



S A I R

Spatial AI & Robotics Lab

# CSE 473/573 - A

## COMPUTER VISION & IMAGE PROCESSING

Chen Wang

Spatial AI & Robotics Lab

Department of Computer Science and Engineering



University at Buffalo The State University of New York

# Introduction

---

**University at Buffalo  
Department of Computer Science and Engineering  
CSE 473/573 - Computer Vision and Image Processing  
Spring 2024**

**Section A: TuTh 9:30AM - 10:50AM**

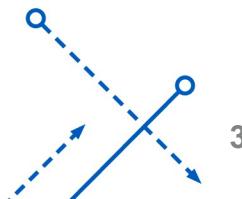
**Location: NSC 216**

**Jan 24, 2024-May 7, 2023**

# Instructor Information

---

<b>Instructor:</b>	<b>Chen Wang, PhD.</b> ( <a href="https://sairlab.org">https://sairlab.org</a> )
<b>Office:</b>	304 Davis Hall
<b>Email:</b>	Prefers to be contacted through Piazza
<b>Office Hours:</b>	Friday 1-2pm
<b>Zoom Link</b>	See Piazza – Resources – Staff, Available when campus closed.



# Special Considerations

---

- Office hours will be a combination of zoom and in person
- Piazza page
  - <https://piazza.com/buffalo/spring2024/cse473573a>
  - Access code: sairlab
- Slides will be put up one day before class.
- Instructor and TAs will monitor Piazza questions

# Course Information

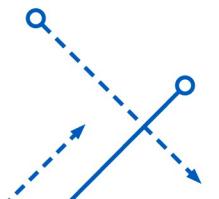
---

- Lectures, Homework's, Quizzes, Projects and Final Exam over 14-week semester.
- This course is an introduction to those areas of Artificial Intelligence that deal with fundamental issues and techniques of **computer vision** and **image processing**.
- The emphasis is on **physical, mathematical, and information-processing aspects** of the vision.
- Topics to be covered include **image formation, edge detection and segmentation, convolution, image enhancement techniques, extraction of features such as color, texture, and shape, object detection, 3-D vision, and their applications**.
- The material is based on graduate-level texts augmented with research papers, as appropriate.

# Overview

---

- The course will move fast
- Self-discipline is important
- The emphasis of the course is to develop practical skills for solving Computer Vision and Image Processing problems
- Fair evaluations: undergraduate and graduate students will be scored separately
- Academic Integrity (AI) will be taken seriously and zero tolerance to any cheating (work on homework and projects independently)



# Prerequisites:

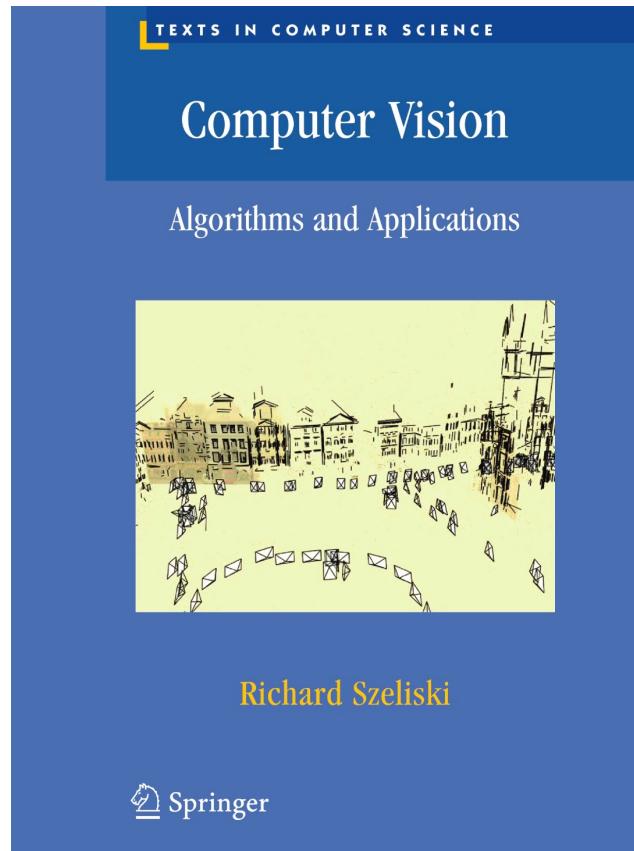
---

- CSE 305:
  - Introduction to Programming Languages
- Knowledge of Linear Algebra
- Programming in Python

# Textbook

---

- Computer Vision: Algorithms and Applications, by Richard Szeliski



# Course Requirements

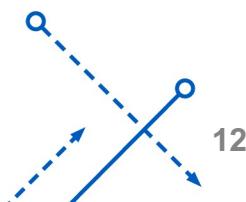
---

- Class attendance and participation is expected
- You are responsible for ALL materials presented in class and assigned to read
- Quizzes will be given during class time only.
- There will be ~~three~~ **two** projects
- Regular deliverables on the project will be graded during the course

# Grading and Submissions

---

- All assignments will be graded out of 100 points and weighted according to the table next page.
- We will drop ONE homework or Quiz Grade, whichever results in a higher overall grade.
- All assignments will be turned in via UB Learns.
- Quizzes will be given online through the UB Learns system.
- For some quizzes and test, you will be required to install the Respondus browser to prevent cheating during online exams.



# Grading

---

<u>Weighting</u>	<u>Assessment / Assignment</u>	<u>Number</u>
20%	Homeworks	4
20%	Quizzes	5
30%	Projects	2
30%	Final	1

Today: Quiz 0

# Late Submission Policy

---

- Completed homework and project deliverables are to be submitted by their deadline (11:59pm).
- For homework, you will have up to 3 days to receive a grade reduced by 50%. No additional late days allowed.
- Projects: You will be allowed a total of 3 grace days throughout the semester.
- Each late day beyond the 3 grace days will reduce your grade by 50%.
- No individual project will be accepted after 3 days late.

# Exam Policy

---

- No makeup exams will be given except in **provably extreme circumstances** and when consistent with University Policy.
- Notify your instructor **1 week prior to the exam** via **e-mail** if you are going to miss it. If it is medically impossible for you to give prior notice, please obtain a **note from a physician detailing the period** (and the reason) you were medically incapable of communicating with the instructor.
- If you miss an exam/quiz because of sickness or similar reasons, visit a physician and obtain a note detailing the period and the reason you were medically incapable of taking the exam/quiz.
- You are responsible for knowing about the exam date. Please plan your travel and other activities accordingly.

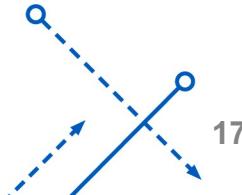
# Regrading for Errors

---

- Assignments, quizzes and exams may be submitted for regrading to correct grading errors.
- Regrade requests are due **no later than one (1) week** after the scores are posted.
- Regrade requests must be **clearly written and attached** to the assignment.
- When work is submitted for regrade, the entire work may be regraded, **which may result in a lower grade**.
- Work done in pencil may not be considered for regrading.

# Questions?

---



# Disabilities

---

- If you have a diagnosed disability (physical, learning, or psychological) that will make it difficult for you to carry out the course work as outlined, or that requires accommodations such as recruiting note-takers, readers, or extended time on exams or assignments, please **advise the instructor during the first two weeks of the course** so that we may review possible arrangements for reasonable accommodations.
- In addition, if you have not yet done so, contact the **Office of Disability Services**.

# Academic Integrity

---

- All work must be your own
  - Do not take the answers, words, ideas or research findings of other people as yours; cite and acknowledge properly, and develop your own ideas.
  - No cheating
  - According to departmental policy, any violation of academic integrity will result in an “F” for the course, and termination of departmental financial scholarship.
  - Tools will be used to check similarity. **Similar submissions will result in “F” for all involved parties.**
- Use of a code from an online repository, e.g., GitHub, must include a proper and clearly visible attribution in your report.

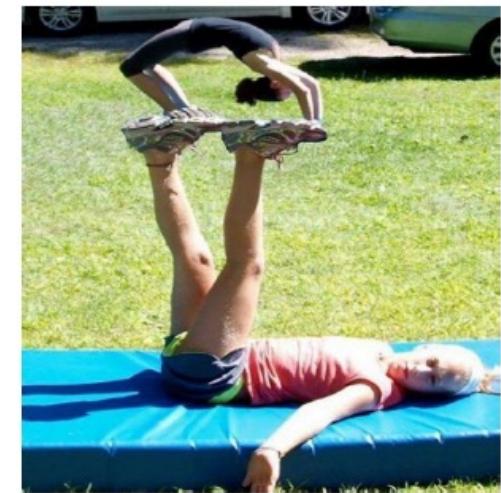
# How to Fail the Course

---

- Don't meet the prerequisites
- Don't pay attention to information on Piazza
- Start your project at the last minute
- Wait until the deadline to submit for the first time
- Don't read the syllabus carefully (Academic Integrity violation and late submission policy etc.)
- **Cheat (please check AI policy of the university and department if you are new)**
- <https://catalog.buffalo.edu/policies/integrity.html>

# Questions?

- Instructors/TA/Graders will stay for a few minutes after each lecture. Simply ask!
- Syllabus and Slides are on Piazza.



# Content

---

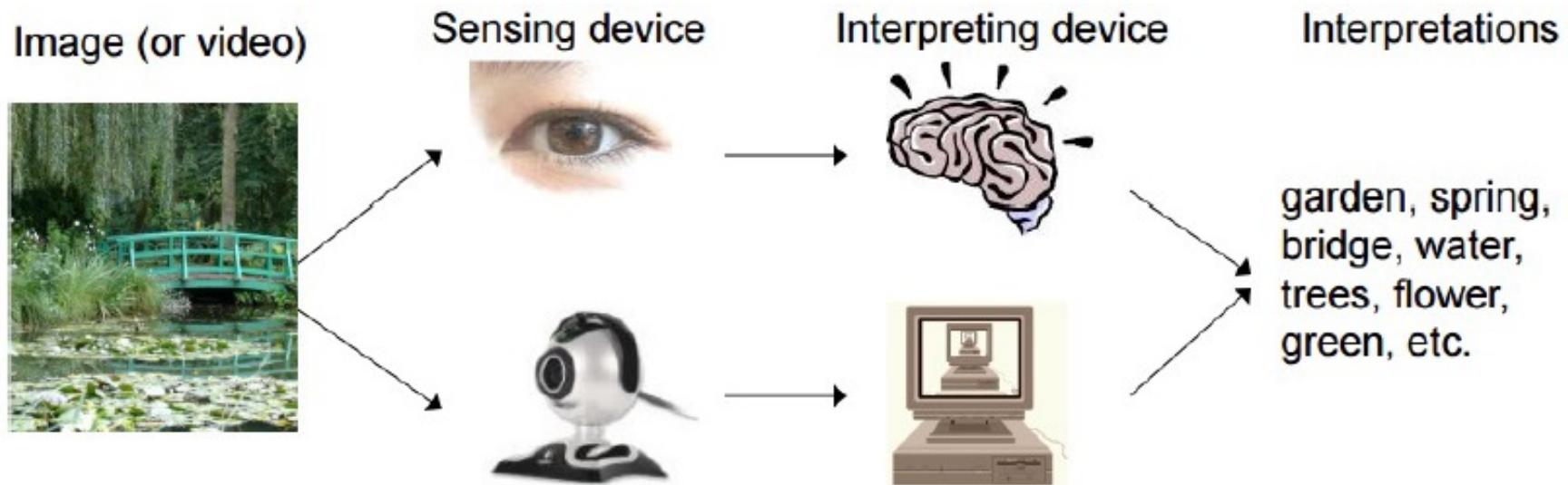
- Definition
- History
- Research Topics
- Applications
- Quiz 0
  - UB Learn, Academic Integrity Letter

