

Computer Vision

AI3604

Instructor

Wei Shen, PhD Advisor

Associate Professor

AI Institute, Shanghai Jiao Tong University

Welcome to AI3604 Computer Vision

- A bit about me
- Course organization
- What is computer vision?
- Course overview

A bit about me

A bit about me

- What do I do outside of work?
Watching live sports
Playing badminton
- I was an assistant research professor at Johns Hopkins University
- Worked with Prof. Alan Yuille



A Bit About My Research

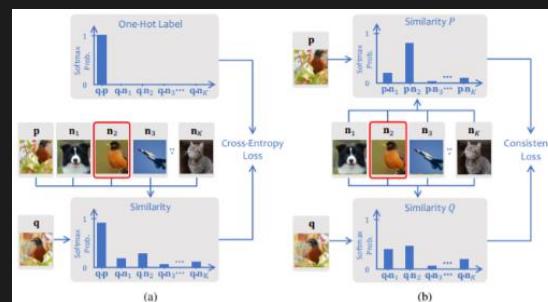
We are interested in pioneering research at the intersection of ...

- Computer Vision,
- Machine Learning,
- Medical Imaging Analysis,

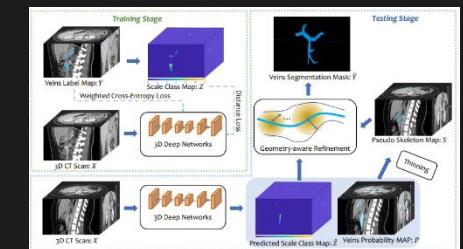
Object Detection and Segmentation



Deep Networks and Beyond



Medical Image Segmentation



Check out my homepage: <https://shenwei1231.github.io/>

Class organization

Course information (1)

- Introductory course
 - Audience: undergraduate
- Required background
 - Programming fundamentals, in particular data structures and Python experience desirable
 - Linear algebra
 - Basic probability
 - Calculus

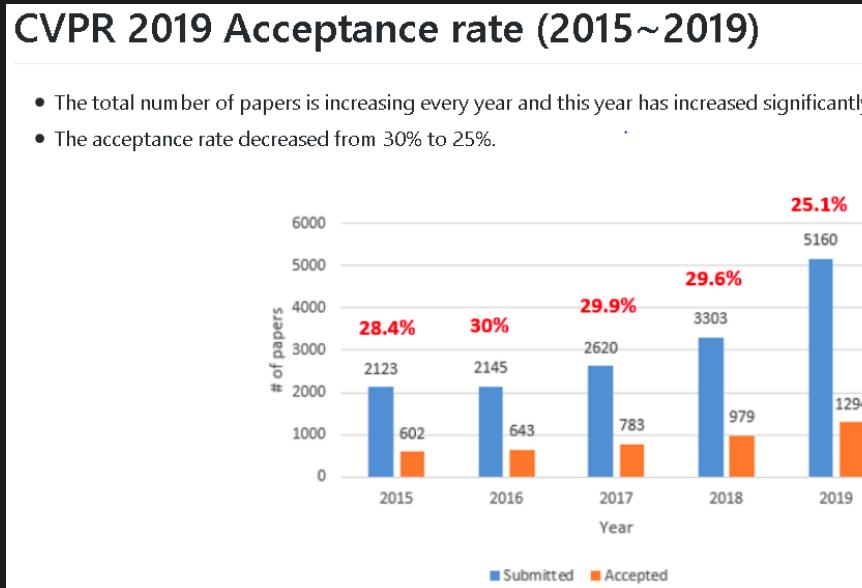
Learning Objectives

- Understand the techniques and algorithms of computer vision
- Know how to use them to address vision problems
- Fearlessly design, build, your vision methods,...
... and reason about pitfalls and design choices
- Gain intuition,...
... but realize you will not become an expert in one course

Learning Objectives

- This is an introductory course to computer vision
- While we will try to cover everything important...
- ... it is impossible to cover everything!

→ Set you up for further research/work in this area

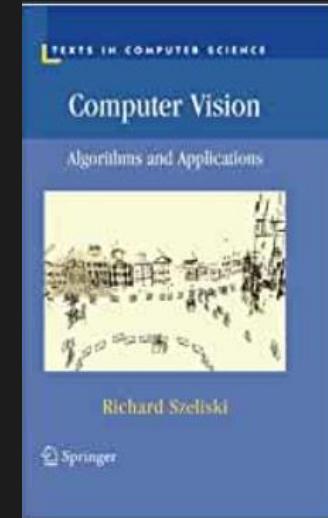


Course information (2)

- Organization
 - Two assignments, Final Project, Final Exam
- Office hours:
 - Friday afternoon or make an appointment
- Course updates:
 - Announcements/material/assignments : CANVAS

Course information (3)

- **Grading**
 - 10% Quiz + Q&A in class
 - 20% Homework
 - 40% Course Project + Presentation
 - 30% Final Exam
- **References:**
 - Optional: R. Szeliski, Computer Vision, Springer (2011).
 - Optional: Forsyth and Ponce, Computer Vision: A Modern Approach, Prentice Hall (2002).
 - Optional: Hartley and Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press (2010).



Homework

- Homework comes in two flavors
 - Written assignments
 - Programming assignments
- Late assignment submissions:
 - You have 3 late days (smallest quantity: “ 1 day ”, largest quantity: 3 days)
 - You need to request late submission before the deadline!

Final Project

- Teamwork of 3-4 students
- I will give some topics, but you can choose your own topic

Deliverables for each team:

- Submit project proposal outline (for our formal review and suggestions)
- Get excited, work on the project
- Give a presentation
- Submit a structured (brief!) final report

More details about the project will be provided on a final project description

Final Exam

- Take it easy
- But remember to flip classroom documents after class.



