CSCI 631
Foundations of Computer Vision

Ifeoma Nwogu

ion@cs.rit.edu

Today

- Logistics
  - Staff, Labs, prerequisites, textbooks, grading etc
- Course Organization
- What is computer vision?
- Why is vision so hard?
- Course goals

Course Staff

- Instructor: Ifeoma Nwogu
  - Office hours:
    - Tues and Thurs
    - 2:20-3:20pm
    - 70-3669
    - also by appointment as needed

- Teaching Assistant: Brij Ketan Shah

LABs

- Two hours of lab time each week
  - Saturdays mornings from 10am till noon (can change)
  - Focused on related software installation and assistance with programming homework
  - TA’s office hours will be at the lab

Prerequisites

- This course is appropriate for students with these essential prerequisites:
  - A good working knowledge of Python, especially Numpy
  - Working knowledge of linear algebra
  - Some vector calculus
- The course does not assume prior imaging experience, computer vision, image processing, or graphics

Textbooks

- No expensive physical textbooks
- Tutorials
  - OpenCV Tutorials as https://docs.opencv.org/master/d9/df8/tutorial_root.html
  - Deep learning tutorials online, especially using Pytorch
- Research papers
  - Programming homework will sometimes be accompanied by related past research papers

Modules

1. First module of the course will be classical methods of computer vision. These involve

   – Directly manipulating of pixels in images (image processing)

   – Extracting hand-engineered features; and

   – Applying basic machine learning techniques for solving problems in computer vision; e.g. classification with SVM, regression, segmentation via clustering features, etc.

Modules cont'd

2. Second module will include the basic architectures of a neural network. These involve
   - Logistic regressor and multi-level perceptron (MLP)
   - CNN for analyzing images
   - RNN and its extensions (LSTM and GRU) for analyzing sequences of images (videos)

- Midterm 1 will happen after Module 2

Modules cont'd

3. Module #3 deals with motion in vision. There are 3 sub-parts to this module:

   i. Estimating the direction and intensity of motion using *optical flow*

   ii. Analyzing motion using statistical models (hidden Markov models; *HMM*)

   iii. Analyzing motion using deep neural networks for sequences such as *LSTM* or *GRU*

   This is a very ambitious module!

Modules cont’d

4. The last module deals with advanced vision algorithms.

- Object detection and segmentation
- Deeper imaging architectures (ResNet; Inceptions; variants)
- We will look at Autoencoders and Transfer learning methods
- And review interpretability of these networks via class activation maps

Midterm 2 will happen after Module #4

Programming

The course will involve programming in Python, specifically the use of Numpy (+ supporting libraries for plotting, reading datasets, etc.) and pyTorch libraries for deep learning.

Grading

- There will be five components to the course grade
  - 5-7 programming assignments (35)
  - 4-5 reading assignments (10)
  - 2 mid-term exams (30)
  - Final conference paper presentation (10)
  - Final project result presentation (15)
  - Active positive participation (extra 5)
- Class participation is strongly encouraged to offset negative performance in any of the above
  - Extra credits will be given for active positive participation on myCourses discussion

Final presentation

- Teams of 2-3 encouraged
  - These are the same teams for the final project
- Detailed review of CVPR-type paper
- Presentation of the paper in class
  - Clearly articulate the relevance/benefits of the work
  - Demonstrate clear understanding of the technical aspect of the work
  - Critical reviews are strongly encouraged

Final project

- Significant implementation of a technique related to the course content
- Teams of 2-3 encouraged
- CVPR-type review article
- Two components :
  - proposal document (no more than 2 pages)
  - final write-up with results (no more than 6 pages CVPR style)

Academic Integrity

- What Is Academic Integrity?
  - https://www.rit.edu/academicaffairs/policiesmanual/d080
  - Pls review; ignorance is not the same as innocence
    - Plagiarism; cheating; duplicate submission for credit
    - Dr. HP Bischoff handles AI problems in the department
    - Deans Committee for Academic Integrity Violations

Mental Health Initiative

Success in this course depends heavily on your personal health and wellbeing. Recognize that stress is an expected part of the college experience, and it often can be compounded by unexpected setbacks or life changes outside the classroom. Your other instructors and I strongly encourage you to reframe challenges as an unavoidable pathway to success. Reflect on your role in taking care of yourself throughout the term, before the demands of exams and projects reach their peak. Please feel free to reach out to me about any difficulty you may be having that may impact your performance in this course as soon as it occurs and before it becomes unmanageable. In addition to your academic advisor, I strongly encourage you to contact the many other support services on campus that stand ready to assist you.

