Lecture 1-1

- Prof. WANG Shiqi (shiqwang@cityu.edu.hk)

- Textbook

Computer Vision: Algorithms and Applications

© 2010 Richard Szeliski, Microsoft Research

http://szeliski.org/Book/

TA information:

Yingwen Zhang:
ywzhang26-c@my.cityu.edu.hk

Siyuan Xu:
siyuanxu9-c@my.cityu.edu.hk

Every image tells a story

Derailment

- "A picture is worth a thousand words"
- Goal of computer vision: perceive the “story” behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What happened?

Can the computer match human perception?

- Yes and no (mainly no)
  - computers can be better at “easy” things
  - humans are much better at “hard” things
- But huge progress has been made
  - Accelerating in the last few years due to deep learning
  - What is considered “hard” keeps changing

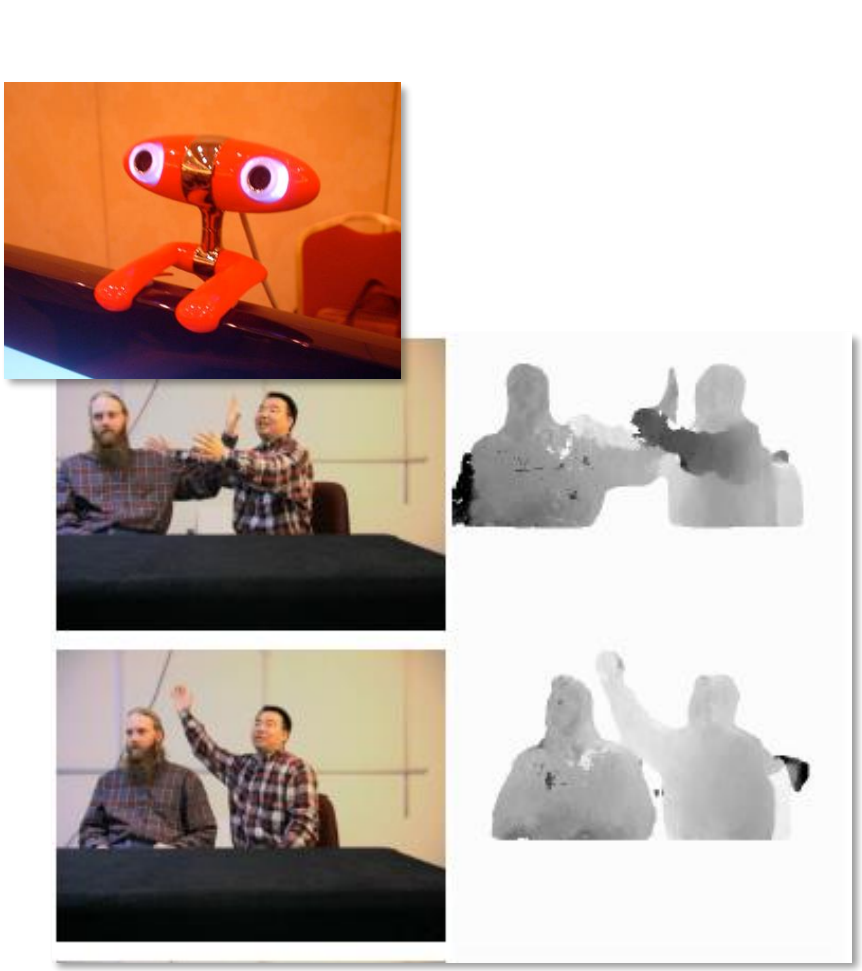
The goal of computer vision

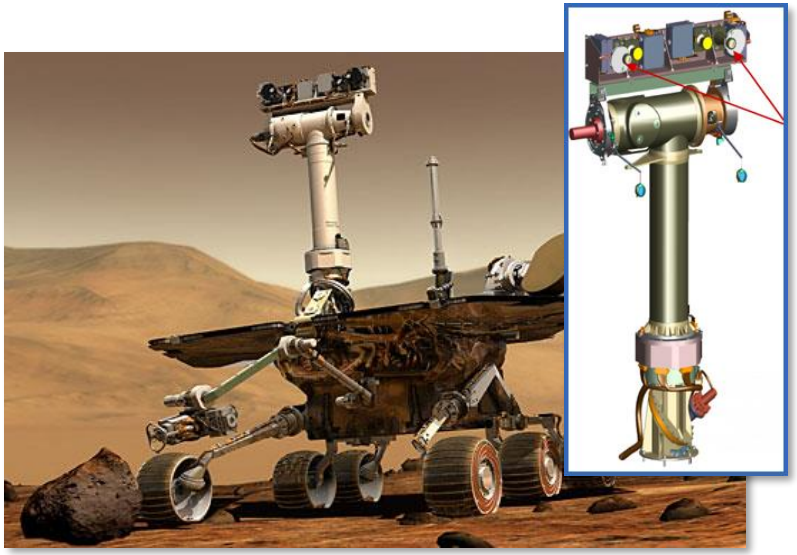
Computers can gain some high-level understanding from digital images/videos