# What is Computer Vision

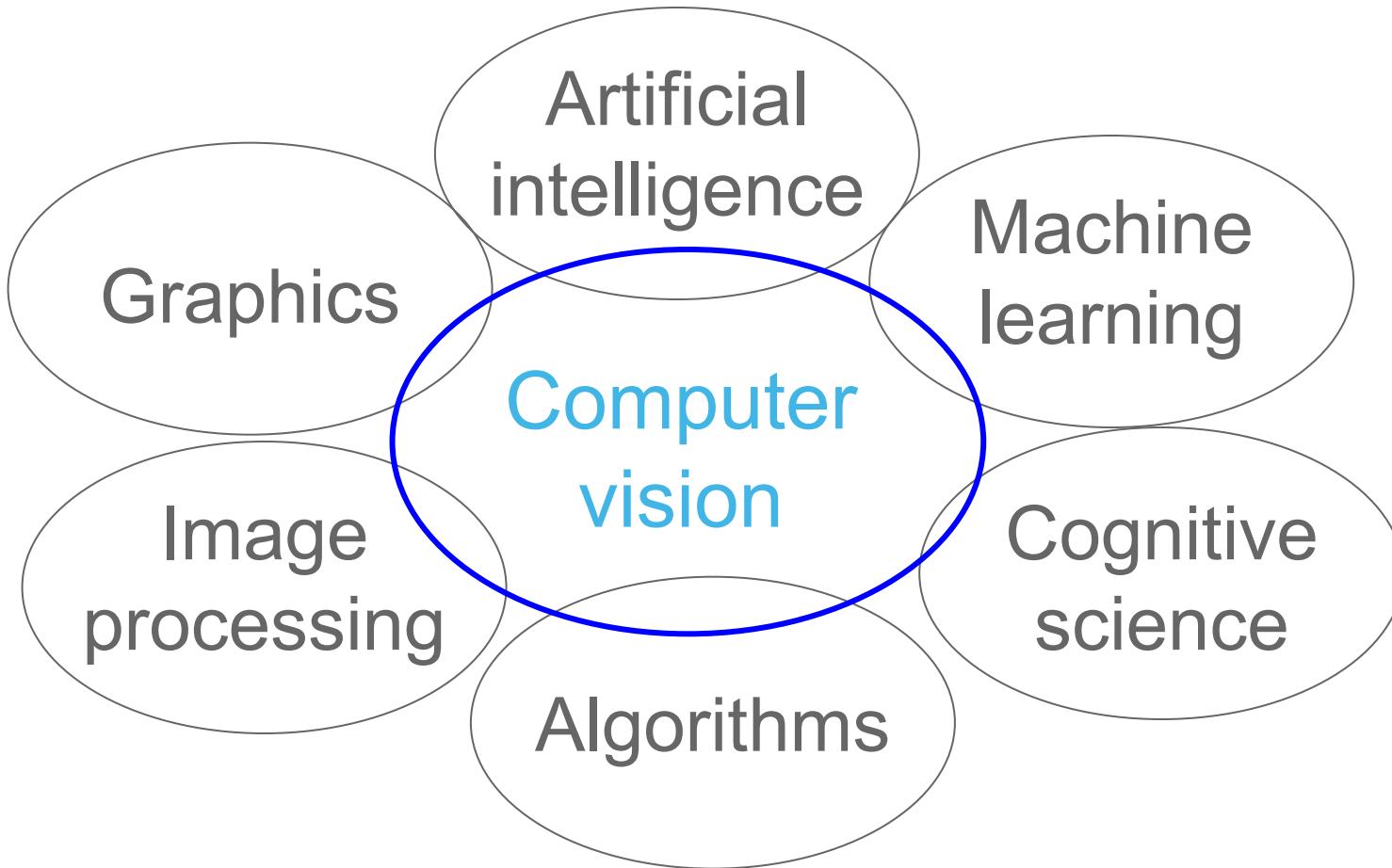
---

- **Computer vision** is a field of **computer science**
  - works on enabling **computers** to see,
  - identify and process images in the same way that human **vision** does, and
  - then provide appropriate output.
- It is like imparting human intelligence and instincts about vision to a **computer**.

# Computer Vision vs Human Vision



# Related disciplines



# (Brief) History of Computer Vision

- In 1966, Marvin Minsky at MIT asked his undergraduate student Gerald Jay Sussman to “spend the summer linking a camera to a computer and getting the computer to describe what is sliding across the table.” We know that the problem is much more difficult than that. (Szeliski et al., 2009)

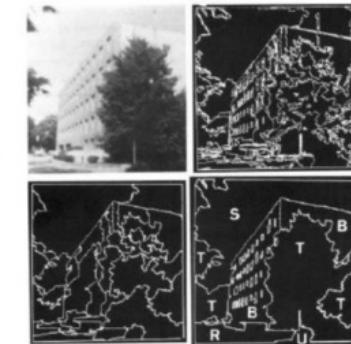


We know that the problem is much more difficult than that. (Szeliski et al., 2009)

...identify and process images in the same way that human **vision** does...

# A Brief History of Computer Vision

- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2010's: Deep learning with ConvNets



Ohta Kanade '78



Turk and Pentland '91

# Timeline



1970

Digital image processing  
Blocks world, line labeling  
Generalized cylinders  
Pictorial structures  
Stereo correspondence

Intrinsic images

Optical flow

Structure from motion

Image pyramids

Scale-space processing

Shape from shading,  
texture, and focus

Physically-based modeling

Regularization

Markov Random Fields

Kalman filters

3D range data processing

Projective invariants

Factorization

Physics-based vision

Graph cuts

Particle filtering

Energy-based segmentation

Face recognition and detection

Subspace methods

Image-based modeling  
and rendering

Texture synthesis and inpainting

Computational photography

Feature-based recognition

MRF inference algorithms

Category recognition  
Learning



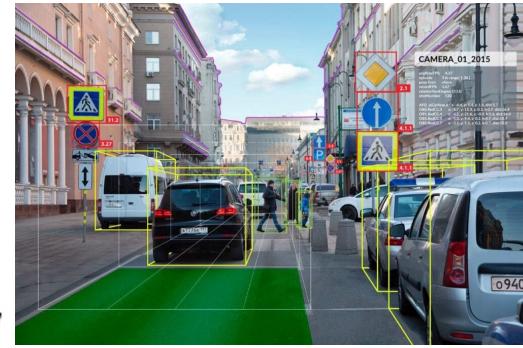
2000

# Computer Vision vs Computer Graphics

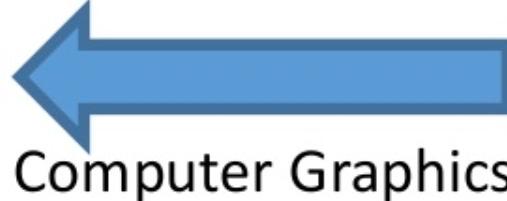
Digital Image Processing  
Computational Photography



Images (2D)



Computer Vision

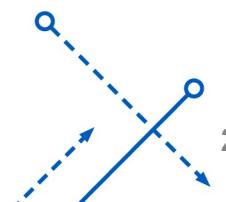


Computer Graphics

Geometry (3D)  
Shape



Photometry  
Appearance

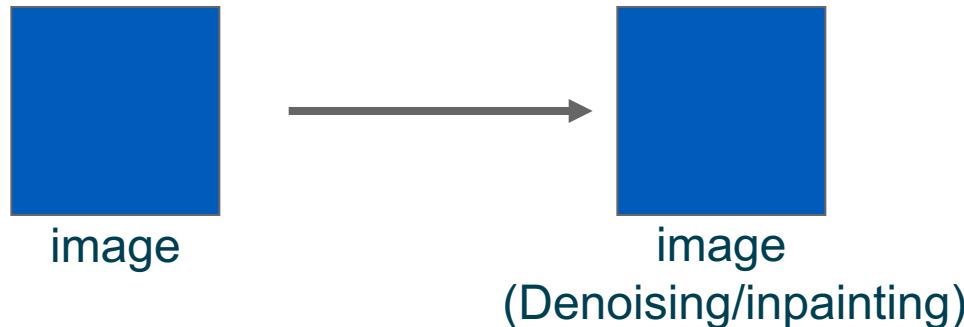


# Example: Visual Effects need CV + CG

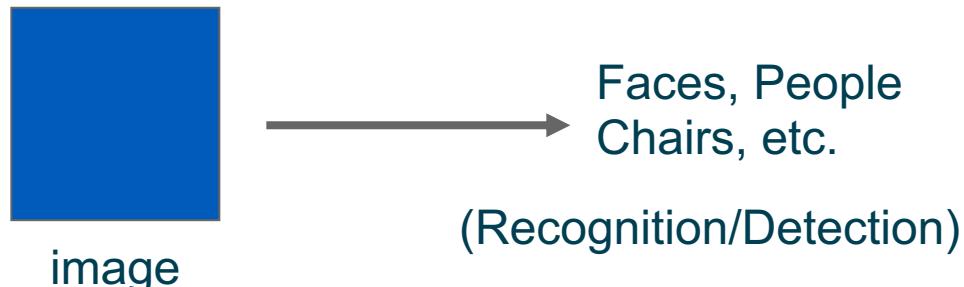


# Image Processing vs. Computer Vision

- Image Processing
  - Research area within electrical engineering/signal processing
  - Focus on syntax, low level features



- Computer Vision
  - Research area within computer science/artificial intelligence
  - Focus on semantics, symbolic or geometric descriptions



# Image & Goal

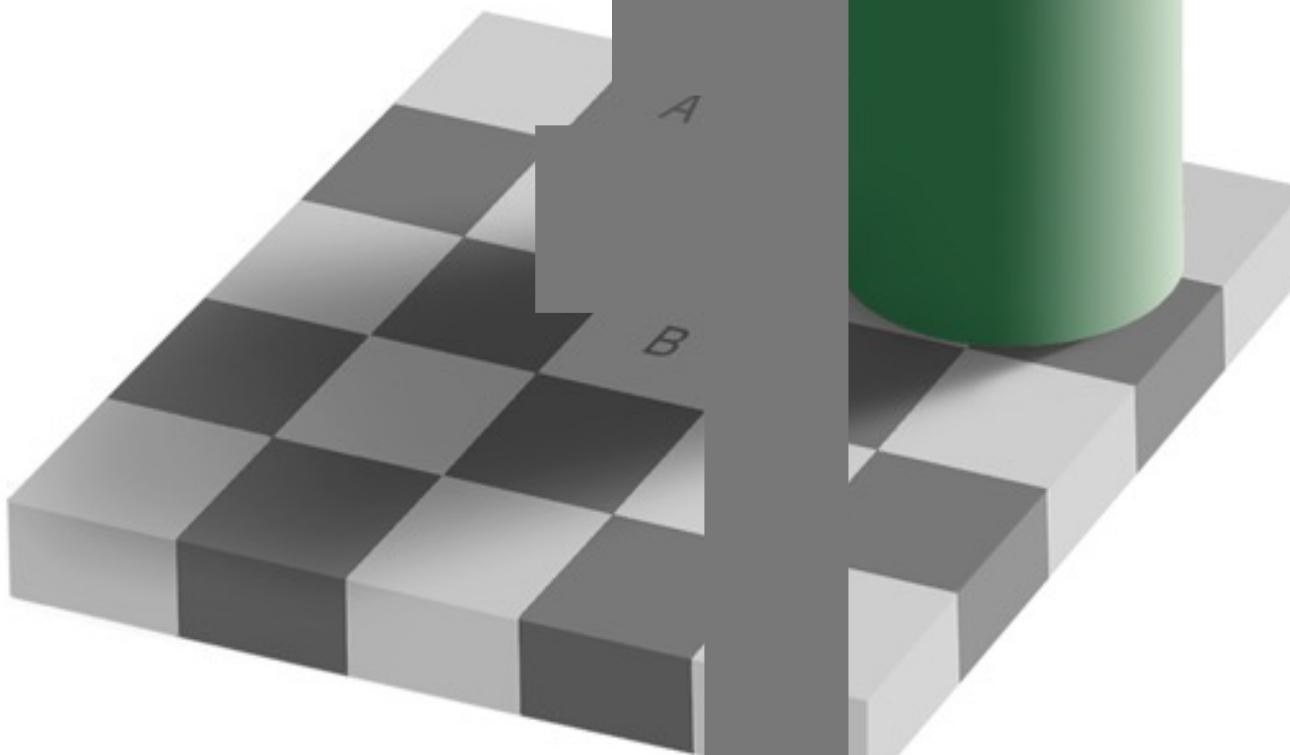


An image is an array of numbers (pixels).

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

- To Bridge the Gap between Pixels and Meaning

# Which one is brighter? A or B?

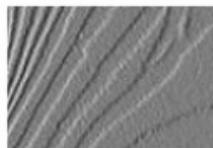


Edward H. Adelson

# Feature Extraction

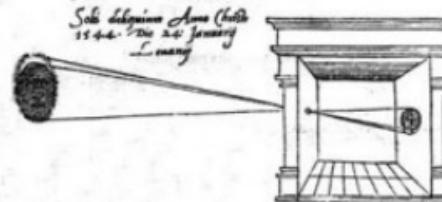


- Basic image processing and image formation

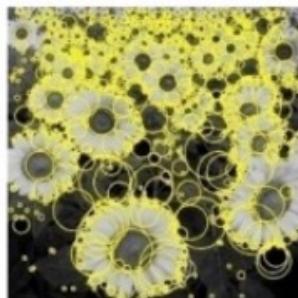


Filtering, edge detection

illum in tabula per radios Solis , quām in celo contin-  
git: hoc est, si in celo superior pars deliquit patiatur, in  
radius apparet inferior deficere, vt ratio exigit optica.



