Introduction to Computer Vision

CSE473/573 - Computer Vision and Image Processing

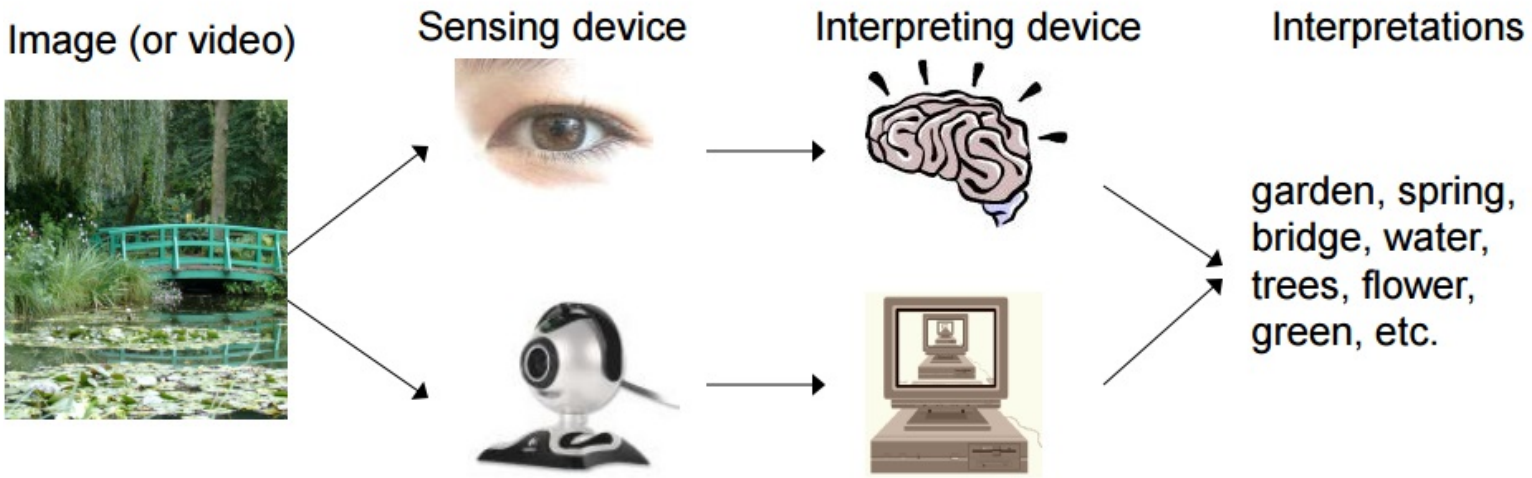
Junsong Yuan

jsyuan@buffalo.edu

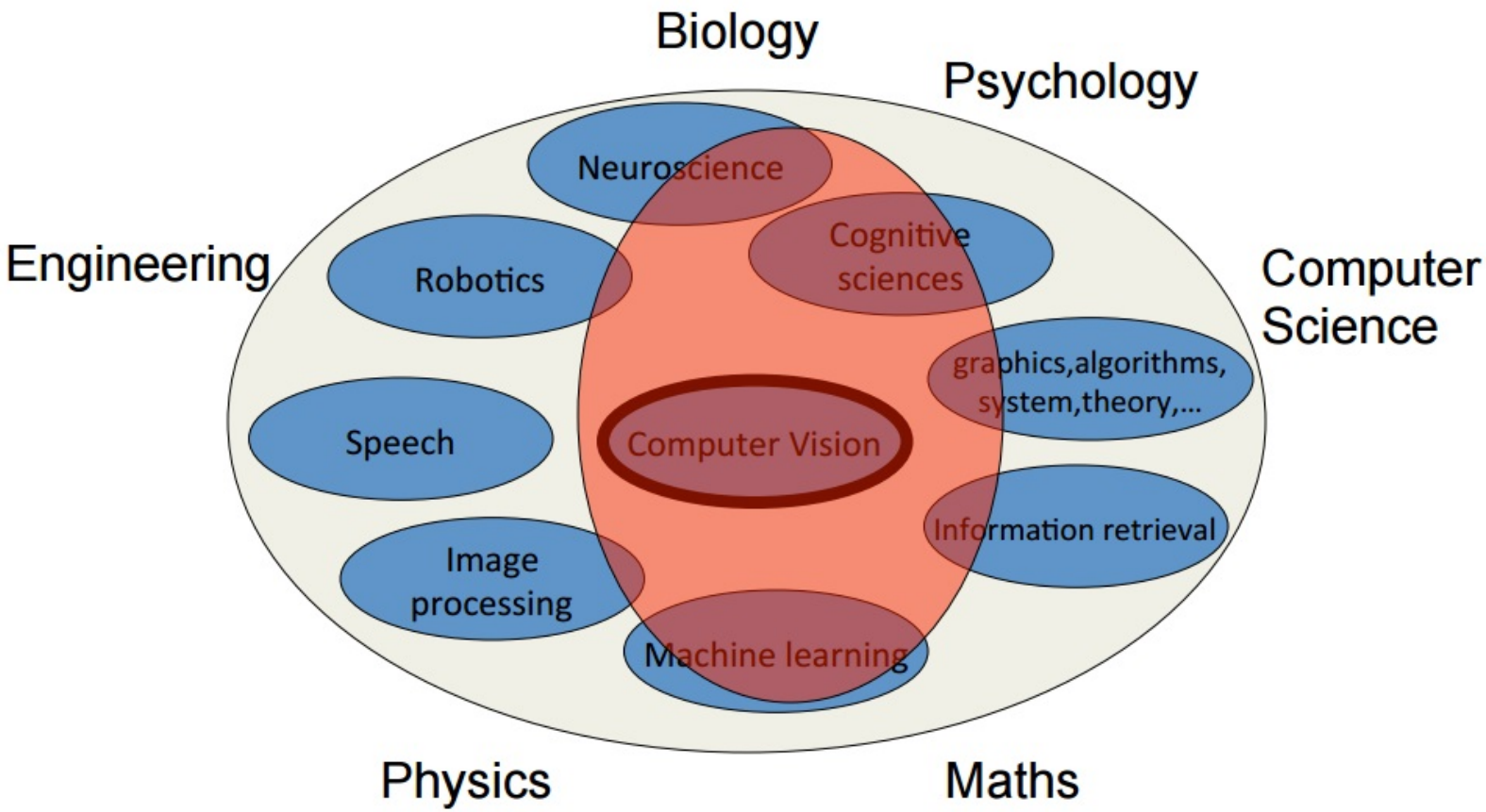
2018 Fall

Overview

Computer Vision - Making Computers See!



What is it related to?



Primary Goal of Computer Vision

- To bridge the gap between pixels and "meaning"

What we see

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

What a computer sees

Source: S. Narasimhan

Human Vision Capabilities

Describe what you see in this image in few words

Source: "80 million tiny images" by Torralba, et al.

Carnivore Predator V.S. Herbivore Preys

Why learn Computer Vision?

- Millions of images being captured all the time

- Loads of useful 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/

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

Automatic check processing

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

Source: S. Seitz

Challenges

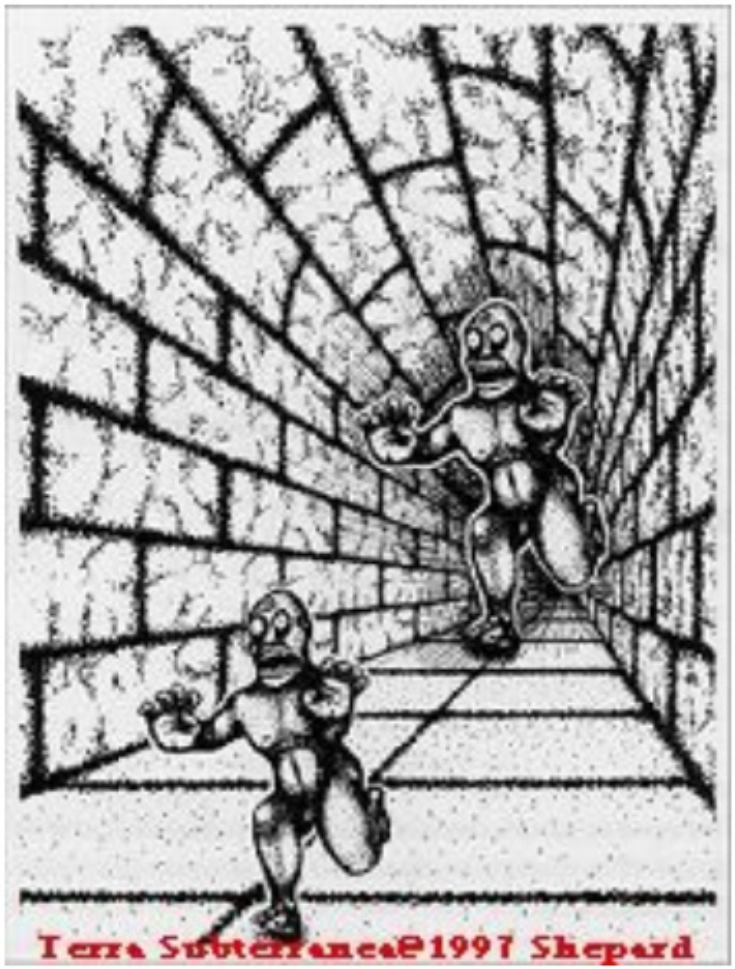
Intra-class variation

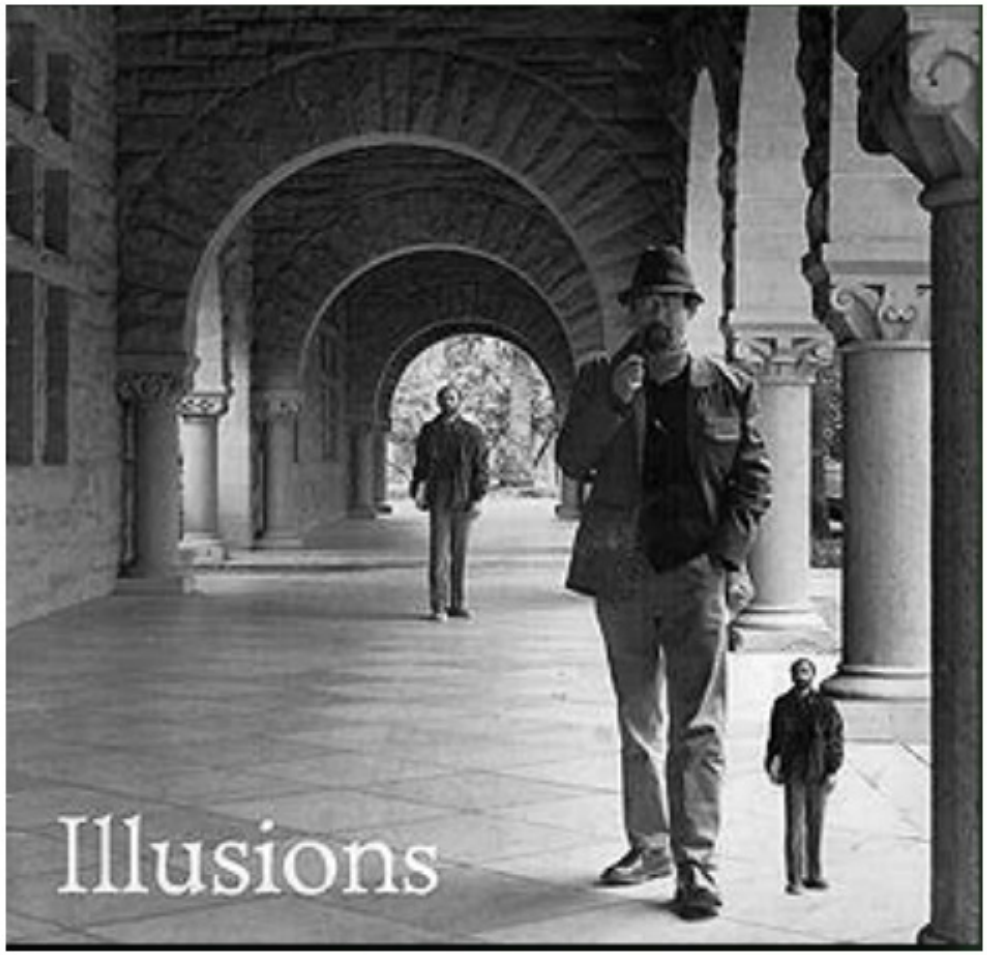
Motion (Source: S. Lazebnik)

