CS 554
Computer Vision

Introduction

Hamdi Dibeklioğlu

Slide Credits: L. van der Maaten

Course Setup

- **Lecture Hours:** Monday 10:40-12:30, Thursday 8:40-10.30
  - While every other week we will have 2+2 hours of lectures, other weeks there will be one 2-hours lecture.
  - **First week only, Thursday lecture will be from 9:40 to 10.30.**
  - It is your responsibility to follow the schedule from the course website.

Reading

- No required textbook, but the following books are recommended:
  - R. Szeliski, Computer Vision: Algorithms and Applications, Springer Science & Business Media, 2010. http://szeliski.org/Book/
  - D.A. Forsyth and Jean Ponce, Computer Vision: A Modern Approach. Prentice Hall Professional Technical Reference, 2002.
  - I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016. http://www.Deeplearningbook.org.

Course Website

- Slides and schedule will be provided on the course homepage:

  http://www.cs.bilkent.edu.tr/~dibeklioglu/teaching/cs554/

Topics

- Human Vision
- Image representation, feature extraction
- Camera geometry and image alignment
- Stereo and 3-D reconstruction
- Motion
- Deep learning
- Face processing
- Object and scene recognition
- Conditional and Markov random fields
- Tracking
- Action/activity recognition

Prerequisites

- Basic calculus and linear algebra
- Basic probability theory and statistics
- Good programming skills (preferably in python, matlab is OK too)

Assessment

- Homework: 20%
  - A single programming homework.
- Midterm: 30%
  - In the second half of the semester, in class.
- Literature Survey & Presentation: 10% + 5%
  - By groups of two students.
- Project: 30%
  - By groups of two students.
- Attendance & In-class participation: 5%
- Schedule and details will be announced on the course page.

Direct F Policy

- **Any of the following will directly result in an F grade:**
  - not submitting a project or homework (including report),
  - not presenting a survey on the pre-scheduled date,
  - being absent in the midterm,
  - being absent in a project presentation,
  - missing more than 12 lecture hours.

Literature Survey and Presentation

- Groups of two will choose a topic related to computer vision, and prepare a survey on it.
- Surveys should be based on about 10 papers (report: 5 pages max).
- You will make a presentation on your survey in class.
  The presentation should be in parallel with your report.

Literature Survey and Presentation

- Survey topics should be confirmed first. Very similar topics to others' will not be allowed (priority: first come, first served).
- To this end, your chosen survey topic and a few lines explanation (indicating group members) should be sent to dibeklioglu@cs.bilkent.edu.tr.
  - Subject line of your email has to be: **cs554_2020s_survey**
  - Deadline: 20 February 2020, 23.59 (Turkey time)

Project

- By group of two students
- Projects related to your research topics are encouraged
- Three stages:
  1) Proposal: one-page description of the project topic and the planning for the project.
     <u>Due: 24 February 2019 (23:59 Turkey time)</u>
  2) Progress report & presentation (report: 2 pages max.)
  3) Final report & presentation (report: 4 pages max.)

Project

- Build a “real” computer vision application and test it on some data
- You may re-use existing code, but indicate what you code you have re-used and what code you have developed yourself

Examples of Projects

- Constructing panorama or 3D reconstruction based on a sequence of images
- Performing face verification for automatic passport control
- Recognizing the identity or expression of faces
- Recognizing gestures for communication via sign language
- Recognizing or detecting objects in images (Pascal VOC / ImageNet)
- Segmenting parts in medical images
- Performing restoration of old, deteriorated photographs
- Performing automatic morphing of images
- Performing video stabilization using object tracking
- Identifying the writer of a piece of handwriting

Computer Vision

Computer Vision

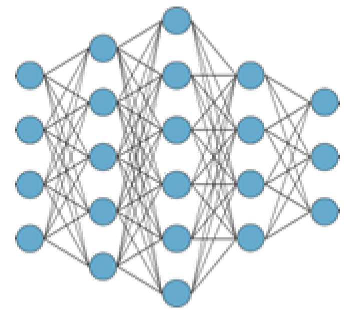
Computer vision tries to get computers to extract information from images

Computer Vision

Scene

Sensor

Interpreter



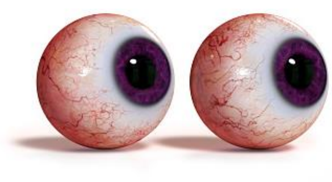
Interpretation

Mountain
Lake
Trees
Sky

Human Vision

Scene

Sensor



Interpreter

Interpretation

Mountain
Lake
Trees
Sky

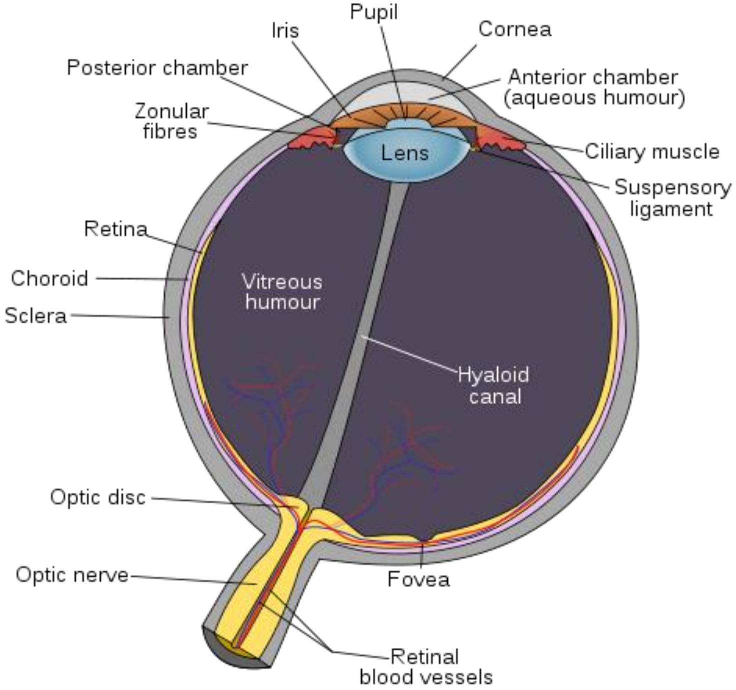
Human Vision

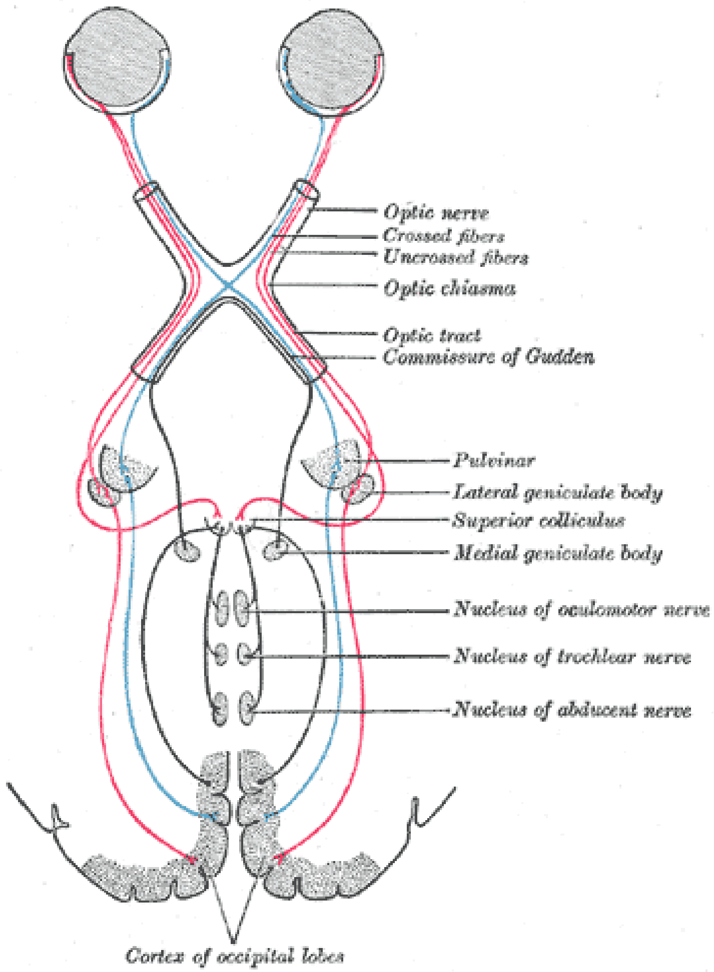
- Can you name some things that influence what we see?