Sic nos exarēt Anno . 1544 . Louani ecliptīm Solis  
obseruauimus , insuenimūs deficere paulò plus q̄ dext-



Feature extraction

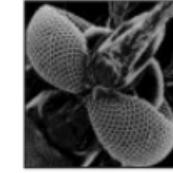
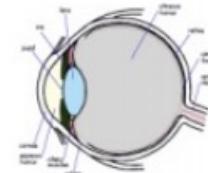
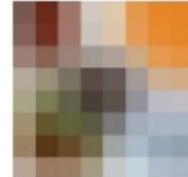
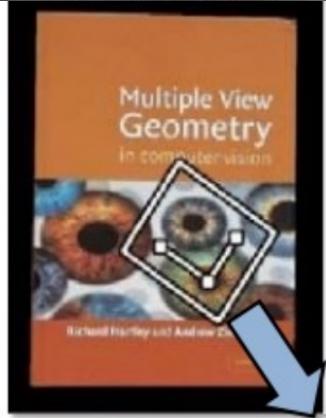


Image formation

# Feature Matching



# Recognition and Detection

**Classification**



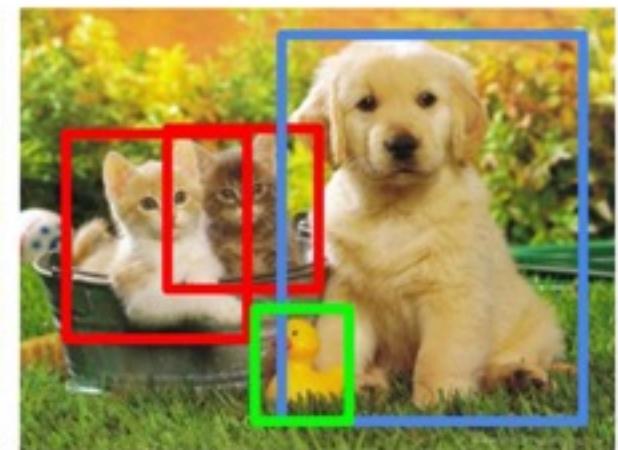
CAT

**Classification + Localization**



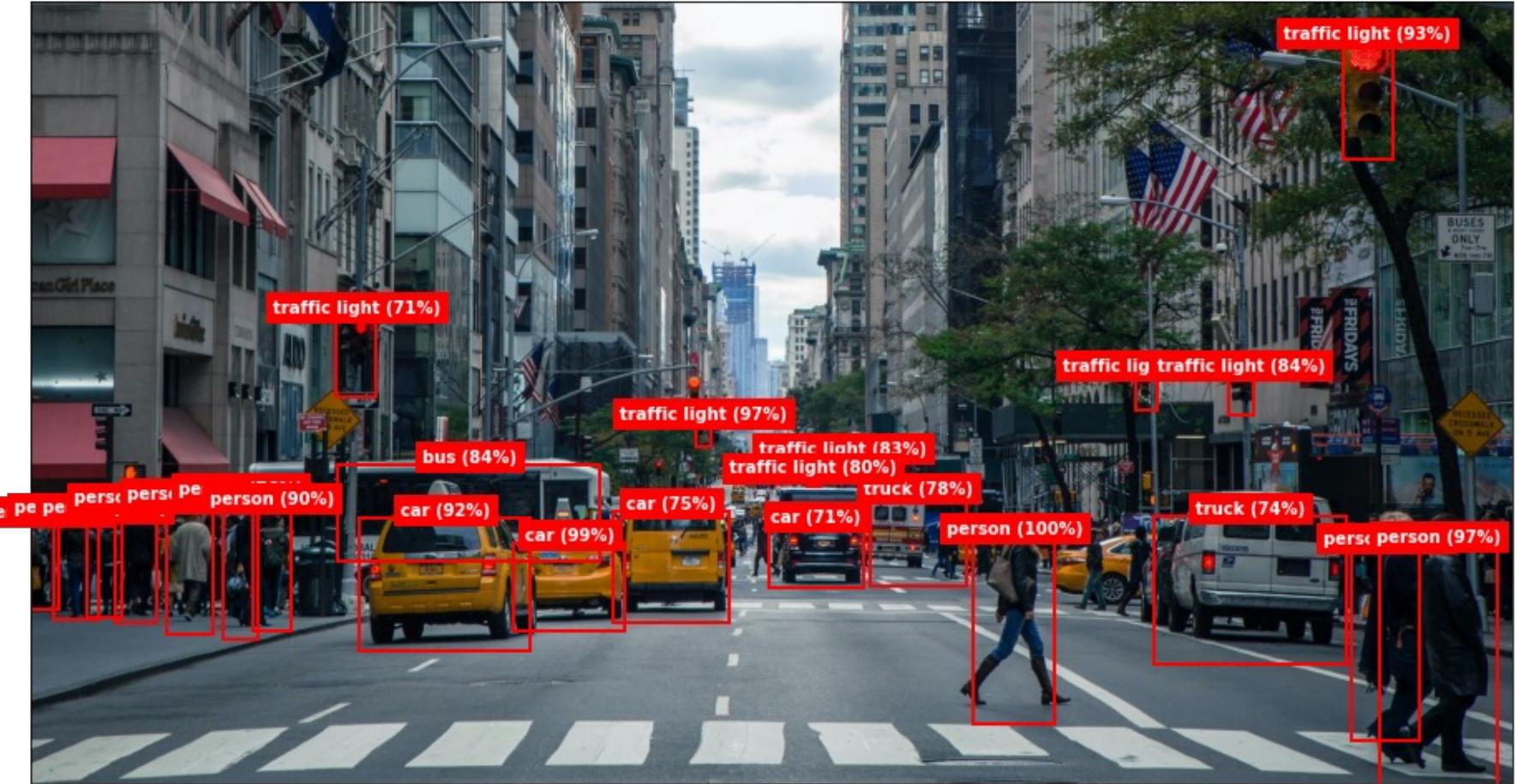
CAT

**Object Detection**



CAT, DOG, DUCK

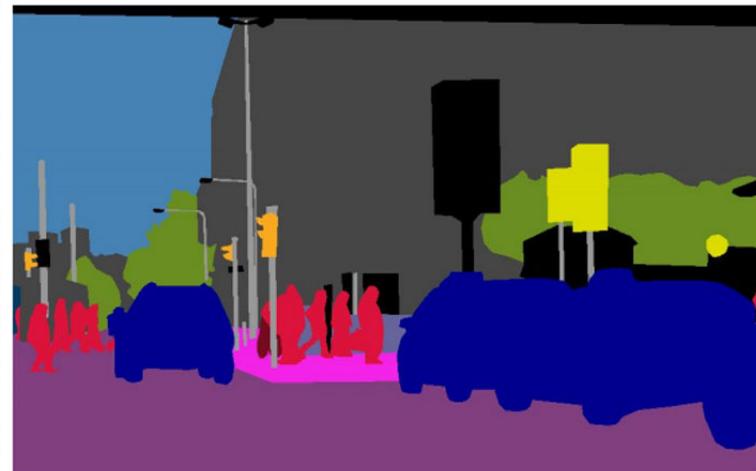
# Detection



# Segmentation



(a) image



(b) semantic segmentation

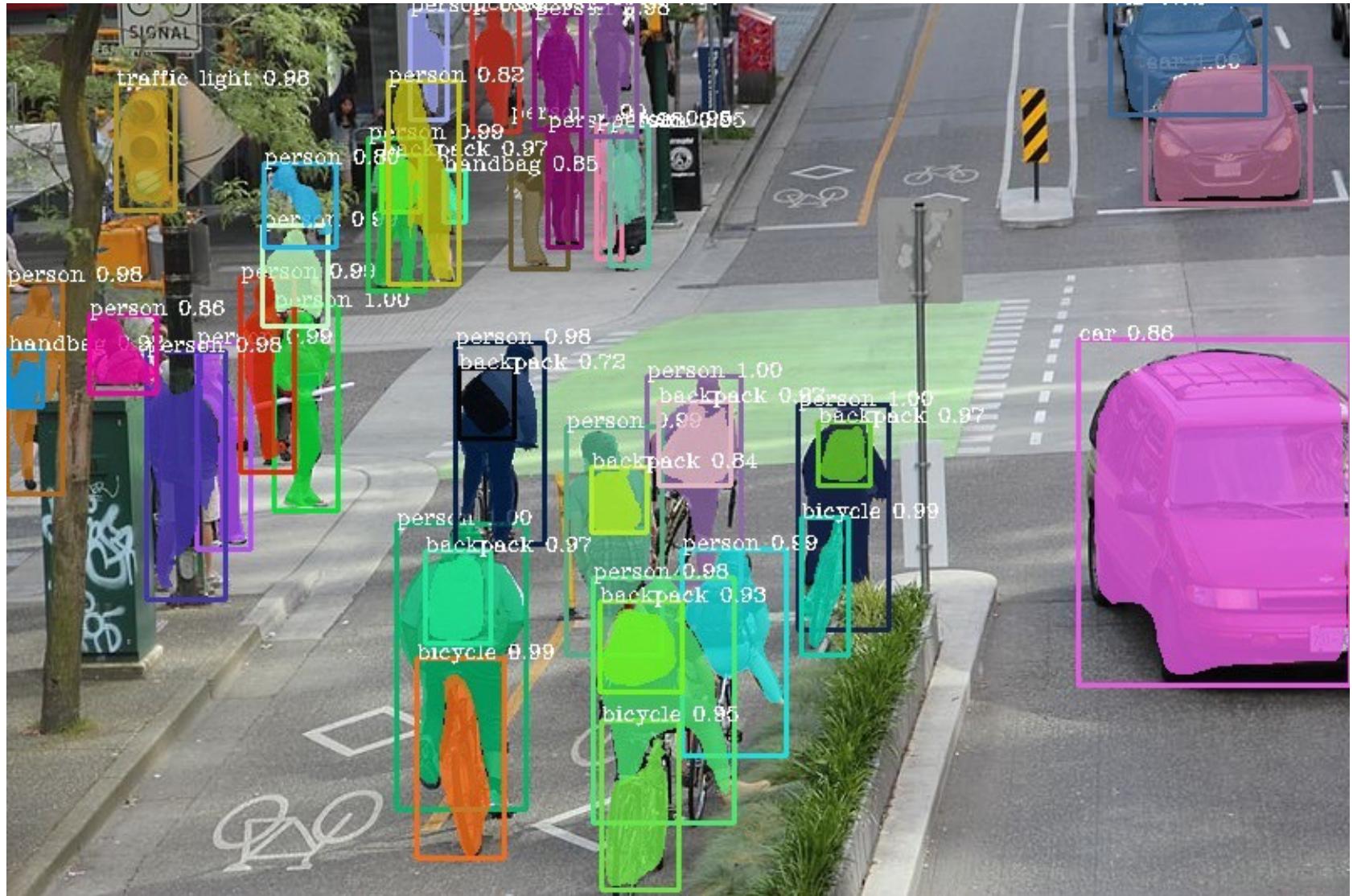


(c) instance segmentation



(d) panoptic segmentation

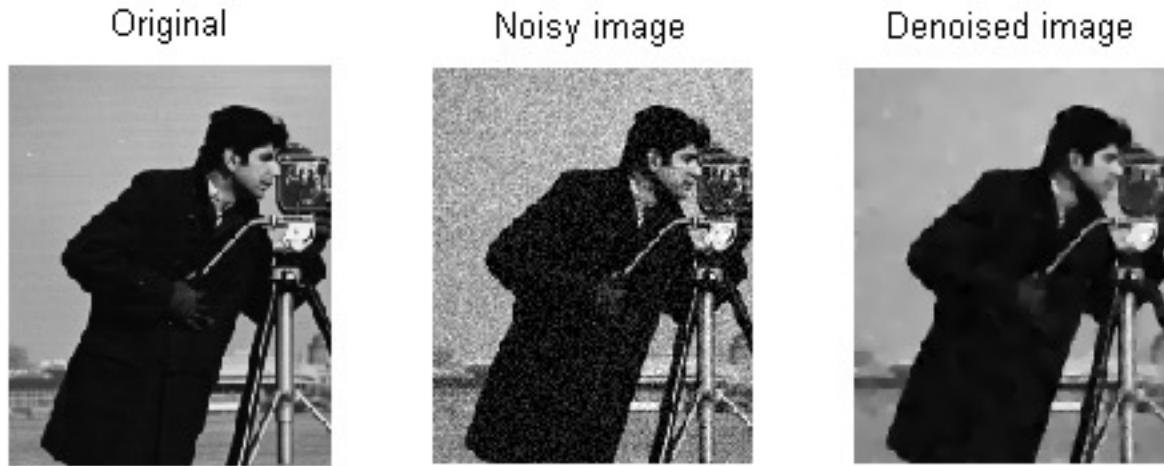
# Detection + Segmentation



# Inpainting



# Denoising (Dehazing, Deblurring)



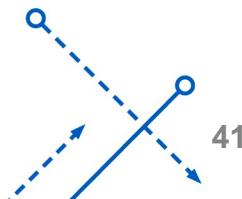
(a) Haze

(b) AOD-Net



(c) GFN

(d) Ours



# Super Resolution

---



# Image Stitching: Panorama



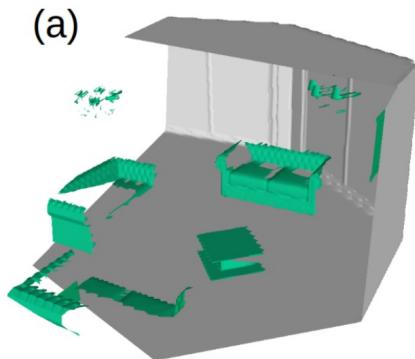
# Example: Google Street View



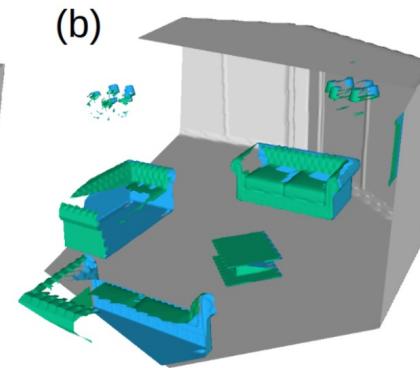
# 3D Reconstruction (RGB/RGBD)



