EECS 4422/5323 COMPUTER VISION

Introduction

Course Website

- wiki.eecs.yorku.ca/course/4422
- Check the 'What's new' page regularly for announcements.

EECS 4422/5323

- This course introduces the fundamental concepts of computer vision, with a balance of theory and practical application.
- Topics:
  - Introduction
  - Image formation
  - Image processing
  - Feature detection & matching
  - Segmentation
  - Dense motion estimation
  - Feature-based alignment
  - 3D - motion
  - 3D - stereo
  - 3D - single view
  - Recognition

Prerequisites

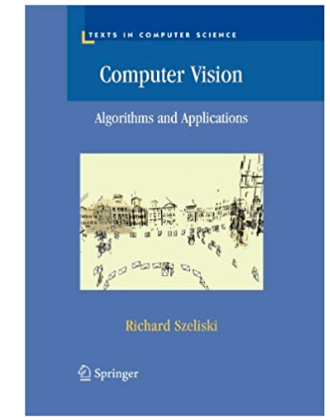
- Multi-variable calculus
- Linear algebra
- Probability
- MATLAB

Instructor Team

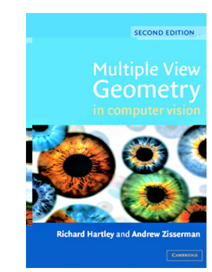
- Instructor
  - James Elder
  - Office: LAS 003G
  - Office hours: W 11:30AM-12:30PM
  - Email: jelder@yorku.ca
- Teaching Assistant
  - Hakki Karaimer
  - Office: LAS 2052
  - Office hours: F 11:00AM - 12:00PM
  - Email: hakkicankaraimer@gmail.com

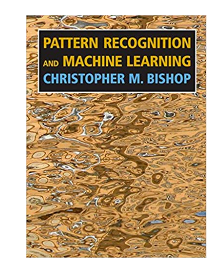
Textbooks

- Required:
  - Computer Vision: Algorithms and Applications, Szeliski R, 2011
    - Available online!



- Supplementary (Optional):
  - Multiple View Geometry in Computer Vision, Hartley R & Zisserman A, 2004
  - Pattern Recognition and Machine Learning, Bishop CM, 2006
  - Computer Vision: Models, Learning and Inference, Prince SJD, 2012
    - Available online!





Lectures

- MW 10:00AM - 11:30AM in Calumet 106.
- Students are responsible for all material covered in lectures.
- Lecture slides will be posted incrementally on this website.

Labs

- Six two-hour labs held in Bergeron 211.
- 4pm-6pm on selected Mondays - see schedule.
- Laptop-based: please bring your own laptop, or sign one out from the department.
- Desks do not have power - please make sure laptop batteries are fully charged or bring a long power cable to plug into the wall.
- Primary software environment: MATLAB.
- The TA will be present at each lab to provide demonstrations and guidance.
- Students are responsible for all material presented in labs.
- Labs 1, 2, 4 and 5 will be associated with the four short assignments.
- Lab 3 will be used to help students prepare for the midterm.
- The final lab will be used to demo projects and to discuss project reports.

Lab	Date
1	Sept 17
2	Oct 1
3	Oct 22
4	Nov 5
5	Nov 19
6	Dec 3

Assignments

- There will be four short assignments that include theory and coding questions
- Due dates can be found on the schedule page.

Midterm

- In class, closed book.
- Lab 3 will be used to help students prepare.

Project

- Students will complete a project involving the implementation of one or more computer vision algorithms of their choosing.

- Typically, projects will involve analytic, implementation and evaluation components.

- EECS 4422 students may decide to implement and evaluate one or more algorithms from the literature (NB: do not select an algorithm for which source code is available online).

- EECS 5323 students are also expected to extend or innovate upon published algorithms.

- Projects will consist of four stages, as follows. Please see the Schedule for deadlines.

  - Project Idea: The student will propose a specific project idea in the form of a short paragraph, submitted via Moodle. The instructor will either accept the topic (possibly with modifications) or raise issues that the student will address in revision of the project idea. The final project selection is subject to the instructor's approval. Additional guidelines for project selection will be provided.

  - Proposal: The proposal is a one-page document submitted via Moodle that includes: (i) Motivation: why is the topic interesting? (ii) Datasets: what datasets will you use to study the problem? (iii) Proposed methods: what algorithm(s) do you plan to implement? (iv) Evaluation methodology and (v) Proposed timeline for completion.

  - Presentation/Demo: A brief slide presentation on the project will be made during one of the last two lecture meetings. There will also be an opportunity to demo the project during the final lab.

  - Final Report: Maximum 8 pages, excluding references, in CVPR 2018 format. More guidance on the structure of the final report will be provided.

Evaluation

- Four assignments (7.5% each)
- Midterm (30%)
- Project (40%)
  - Proposal (10%)
  - Presentation/Demo (10%)
  - Report (20%)

Readings

- Today
  - Chapter 1. Introduction (pp. 2-27)
- Next meeting
  - Chapter 2. Image Formation (pp. 30-59)

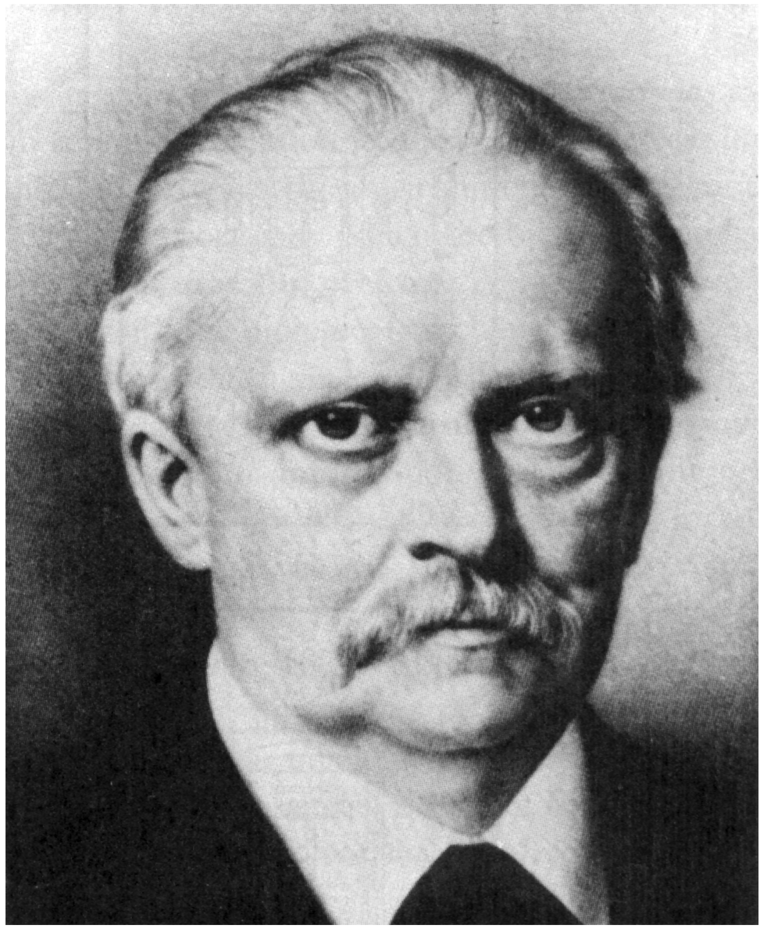
What is Vision?

An application of geometry?

Hasan Ibn al-Haytham (Latinized Alhazen)

- Two major theories on vision prevailed in classical antiquity.
- The emission theory was supported by such thinkers as Euclid and Ptolemy, who believed that sight worked by the eye emitting rays of light.
- The intromission theory supported by Aristotle and his followers, had physical forms entering the eye from an object.

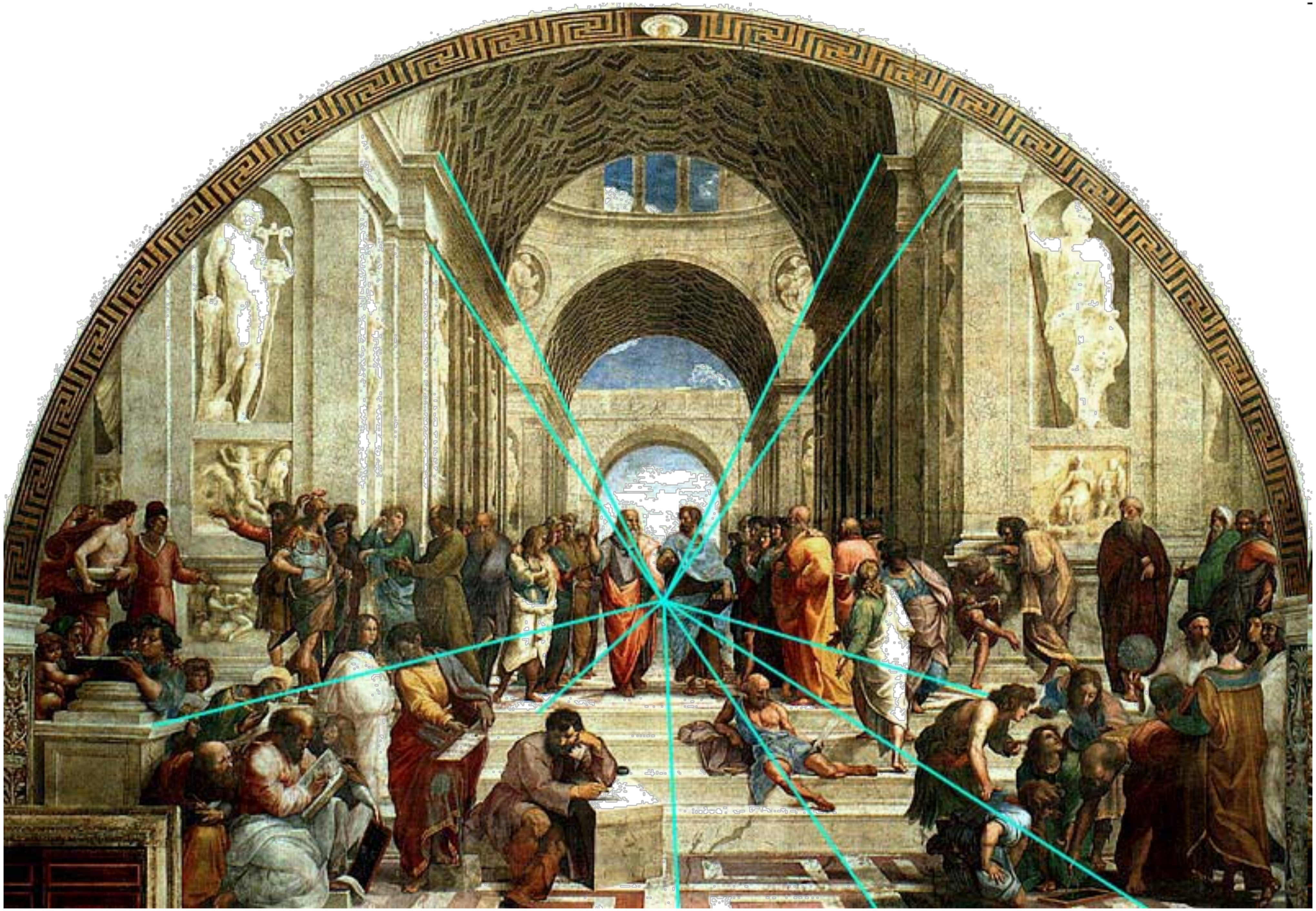
An application of probability?



Hermann von Helmholtz

- Percepts are the result of the brain's attempt to identify the underlying causal factors that produced the image.
- Any given image is consistent with multiple possible interpretations (vision is ill-posed).
- In general, the brain chooses the interpretation that is more probable (likelihood principle).

Linear Perspective



‘The School of Athens’ by Raphael (1518), Stanze di Raffaello, in the Apostolic Palace in the Vatican.

Ecological Statistics

JJ Gibson

Egon Brunswik

Ecological Statistics

- The brain evolved to use sensory signals to provide selective advantage within specific ecological niches.

