TDS3651
Visual Information Processing

Course Information
[Trimester 2310]

Faculty of Computing and Informatics
Multimedia University

Some graphics and materials are taken from UTexas at Austin CS376 & CS381V notes, Stanford Uni's CS131 notes and Gonzalez & Wood's DIP textbook

Course Information

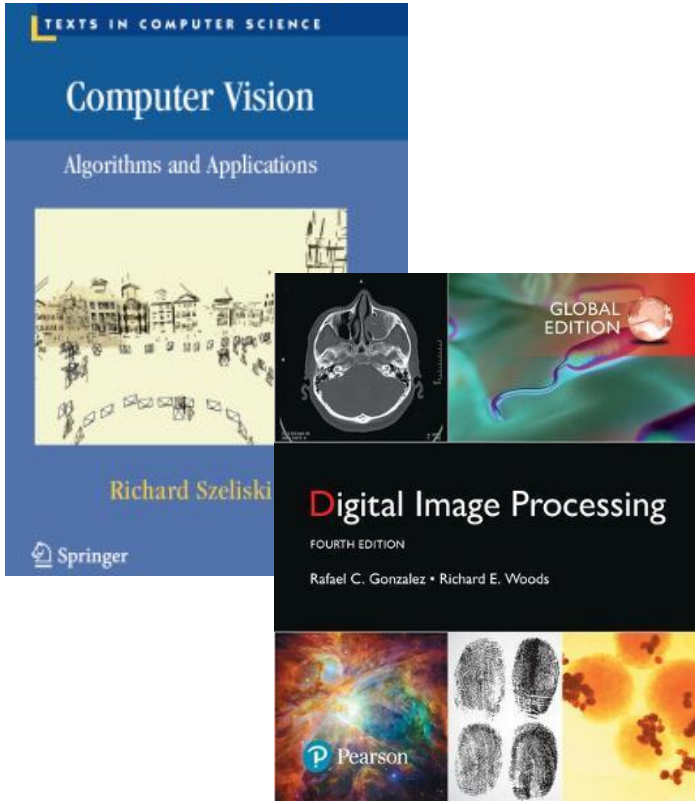
- **Visual Information Processing TDS3651**
  - Specialization Elective for Data Science specialization.
  - Elective for all other specialization.

- **Pre-requisites**
  - OOPDS (TCP1201)
  - “Other pre-requisites” – Decent programming skills, Like problem-solving

Course Instructor

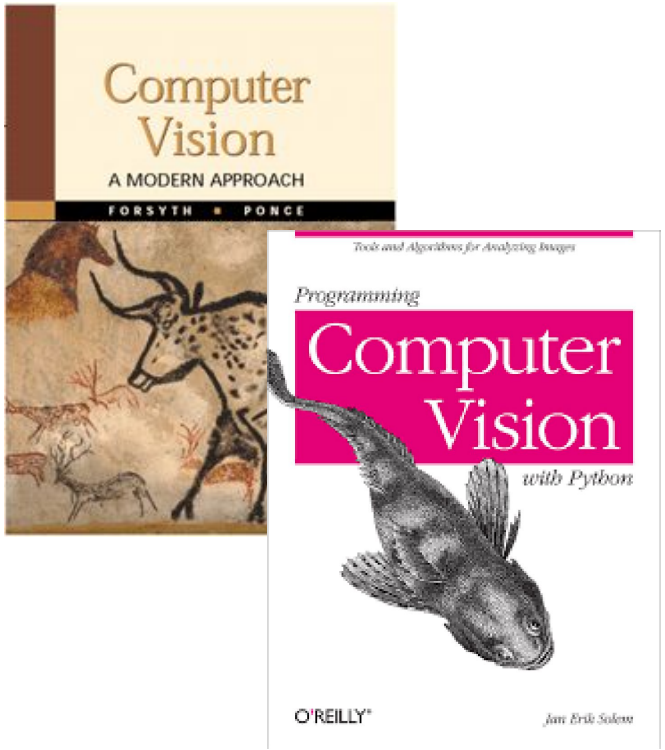
- **Noramiza Hashim**
  - Email: *noramiza.hashim@mmu.edu.my*
  - Office: BR2007

Main Reference



- **Digital Image Processing**, 4th Ed. – Gonzalez & Woods (2018)
- **Computer Vision: Algorithms and Applications**, R. Szeliski (2011)

Other References



- **Computer Vision – A Modern Approach** – Forsyth & Ponce, 2nd Ed. (2011)
- **Programming Computer Vision with Python** – Solem (2012)
- **Concise Computer Vision** – Klette (2014)

Topics overview

- Image Filtering, Morphological Operations
- Low-level Features – Edges, Color, Texture
- Local Invariant Features – Detectors and Descriptors
- Feature Indexing & Encoding
- Clustering and Region Segmentation
- Deep Learning in Computer Vision
- Motion processing in video
- Applications

Goal of this course: Introduction to primary topics, Hands-on problem solving, Motivation for research

Software/Tools

- **Python 3 (3.11 latest)**
  - Google Colaboratory [online browser]
  - Anaconda package [local installation] (https://www.anaconda.com/products/individual)
    - Python + useful libraries like scipy, numpy, matplotlib)
    - Spyder IDE
    - Jupyter notebooks
- **OpenCV – any 4.x version (4.8.1 latest)**
  - pip install opencv-python

Course Learning Outcome

CLO1 : Describe fundamental concepts and techniques in image processing and computer vision with their usages

CLO2 : Apply existing computer vision algorithms and schemes to process visual information in specific tasks

CLO3 : Design solutions and techniques to solve real-world visual processing problems

CLO4 : Demonstrate the application of various image processing algorithms for a variety of image processing tasks with effective communication

Coursework Grading

100% Coursework

- **Quiz (20%)**
  - Format: problem set homework / MCQ, class quizzes, pop lab quizzes, etc.
- **Assignment (20%)**
  - Task specific assignment
- **Test (20%)**
  - MCQ / discussion based questions
- **Project (40%)**
  - Visual information system
  - Programming, report, presentation

Course Policies

- **Late Policies**
  - -10% (of the submission mark) per day late
  - Late due to emergency cases must be supported by valid documentations as proof (accident reports, etc.)
  - Disagreements with groupmates should be resolved before submission (not used as excuse)
- **University-wide Policies**
  - Plagiarism is a serious offence. If found, immediate ZERO given to assessment with no negotiation.
  - Attendance to be taken as per university rules

Recommended Python Primers

- Justin Johnson (Stanford Uni)'s **Python Numpy tutorial**

  http://cs231n.github.io/python-numpy-tutorial/

  iPython notebook:

  https://github.com/kuleshov/cs228-material/blob/master/tutorials/python/cs228-python-tutorial.ipynb

- **SciPy Lecture Notes** – One document to learn numerics, science, and data with Python

  http://www.scipy-lectures.org/index.html

- **OpenCV Python Tutorials** (make sure version is correct!)