Course goals

1. Be aware of the value and ethical problems of computer vision to society
2. Understand CV pre-deep learning methods and when old tricks might come in handy
3. Understand the basics of deep learning from first principles
4. Obtain a level of comfort implementing advanced deep learning algorithms for vision
5. Understand how to analyze moving data

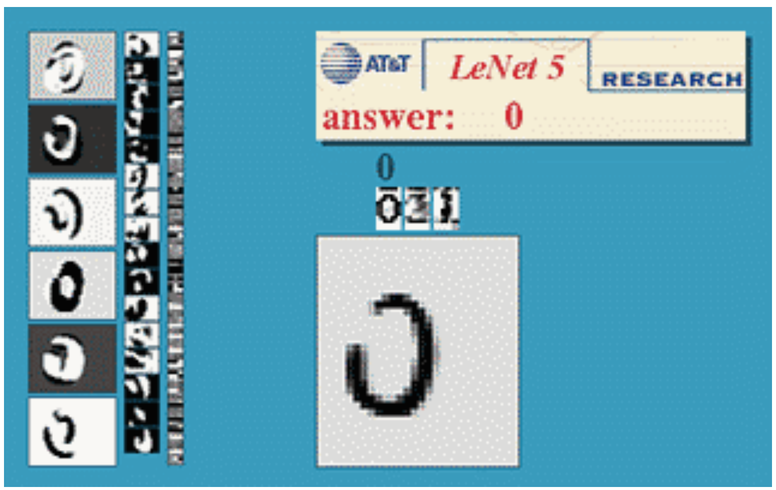
Examples of mainstream vision applications

The next slides show some examples of what current vision systems can do

Optical character recognition (OCR)

Technology to convert scanned docs to text

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



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

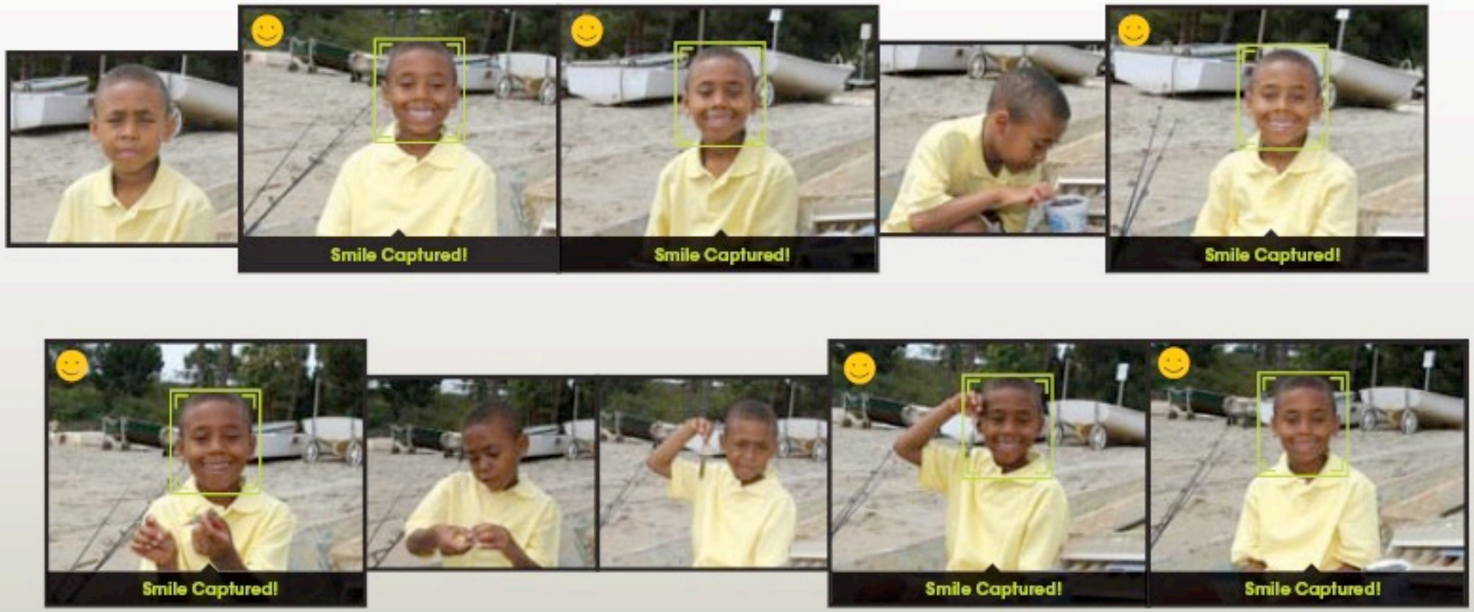
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face detection and recognition

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.