Programming resources

- Python tools (see TA for help)
 - Numpy
 - Pytorch
 - Matplotlib

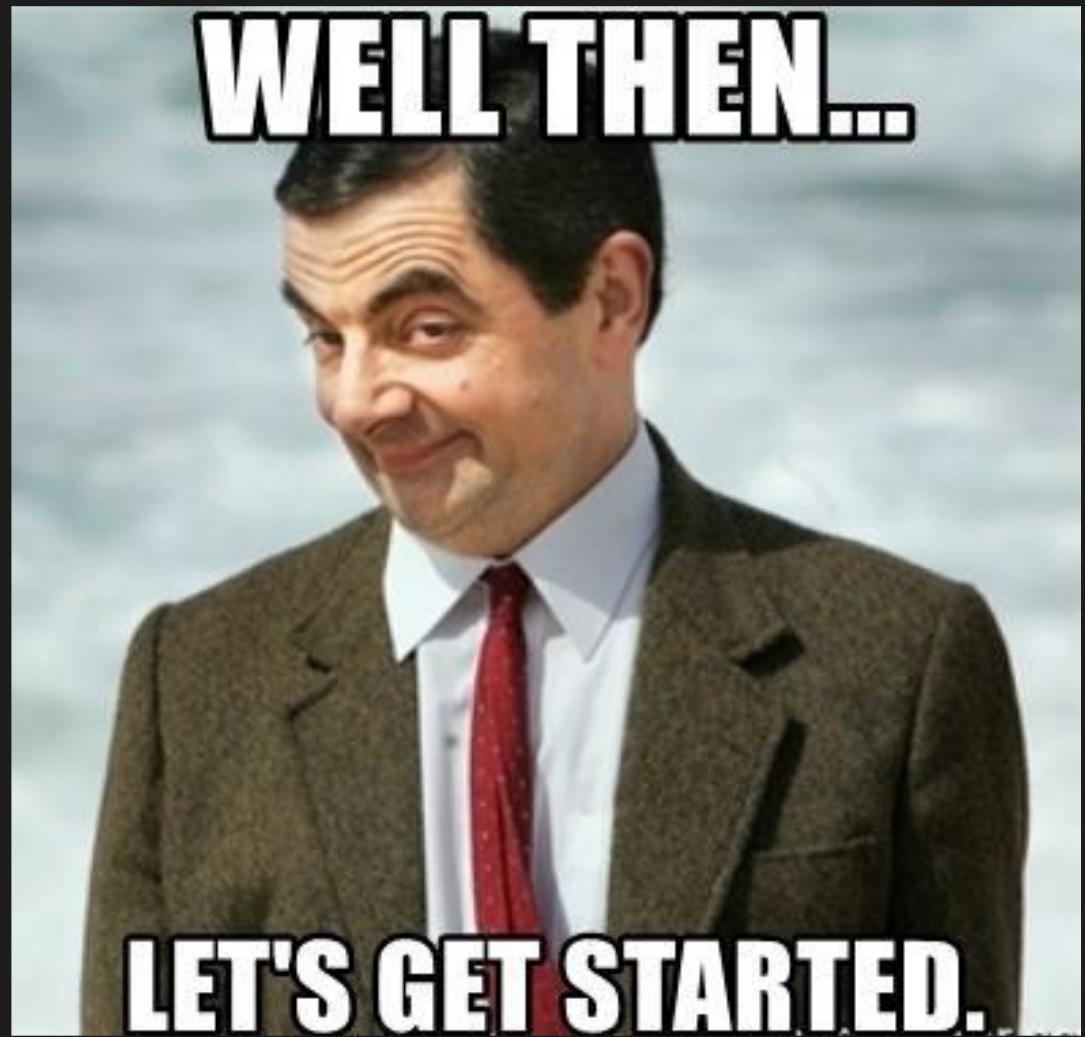
Google Colaboratory

<https://colab.research.google.com/notebooks/welcome.ipynb#recent=true>

Kaggle

- OpenCV

**Let's start
our course!**

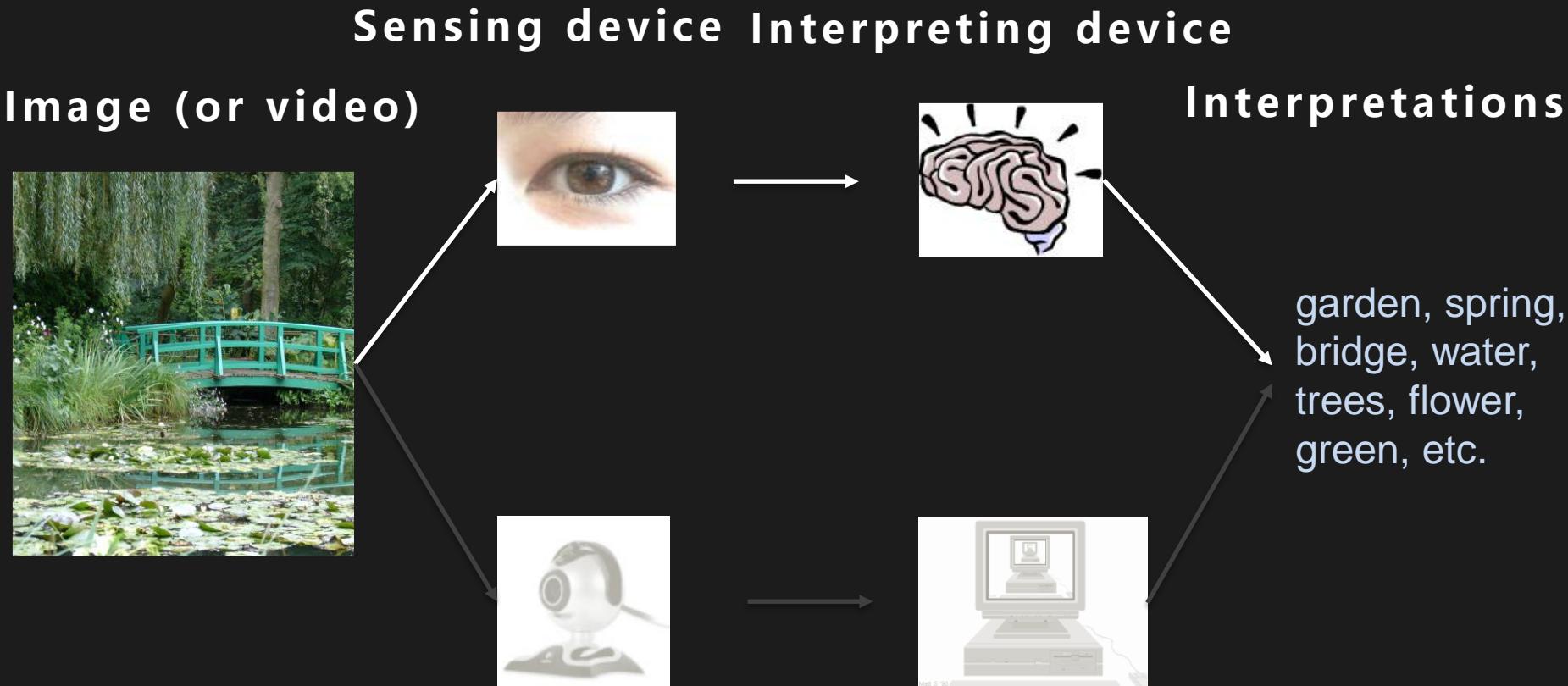


What is Computer Vision?

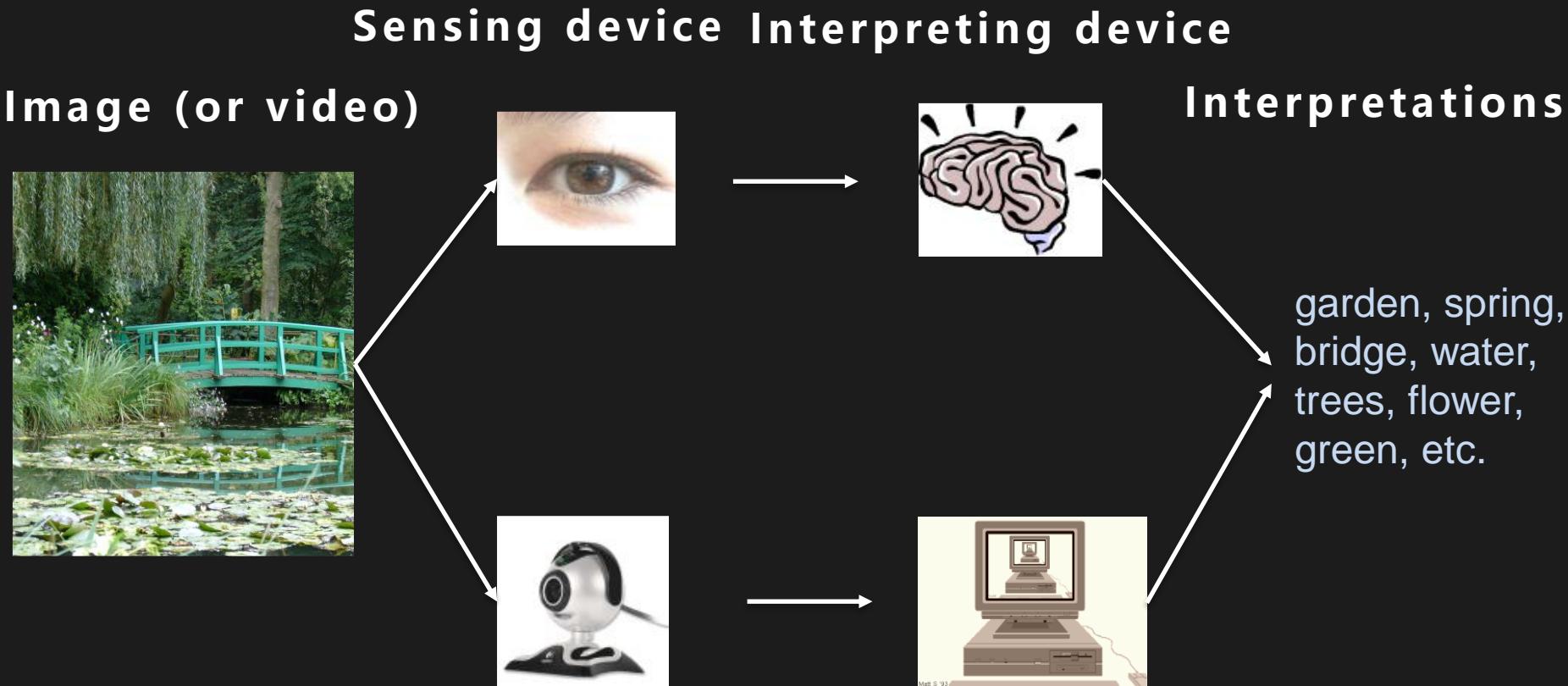


What is Computer Vision?

What is (Computer) Vision?

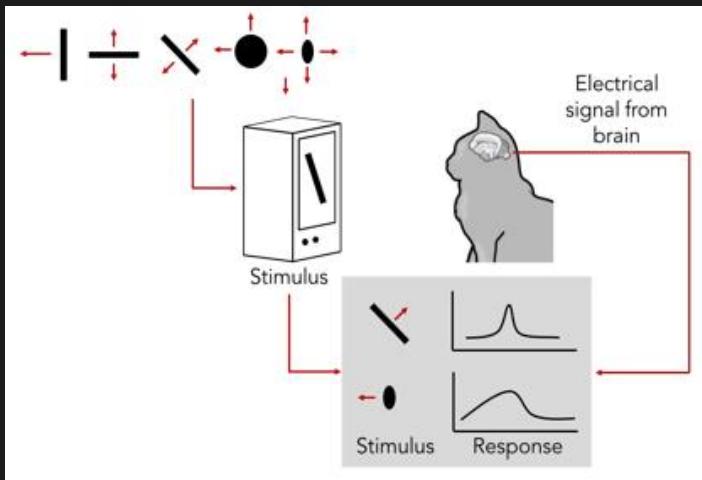


What is (Computer) Vision?



A bit of history

Biological Vision



David Hubel (1926-2013) Torsten Wiesel (1924-)

1981: Nobel Prize In medicine

David Hubel and Torsten Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. JOURNAL OF PHYSIOLOGY-LONDON, no. 1 (1962).

The First Digital Image



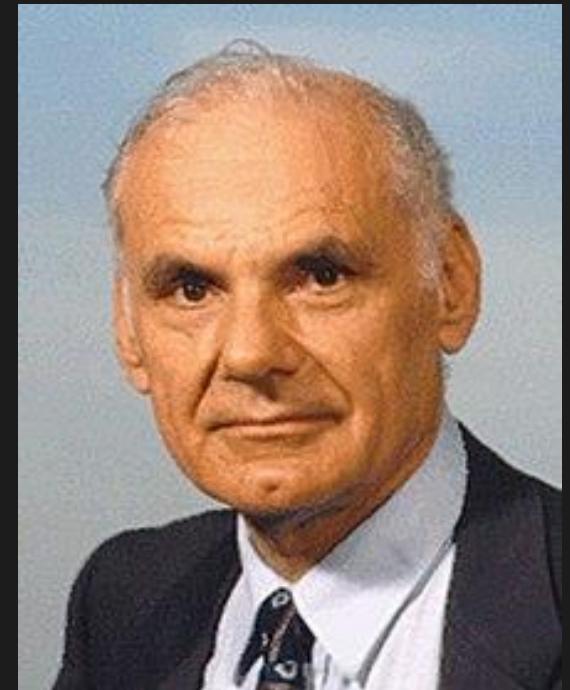
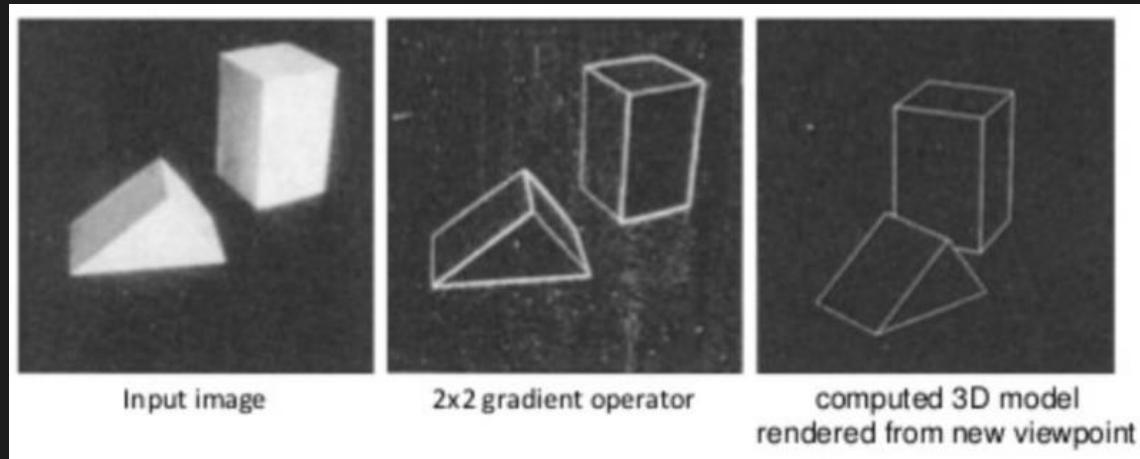
5cm by 5cm
30,976 pixels



Russell Kirsch (1929-2020)