JJ Gibson

Egon Brunswik

Ecological Statistics

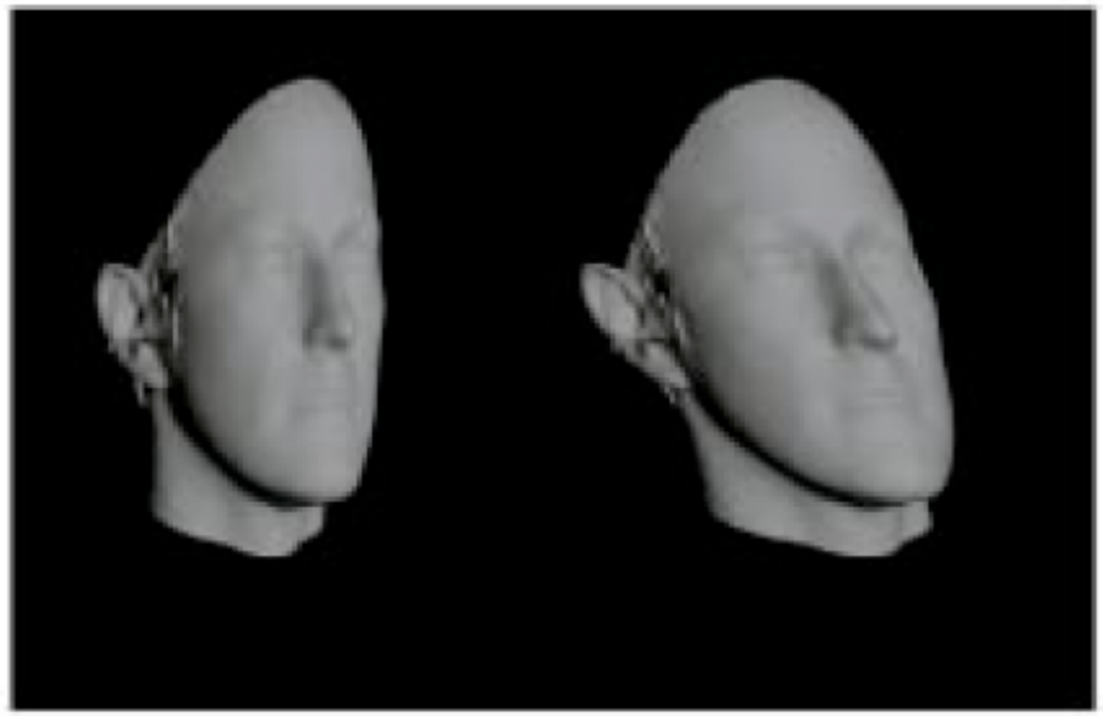
- The brain evolved to use sensory signals to provide selective advantage within specific ecological niches.
- Understanding vision thus requires the study not only of the brain but also of the **statistics of natural stimuli** in the sensory environment.

JJ Gibson

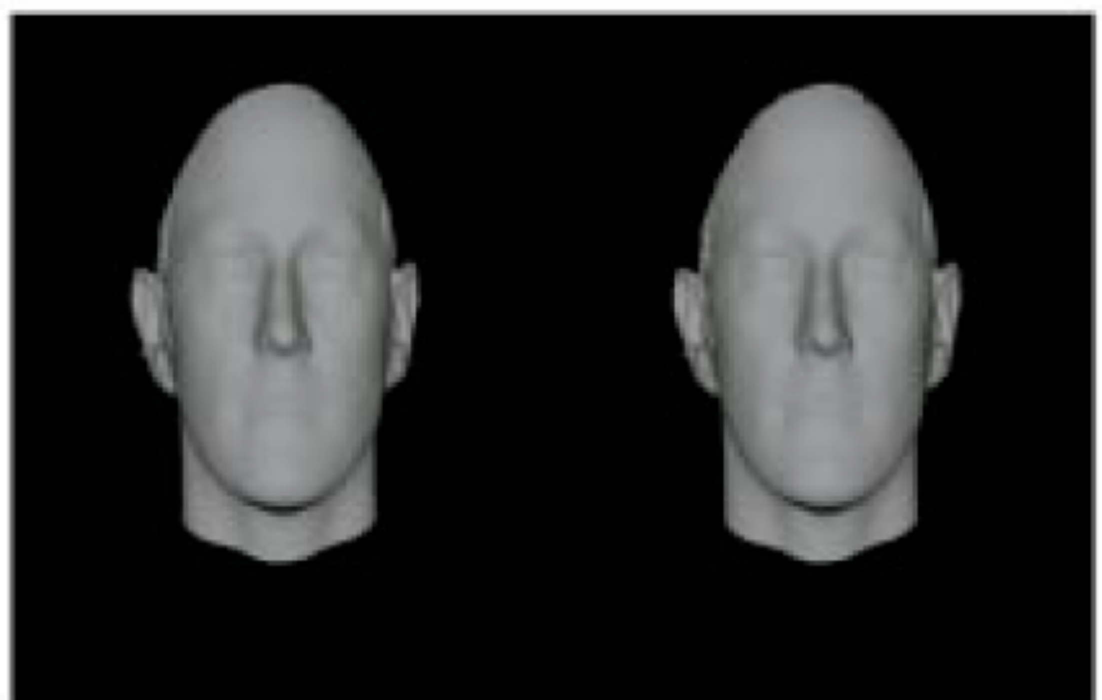
Egon Brunswik

Vision is Ill-Posed (Shape & Illumination)

Different Objects

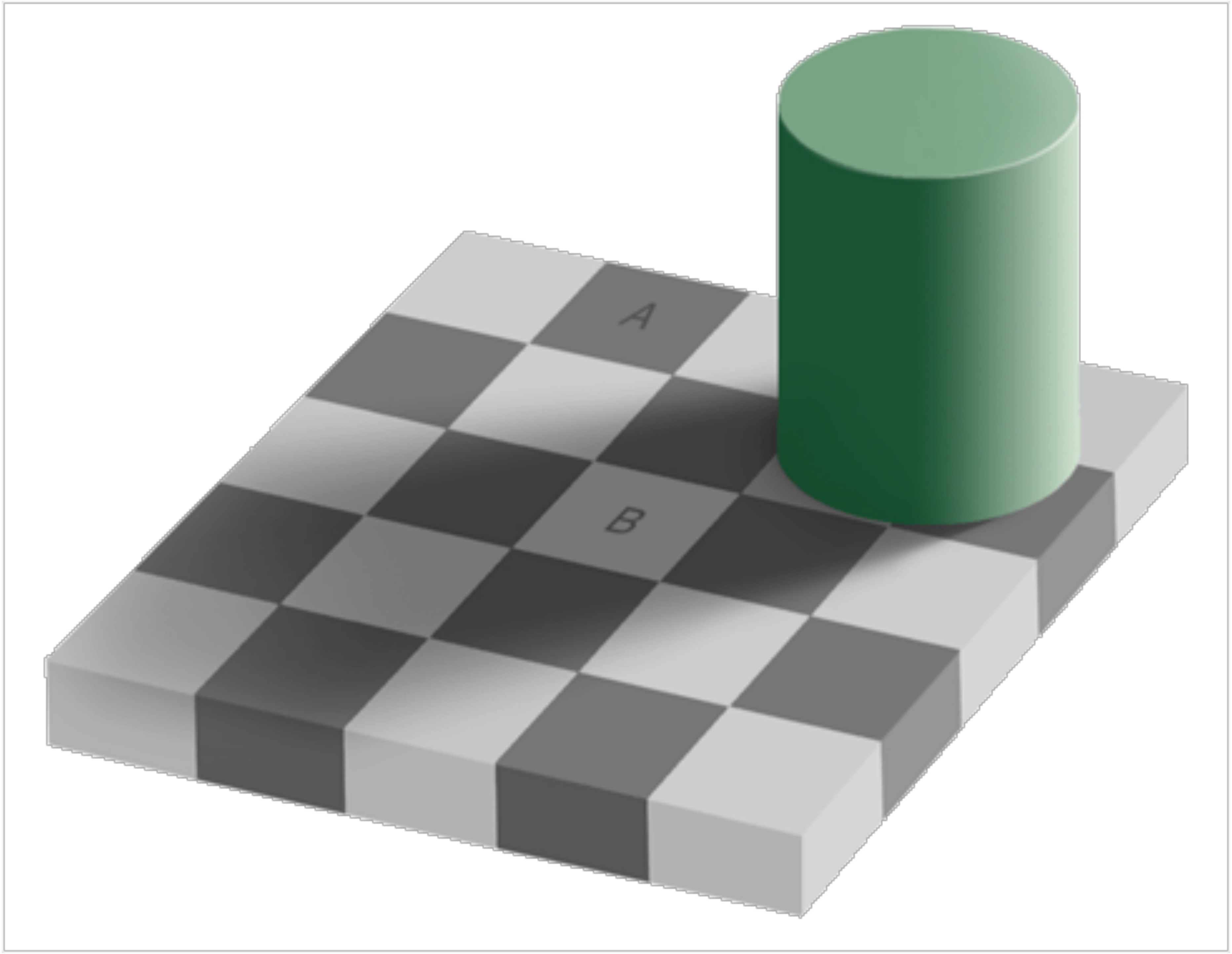


Similar Images



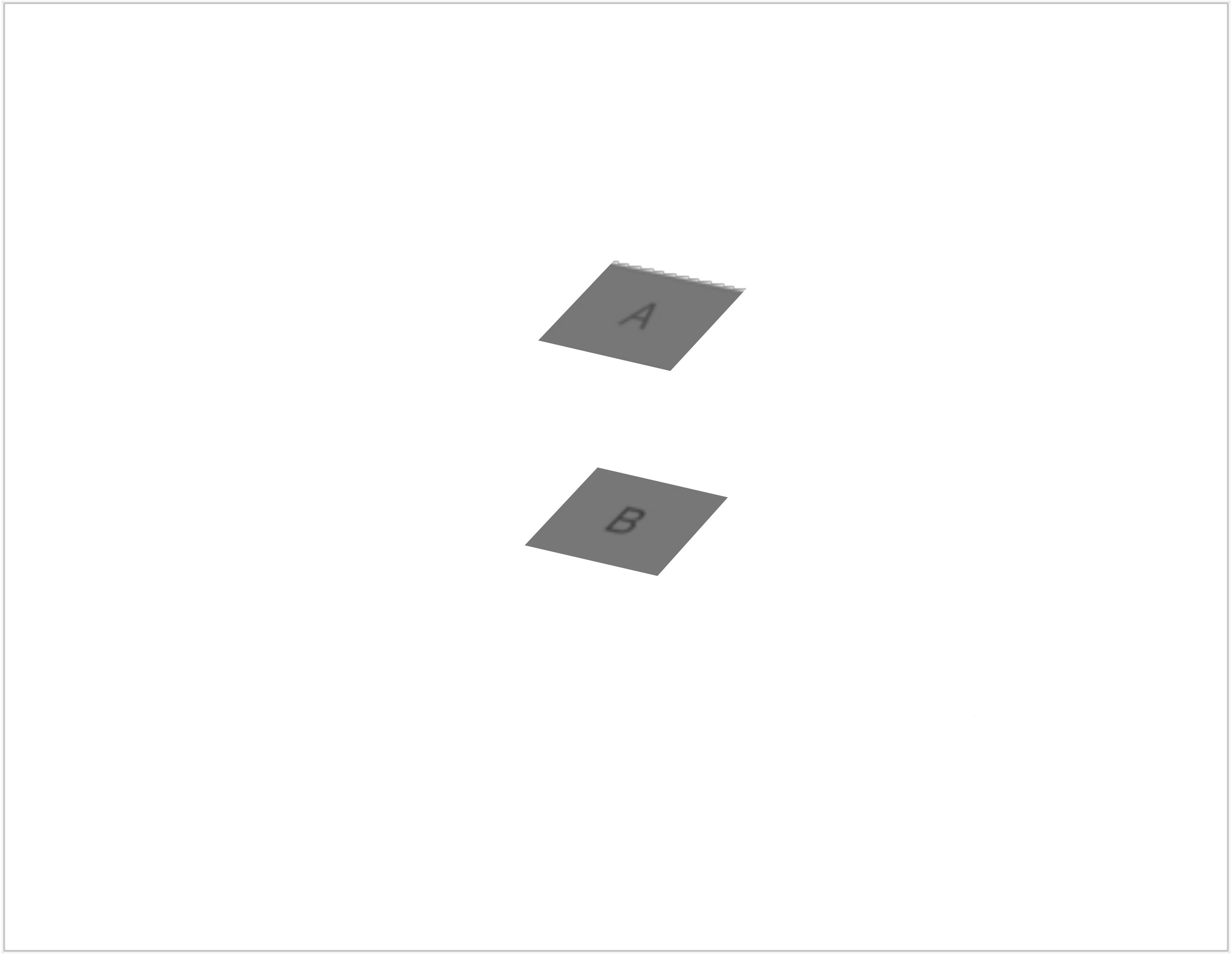
From Kersten et al., 2004

Vision is Ill-Posed (Reflectance & Illumination)