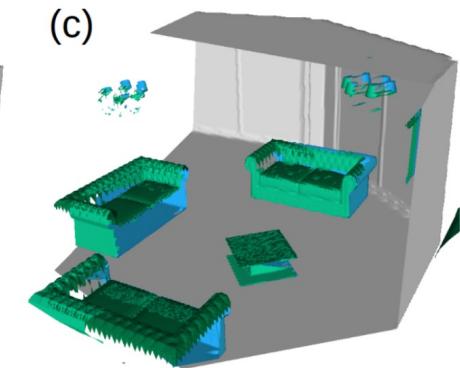
RGB Image



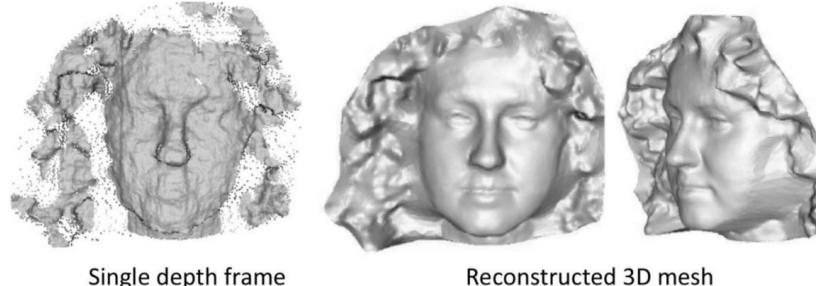
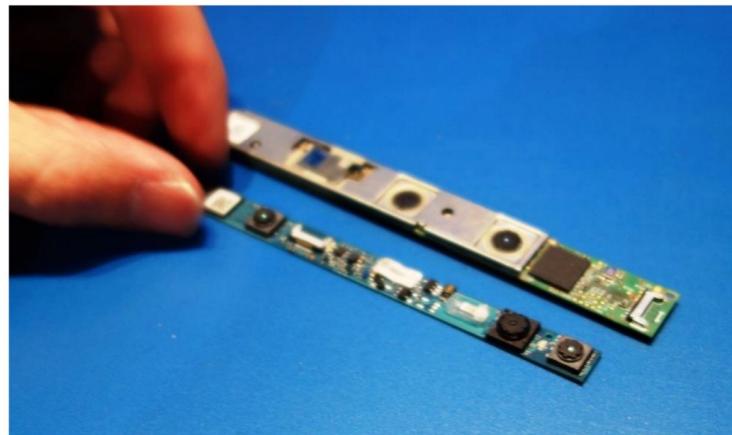
2.5D Object Surfaces



Multi-layer Surfaces



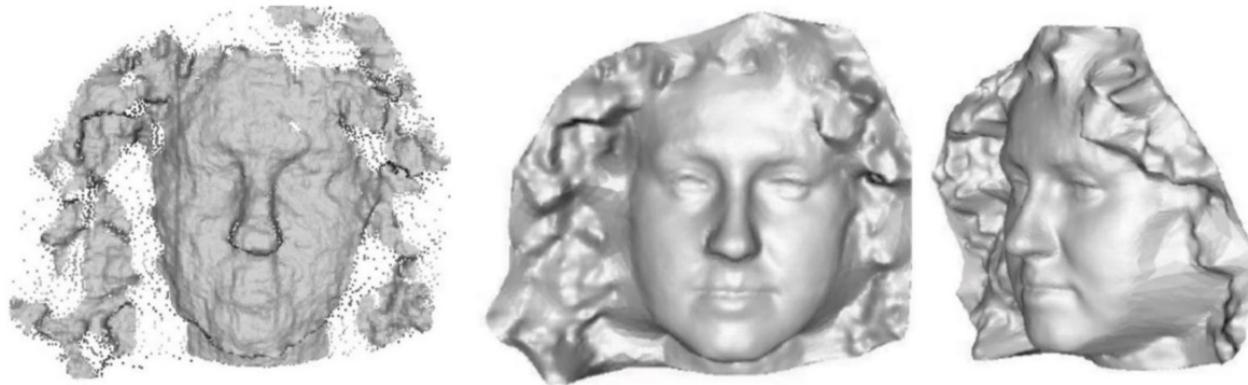
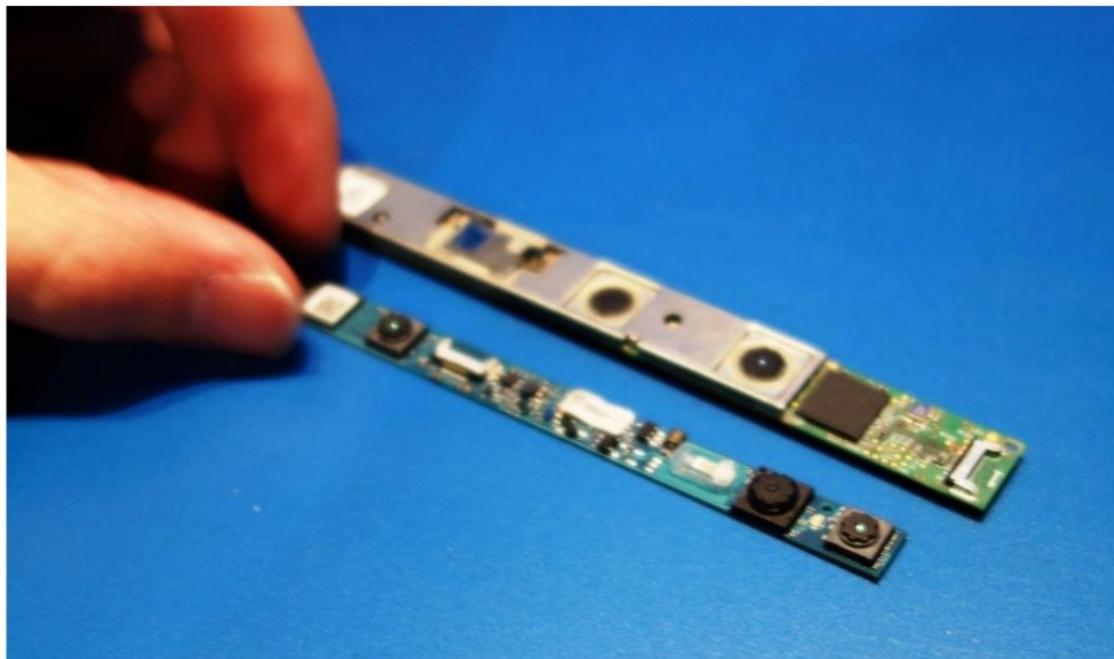
Multi-layer and  
Virtual-view Surfaces



Single depth frame

Reconstructed 3D mesh

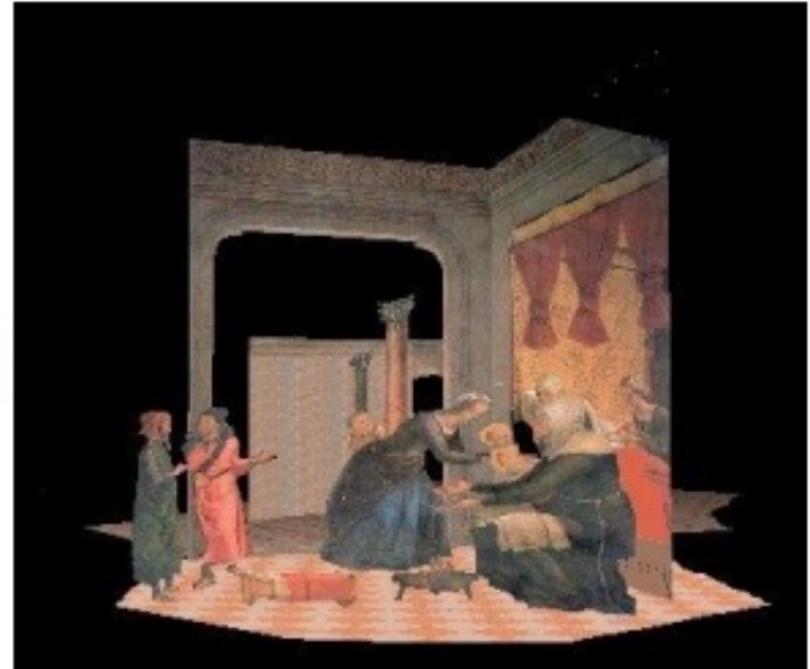
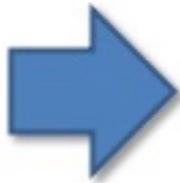
# Shape Reconstruction using depth sensors



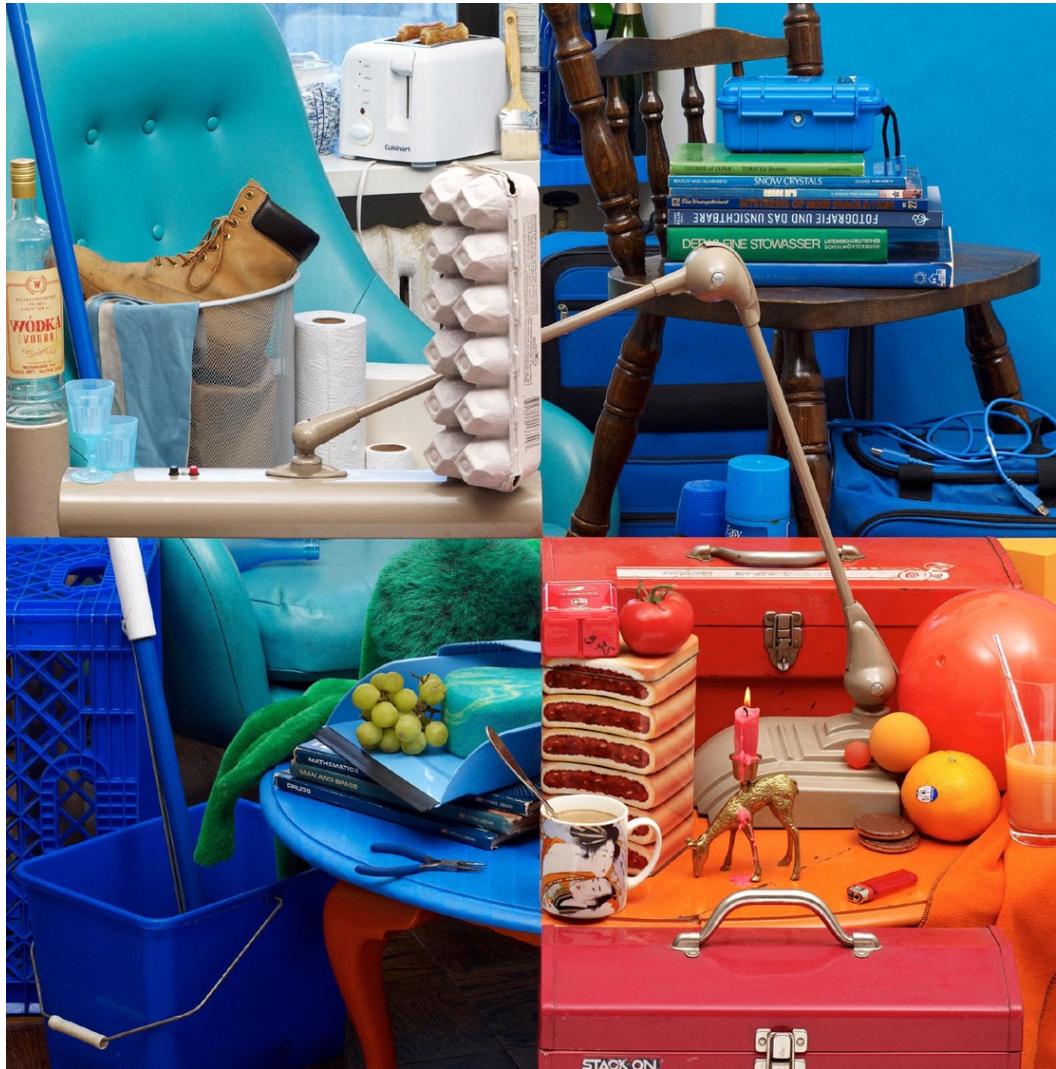
Single depth frame

Reconstructed 3D mesh

# Single View Modeling



# Can you do single view modeling?



# Can you do single view modeling?



# Visual Question and Answering (VQA)



What color are her eyes?  
What is the mustache made of?