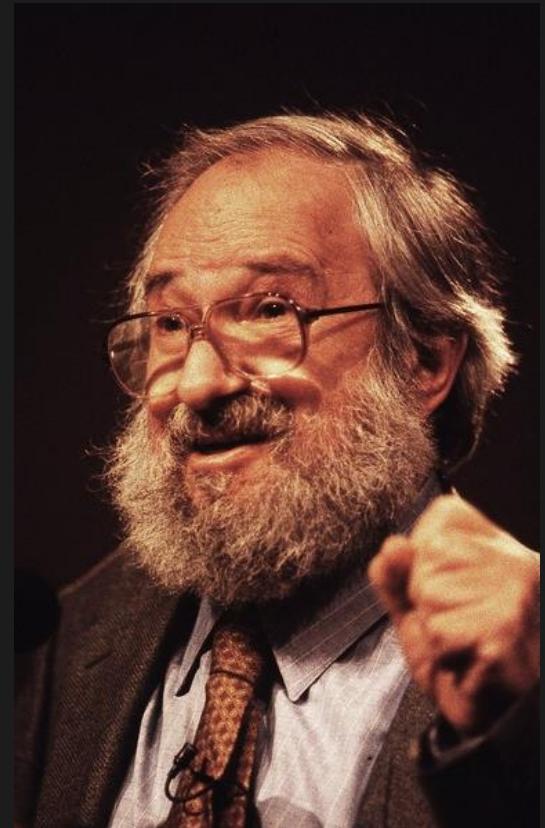
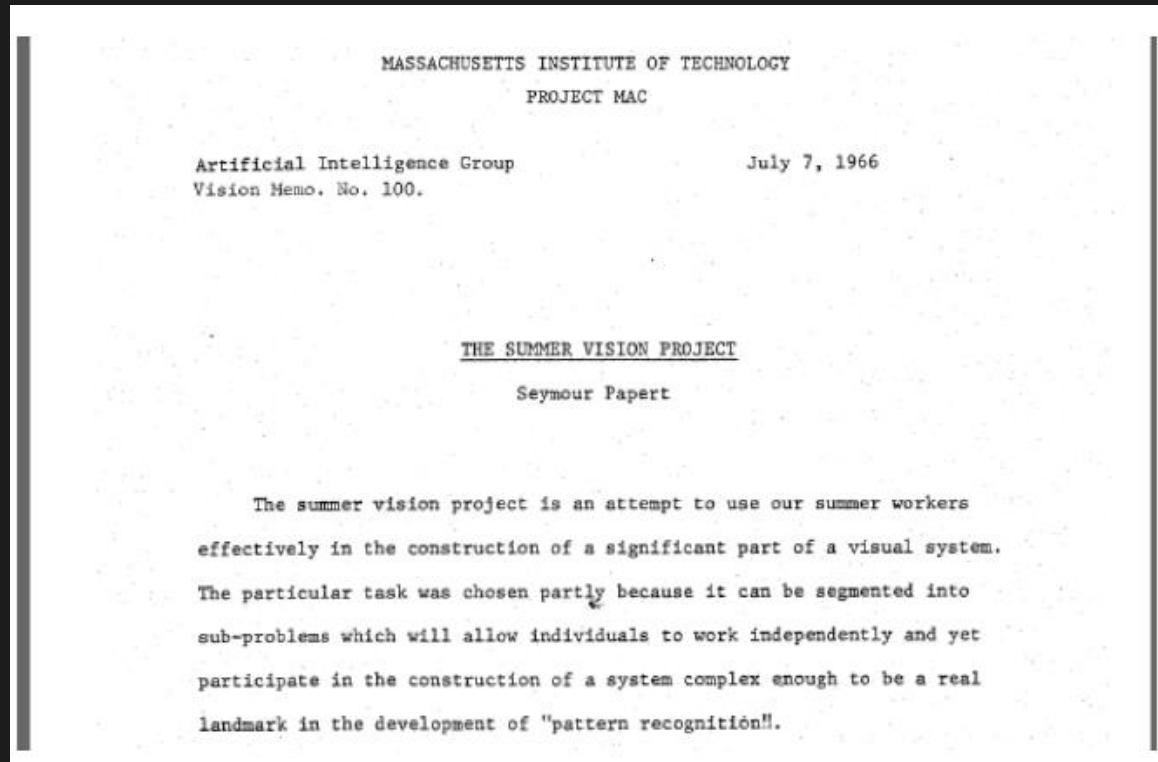
The First Thesis for Computer Vision

Machine Perception of Three-Dimensional Solids



Lawrence Roberts
(1937-2018)

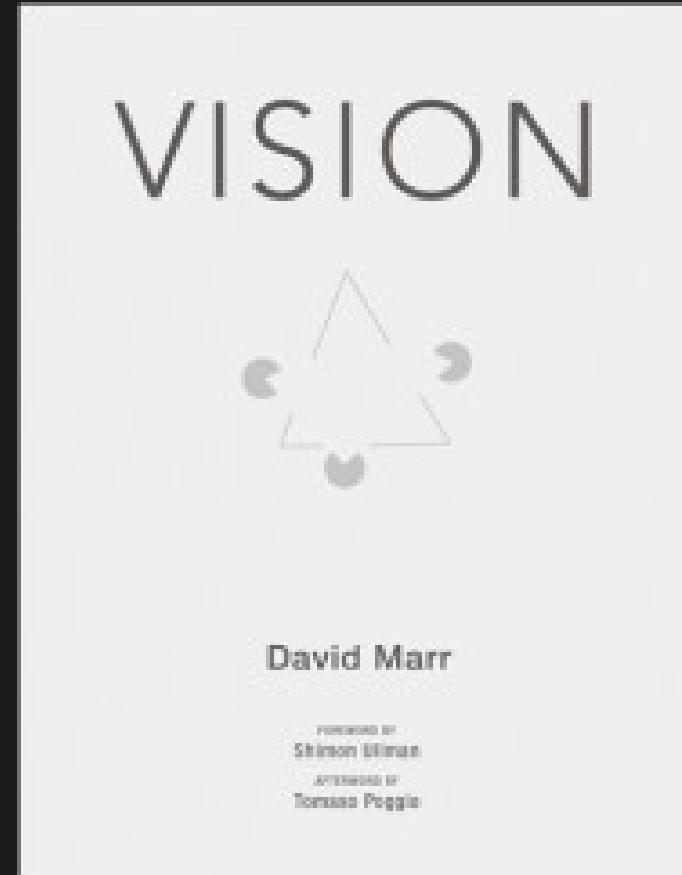
The First Project for Computer Vision



Seymour Papert
(1928-2016)



David Marr
1945-1980
Professor of
Psychology at MIT



A general framework for
understanding visual
perception

Vision as Information Processing

Marr's Tri-Level Hypothesis:

Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Vision as Information Processing

A two-dimensional visual array

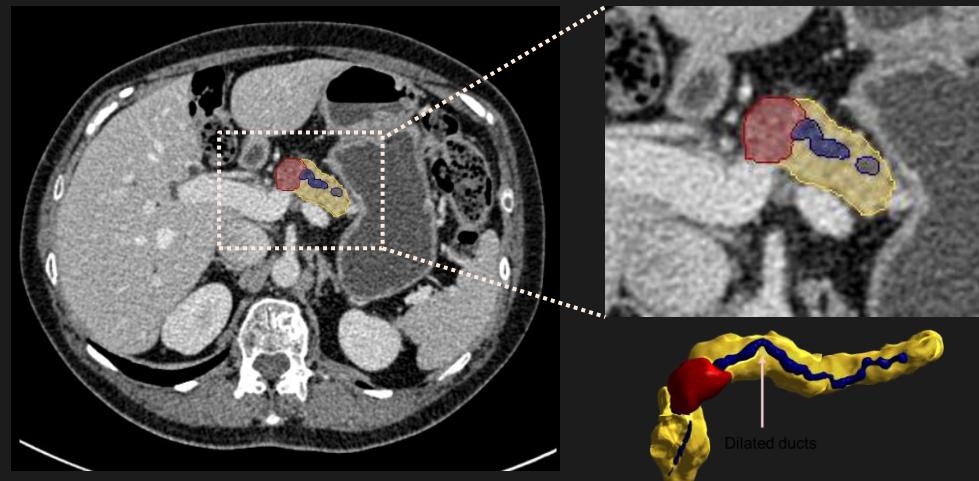


A three-dimensional description of the world as output

1. a ***primal sketch*** of the scene, based on feature extraction of fundamental components of the scene, including edges, regions, etc
2. a ***2.5D sketch*** of the scene, where textures are acknowledged
3. a ***3D model***, where the scene is visualised in a continuous, 3-dimensional map.

Computer vision?

- Recovering properties from the real world using images
 - Examples:
 - Detecting/recognizing objects in a scene
 - Reconstructing a 3D model of the imaged buildings
- Overlaps with a few other areas from computer Science, e.g.
 - Image processing
 - Machine learning/pattern recognition
 - Graphics



But What Really Is Computer vision?

- Vision is
 - ... automating human visual processes
 - ... an information processing task
 - ... inverting image formation
 - ... inverse graphics
- ... really useful!

So Why Is It So Hard?

- The Human Visual System has no problem interpreting subtle variations and correctly segmenting the object from the background



So Why Is It So Hard?

- The Human Visual System has no problem interpreting subtle variations and correctly segmenting the object from the background

What can we infer?

- It's a flower.
- The flower is pink.
- There are water droplets on the flower. Perhaps it rained earlier?
- It's daytime and sunny.
- ...



So Why Is It So Hard?

- The Human Visual System has no problem interpreting subtle variations and correctly segmenting the object from the background

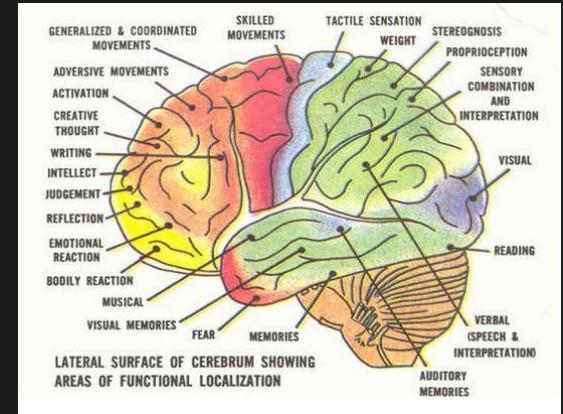
What can the machine infer?

- It's a flower.
- ...
- That's mostly it!



Why is vision challenging...?

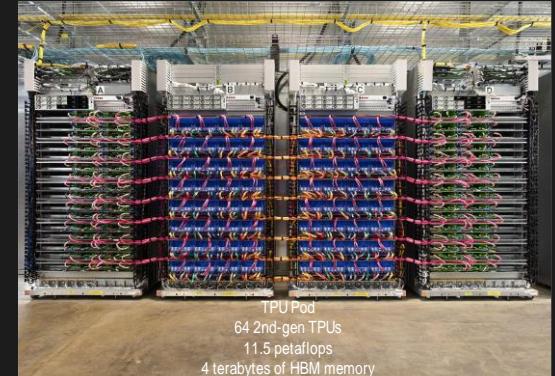
- When you see an image, you see:
 - The objects in the image
 - The context of the scene
 - Your prior knowledge and experiences automatically apply to what you see to deduce meaning
 - ...and much much more...
- When a machine sees an image, it sees:
 - Numbers
 - ...and...
 - ...well...
 - That's it



- Approx. 10^{11} Neurons
- Approx. 10^{14} Synapses
- Firing rates 100-1000 Hz
- Asynchronous, distributed
- Consumes 20W
- Diverse (everyone is different!)

Why is vision challenging...?