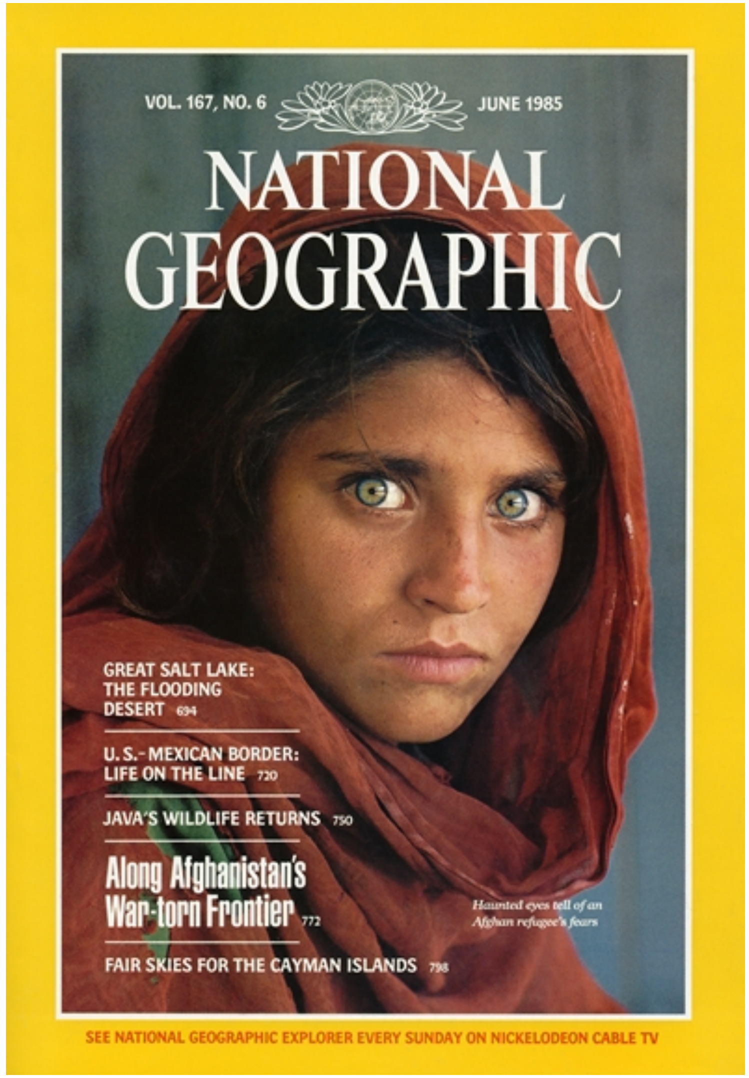
[Sony Cyber-shot® T70 Digital Still Camera](#)

Object recognition (in supermarkets)

LaneHawk by EvolutionRobotics

"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… "

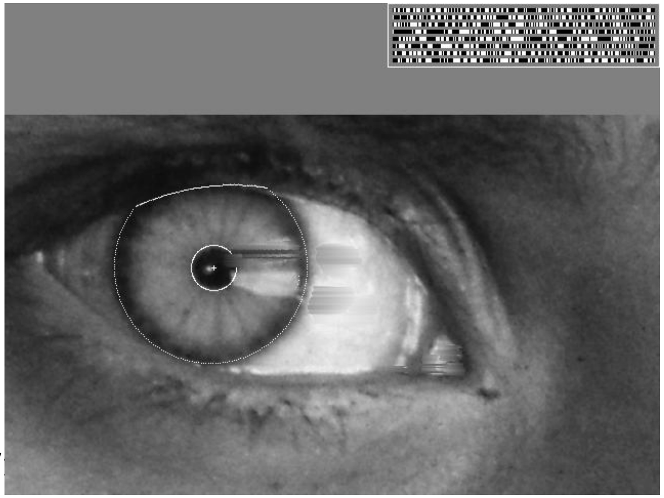
Face recognition

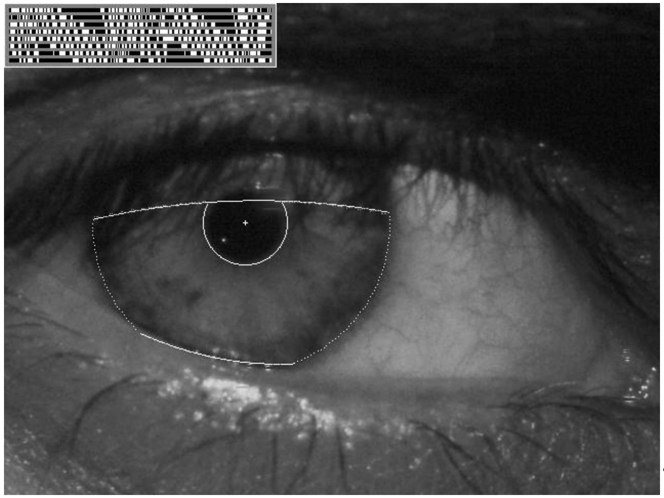


Who is she?