(Adelson, 1995)

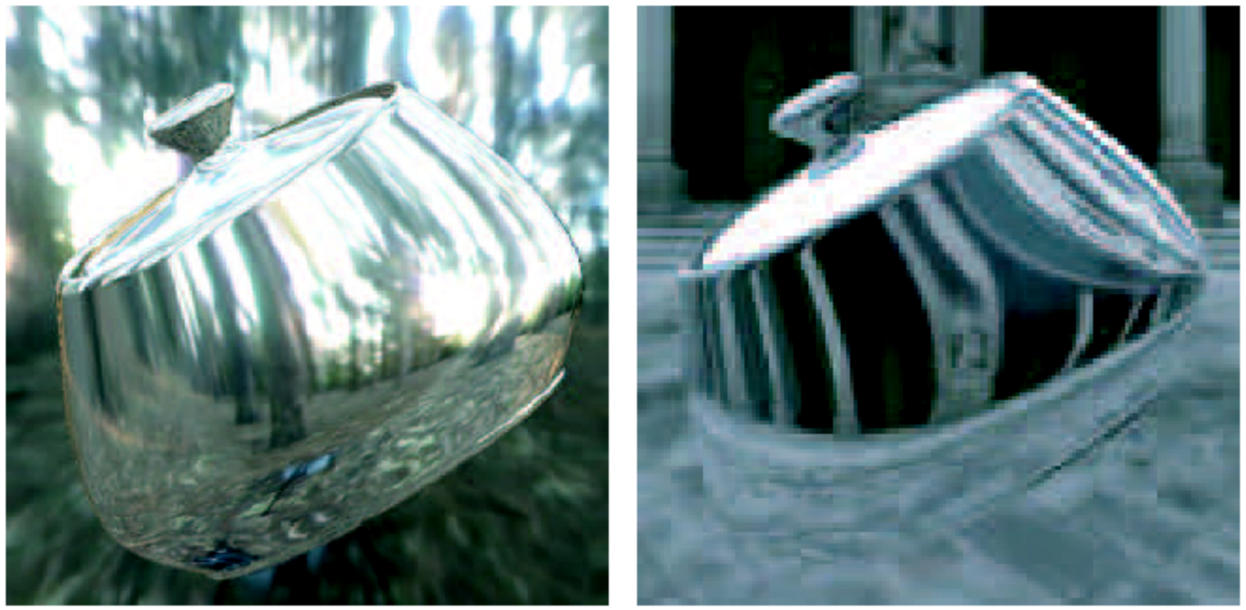
Vision is Ill-Posed (Reflectance & Illumination)



(Adelson, 1995)

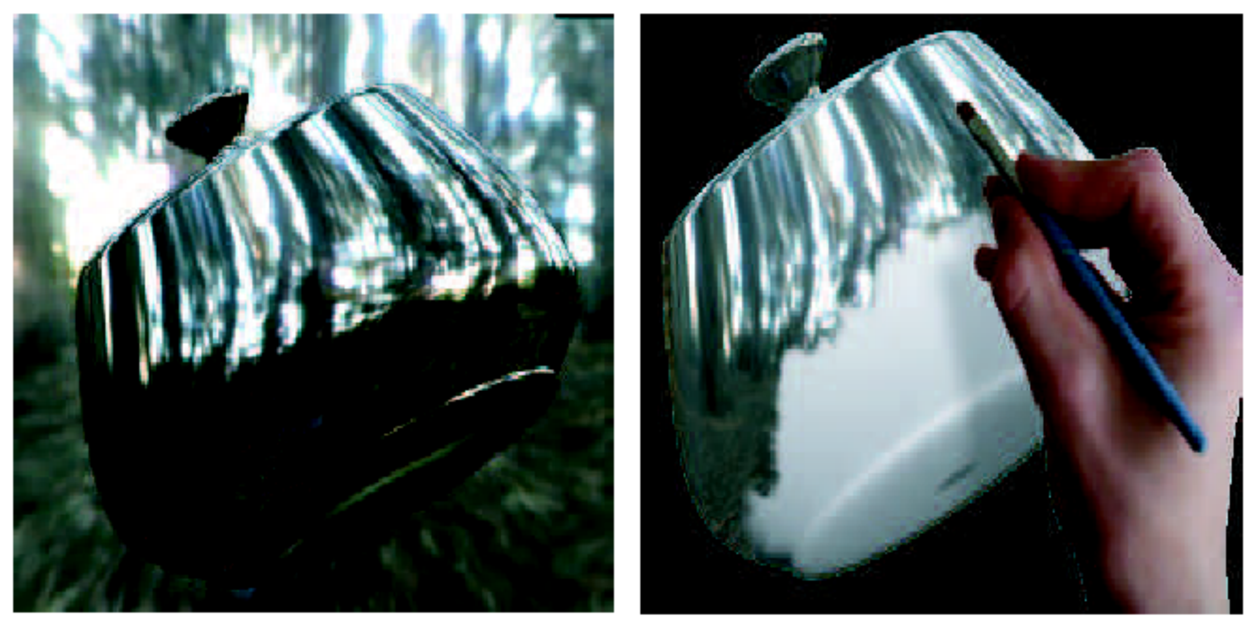
Vision is ill-posed (Reflectance & Illumination)

YORK U UNIVERSITÉ UNIVERSITY



1:N Mapping

N:1 Mapping



From Kersten et al., 2004

Art and Illusion

- Artists use the ill-posedness of images to great effect, usually by playing off two solutions that are both highly improbable.

Julian Beever

Julian Beever

Julian Beever

Liu Bolin

Liu Bolin

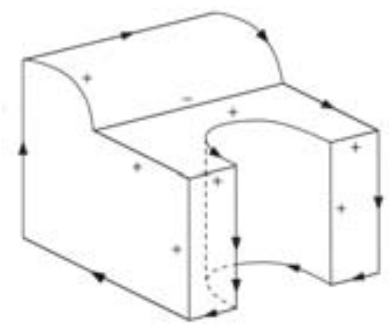
Liu Bolin

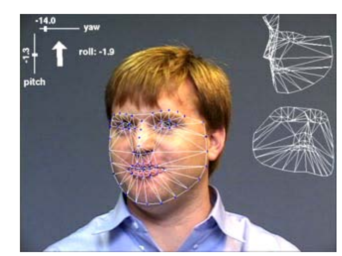
Summary

- Geometry
  - Images are 2D, but vision is 3D
- Likelihood Principle
  - Vision is ill-posed: multiple possible explanations for every observation.
  - The brain tries to determine the most probable.
- Ecological statistics
  - To make these judgements, the brain must use the ecological statistics of natural stimuli in the visual environment.
- Illusions
  - We perceive what is *visually* more probable.

History

- Evolution of ideas
- Don't reinvent the wheel
- Give credit

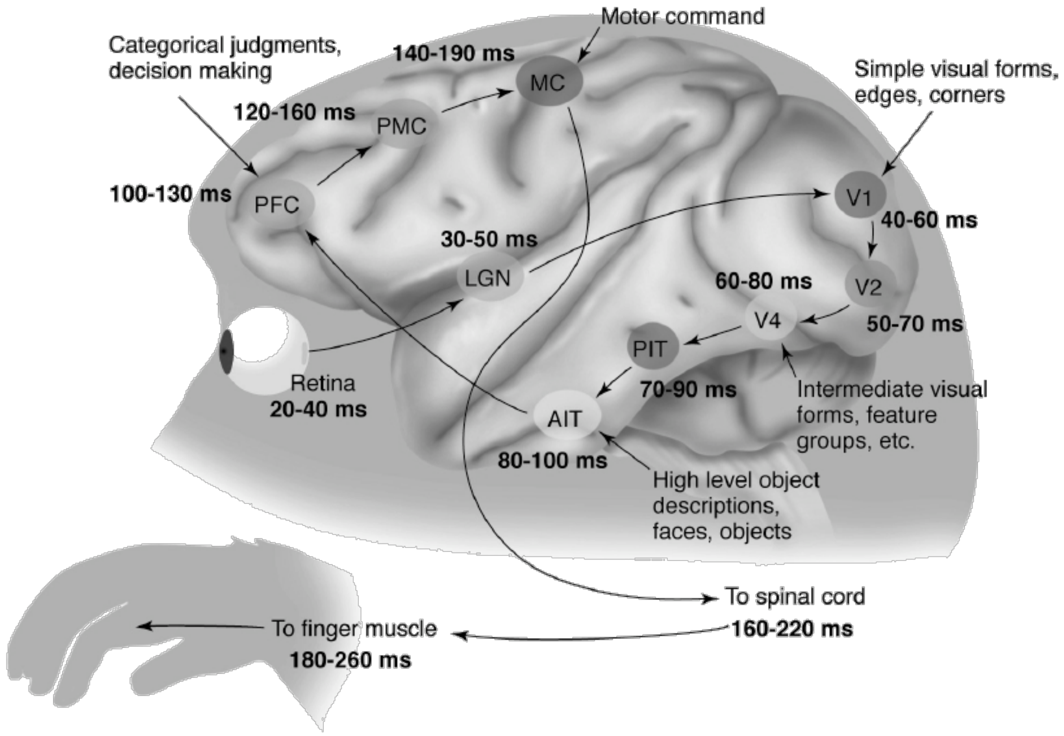




Digital image processing
Blocks world, line labeling
Generalized cylinders
Pictorial structures
Stereo correspondence
Intrinsic images
Optical flow
Structure from motion
Image pyramids
Scale-space processing
Shape from shading, texture, and focus
Physically-based modeling
Regularization
Markov Random 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

