COSC428 Computer Vision

Overview

- Perception
- Image processing
- Multi-camera
- 3D vision
- Motion
- **Deep learning**

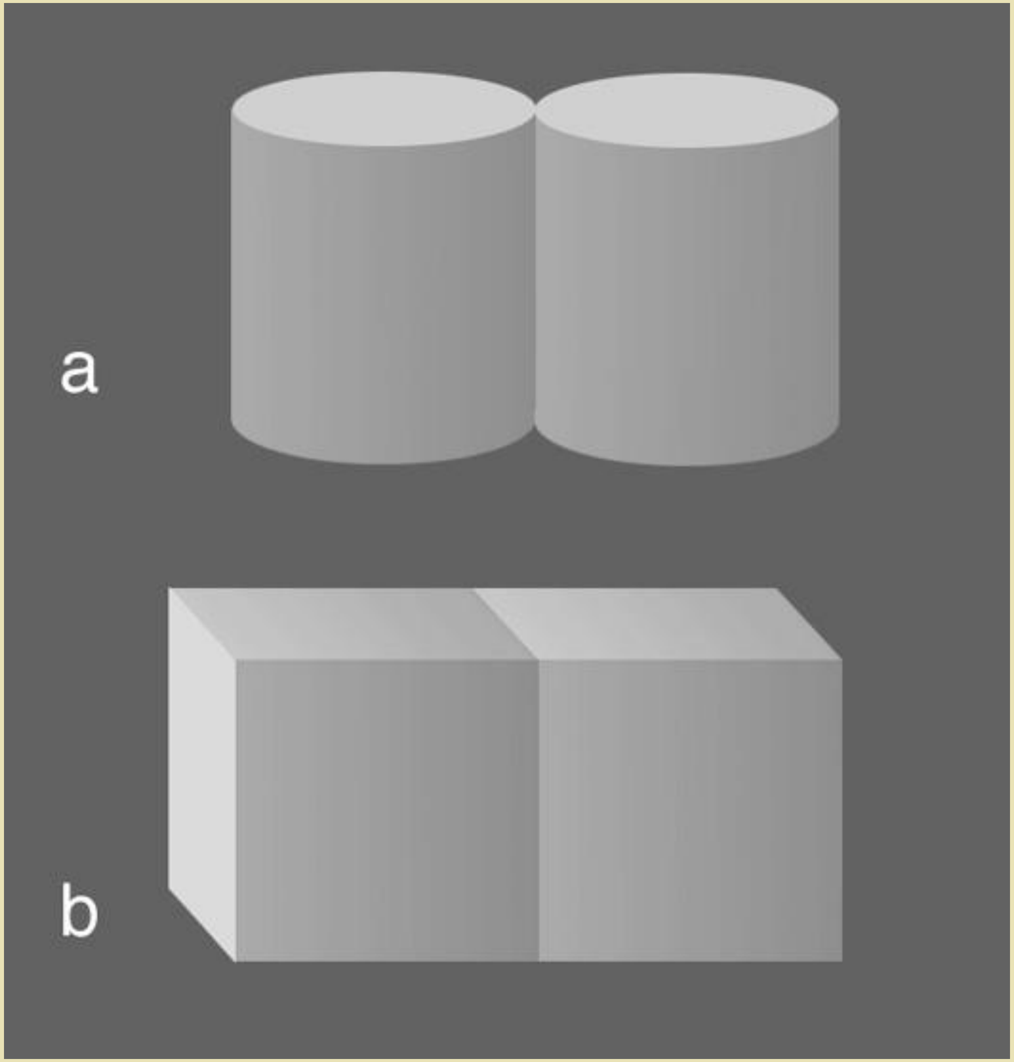
Computer Vision

The goal of

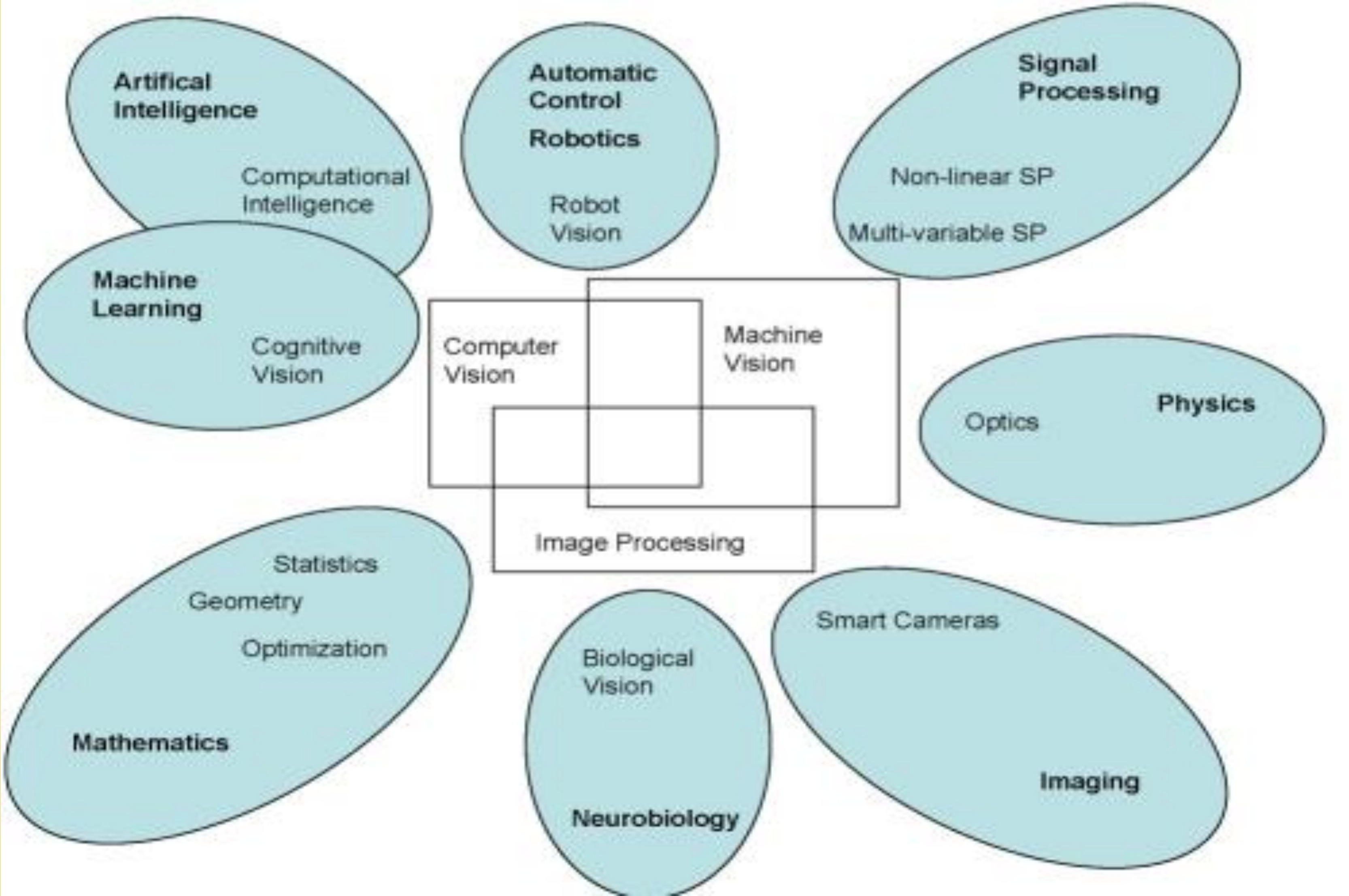
- computer vision
- machine vision
- robot vision
- drone vision
- (deep learning)

is to recognise objects and their motion.

Vision is Inferential



Computer Vision

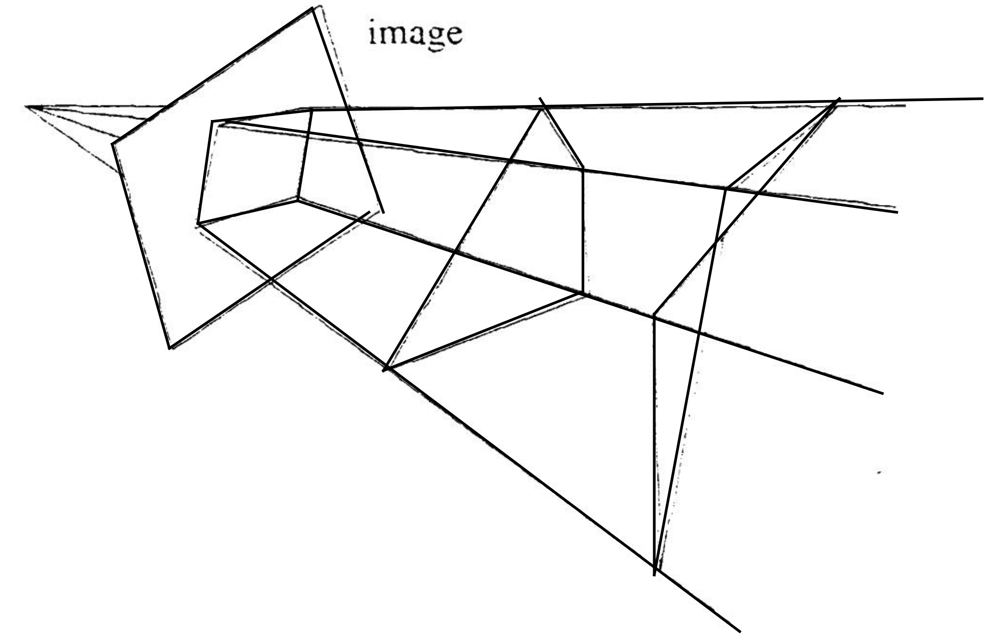


Intelligence and Perception

- First to understand how we perceive the world then to teach the machine to interpret the world based on primitive data it has received
- ***Perceptual modalities***
  - **Tactile – touch**
  - **Gustatory – taste**
  - **Visual – sight**
  - **Auditory – hearing**
  - **Olfactory – smell**
- Different levels of processing for bridging different gaps - Low to high

Difficulties in vision computing
- the sensory gap

- The sensory gap is the gap between the object in the world and the information in a (computational) description derived from a recording of that scene.



An infinite number of 3D drawings can give rise to the same image (C1)

The goal of the shape-from-contour module is to derive information about the orientation of the various different faces.

Difficulties in vision computing
- The semantic gap

- The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.
- The higher the level of interpretation, the more domain knowledge and its management are required.

How Do We Recover 3-D Information?

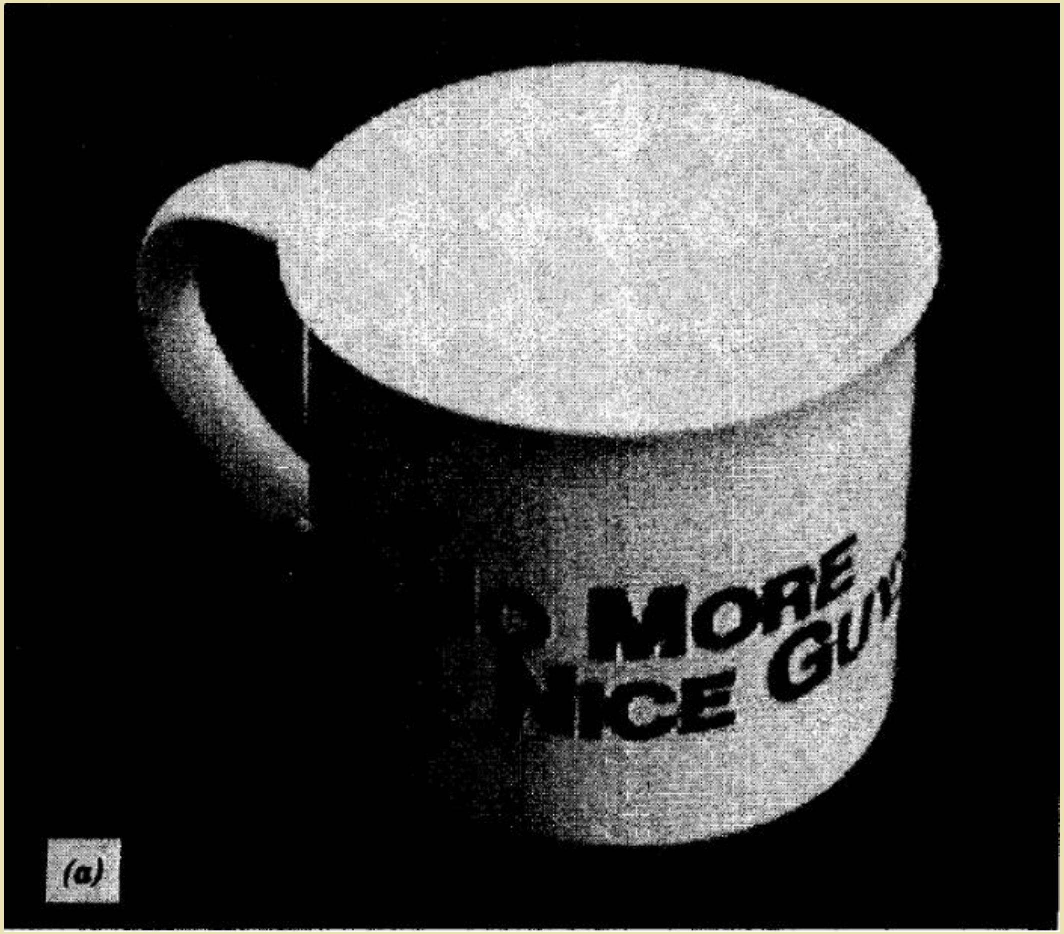
- There are number of cues available in the visual stimulus
  - Motion
  - Stereo
  - Texture
  - Shading
  - Contour
  - Time of flight
- Each of these cues relies on background assumptions about physical scenes in order to provide unambiguous interpretation.

Define the problem

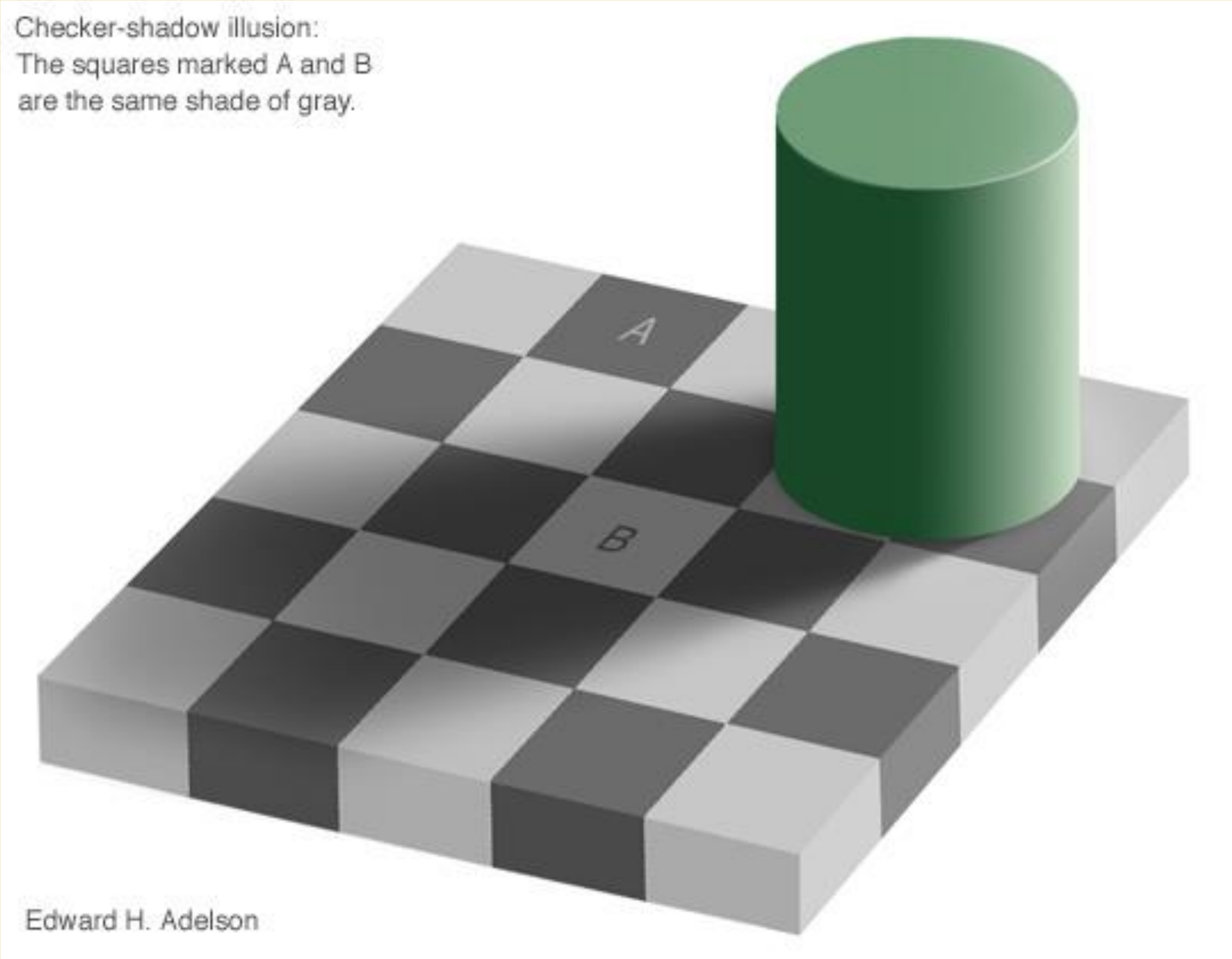
- Simple example - Grey images
- Image Intensities (brightness) are discretely sampled, and the sampled values are quantized to a discrete set of values (integers). The elements of the resulting array of numbers are called pixels, and their values are called grey levels.
  - A function giving the grey level at every point on the image plane.
  - Grey level varies from 0 (black) to 255 (max brightness, maximum response the eye can make)
  - Assign coordinates to the surface the image is produced on
    - Using two dimensional array
    - Pixel – receptor in the retina