  https://docs.opencv.org/4.8.0/d6/d00/tutorial_py_root.html

TDS3651
Visual Information Processing

LECTURE 1

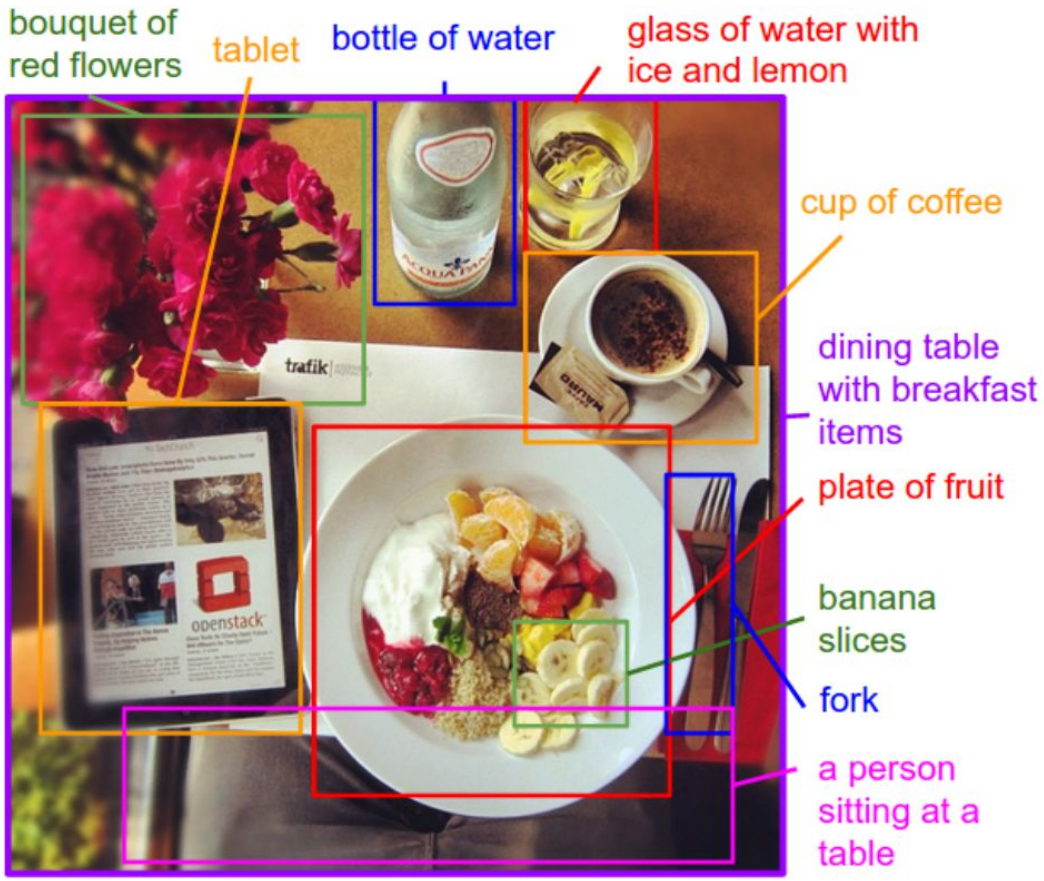
Introduction

Faculty of Computing and Informatics
Multimedia University

prepared by John See; Lai-Kuan, Wong
modified by Yuen Peng, Loh

Some graphics and materials are taken from UTexas at Austin CS376 & CS381V notes, Stanford Uni's CS131 notes and Gonzalez & Wood's DIP textbook

What can the “visual” world tell us?



Object/Scene classification + Visual captioning with NLP

Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3128-3137).

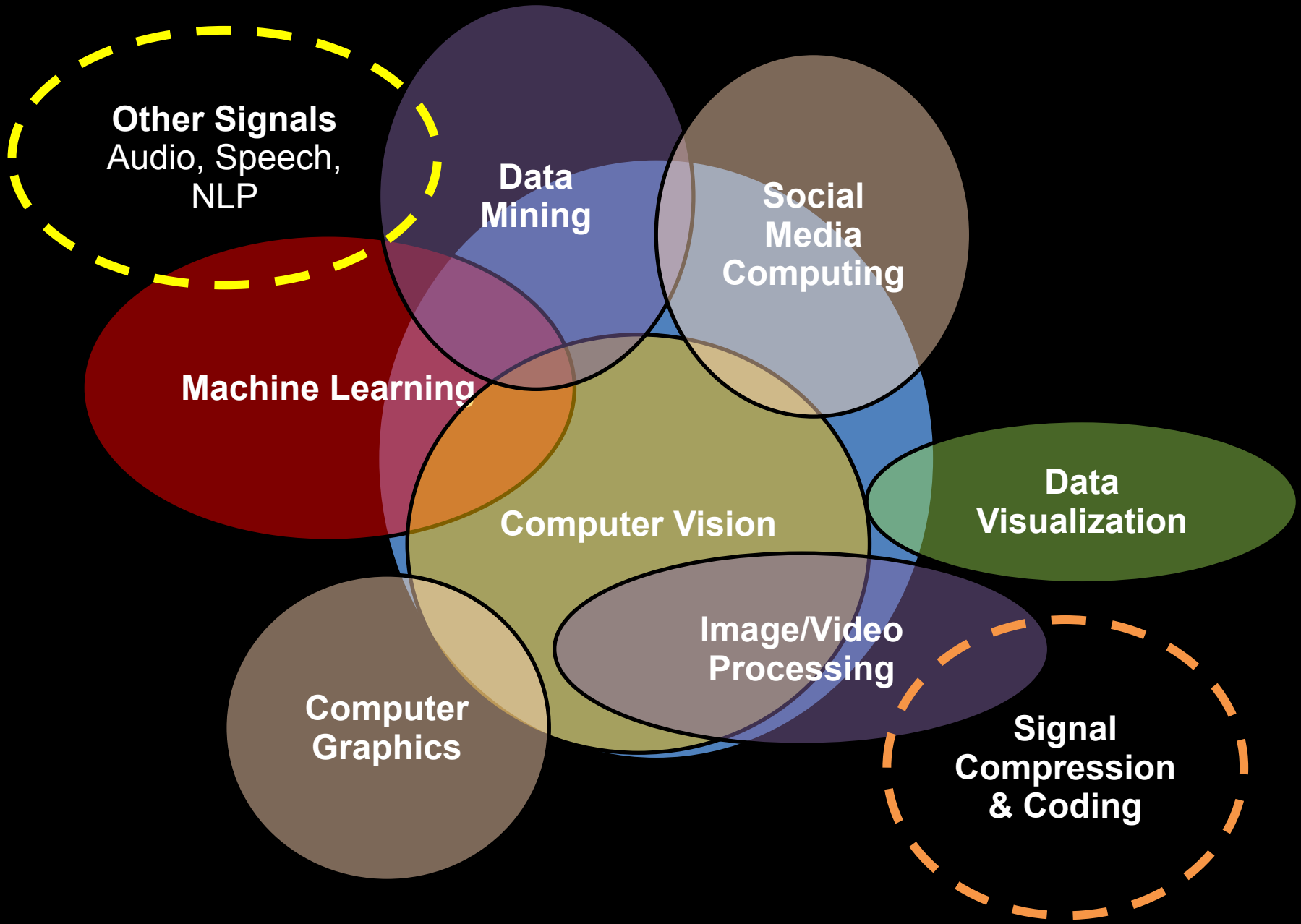
Can it drive cars?

Scene understanding + Autonomous driving

Can it tell us more about others?