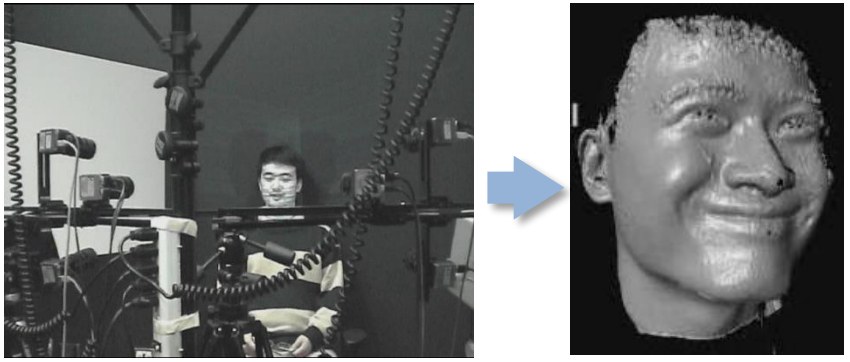
- wikipedia

The goal of computer vision

- Compute and understand the physical world



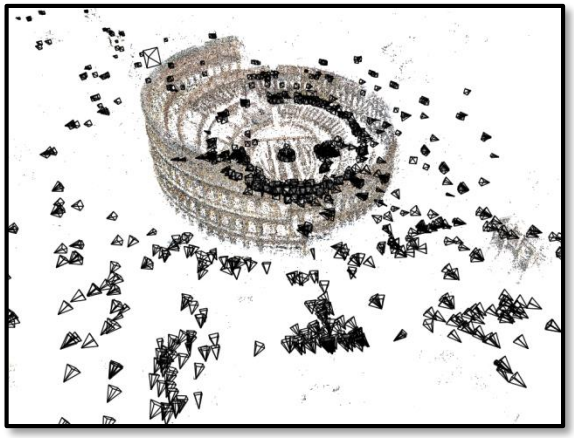




The goal of computer vision

- Reconstruct 3D model from crowdsourcing

Internet Photos

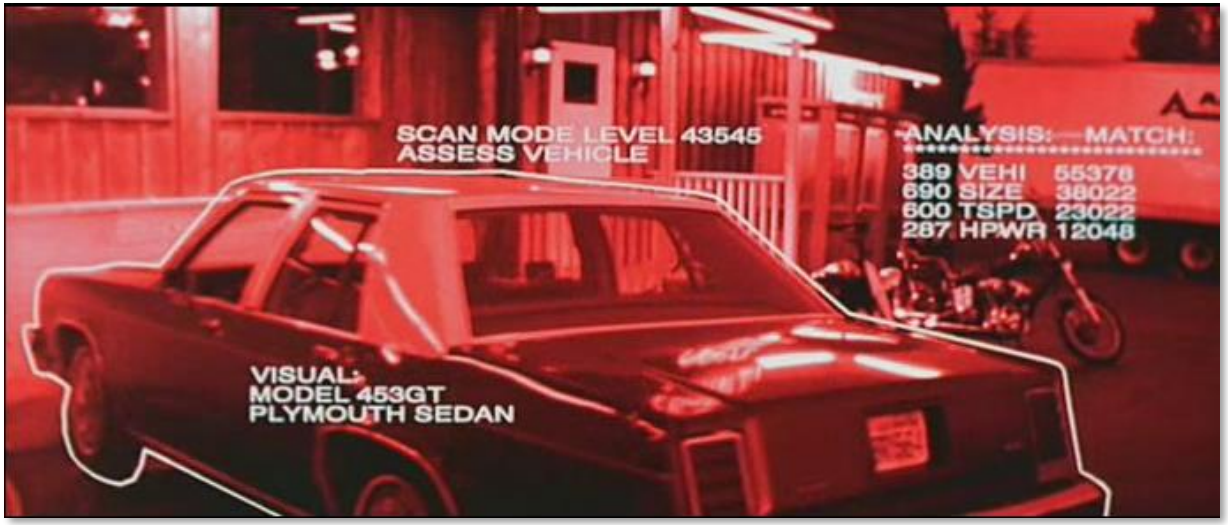


Reconstructed 3D cameras and points

Dense 3D model

The goal of computer vision

- Recognize objects and people



Terminator 2, 1991

The goal of computer vision

- Improve photos ("Computational Photography")

Super-resolution (source: 2d3)

Haze removal

Inpainting / image completion (image credit: Hays and Efros)

Why study computer vision?

