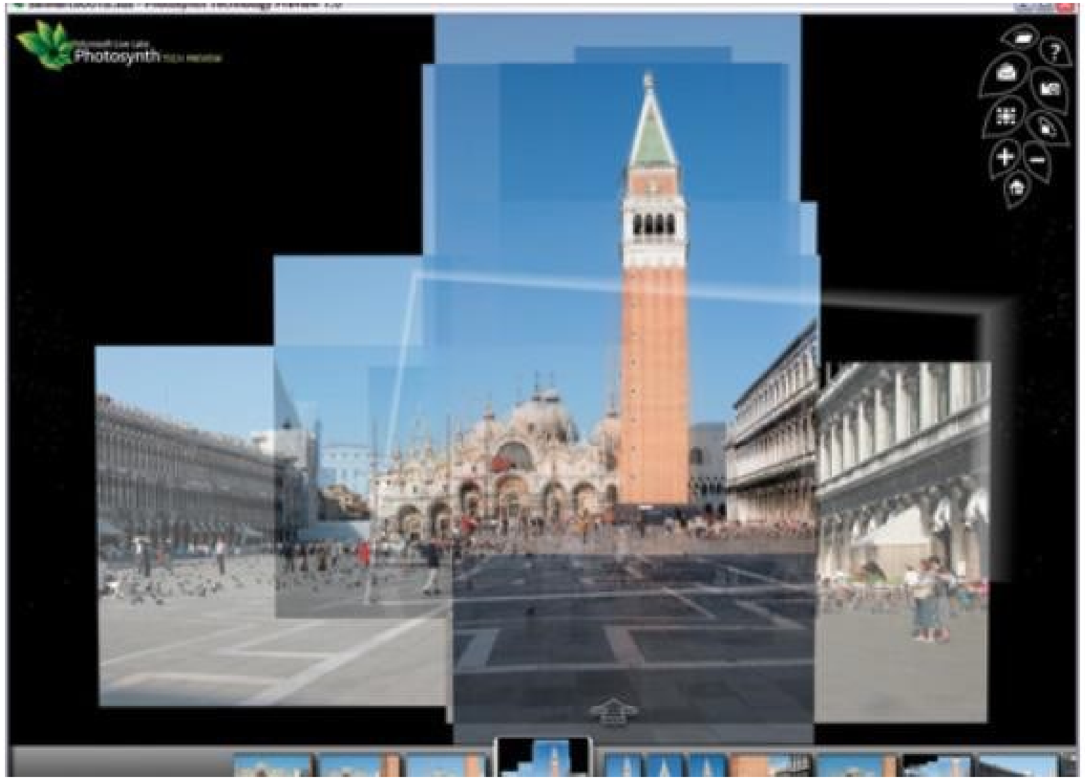
- (Most) vertebrates have 2 eyes
- Eyes in front -> binocular stereopsis
  - 2 images separated in space
- Try it: look at something in classroom, close one eye, then the other

http://oaklanddiscovery.blogspot.com/2012/01/stereoscope-viewer.html

Viewmaster/Amazon.com

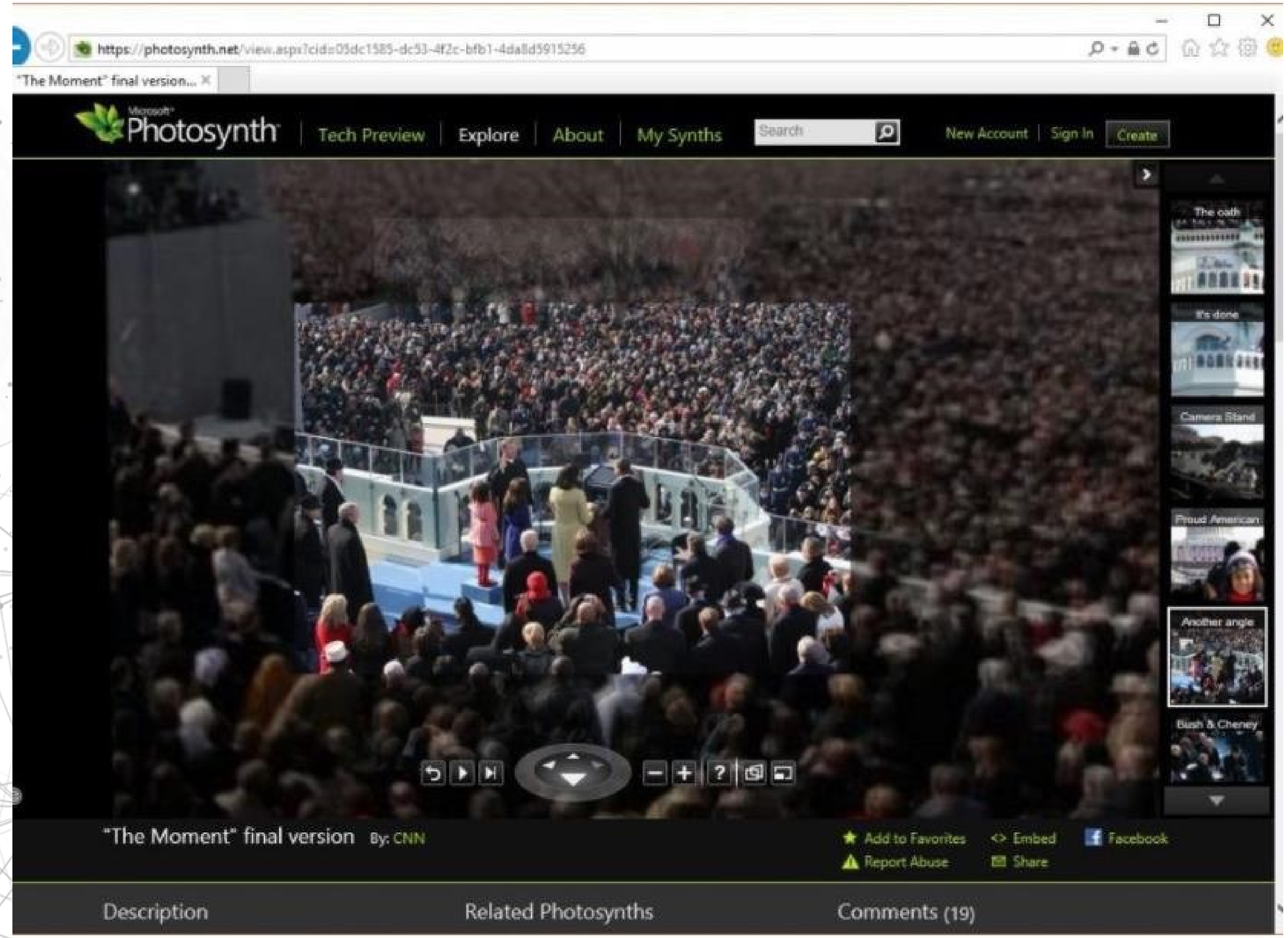
Multiple Views (Photosynth)