Human Vision

- Can you name some things that influence what we see?
  - What objects are where in the world (and how they are deformed)
  - Lighting conditions of the environment
  - Position and orientation of the eyes (viewpoint)
  - Your own brain!!!

Human Vision



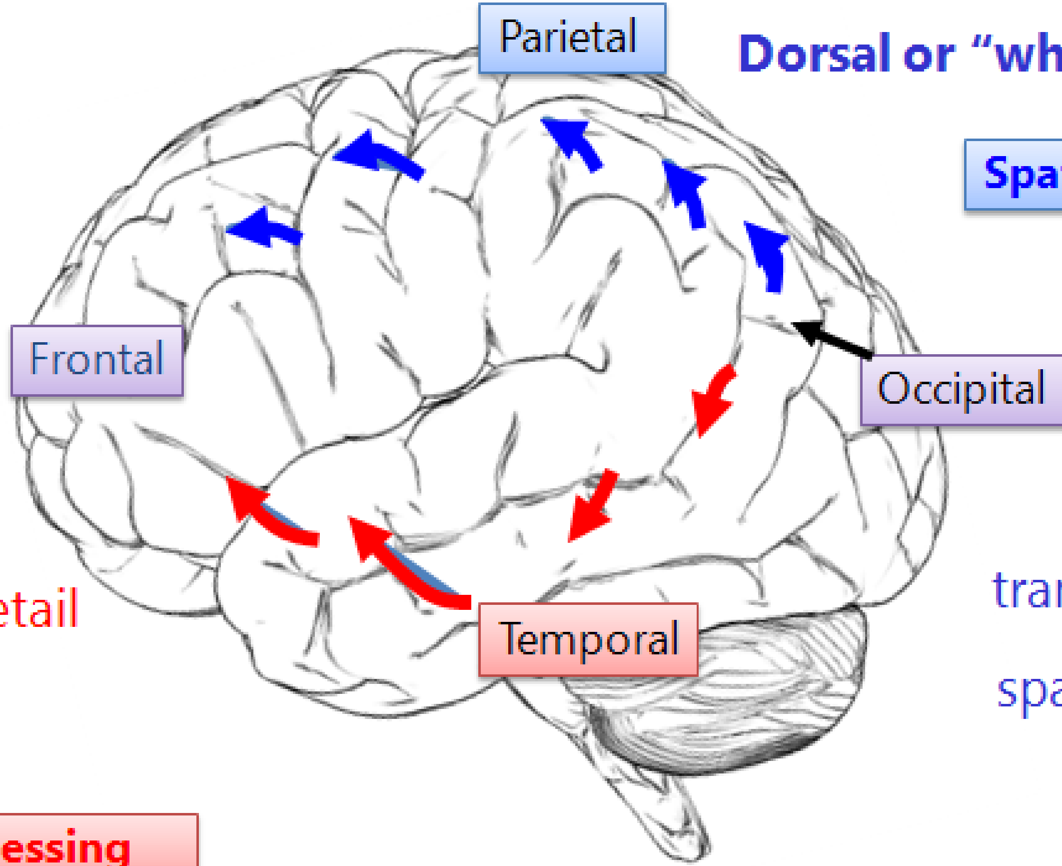


Human Vision

Dorsal or “where” stream

Spatial processing

location
movement
spatial transformations
spatial relations



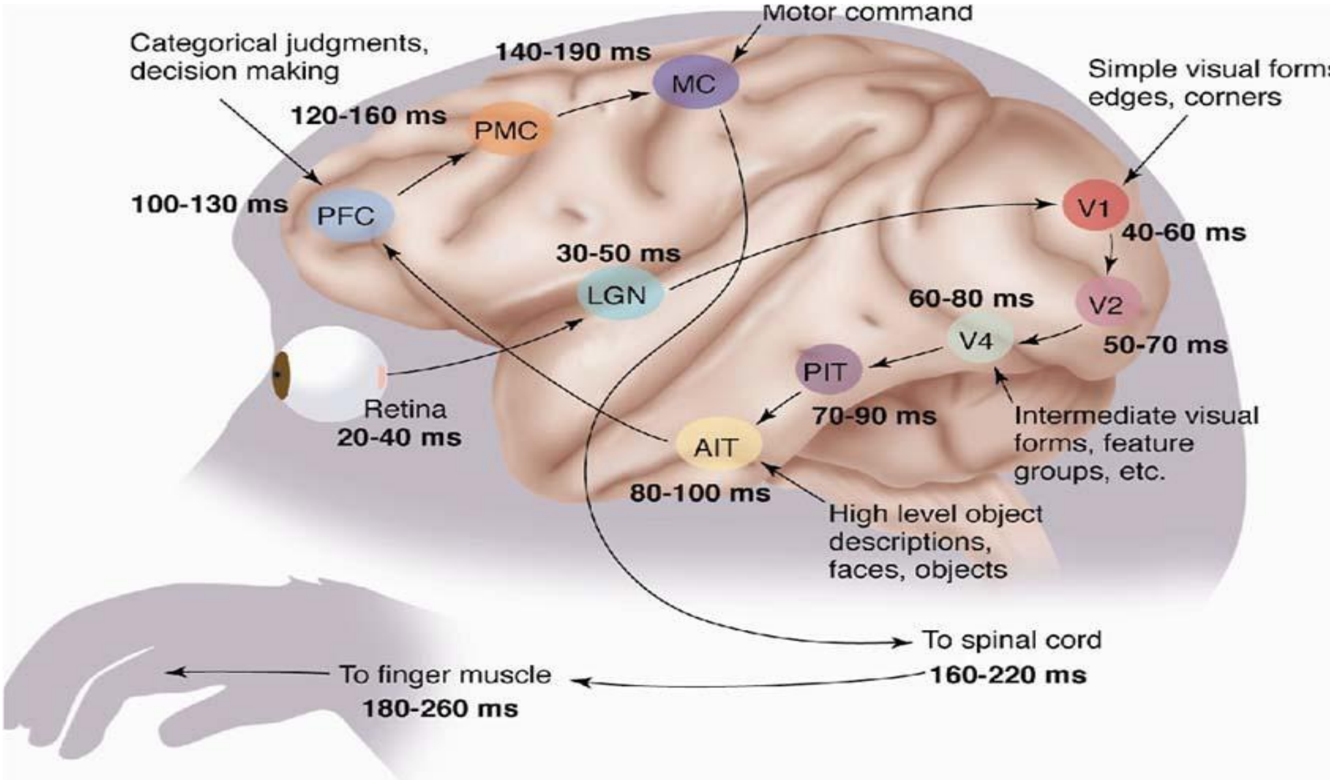
color
texture
pictorial detail
shape
size

Object processing

Ventral or “what” stream

Human Vision

- The ventral (recognition) pathway in the visual cortex



[picture from Simon Thorpe]

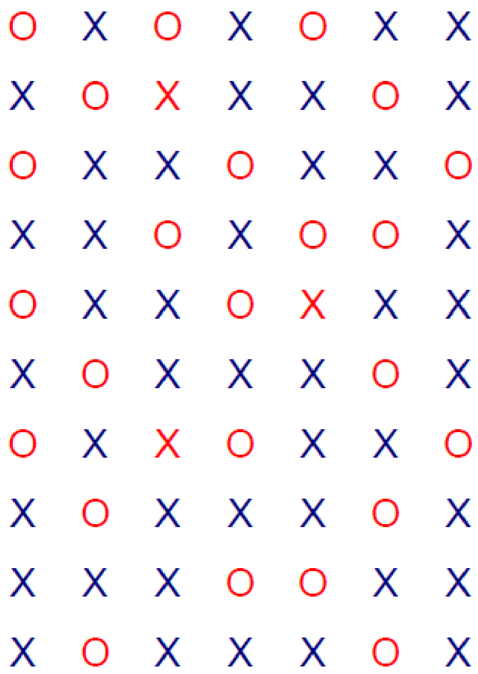
Human Vision

Human Vision

- Count the red crosses

Human Vision

- Count the red crosses



Human Vision

Human Vision

Human Vision

- Computer vision is much more difficult than you might initially think:
  - Your brain needs 25% of the cortex just to solve vision
  - The general computer vision problem is still largely unsolved