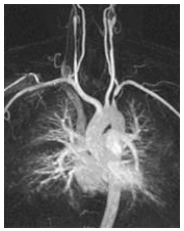
- Billions of images/videos captured per day

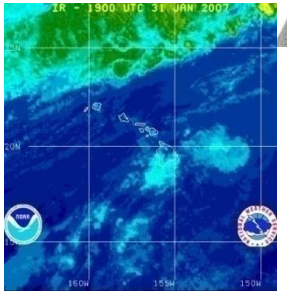









- Huge number of useful applications

Optical character recognition (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

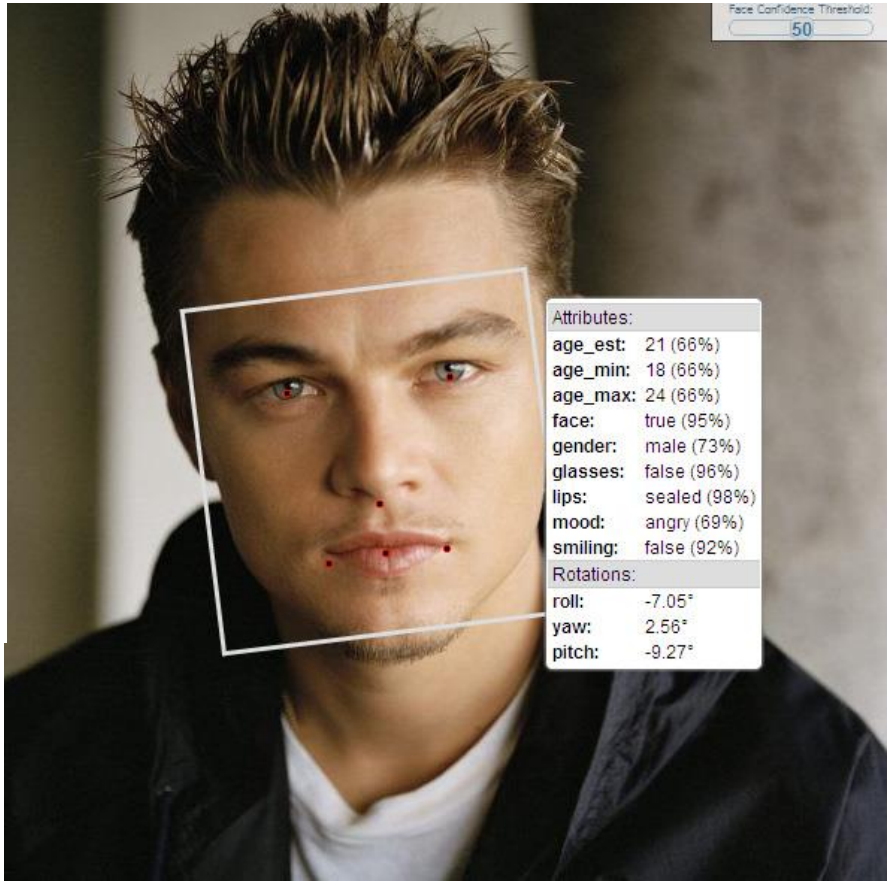