The Difficulty in Vision Computing – Taking the Human Visual System for Granted

- The processing capability of human visual systems is often taken for granted
- The subtlety and difficulty of describing the exact operation of the subconscious functions presents significant difficult in developing algorithms to emulate human visual behaviour
- If we are a computer…


MORE
ICE GU
(a)

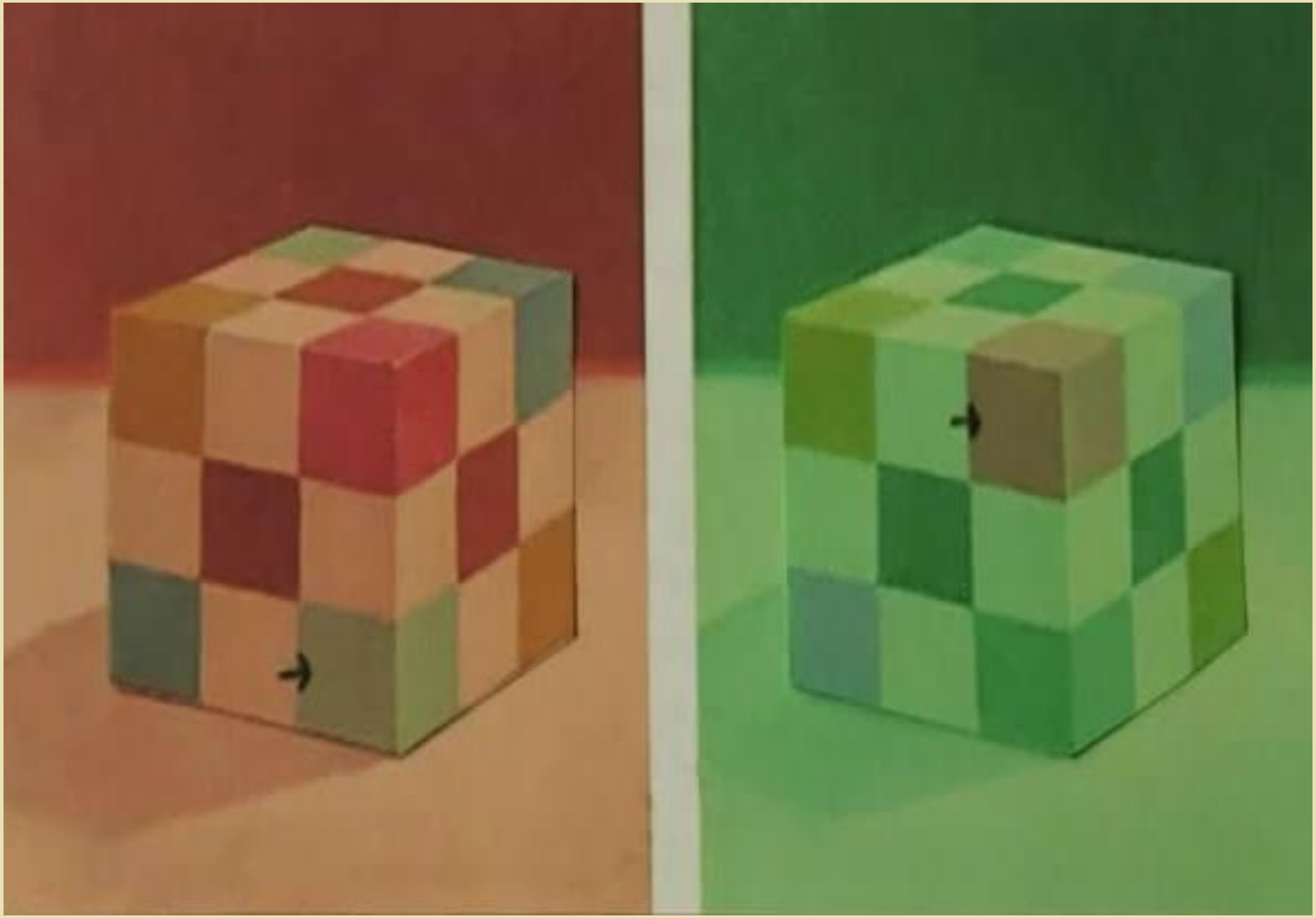
Vision is inferential: Light



(http://www-bcs.mit.edu/people/adelson/checkershadow_illusion.html)

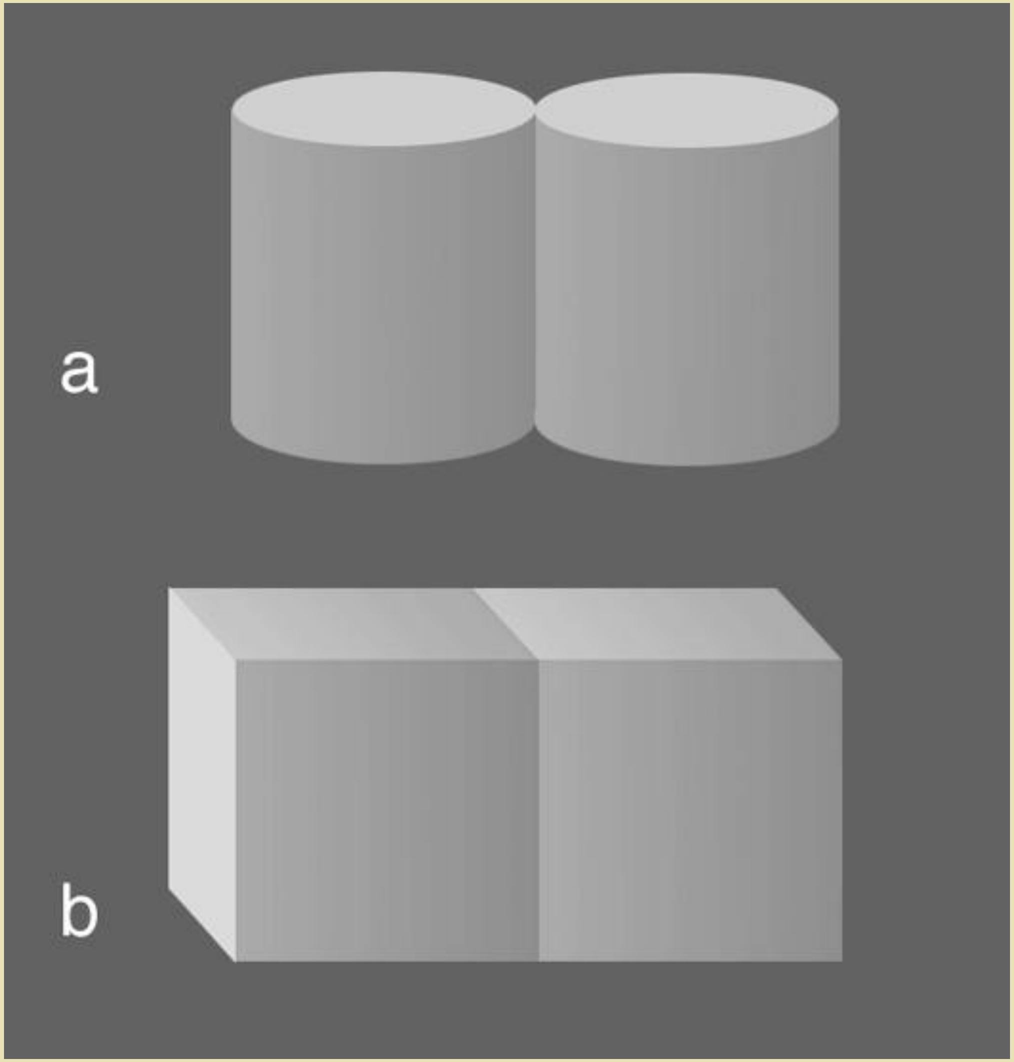
Is this dress
black & blue
or gold & white?

Other optical illusions



varying light source colour

Vision is Inferential



Vision is Inferential: Prior Knowledge

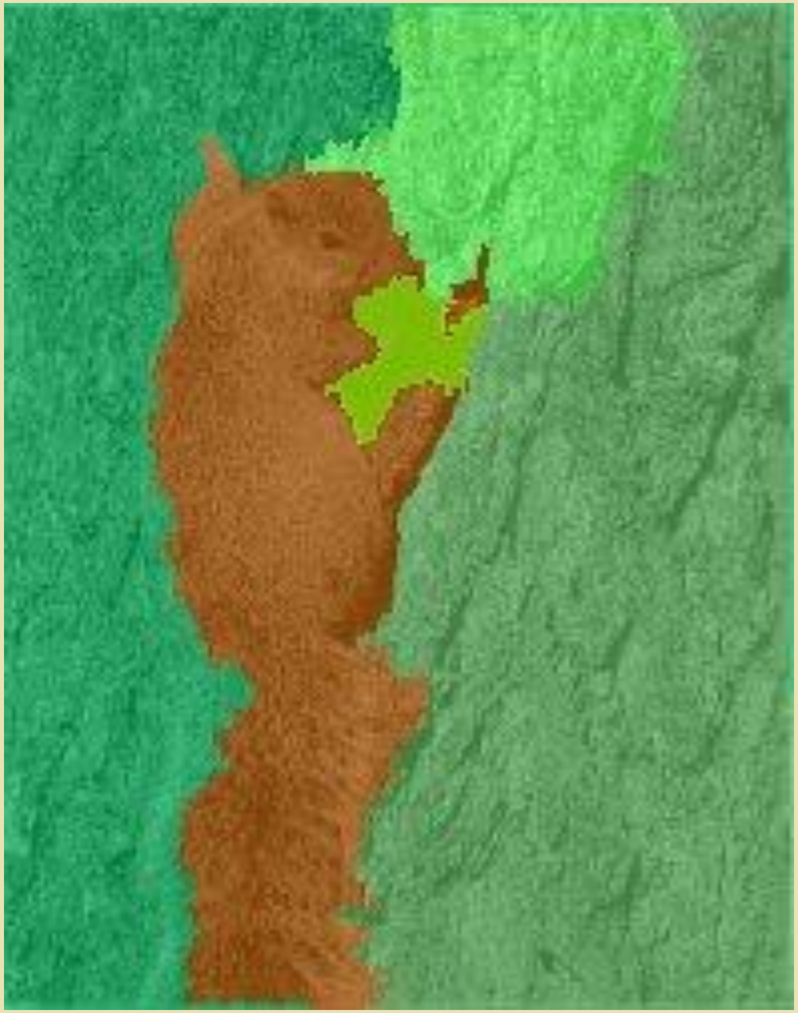
Computer Vision

- Inference → Computation
- Building machines that see
- Modeling biological perception

Boundary Detection: Local cues

Boundary Detection

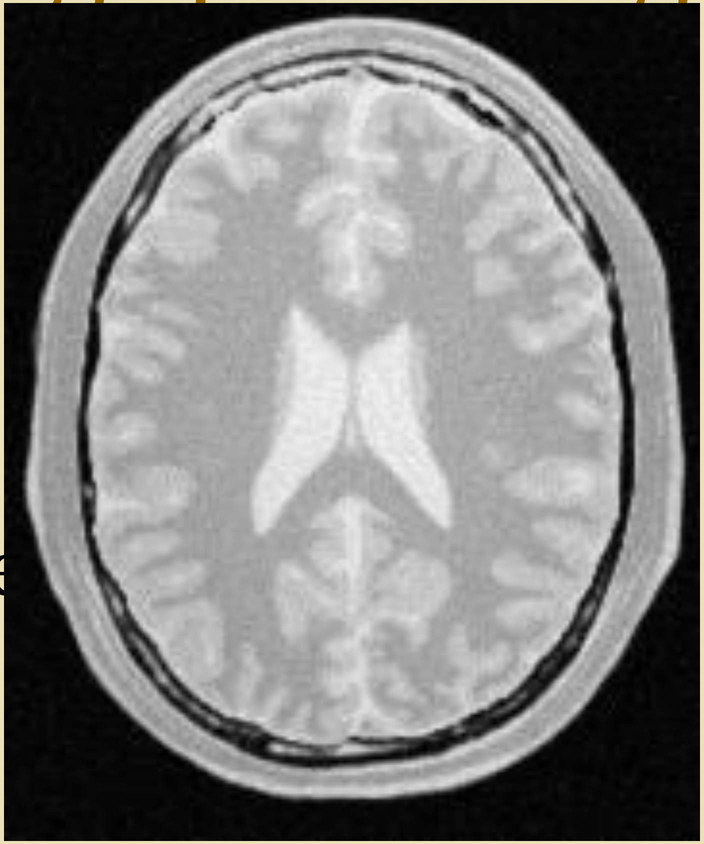
http://www.robots.ox.ac.uk/~vdg/dynamics.html



Computer Vision Processing

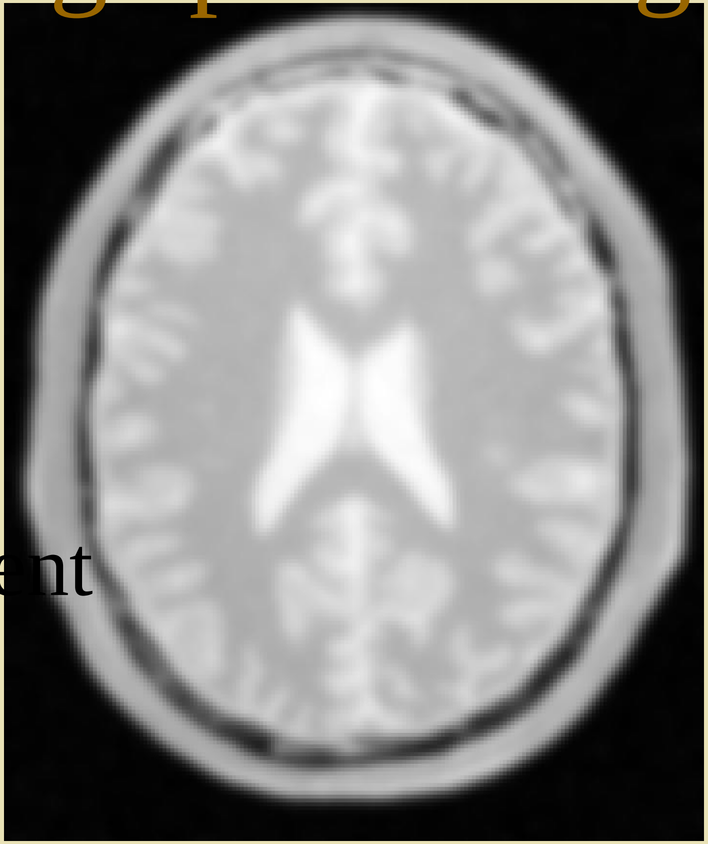
- Simulate human image perception
- pre-processing:
  - Noise removal, contrast enhancement etc.
- Early processing (Low level) – find useful info from raw images
  - Colour, edges, shape, texture detection
- Late processing (High level) – find objects and meanings from the useful info
  - Objects
  - Spatial relationship
  - Meanings
- The higher level processing, the more domain knowledge needed

Low level image processing



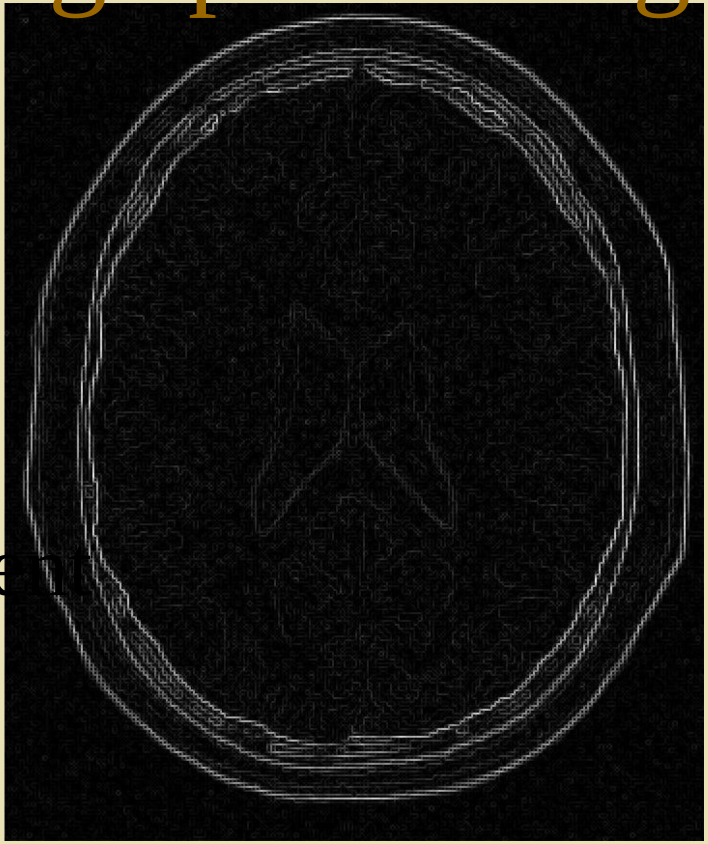
- Image compression
- Noise reduction
- Edge extraction
- Contrast enhanceme
- Segmentation
- Thresholding
- Morphology
- Image restoration

Low level image processing



- Image compression
- Noise reduction
- Edge extraction
- Contrast enhancement
- Segmentation
- Thresholding
- Morphology
- Image restoration