- When you see an image, you see:
 - The objects in the image
 - The context of the scene
 - Your prior knowledge and experiences automatically apply to what you see to deduce meaning
 - ...and much much more...
- When a machine sees an image, it sees:
 - Numbers
 - ...and...
 - ...well...
 - That's it

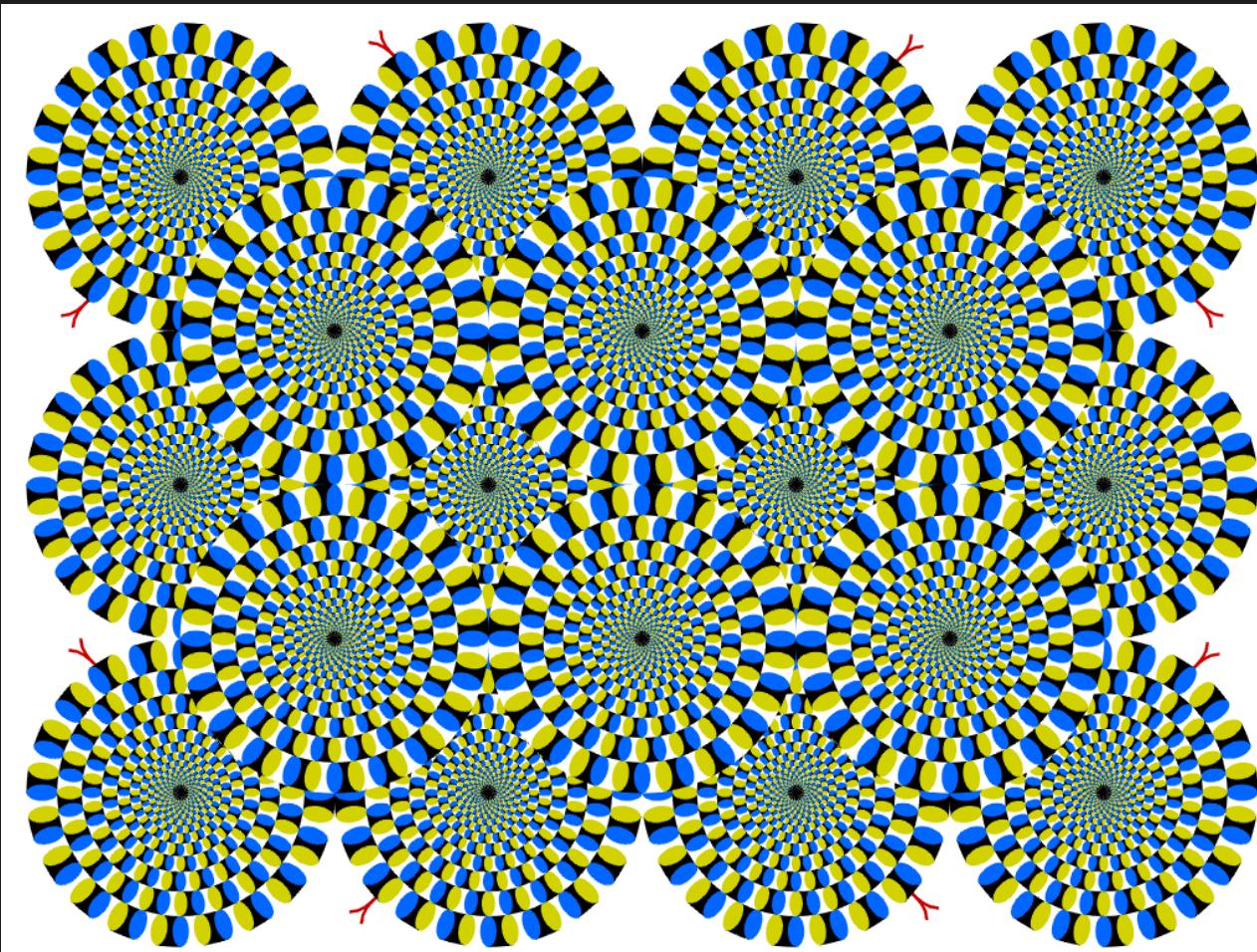


- 11.5×10^{15} floating point ops
- 4×10^{12} memory locations
- Gigahertz clock
- Synchronous
- Consumes hundreds of kilowatts
- Each is exactly the same

So Why Is It So Hard?

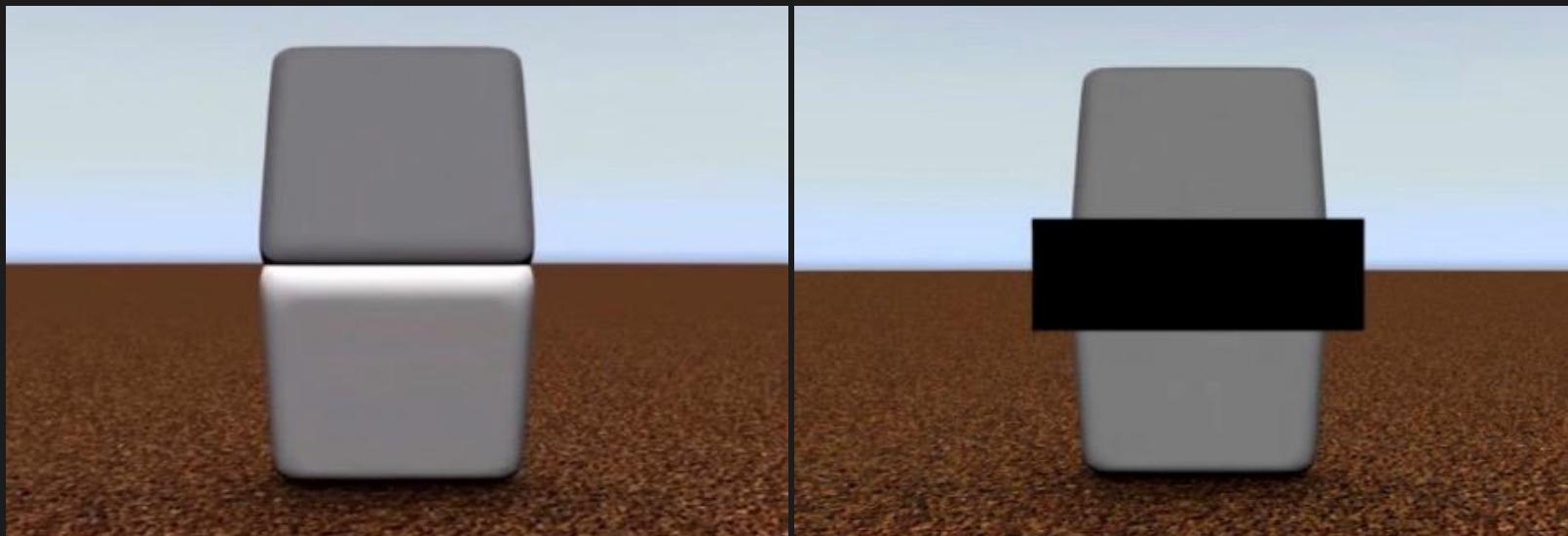
- Basically we are solving an “Inverse Problem”
 - Recover some unknowns given insufficient information to fully specify the solution
- So what do we do?
 - Use physics-based and probabilistic models to disambiguate between potential solutions
 - Let’s be honest – we also use a bunch of “tricks”!

Can computers match (or beat) human vision?

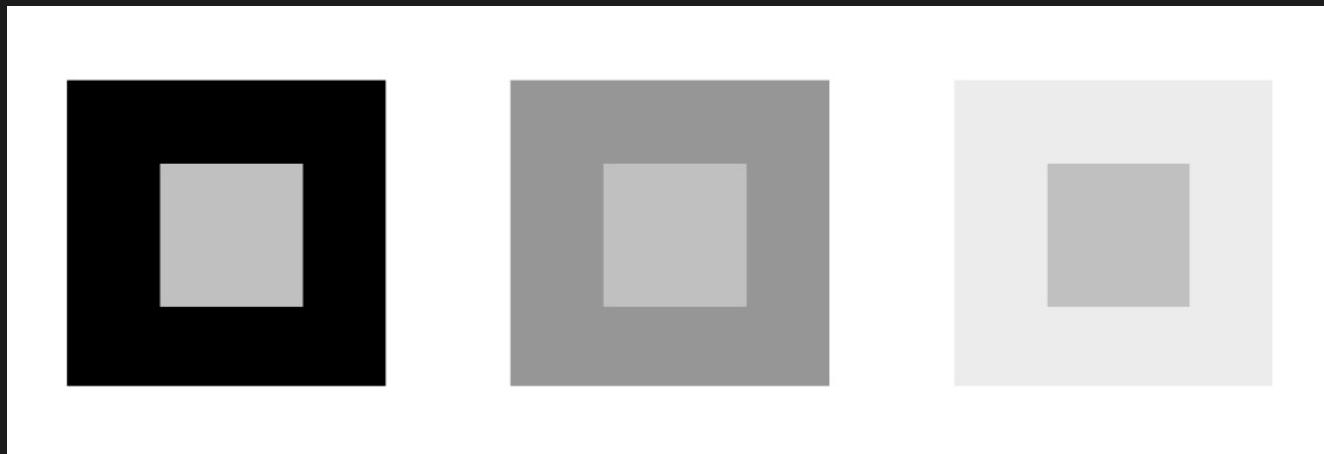


- Yes and no (but mostly no!)
 - humans are much better at “hard” things
 - computers can be better at “easy” things

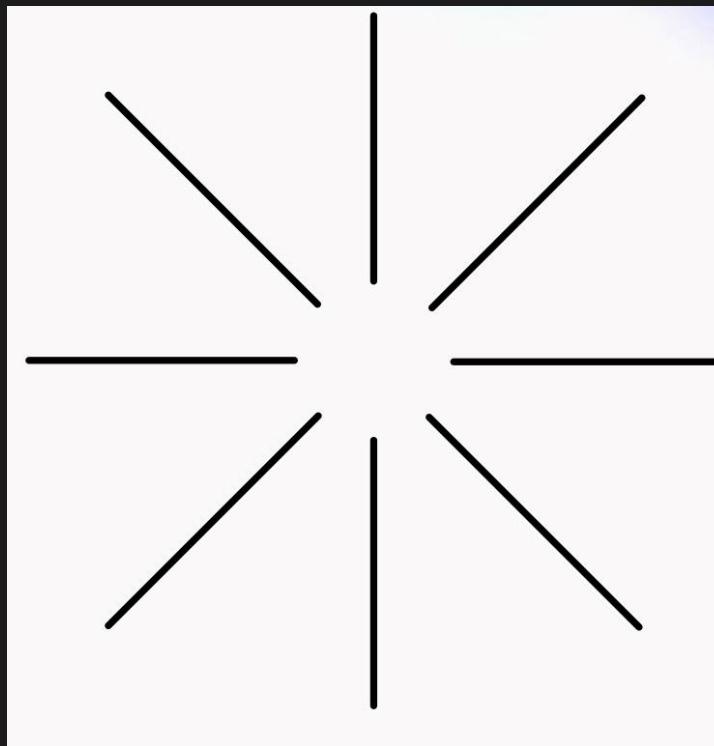
Optical Illusions



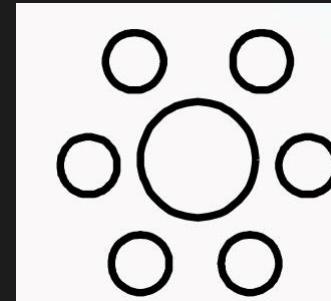
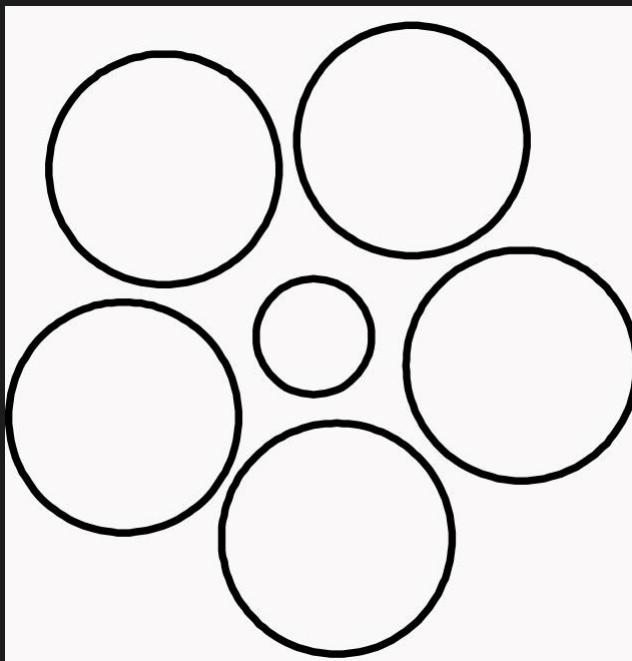
Optical Illusions