Automatic check processing

Source: S. Seitz

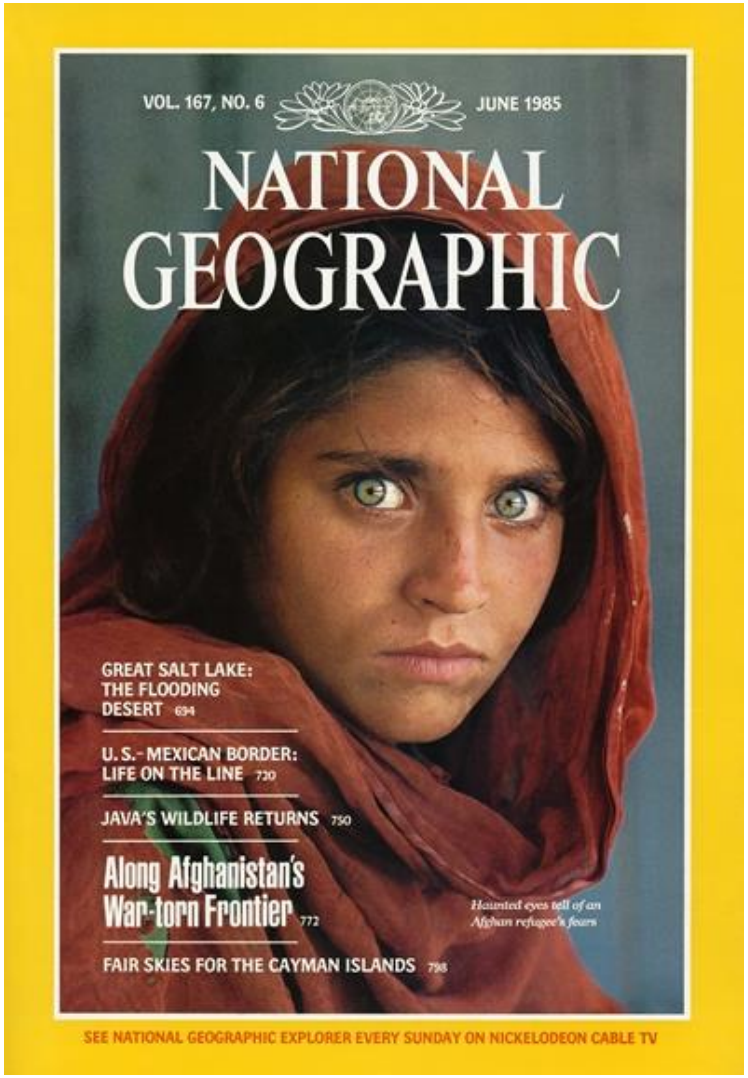
Face detection

- Nearly all cameras detect faces in real time

Face Recognition



Face recognition

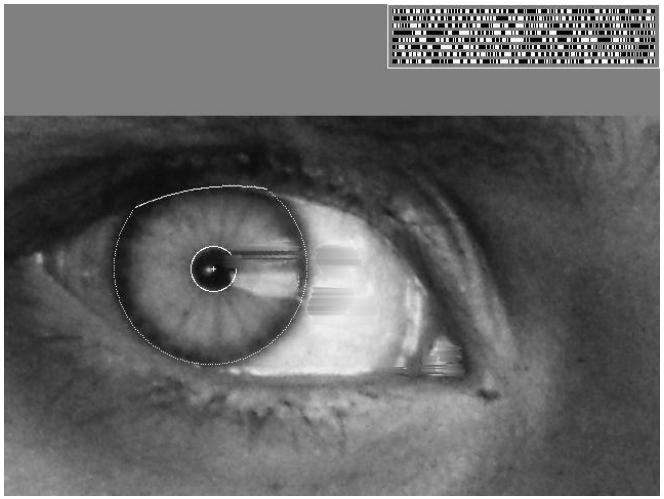


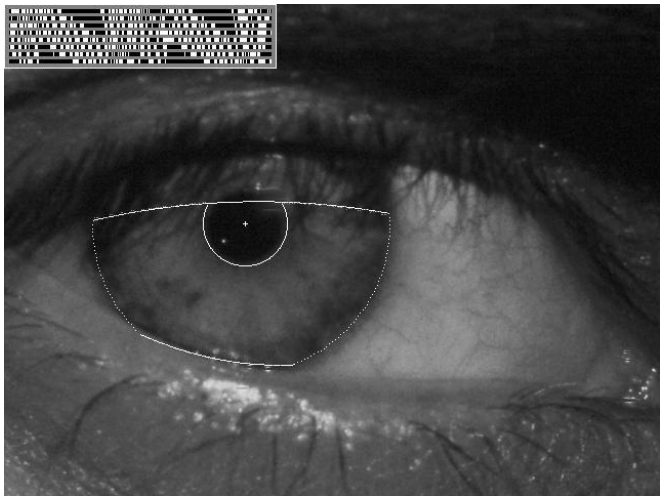
Who is she?