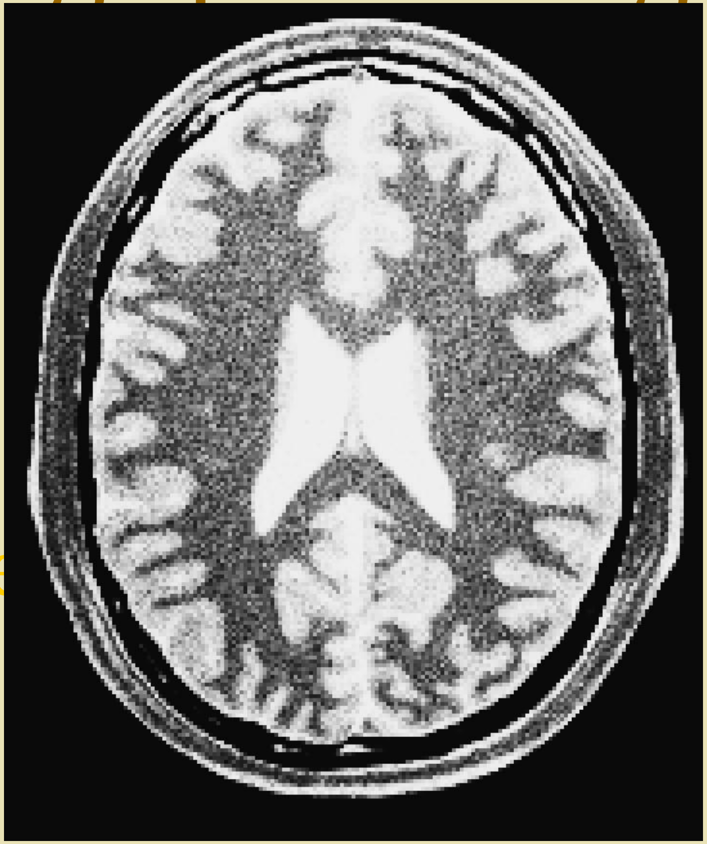
Low level image processing



- Image compression
- Noise reduction
- Edge extraction
- Contrast enhancement
- Segmentation
- Thresholding
- Morphology
- Image restoration

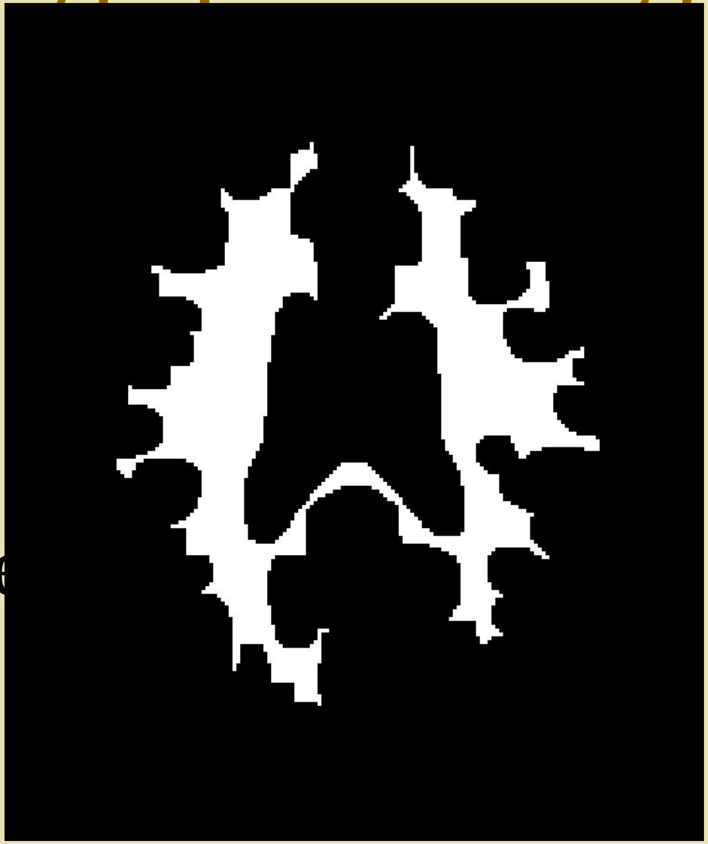
Low level image processing

- Image compression
- Noise reduction
- Edge extraction
- Contrast enhancement
- Segmentation
- Thresholding
- Morphology
- Image restoration

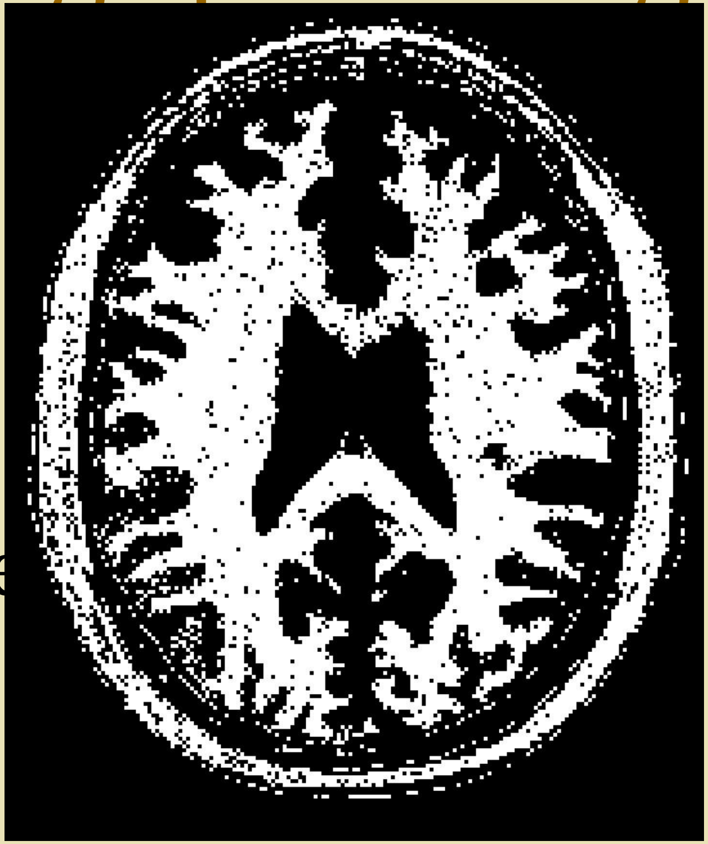


Low level image processing

- Image compression
- Noise reduction
- Edge extraction
- Contrast enhanceme
- Segmentation
- Thresholding
- Morphology
- Image restoration

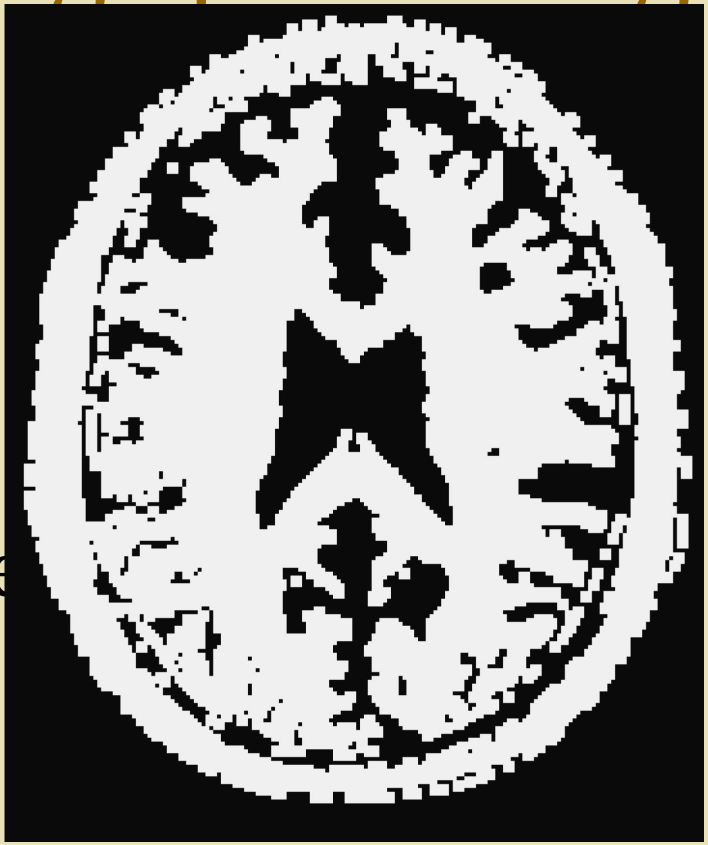


Low level image processing



- Image compression
- Noise reduction
- Edge extraction
- Contrast enhanceme
- Segmentation
- Thresholding
- Morphology
- Image restoration

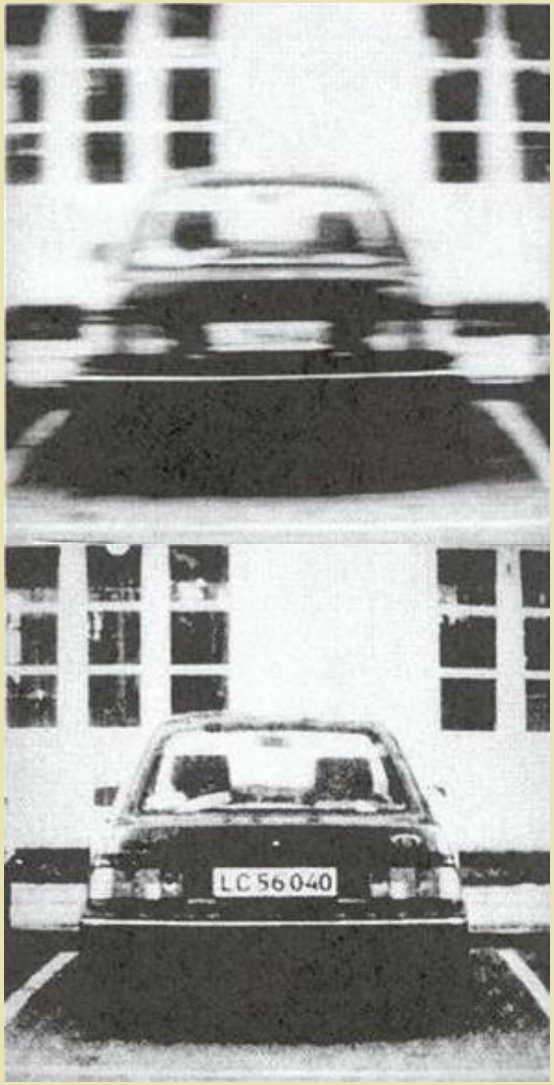
Low level image processing



- Image compression
- Noise reduction
- Edge extraction
- Contrast enhanceme
- Segmentation
- Thresholding
- Morphology
- Image restoration

Low level image processing

- Image compression
- Noise reduction
- Edge extraction
- Contrast enhancement
- Segmentation
- Thresholding
- Morphology
- Image restoration



Deep Learning

Deep Learning demo

(using Tiny YOLO on a Tensorflow architecture)

https://modeldepot.github.io/tfjs-yolo-tiny-demo

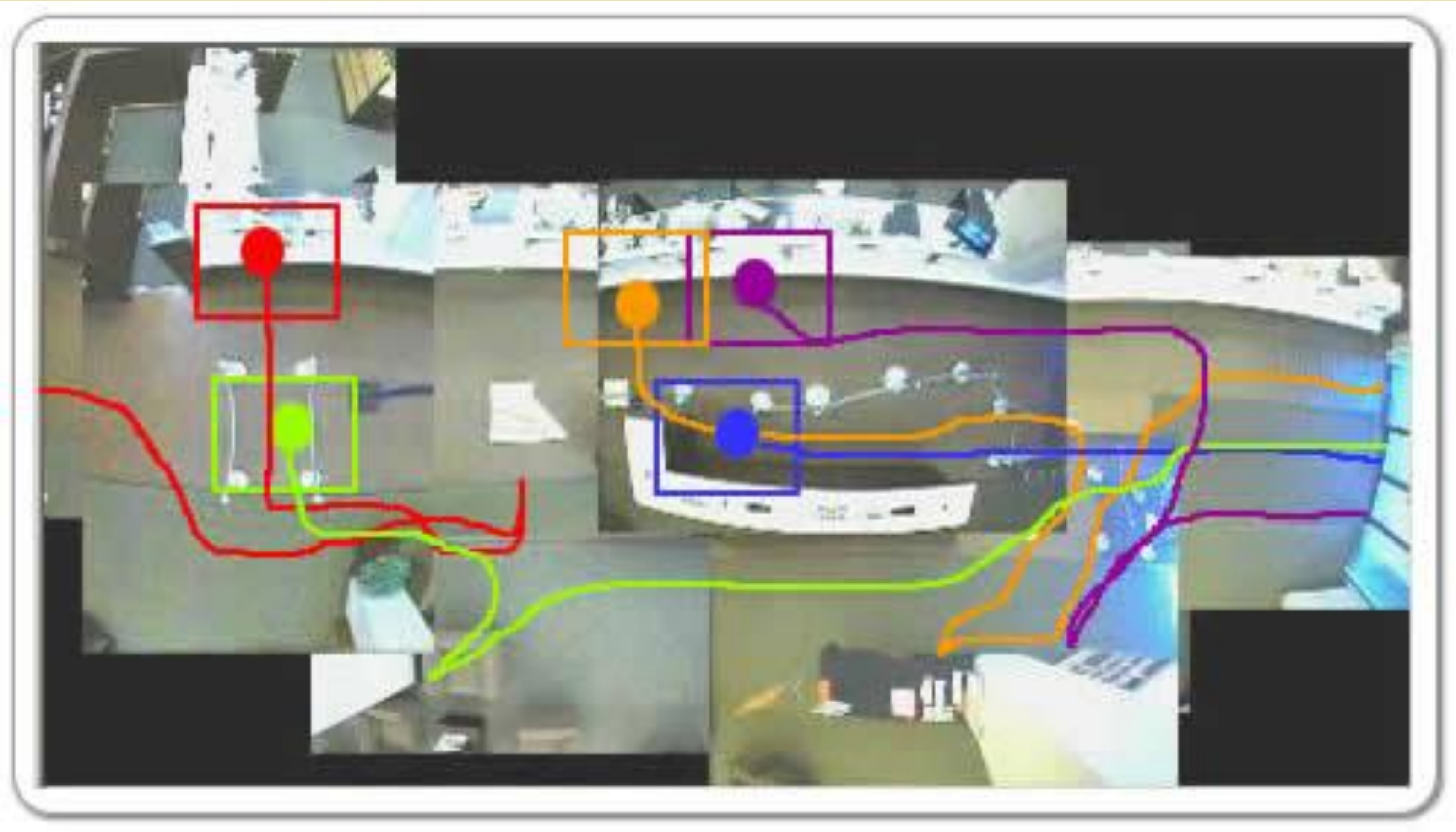
Why is Vision Interesting?

- Psychology
  - ~ 50% of cerebral cortex is for vision.
  - Vision is how we experience the world.
- Engineering
  - Want machines to interact with world.
  - Digital images are everywhere.

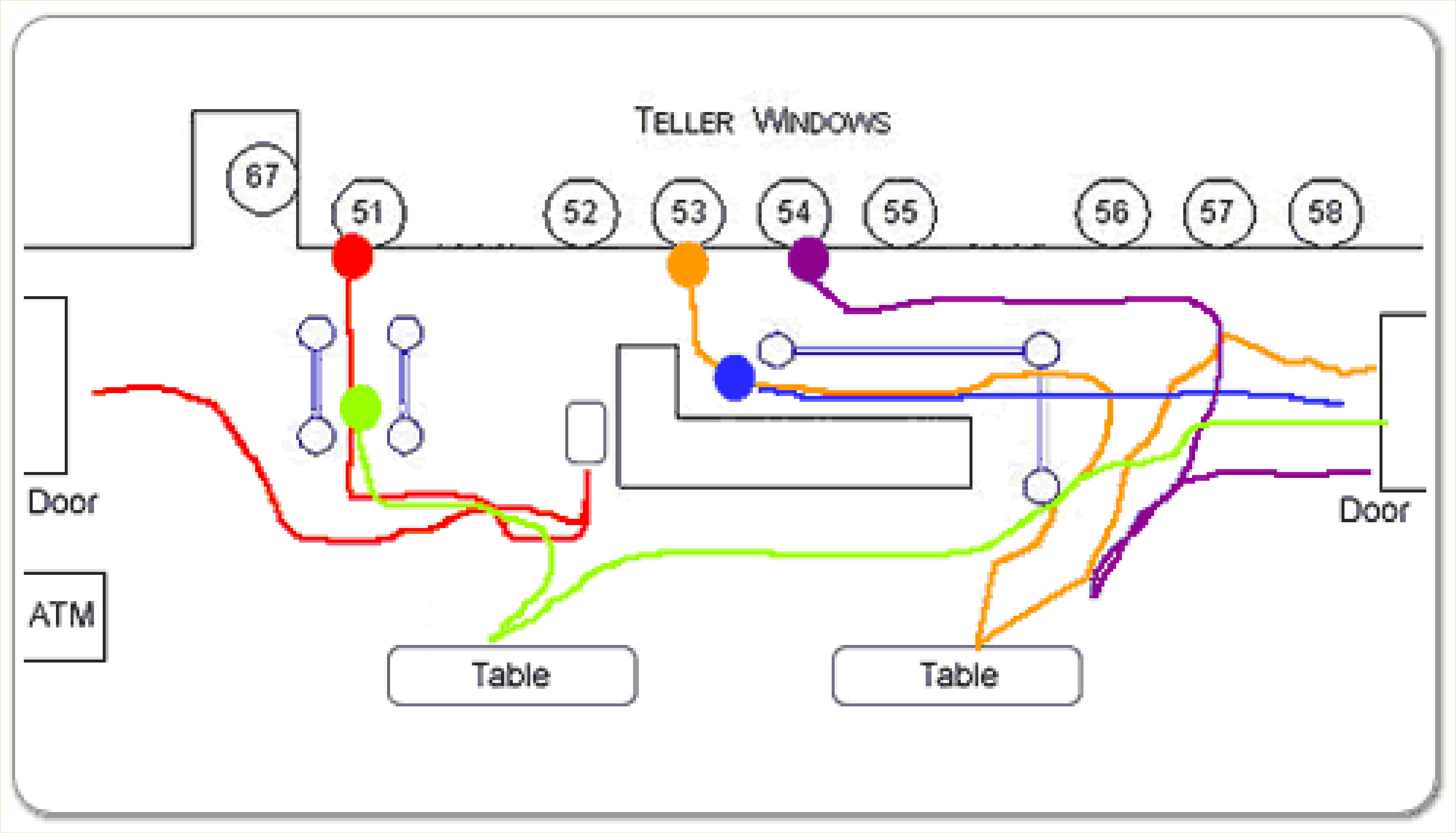
Tracking

(Comaniciu and Meer)