Face recognition + Social Network Mining + NLP

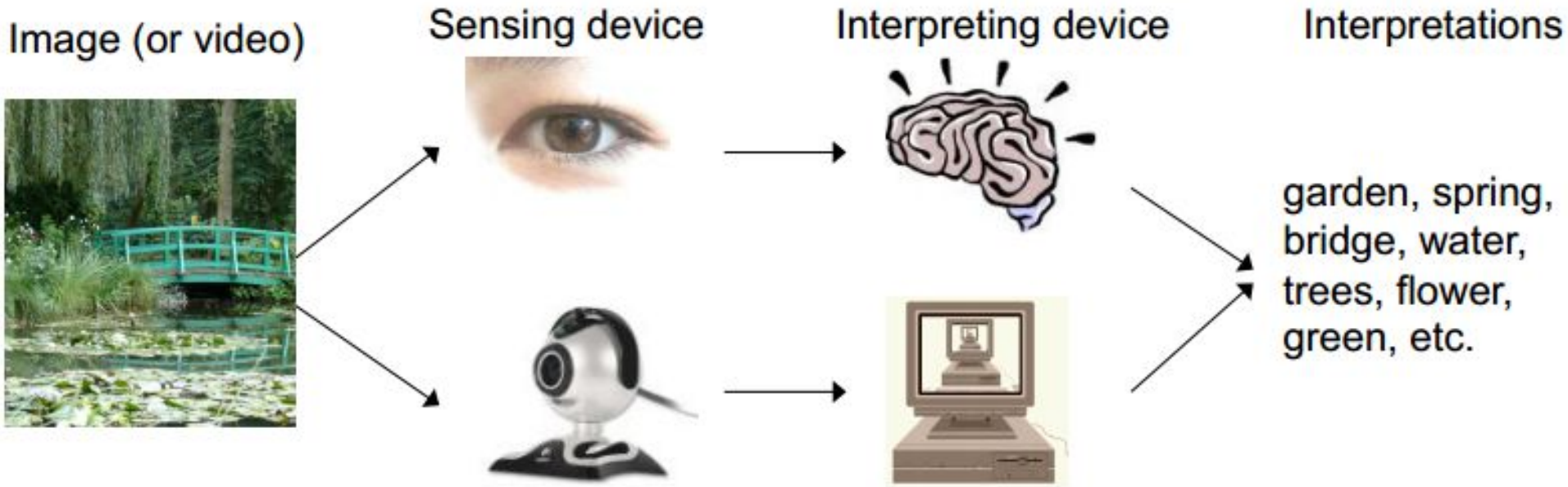
Visual Information Processing



What is computer vision?

Good enough?

What is computer vision?



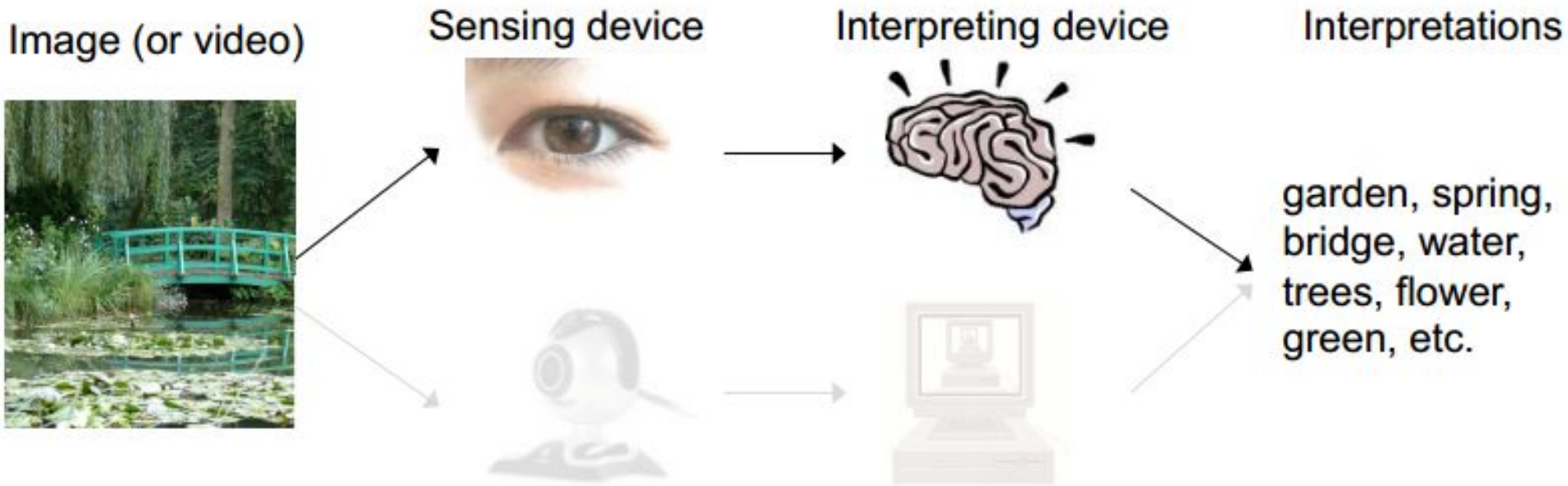
Computer Vision: The Goal

What we see

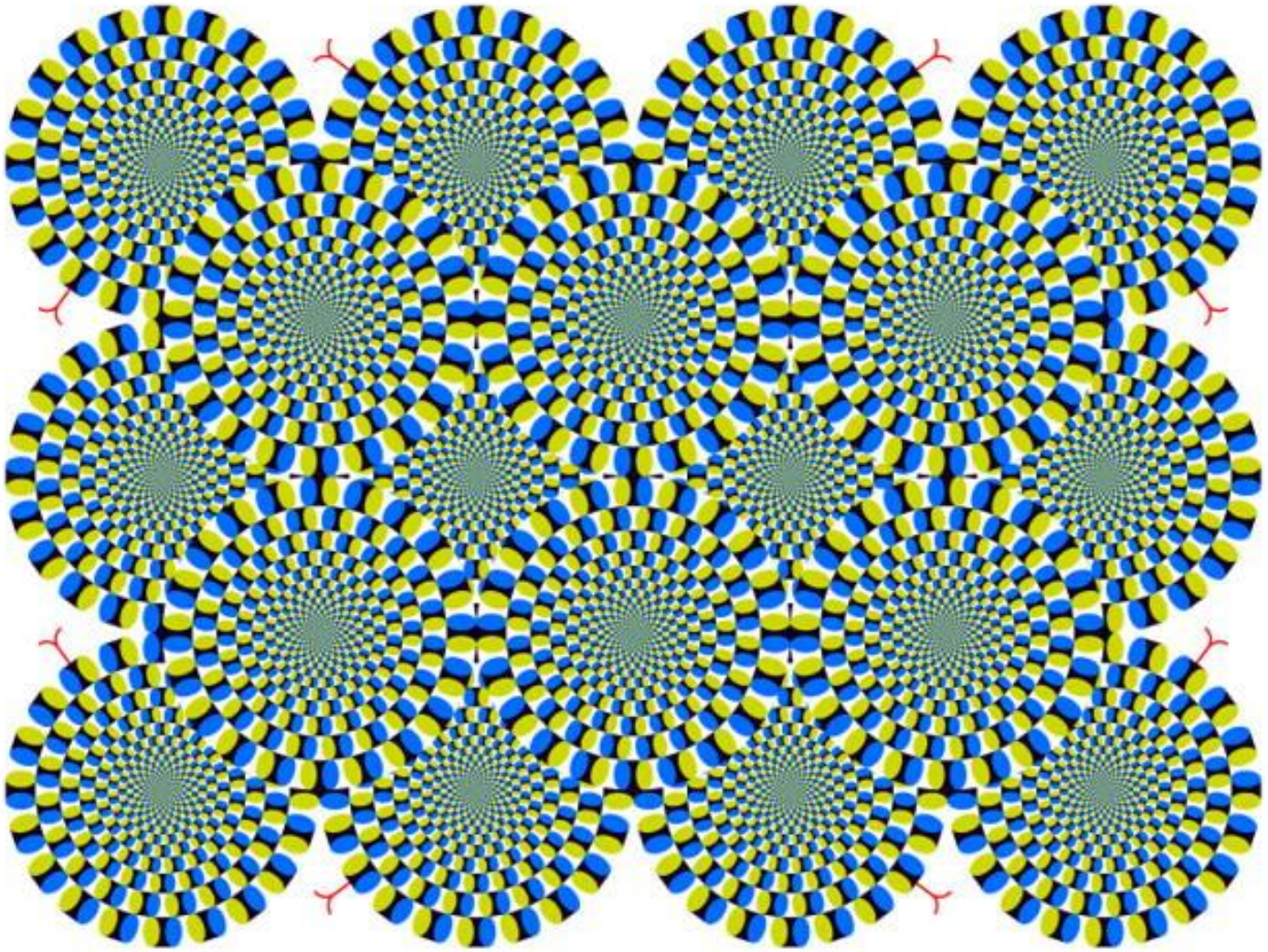
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What the computer sees