Source: S. Seitz

Vision-based biometrics

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



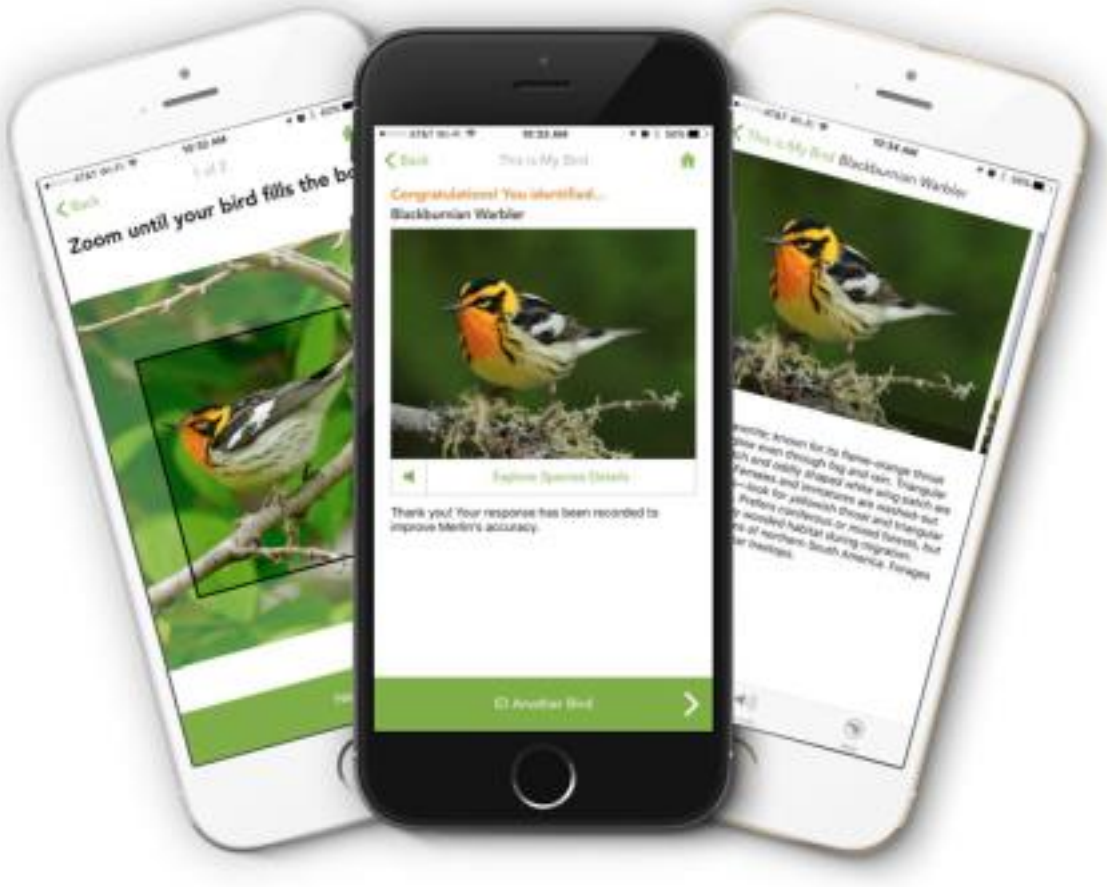


Source: S. Seitz

Object recognition (in mobile phones)

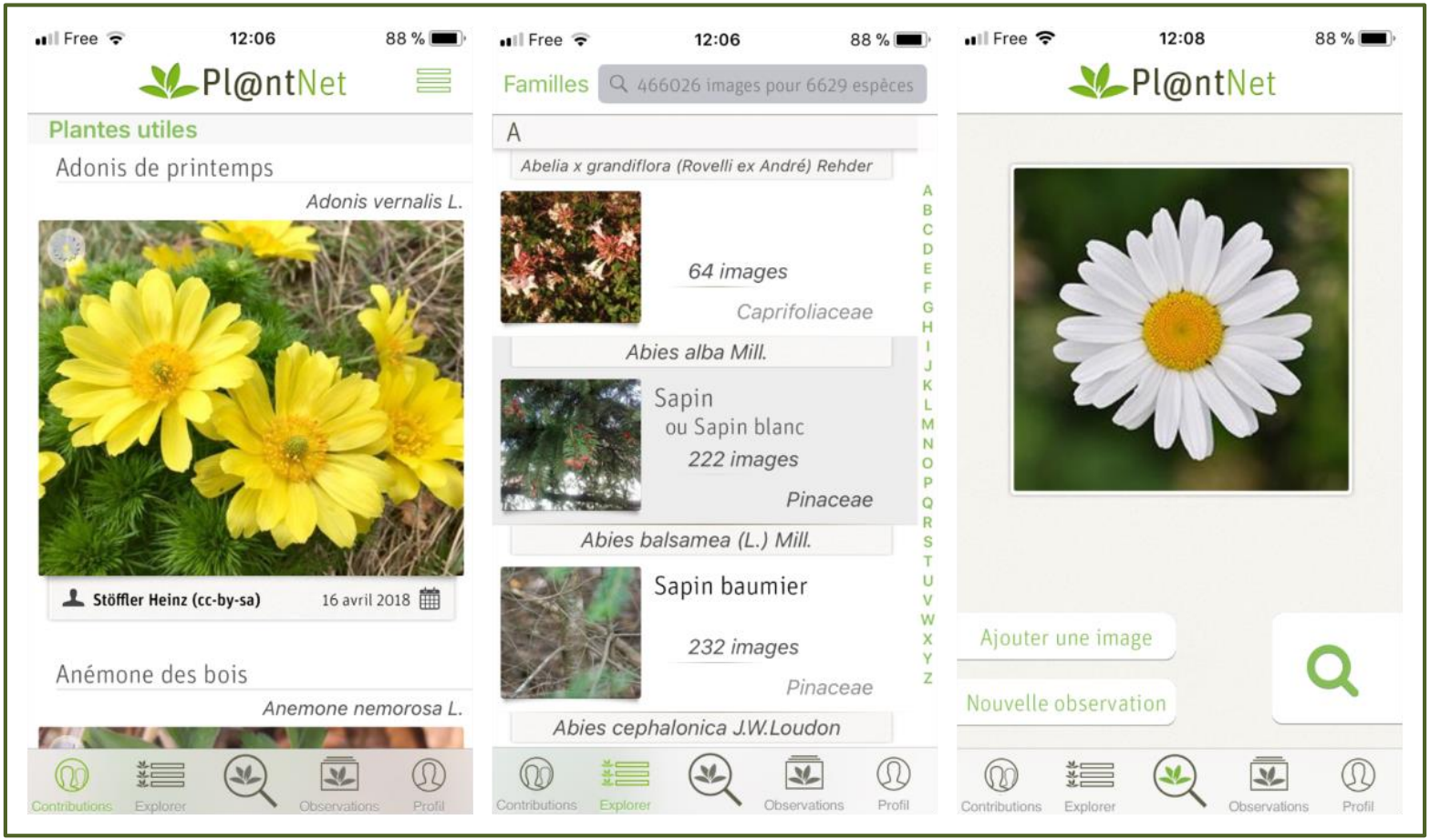
Source: S. Seitz

Bird Identification



Merlin Bird ID (based on Cornell Tech technology!)