- Main problem: vision needs to deal with enormous variations in the signal
  - Some of these variations are relevant and others should be ignored

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

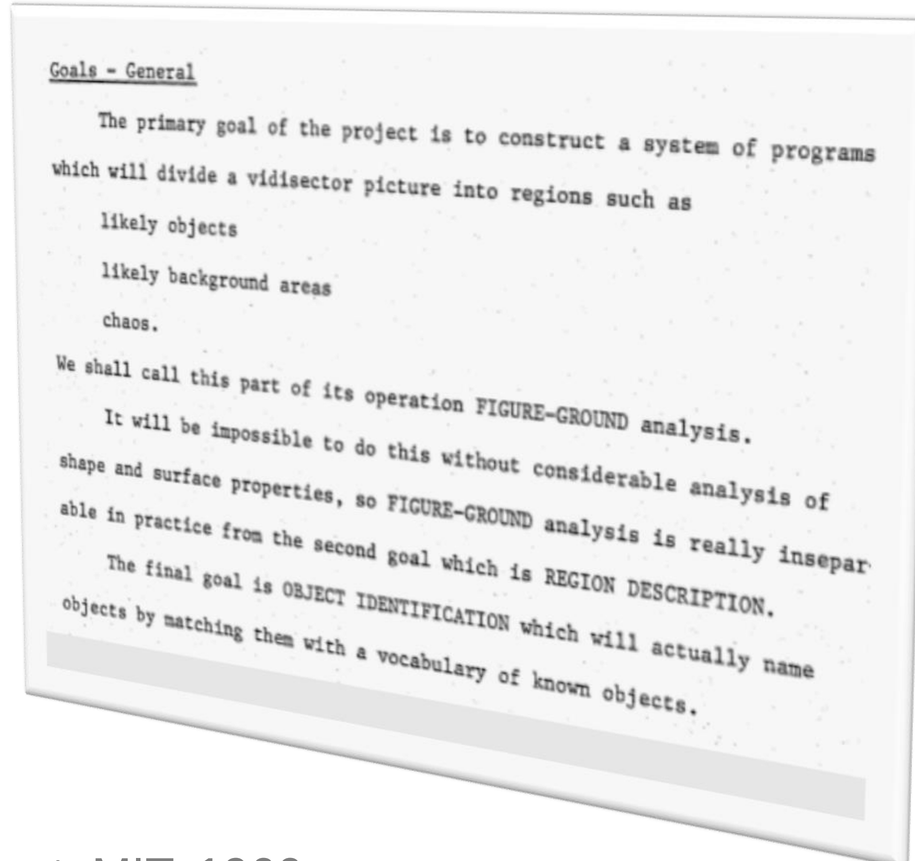
Source: S. Narasimhan

Computer Vision

According to Hollywood

Computer Vision: History

- An MIT undergraduate summer project*, in 1966, aimed to solve background/foreground segmentation and object detection/classification.
- It has been 54 year and we still work on the same problems.


Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

likely objects

likely background areas

chaos.

We shall call this part of its operation FIGURE-GROUND analysis.

It will be impossible to do this without considerable analysis of shape and surface properties, so FIGURE-GROUND analysis is really insepar able in practice from the second goal which is REGION DESCRIPTION.

The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects.

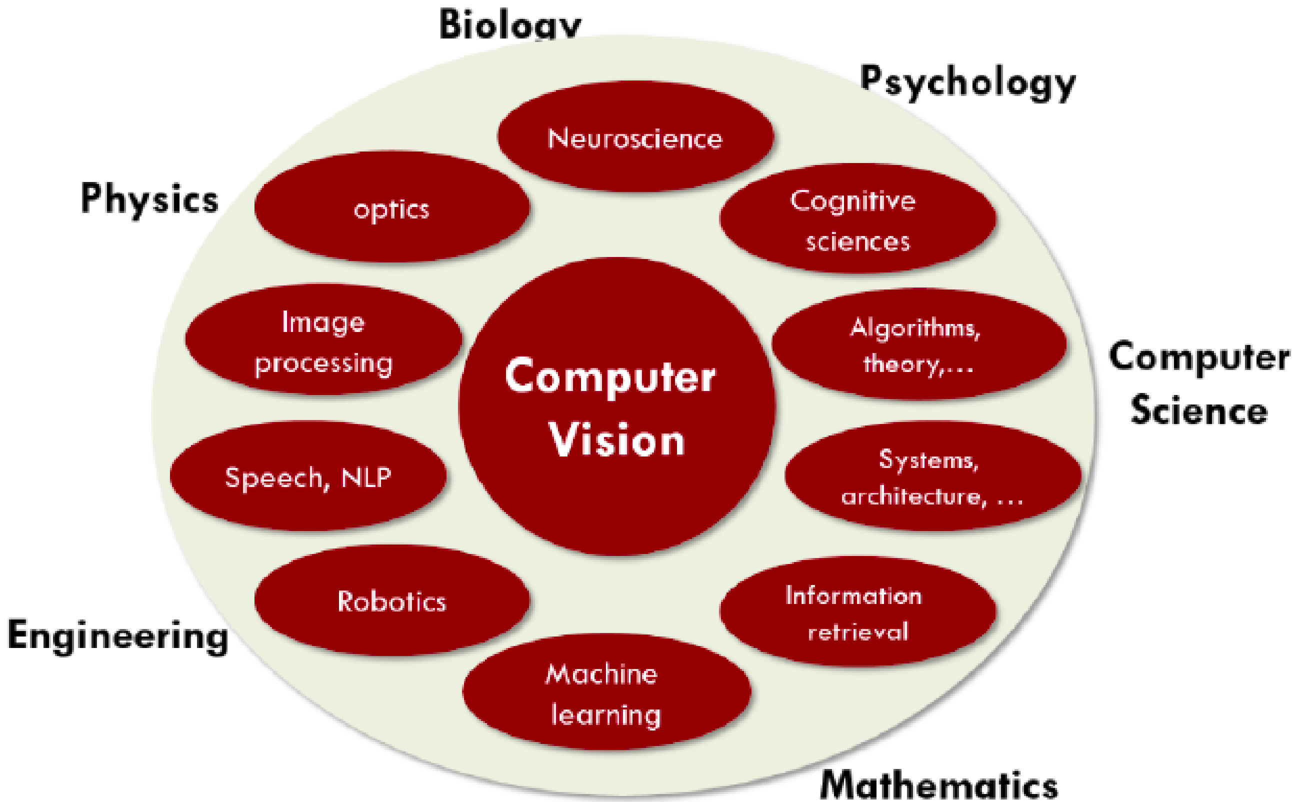
*Seymour A Papert. The summer vision project, MIT, 1966.

Computer Vision

- Making useful decisions about real physical objects and scenes based on images (Shapiro & Stockman, 2001)
- Extracting descriptions of the world from pictures or sequences of pictures (Forsyth & Ponce, 2003)
- Analyzing images and producing descriptions that can be used to interact with the environment (Horn, 1986)
- Designing representations and algorithms for relating images to models of the world (Ballard & Brown, 1982)

Slide Credit: Rajesh Rao

Computer Vision



Slide Credit: Olivier Moindrot

Computer Vision

- How do we describe the variations within the *class* “chair”?