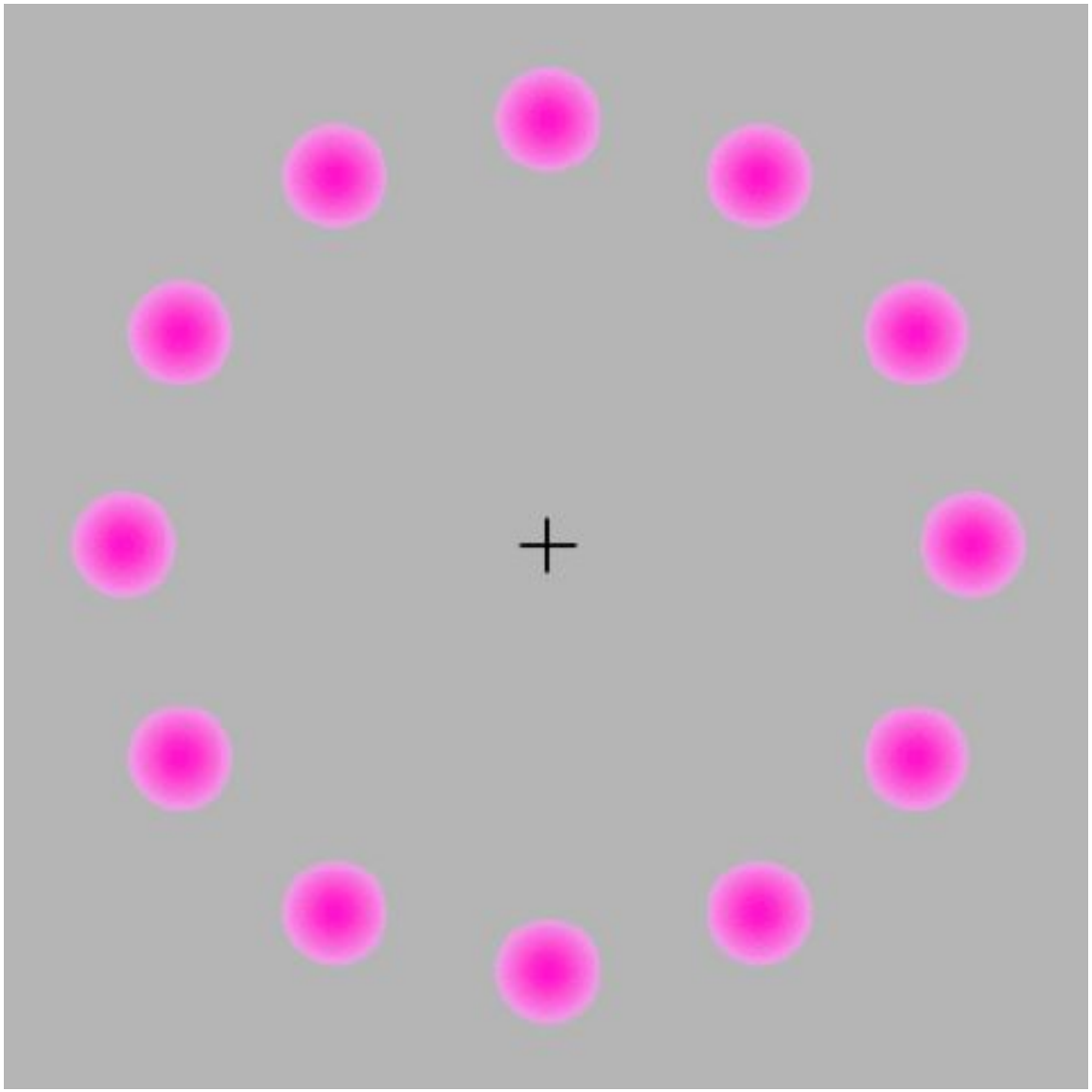
What is computer vision?



Human vision is superb

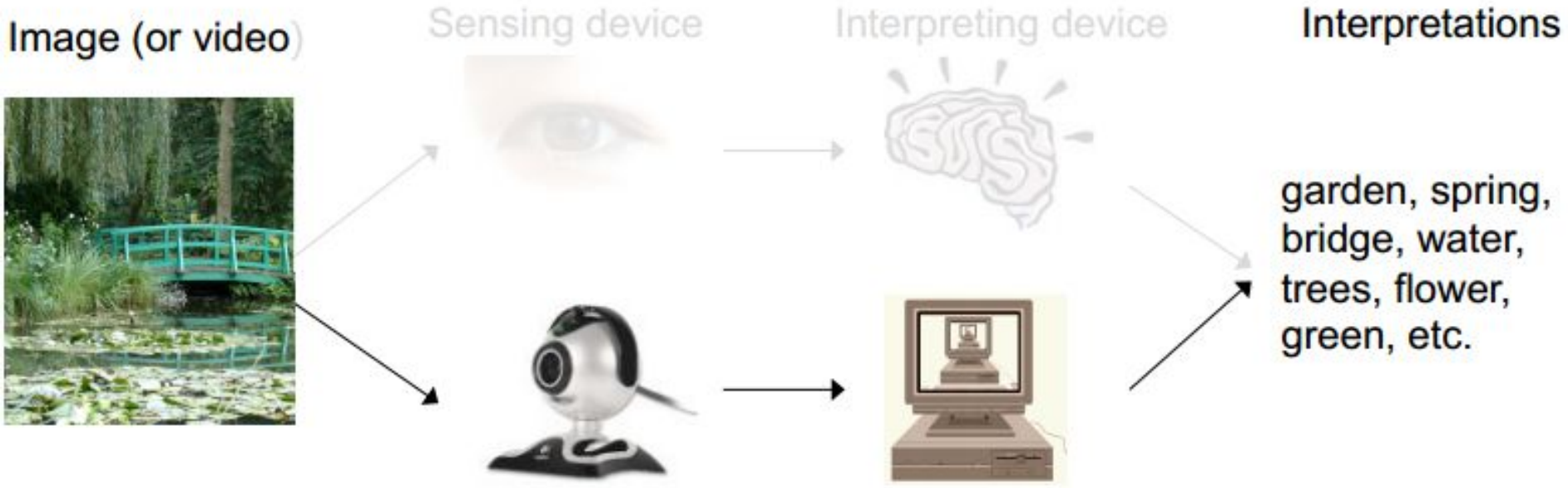


Troxler’s effect



Is that the same person?

What is computer vision?