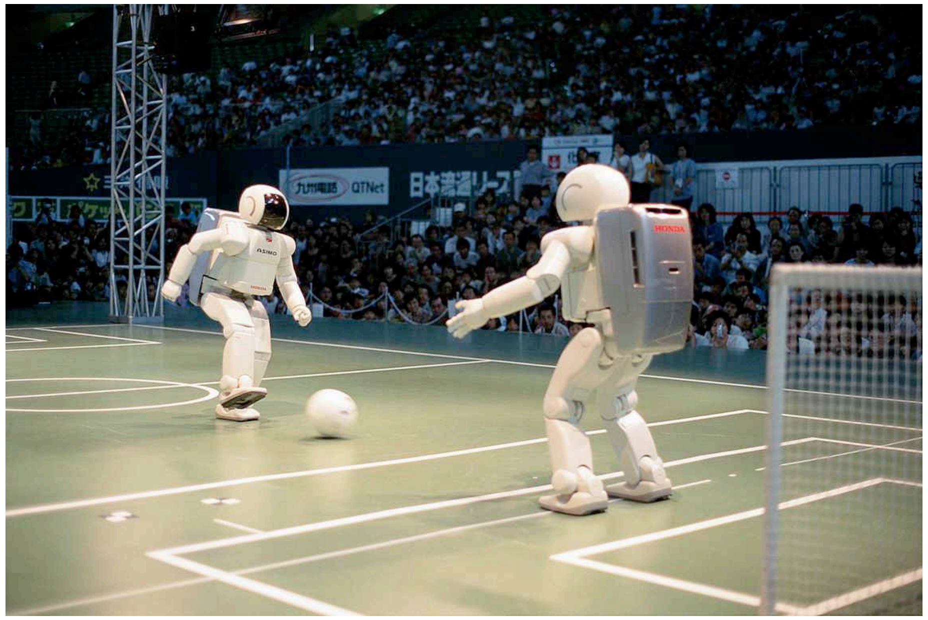
1970
1980
1990
2000

Interdisciplinary Studies: Human and Computer Vision



Reverse Engineering

- Features
- Representations
- Cue Integration



Theory

- Algorithms
- Learning
- Testing

DNNs as Models of Primate Ventral Stream

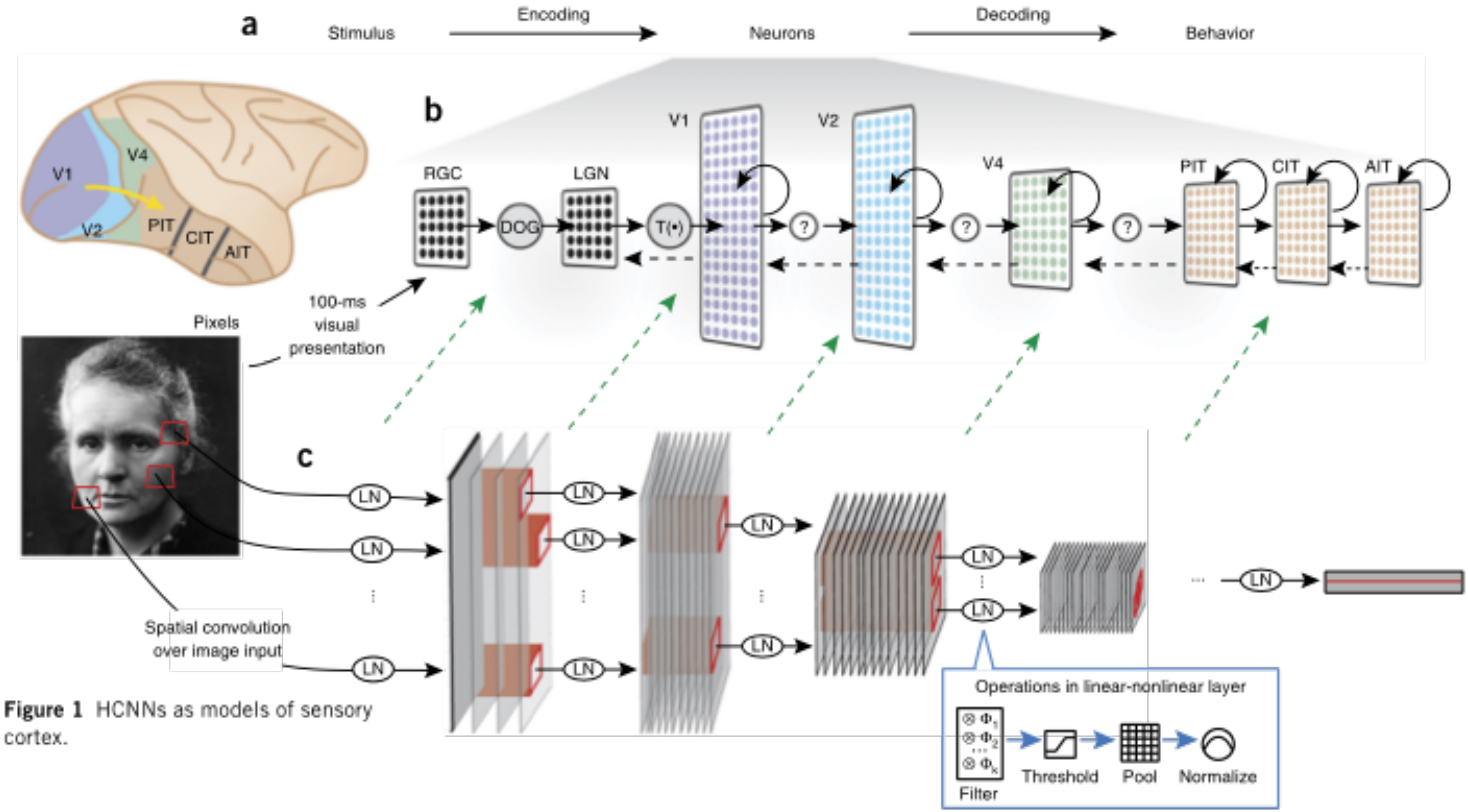
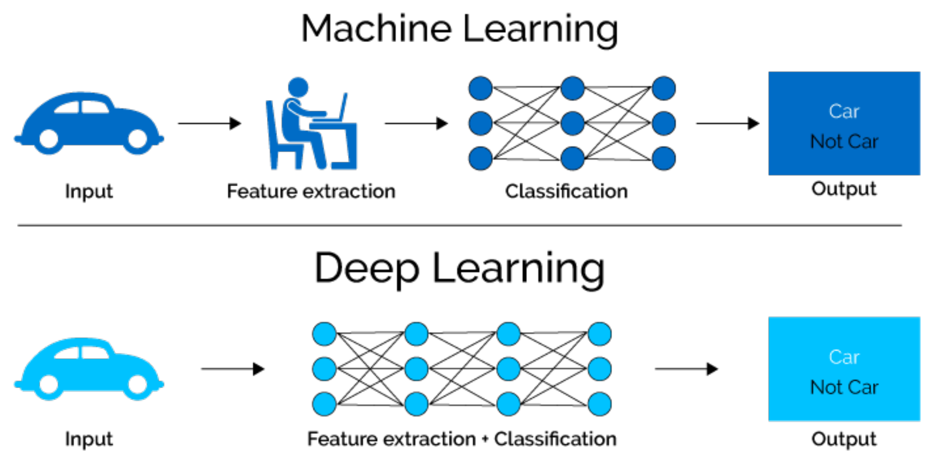


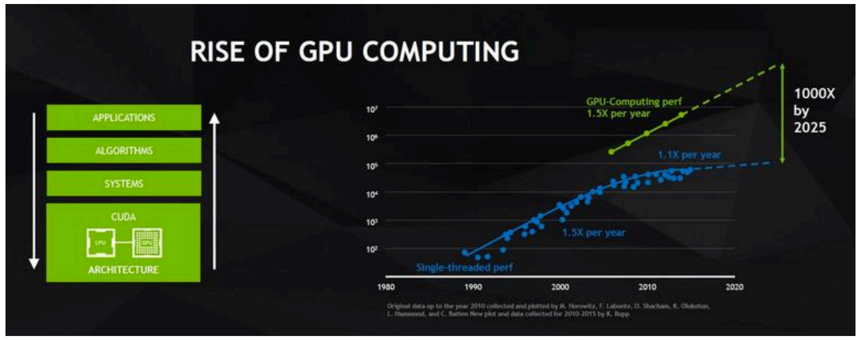
Figure 1 HCNNs as models of sensory cortex.

From Yamins & DiCarlo, 2016

Computer Vision Today

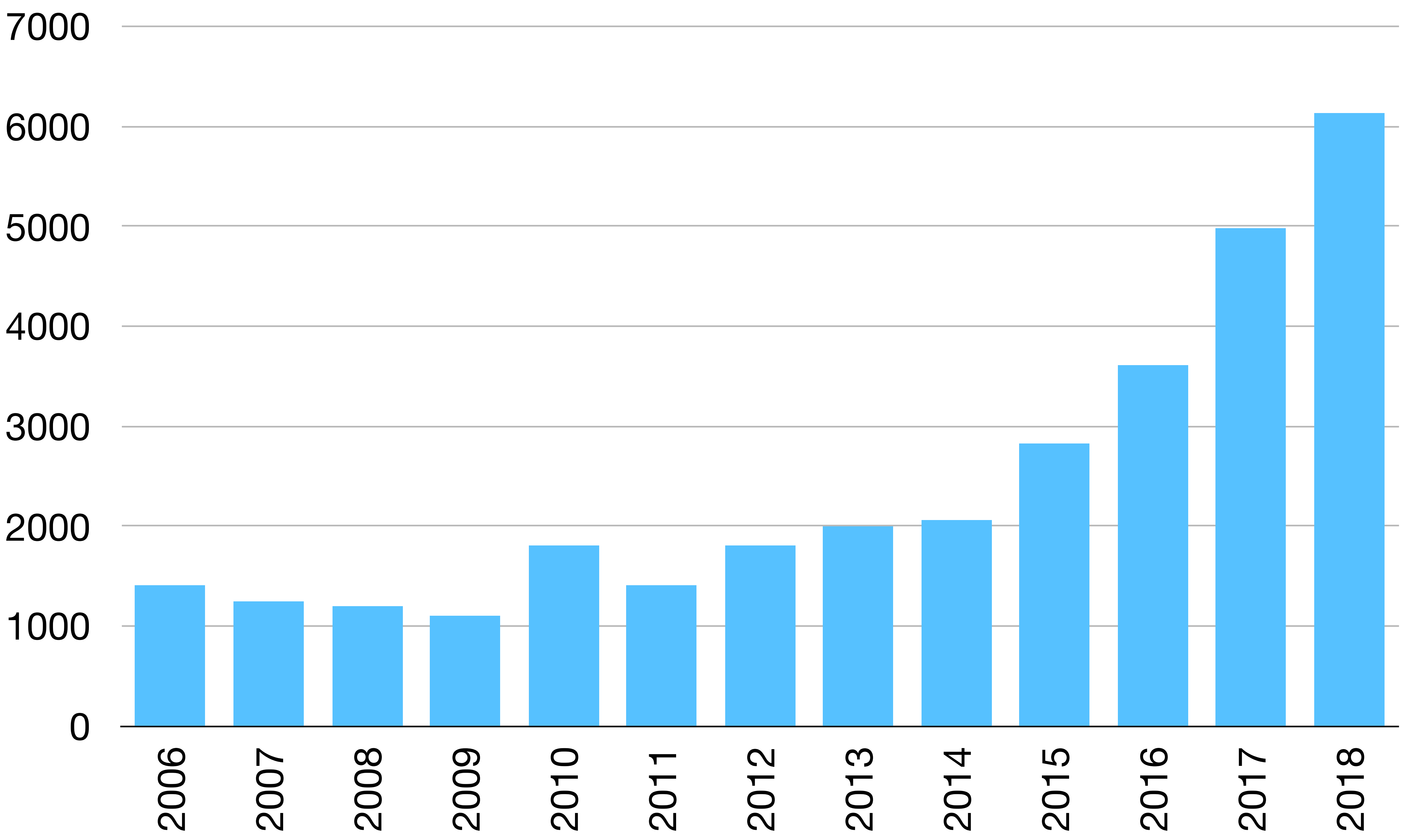
- Growth in Prestige
- Deep Learning
- Big Data
- Industry
- Entrepreneurship