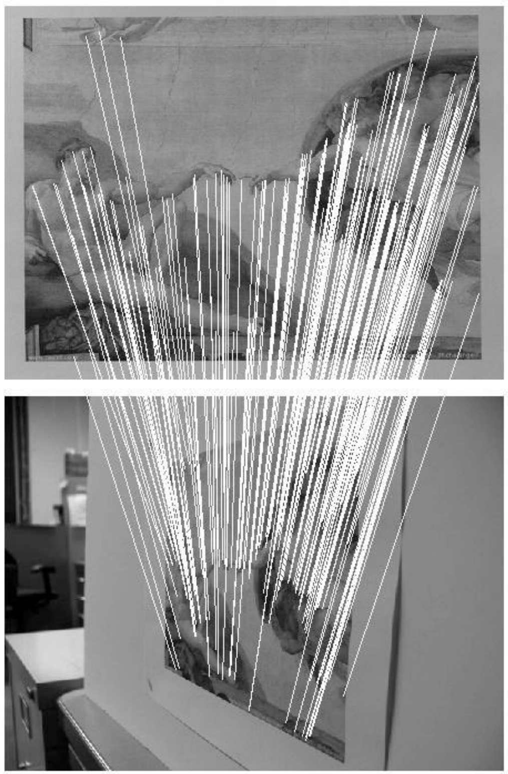
- Invariance to some variations can be obtained using hand-crafted models
- We generally try to *learn* invariance to the remaining variations from *examples*

Object Recognition

- Observation: chairs contain relatively lots of *edges*

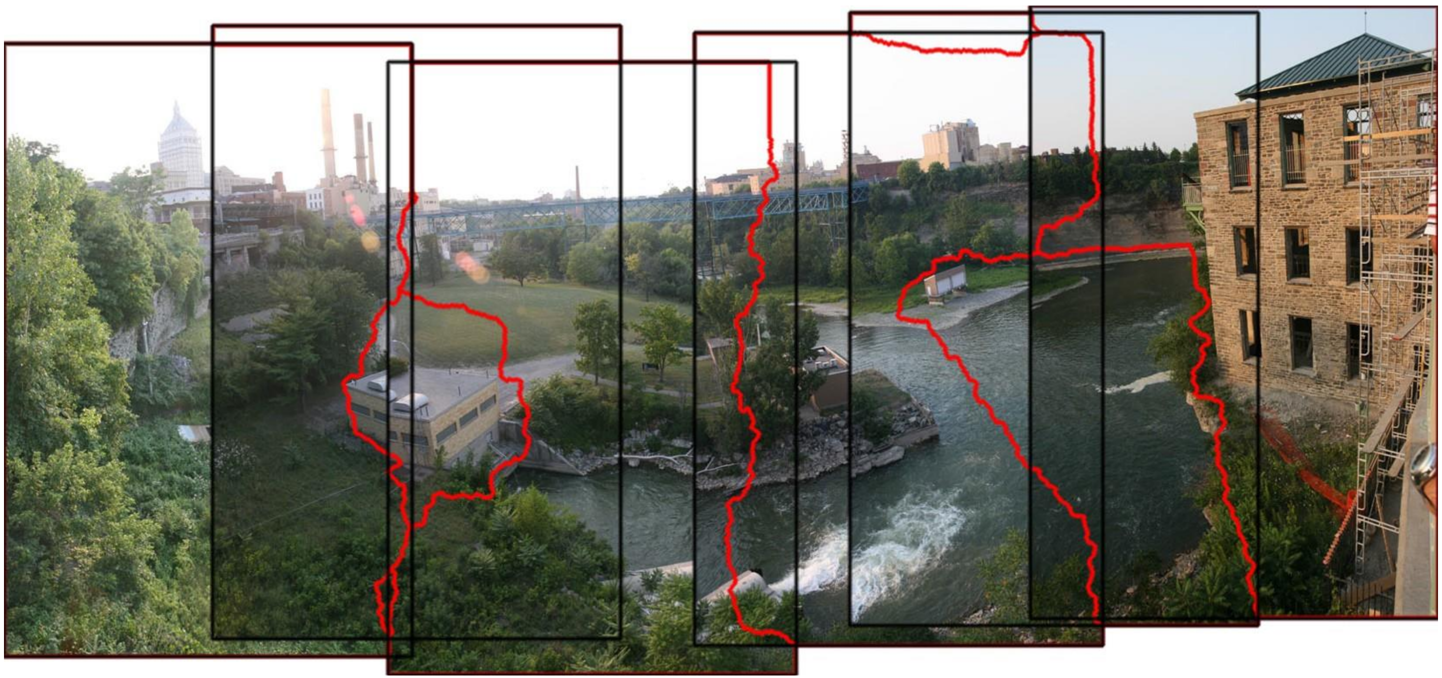
What will you learn in this course?

- Feature point detection and matching:



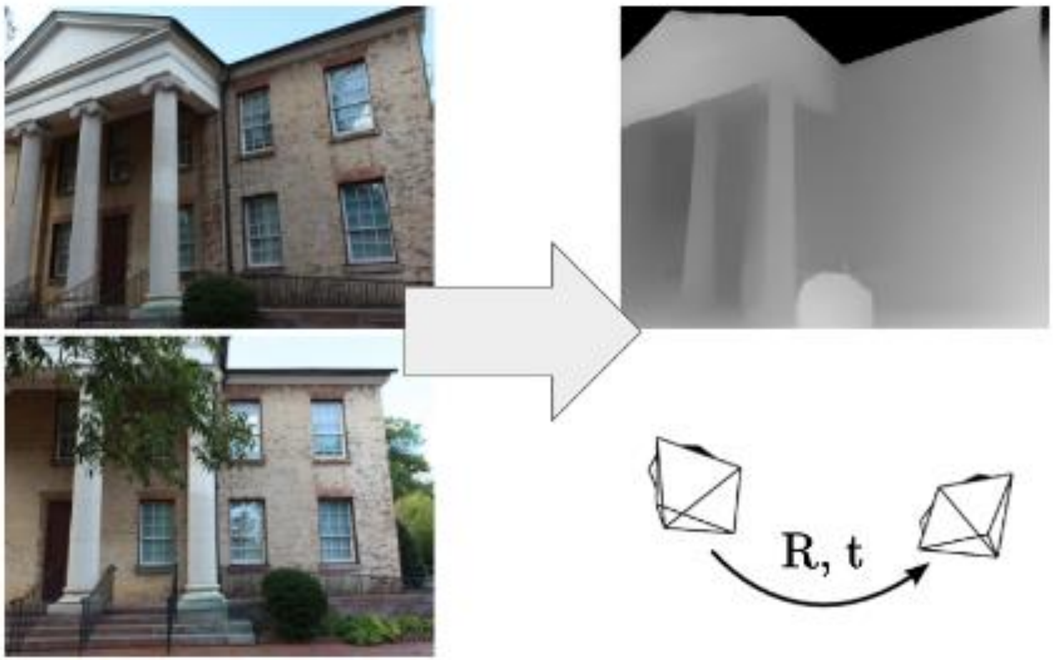
What will you learn in this course?

- Automatic image stitching:



What will you learn in this course?

- Stereo Vision & 3D Reconstruction:

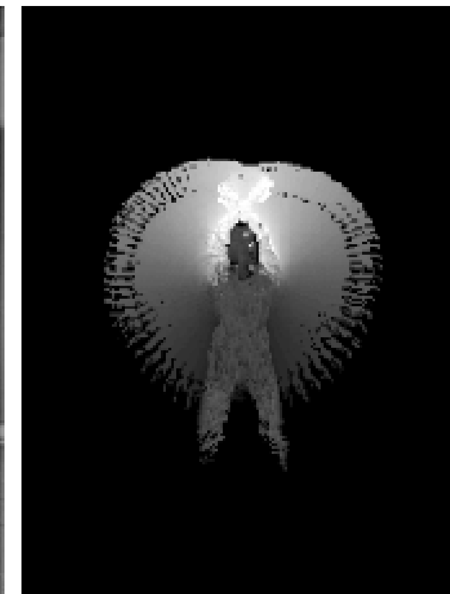


What will you learn in this course?