Fields that deal with Images/Videos

- Computer Graphics: Creation and synthesis of images
- Image/Video Processing: Enhancement or manipulation of an image or video
  - Image ⇒ Image | Video ⇒ Video
- Computer Vision: Analysis and understanding of image/video content
  - Image/Video ⇒ Information (for further use)

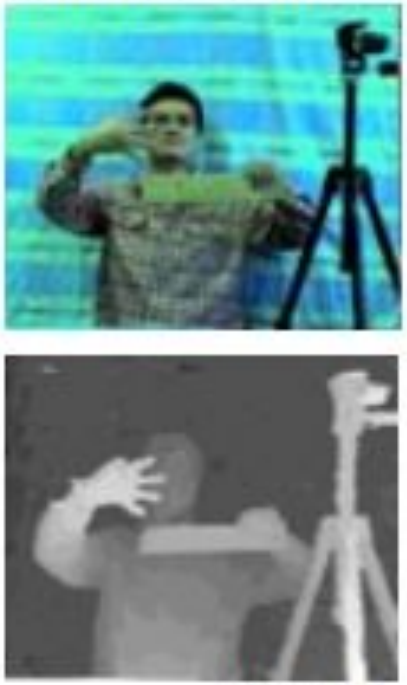
Computer Vision

- Automatic understanding of images and video

1. **Computing properties of the 3D world from visual data *(measurement)***

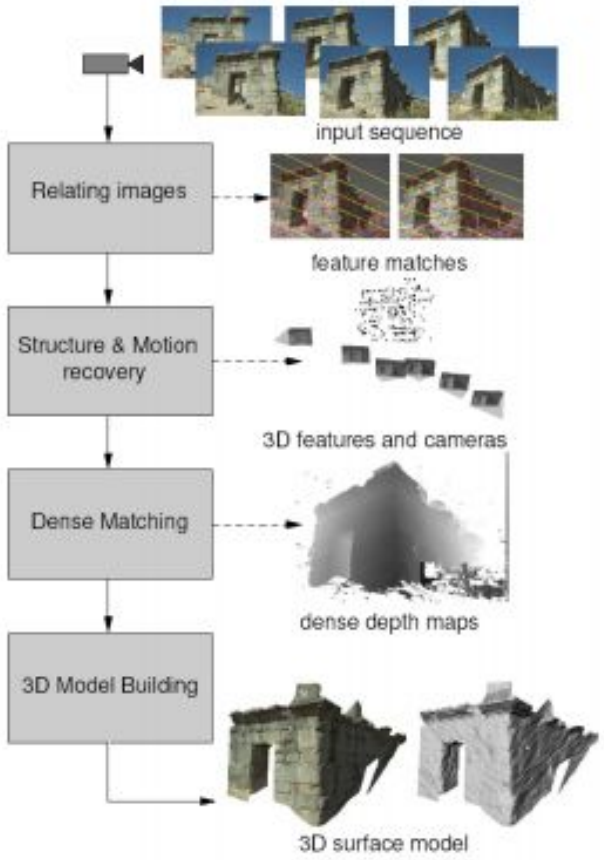
For measurement

Real-time stereo



Wang et al.

Structure from motion



Tracking

Demirdjian et al.

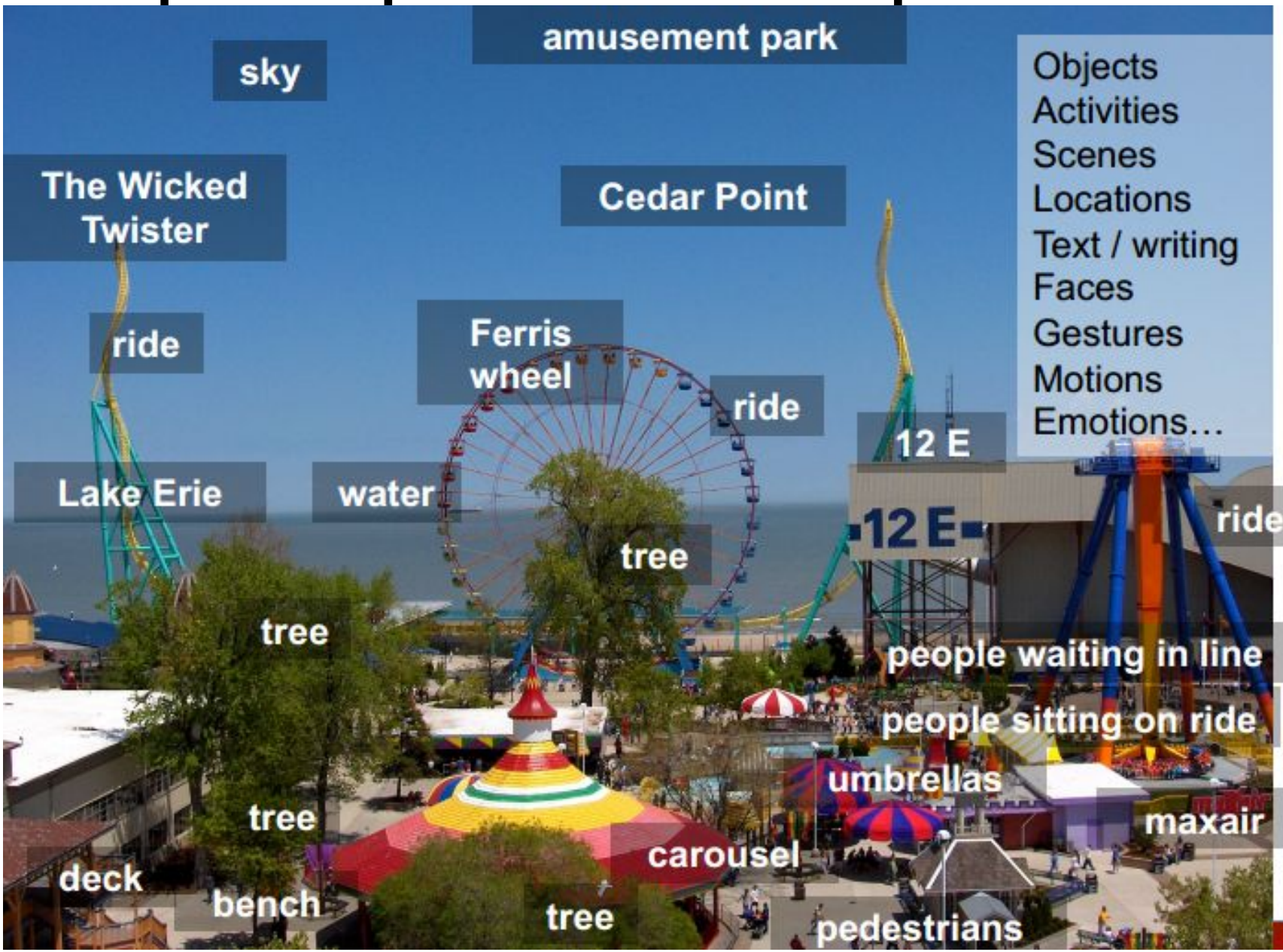
3D reconstruction from image views

Computer Vision

- Automatic understanding of images and video

1. Computing properties of the 3D world from visual data ***(measurement)***
2. **Algorithms and representations to allow a machine to recognize meaningful entities (objects, people, scenes, activities, etc.)**
   (perception and interpretation)