Vision-based biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the [story]





Login without a password…

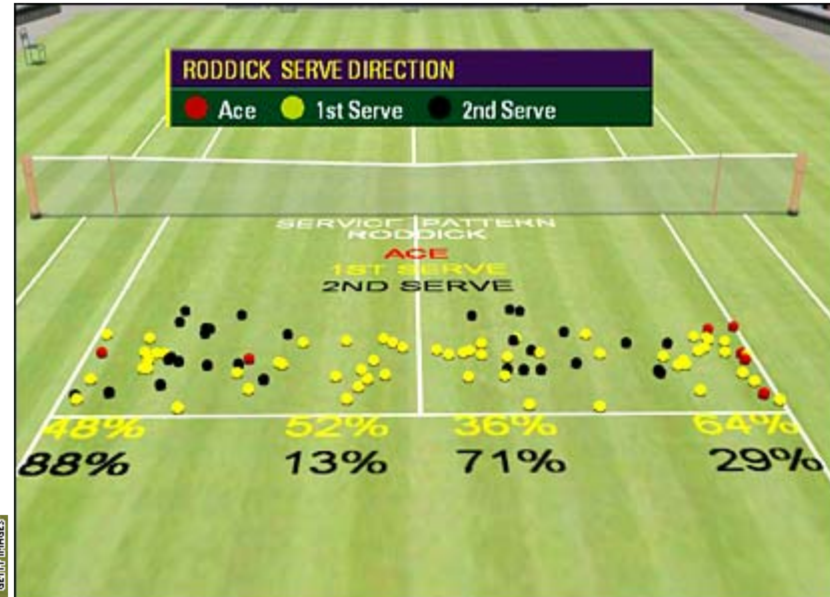
Fingerprint scanners on
many new laptops,
other devices

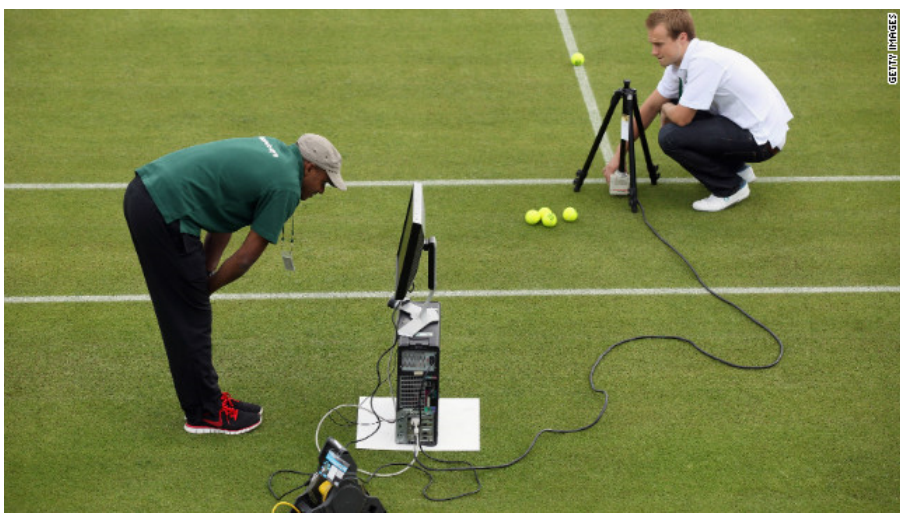
Face recognition systems now more
widely used
http://www.sensiblevision.com/

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

Sports





Brief explanation on how hawk-eye works can be found [here]()

Smart cars

Tesla Autopilot system is considered to be an SAE level 2 system

Level 0: Automated system issues warnings but no sustained vehicle control.

Level 1 ("hands on"): Driver and the automated system share control of the vehicle. e.g. Adaptive Cruise Control

Level 2 ("hands off"): Automated system takes full control (accelerating, braking, and steering). Driver must monitor the driving and be prepared to intervene at any time

Level 3 ("eyes off"): The driver can safely turn their attention away from the driving tasks, e.g. the driver can text or watch a movie. Vehicle will handle situations like emergency braking.

The 2018 Audi A8 Luxury Sedan was the first commercial car to claim to be capable of level 3 self driving.

Level 4 ("mind off"): As in level 3, but no driver attention is ever required for safety, i.e. the driver may safely go to sleep or leave the driver's seat. Self driving is supported only in limited spatial areas (geofenced) or under special circumstances, like traffic jams.

Level 5 ("steering wheel optional"): No human intervention is required at all. An example would be a robotic taxi.