Background clutter

Occlusion

Example Challenge - Depth perception from Single Image



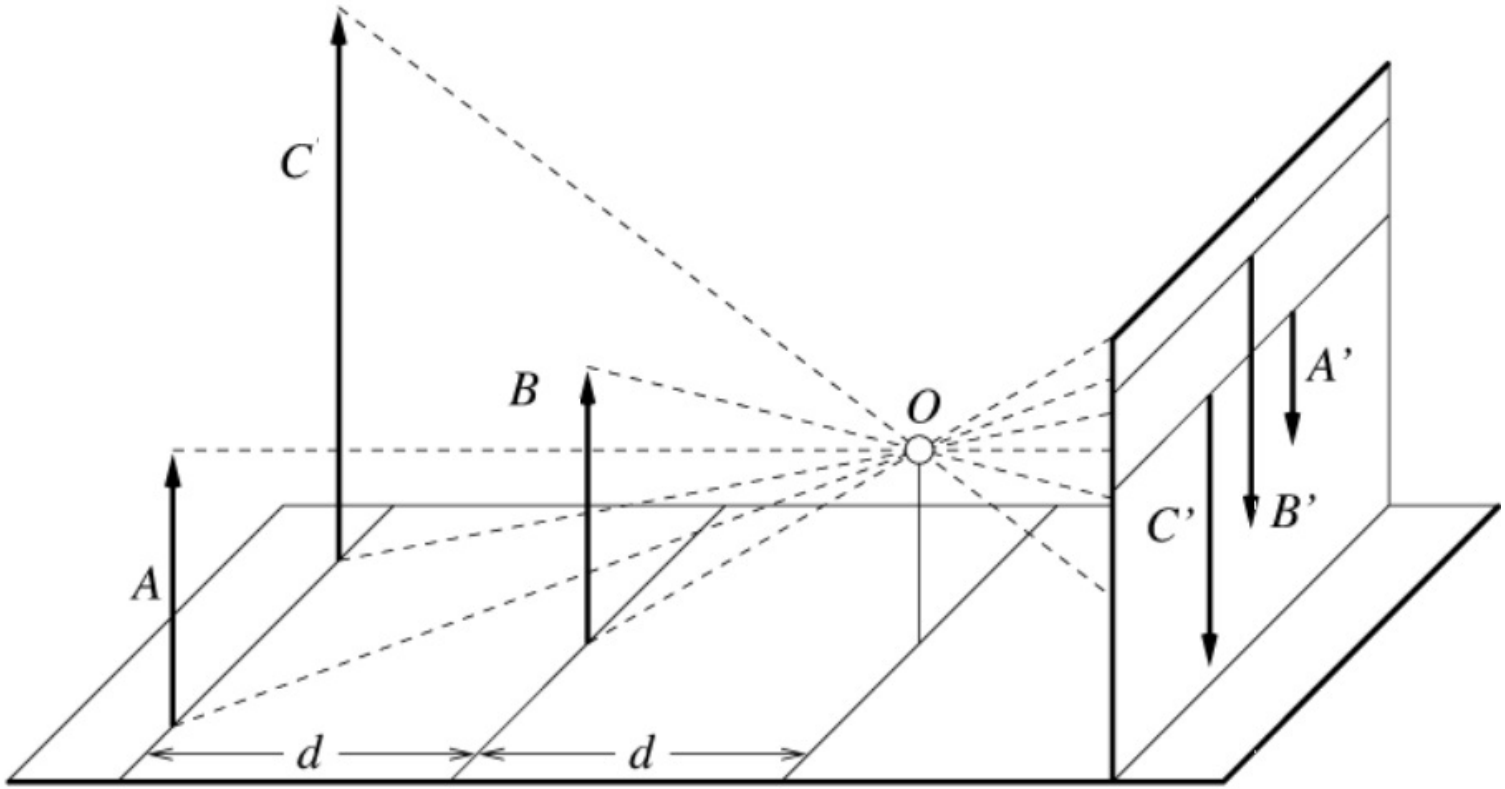


Visual Cues for Depth Perception

Source: S. Lazebnik

Understand the properties of Image formation

- Distant objects appear smaller



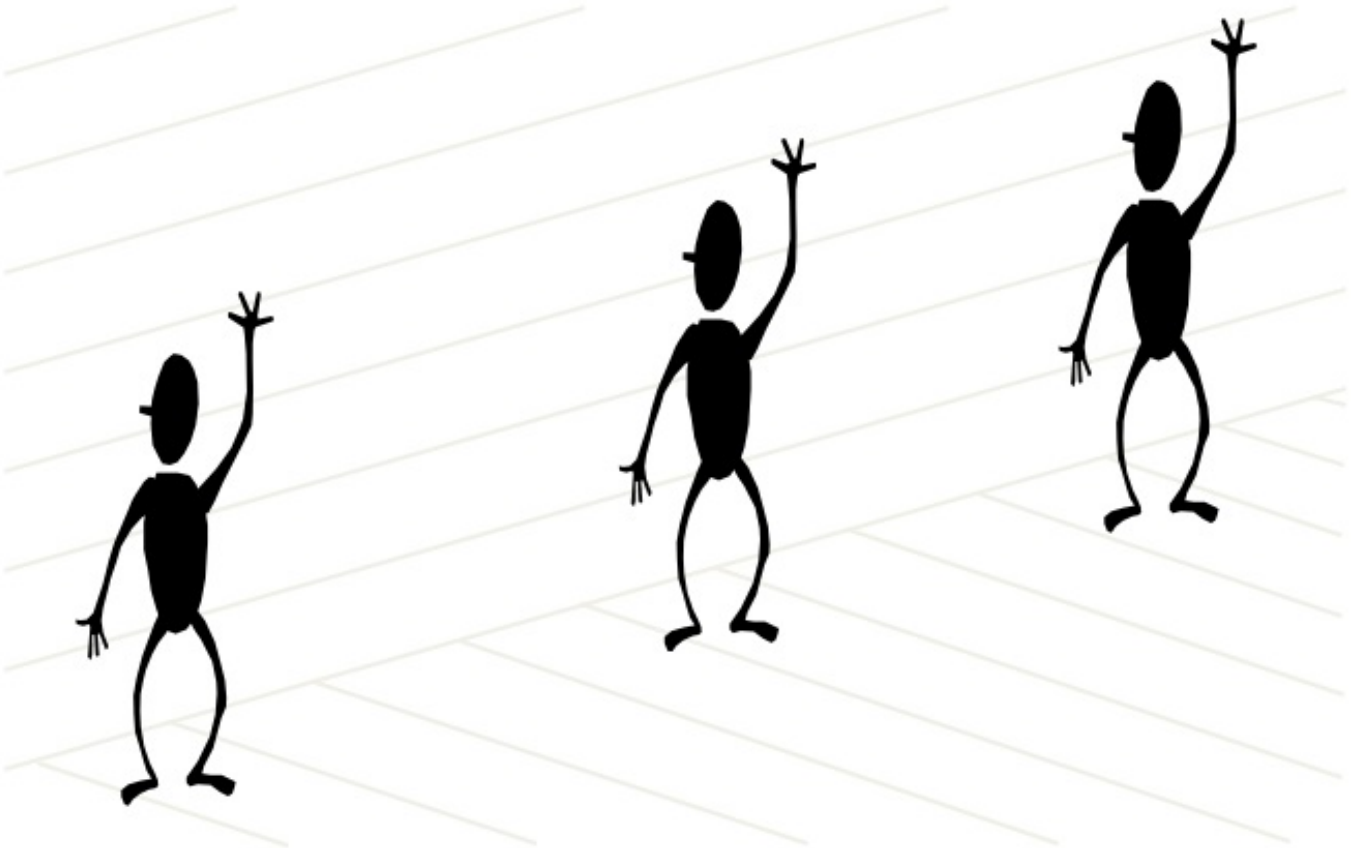
Source: D. Forsyth, S. Savarese slides.

Use cues from scene geometry

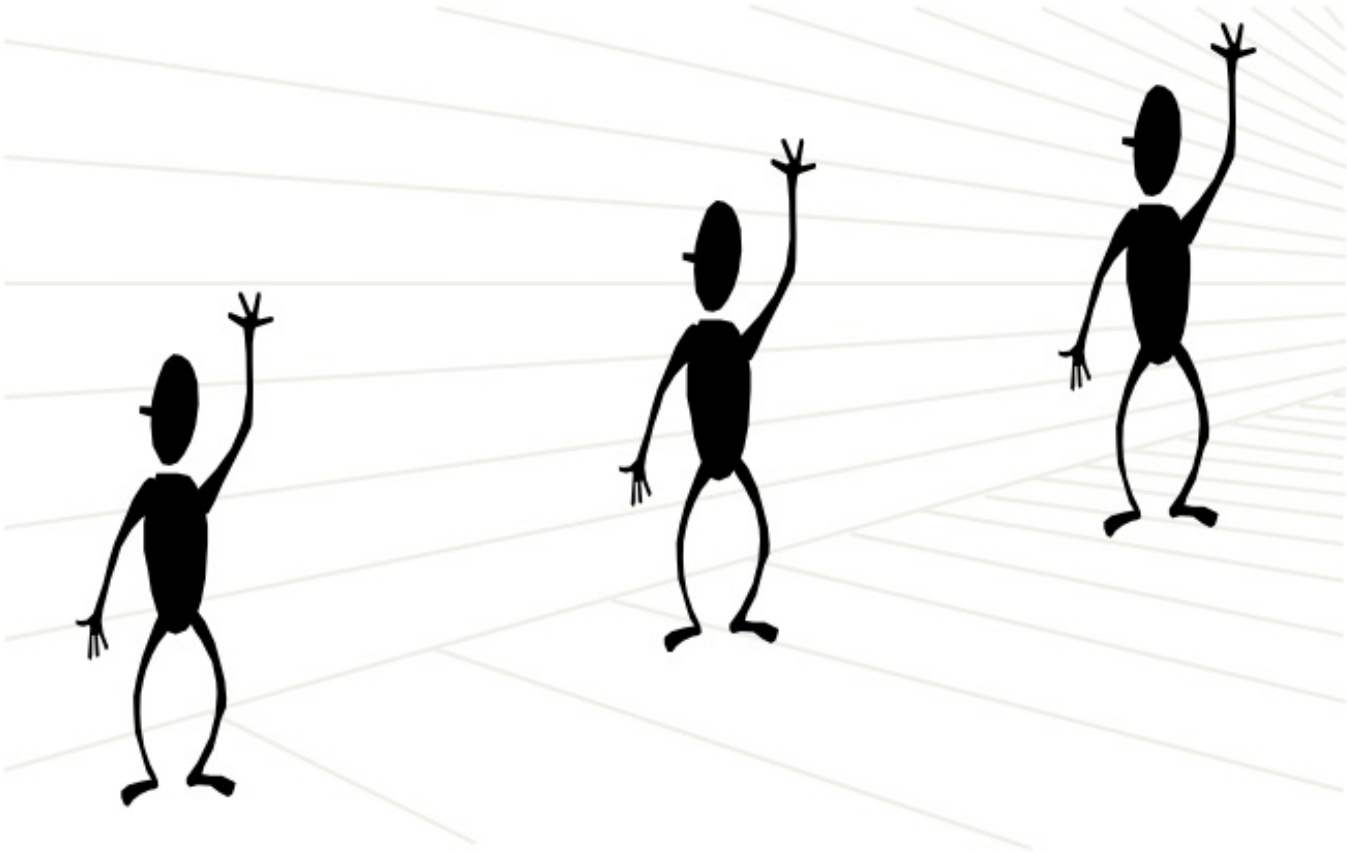
- Points project to points
- Lines project to lines

- Angles are not preserved.