- Object detection and recognition:

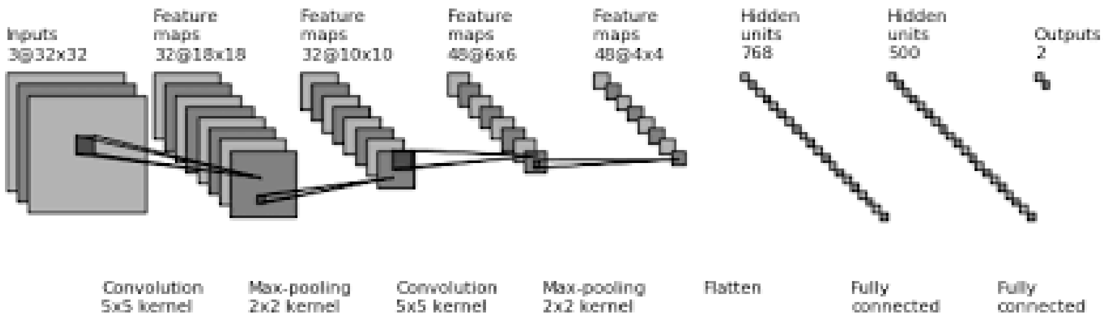
What will you learn in this course?

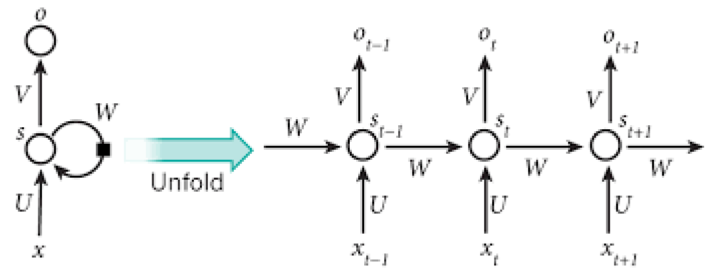
- Motion and Flow Analysis:



What will you learn in this course?

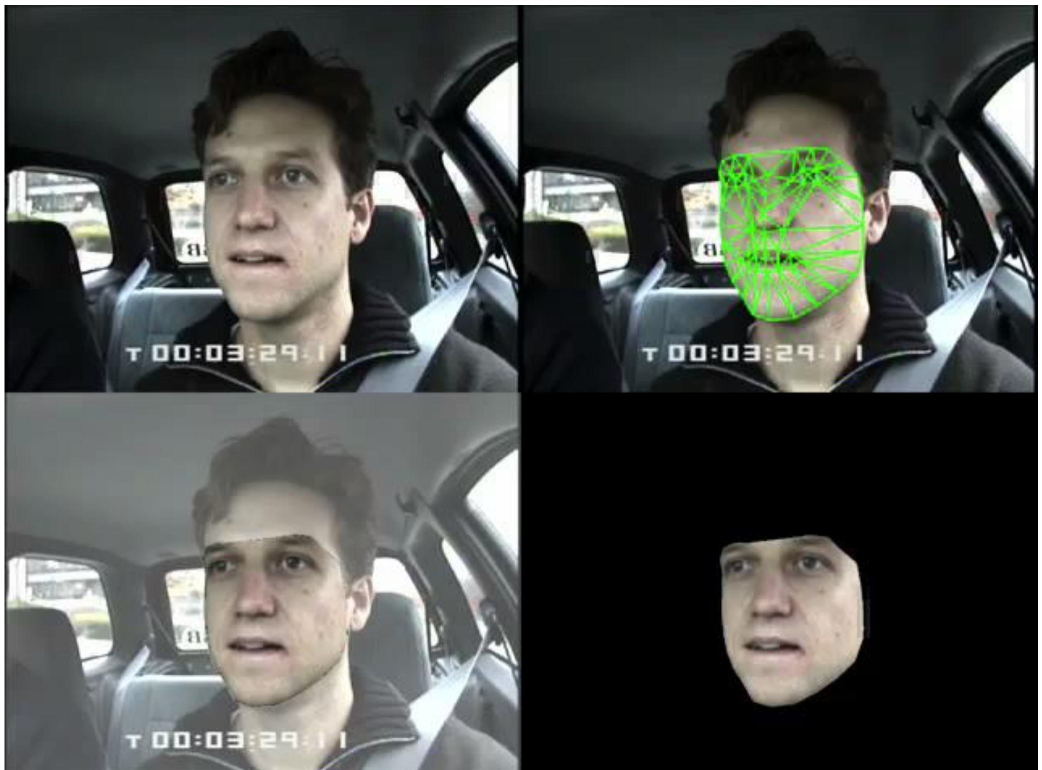
- Basics of Convolutional and Recurrent Models:





What will you learn in this course?

- Face Processing:

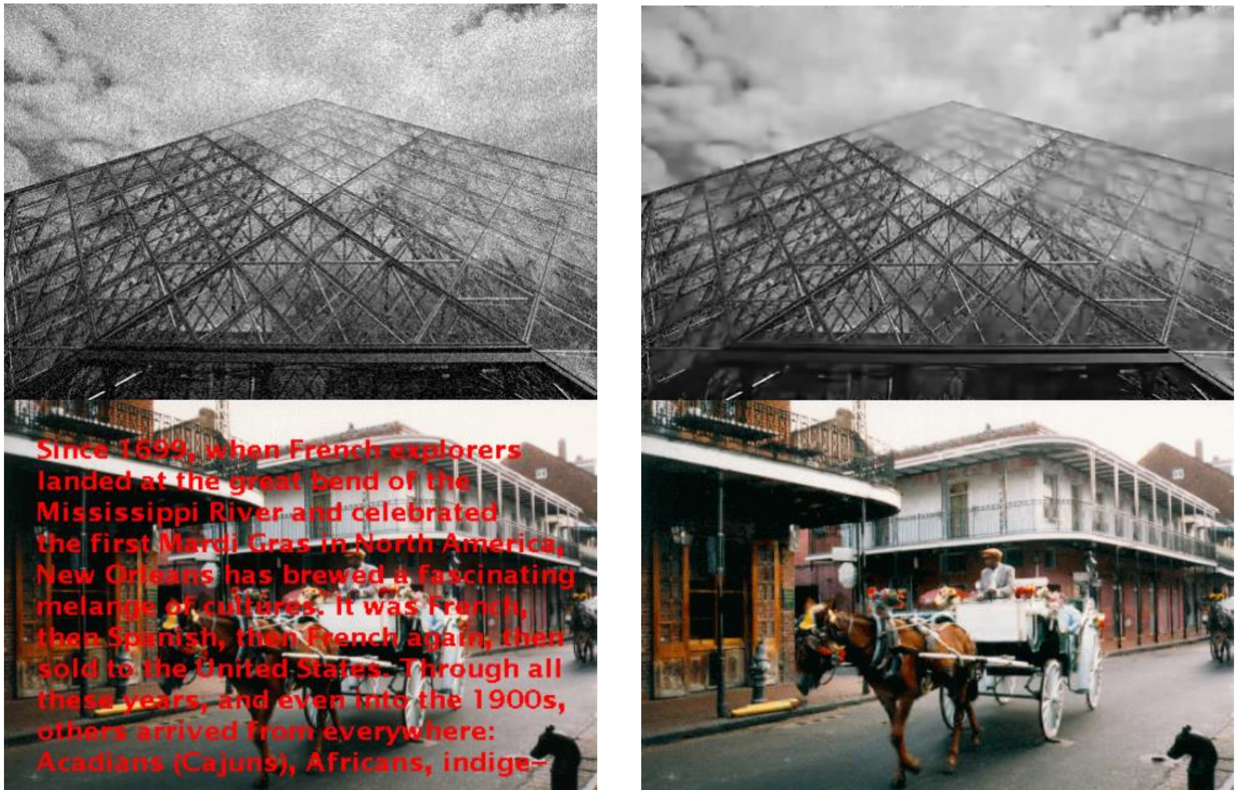


What will you learn in this course?

- Tracking:

What will you learn in this course?

- Image denoising and inpainting:



What will you learn in this course?