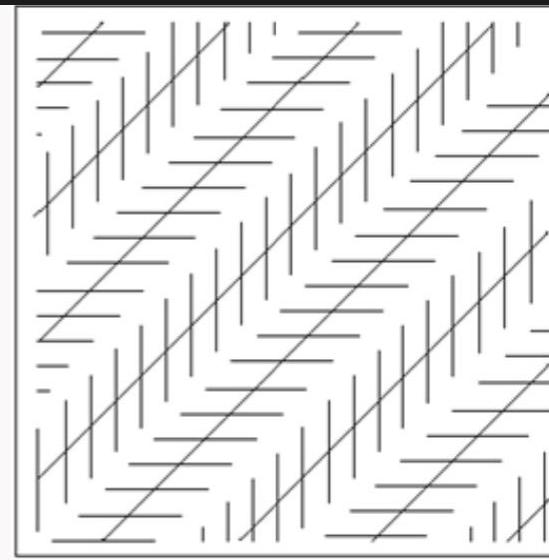
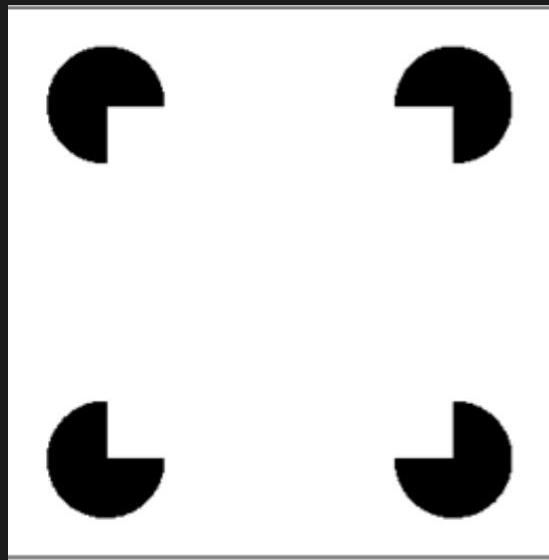
Optical Illusions



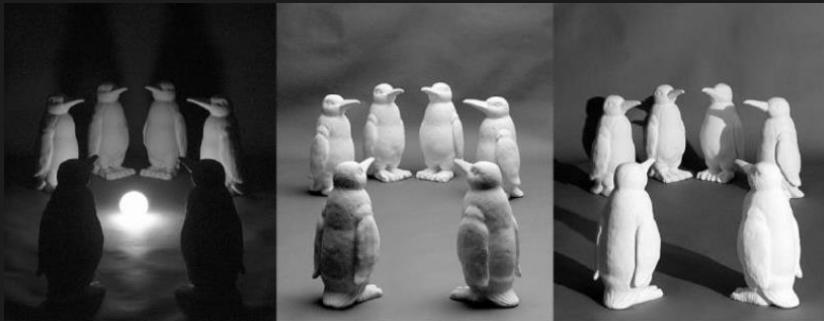
Optical Illusions



Optical Illusions



Why is vision challenging...?



Illumination and viewpoint



Intra-class variation



Scale and depth

Why is vision challenging...?



Shadows

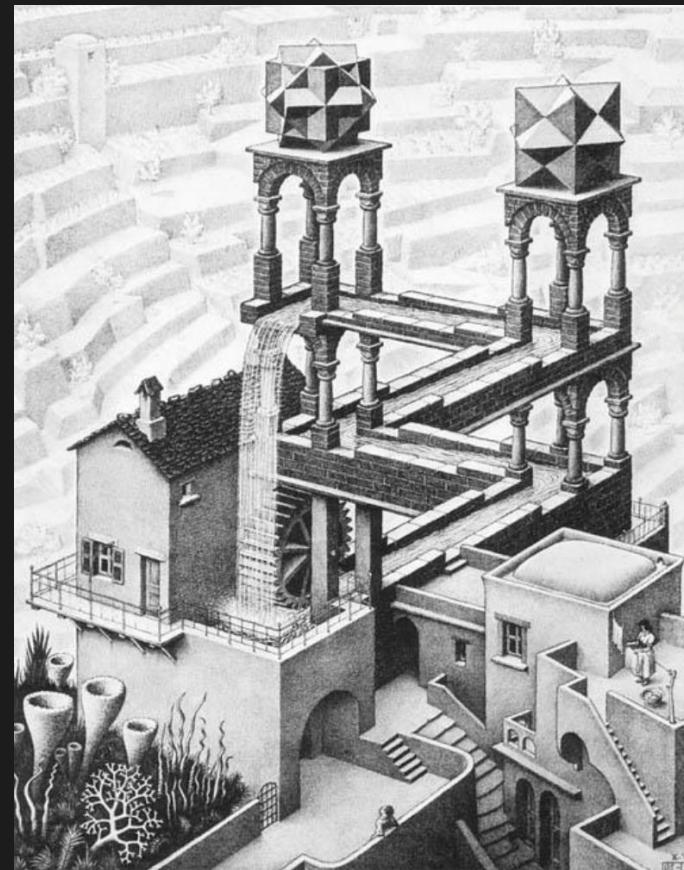


Clutter

Why is vision challenging...?



Occlusions



Illusions

Why is vision challenging...?



Why is vision challenging...?



Why Vision?

Every picture tells a story



Why Vision?

- Images and video are everywhere!



Image modalities (1)



Digital cameras



Infrared cameras

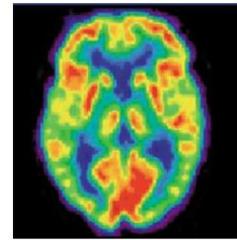
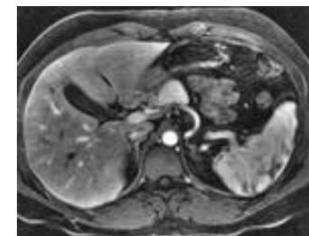


Time-of-flight
cameras



Kinect

Image modalities (2)



Ultrasound Endoscopy

X-rays

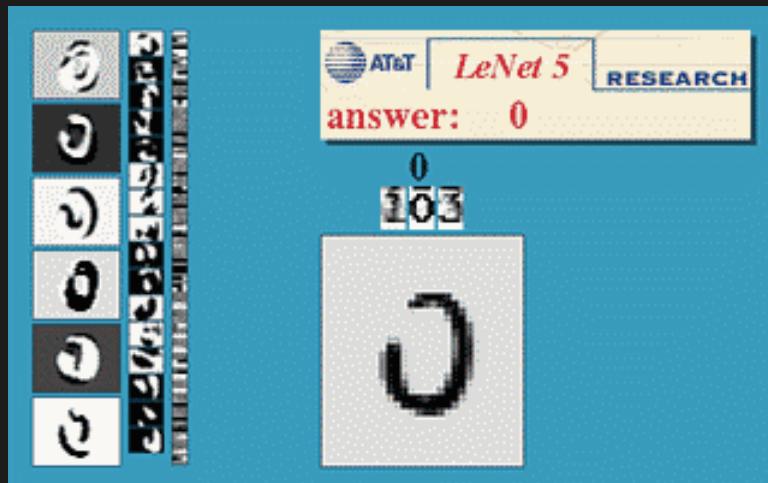
Computed
tomography

Magnetic
resonance

Positron
emission
tomography

Application examples

Optical character recognition (OCR)



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



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

Application examples

Face Detection



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

Application examples

Smile Detection

The Smile Shutter flow

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



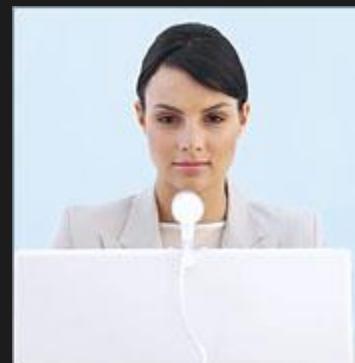
Sony Cyber-shot® T70 Digital Still Camera

Application examples

Login without a password...



Fingerprint scanners on
many new laptops,
other devices



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

Application examples

Object recognition (in mobile phones)



- This is becoming real:
 - **Lincoln** Microsoft Research
 - Point & Find, Nokia
 - SnapTell.com (now amazon)

Application examples

Special effects: shape capture



The Matrix movies, ESC Entertainment, XYZRGB, NRC

Application examples

Sports



Application examples

The image shows a screenshot of the Mobileye website. At the top, there are tabs for "manufacturer products" and "consumer products". Below this, a banner features the slogan "Our Vision. Your Safety." with an overhead view of a car showing three cameras: a "rear looking camera" at the back, a "forward looking camera" at the front, and a "side looking camera" on the side. To the right, there's a "News" section with a link to "Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System" and another to "Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end". Below the news is a "Events" section with a link to "Mobileye at Equip Auto, 1999-2002". The main content area contains three cards: one for "EyeQ Vision on a Chip" showing a chip image, one for "Vision Applications" showing a pedestrian crossing, and one for "AWS Advance Warning System" showing a driverless car simulation with a 3D map and distance markers.

- Mobileye
 - Vision systems currently in high-end B
 - By 2010: 70% of car manufacturers.