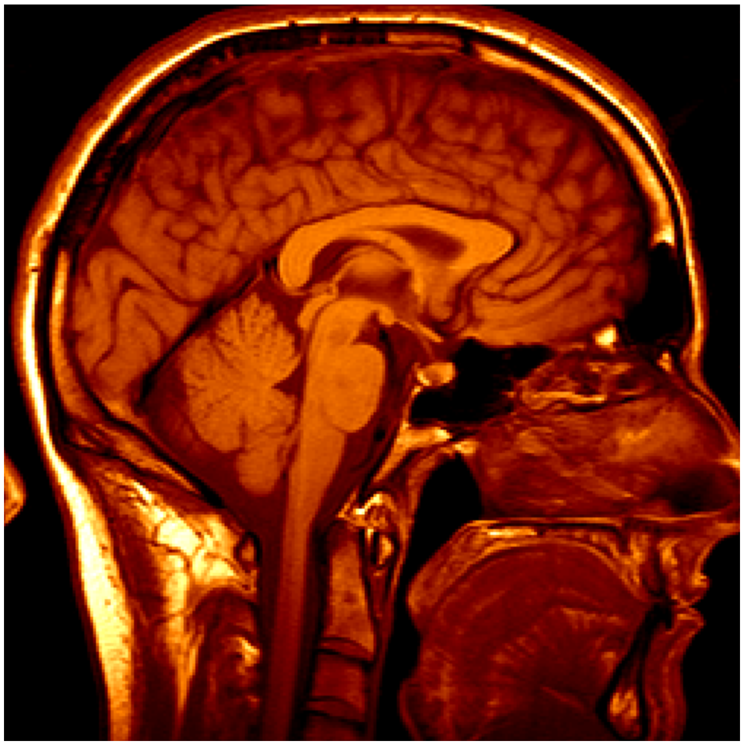
Vision-based interaction (and games)

Nintendo Wii has camera-based IR tracking built in. See Lee's work at CMU on clever tricks on using it to create a multi-touch display!

Digimask: put your face on a 3D avatar.

"Game turns moviegoers into Human Joysticks", CNET
Camera tracking a crowd, based on this work.

Medical imaging



3D imaging
MRI, CT

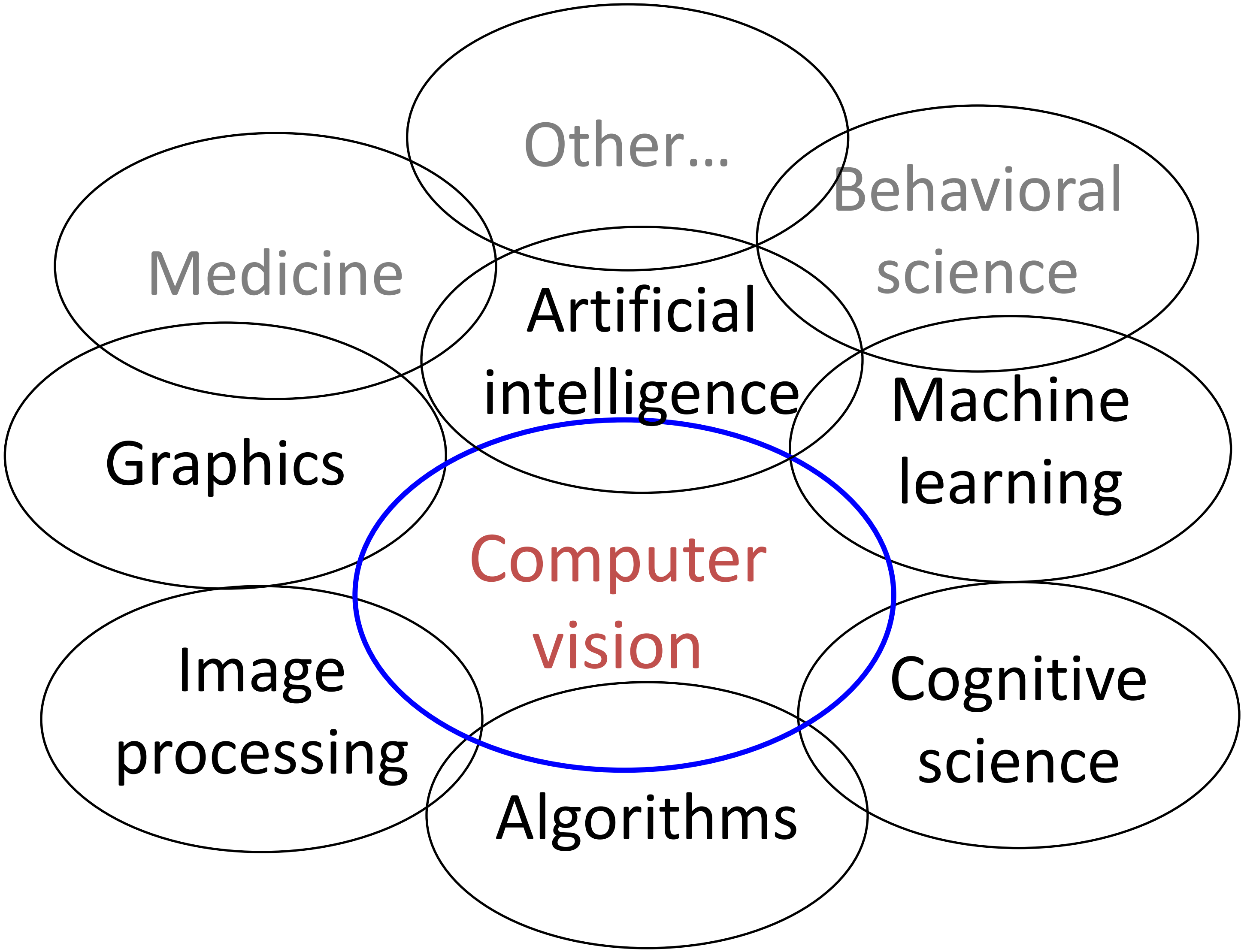
Image guided surgery
[Grimson et al., MIT]