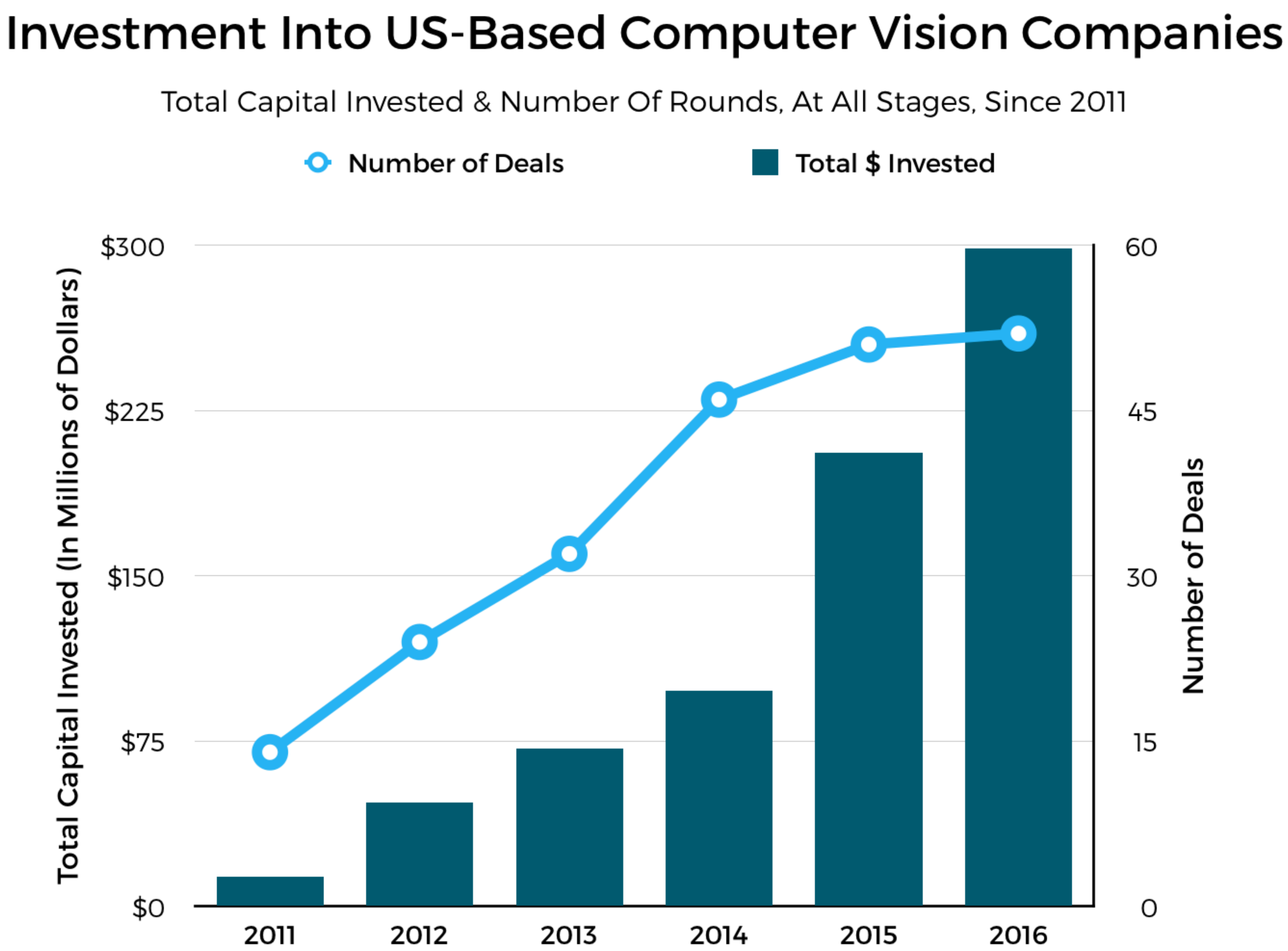
Shelf-Scanning Robot (Walmart)

CVPR Attendance

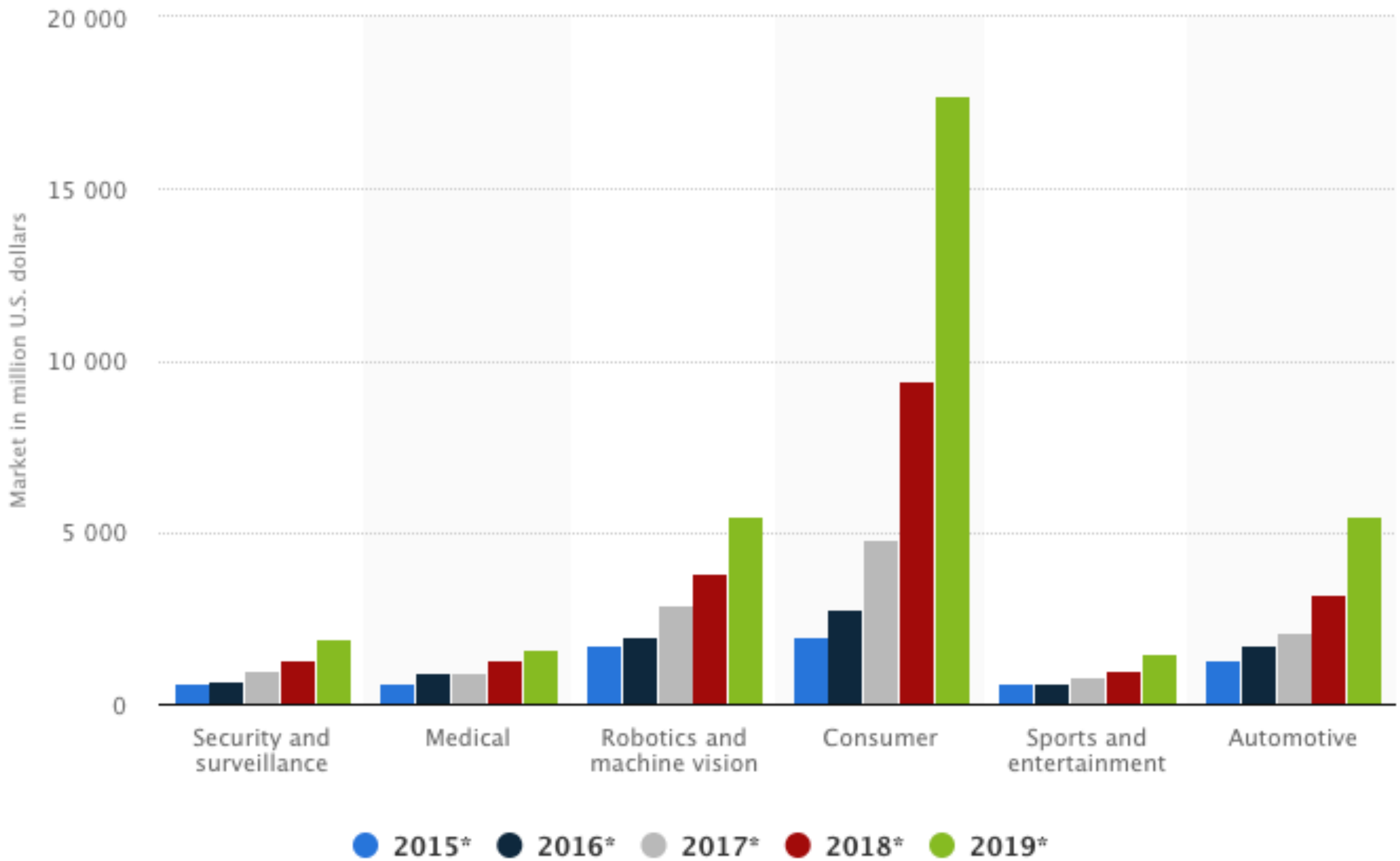


CVPR 2018

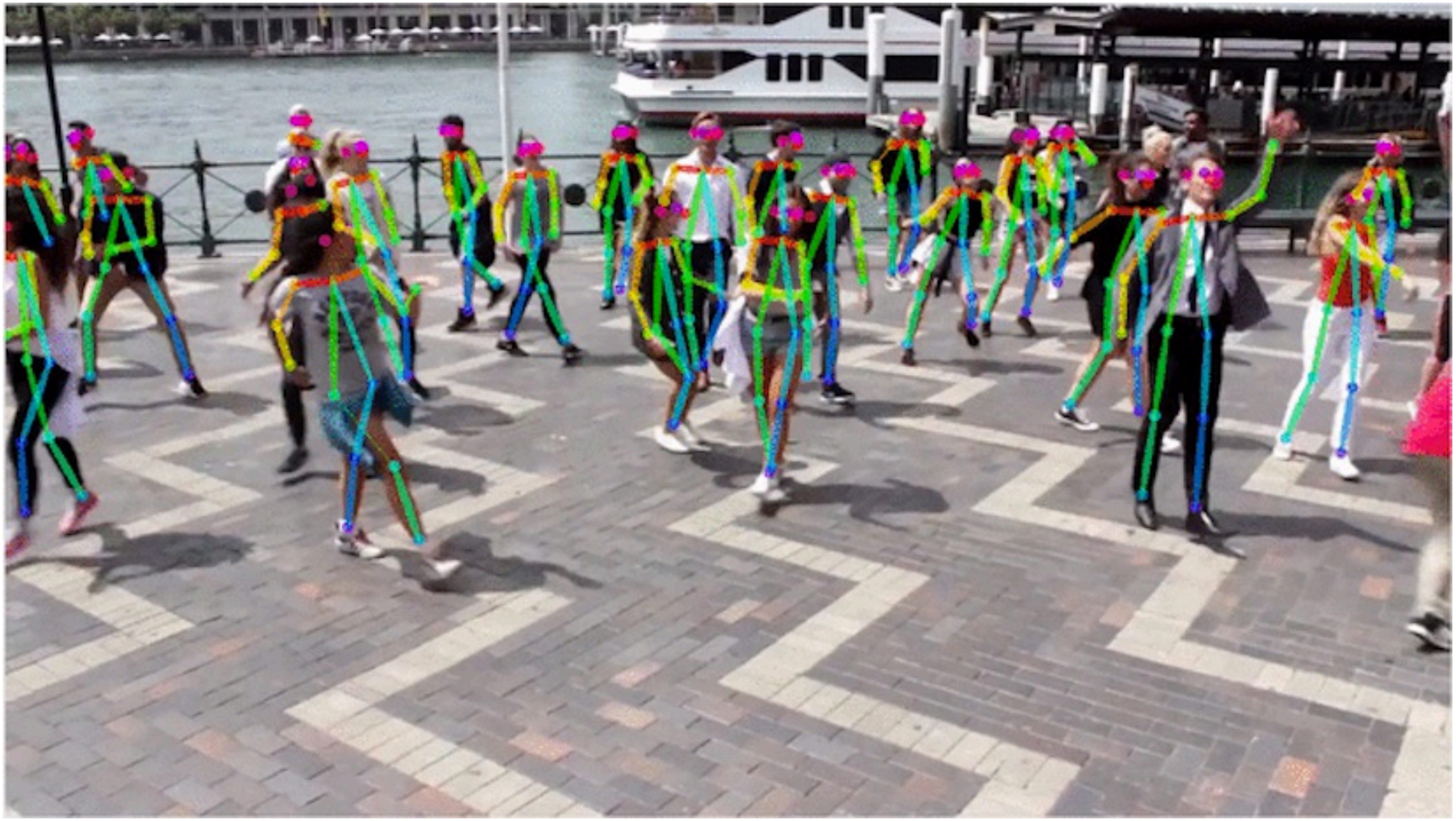
- 3,309 submissions
- 149 sponsors and exhibitors
- $2M+ in sponsorships/exhibitions
- 6,128 attendees


Investment Into US-Based Computer Vision Companies
Total Capital Invested & Number Of Rounds, At All Stages, Since 2011
Number of Deals
Total $ Invested
Total Capital Invested (In Millions of Dollars)
$300
$225
$150
$75
$0
60
45
30
15
0
Number of Deals
2011
2012
2013
2014
2015
2016