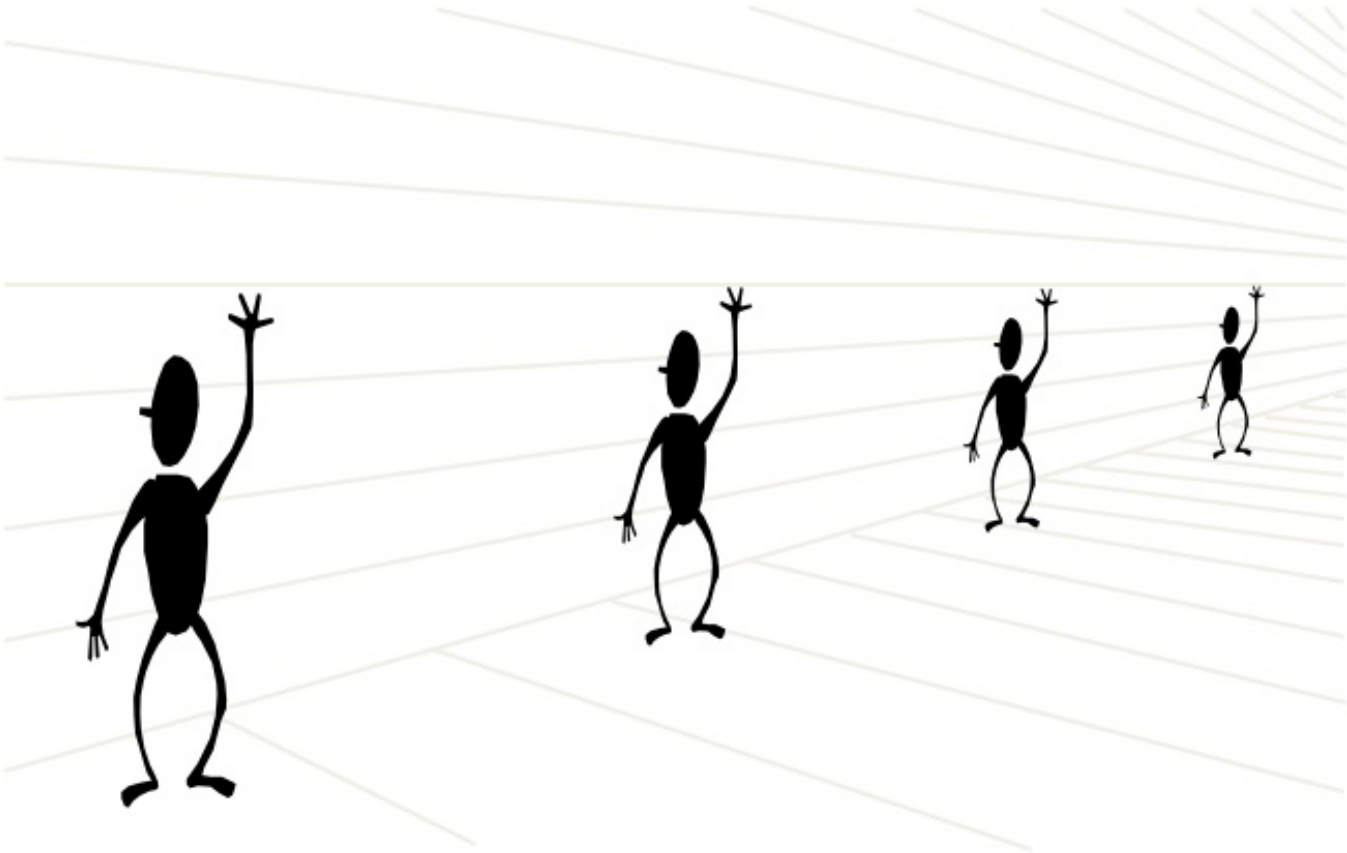
Use cues from scene geometry



Use cues from scene geometry

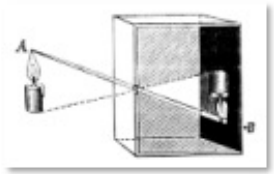


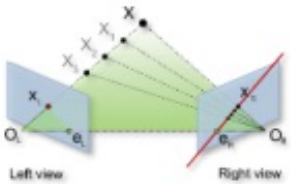
Use cues from scene geometry



Still, Depth from single Image is ambiguous. Why?

Computer Vision Overview



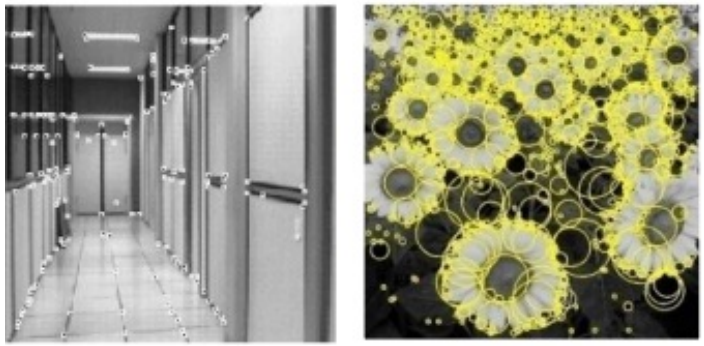


1. Low-level vision
   - image processing, edge detection, feature detection, cameras, image formation
2. Geometry and algorithms
   - projective geometry, stereo, structure from motion,
3. Recognition
   - face detection / recognition, category recognition, segmentation

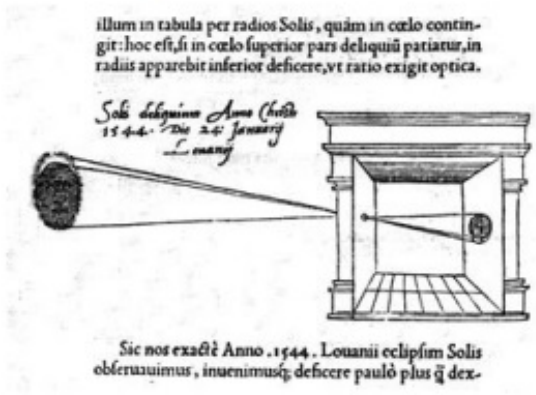
Image processing and Feature Extraction

- Basic image processing and image formation

Filtering, edge detection



Feature extraction



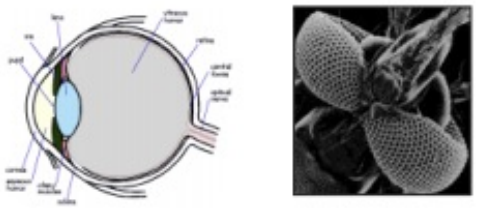
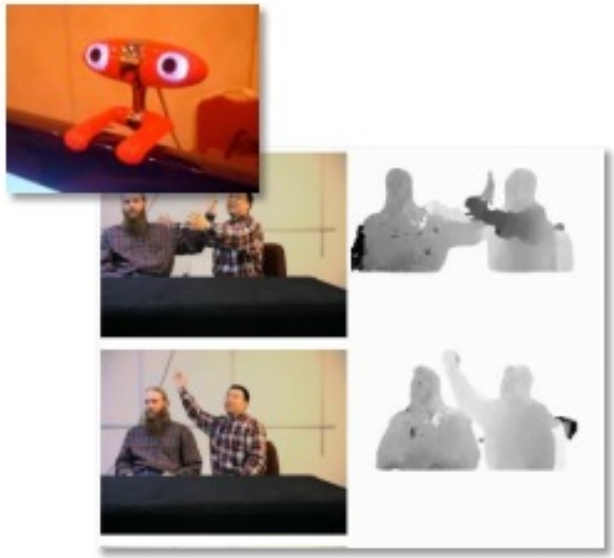