Tracking



Tracking

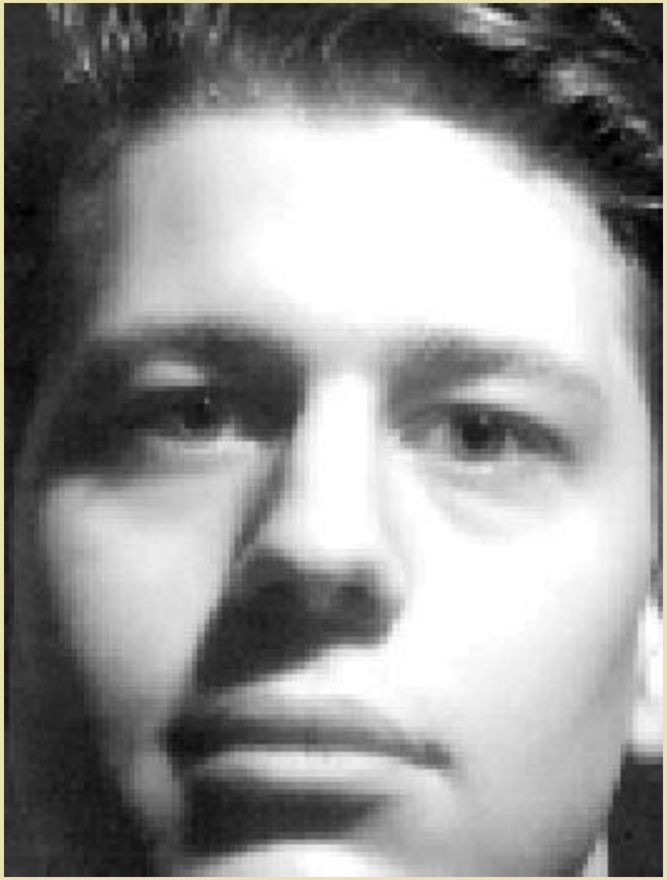


Teller Windows
67
51
52
53
54
55
56
57
58
Door
Door
ATM
Table
Table

Pose Determination

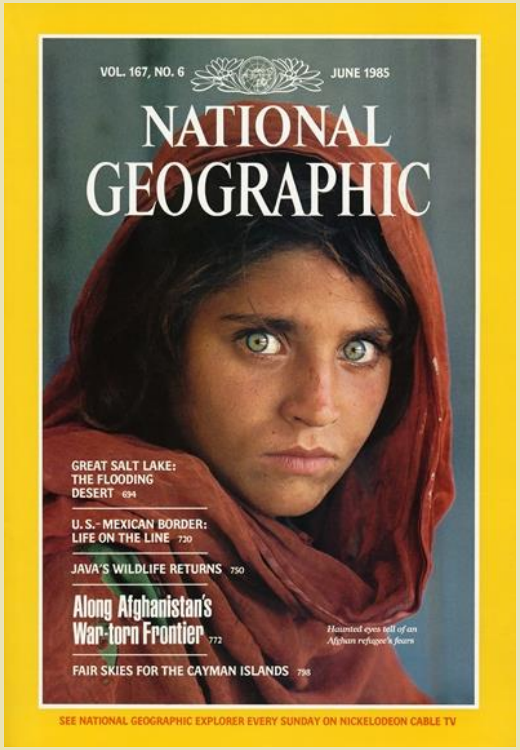
Visually guided surgery

Recognition - Shading

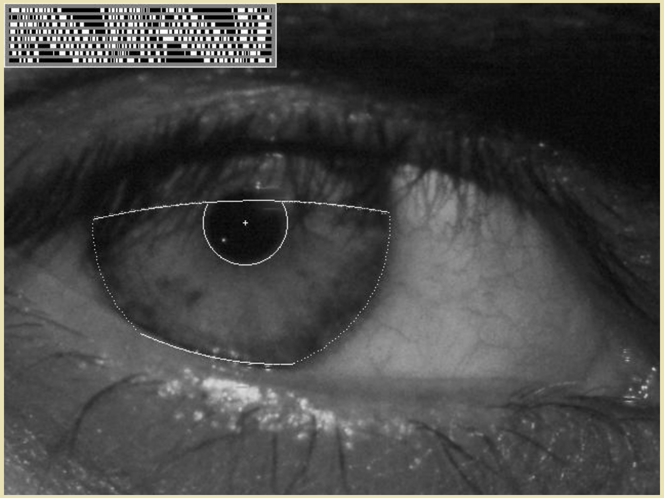


Lighting affects appearance

Vision-based biometrics



"*How the Afghan Girl was Identified by Her Iris Patterns*"



Special effects: shape capture


The Matrix movies, ESC Entertainment, XYZRGB, NRC

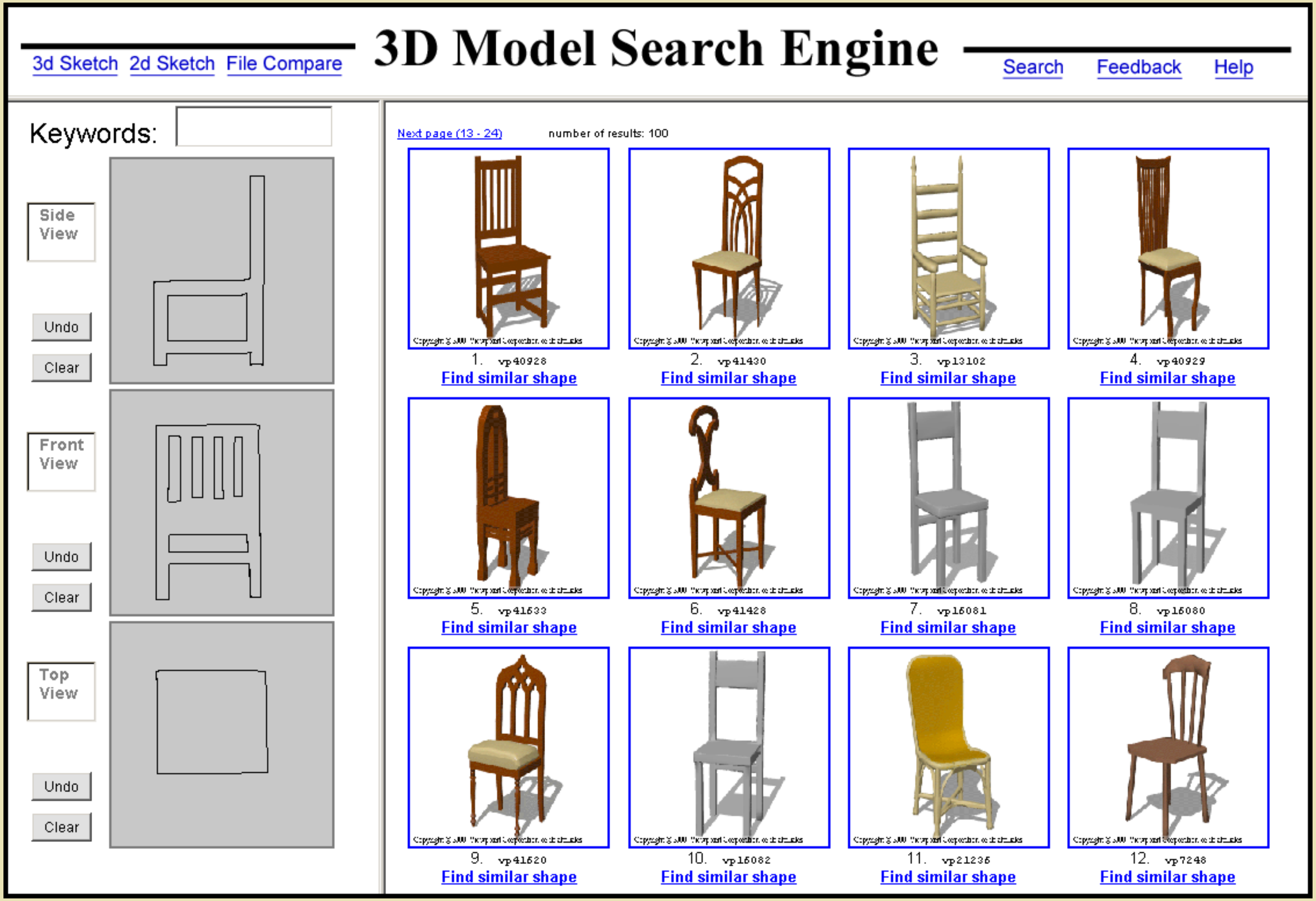
Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Sports

Sportvision first down line

Classification



(Funkhauser, Min, Kazhdan, Chen, Halderman, Dobkin, Jacobs)

Vision depends on:

- Geometry
- Physics
- The nature of objects in the world

  (This is the hardest part).

Approaches to Vision

Modeling + Algorithms

- Build a simple model of the world
  (eg., flat, uniform intensity).
- Find provably good algorithms.
- Experiment on real world.
- Update model.

Problem: Too often models are simplistic or intractable.

Engineering

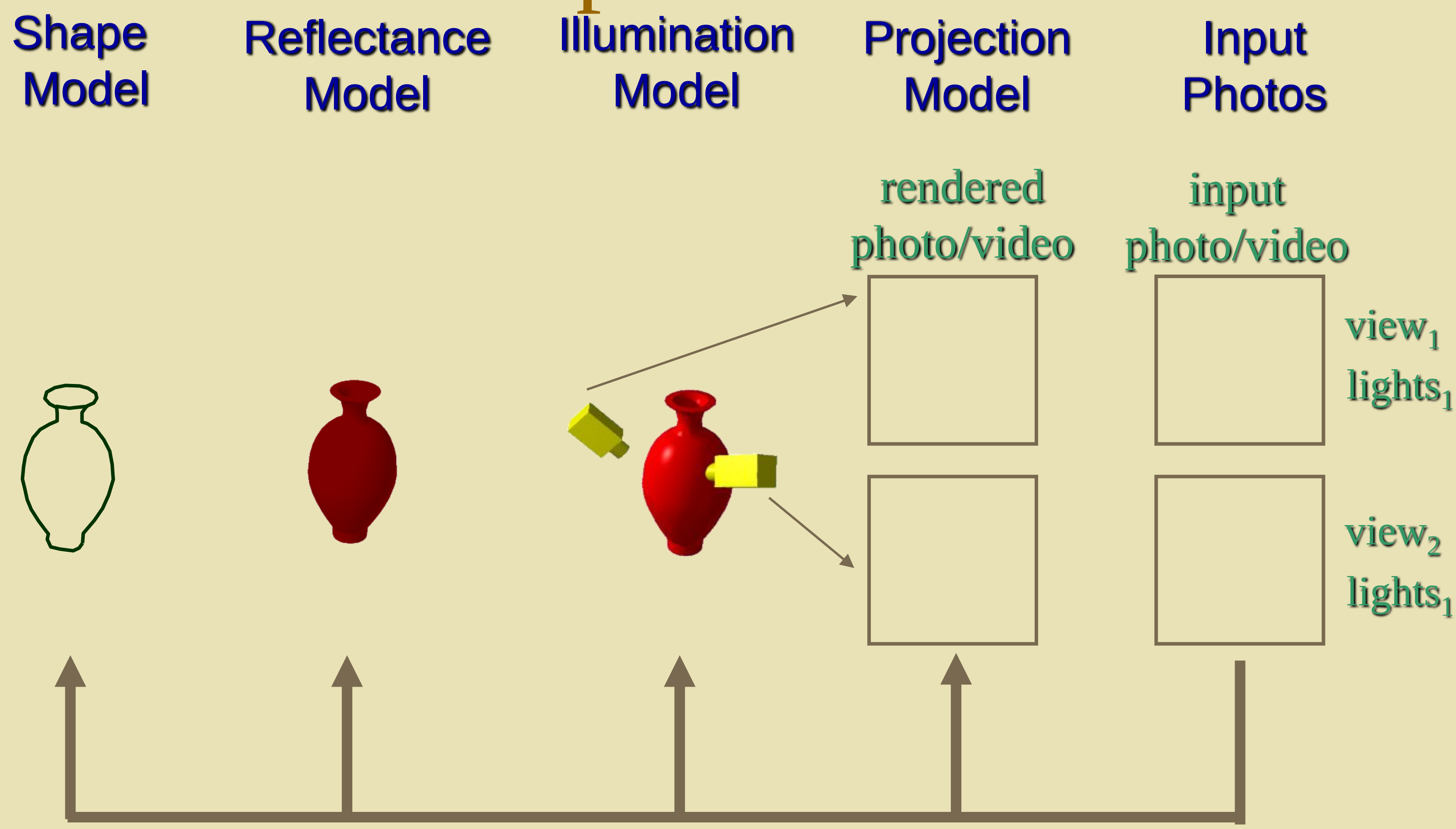
- Focus on definite tasks with clear requirements.
- Try ideas based on theory and get experience about what works.
- Try to build reusable modules.

Problem: Solutions that work under specific conditions may not generalize.

Related Fields

- Graphics. "Vision is inverse graphics".
- Visual perception.
- Neuroscience.
- AI / machine learning
- Math: eg., geometry, stochastic processes.
- Optimization.

Computer Graphics vs. Computer Vision



Is Graphics the Inverse of Vision?

Forward Modeling
(a.k.a. "traditional" graphics)

geometry, physics
computer algorithms

↓

synthetic
photos

Inverse Modeling
(a.k.a. "traditional" vision)

real
photos

↓

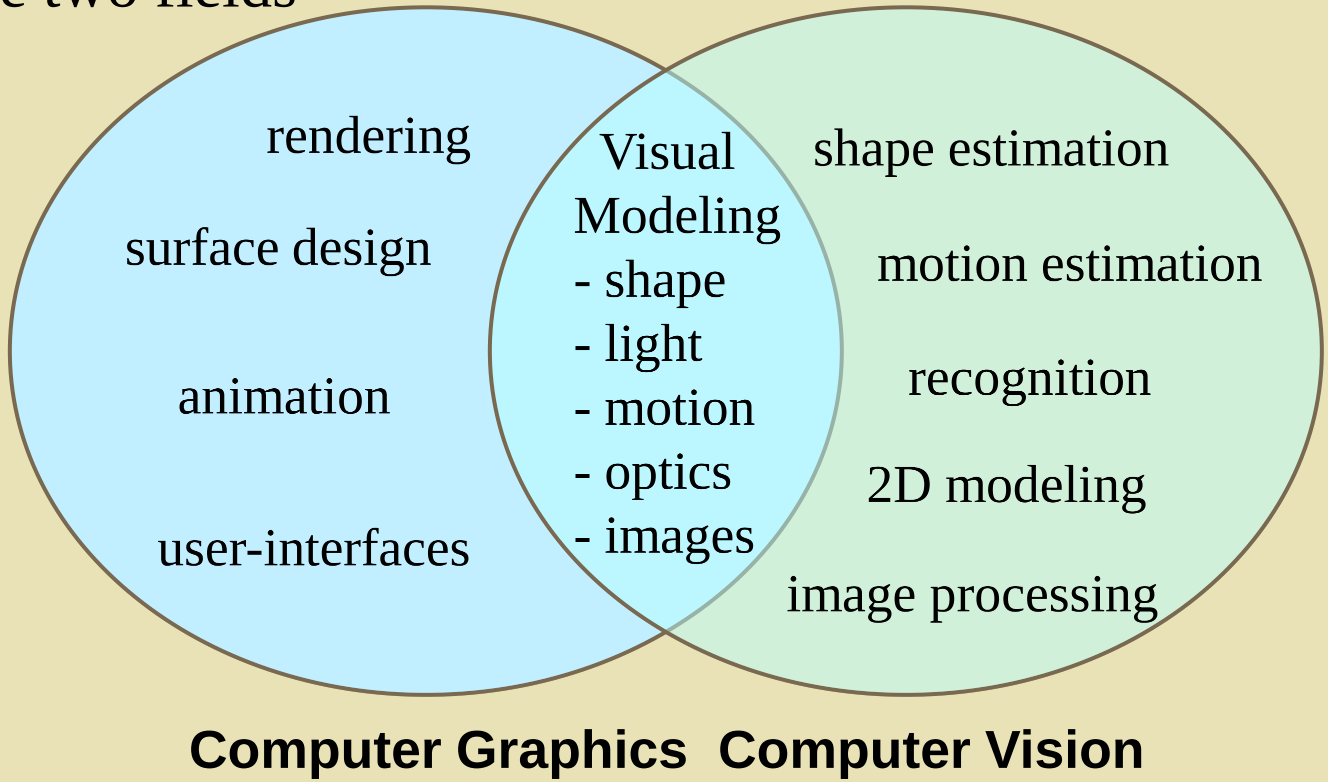
geometry, physics
computer algorithms

↓

synthetic
photos

Answer: Not Exactly...