For perception & interpretation

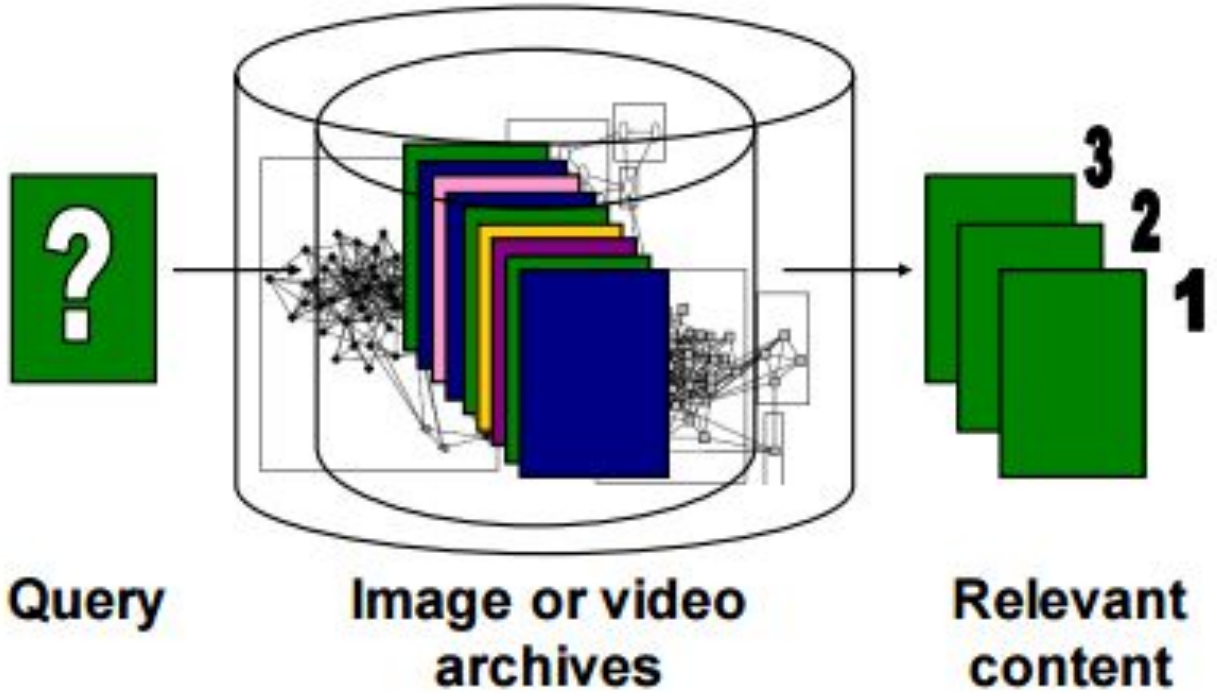


Computer Vision

- Automatic understanding of images and video

1. Computing properties of the 3D world from visual data ***(measurement)***
2. Algorithms and representations to allow a machine to recognize meaningful entities (objects, people, scenes, activities, etc.)
   (perception and interpretation)
3. **Algorithms to mine, search and interact with visual data** ***(search and organization)***

For visual search & organization

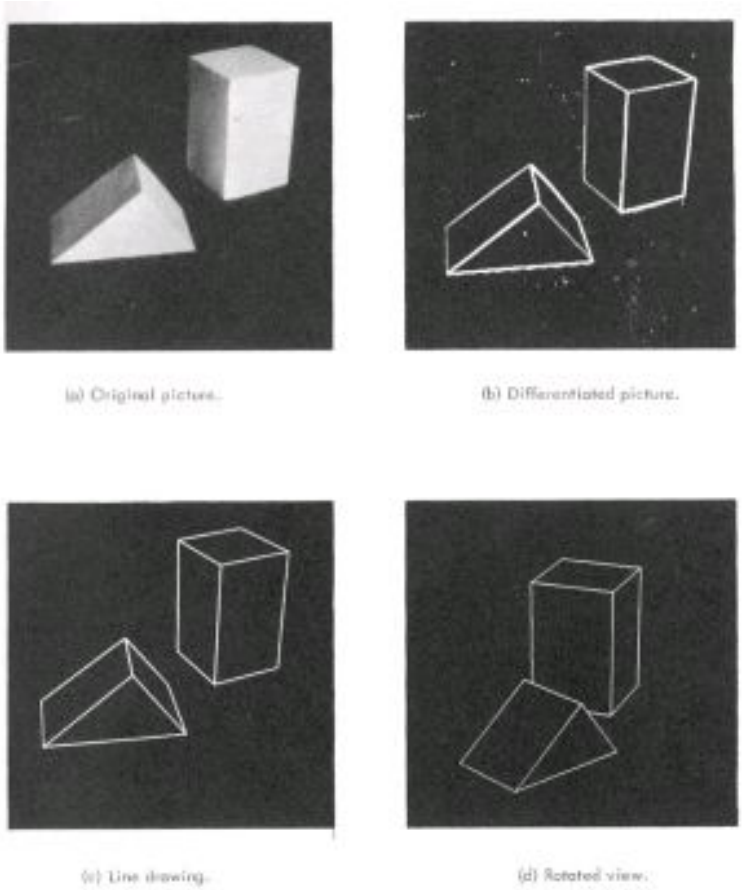


Course Focus

- Automatic understanding of images and video

1. Computing properties of the 3D world from visual data *(measurement)*
2. Algorithms and representations to allow a machine to recognize meaningful entities (objects, people, scenes, activities, etc.) *(perception and interpretation)*
3. Algorithms to mine, search and interact with visual data *(search and organization)*

Visual data in 1963



(a) Original picture. (b) Differentiated picture.

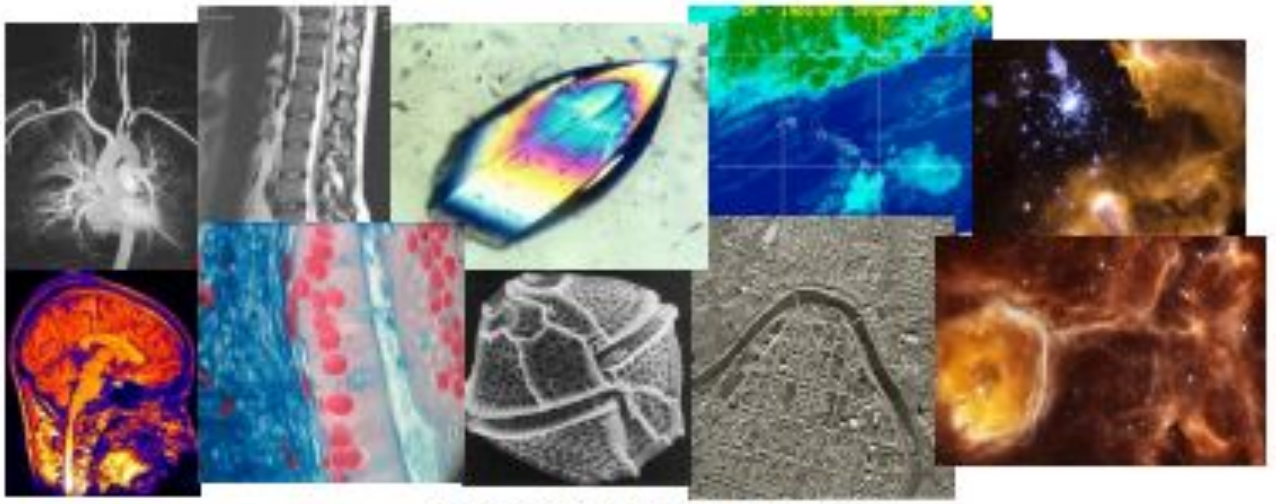
(c) Line drawing. (d) Rotated view.