How many slices of pizza are there?  
Is this a vegetarian pizza?



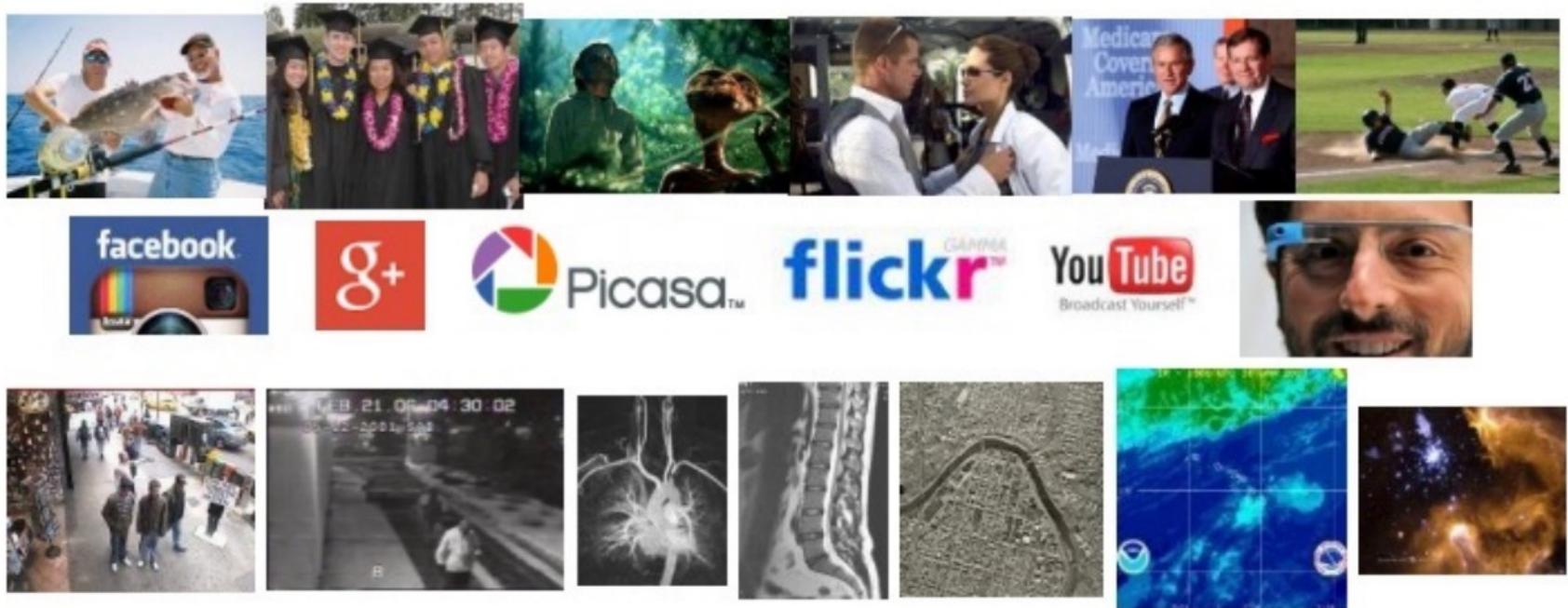
Is this person expecting company?  
What is just under the tree?



Does it appear to be rainy?  
Does this person have 20/20 vision?

# Why learn Computer Vision?

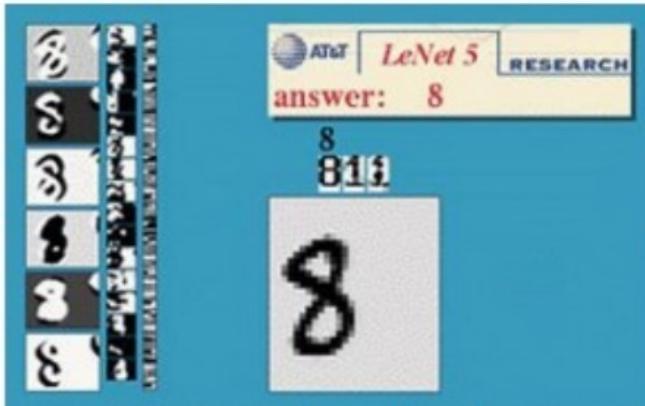
- One third of human brain devoted to vision.
- Millions of images being captured all the time.



What are some other Applications?

# Everyday Applications - OCR

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



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



Automatic check processing

4YCH428

4YCH428

4YCH428

License plate readers

[http://en.wikipedia.org/wiki/Automatic\\_number\\_plate\\_recognition](http://en.wikipedia.org/wiki/Automatic_number_plate_recognition)



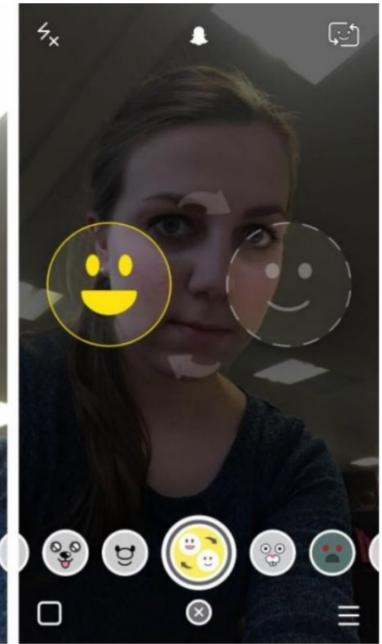
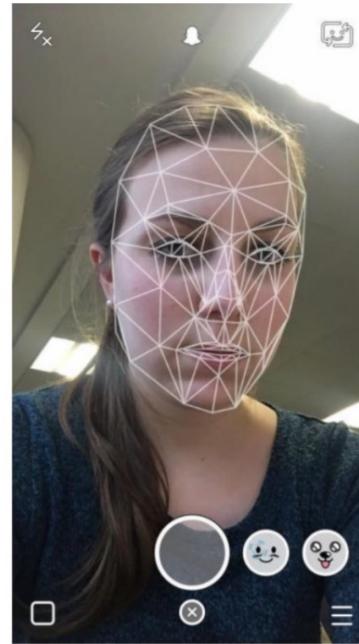
Sudoku grabber  
<http://sudokugrab.blogspot.com/>

Source: S. Seitz

# Object Recognition



# Face Detection



# Face Recognition



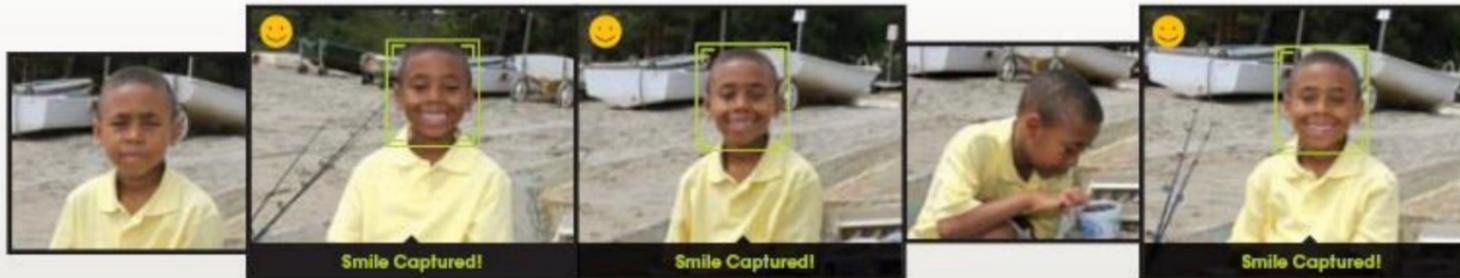
How to solve this problem?



# Smile Detection

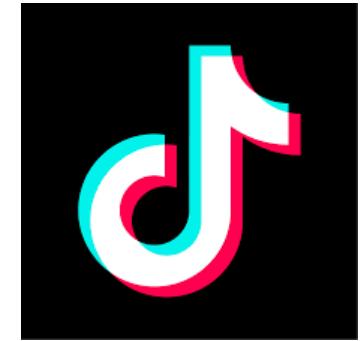
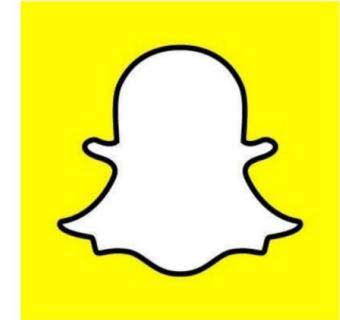
## The Smile Shutter flow

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.

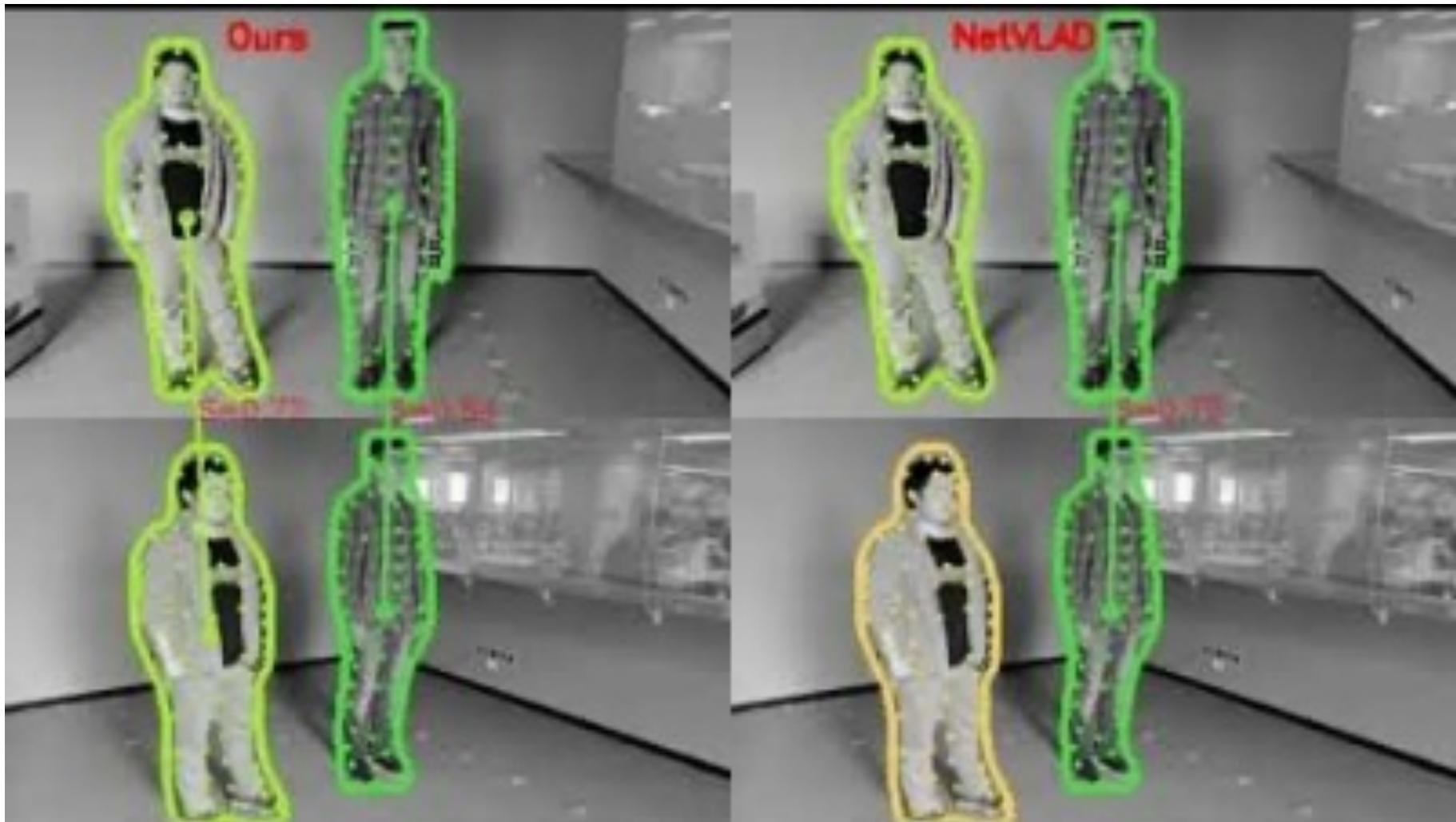


# Entertainment

---



# Person Re-identification



# Search and Rescue (Few-shot Detection)



# Autonomous Vehicle Navigation



# Smart cars

►► manufacturer products

consumer products ◀◀

News

## Our Vision. Your Safety.

rear  
looking  
camera



forward  
looking  
camera

side looking camera

### › EyeQ Vision on a Chip



[› read more](#)