- Traditionally CG & CV were posed as inverse of each other
- A great deal of cross-fertilization & intersection recently
- Visual Modeling Course: Study major topics at intersection of the two fields



Vision & Graphics

- **computer vision impacts graphics**
  - image-based rendering
  - model acquisition
  - motion capture
  - perceptual user interfaces
  - special effects
  - image editing

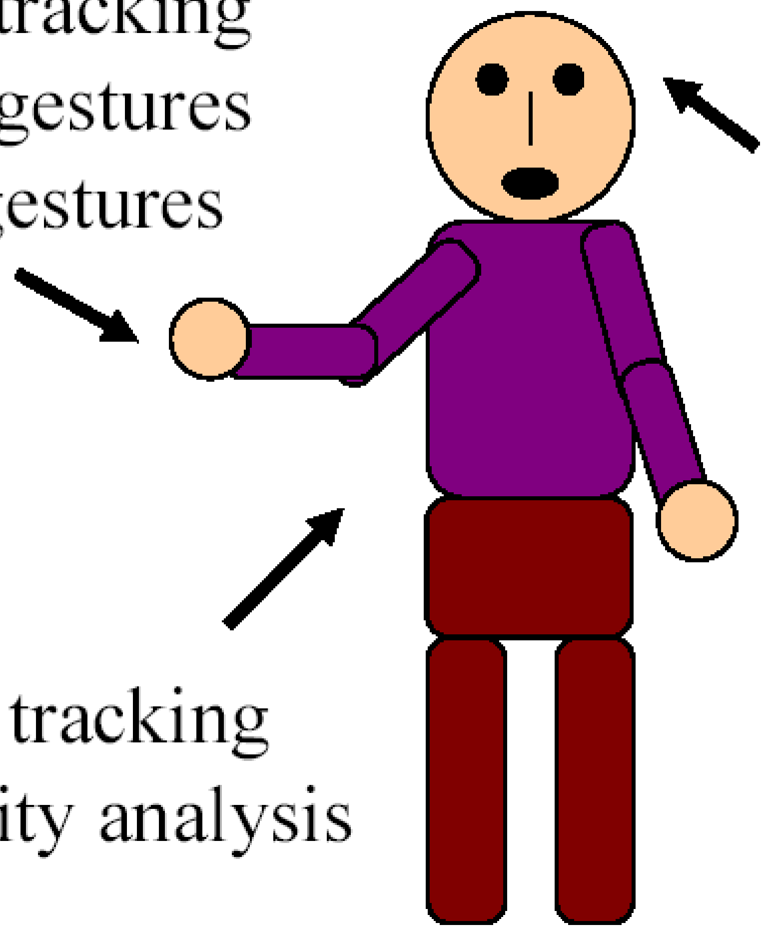
- **graphics impacts computer vision**
  - reflectance
  - transparency
  - shape modeling

- **Cross-fertilization has already enhanced significantly our understanding of the principles underlying the two fields**

Relationship of Computer Vision to Other Fields

Human Computer Interfaces – Vision Based Interfaces

Hand tracking
Hand gestures
Arm gestures

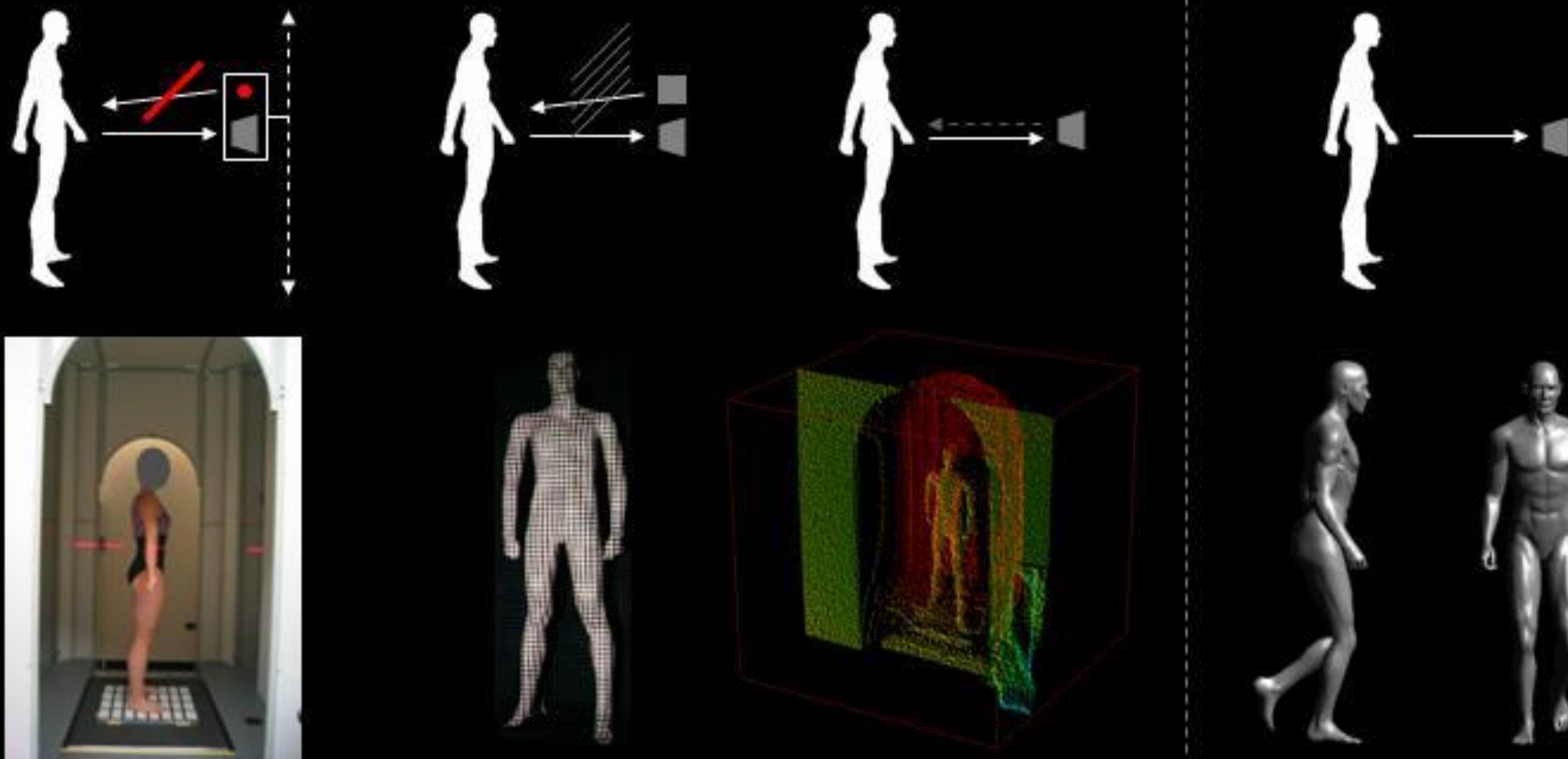


Head tracking
Gaze tracking
Lip reading
Face recognition
Facial expression

Body tracking
Activity analysis

Vision systems

- active
- passive



laser scanner | structured light | time-of-flight | images

Depth Imaging Technologies

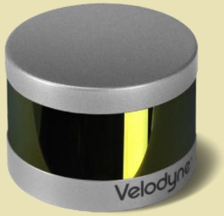
Stereo Vision

Active Stereo

Structured Light

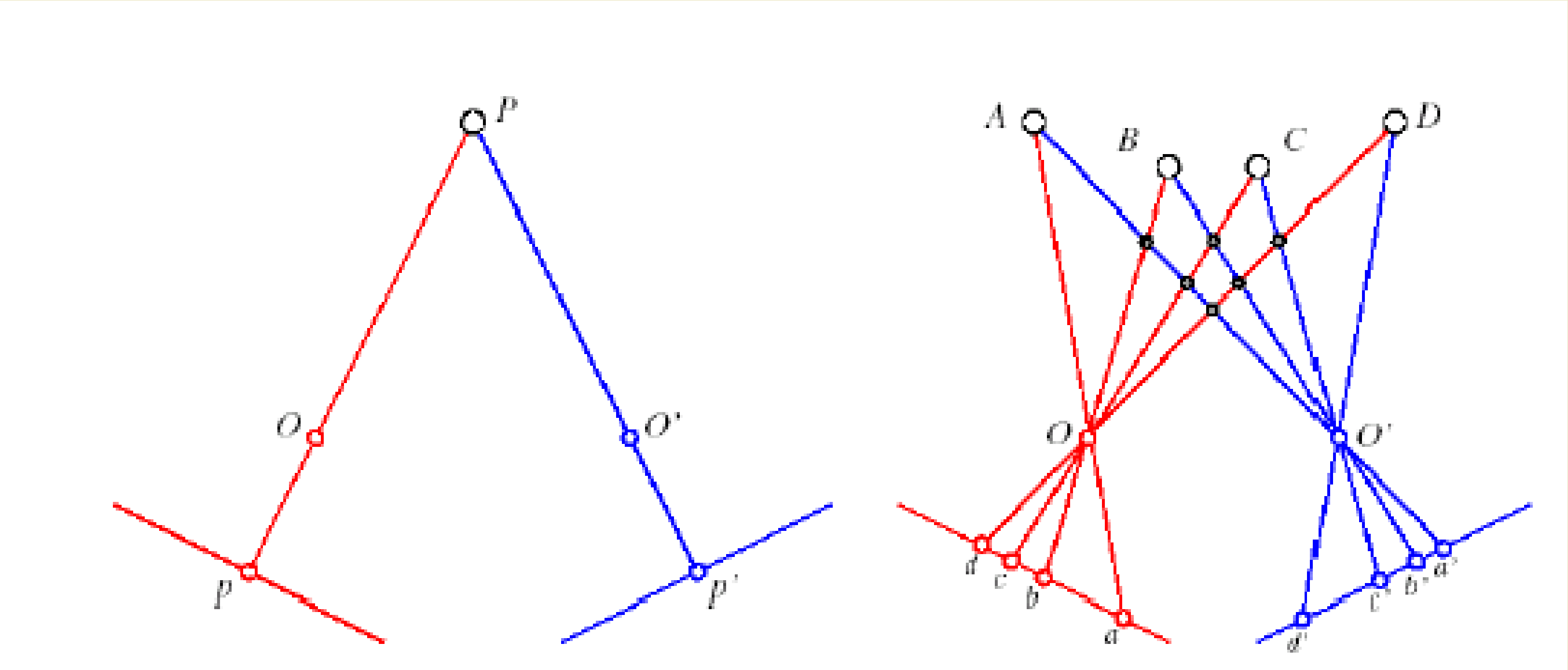
Time Of Flight
(RF-Modulated)

Time Of Flight
(Direct Timing)



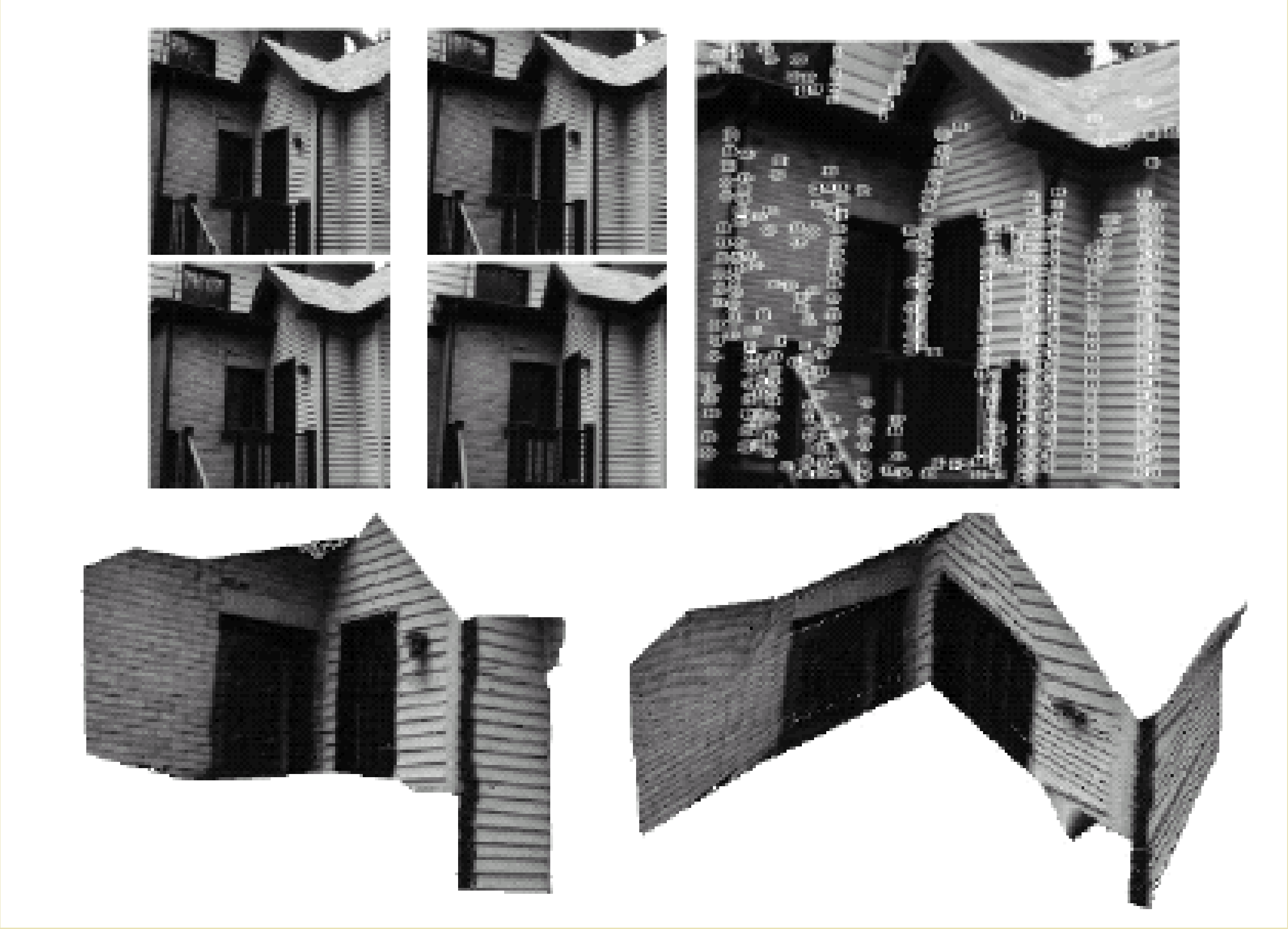
Intel Realsense D435 camera: colour stereo=RGB-D

Stereo



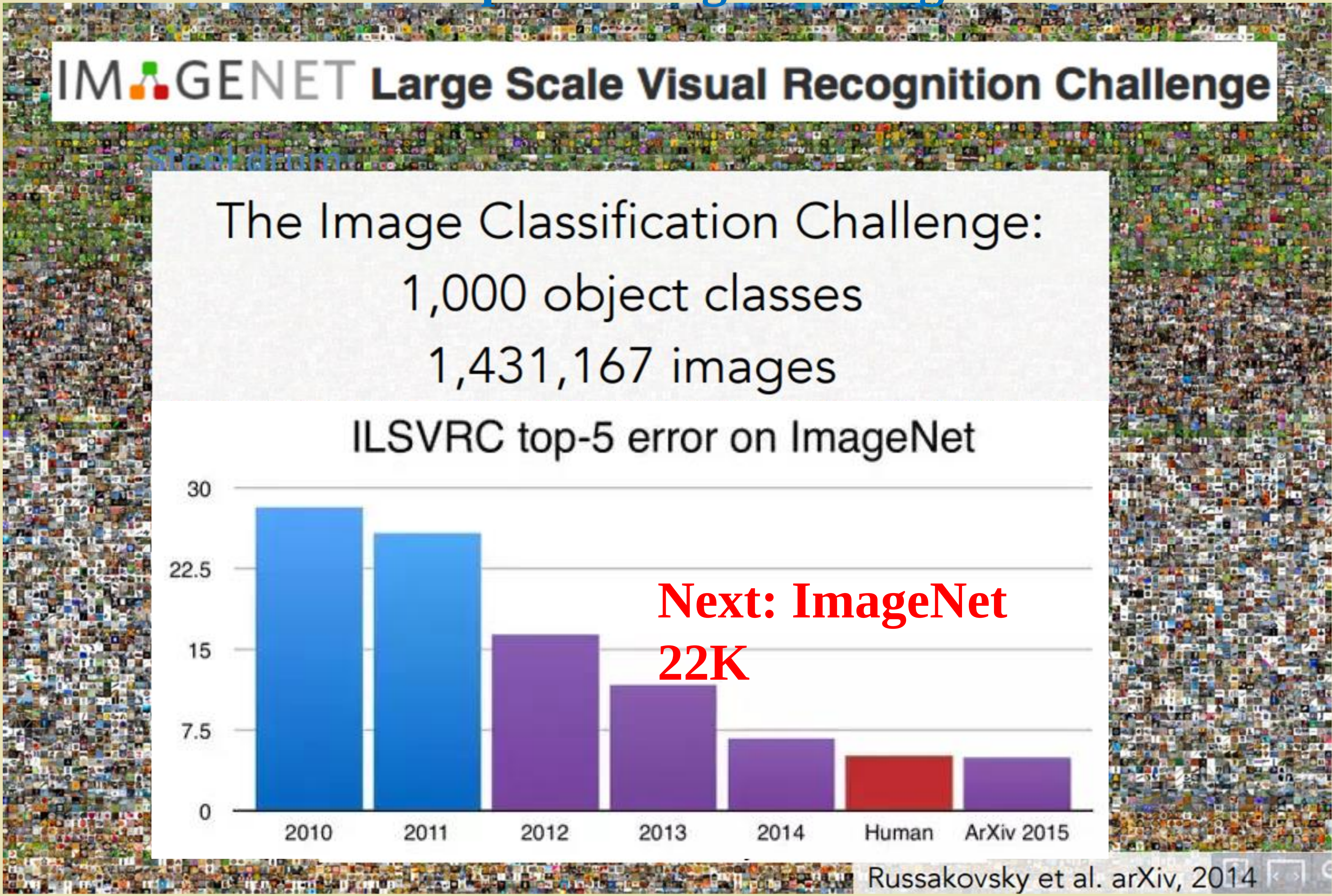
http://www.ai.mit.edu/courses/6.801/lect/lect01_darrell.pdf