https://lazowska.cs.washington.edu/tr.photosynth/

Another Photosynth Example



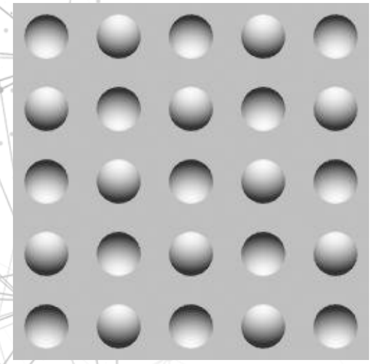
Texture

Can be used to estimate distance and viewing angle too

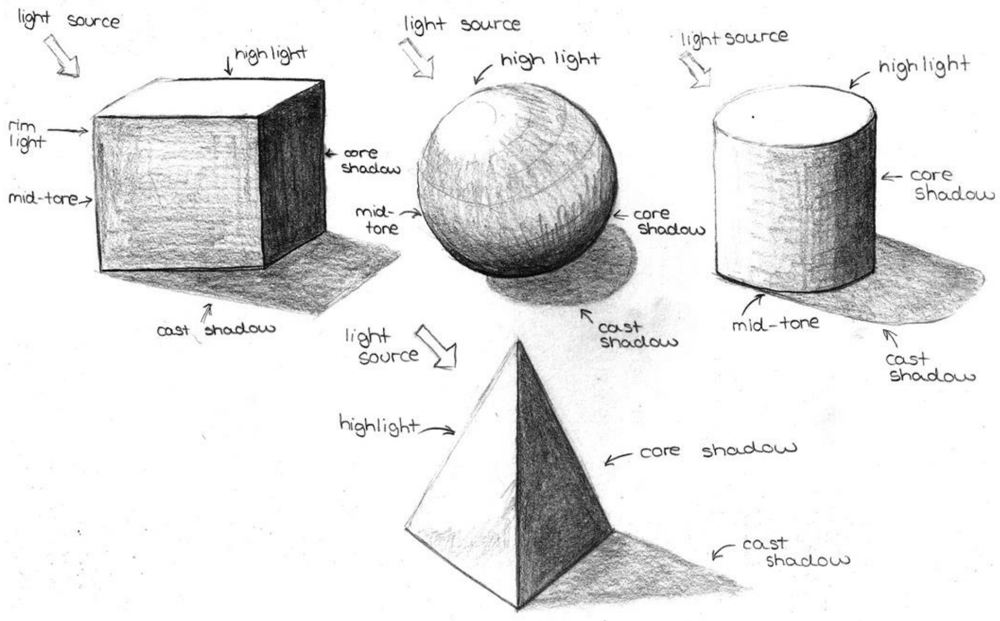
Shading

Get an idea of shape from shading

Humans do it easily



http://www.ru.is/faculty/thorisson/courses/v2008/gervigreind/perception1.html

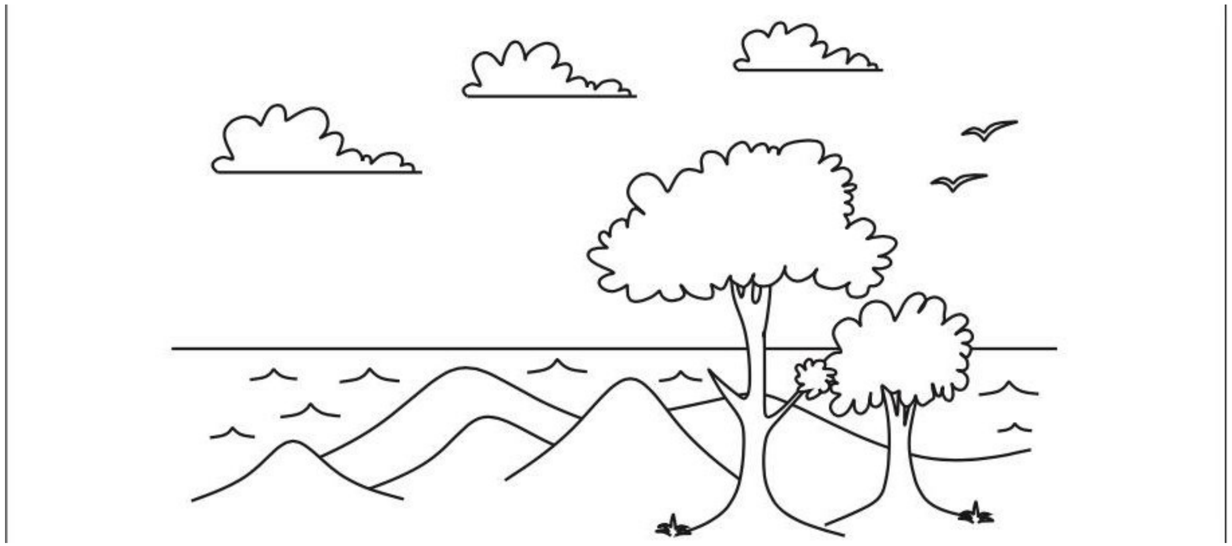


https://mrcolemita.files.wordpress.com/2017/08/shading-terms.jpg

Contour

Line drawing to 3D perception – how?

- Occluding contours (hills)
- Figure (nearer)-[back]ground (farther)
- Position on ground-plane
- …