More recent applications

- Many more applications of advanced vision algorithms making their way to the public:
  - Video-conferencing-based communication coach
  - Face-to-face interaction quality, e.g. between therapist and client; student and mentor; etc
  - Automated warehouses (Amazon robots, etc)
  - Robot knife for surgery
  - Cassava leaf health monitoring
  - So may more….

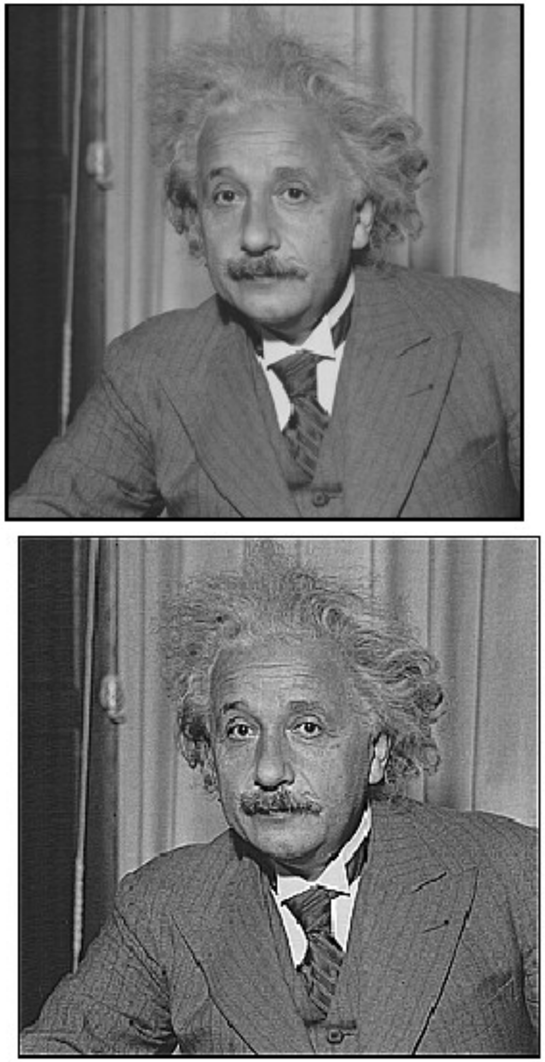
 8/26/19

Related disciplines



Linear Filtering

Image Filters



Smooth/Sharpen Images... ***Find edges...*** ***Find waldo...***

Images as functions

- We can think of an image as a function, f, from R^2 *to* R:
  - $f(x, y)$ gives the intensity at position (x, y)
  - Realistically, we expect the image only to be defined over a rectangle, with a finite range:
    - f: $[a,b] \times [c,d] \rightarrow [0, 1.0]$