### › Vision Applications

Road, Vehicle,  
Pedestrian Protection  
and more



[› read more](#)

### › AWS Advance Warning System



[› read more](#)

› [Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System](#)

› [Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end](#)

[› all news](#)



Events

› [Mobileye at Equip Auto, Paris, France](#)

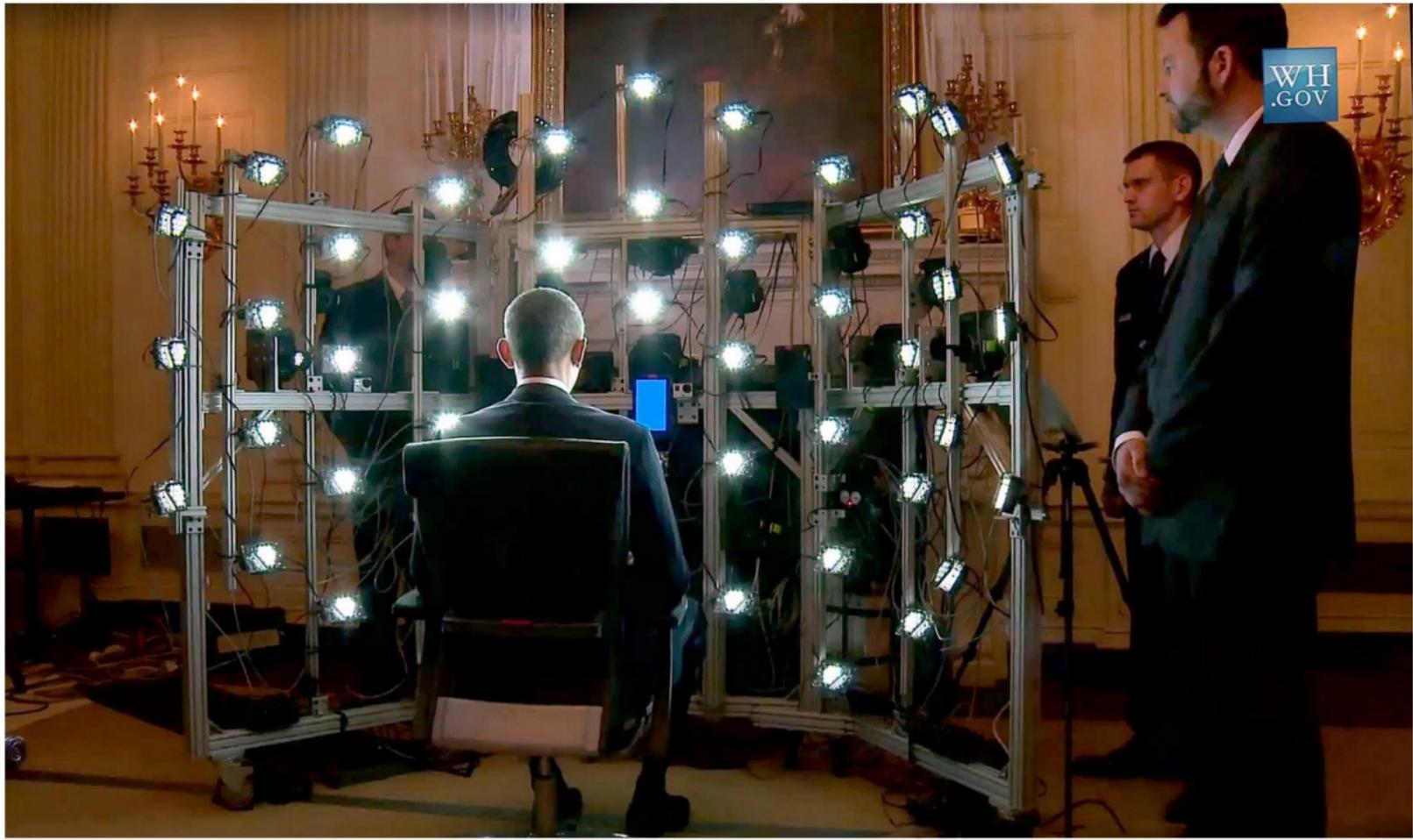
› [Mobileye at SEMA, Las Vegas, NV](#)

[› read more](#)

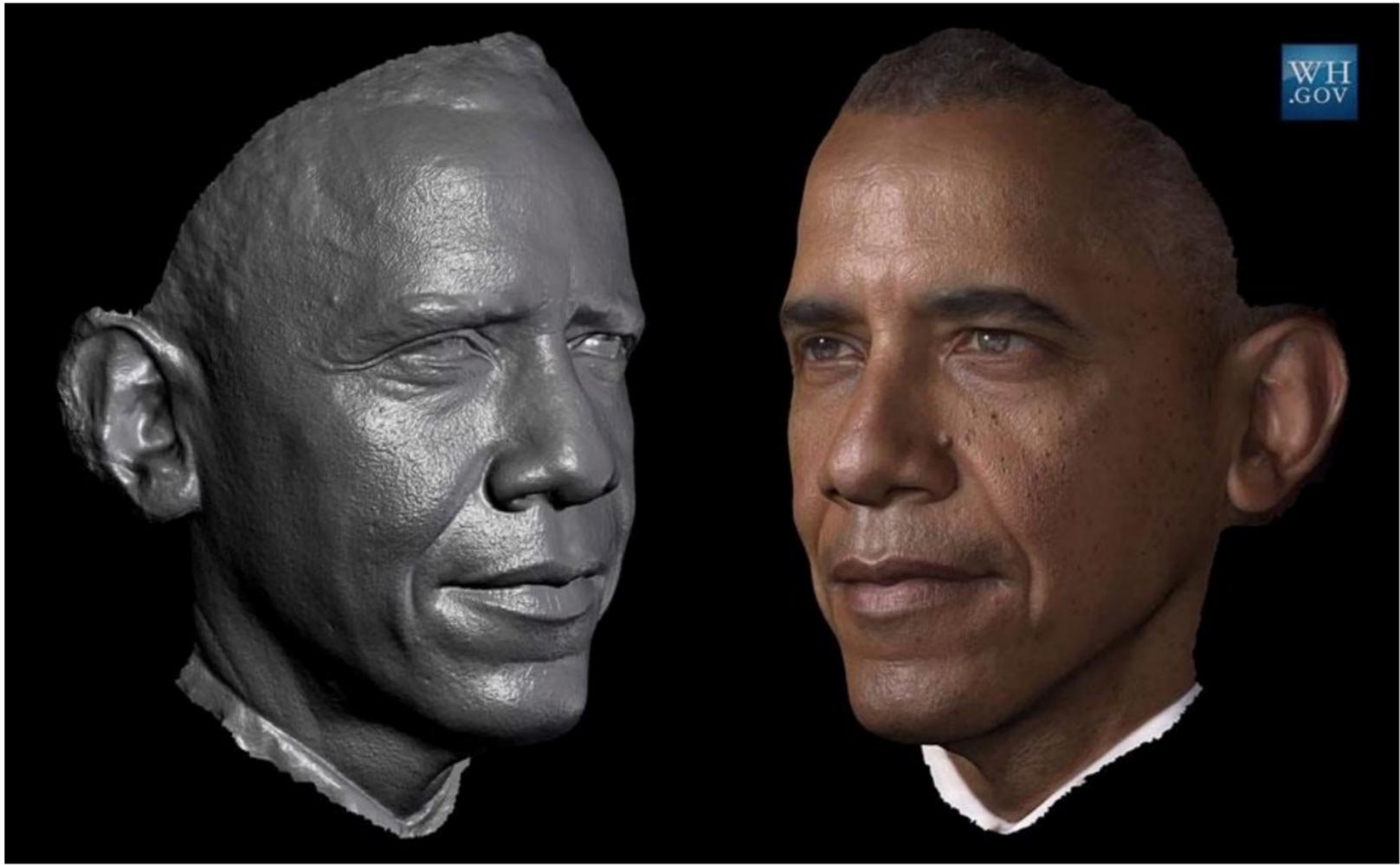
Tesla?