Geometric Structure of Scenes

Locating horizon in scene ➔

relative position of people in scene, using sizes

The Rialto Bridge/Canaletto ca 1740

Consider techniques to use 2D to represent 3D

Paul Allen's collection

Detail …1

Detail …2

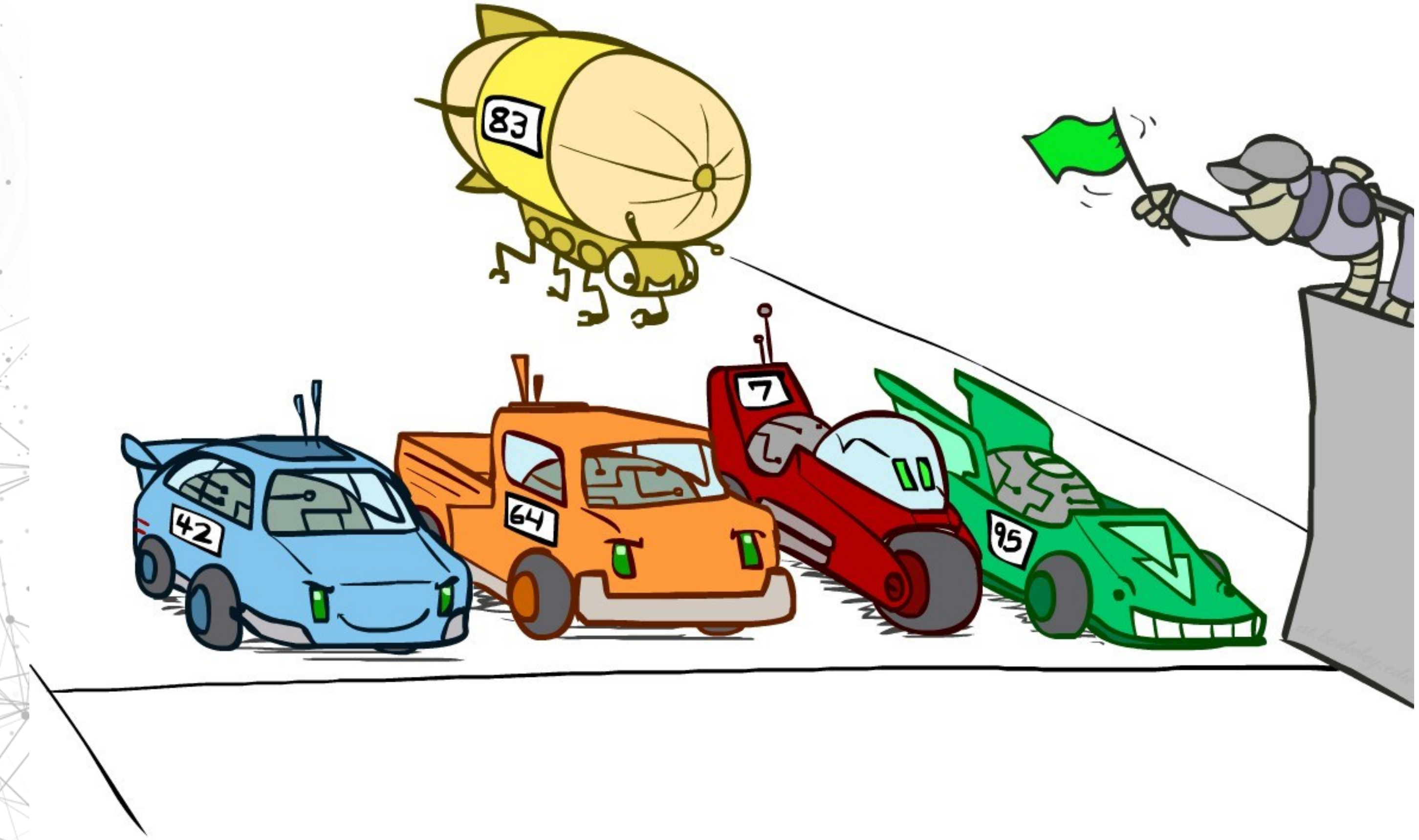
Generating Images

https://www.youtube.com/watch?v=F1X4fHzF4mQ

DALL-E 3

https://www.youtube.com/watch?v=sqQrN0iZBs0

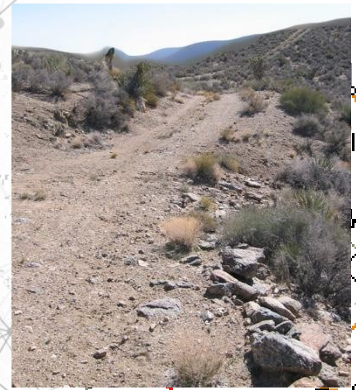
Autonomous Driving

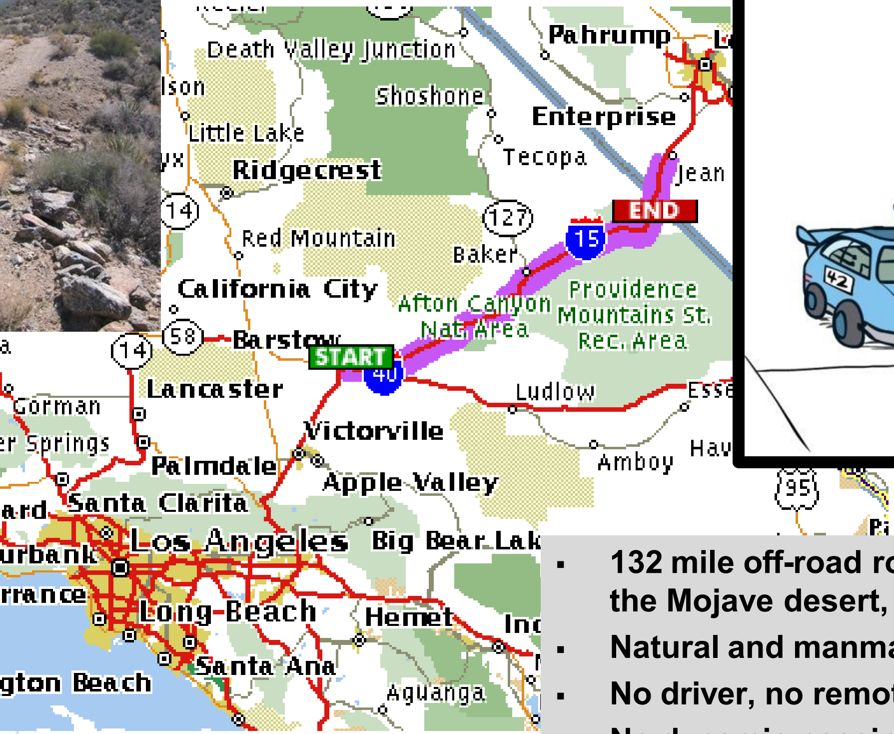


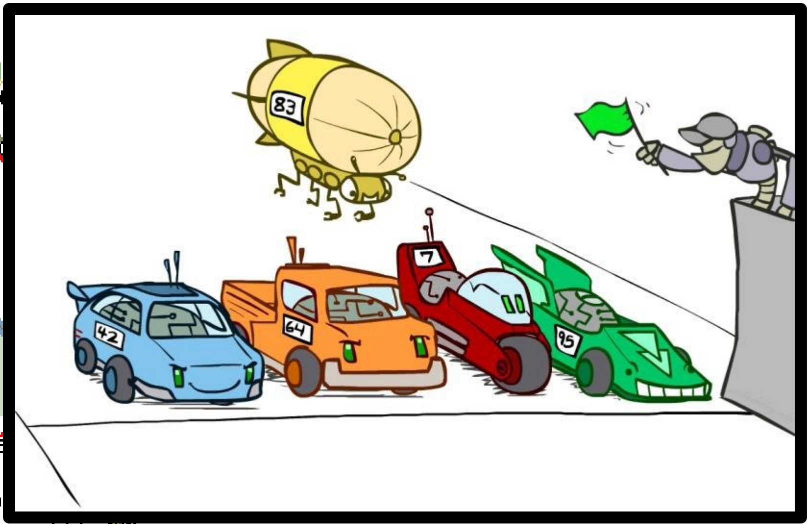
Sebastian Thrun / TED 2011

Click here to view video in new browser

Grand Challenge 2005: Barstow, CA, to Primm, NV



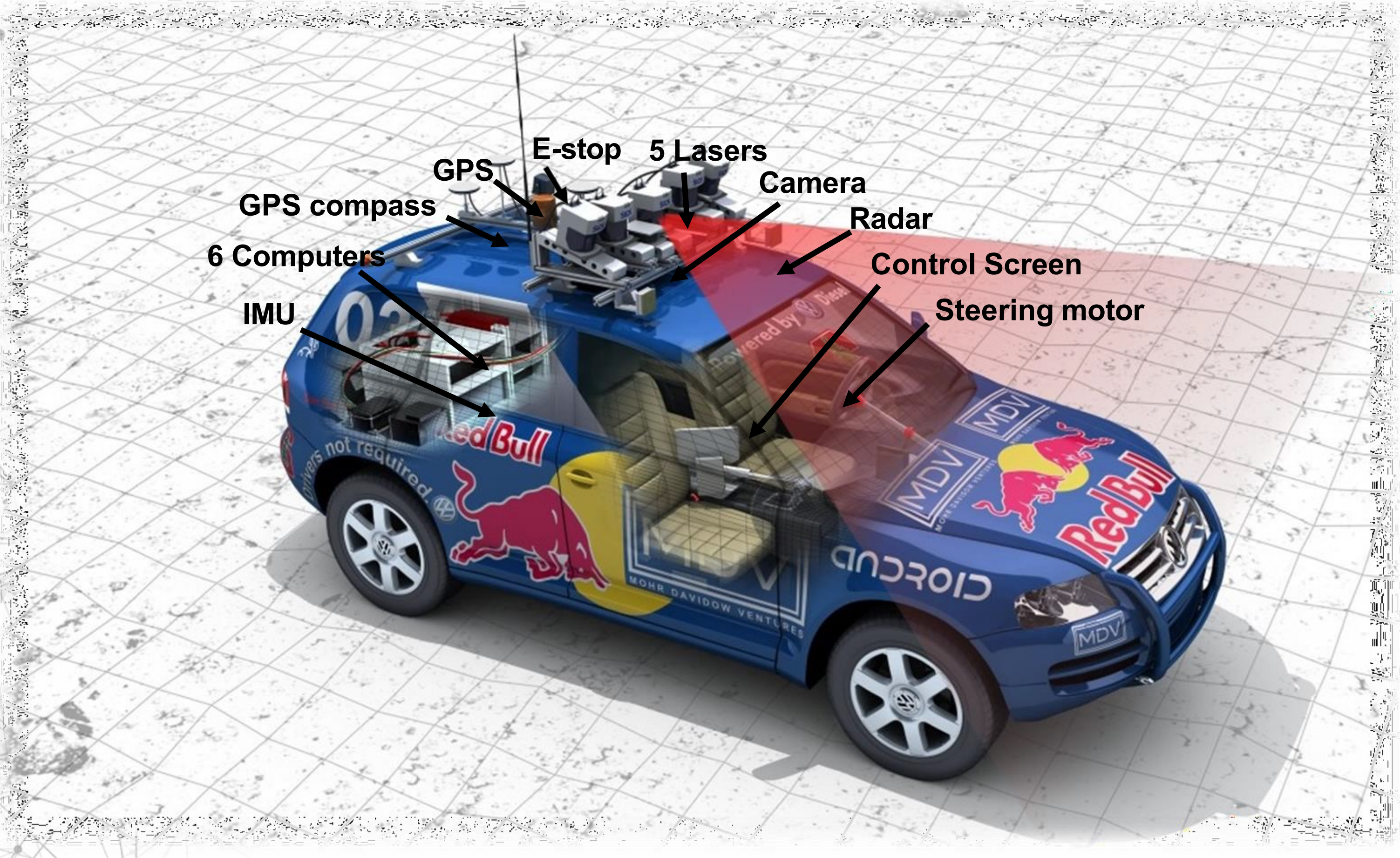




- 132 mile off-road robot race across the Mojave desert, < 10 hrs
- Natural and manmade hazards
- No driver, no remote control
- No dynamic passing