Plant Identification



Pl@ntNet is a research and educational initiative on plant biodiversity supported by [Agropolis Foundation] since 2009.

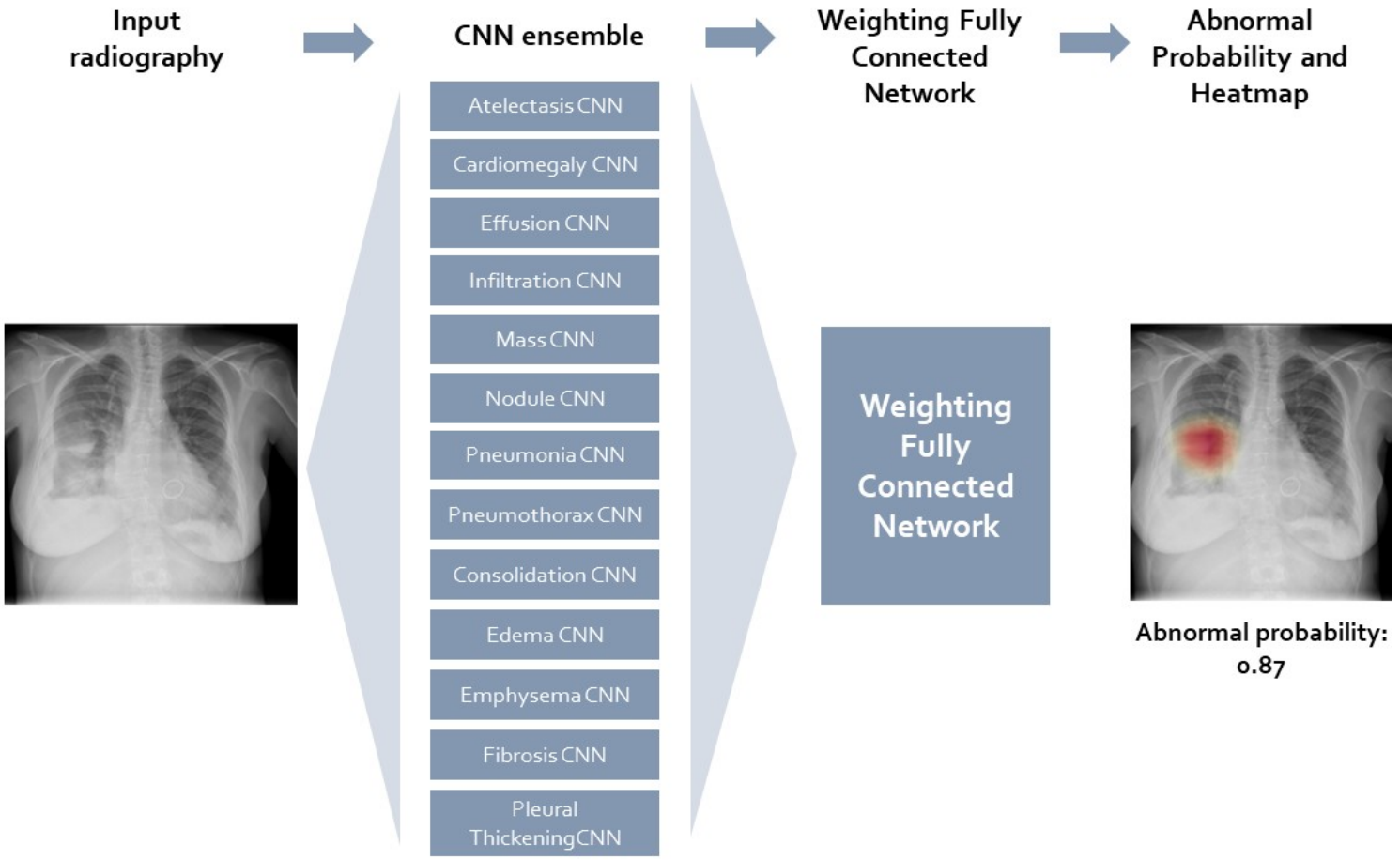
Marine Mammal Recognition

Under-water fish counting

Amazon Picking Challenge
http://www.robocup2016.org/en/events/amazon-picking-challenge/

Robomart

Medical imaging



Healthcare

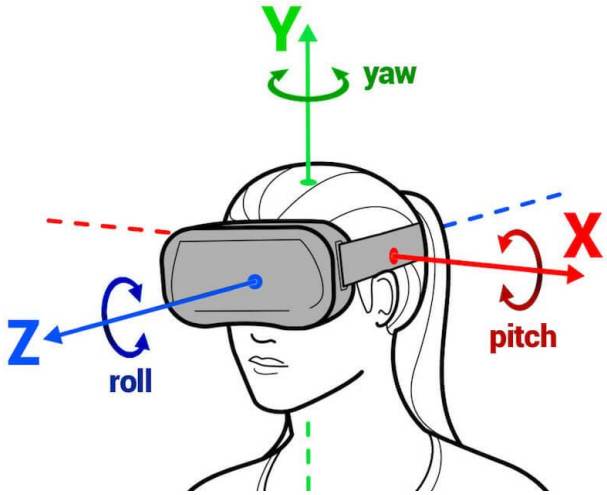
Gist – Chili fish head
Color moment – Braised pork
FC7 – Steamed chicken feet