Image formation

Feature Matching

Scene Geometry

Projective geometry



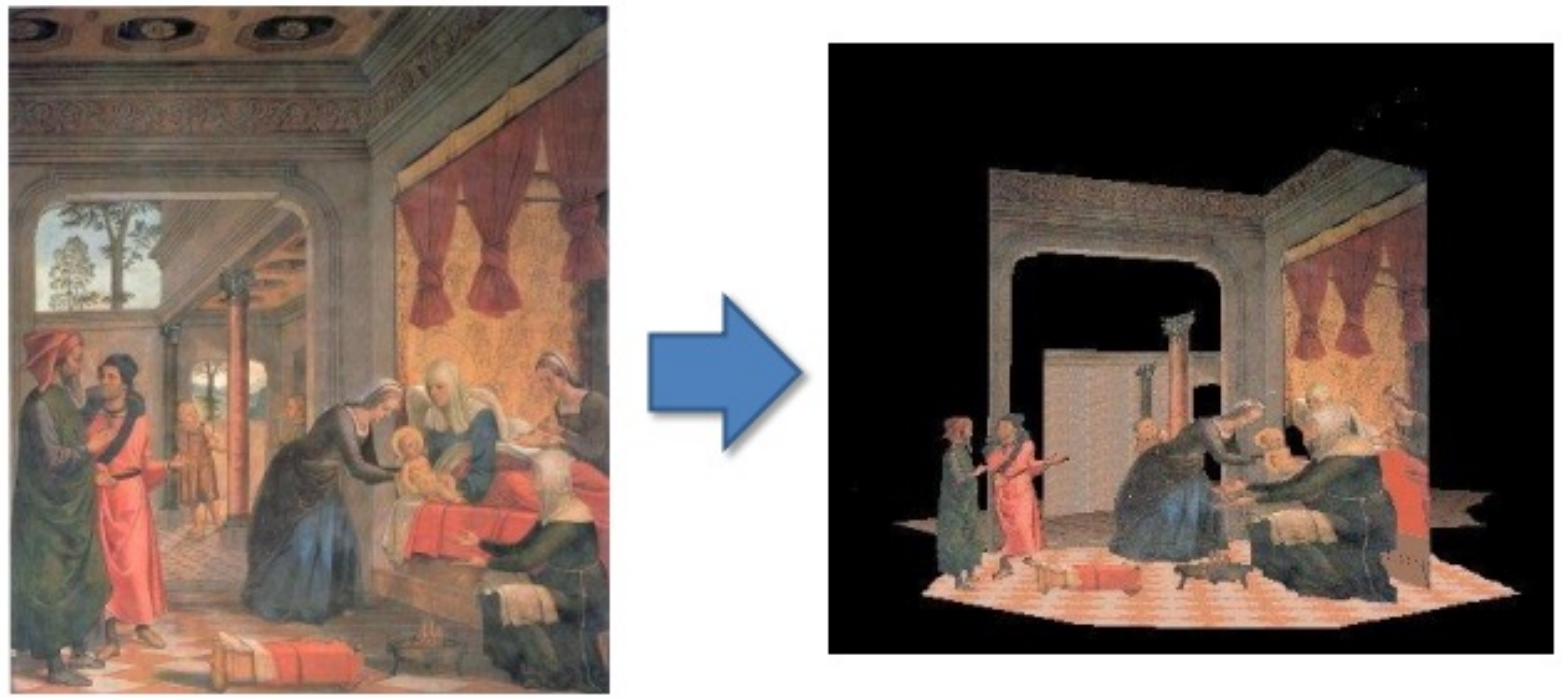
Stereo

Multi-view stereo

Structure from motion

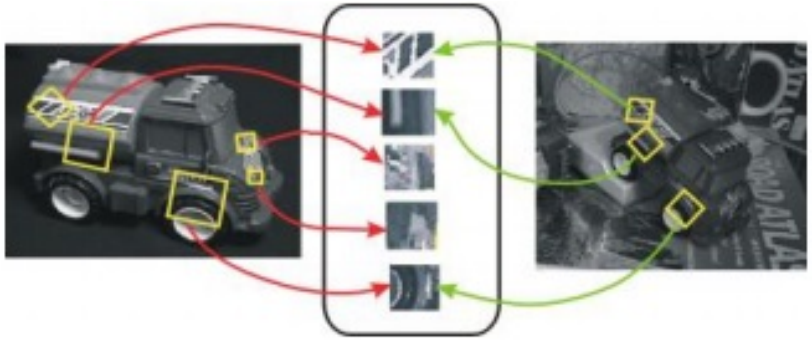
Panorama

Single View Modeling



Detection and Recognition

Face detection and recognition

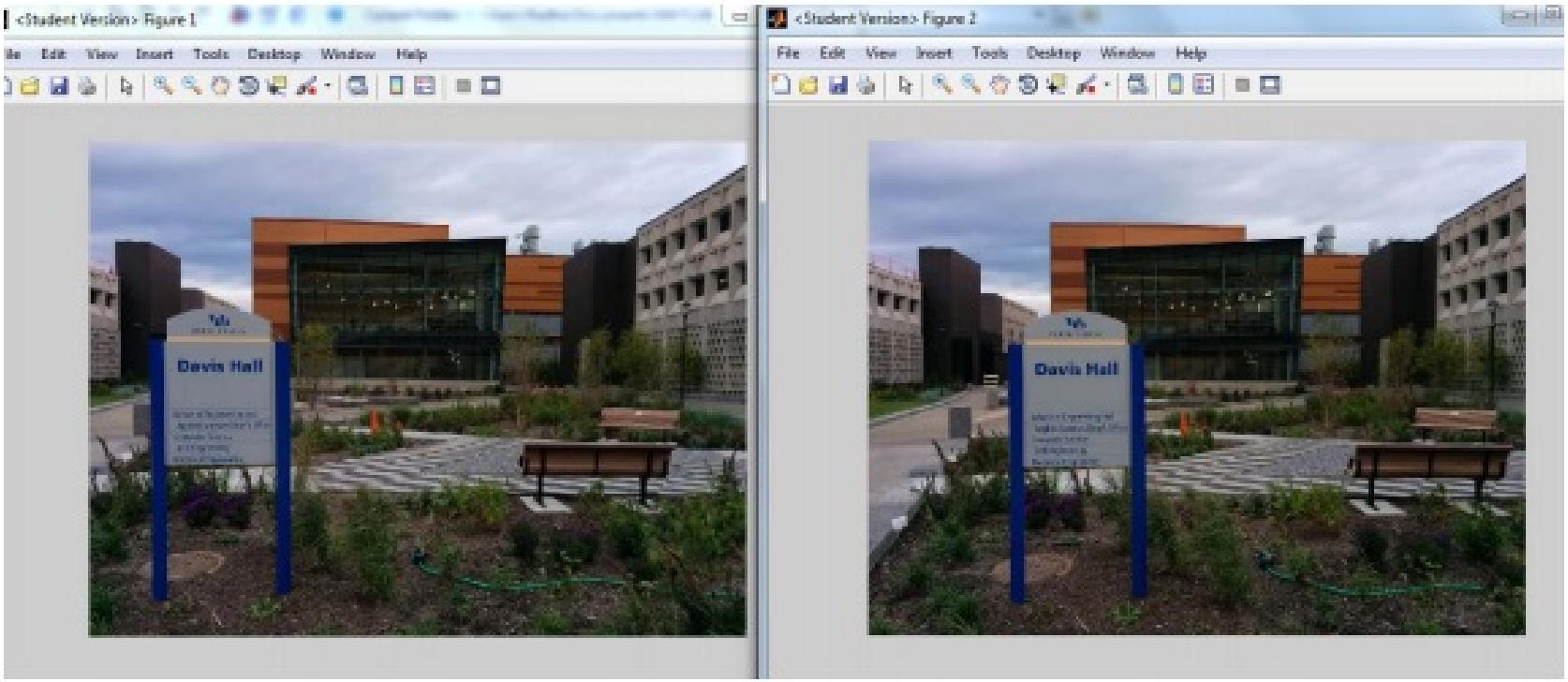


Single instance recognition

Category recognition

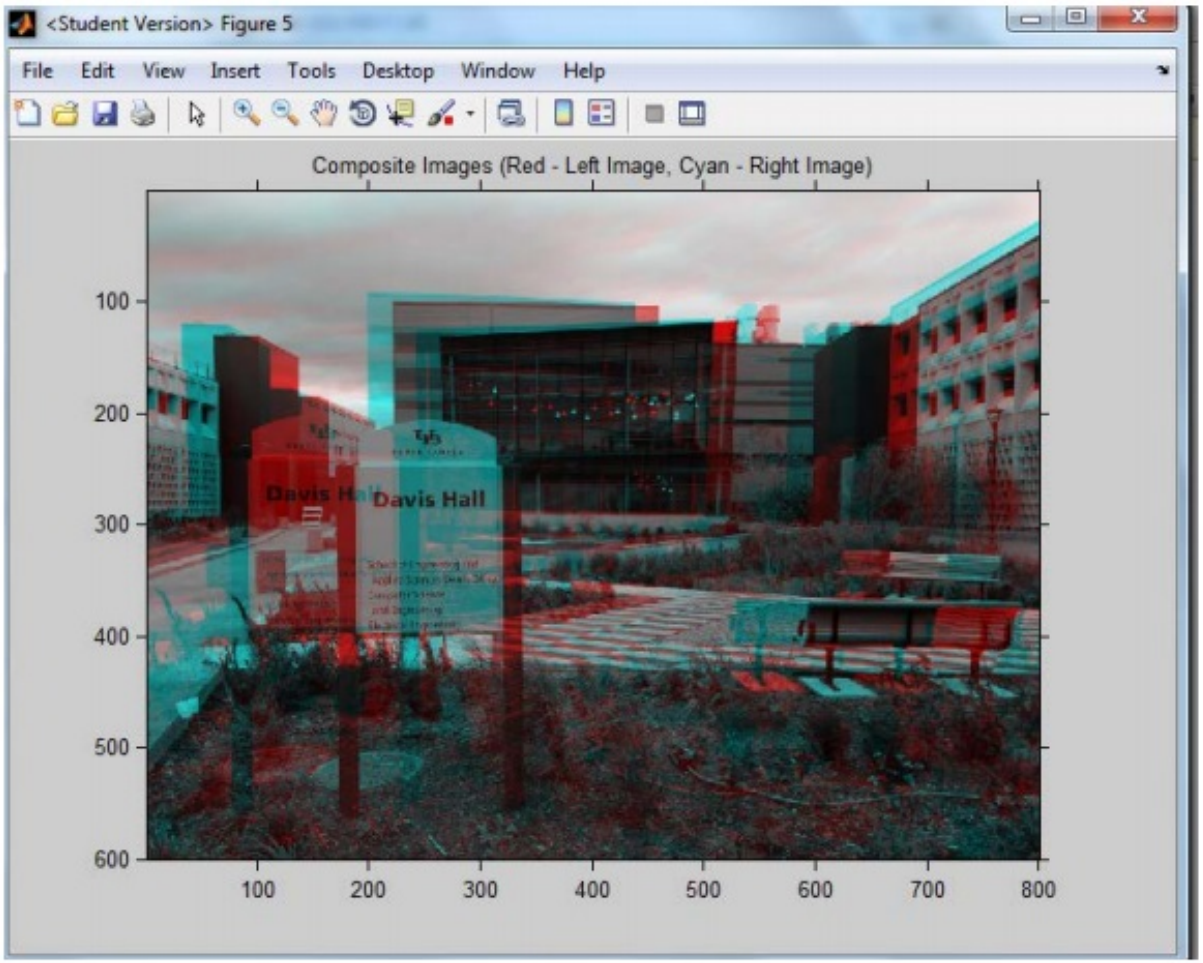
Specific Applications - Pedestrian Detection and Tracking

Depth from Two Images

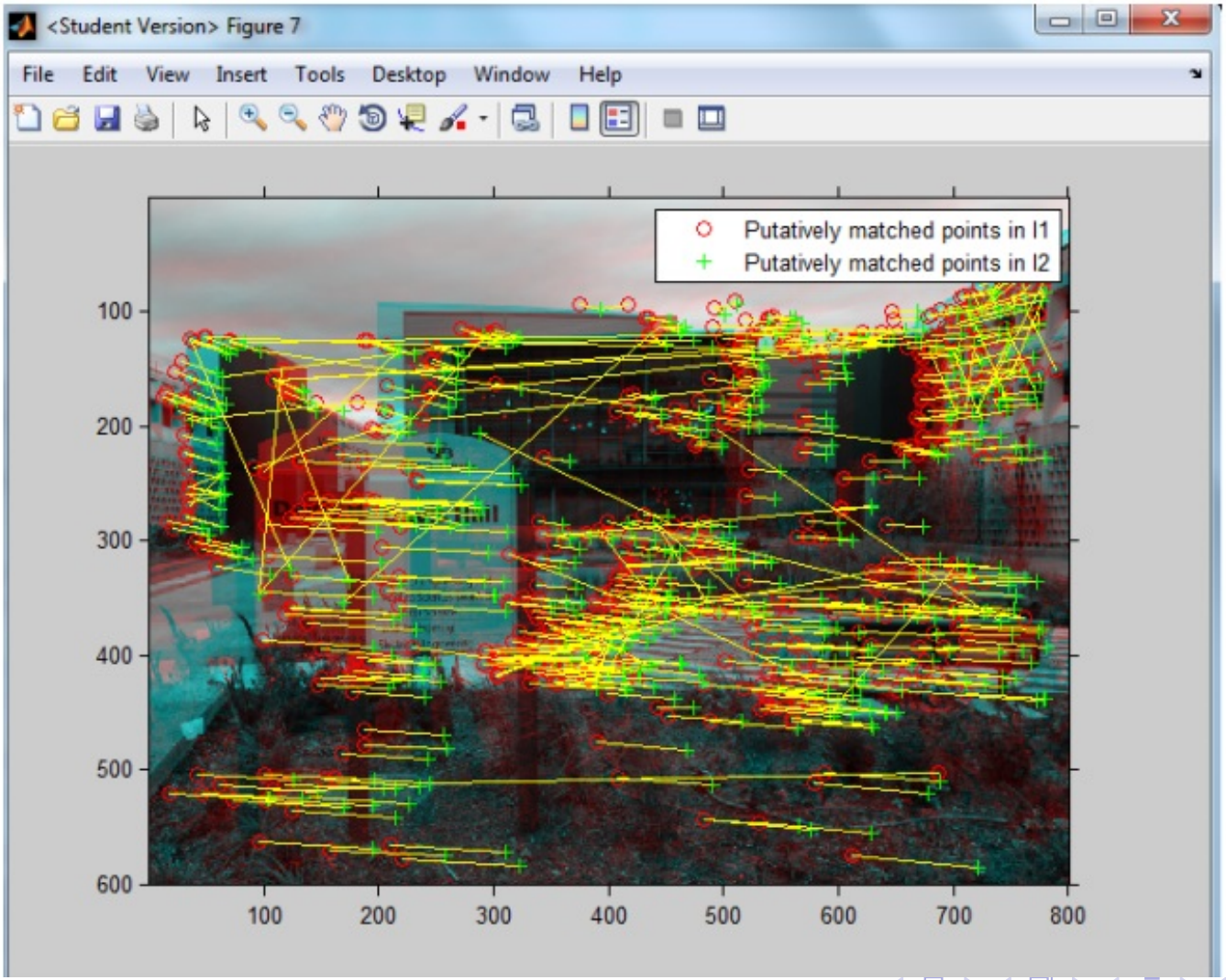


Depth from Two Images - Feature Matching

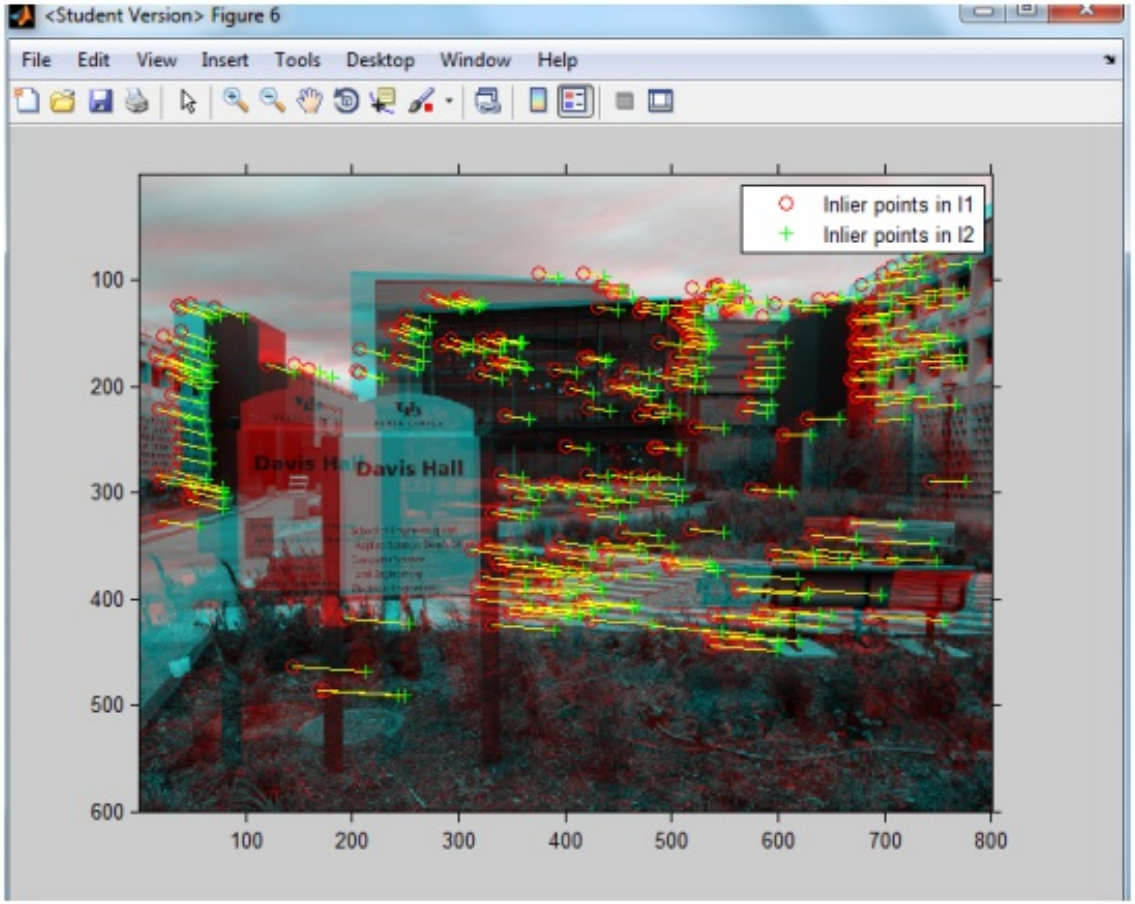
Superposing the two input images on each other and compositing