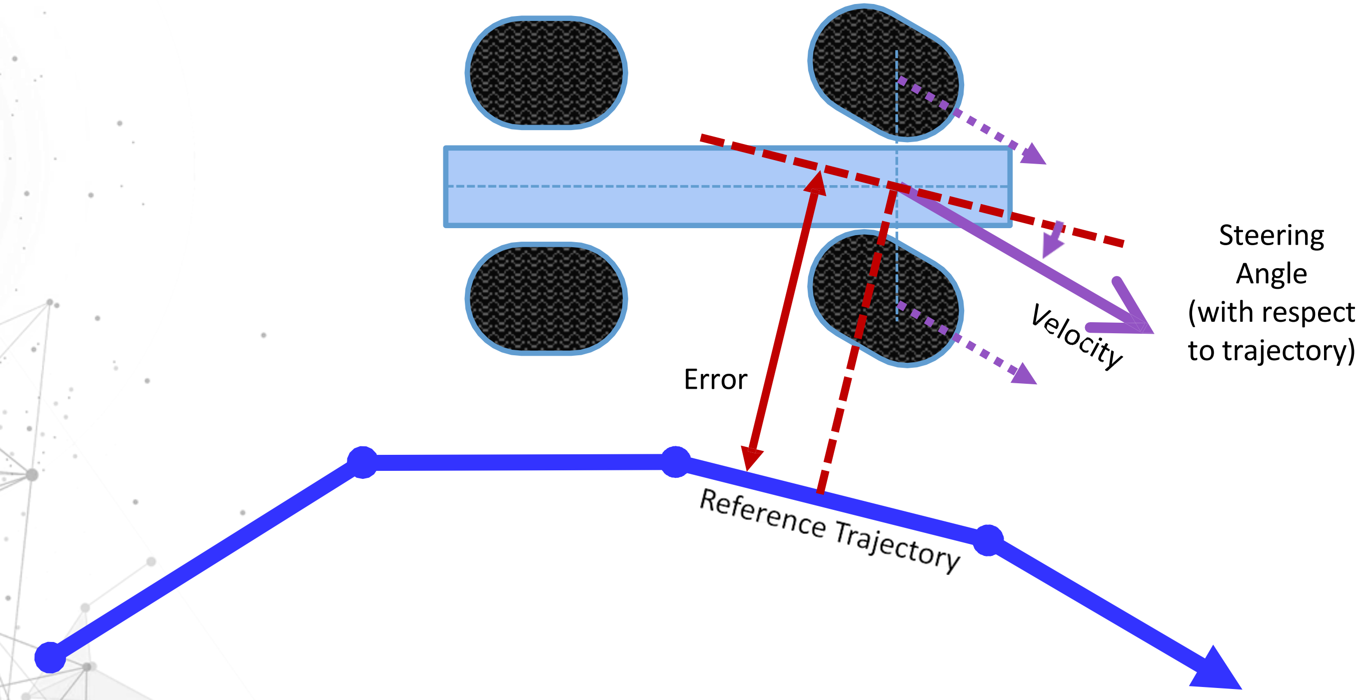
Grand Challenge 2005
Nova Video

An Autonomous Car

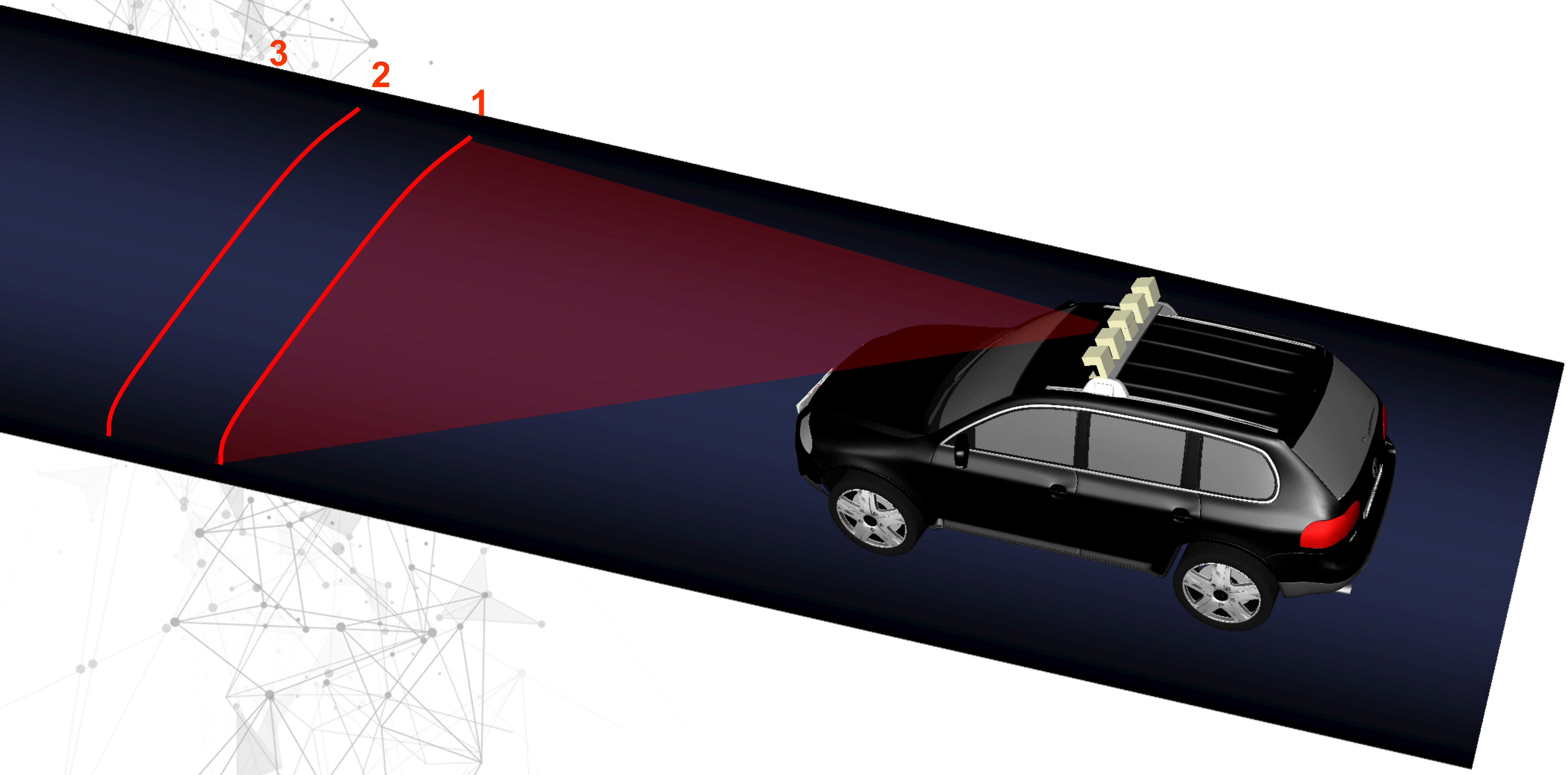


IMU: Inertial Measurement Unit

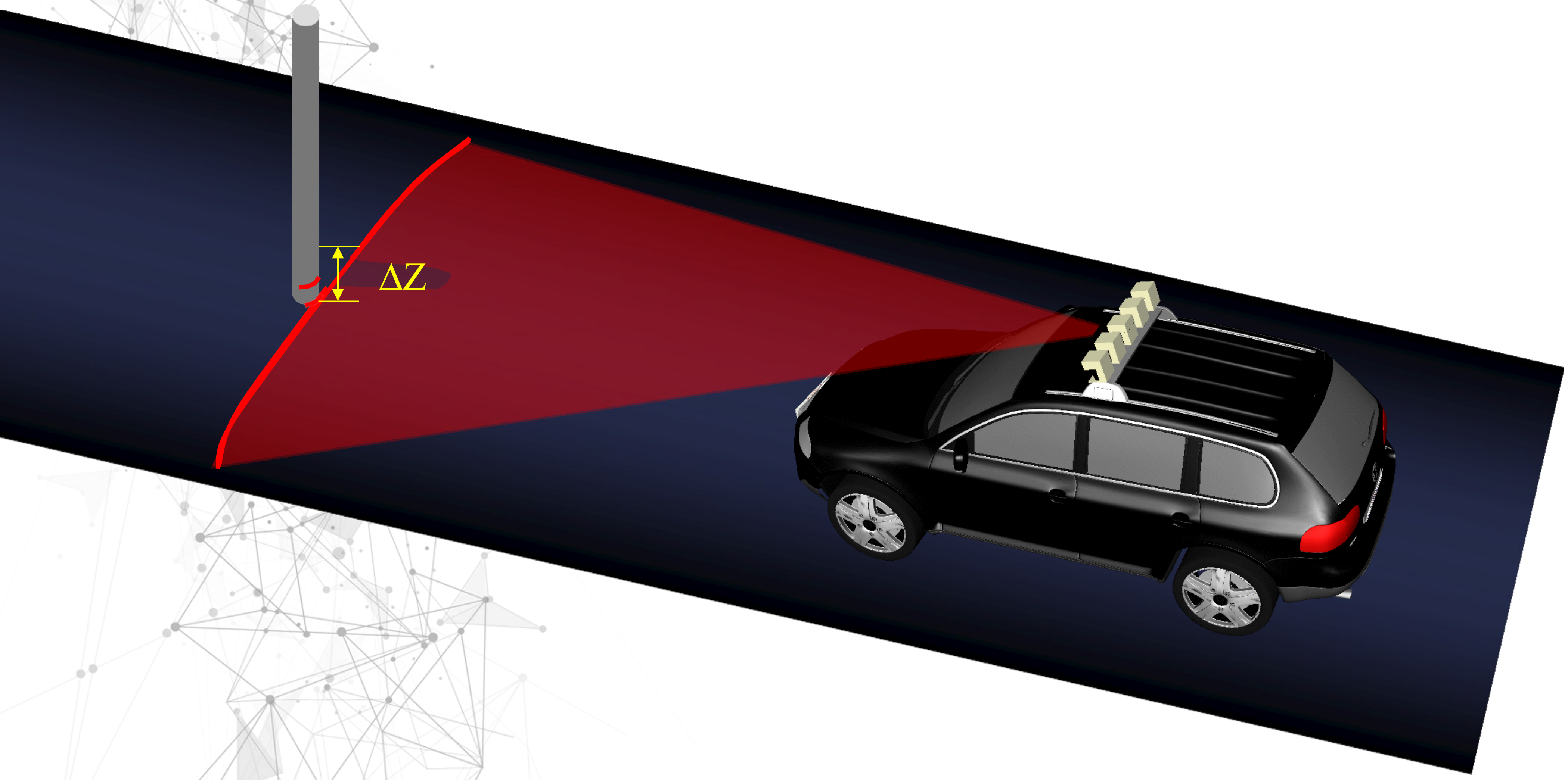
Actions: Steering Control



Laser Readings for Flat / Empty Road

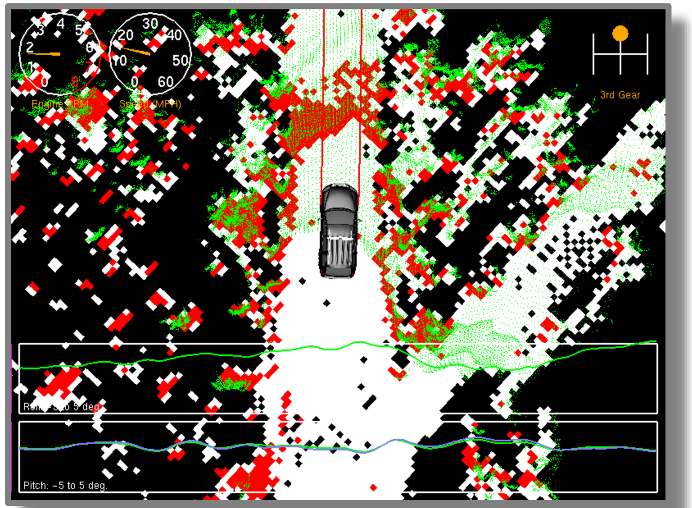


Laser Readings for Road with Obstacle



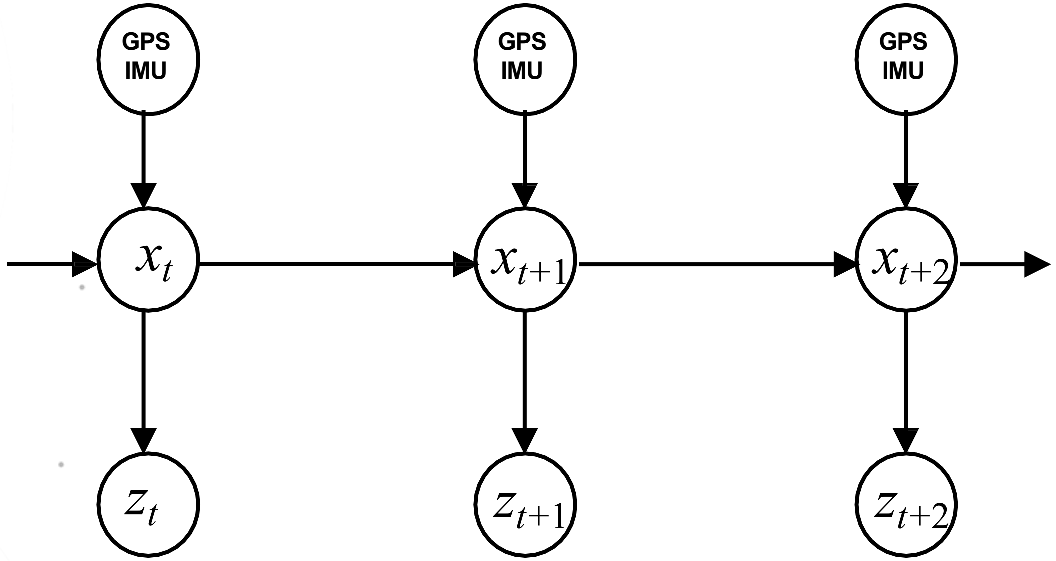
Obstacle Detection

Trigger if $|Z^i - Z^j| > 15\mathrm{cm}$ for nearby z^i, z^j

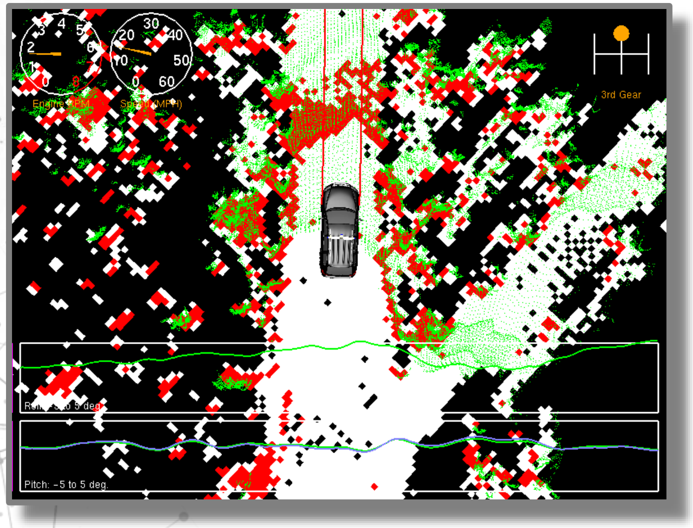


Raw Measurements: 12.6% false positives

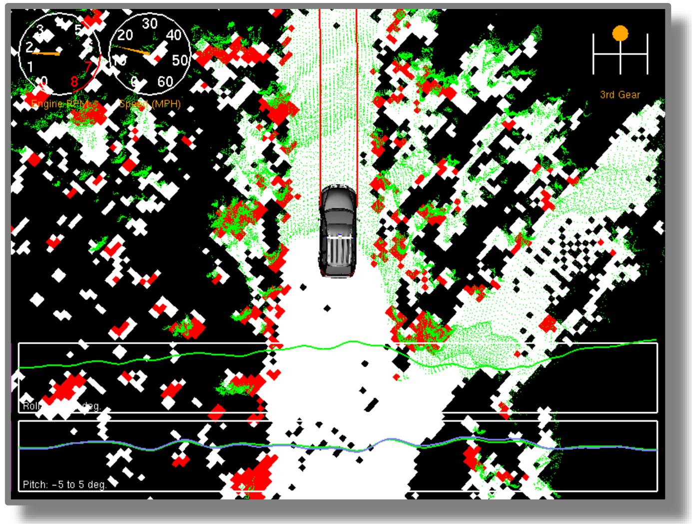
Probabilistic Error Model



Hidden Markov Models for Detection

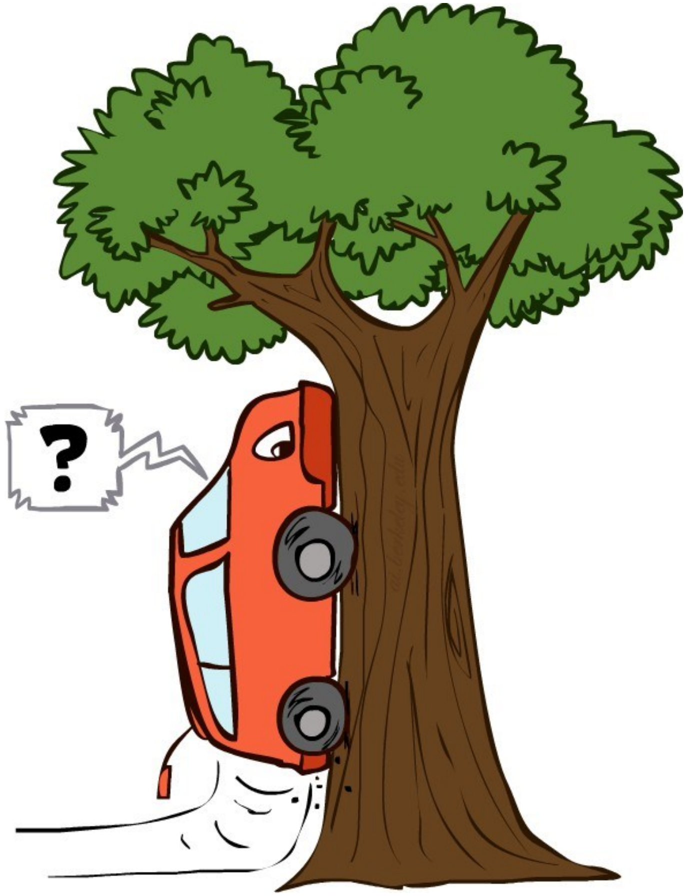


Raw Measurements:
12.6% false positives