AlexNet – Kung Pao Chicken
VGG – Kung Pao Chicken

Multi-task VGG – Kung Pao Chicken
[chicken, chili, peanut]

Region-based Multi-task VGG
chicken: *dice, stir-fry*
chili: *dry*
peanut: *roasted*

Virtual & Augmented Reality



6DoF head tracking

Hand & body tracking

3D scene understanding

3D-360 video capture

Current state of the art

- This is a very active research area, and rapidly changing
  - More apps in the next 5 years??
- To learn more about vision applications and companies
  - David Lowe maintains an excellent overview of vision companies
    - http://www.cs.ubc.ca/spider/lowe/vision.html

Why is computer vision difficult?

Viewpoint variation

Illumination

Scale

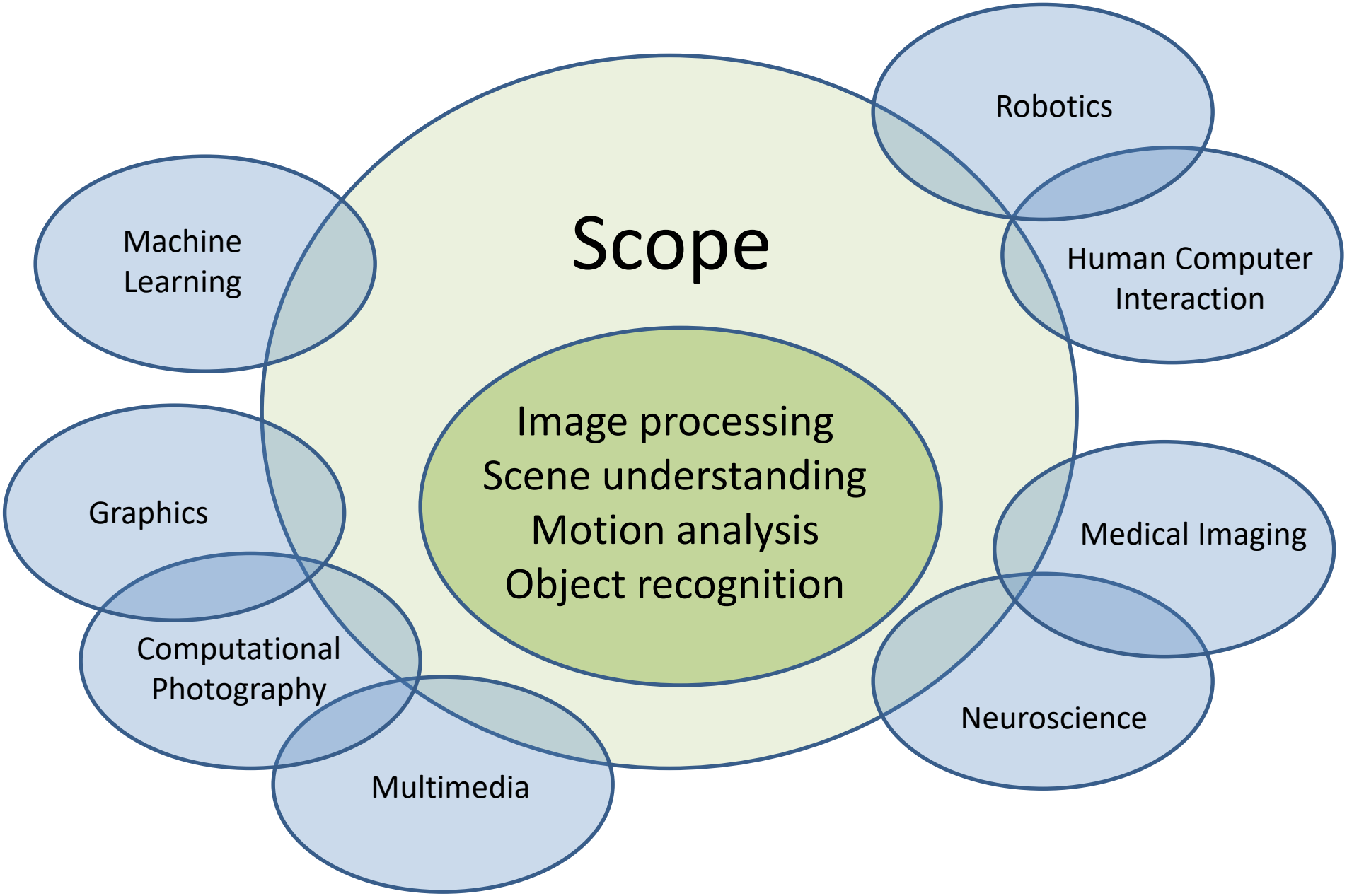
Why is computer vision difficult?