L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Visual data in 2016

Google Image Search Picasa™ flickr™ webshots picsearch™ YouTube Broadcast Yourself™

Surveillance and security

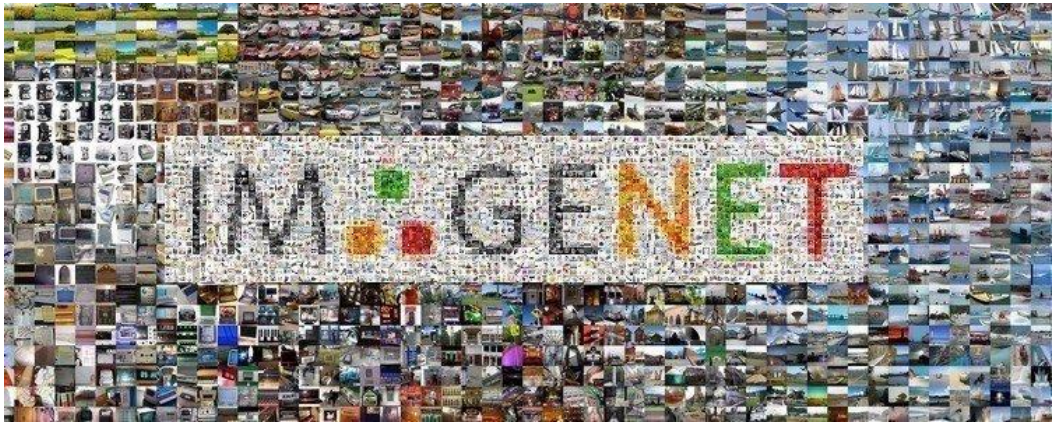


Medical and scientific images

Visual data now

Labeled Faces in the Wild

Microsoft Common Objects in Context (COCO)



PRINCETON UNIVERSITY

Stanford University

ImageCLEF/LifeCLEF
Cross Language Evaluation Forum (CLEF)

Special effects: Shape & motion capture

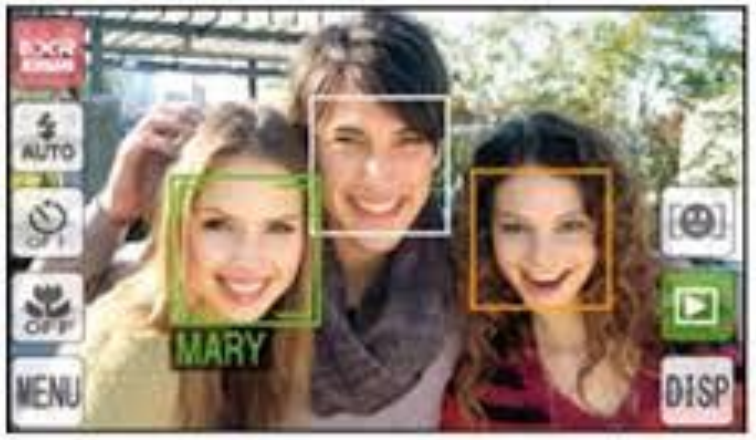
Urban modeling: Microsoft Photosynth

Pix camera app for iOS.

Face detection

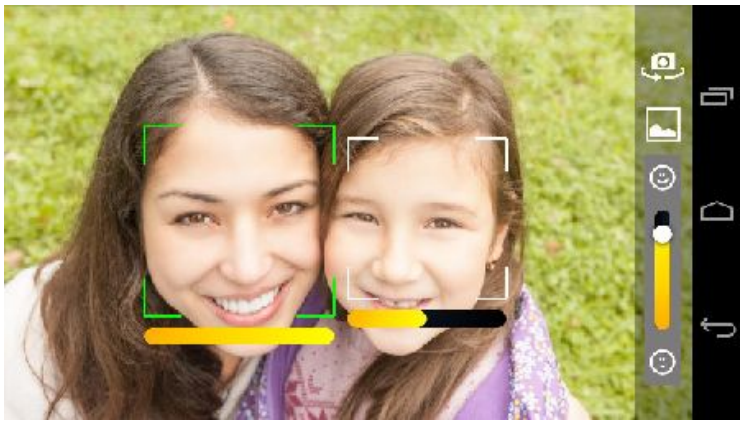
Many digital cameras and smartphone cameras now can detect faces

Face detection…and MORE!



Name recognition and Priority Ranking (Fujifilm)

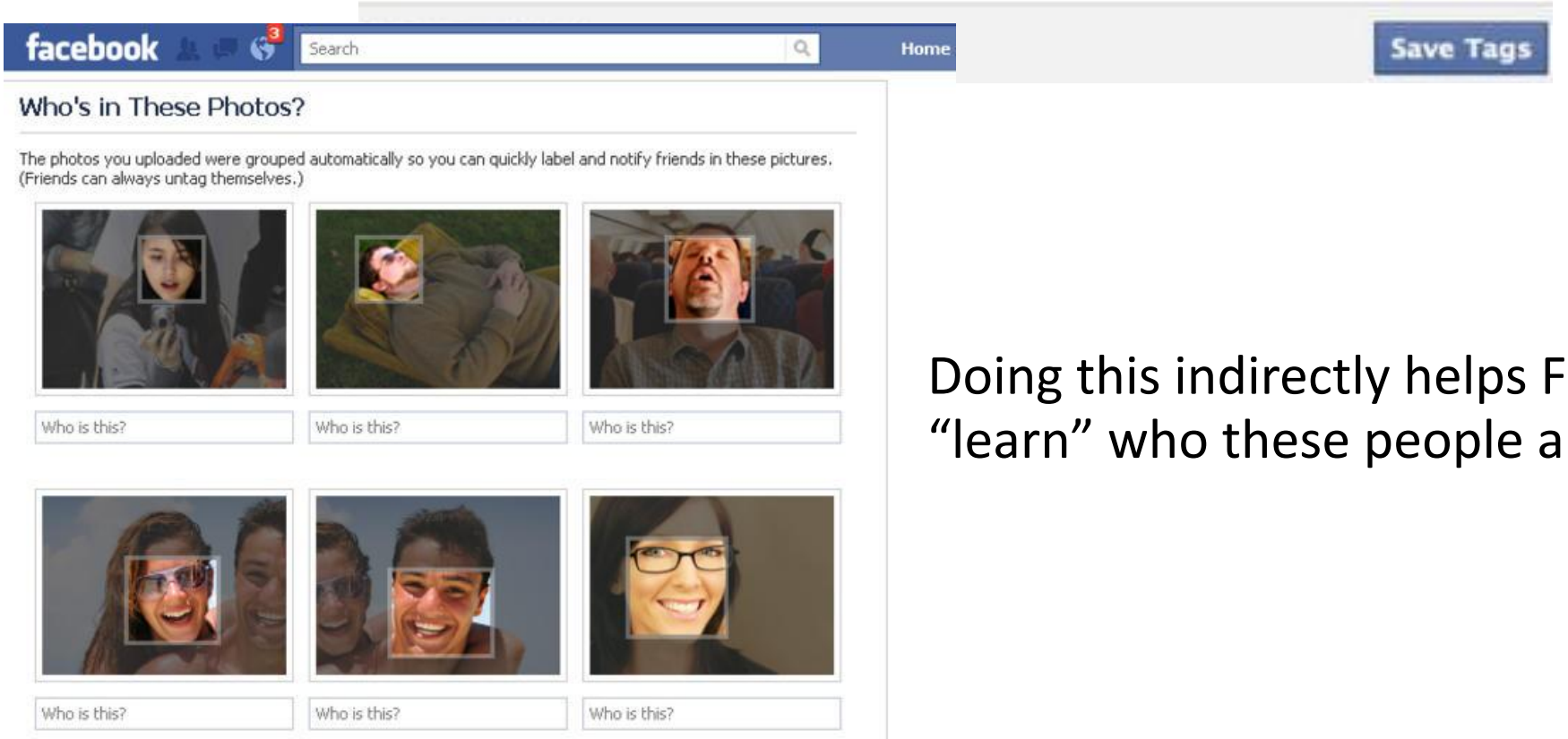
Blink detector (Nikon)