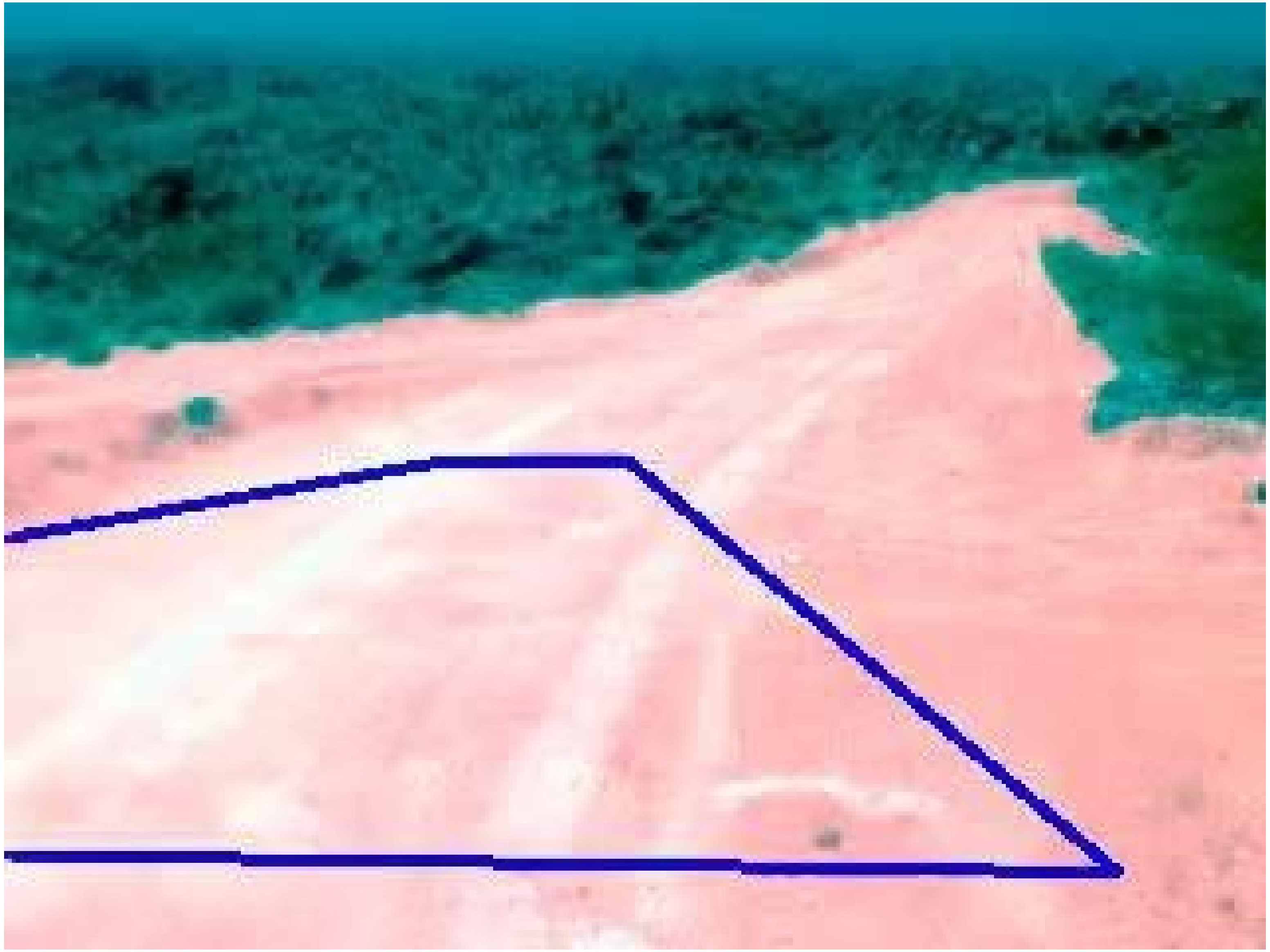
HMM Inference:
0.02% false positives

Sensors: Camera

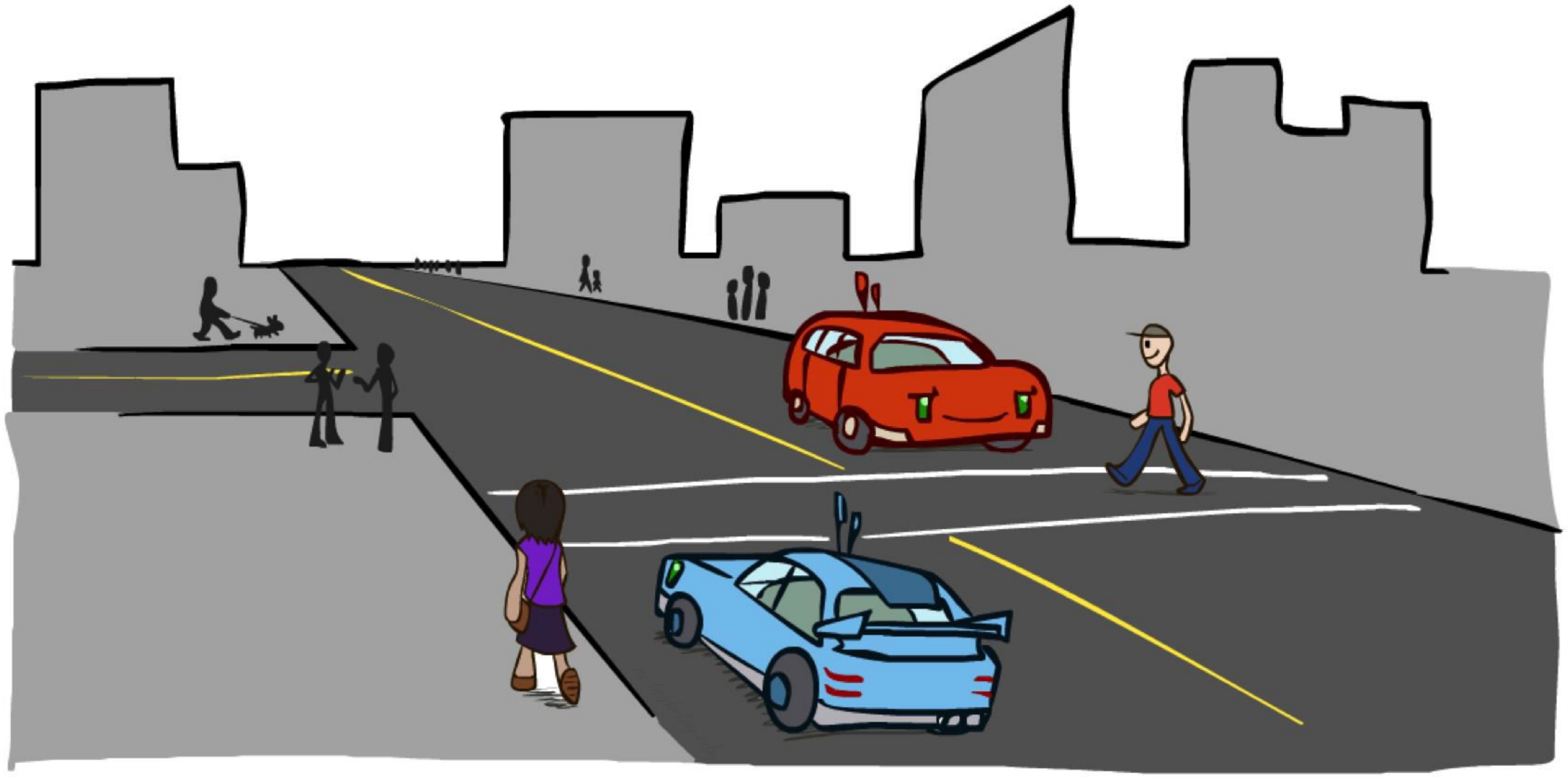


Vision for a Car

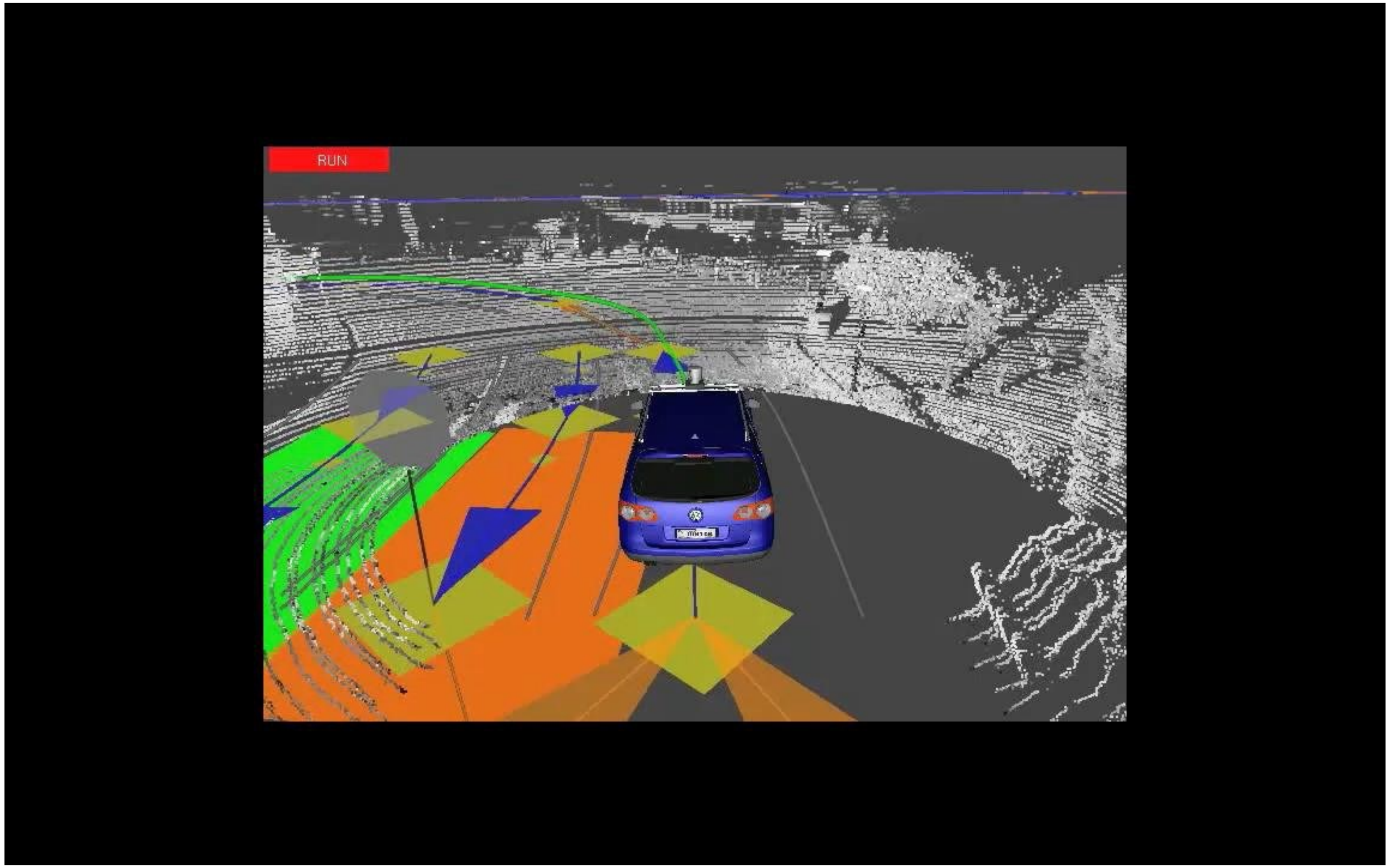
Self-Supervised Vision



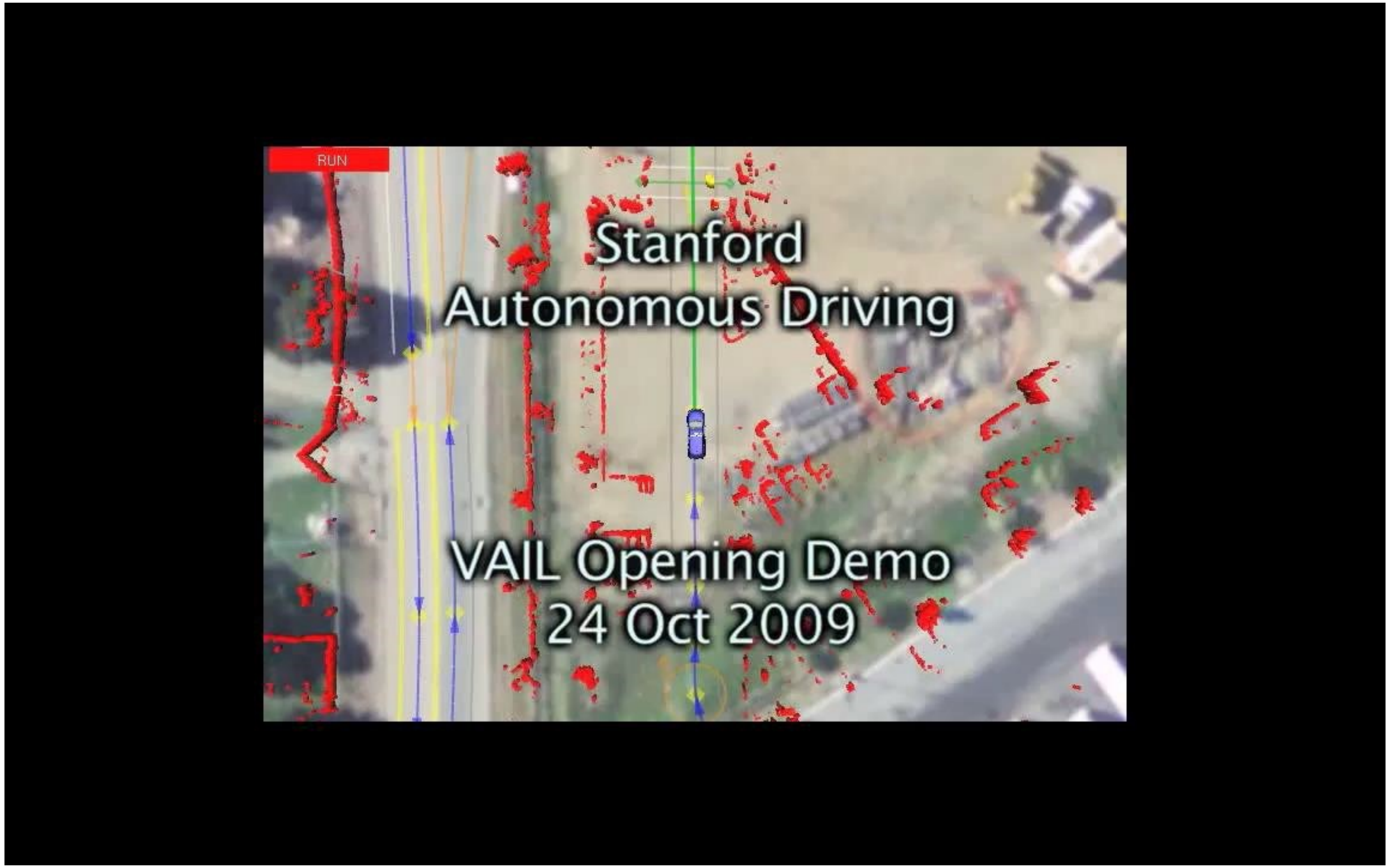
Urban Environments



Sensors
Laser Readings in Urban Scenes



Environmental Tracking



Google Self-Driving Car

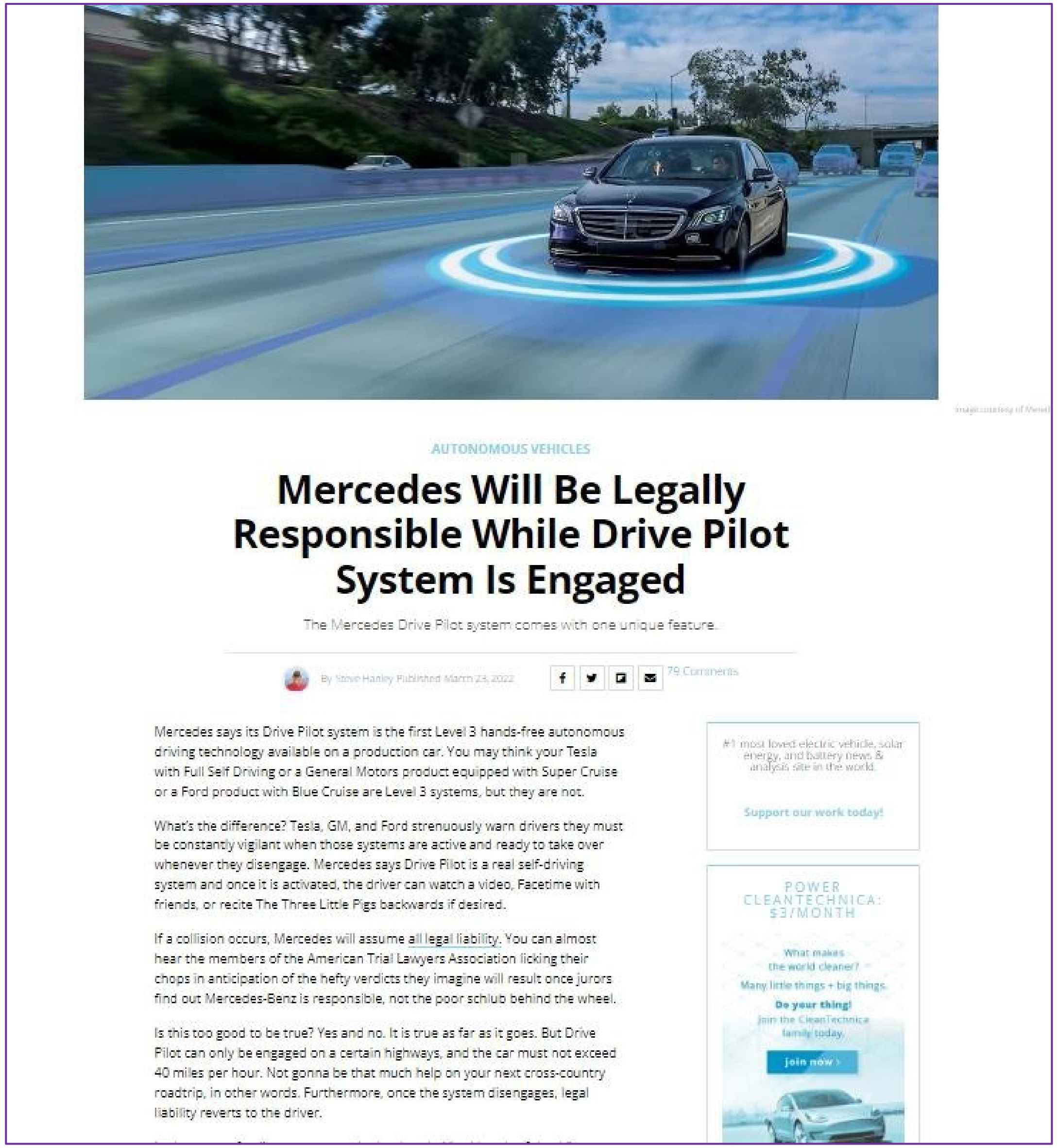
“Pulling over” an autonomous car

Interacting with a Cruise Autonomous Vehicle: A Guide for First Responders

In case of accidents, who is responsible?

- Car manufacturer?
- Software developer?
- Car owner?
- Car operator (Uber/Lyft etc.)?
- Passenger(s)?

- More during Lecture 20



AUTONOMOUS VEHICLES

Mercedes Will Be Legally Responsible While Drive Pilot System Is Engaged