Structure from Motion / 3D Reconstruction



http://www.ai.mit.edu/courses/6.801/lect/lect01_darrell.pdf

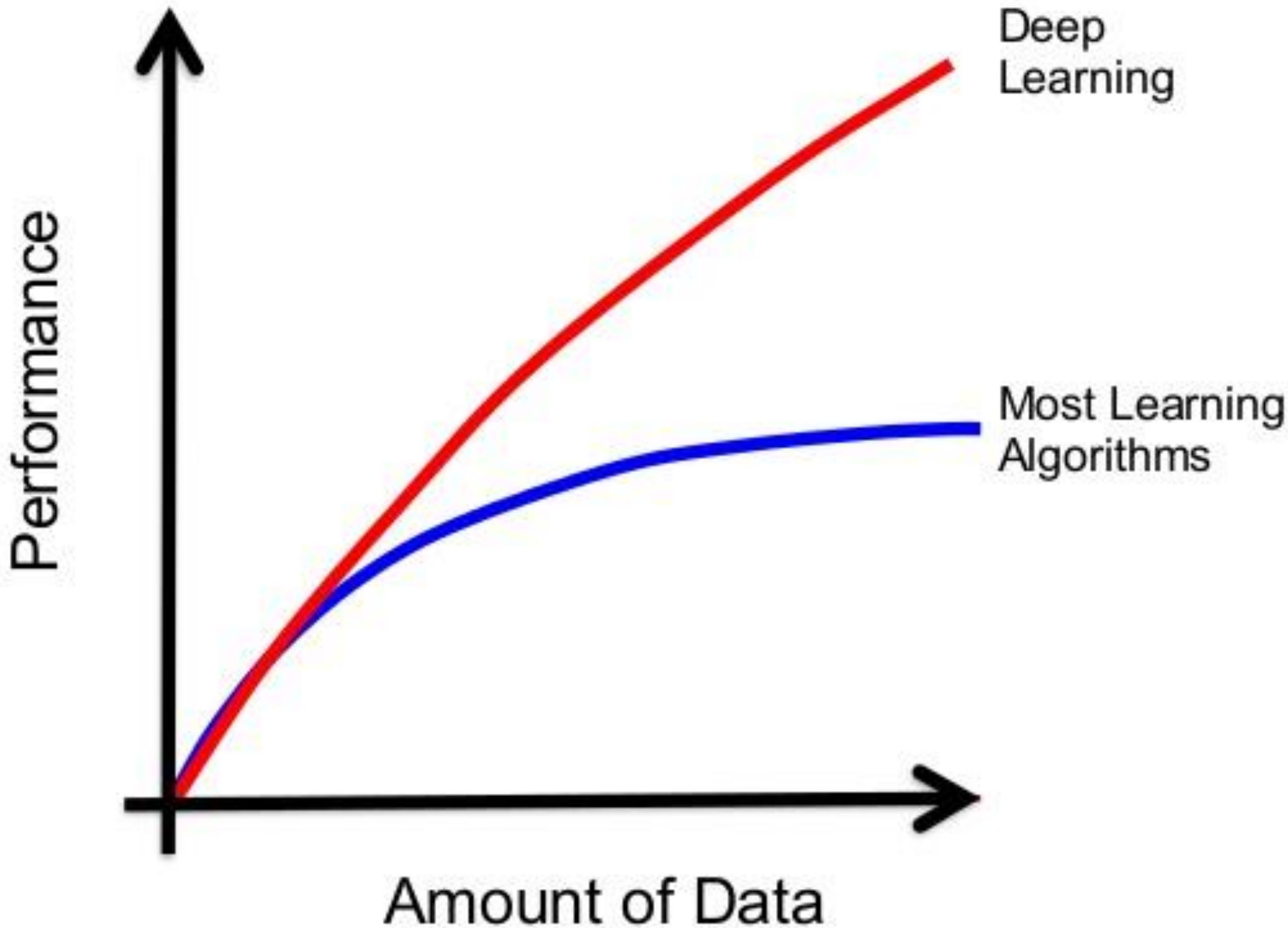
Smart Computer Vision needs Deep Learning Learning



Next: ImageNet 22K

Deep Learning Learning

BIG DATA & DEEP LEARNING



Deep Learning in Real-time

- ~NZ$100 Neural Network Compute Stick from Movidius (Intel) has 100Gflops of computing power.
- Even runs complex deep learning in real-time on a Raspberry Pi.

Why is Computer Vision Interesting?

- Psychology
  - ~ 50% of cerebral cortex is for vision.
  - Vision is how we experience the world.
- Engineering
  - Want machines to interact with world.
  - Digital images are everywhere.

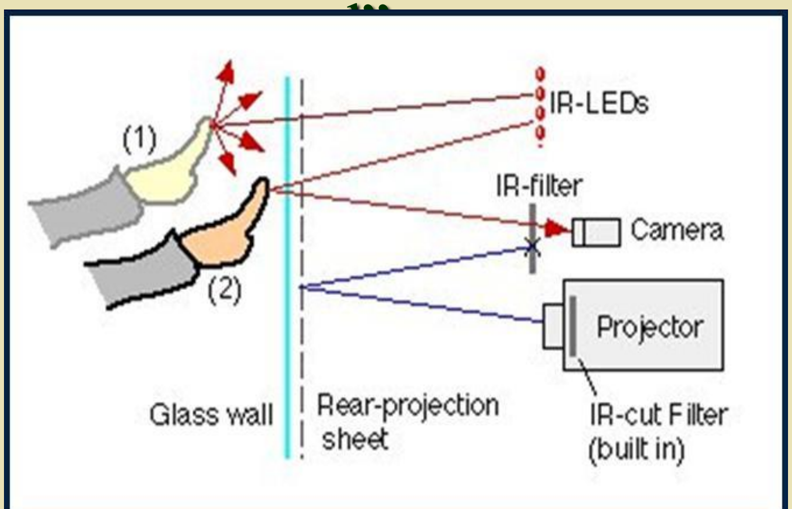
Large-scale Display for Collaborative Interactions

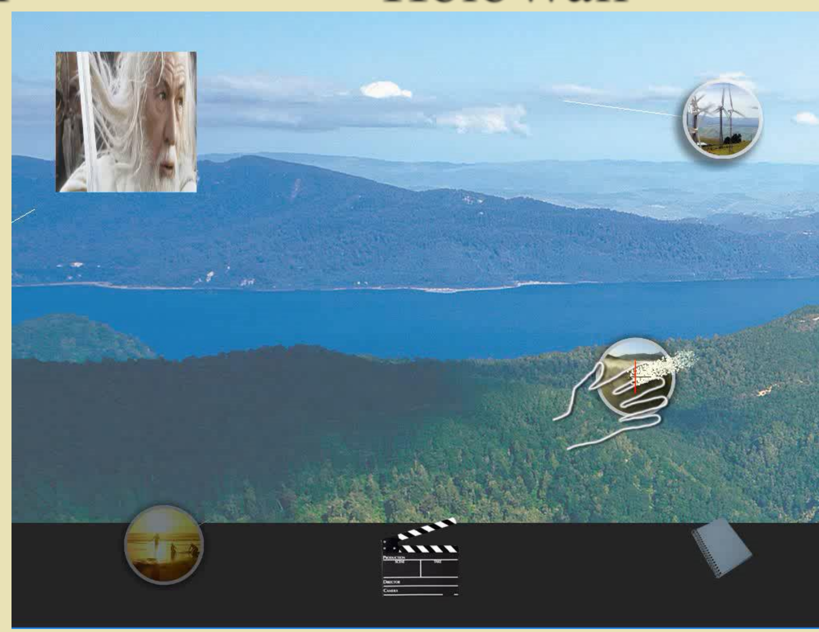
(Over 1 million at Expo NZ exhibit)

Touch sensitive
“SMARTBoar
d”

Single hand
“DynaWall”

Multi-hand
“HoloWall”





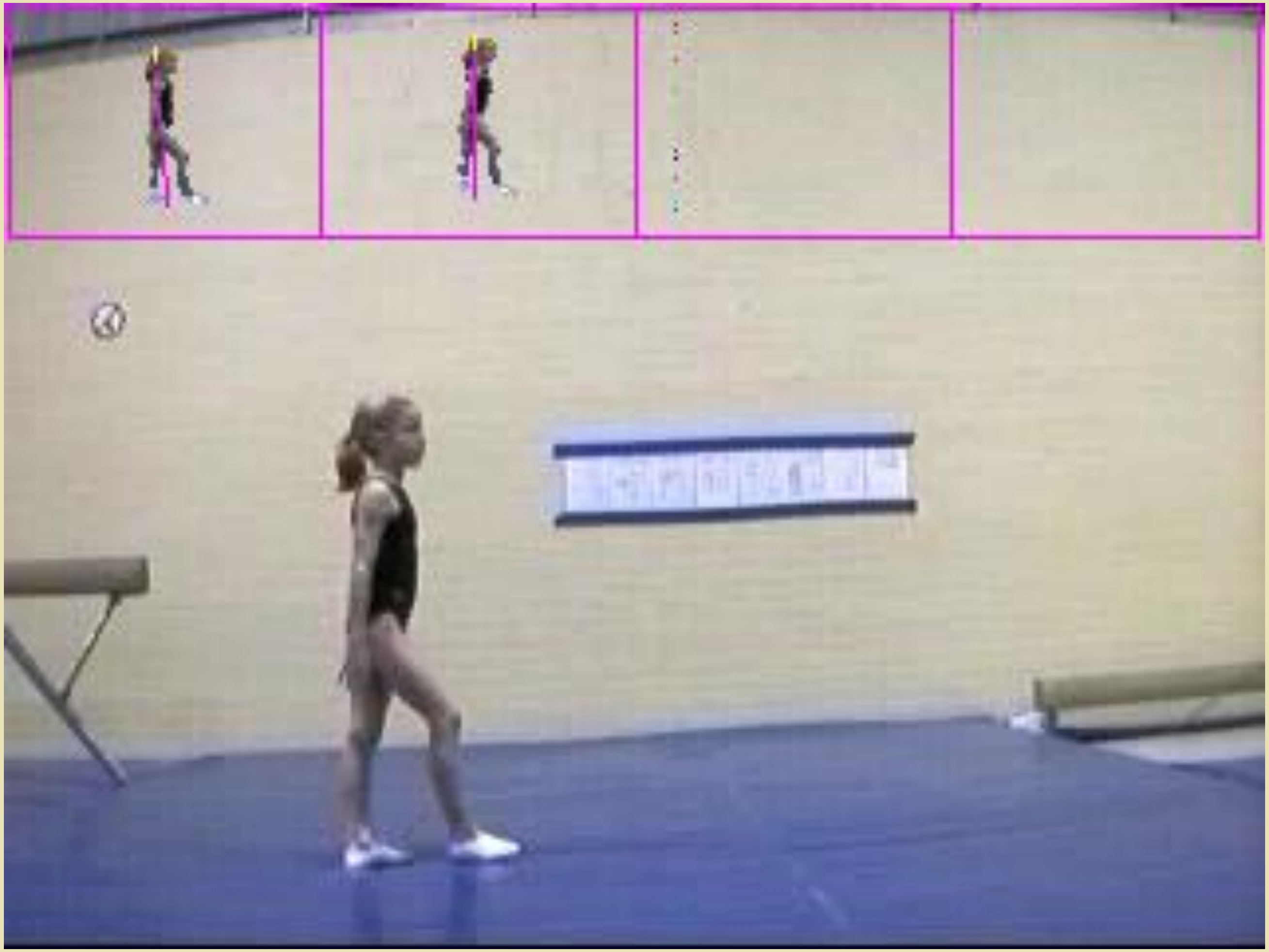
Multi-hand gesture based collaboration

Instead of
tracking with joint markers
to animate Gollum...

Use Computer Vision with stereo cameras to animate and interact

(Over 30,000 at Boston Museum of Science and Technology)

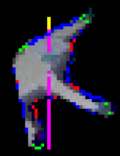
Recognising Actions

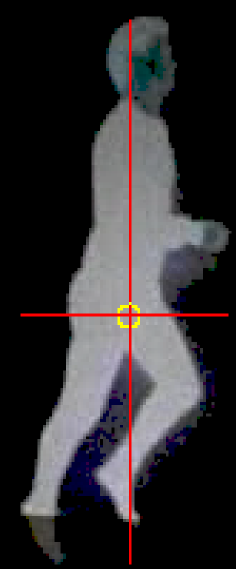


Skill recognition

95.5% using the Microsoft / Cambridge University Engineering Dept. HMM Tool Kit
Viterbi alignment on the training data, Baum-Welch re-estimation, context model
(won best IEEE Transactions journal paper)

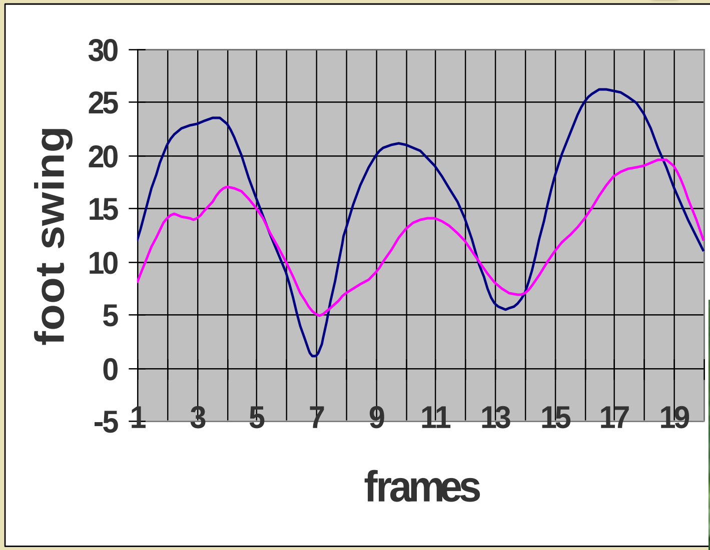
Analysing human motion





Real-time Parkinsonism gait analysis enables treatment to be quantified.

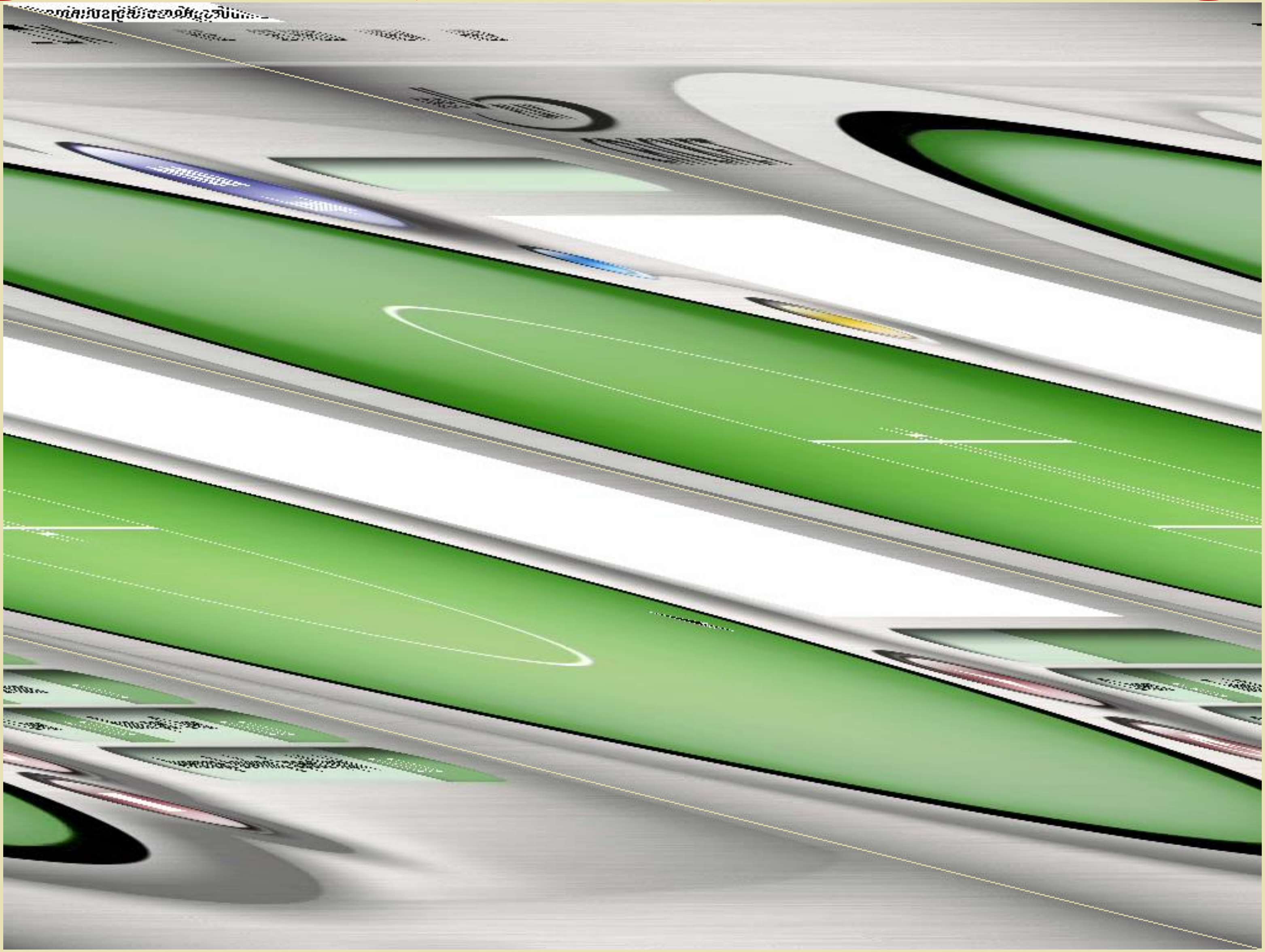
Asymmetrical leg swing typical of Parkinsonism gait.