Intra-class variation

Motion (Source: S. Lazebnik)

Background clutter

Occlusion


Scope
Image processing
Scene understanding
Motion analysis
Object recognition
Robotics
Human Computer Interaction
Machine Learning
Graphics
Computational Photography
Multimedia
Medical Imaging
Neuroscience

Large Models for Computer Vision

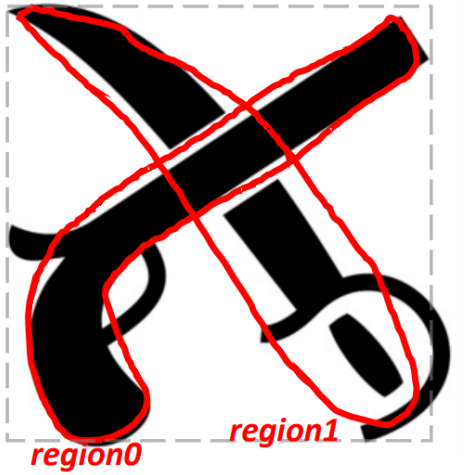
how many people are on the street?

There are three people on the street.

how many people riding a bike?

There are two people riding a bike.

https://llava.hliu.cc/

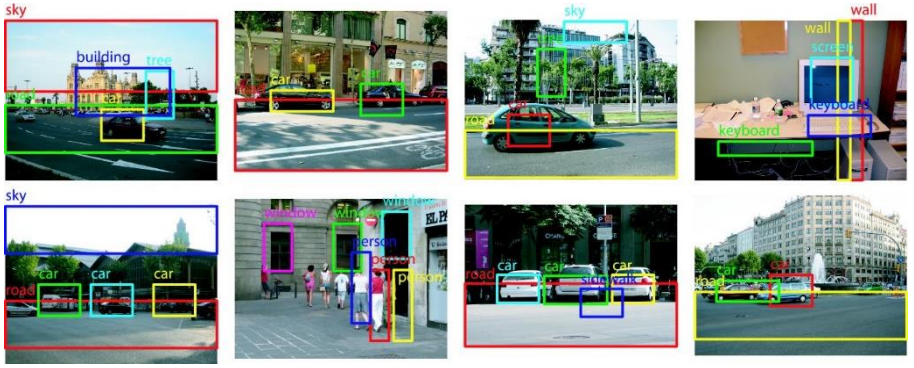


FERRET: REFER AND GROUND ANYTHING ANYWHERE AT ANY GRANULARITY

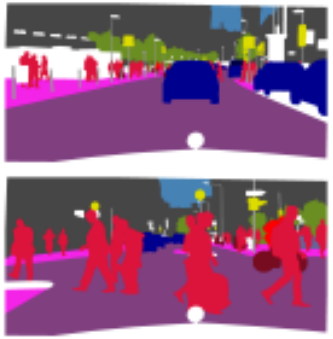
AIGC: Which one is real?

Suggested Topics

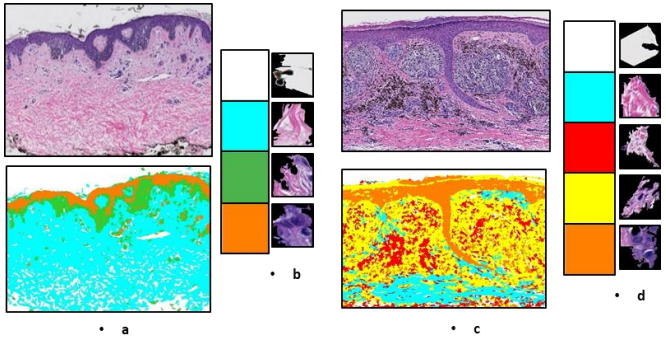
Object/face recognition



Detection/Tracking



Segmentation



(Medical) Image Processing/Enhancement/…

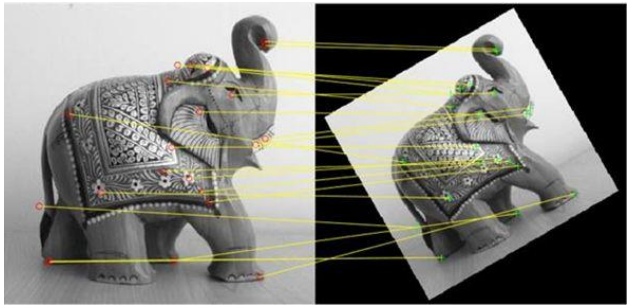


Image registration

You may not follow these suggested topics, and you could choose your own preference topic. ☺

Course information

- Prerequisites
  - A good working knowledge of programming
  - Data structure and algorithm
  - Some math: linear algebra, vector calculus
- Grading
  - Assignments (30%)
  - Group project (20%)
  - Final exam (50%)