The Mercedes Drive Pilot system comes with one unique feature.

By Steve Hanley Published March 23, 2022 79 Comments

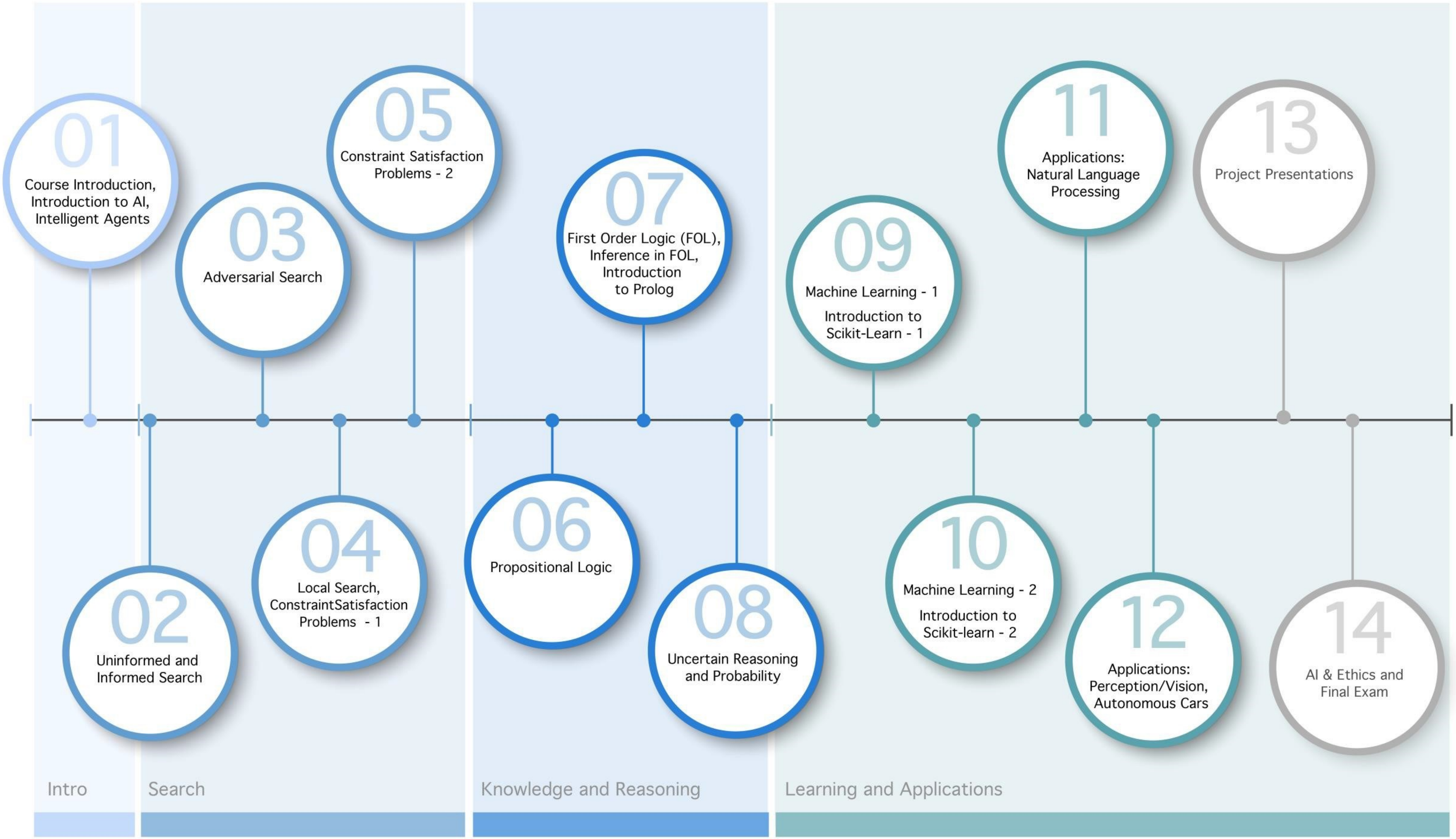
Mercedes says its Drive Pilot system is the first Level 3 hands-free autonomous driving technology available on a production car. You may think your Tesla with Full Self Driving or a General Motors product equipped with Super Cruise or a Ford product with Blue Cruise are Level 3 systems, but they are not.

What's the difference? Tesla, GM, and Ford strenuously warn drivers they must be constantly vigilant when those systems are active and ready to take over whenever they disengage. Mercedes says Drive Pilot is a real self-driving system and once it is activated, the driver can watch a video, Facetime with friends, or recite The Three Little Pigs backwards if desired.