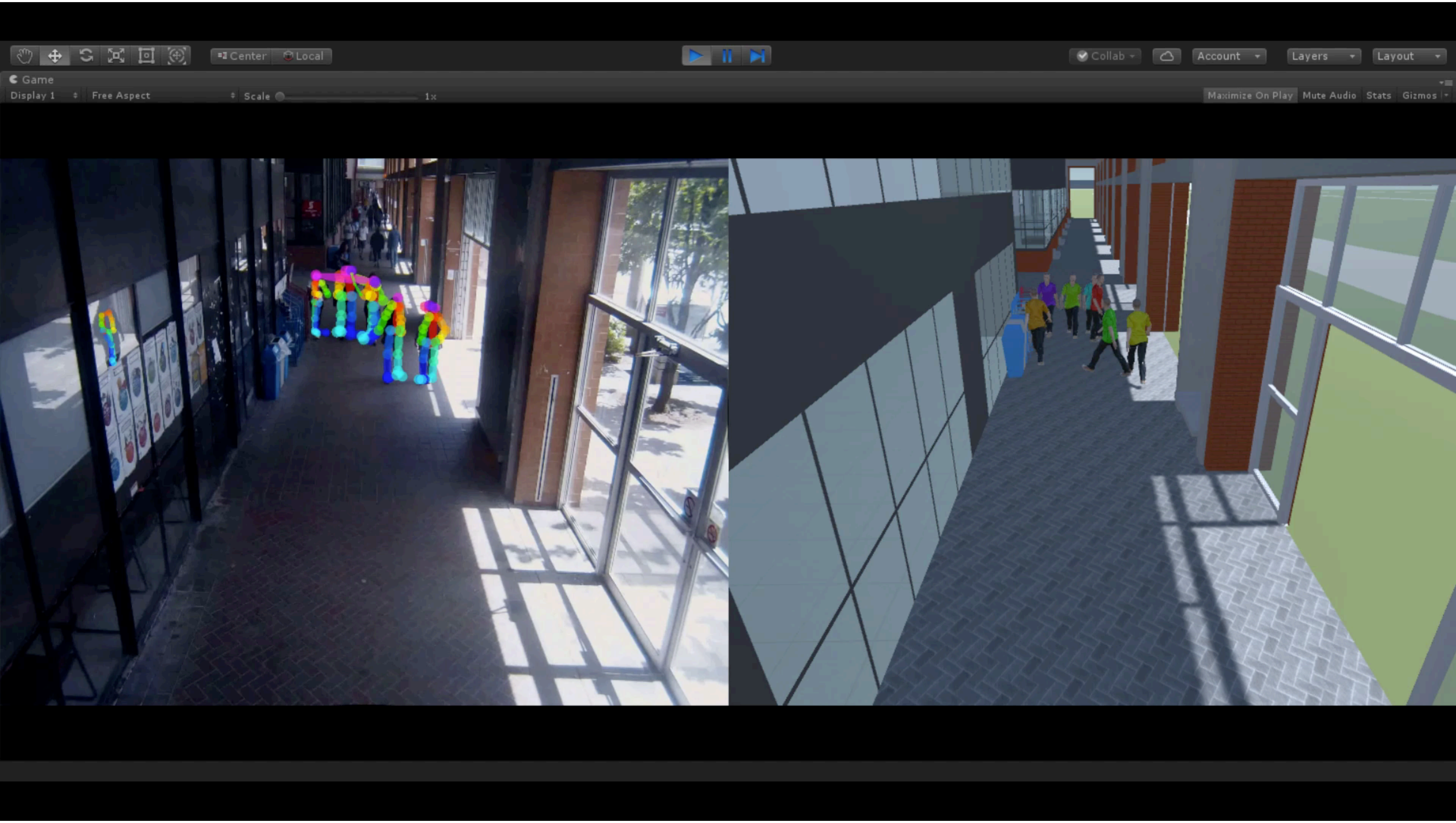
Industrial Computer Vision by Sector



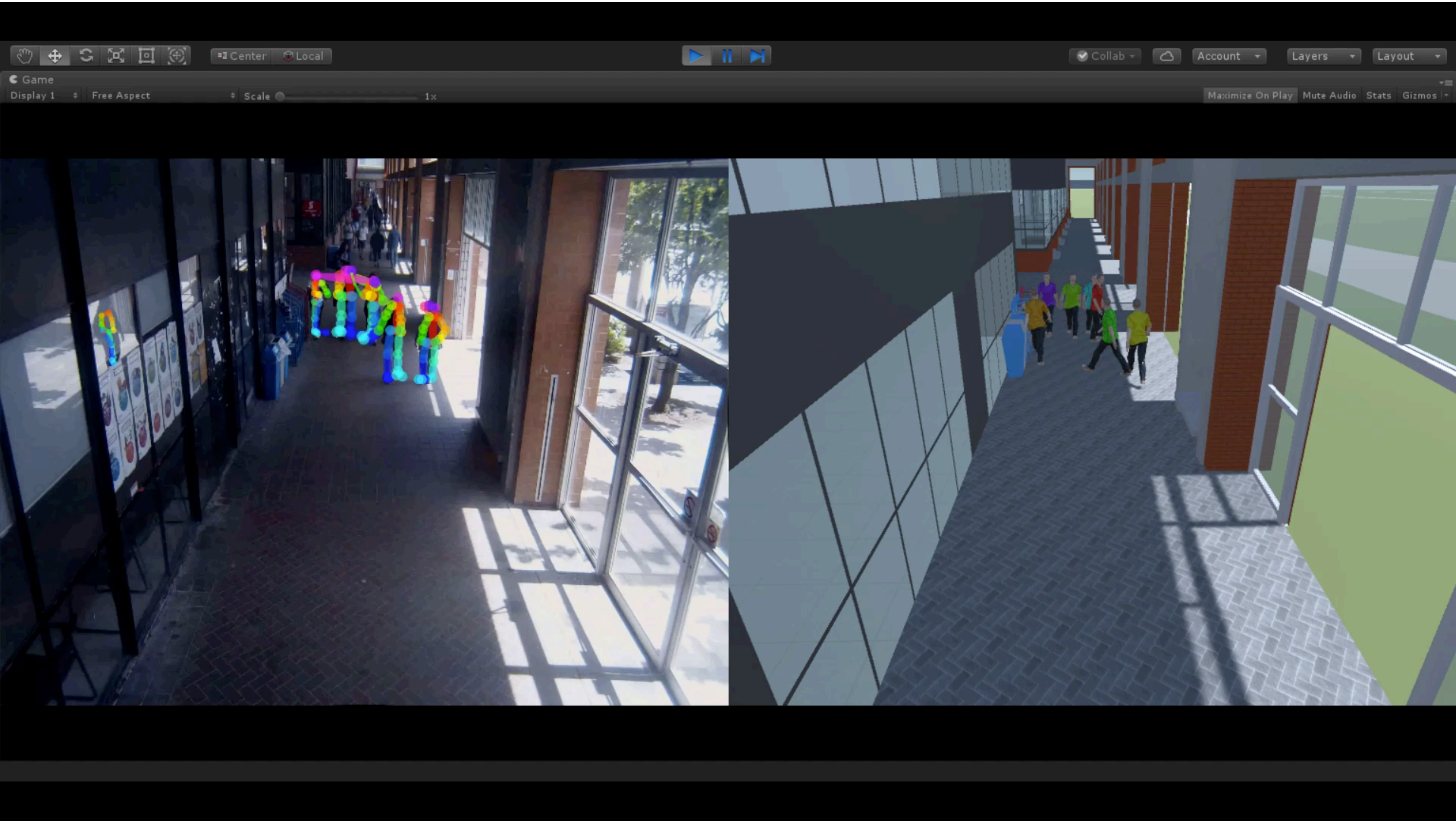
Impressive Technical Achievements



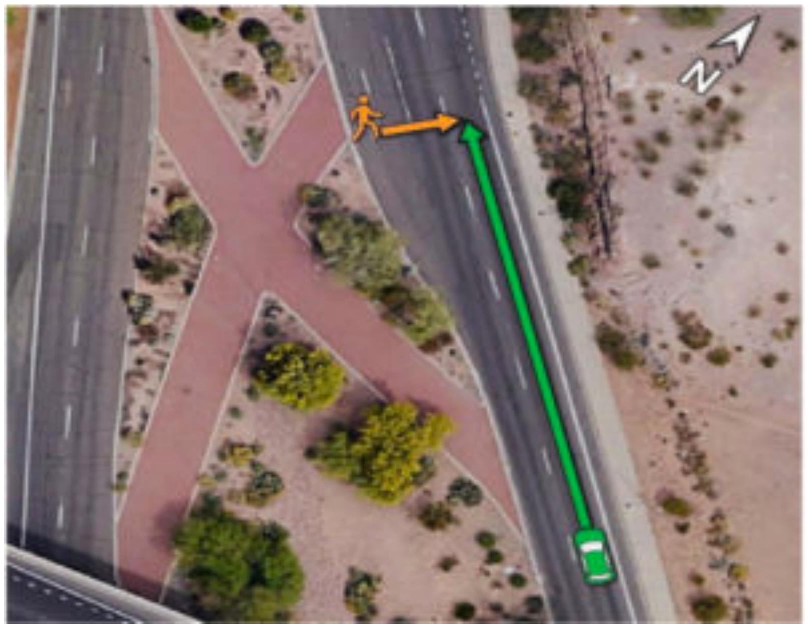
3D Virtualization



3D Virtualization

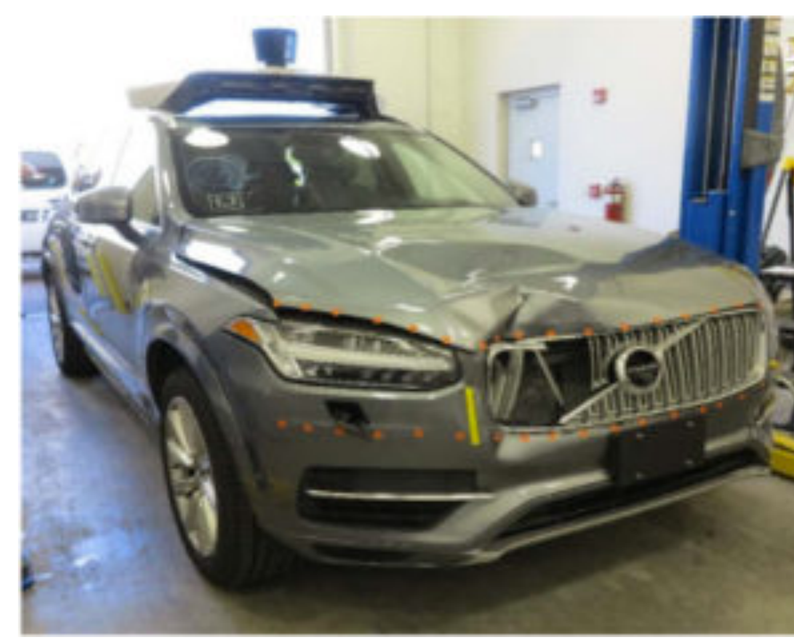


Computer Vision Matters



March 18th, 2018

…the pedestrian was detected about 6 seconds before impact but the system had trouble identifying it. It was first classified as an unknown object, then a vehicle, then a bicycle – but even then it couldn't work out the object's direction of travel. At 1.3 seconds before impact, the system realized that it needed to engage an emergency braking maneuver but this maneuver had been earlier disabled to prevent erratic vehicle behaviour on the roads.