% correlation	Abnormal gaits	False negative	Normal gaits	False positive
Leg amplitude	55	0	33	11
Arm amplitude	50	6	33	11
Gait period	28	28	44	0
L/R asymmetry	50	6	44	0

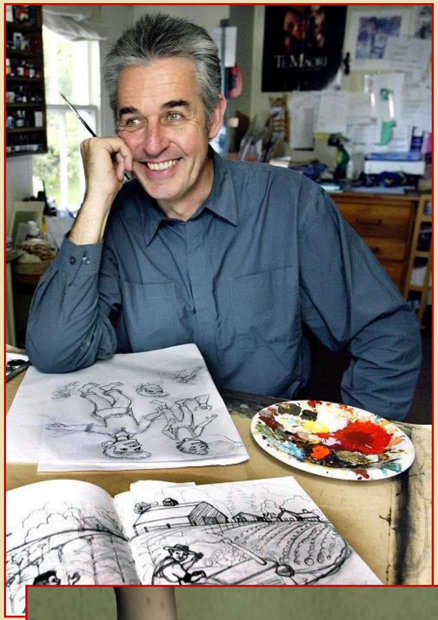
Correlation of swing amplitude, period, and left-right asymmetry

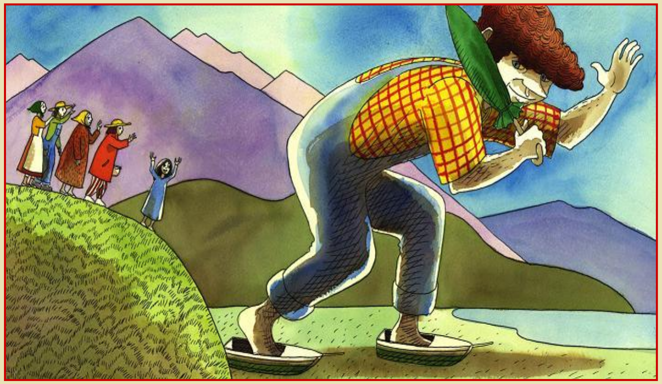
Tracking cricket balls to automatically generate Wagon wheels, scores, line and length



MagicBook project

- Children's Augmented Reality book
  - Collaboration with Gavin Bishop
  - Natural feature tracking
  - Educational tool





Boss - 1st prize DARPA Urban Challenge

- Carnegie Mellon/GM Team

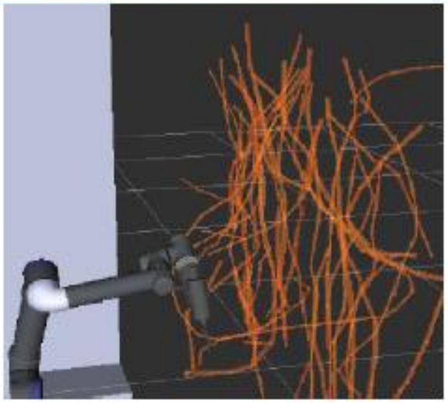
First Grand Challenge for 15 autonomous cars over a 142 mile desert course - **no one finished**

Grand Challenge, four completed a 132-mile desert route under the required 10 hour limit. DARPA awarded the $2 million prize to Stanford University

DARPA Urban Challenge - 96 km urban area course, to be completed within 6 hours obeying all traffic regulations while negotiating with other traffic & obstacles and merging into traffic - $2 million winner was Carnegie Mellon Uni

Computer vision for precision agriculture

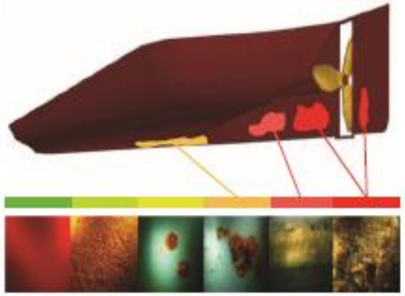
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND



prune vines

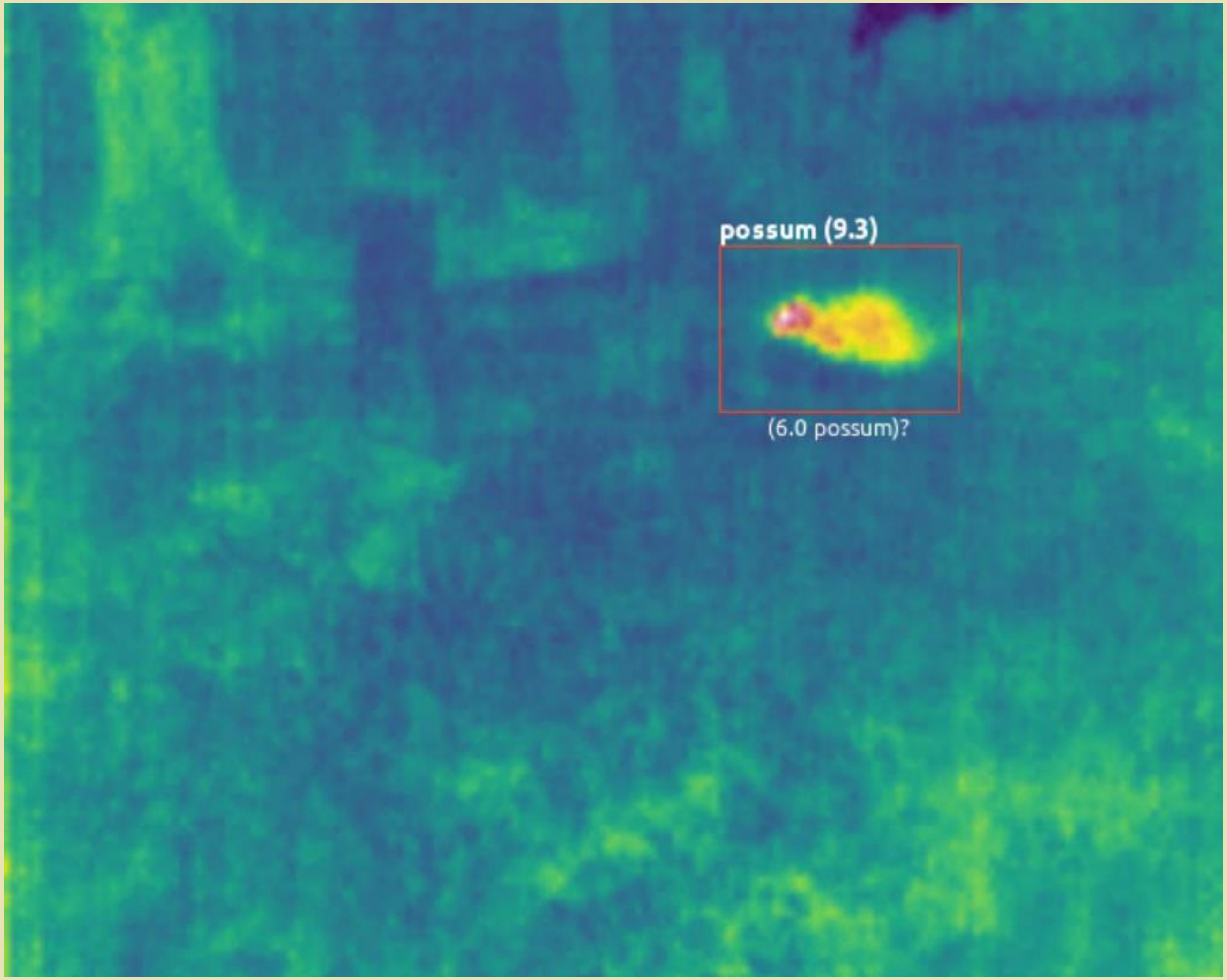
perch & prune forests

remote sensing++



detect biofouling

recognise activity

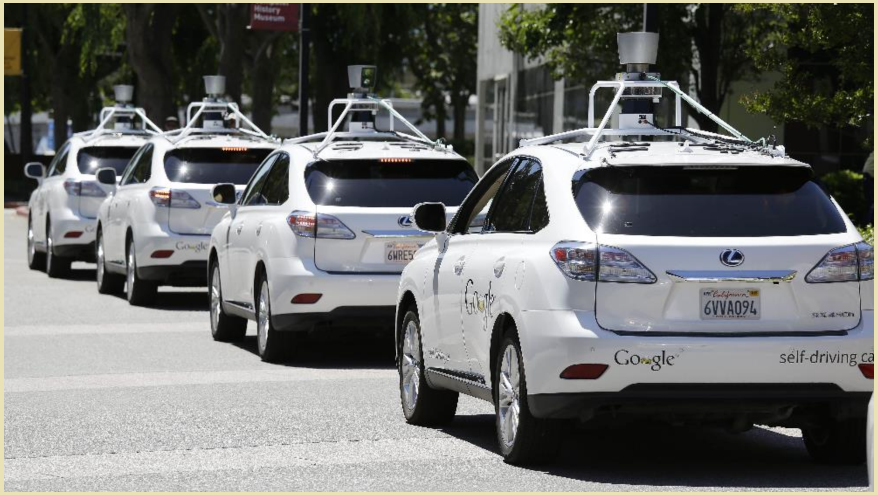

possum (9.3)
(6.0 possum)?

Self Driving Cars

- **Now?** Cars are already driving themselves on roads in California, Texas, Arizona, Washington, Pennsylvania, and Michigan - but restricted to specific test areas and driving conditions.

- **Autonomous?** Self-drive only in specific conditions - or - don't need steering wheel & brake pedal.

- **When?** 2032 to 2047 according to Mary Cummings, a professor at Duke University. But what she means by autonomous is a car that "**operates by itself under all conditions**, period."



Pruning Forests – NZ's 3rd largest export industry

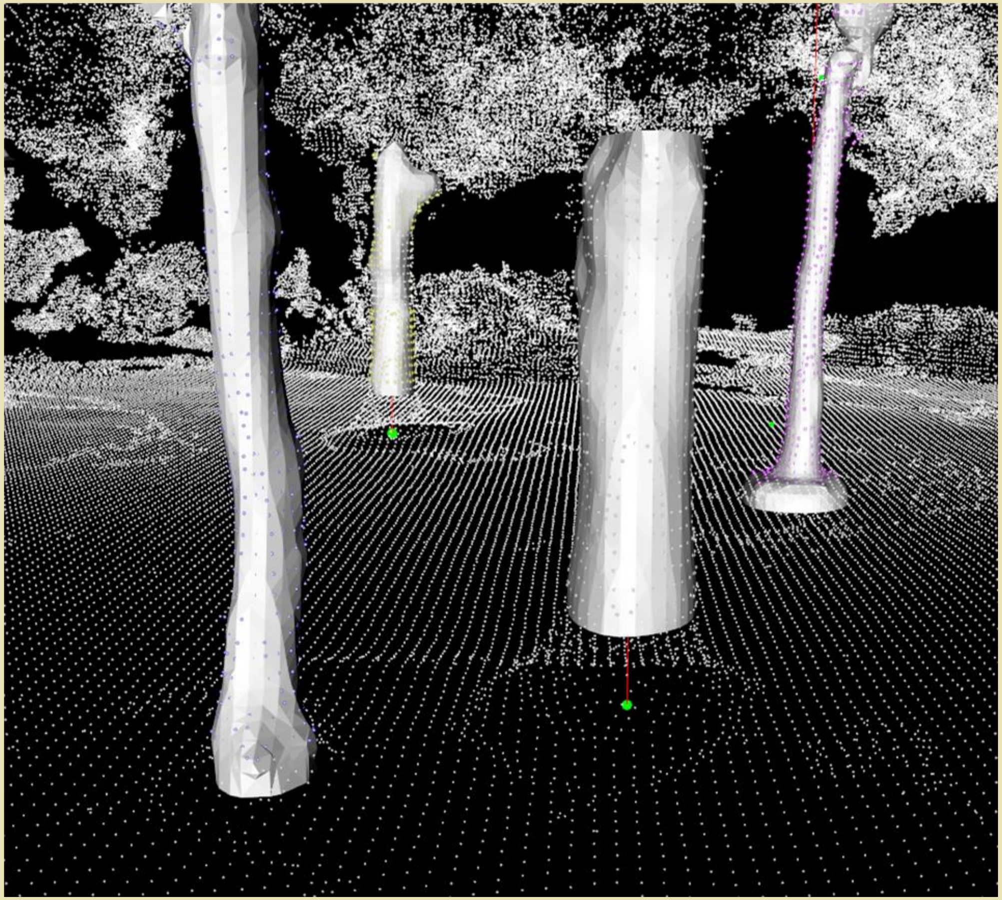
Autonomously prune pine trees using drones with depth sensors and electric pruner.

- **Collaborative UAV swarms** need **3D computer vision, swarm intelligence**, ad-hoc wireless communications, software engineering to certify commercially airworthy avionics code, flight control, etc. Depth sensors enable 99% accurate models of trees to support 3D navigation to each branch/cut point and an AI based expert pruner system.

Pruning Forests – NZ’s 3rd largest export industry

Deep learning to recognise trees

3D surface reconstruction for volume, etc

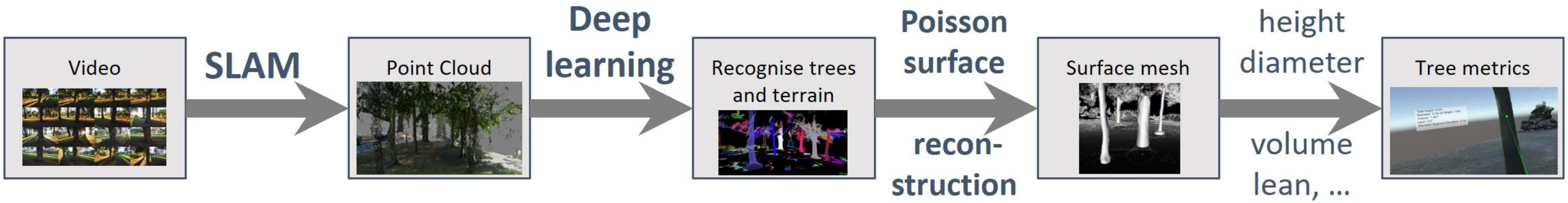


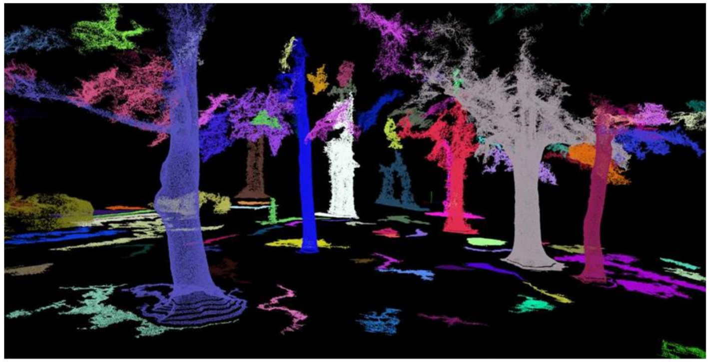
Autonomously generate semantic models

Intel D435 colour/depth camera

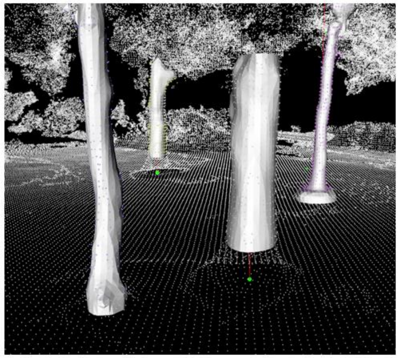
Colour & depth video of trees and terrain