If a collision occurs, Mercedes will assume all legal liability. You can almost hear the members of the American Trial Lawyers Association licking their chops in anticipation of the hefty verdicts they imagine will result once jurors find out Mercedes-Benz is responsible, not the poor schlub behind the wheel.

Is this too good to be true? Yes and no. It is true as far as it goes. But Drive Pilot can only be engaged on a certain highways, and the car must not exceed 40 miles per hour. Not gonna be that much help on your next cross-country roadtrip, in other words. Furthermore, once the system disengages, legal liability reverts to the driver.

https://cleantechnica.com/2022/03/23/mercedes-will-be-legally-responsible-while-drive-pilot-system-is-engaged/


01
Course Introduction, Introduction to AI, Intelligent Agents
02
Uninformed and Informed Search
03
Adversarial Search
04
Local Search, ConstraintSatisfaction Problems - 1
05
Constraint Satisfaction Problems - 2
06
Propositional Logic
07
First Order Logic (FOL), Inference in FOL, Introduction to Prolog
08
Uncertain Reasoning and Probability
09
Machine Learning - 1
Introduction to Scikit-Learn - 1
10
Machine Learning - 2
Introduction to Scikit-learn - 2
11
Applications: Natural Language Processing
12
Applications: Perception/Vision, Autonomous Cars
13
Project Presentations
14
AI & Ethics and Final Exam
Intro
Search
Knowledge and Reasoning
Learning and Applications