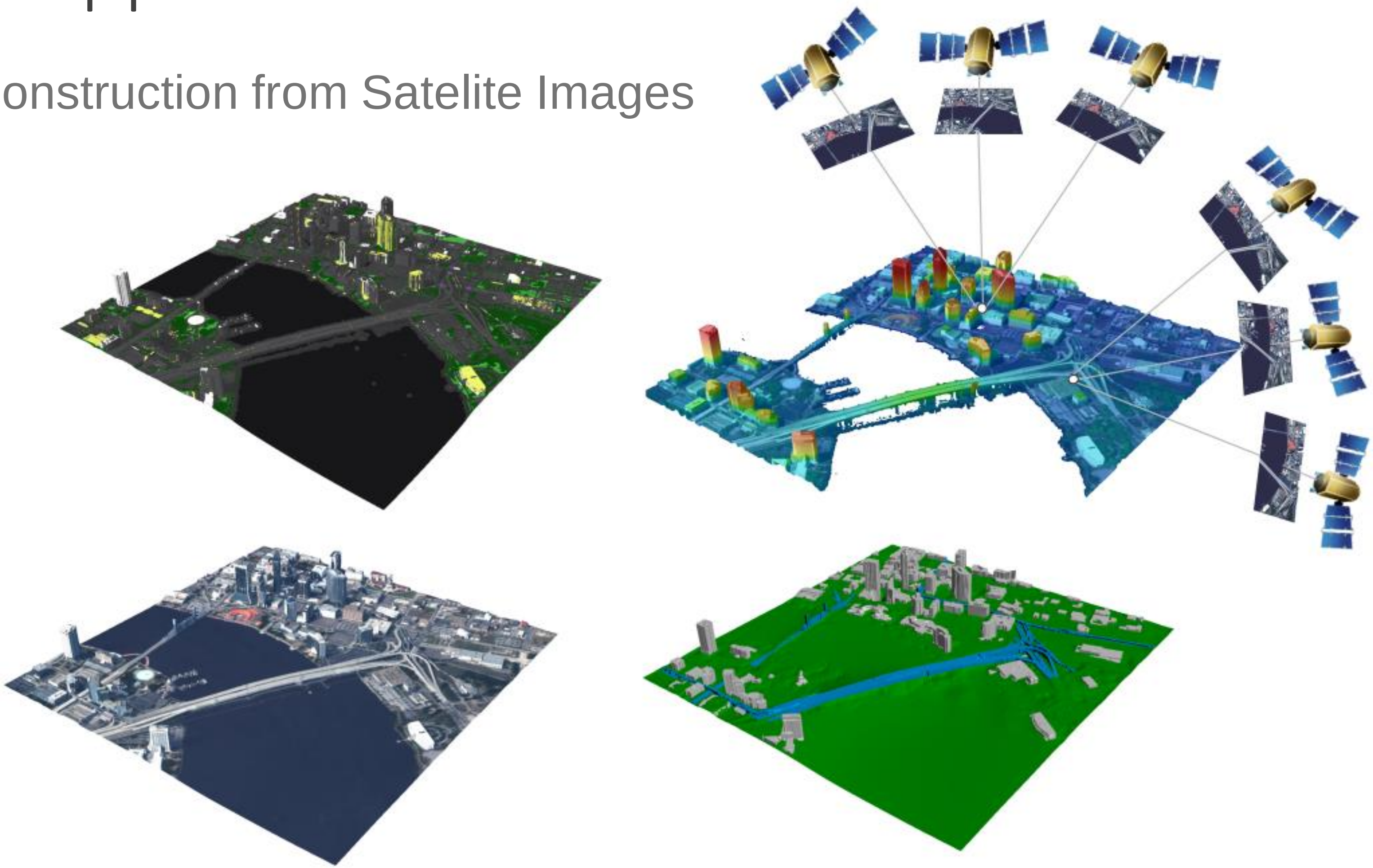
- Action/activity recognition:

Real-life Applications

- Face Detection / Smile Shutter:

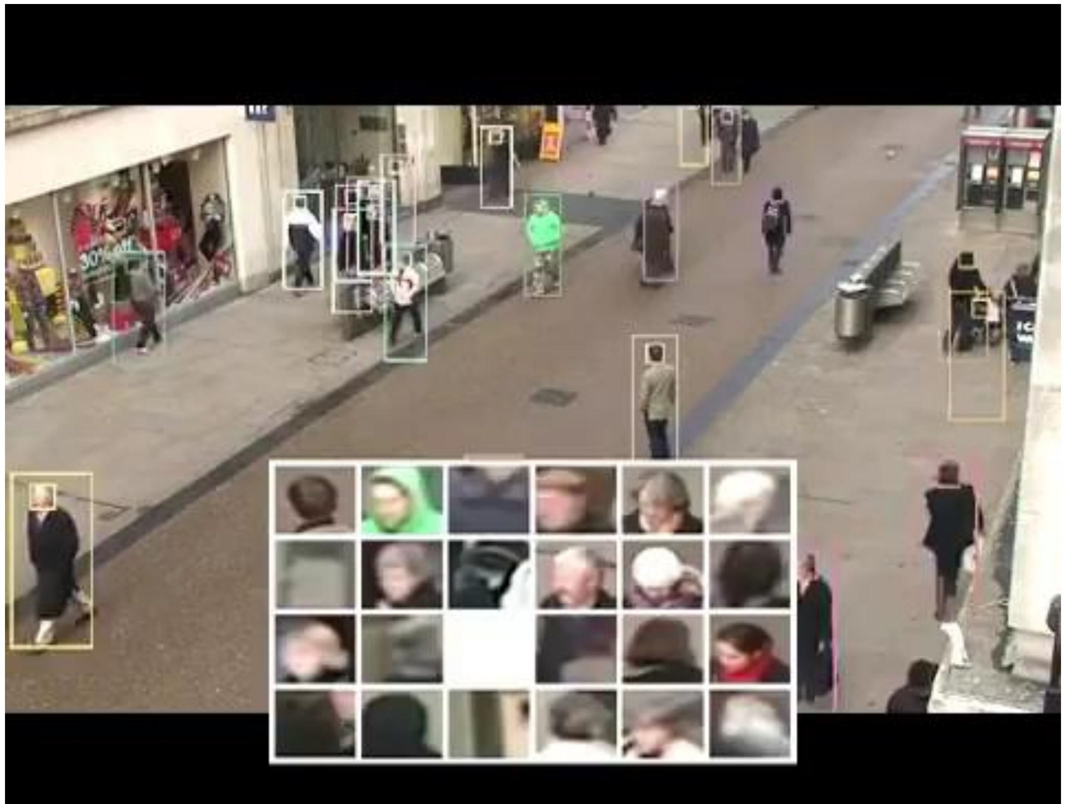
Real-life Applications

- 3D Reconstruction from Satelite Images



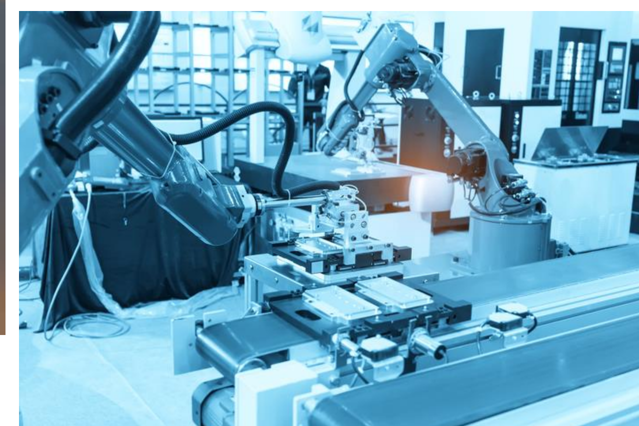
Real-life Applications

- Surveillance



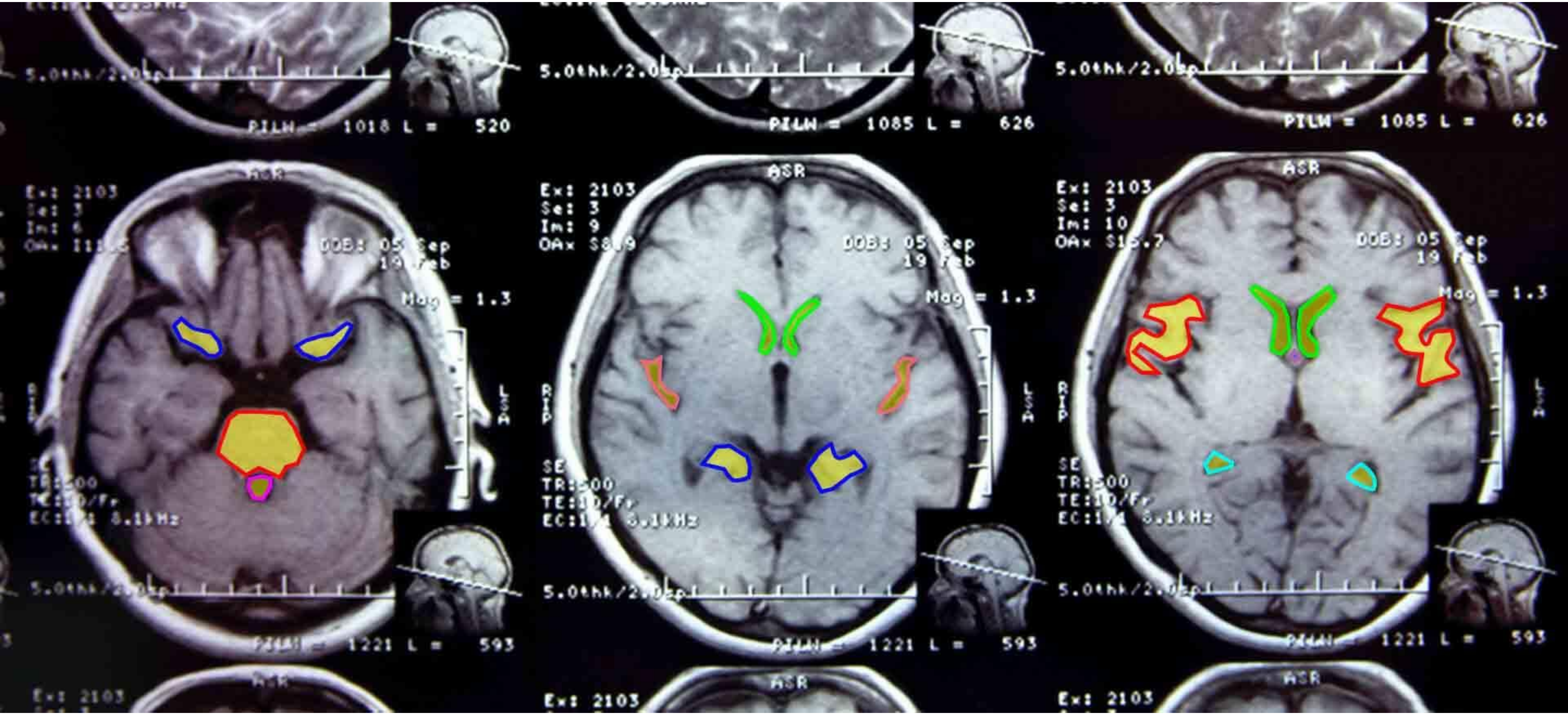
Real-life Applications

- Robotics



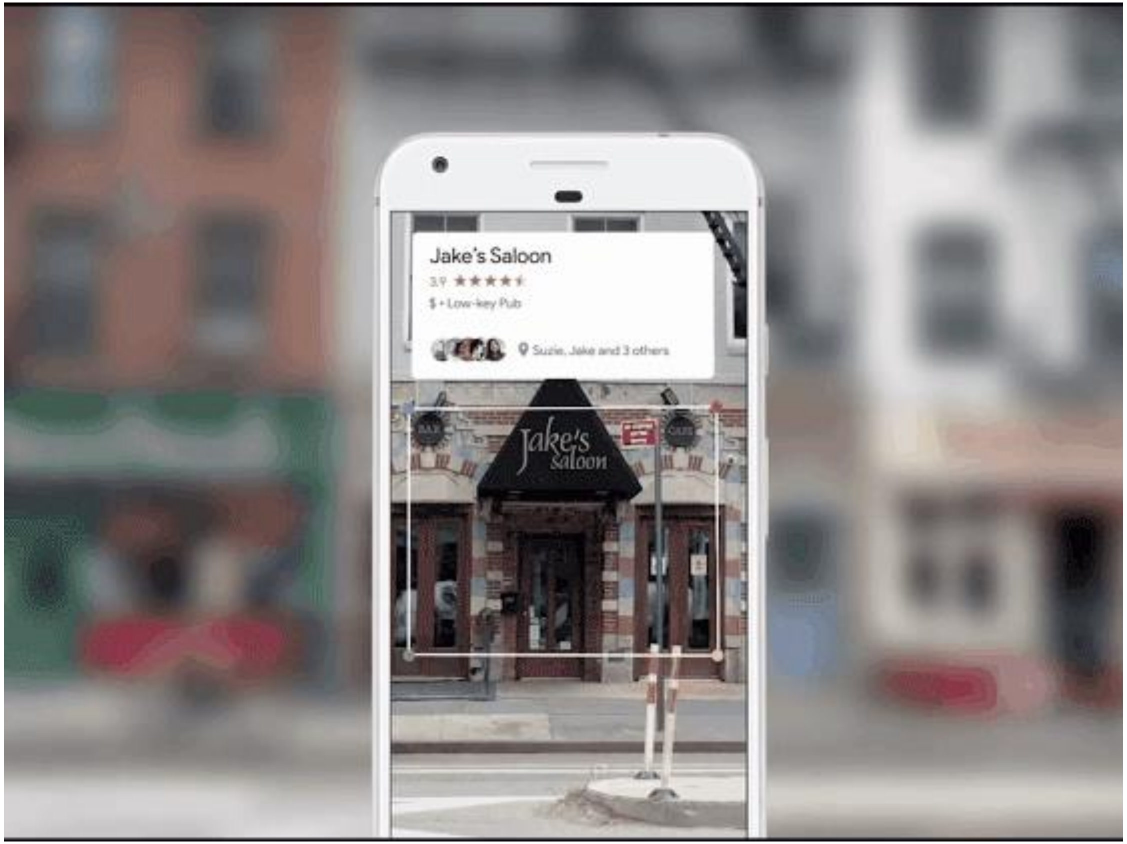
Real-life Applications

- Medical Image Analysis



Real-life Applications

- Visual Search (https://lens.google.com/)



Real-life Applications

- Gaming / Human Computer Interaction

Real-life Applications

- Special Effects

Real-life Applications

- Sports Analytics

Real-life Applications

- Autonomous vehicles:

What have we learned so far?

- What we perceive visually is largely a construction of the brain
- Vision is difficult...
- ...but lots of fun!

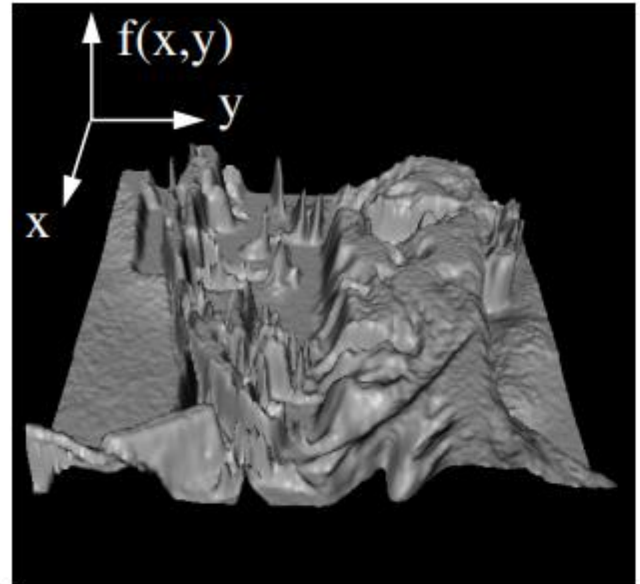
Basics: What is an image?