Application examples

Vision in space



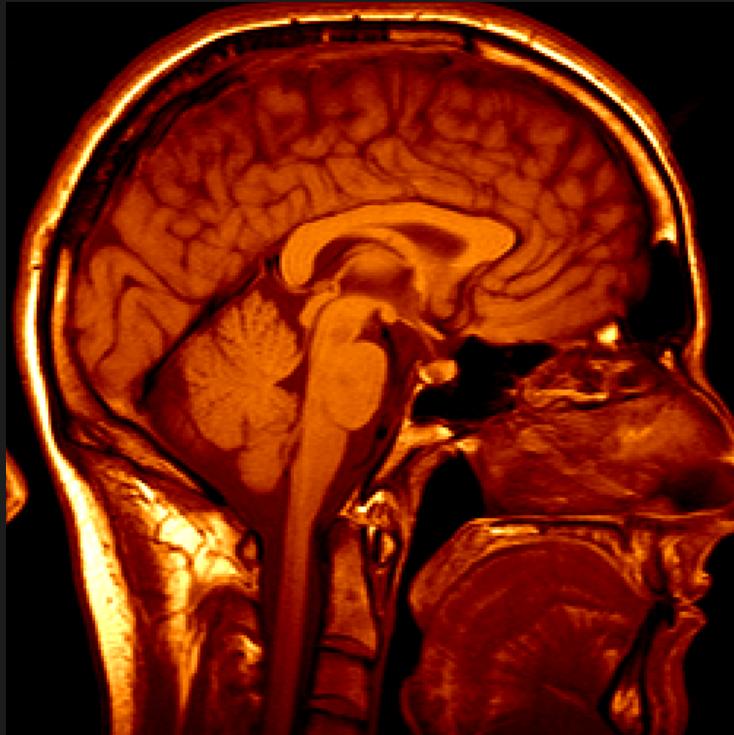
Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “Computer Vision on Mars” by Matthies et al.

https://www.bilibili.com/video/BV1Vb4y1f7SQ/?vd_source=e1e4c8fc77bdec98956285ddf0e0b8c4

Application examples

Medical imaging



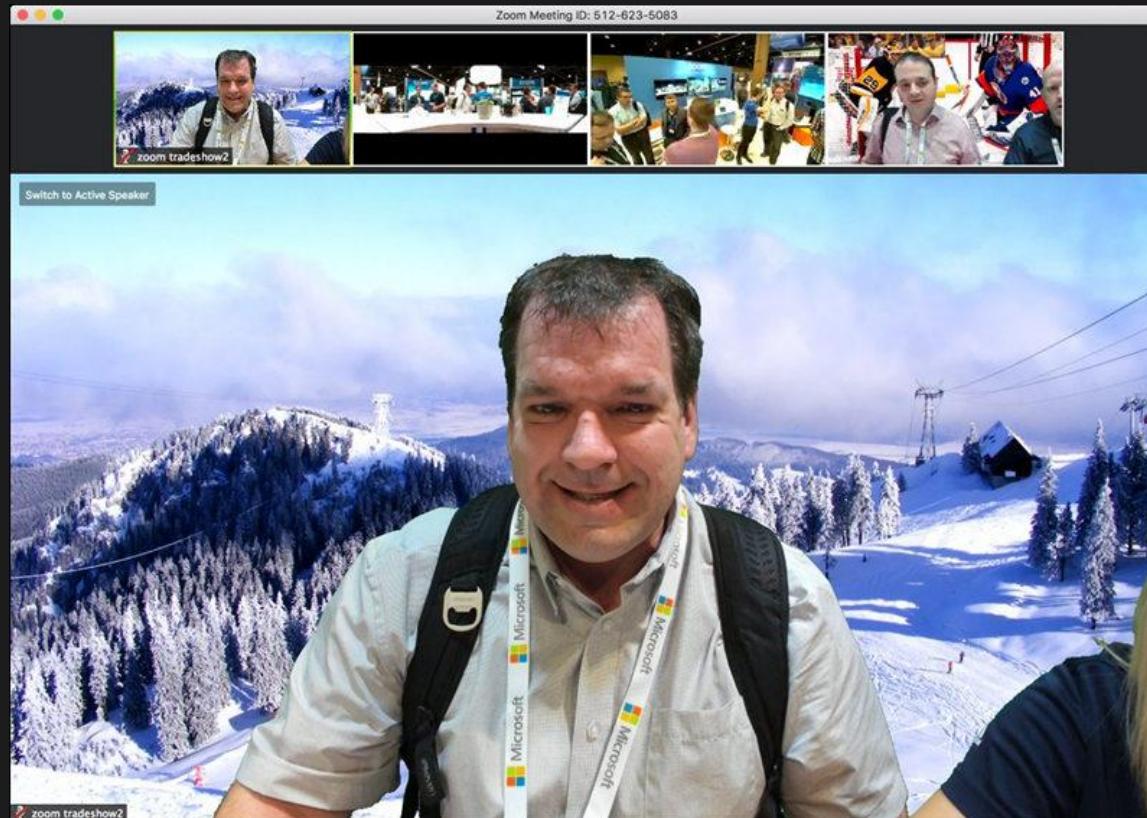
3D imaging
MRI, CT



Image guided surgery
Grimson et al., MIT

Application examples

Virtual Meeting – Foreground Segmentation



Class Overview

Low-level Vision

- Edge Detection
- Corner Detection
- Feature Detection

- Image Classification
- Semantic Segmentation
- Object Detection

High-level Vision

- Camera Calibration
- Image Alignment
- Motion Estimation
- Stereo Vision

3D Reconstruction

Learning

Topic Relation in This Course

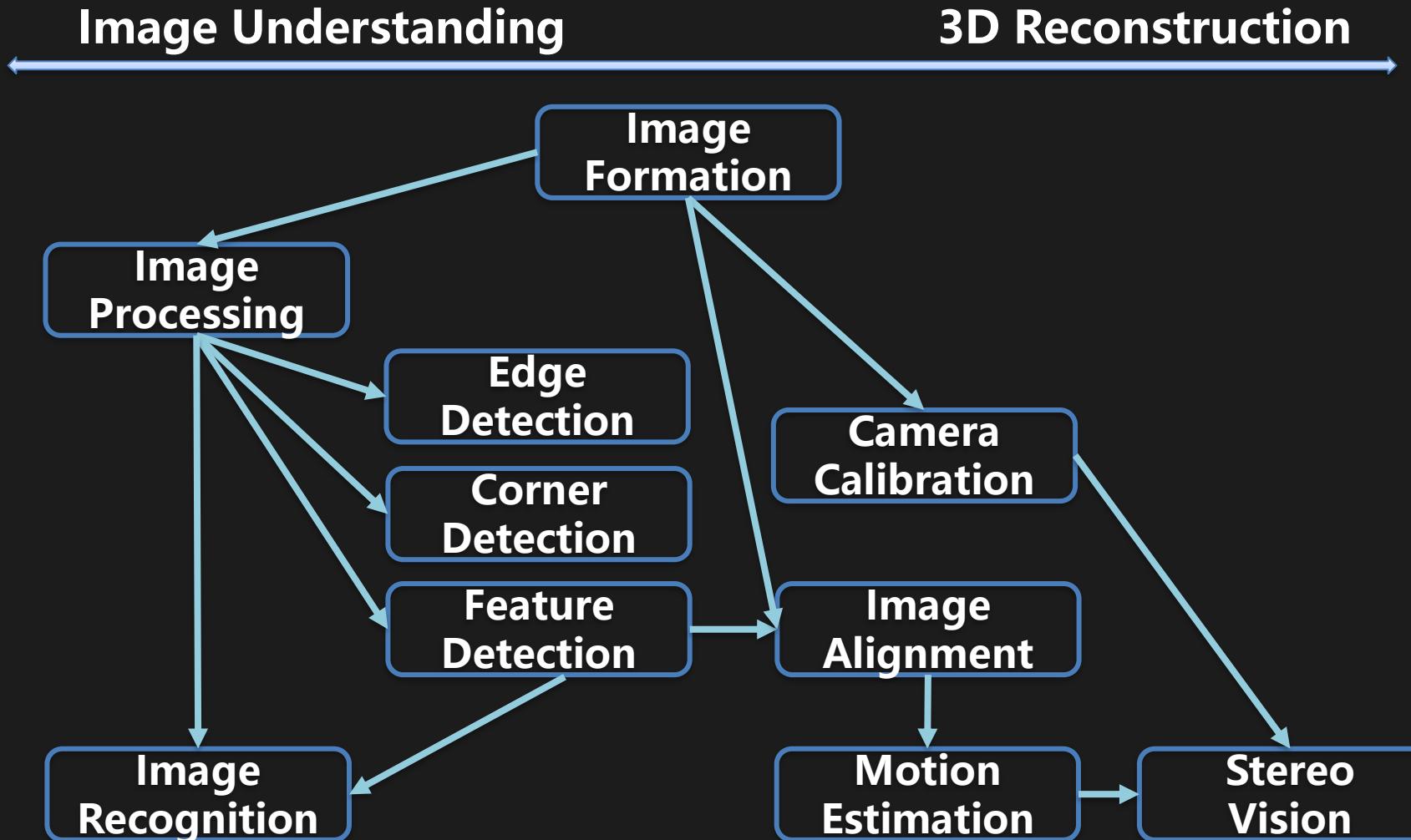
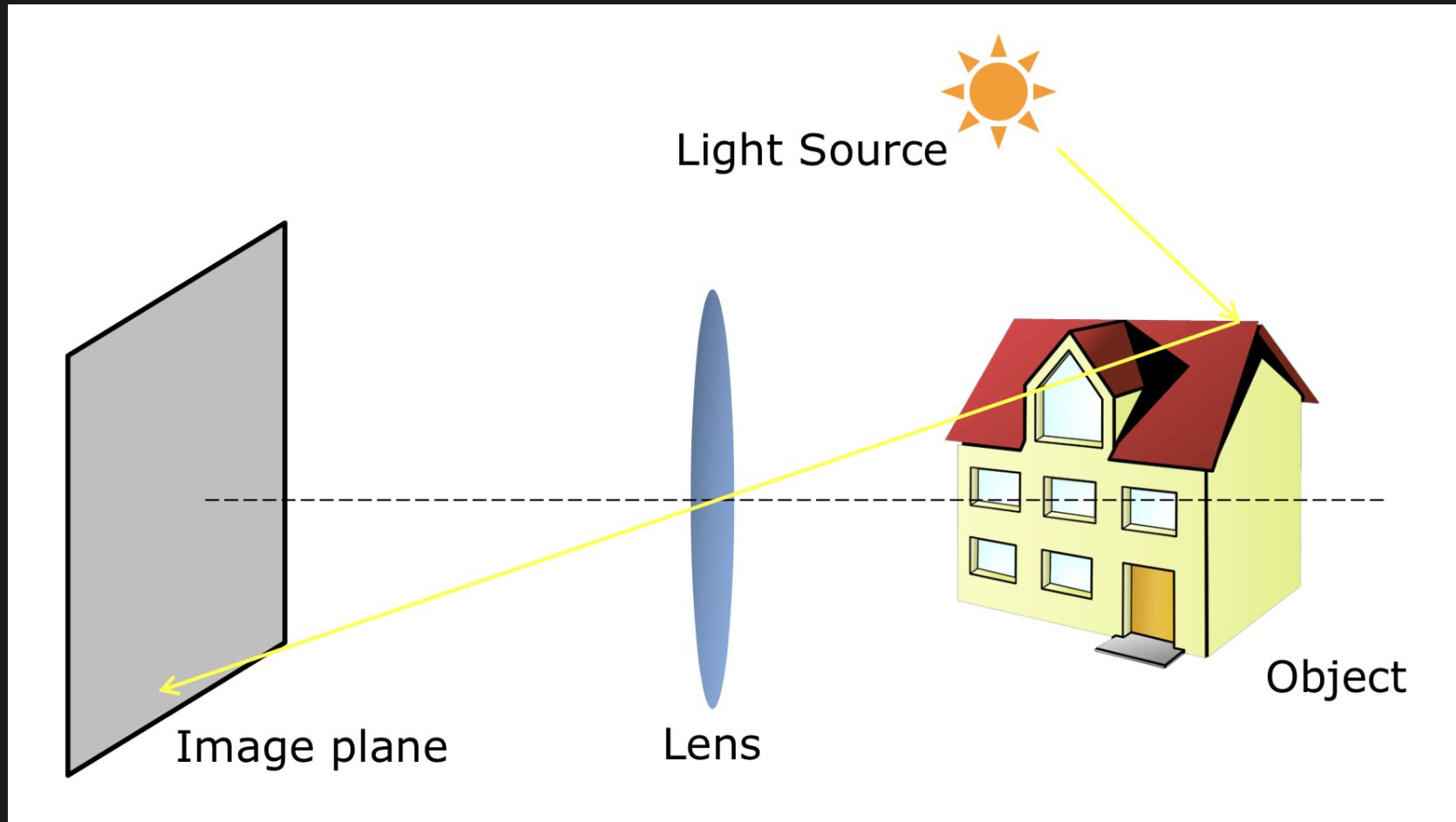