Topics

- Image formation
- Image processing
- Feature detection & matching
- Segmentation
- Dense motion estimation
- Feature-based alignment
- 3D - Motion
- 3D - Stereo
- 3D - Single View
- Recognition

Image Formation

- Geometric primitives and transformations
- Photometric image formation
- The digital camera

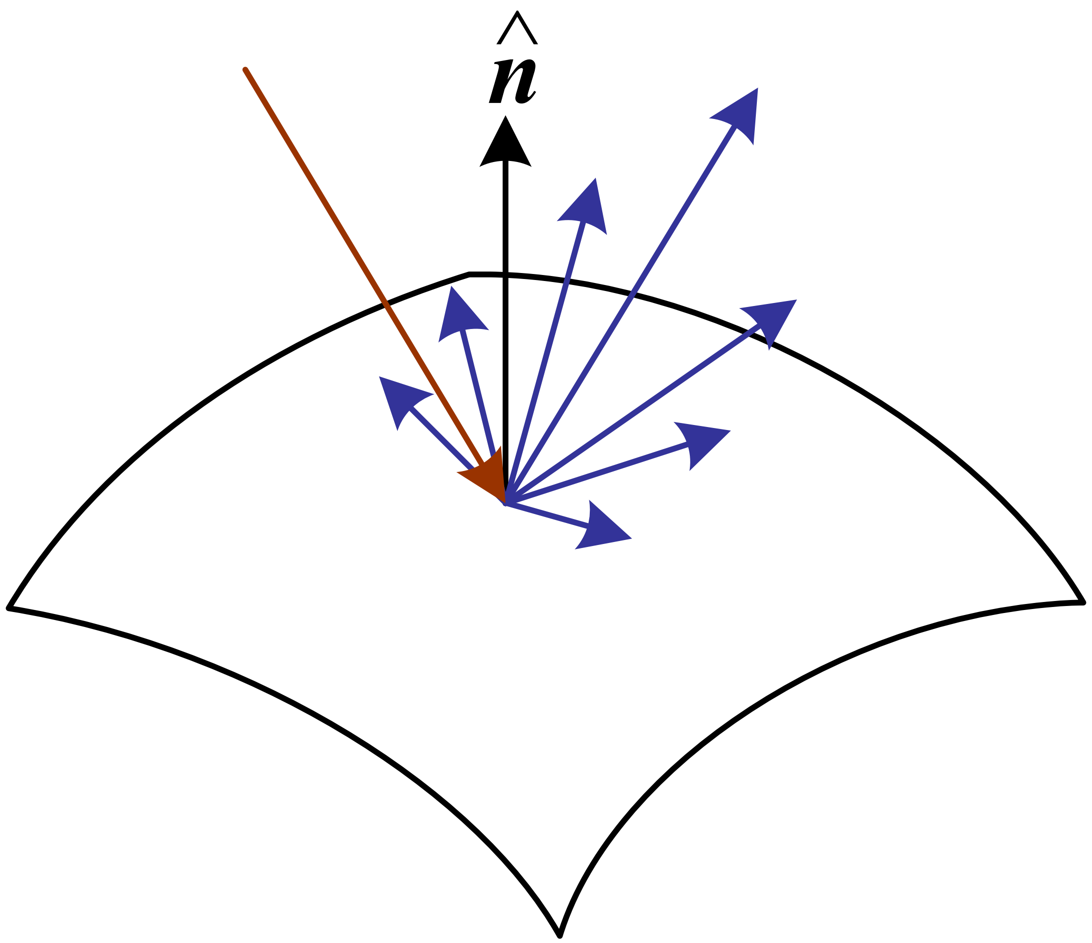
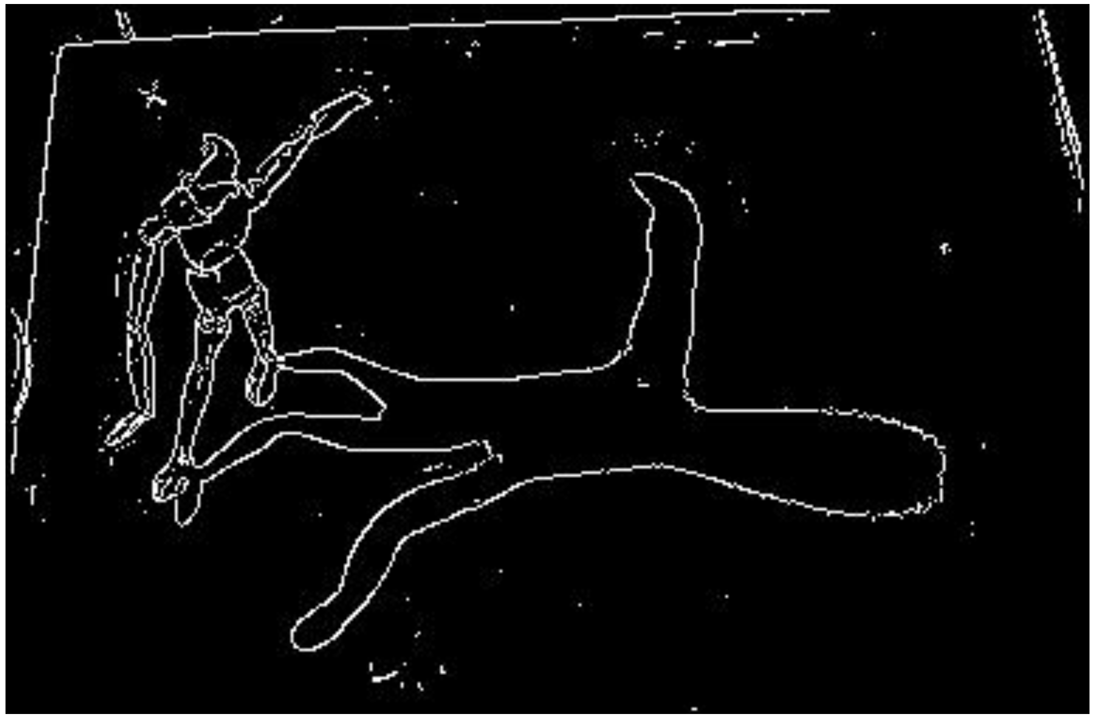


Image Processing

- Point operators
- Linear filtering
- More neighbourhood operators
- Fourier transforms
- Pyramids and wavelets
- Geometric transformations
- Global optimization