Shashua

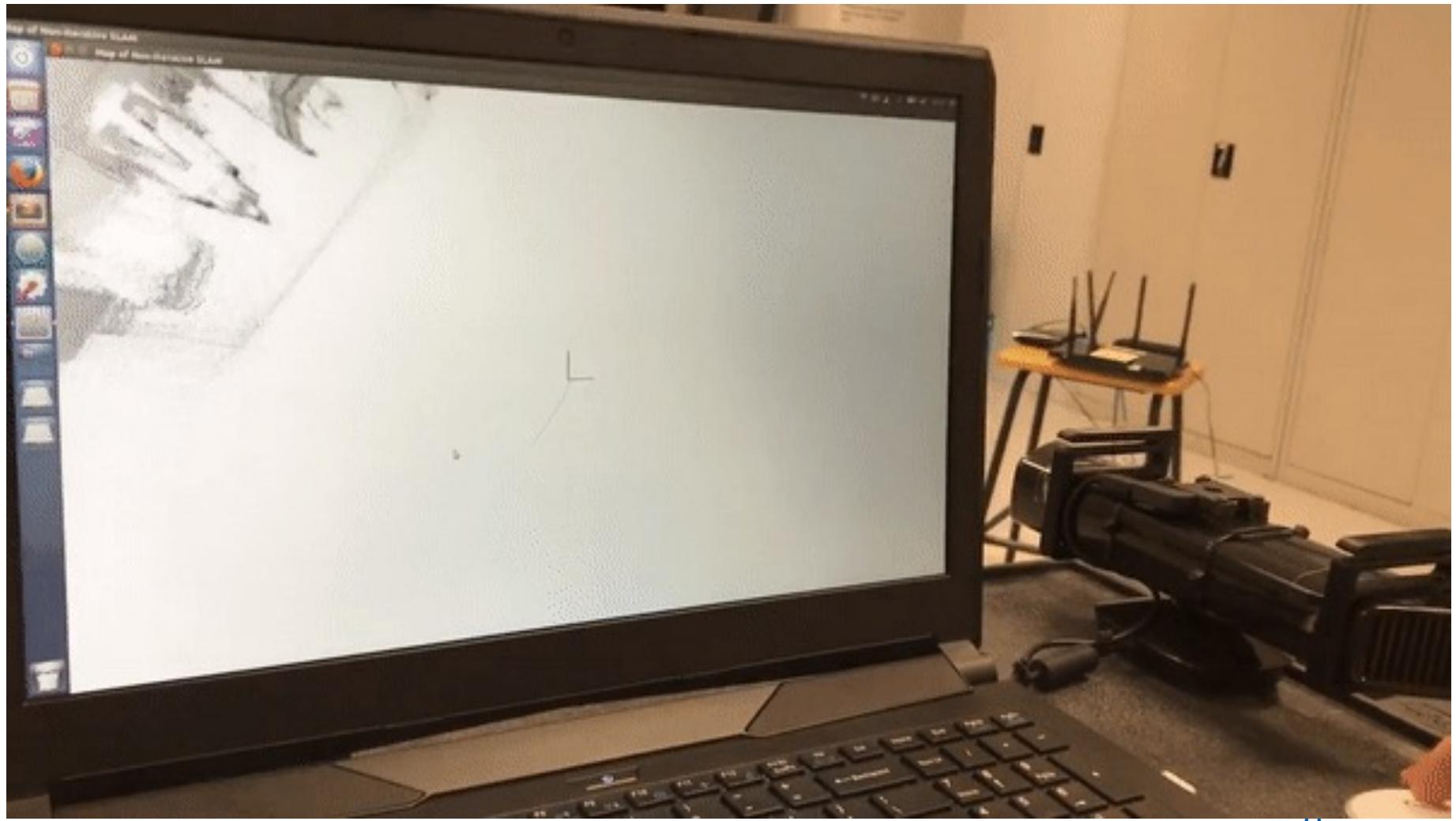
# Multi-Camera



# 3D Scanning



# Room Reconstruction



# Simultaneously Localization and Mapping (SLAM)

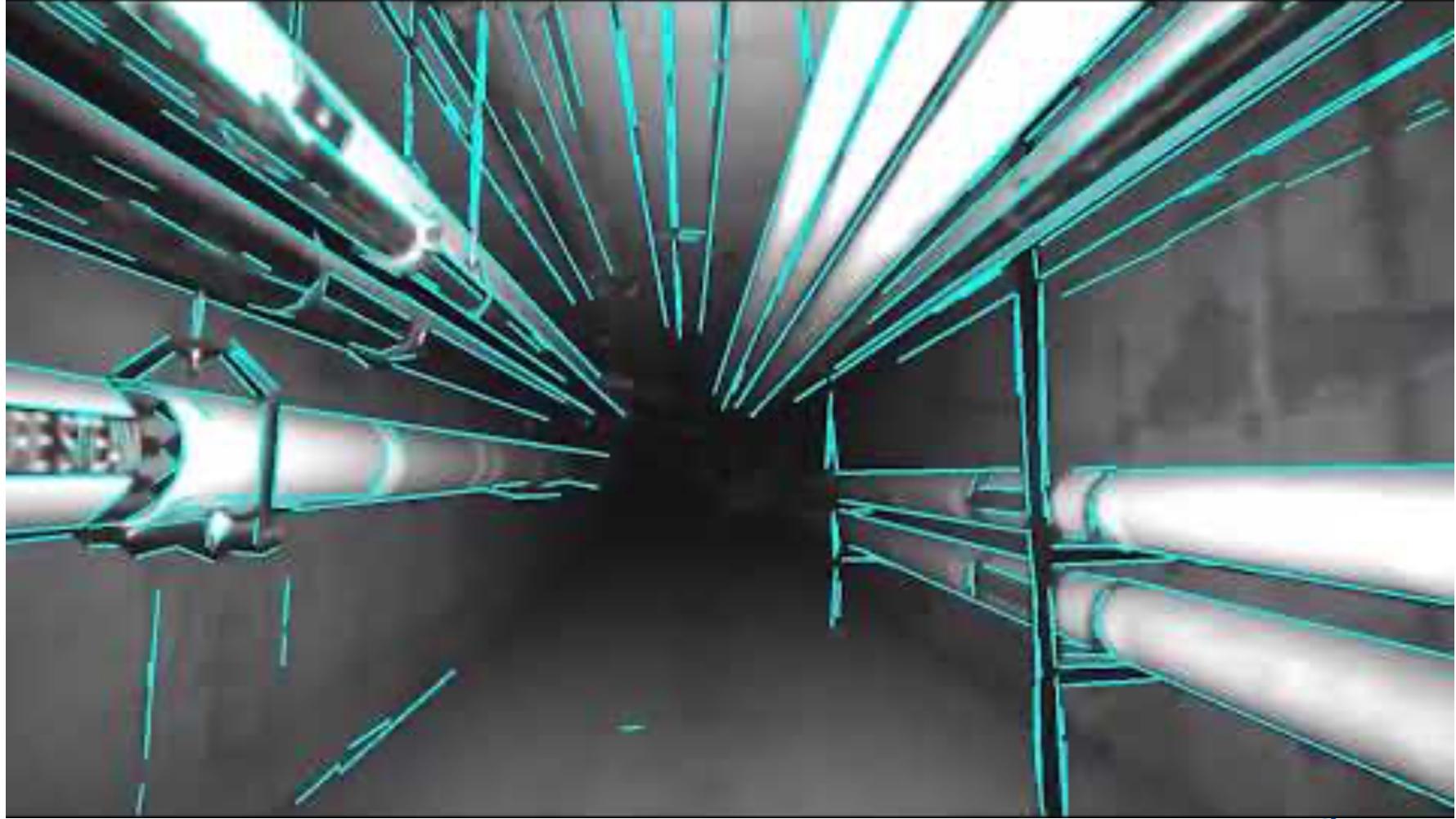


Fast-LOAM



Credit: Han Wang

# Line Detection



Credit: Xiao Lin

# Visual Odometry



NANYANG  
TECHNOLOGICAL  
UNIVERSITY  
SINGAPORE

**Geek+**  
Moving the world intelligently



**SAIR**  
Spatial AI & Robotics Lab

## AirVO: An Illumination-Robust Point-Line Visual Odometry

Kuan Xu<sup>1</sup>, Yuefan Hao<sup>2</sup>, Shenghai Yuan<sup>1</sup>, Chen Wang<sup>3</sup>, and Libin Xie<sup>1</sup>, *Fellow, IEEE*

Kuan Xu, Shenghai Yuan and Libin Xie are with School of Electrical and Electronic Engineering, Nanyang Technological University.  
Yuefan Hao is with the Geek+ Corp., Beijing, China.

Chen Wang is with the Department of Computer Science and Engineering, State University of New York at Buffalo.

# Indoor Localization

---



## AirLoc: Object-based Indoor Relocalization

Aryan<sup>1</sup>, Bowen Li<sup>2</sup>, Sebastian Scherer<sup>2</sup>, and Chen Wang<sup>2,3</sup>

---

<sup>1</sup>The Department of Electronics and Communication, Delhi Technological University, Delhi, India

<sup>2</sup>The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA

<sup>3</sup>The Department of Computer Science and Engineering, State University of New York at Buffalo, NY 14260, USA

# Warehouse Robot



# Vision in Space

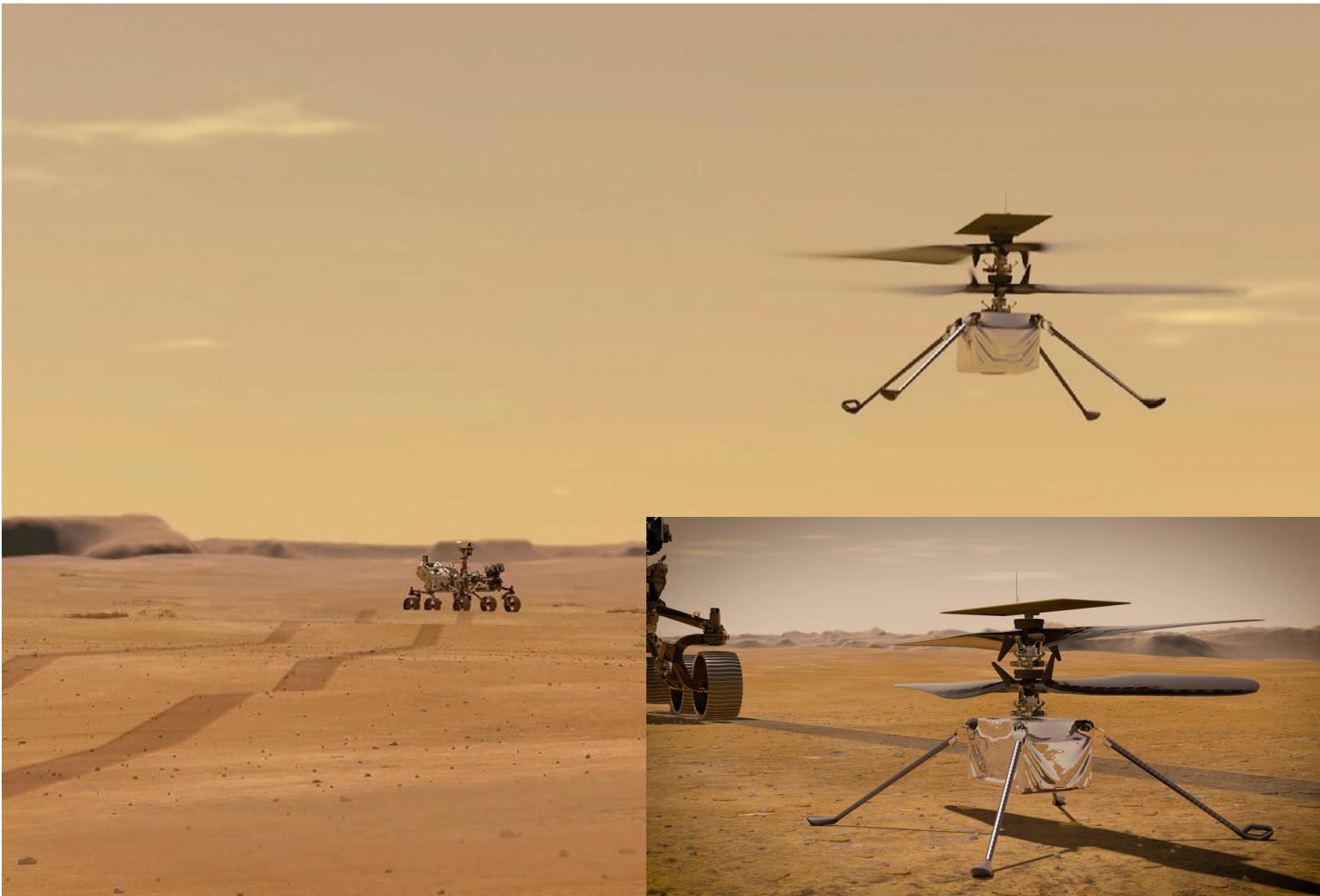


[NASA'S Mars Exploration Rover Spirit](#) captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

## 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.

# Mars Exploration



Feb. 18, 2021, Jezero Crater, Mars

Perseverance rover and Ingenuity Helicopter on Mars

# Techniques behind Mars Helicopter

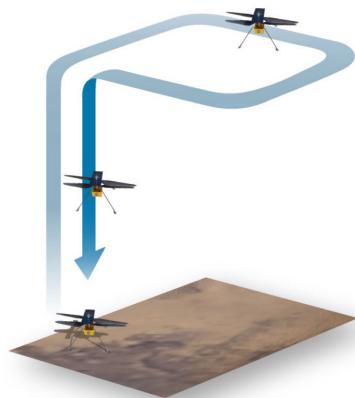
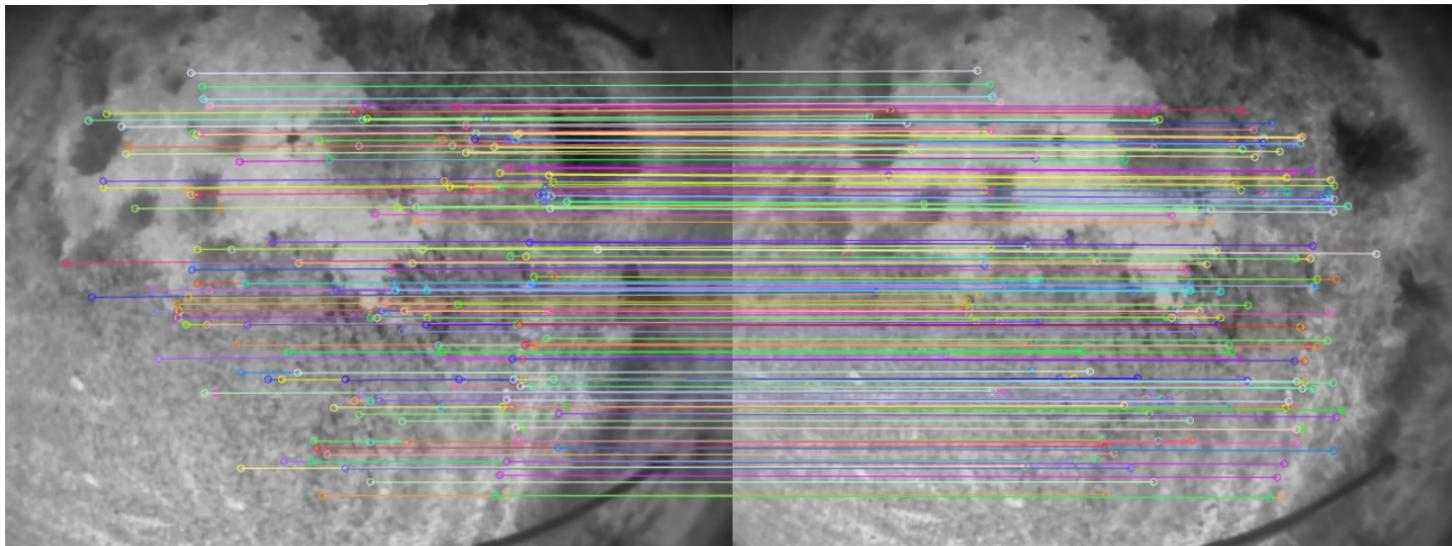
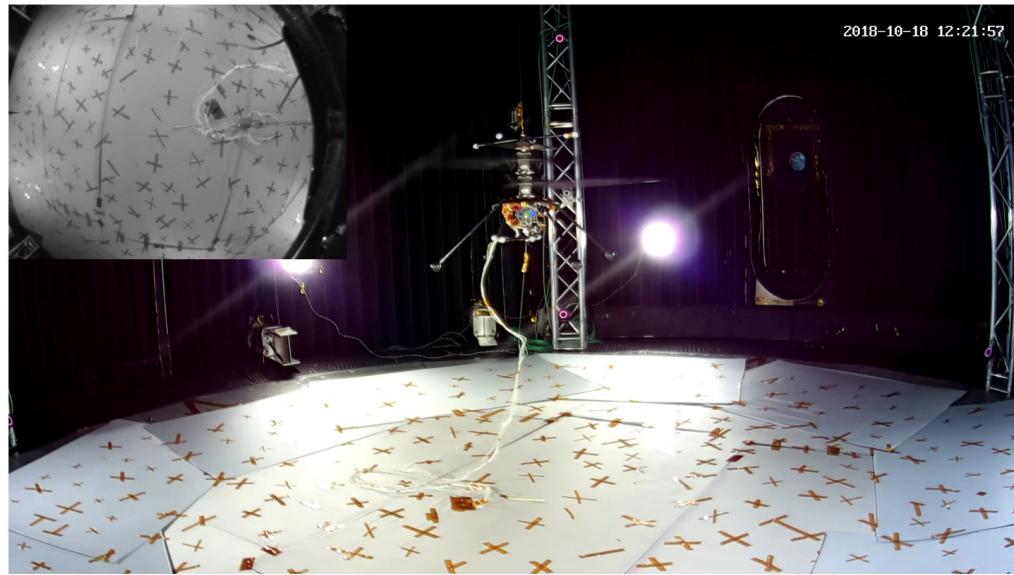