Image Formation and Optics

- Where do Images come from?



Projection of 3D world on a 2D Plane

Computer Vision, SJTU, Wei Shen

Image Processing

- Basic operations, linear filters, non-linear filters, binary image processing



32

16

8

4

2

Number of gray levels

Edge Detection

- 1st Derivative, 2nd Derivative, Canny, Hough Transform



Image (I)



$\partial I / \partial x$



$\partial I / \partial y$



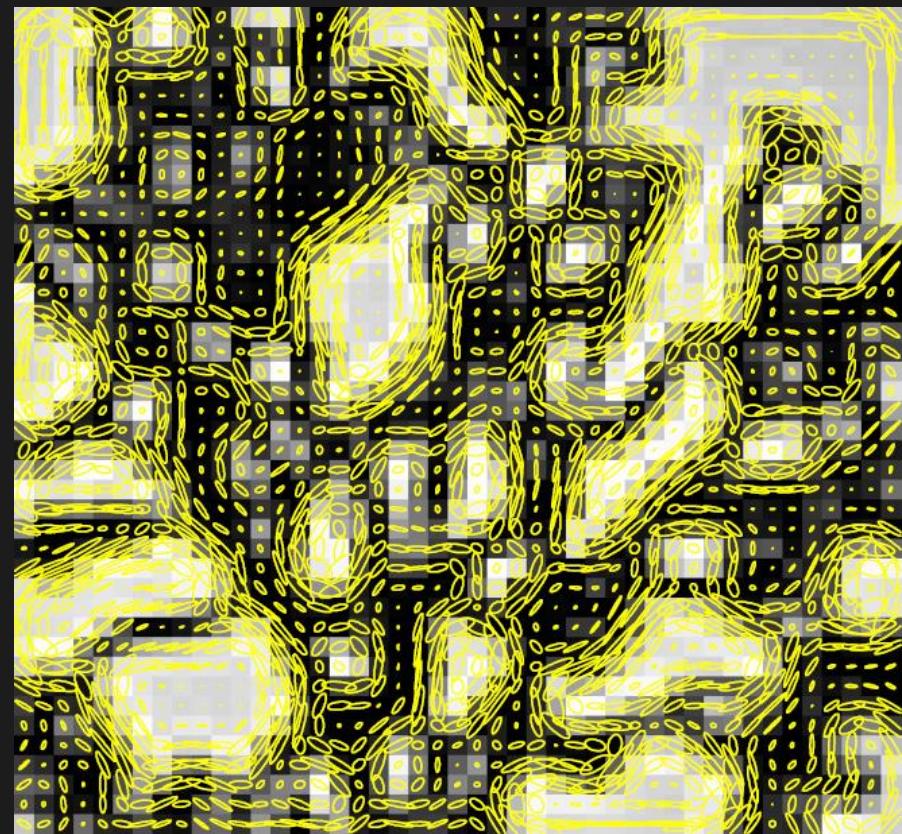
Gradient Magnitude



Thresholded Edge

Corner Detection

- Harris corner detector, cross-correlation



Feature Descriptors

- Feature detection and matching, SIFT

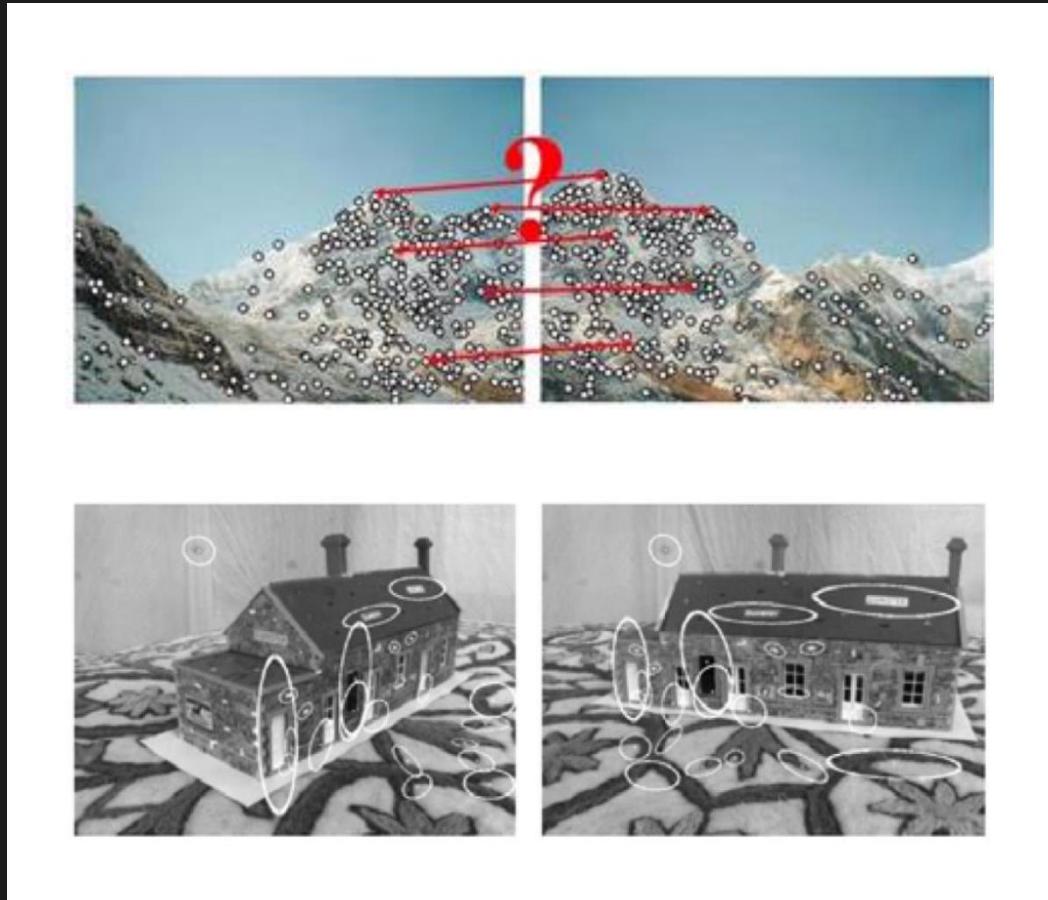
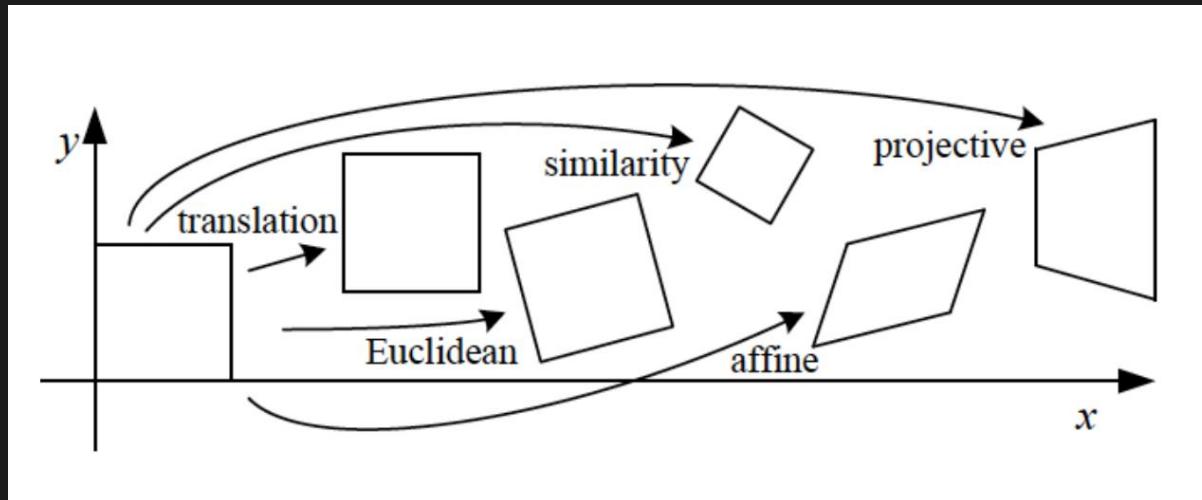


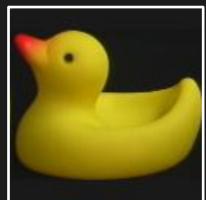
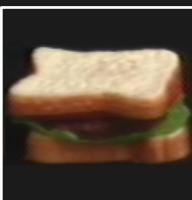
Image Alignment

- 2D transformations, homographies



Visual Recognition

- Object recognition by appearance matching, Bag-of-features model



100 Objects Database

Recognition

Object



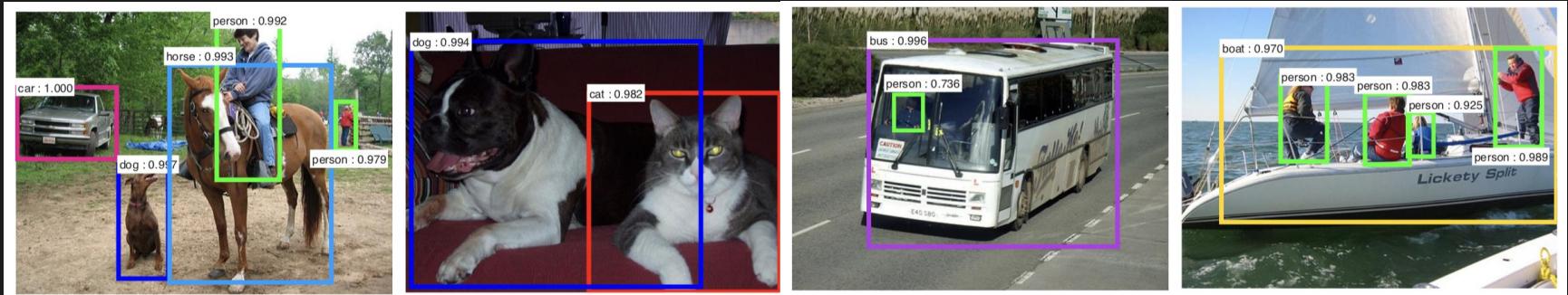
Bag of 'words'

Semantic Segmentation



Chen et al.