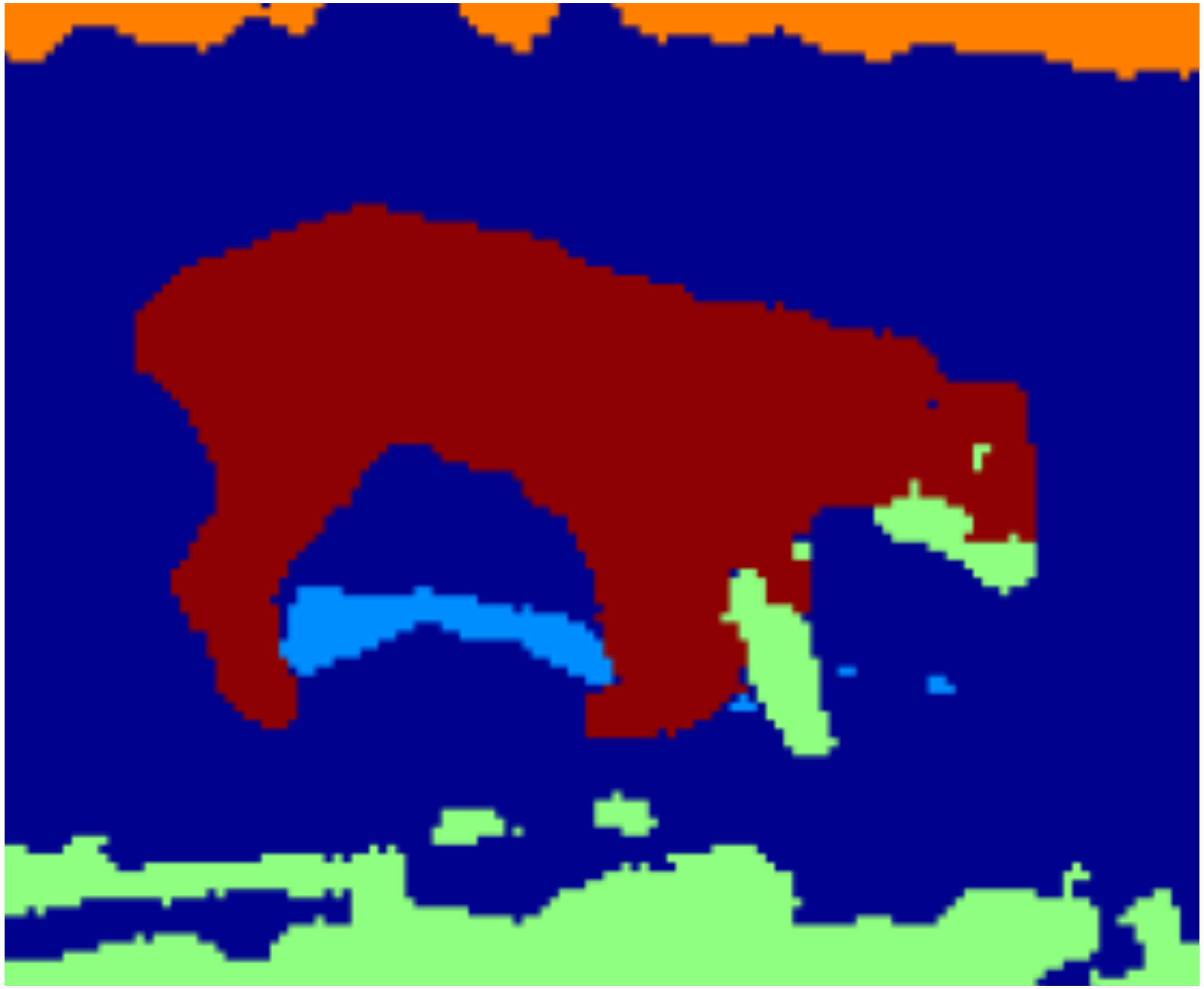
Feature Detection and Matching

- Points and patches
- Edges
- Lines



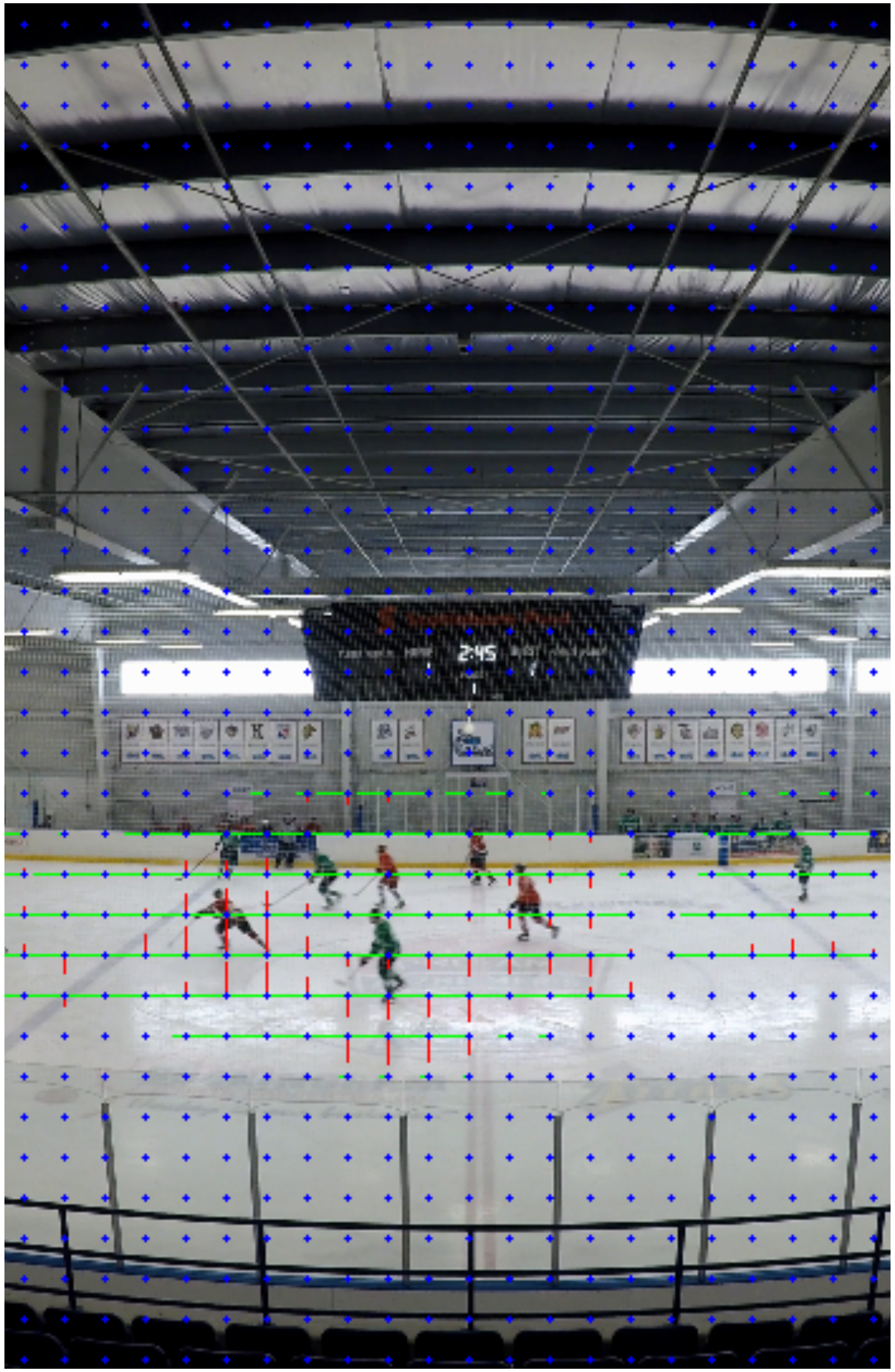
Segmentation

- Active contours
- Split and merge
- Mean shift and mode finding



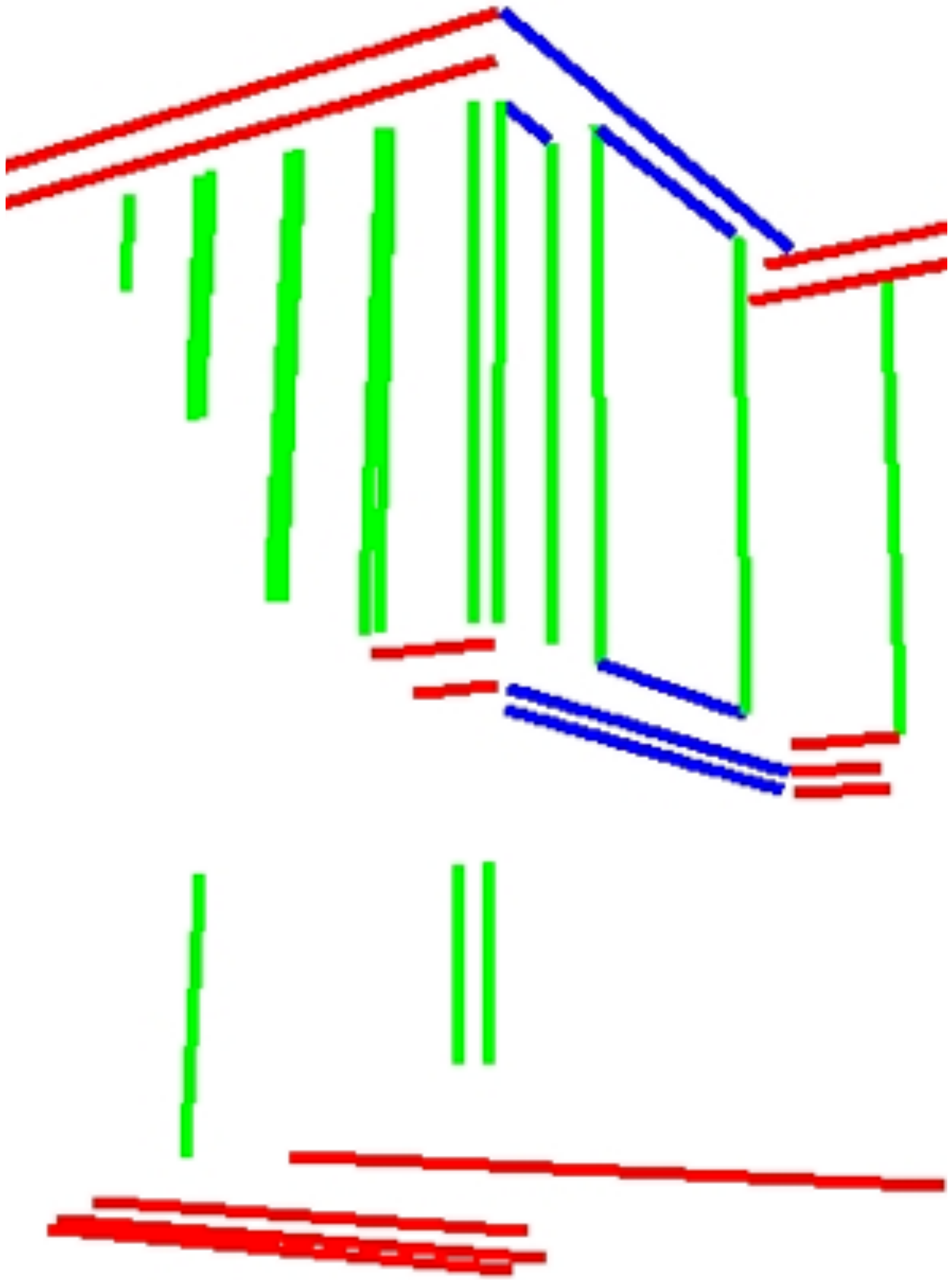
Dense Motion Estimation

- Translational alignment
- Optical flow



Feature-Based Alignment

- 2D and 3D feature-based alignment
- Pose estimation
- Geometric intrinsic calibration

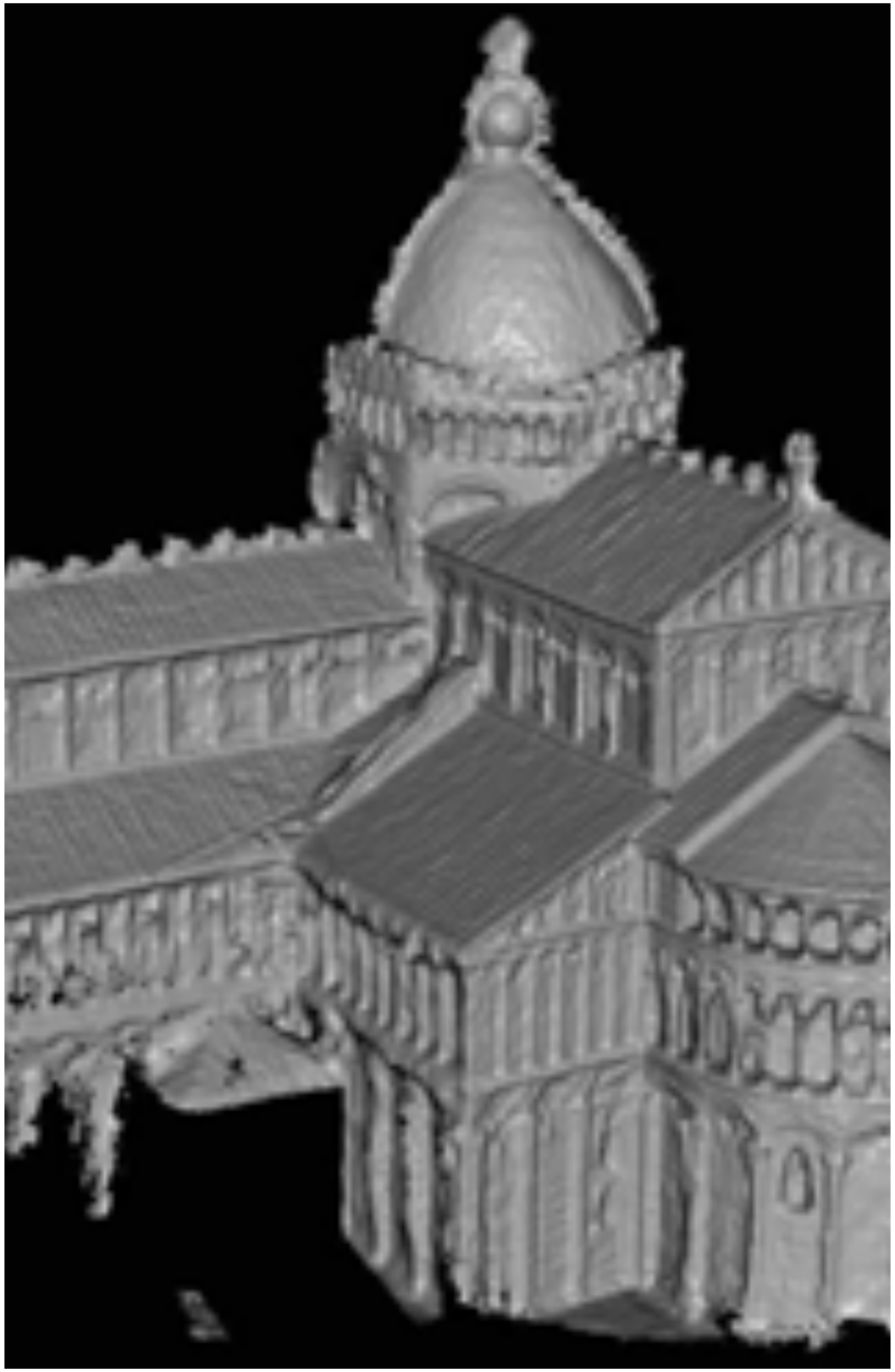


3D - Motion

- Triangulation
- Two-frame structure from motion

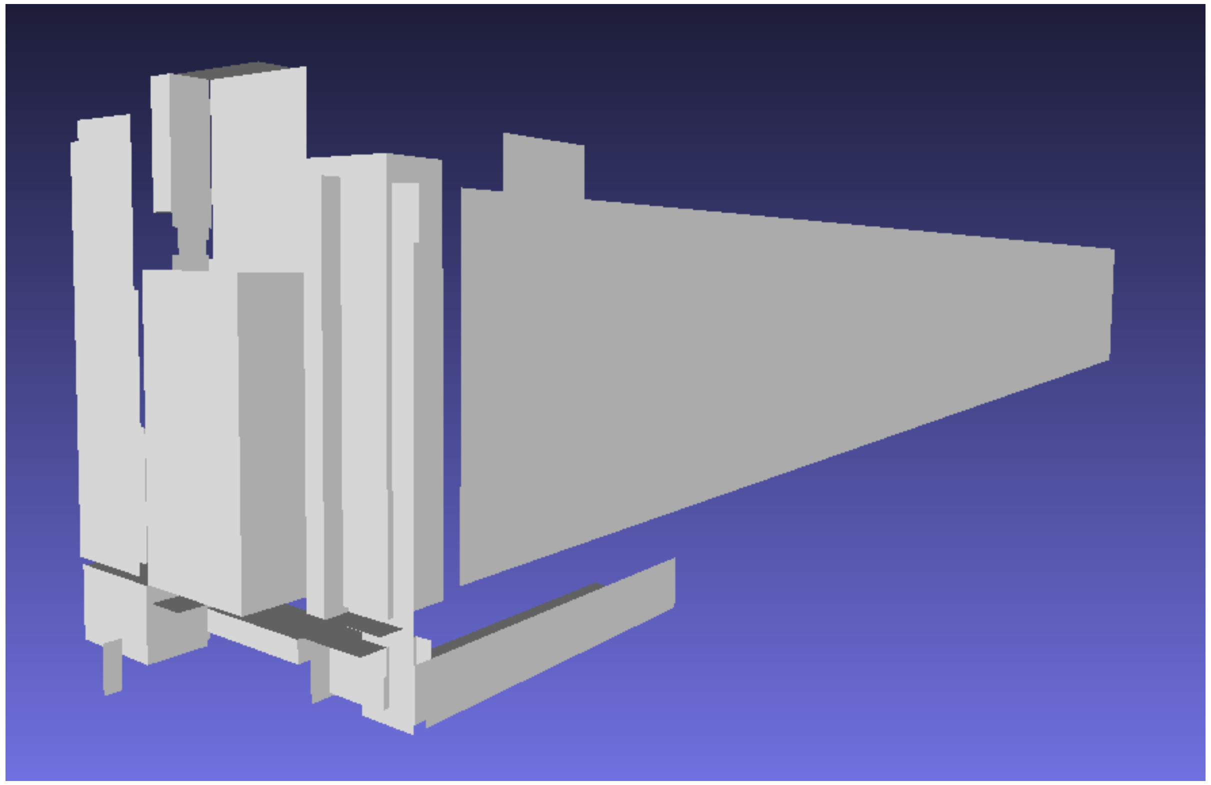
3D - Stereo

- Epipolar geometry
- Sparse correspondence
- Dense correspondence
- Local methods
- Global optimization



3D - Single View

- Shape from X
- Surface representations
- Point-based representations
- Volumetric representations



Recognition

- Object detection
- Face recognition
- Instance recognition
- Category recognition