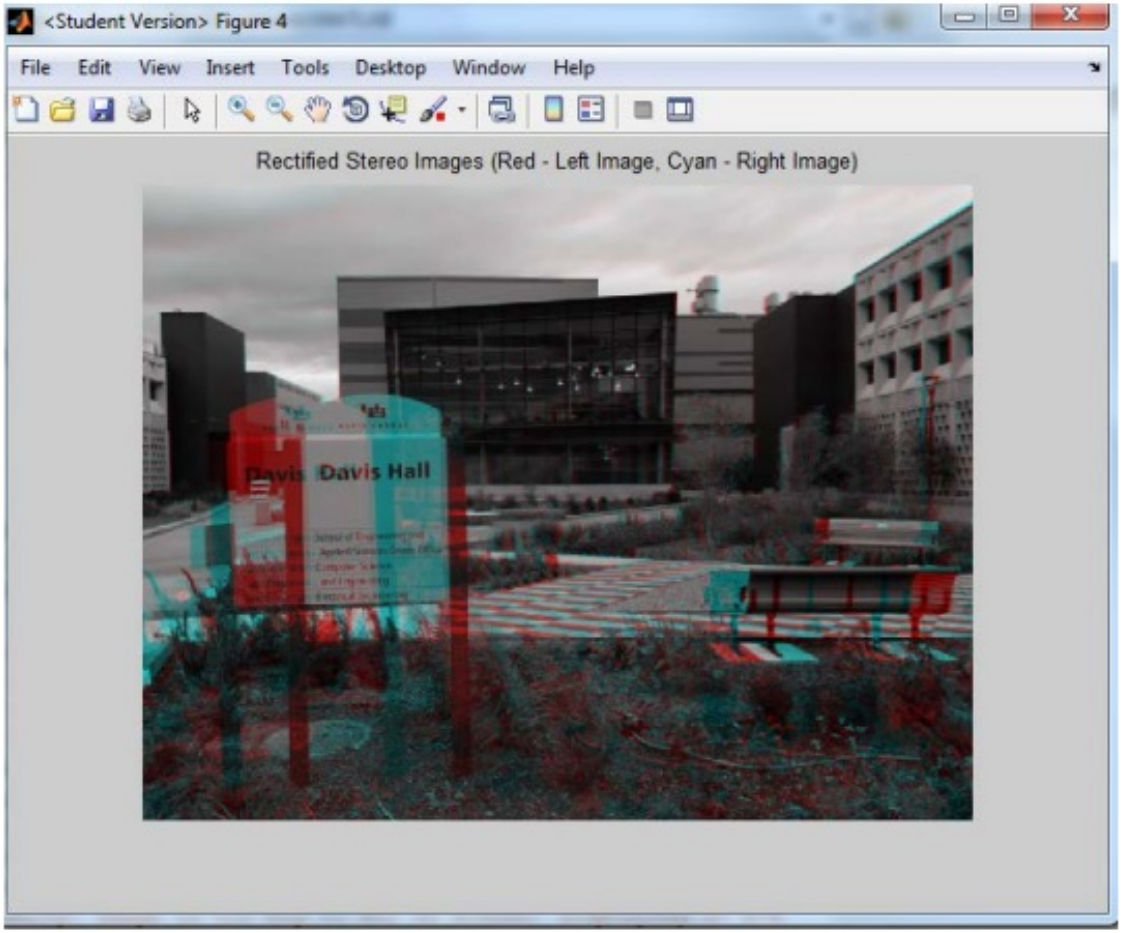
Depth from Two Images - Rectification



Depth from Two Images



Depth from Two Images - Disparity Estimation



View Synthesis

Is it possible to synthesize views from the locations where the cameras are removed? i.e Can we synthesize view from a virtual camera

Course Overview

- Seniors/Grads
- Linear algebra
- Probability?
- Graphics course?
- Vision/image processing course before?
- Machine learning?

Course Logistics

- Enroll in Piazza Page for CSE473/573
- 14 Weeks - Covering the core concepts of Computer Vision. Full attendance recommended
- Download and Configure Enthought Canopy/ Anaconda with OpenCV and Python 3
- **Textbook**: Computer Vision: Algorithms and Applications, by Richard Szeliski

Projects

- **Three** projects - Covering three important aspects (Low, Mid and high level) of Computer Vision
- Using Python and libraries from OpenCV (When recommended)
- Time frame of 3 to 4 weeks per project
- Active support on Piazza, Recitation and Office Hours
- 45% of Course Credit
- Individual Project

Homework

- **Three** Homework
- Active support on Piazza, Recitation and Office Hours
- 15% of Course Credit

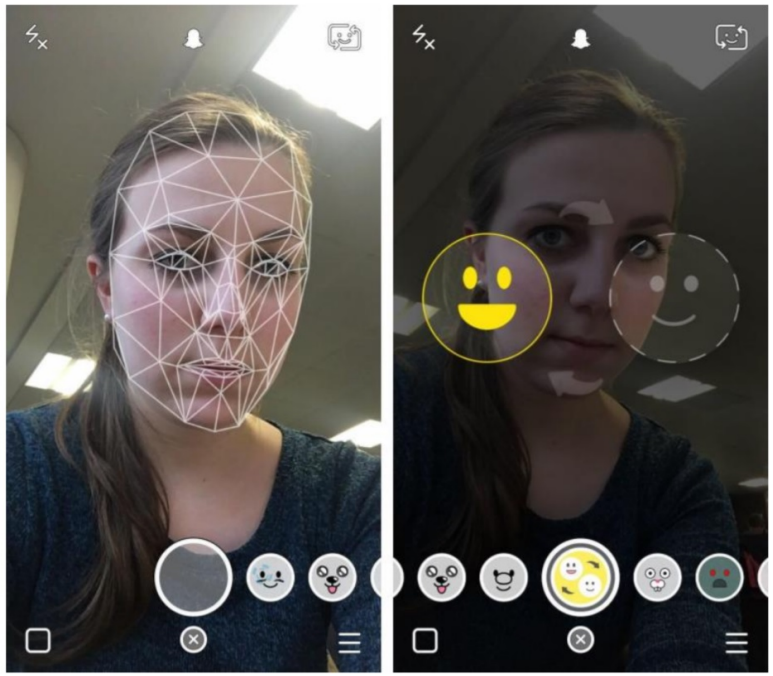
Final

- Questions will be straight-forward and will be prepared from the content presented on lecture slides and concepts on homework/projects
- Final - Dec. 12th - 40% of Course Credit

Vision and Expectations of the Course

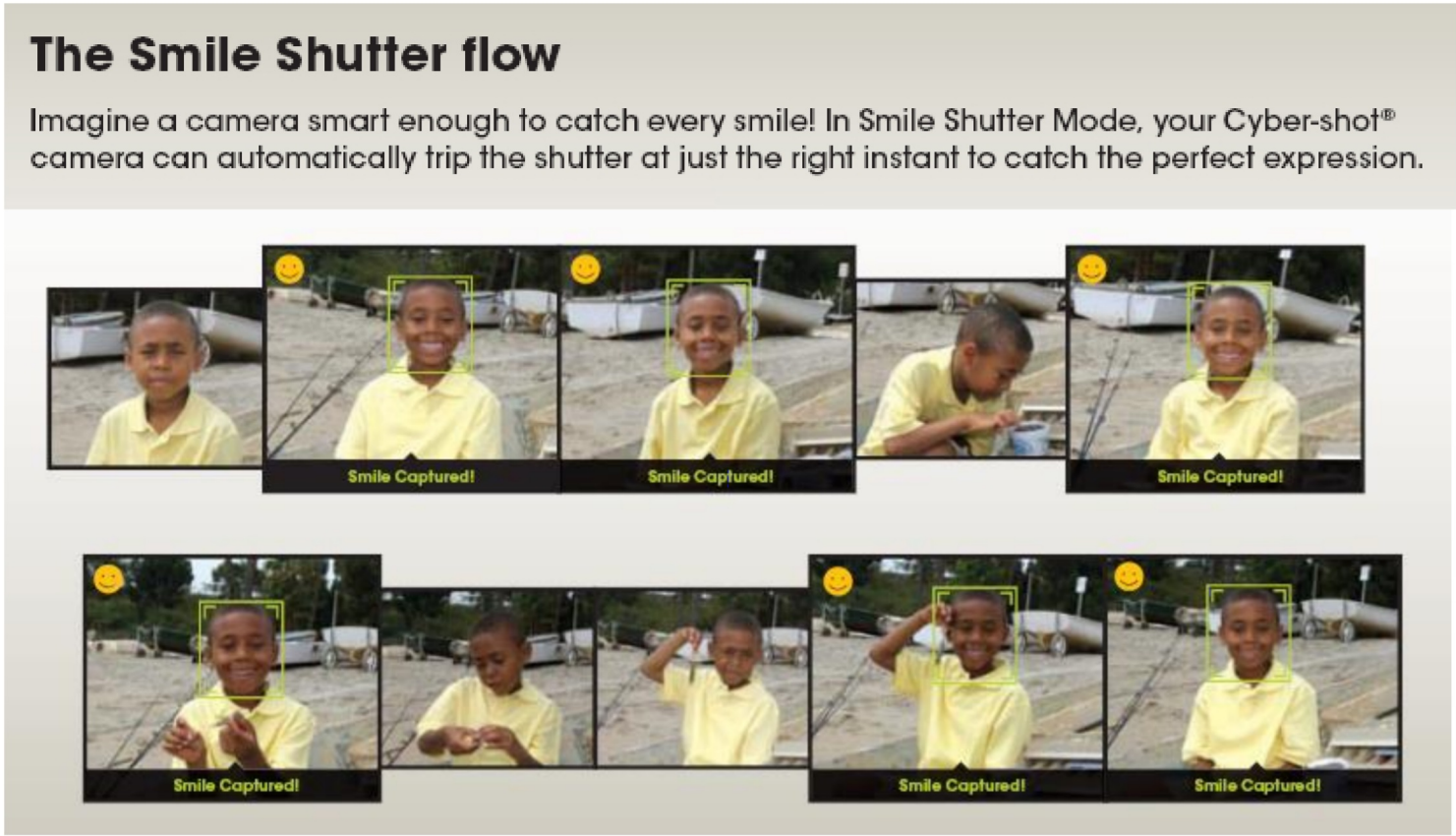
- You will be encouraged and supported to learn the course material at a moderate pace
- The emphasis of the course is to develop practical skills for solving Computer Vision problems
- Academic Honesty will be taken seriously and will be rewarded. Please check the departmental website for penalties