- Assume an image as a function, f, from $\mathbb{R}^2$ to $\mathbb{R}$:
  - $f(x, y)$ gives the intensity at position (x, y)
  - Realistically, an image is defined over a rectange:
    - $f: [a, b] \times [c, d] \rightarrow [0,1]$

- Color image = Three functions combined together:
  - $f(x, y) = \begin{bmatrix} r(x, y) \\ g(x, y) \\ b(x, y) \end{bmatrix}$

Slide Credit: Rajesh Rao

Basics: An image as a function

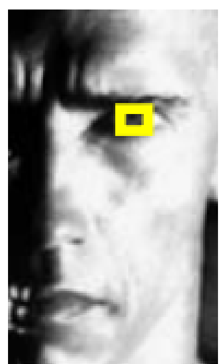


Bright regions are high, dark regions are low

Slide Credit: Rajesh Rao

Basics: Digital Images

- In computer vision we operate on digital (discrete) images:
  - Sample the 2D space on a regular grid
  - Quantize each sample (round to the nearest integer)
  - Each sample is a pixel (picture element)
  - If we assume each pixel as 1 byte, values range from 0 to 255



y

x

62	79	23	119	120	105	4	0
10	10	9	62	12	78	34	0
10	58	197	46	46	0	0	48
176	135	5	188	191	68	0	49
2	1	1	29	26	37	0	77
0	89	144	147	187	102	62	208
255	252	0	166	123	62	0	31
166	63	127	17	1	0	99	30

Slide Credit: Rajesh Rao

Basics: Image processing

- Range transformation (pixel processing):

$$g(x, y) = h\big(f(x, y)\big)$$

- Example?

Basics: Image processing

- Range transformation (pixel processing):

$$g(x, y) = h\big(f(x, y)\big)$$

- Example: Noise filtering

$\boldsymbol{f}$

$\xrightarrow{\boldsymbol{h}}$

$\boldsymbol{g}$

Basics: Image processing

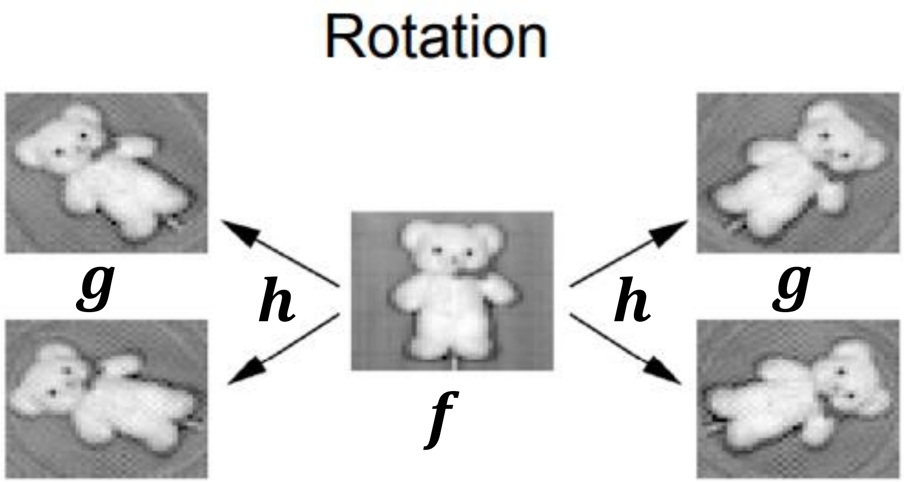
- Domain transformation (geometric transform):
$$g(x, y) = f\left(h_x(x, y), h_y(x, y)\right)$$
- Example: ?

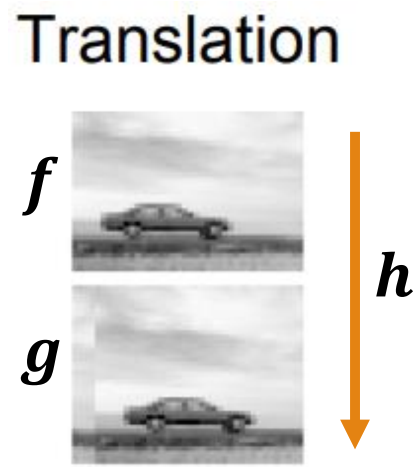
Basics: Image processing

- Domain transformation (geometric transform):

$$g(x, y) = f\left(h_x(x, y), h_y(x, y)\right)$$

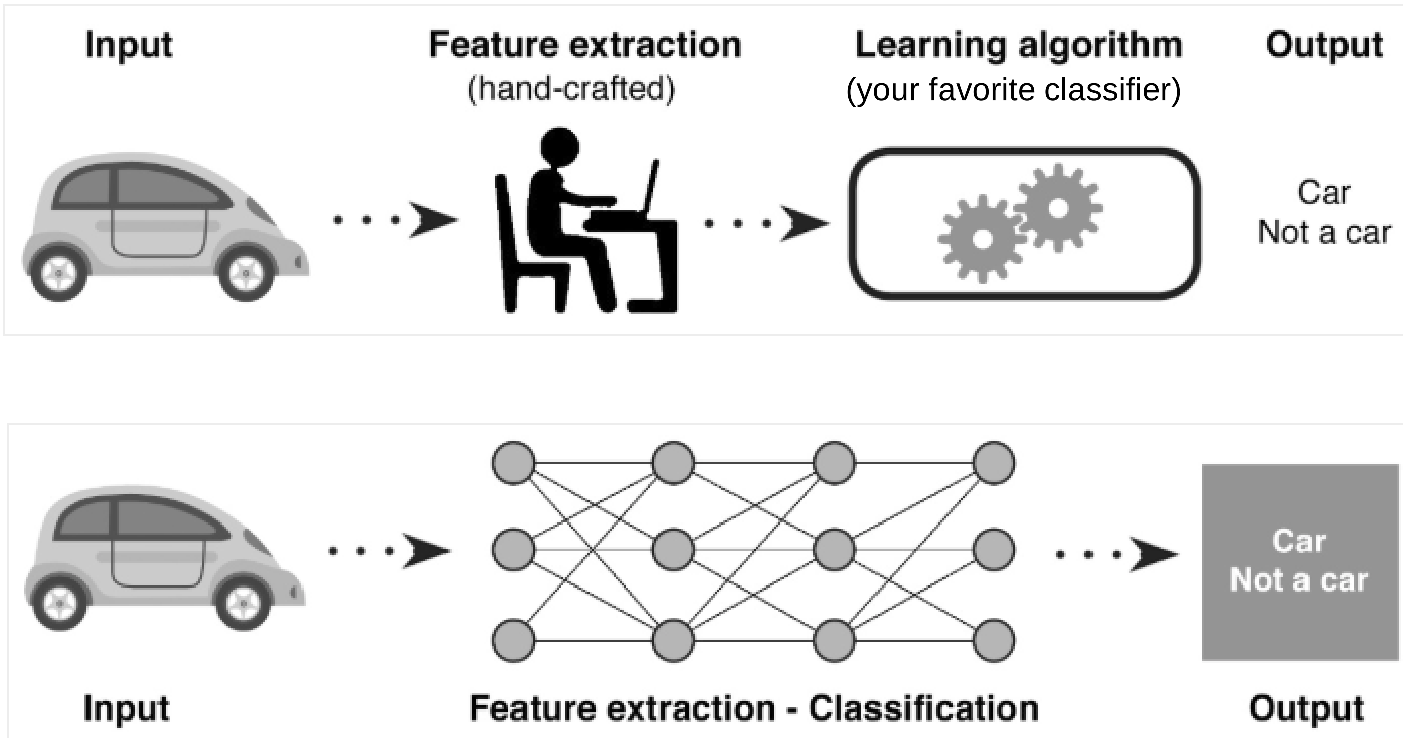
- Example: Rotation / Translation





Slide Credit: Rajesh Rao

Basics: Analysis Pipeline



Reading material: Section 1 of Szeliski