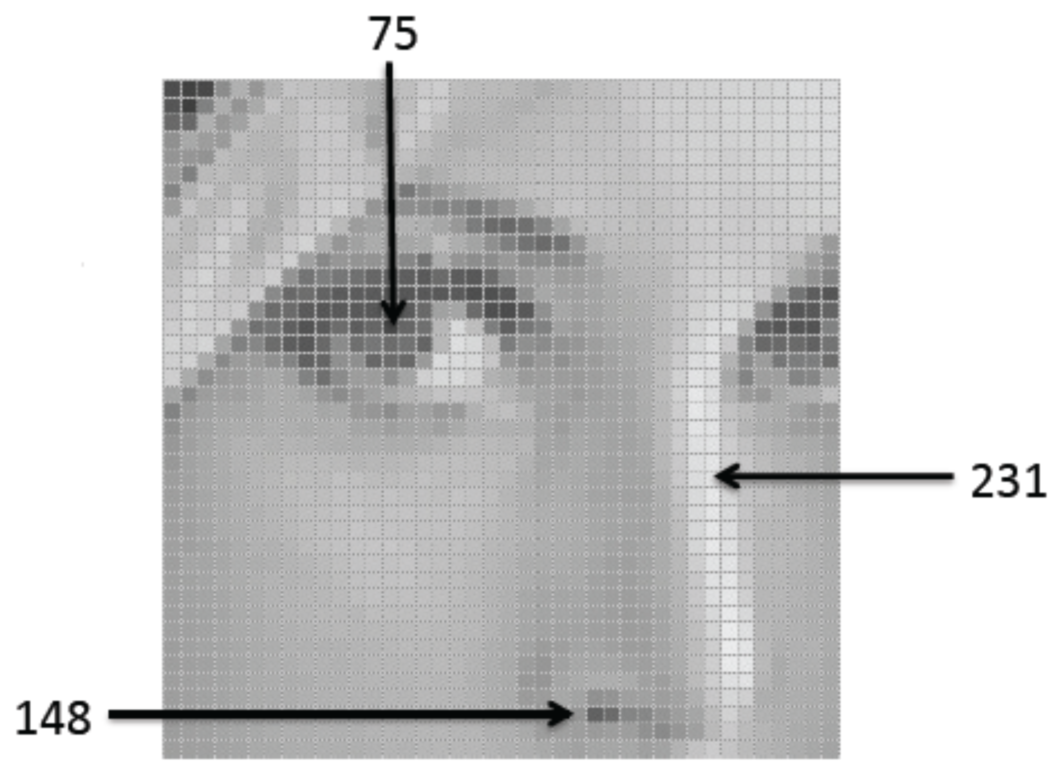
- A color image is just three functions pasted together. We can write this as a “vector-valued” function:

$$f(x,y) = \begin{bmatrix} r(x,y) \\ g(x,y) \\ b(x,y) \end{bmatrix}$$

Source: S. Seitz

Images as functions

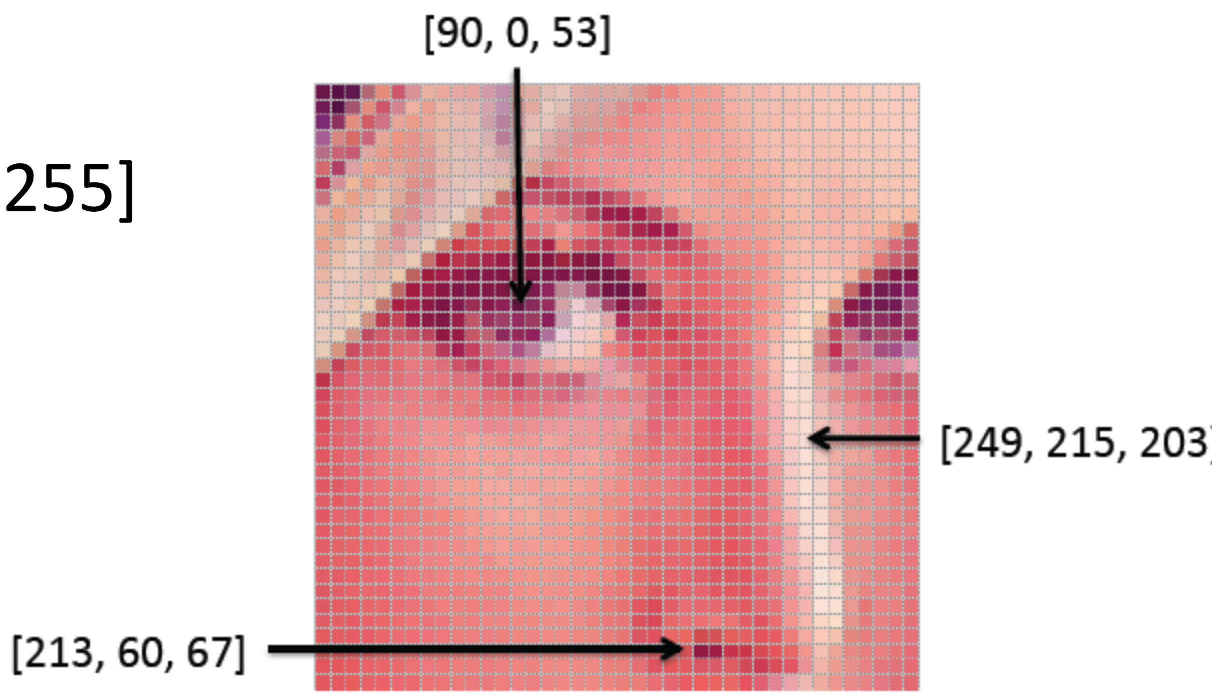
- An image contains discrete number of pixels, e.g.
  - "grayscale" (or "intensity"): [0,255]