More Applications - Face Detection



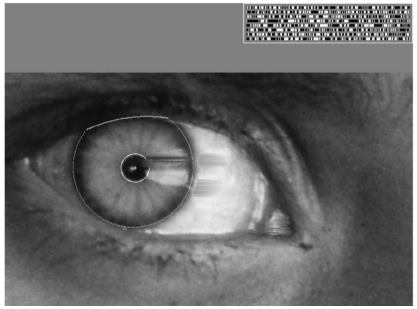
Applications of Face Detection

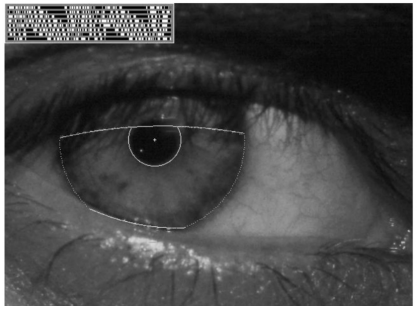
Smile Detection



Biometrics and Image Processing

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

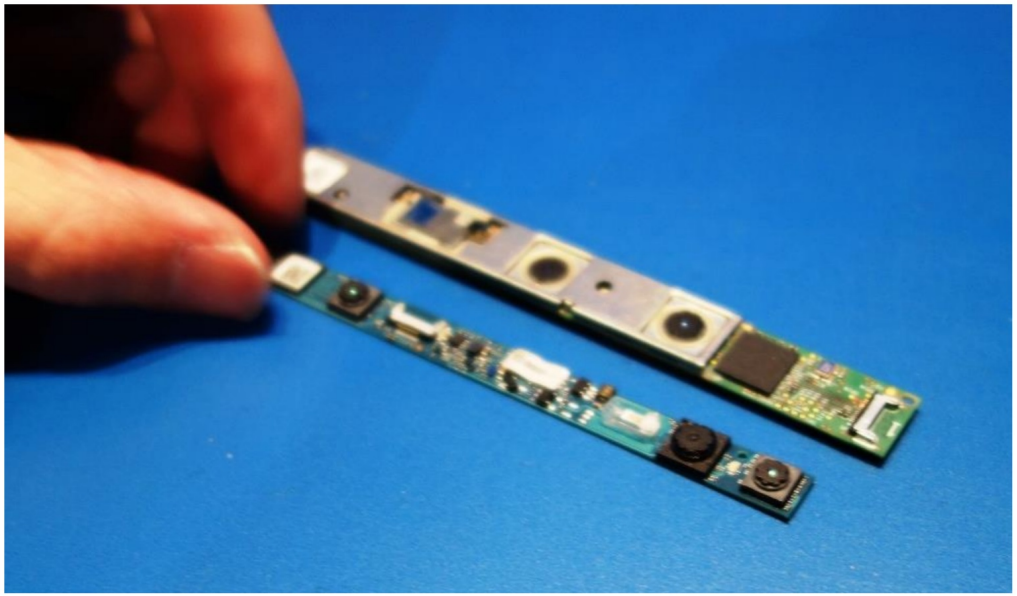


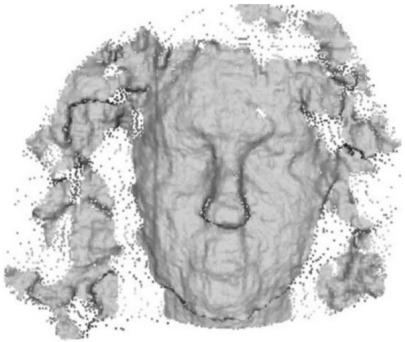


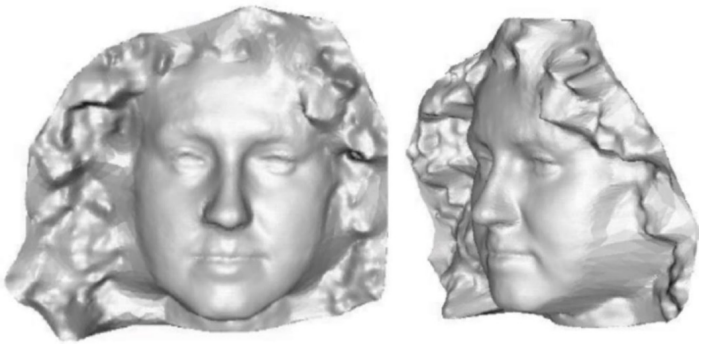
Face Recognition

How to solve this problem?

Shape Reconstruction using depth sensors





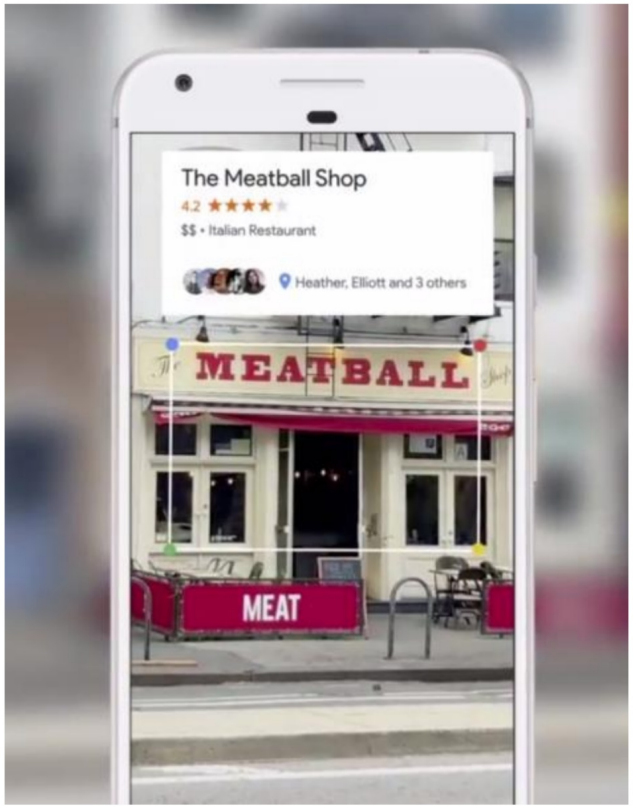


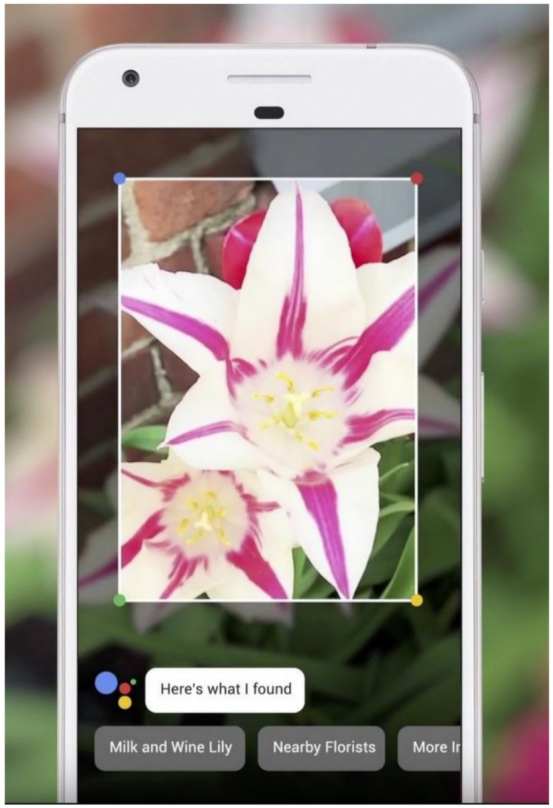
Single depth frame

Reconstructed 3D mesh

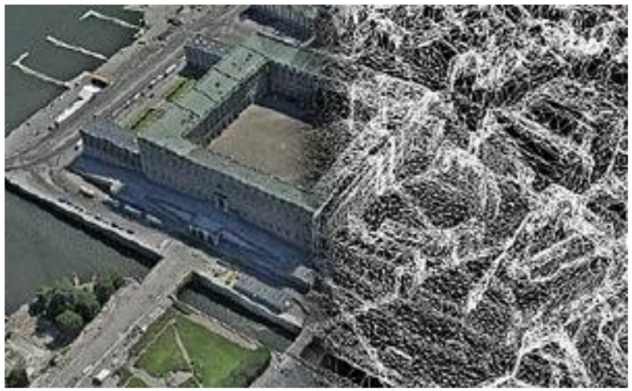
Visual Search

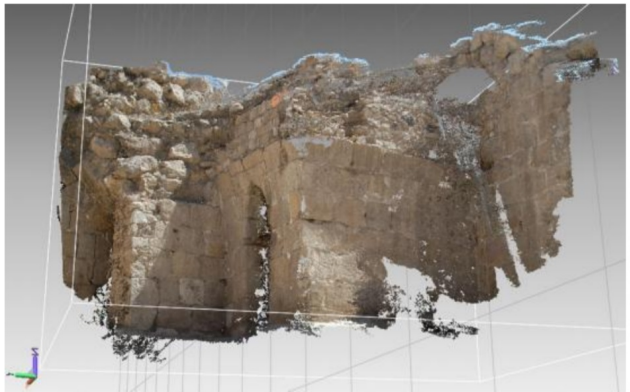
e.g., Google Lens





Building Rome in a Day!