Object Detection



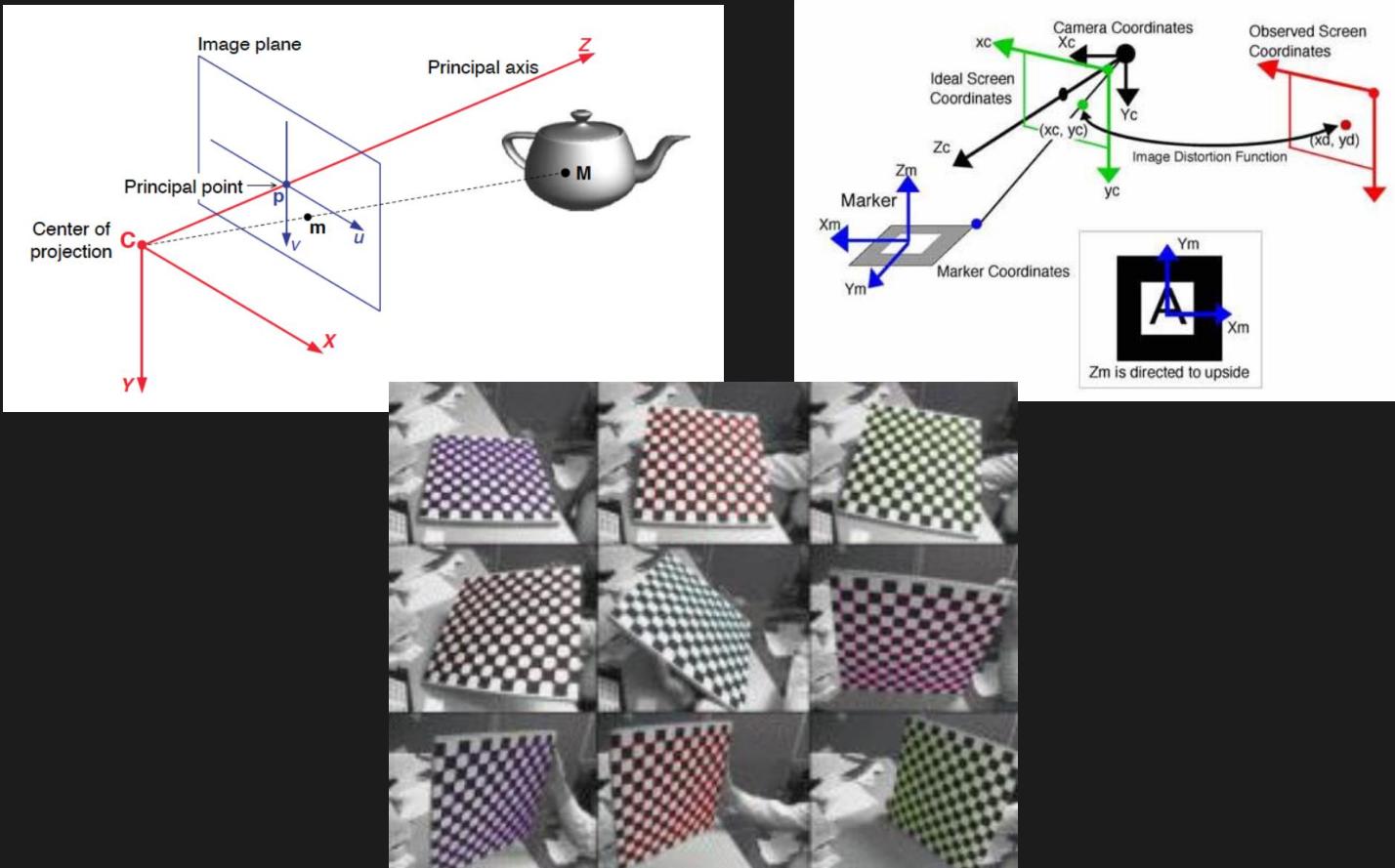
Ren et al.



He et al.

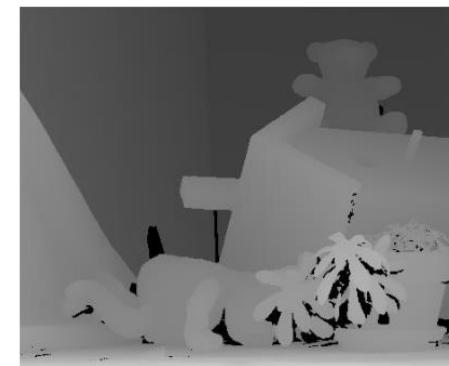
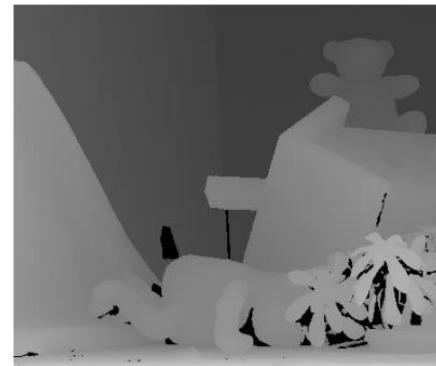
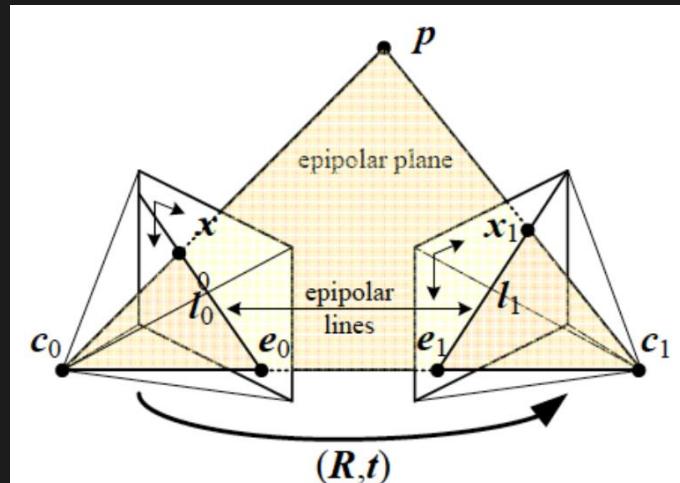
Camera Calibration

- Camera model, calibration method



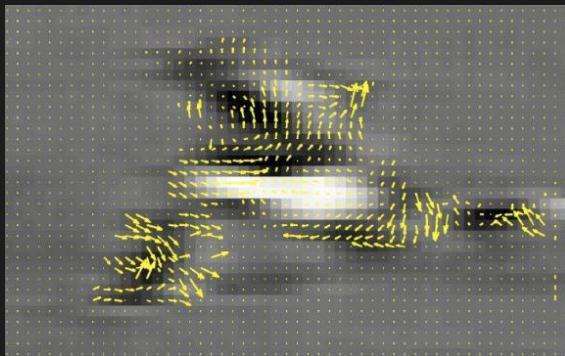
Stereo Vision

- Epipolar geometry, depth estimation

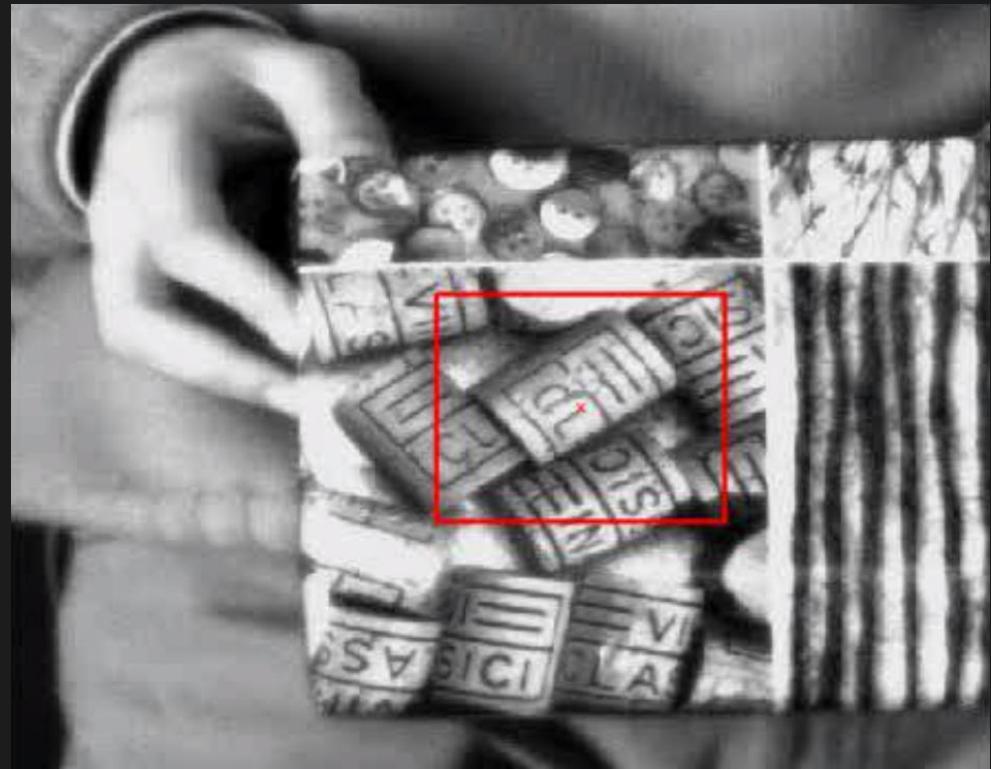


Motion & Tracking

- Optical flow, feature tracking, applications, SSD template tracking



Efros et al.



INRIA

A few videos...

Videos

- Mobile fusion

MobileFusion

Peter Ondrúška, Pushmeet Kohli and Shahram Izadi

ISMAR 2015



UNIVERSITY OF
OXFORD

Microsoft Research

Videos

- Real-time semantic segmentation

Videos

- Real-time 3D Object Detection

Monocular Quasi-Dense 3D Object Tracking

Hou-Ning Hu, Yung-Hsu Yang, Tobias Fischer, Trevor Darrell,
Fisher Yu, and Min Sun

Videos

- Real-time human pose estimation

Real-time Multi-Person 2D Pose Estimation Using Part Affinity Fields

Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh
Carnegie Mellon University

Videos

- Robotics



Videos

- AI Agriculture

Videos

- AI Agriculture



What is next?

- Images and image formation
 - Basic image operations
 - Linear and non-linear filtering
-
- Please check:
 - Access to CANVAS
 - Access to Python