3D point cloud

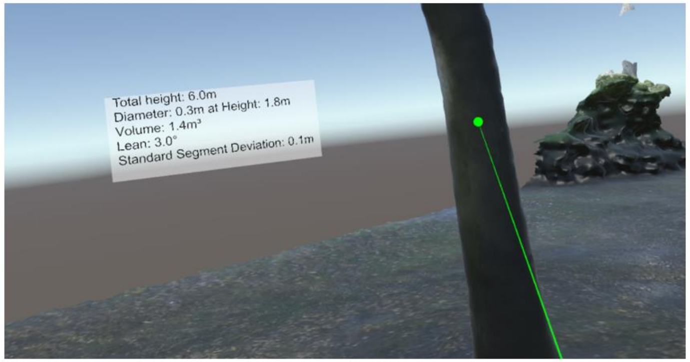




Recognised/segmented trees and terrain

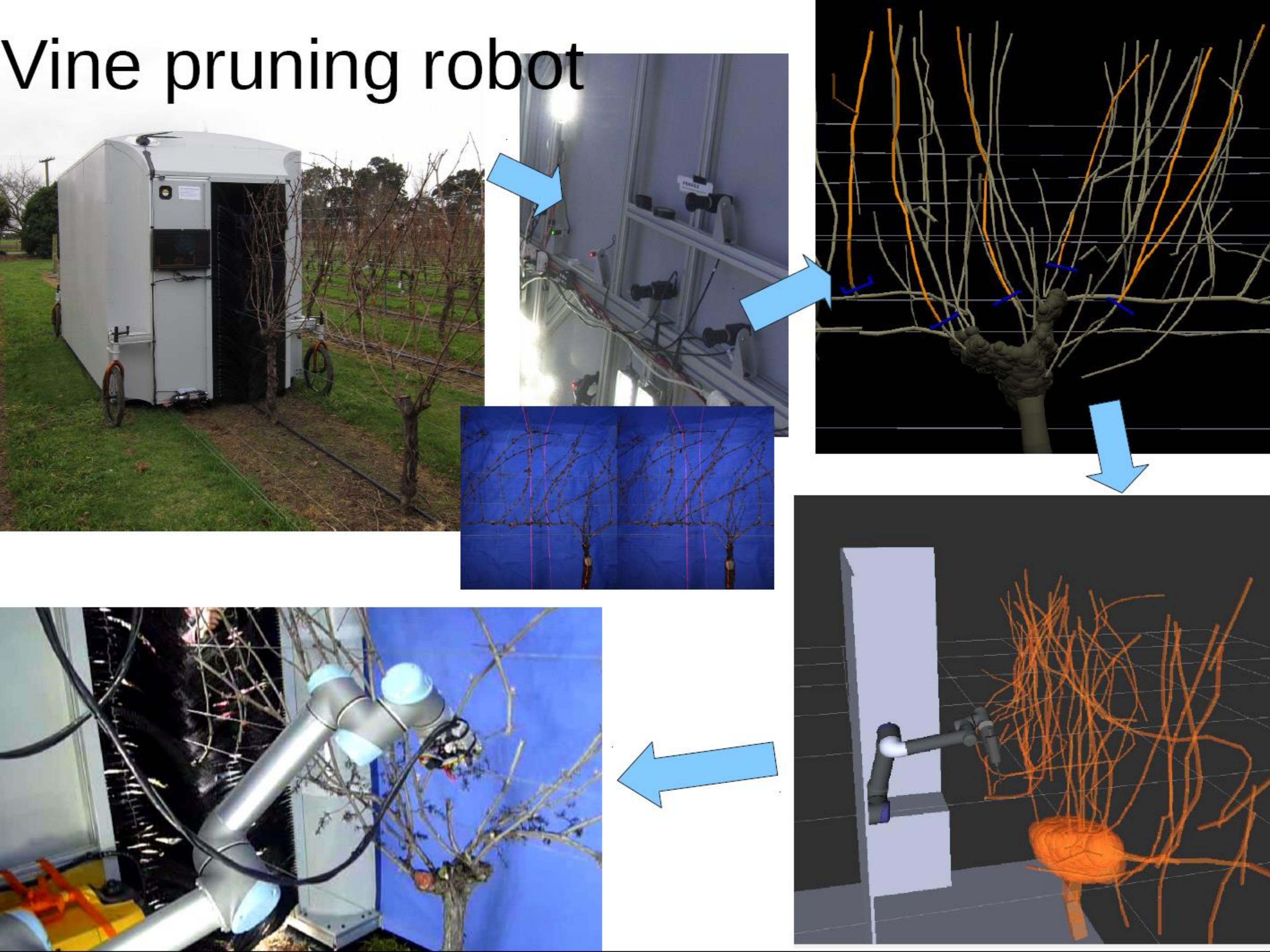


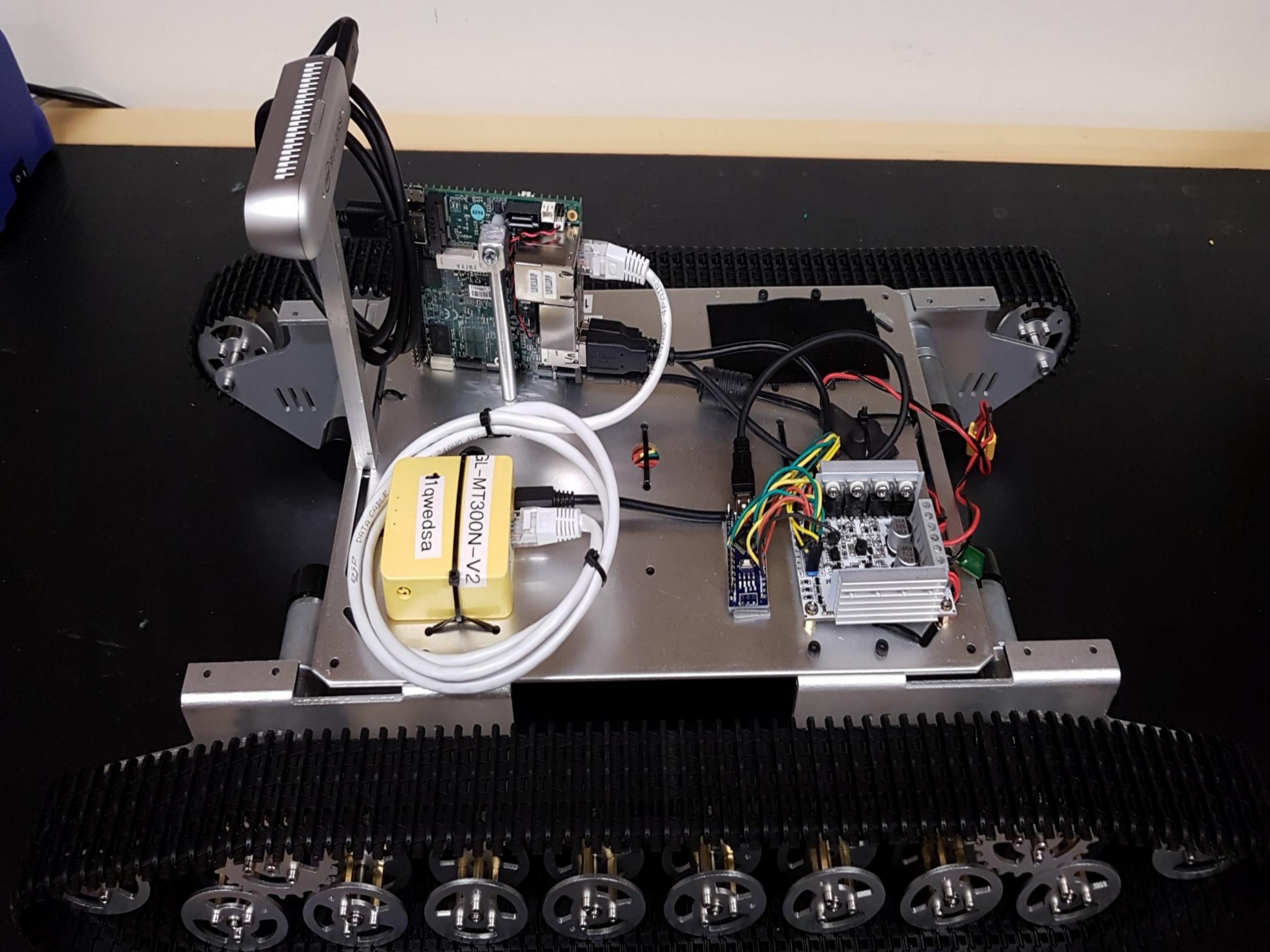
Surface mesh



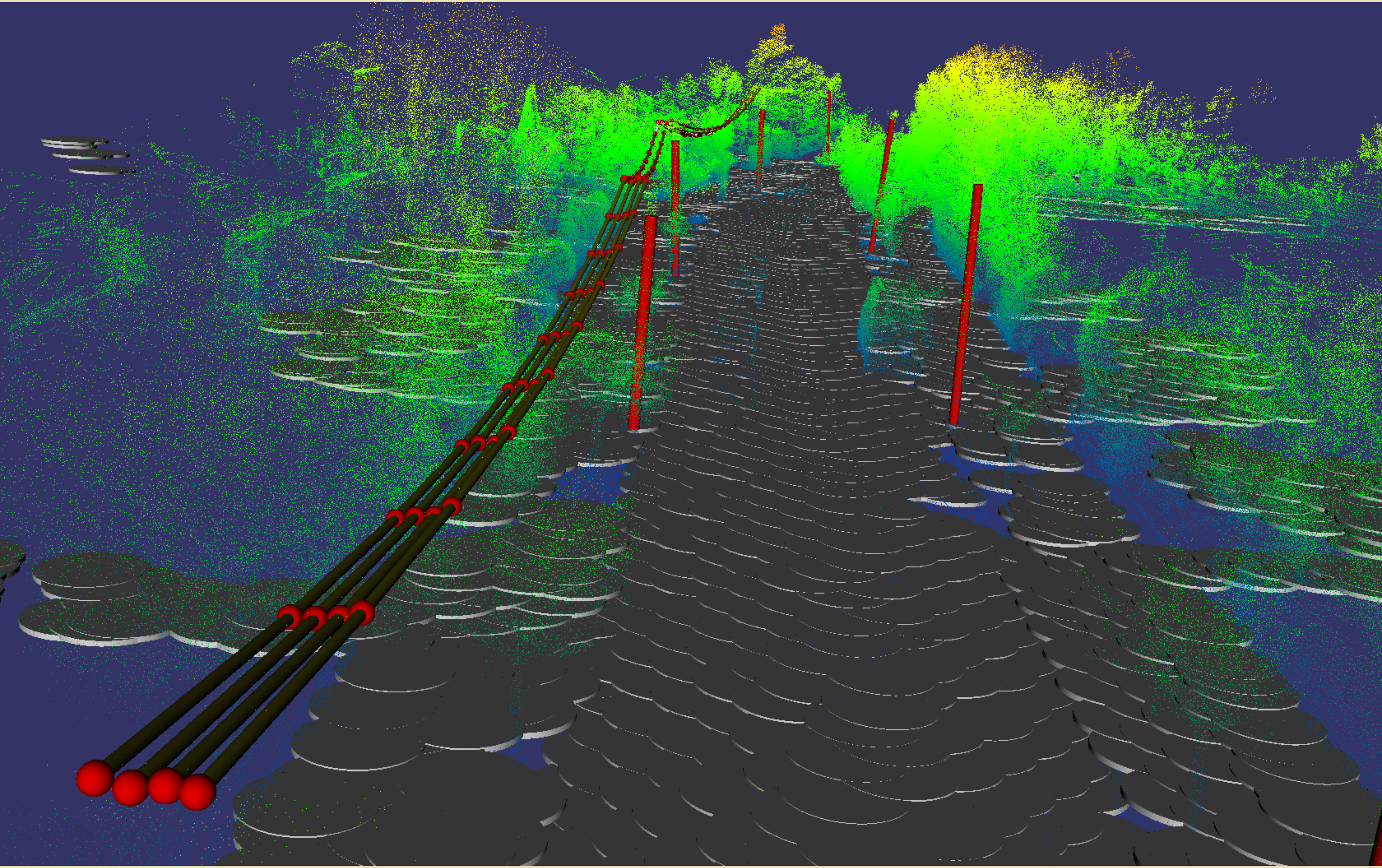
Tree metrics

Vine pruning robot

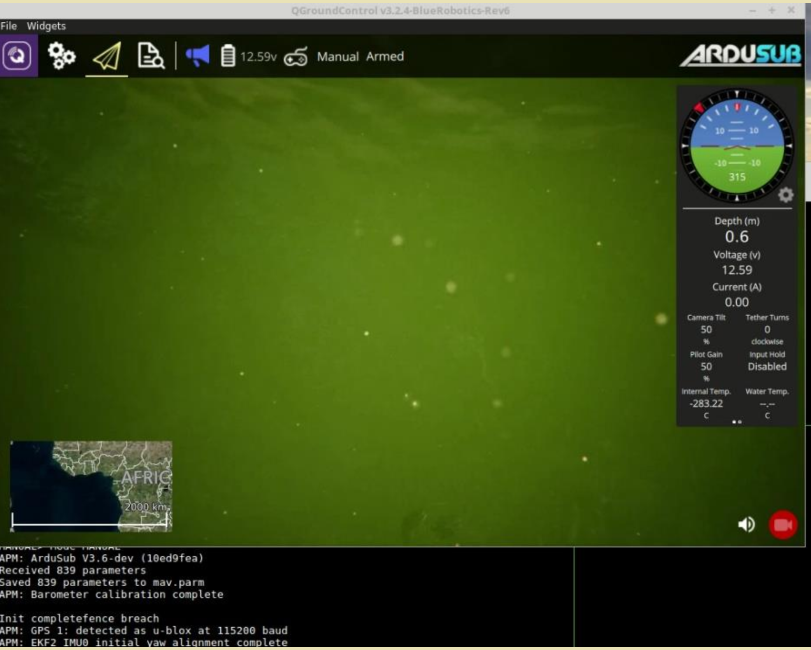


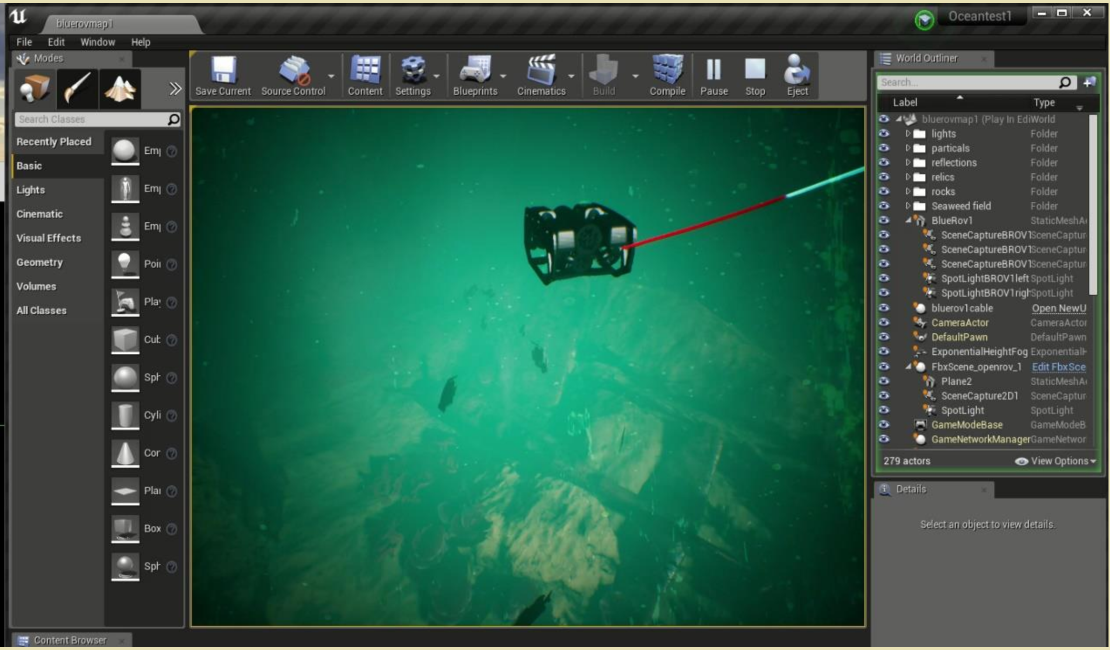

GL-MT300N-V2
11qwedsa

Recognising Poles, Wires and Trees

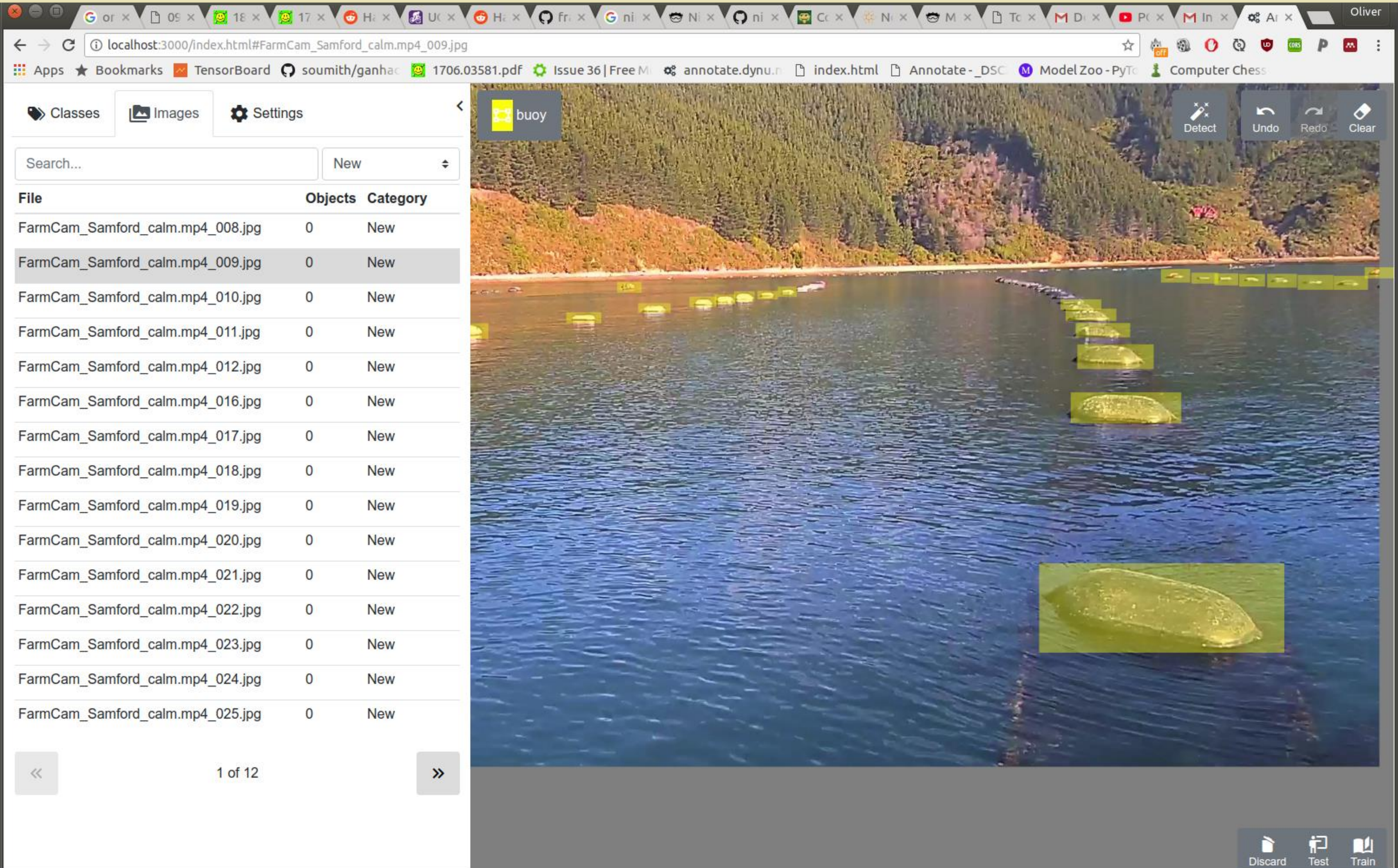


Autonomous Underwater Drones





Autonomous Underwater Drones



COSC428 Computer Vision

Overview

- Perception
- Image processing
- Multi-camera
- 3D vision
- Motion
- **Deep learning**

Example exam question

- Briefly describe advantages and/or disadvantages of the following four different types of camera technologies **for acquiring image depth values**.

  (1 mark for each advantage *or* disadvantage cited)

- structured light camera [3 marks]
- time-of-flight camera [3 marks]
- stereo camera [3 marks]
- LIDAR [3 marks]

Example exam question

For all depth cameras, reflective (e.g. wet) surfaces can cause noisy depth values.

Structured light:

- Cannot work in direct sunlight because the strong infra-red sunlight interferes with the low intensity projected infra-red camera light (low signal-to-noise ratio)
- Cannot work closer than 0.5m because the projected pattern of dots become too close together in the image.
- Cannot work further away than about 3.5m because the projected dots become too far apart and the intensity is too low.
- Motion blur occurs for fast motion because of the low intensity of the projected infra-red pattern of dots.
- Accuracy decreases with distance

Example exam question

Time of flight camera:

- Cannot work in direct sunlight because the strong infra-red sunlight interferes with the low intensity infra-red camera light (low signal-to-noise ratio)
- Limited range due to low intensity infra-red light
- Accuracy is independent of distance

Example exam question

Stereo camera:

- Potential for highest resolution
- Colour is also available for each pixel (as well as depth)
- Works well in direct sunlight.
- Noisy depth values in low ambient light.
- Works for motion (if well illuminated)
- Accuracy decreases with distance
- Many gaps in depth values in image regions without features (i.e. regions of uniform colour/intensity). Depth accuracy can be increased using higher resolution cameras.
- Depth accuracy over longer distances can be increased using a wider baseline.
- Pre-calibrated commercial cameras with good depth accuracy are expensive.
- Cheap cameras (e.g. webcams) need extensive calibration for useful depth accuracy.

Example exam question

LIDAR (Light Detection And Ranging):

- Good range (e.g. used for mapping ground from aircraft).
- Accuracy is independent of distance
- Works well in direct sunlight
- Low resolution
- Low frame rate
- Has moving parts (e.g. motor rotating mirror)