Fig. 2 Illustration of a Mars Helicopter flight, beginning and ending in the same pre-inspected safe area



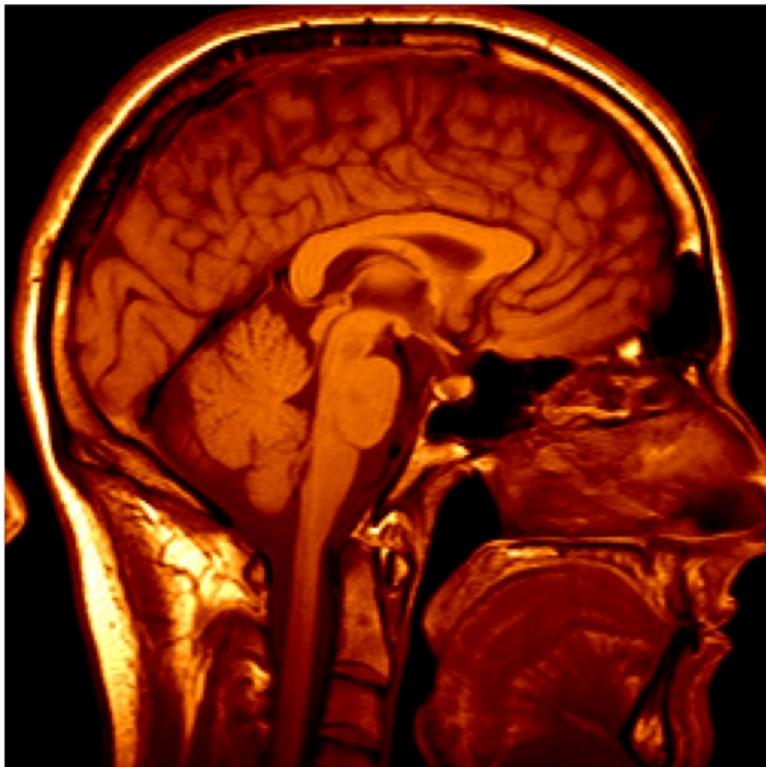
# Virtual/Augmented/Mixed Reality



# Video Surveillance



# Medical Imaging

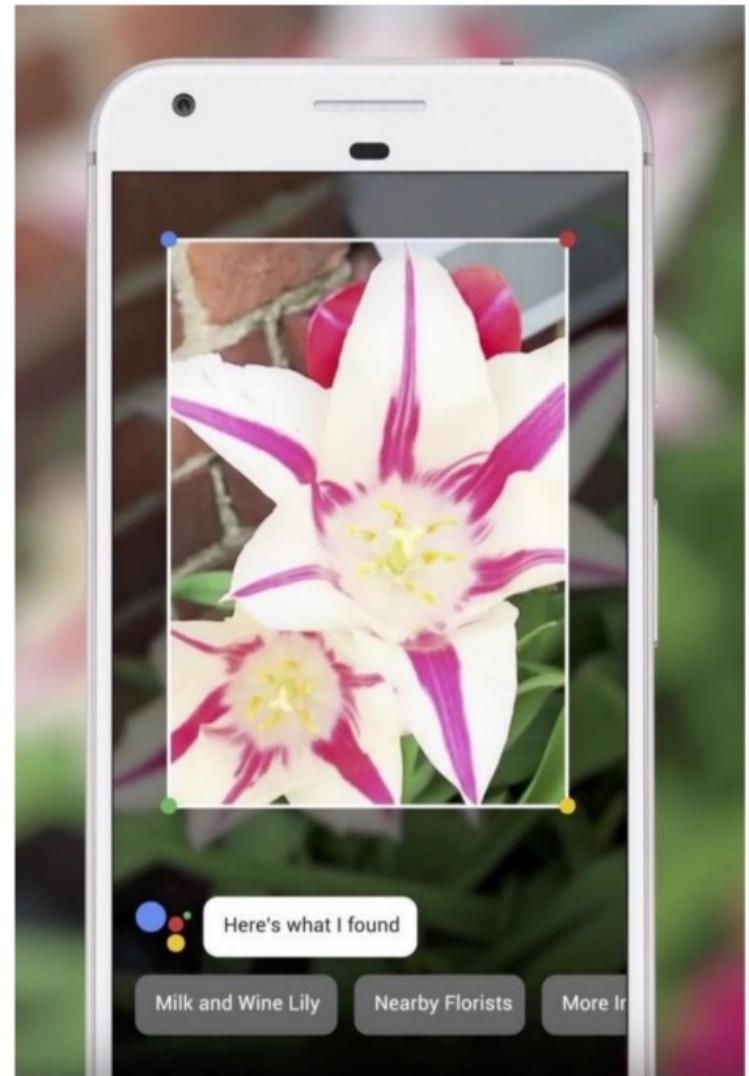
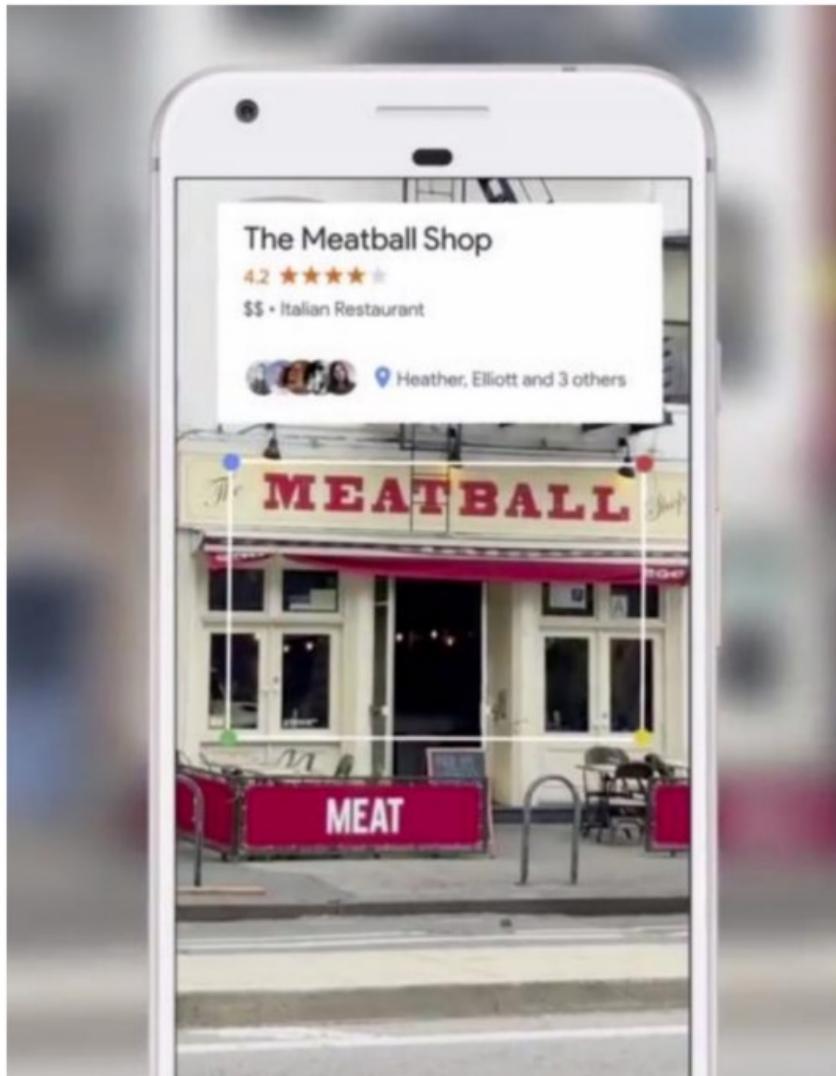


3D imaging  
MRI, CT

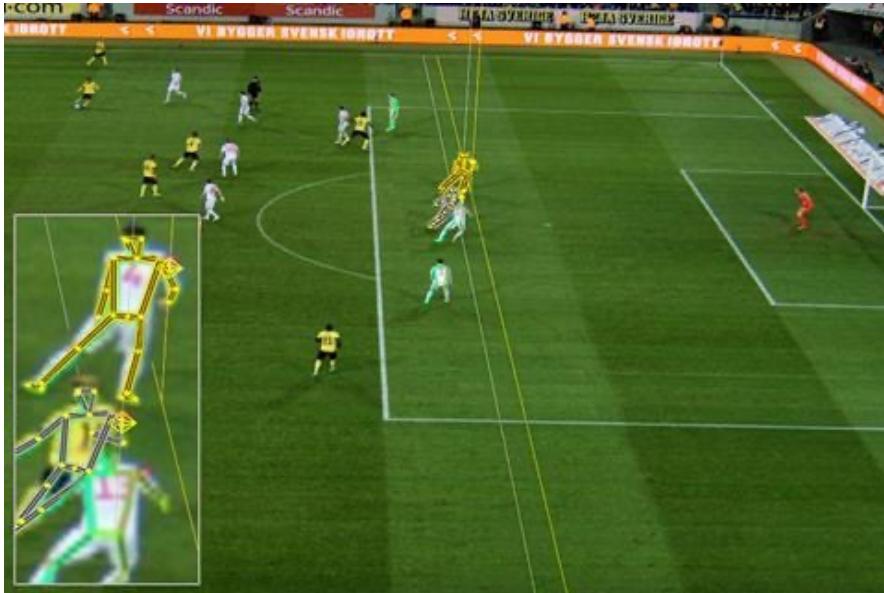


Image guided surgery  
Grimson et al., MIT

# Visual Search: Google Lens



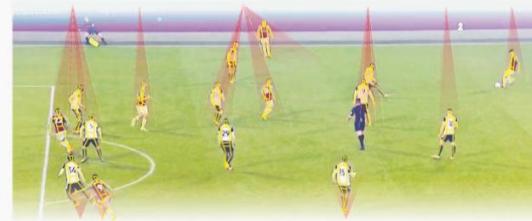
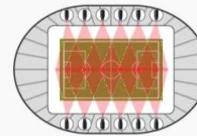
# Sports and Advertising



## Explained: How semi-automated offside will work

Fifa are planning to have semi-automated offside decisions using the latest technology at the 2022 World Cup in Qatar which would replace the current VAR offside system. Hawk-Eye believes its system, detailed below, is ready to go

- 1 Hawk-Eye will have 12 cameras around the pitch



- 2 Artificial Intelligence monitors 29 points on each player's body, using a special 'skeletal player tracking system'



- 3 AI will also track the ball's movement and the exact moment passes are made

- 4 Algorithms will be able to calculate whether an offside has occurred within 0.5 seconds of real time



- 5 The information is sent to the video assistant referee to check before the final decision is made