Source: F.F Lei

Images as functions

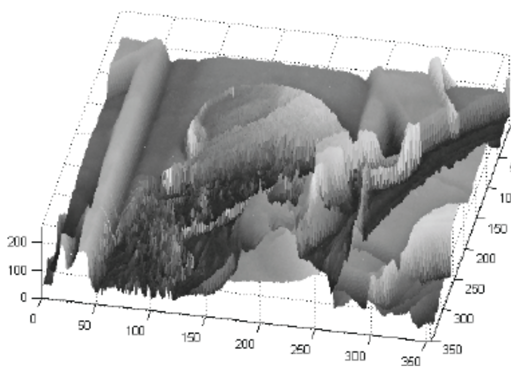
- An image contains discrete number of pixels, e.g.
  - "grayscale" (or "intensity"): [0,255]
  - Color
    - RGB: [R, G, B]
    - Lab: [L, a, b]
    - HSV: [H, S, V]



Source: F.F Lei

Images as discrete functions

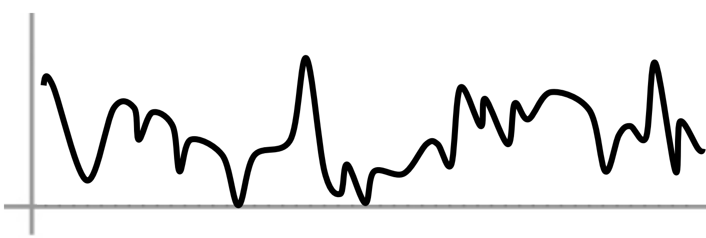
- In computer vision we operate on **digital (discrete)** images:
  - **Sample** the 2D space on a regular grid
  - **Quantize** each sample (round to nearest integer)
- Image thus represented as a matrix of integer values.

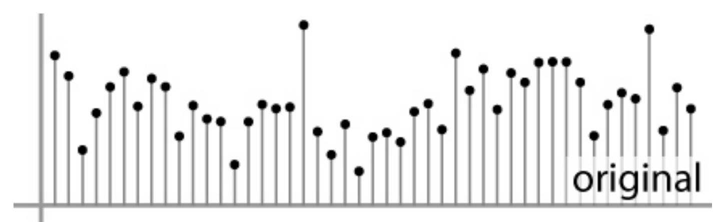


$j \rightarrow$, $i \downarrow$

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

2D





1D

Images as discrete functions

- Cartesian coordinates

$$f[n,m] = \begin{bmatrix} \ddots & & \vdots & & \\ & f[-1,1] & f[0,1] & f[1,1] & \\ \dots & f[-1,0] & \underline{f[0,0]} & f[1,0] & \dots \\ & f[-1,-1] & f[0,-1] & f[1,-1] & \\ & & \vdots & & \ddots \end{bmatrix}$$

Notation for discrete functions

Motivation

Filtering:

– Form a new image whose pixels are a combination original pixel values

Goals:

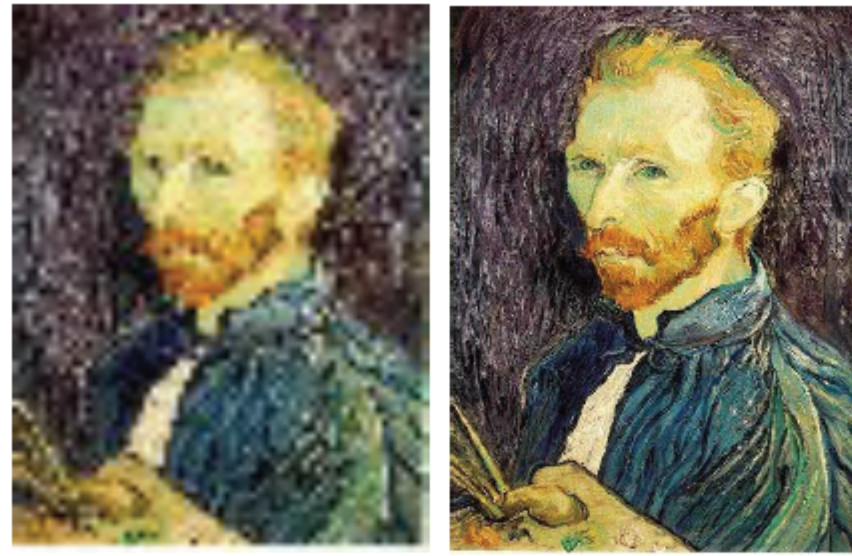
– Extract useful information from the images
  - Features (edges, corners, blobs…)

– Modify or enhance image properties:
  - super-resolution; in-painting; de-noising

De-noising

Salt and pepper noise

Super-resolution



In-painting

Bertamio et al

Source: F.F. Li

Questions