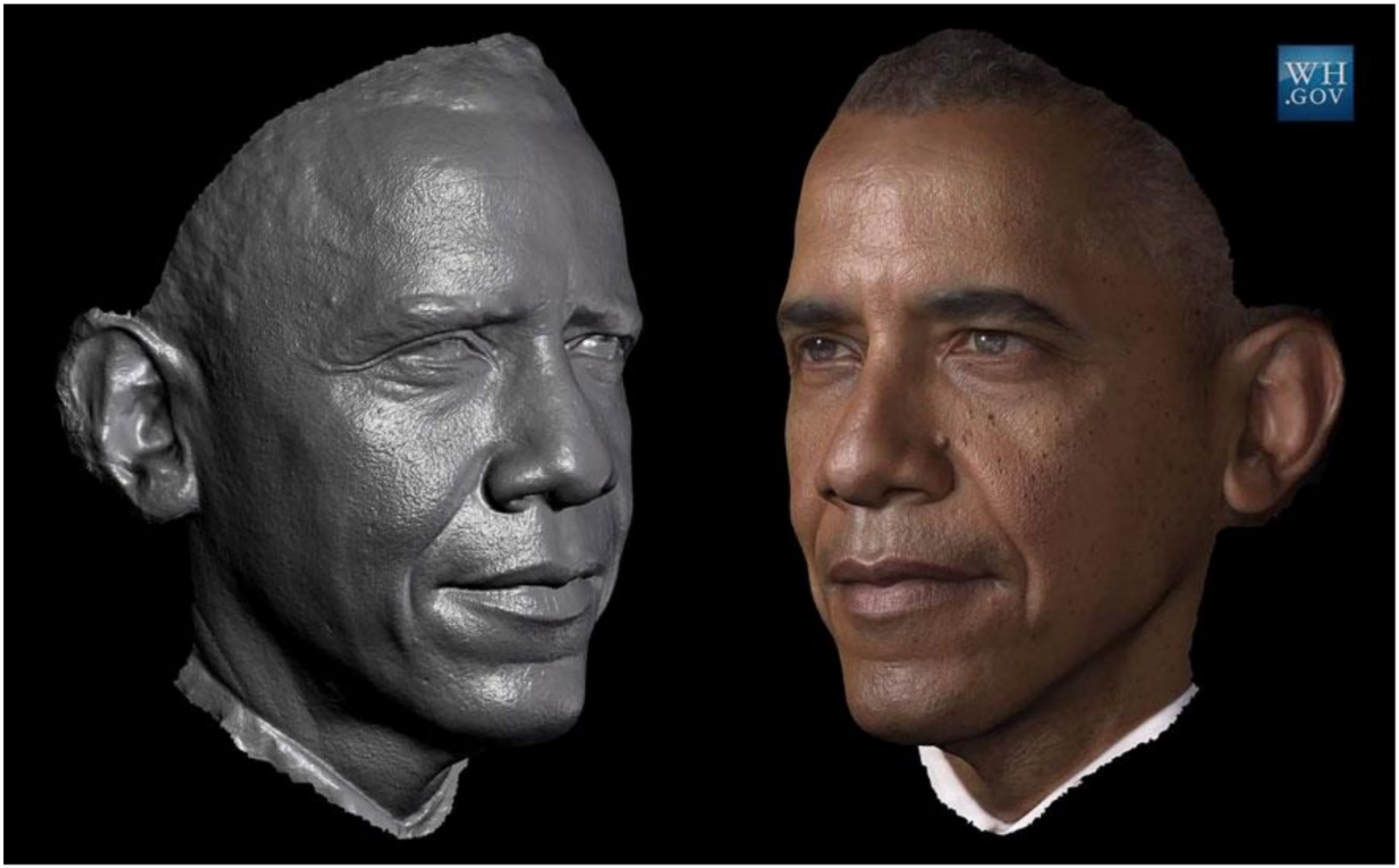




Building Rome in a Day: Agarwal et al. 2009

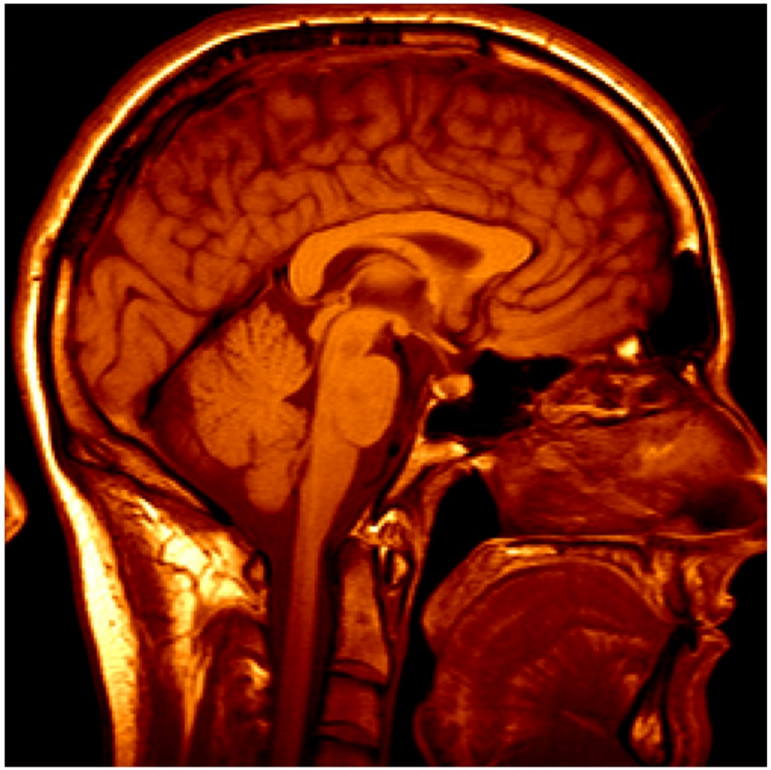
Multi-Camera Applications

3D Scanning



Visual Effects

Medical Imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

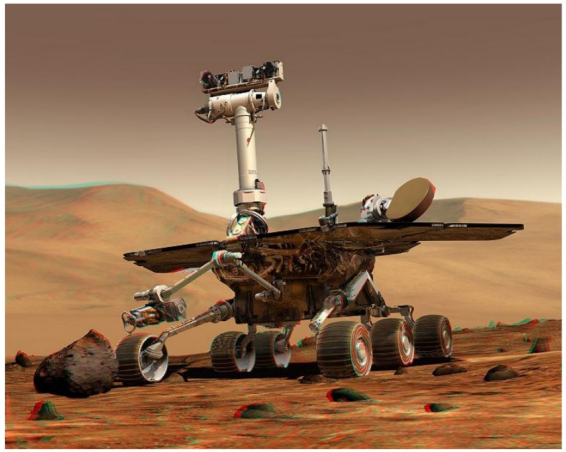
Autonomous Vehicle Navigation

Google Street View

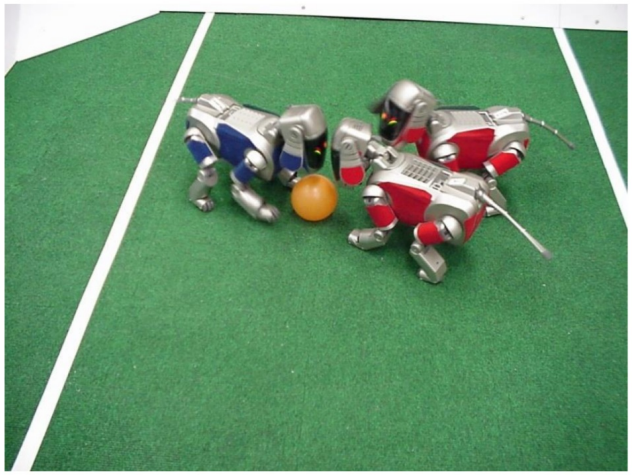
Machine Vision

Vision-guided robots position nut runners on wheels

Robots and Vision



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

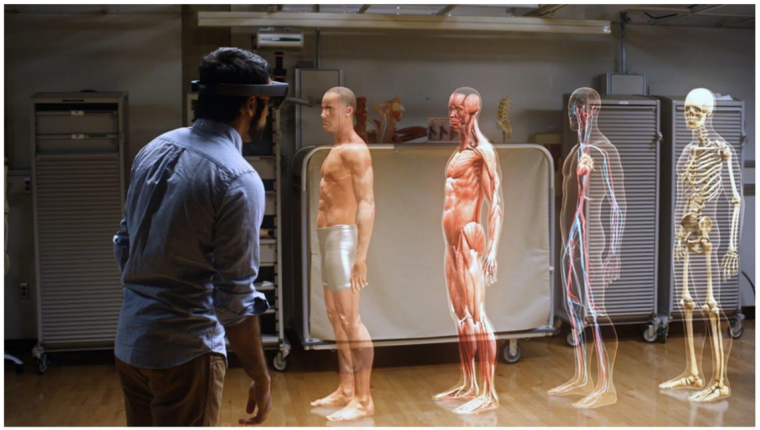


http://www.robocup.org/

Saxena et al. 2008
STAIR at Stanford

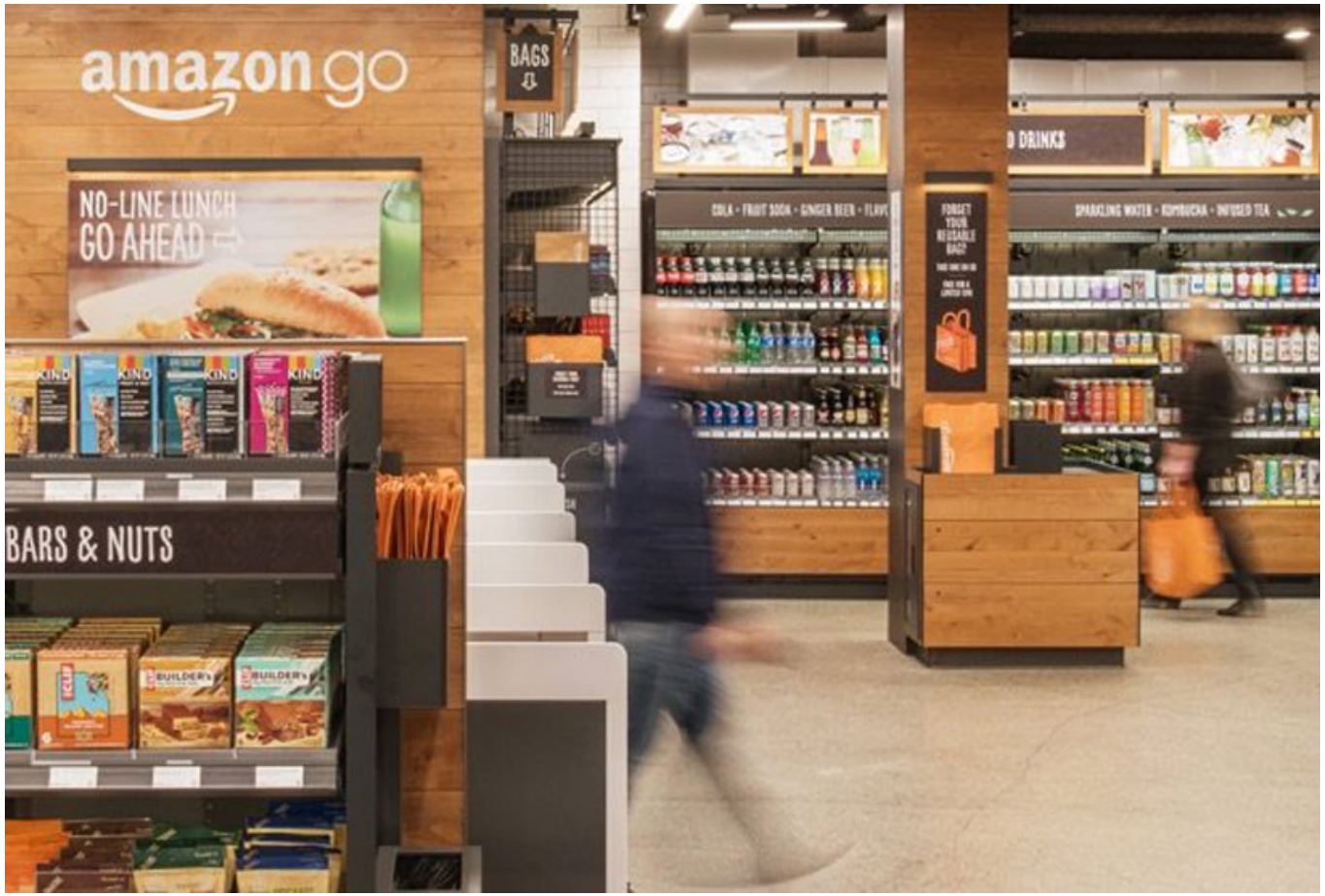
amazon Prime Air

Augmented Reality



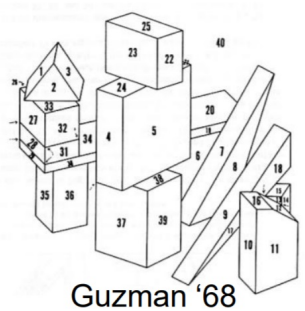
MS HoloLens, Oculus, Magic Leap,
ARCore / ARKit

Multi-Sensor Fusion Applications

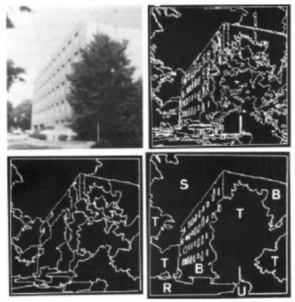


A Brief History of Computer Vision

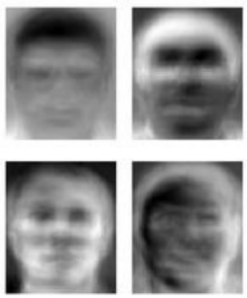
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis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



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

Image Formation

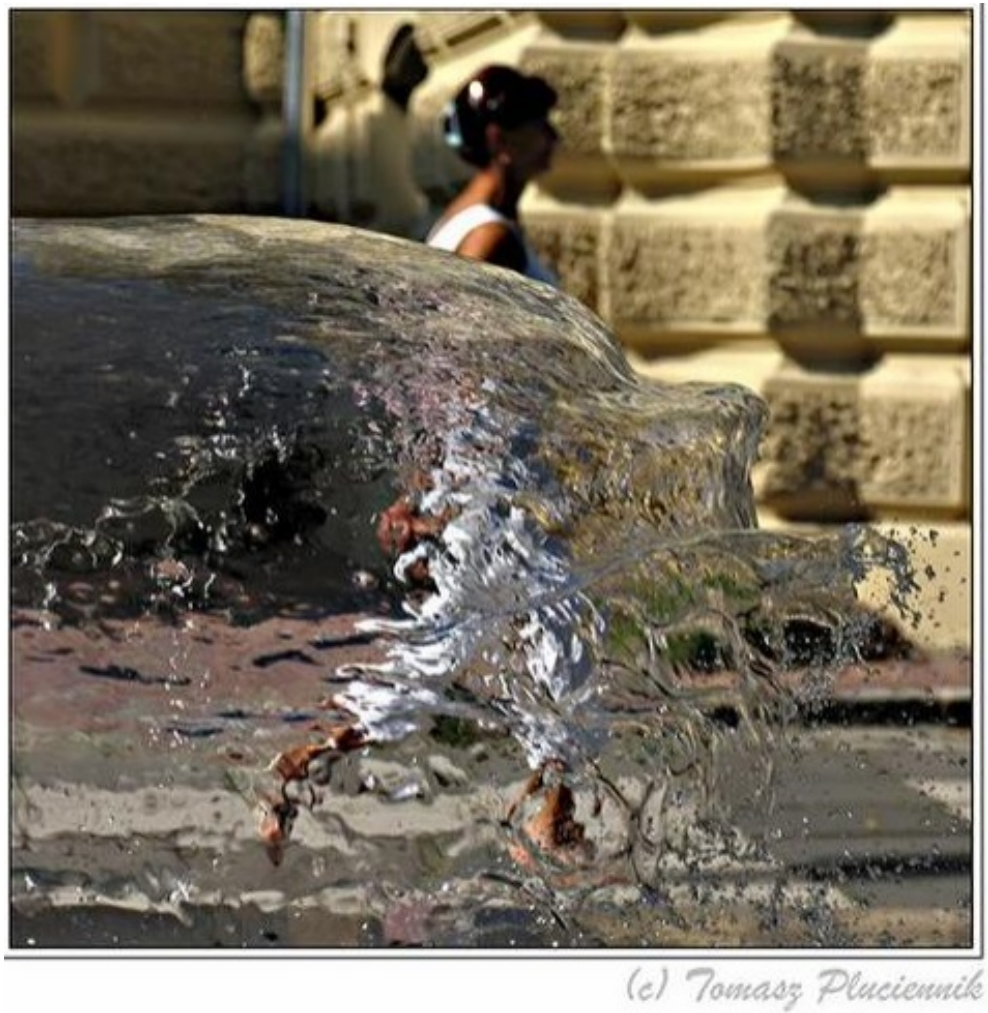


Image Formation Process

All the images are a resultant of the following entities

- Lighting Conditions
- Scene Geometry
- Surface Properties
- Camera properties

Image Formation Process

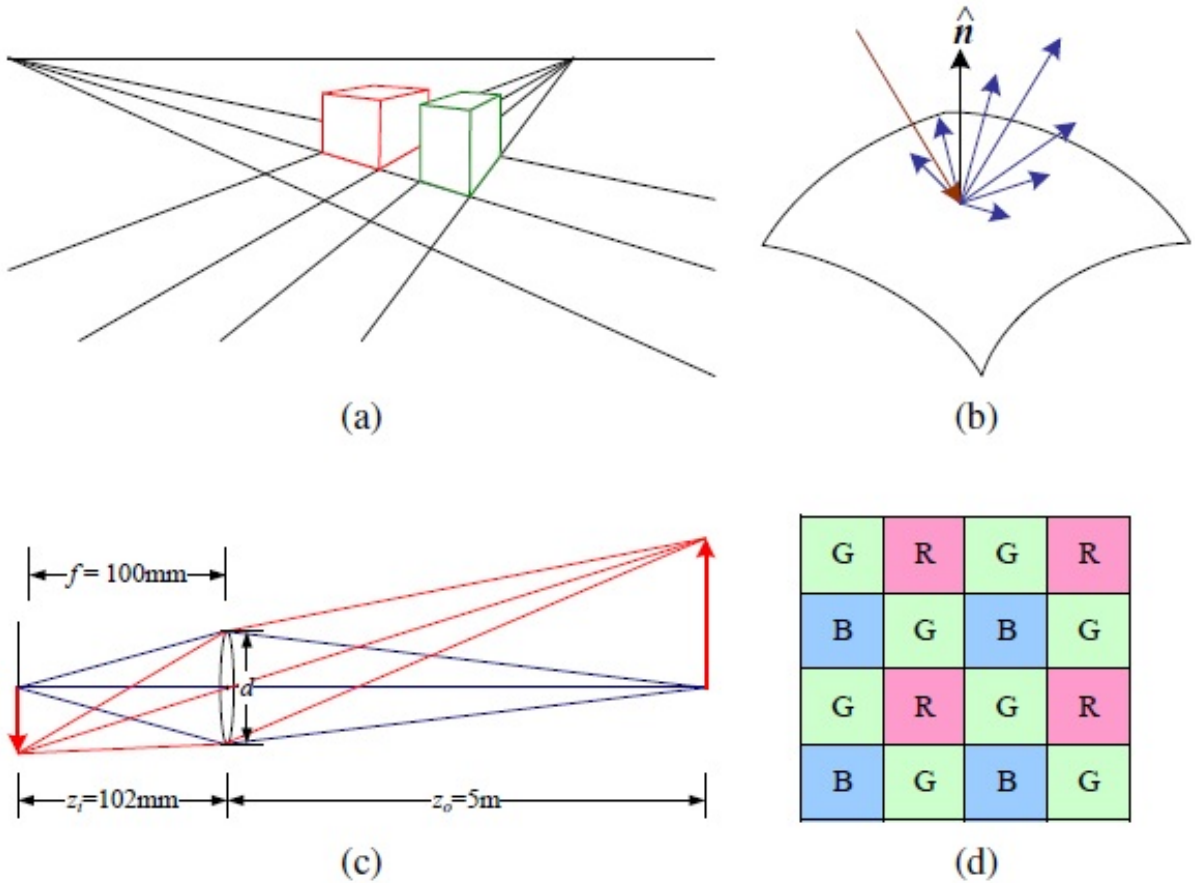


Figure 2.1 A few components of the image formation process: (a) perspective projection; (b) light scattering when hitting a surface; (c) lens optics; (d) Bayer color filter array.

3D Co-ordinate Transformations

DoF - Degrees of Freedom

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\left[\begin{array}{c\|c} \boldsymbol{I} & \boldsymbol{t} \end{array}\right]_{3\times4}$	3	orientation	
rigid (Euclidean)	$\left[\begin{array}{c\|c} \boldsymbol{R} & \boldsymbol{t} \end{array}\right]_{3\times4}$	6	lengths	
similarity	$\left[\begin{array}{c\|c} s\boldsymbol{R} & \boldsymbol{t} \end{array}\right]_{3\times4}$	7	angles	
affine	$\left[\begin{array}{c} \boldsymbol{A} \end{array}\right]_{3\times4}$	12	parallelism	
projective	$\left[\begin{array}{c} \tilde{\boldsymbol{H}} \end{array}\right]_{4\times4}$	15	straight lines	

Photometric Image Formation

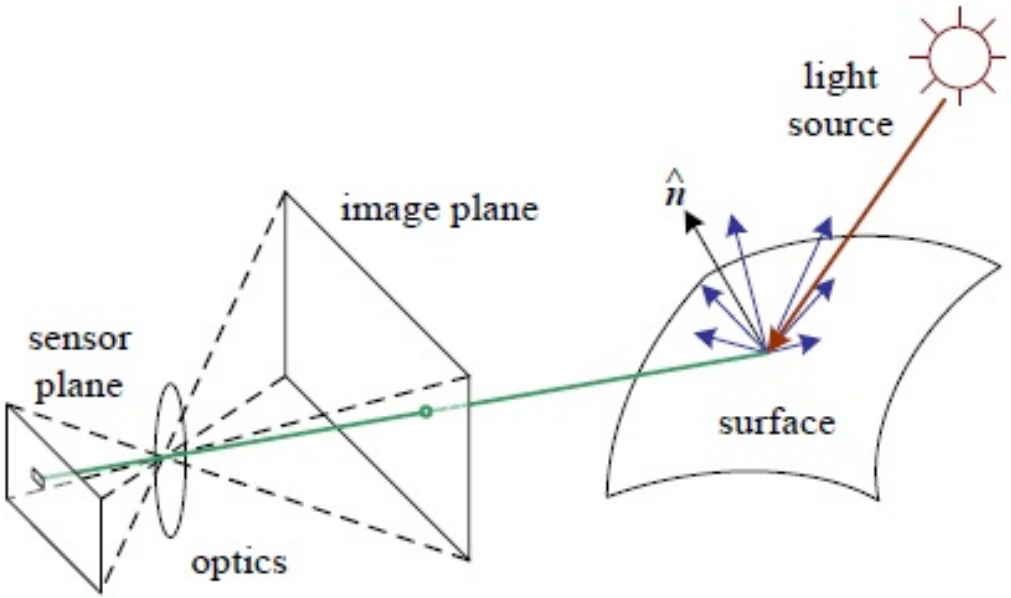


Figure 2.14 A simplified model of photometric image formation. Light is emitted by one or more light sources and is then reflected from an object's surface. A portion of this light is directed towards the camera. This simplified model ignores multiple reflections, which often occur in real-world scenes.

Image Capture pipeline

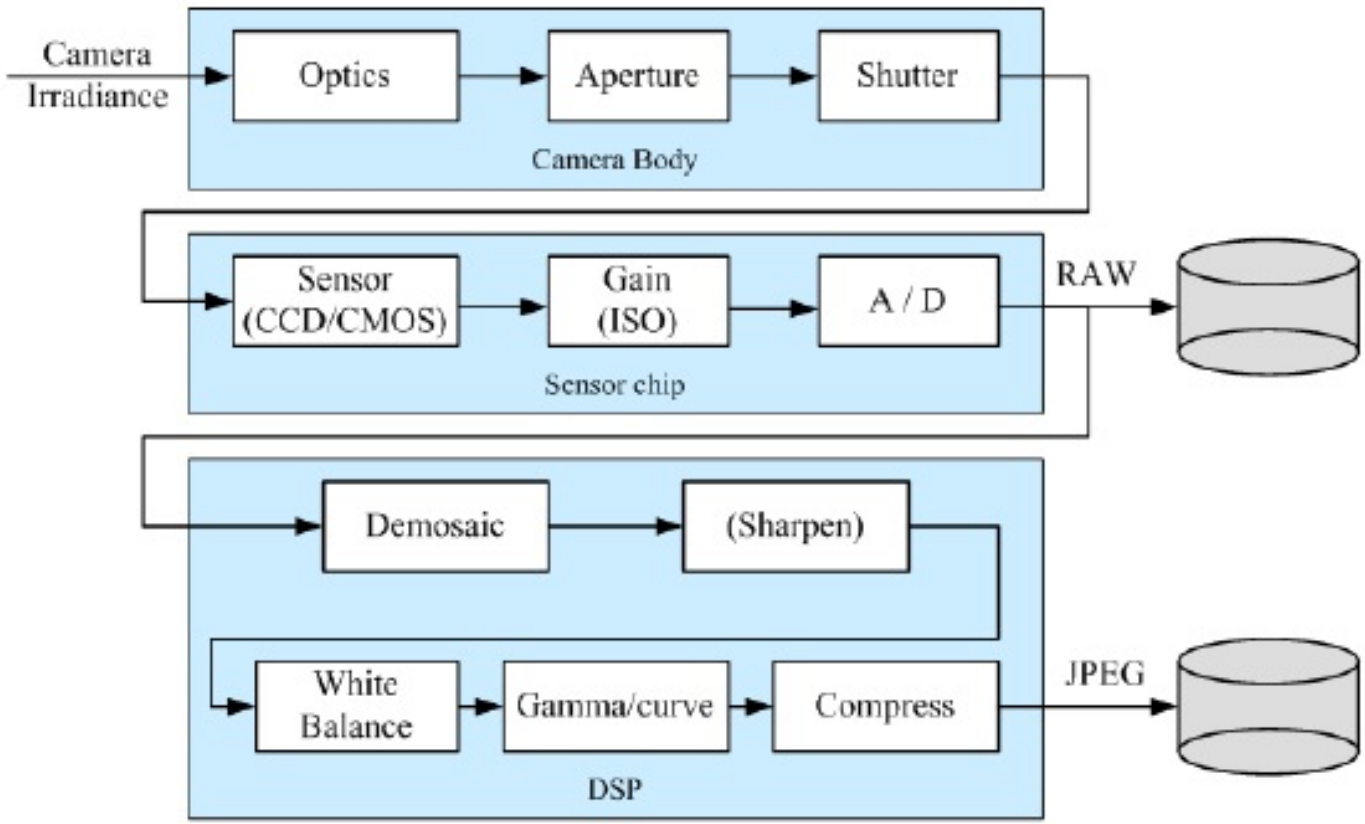


Figure 2.23 Image sensing pipeline, showing the various sources of noise as well as typical digital post-processing steps.

Slide Credits

Radhakrishna Dasari, Fei Fei Li, S Seitz, Noah Snavely, James Hays, Derek Hoiem, Richard Szeliski and Svetlana Lazebnik