Smile detector (Android camera app, Sony)

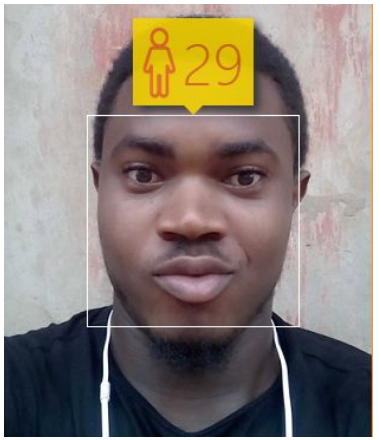
Face recognition: Facebook (where else)

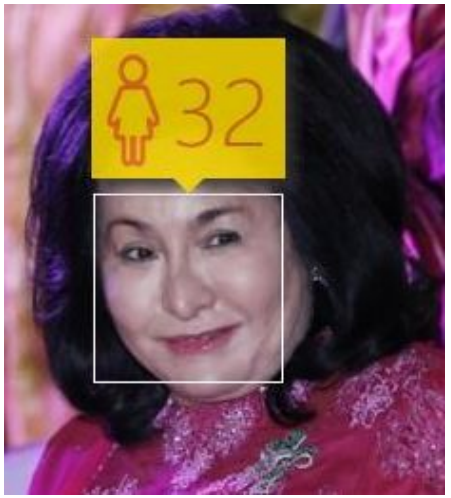
SCARY

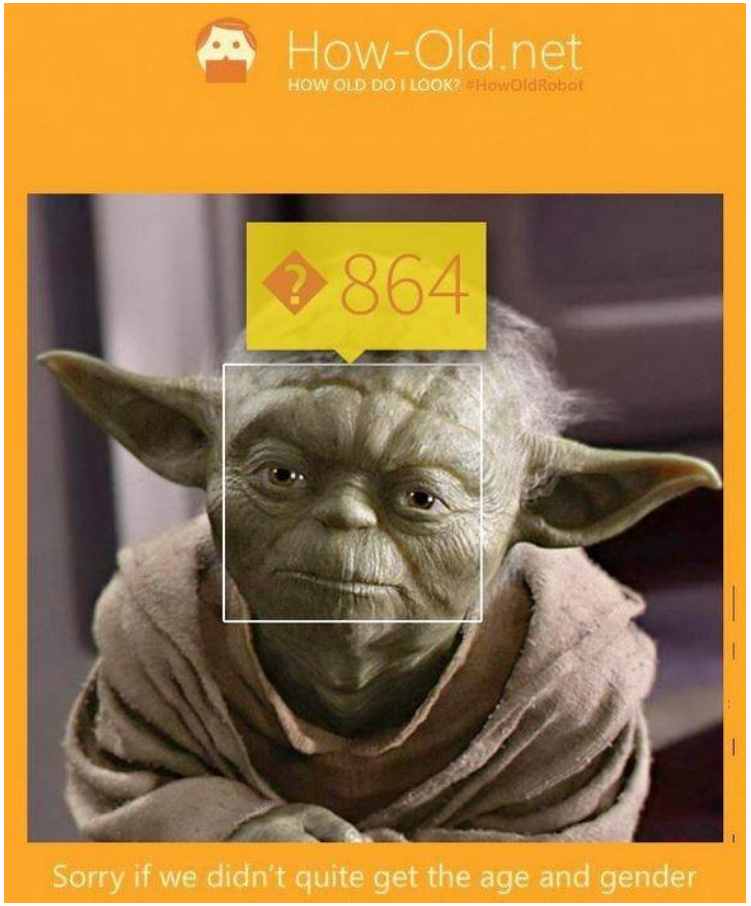


Doing this indirectly helps FB to "learn" who these people are!

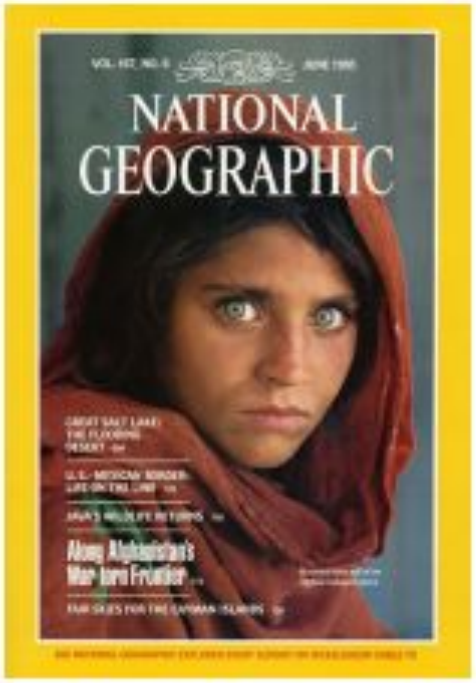
Age estimation: How-Old.net



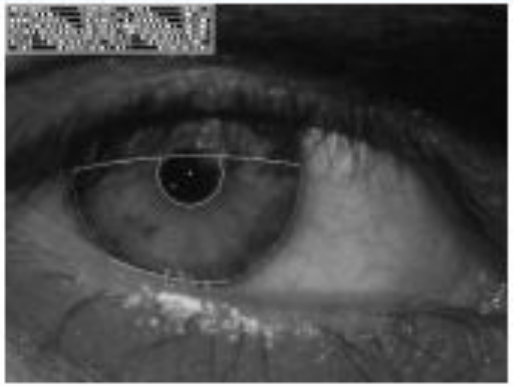




Biometrics



[How the Afghan Girl was Identified by Her Iris Patterns]

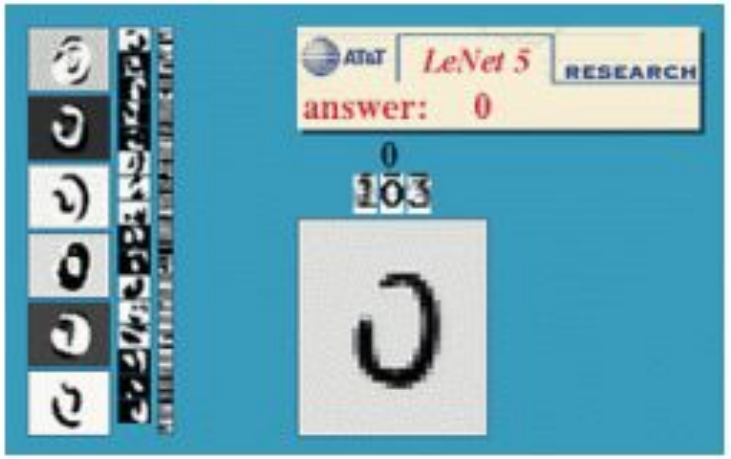


Source: S. Seitz

Biometrics: Fingerprint

Optical Character Recognition (OCR)

OCR technology converts scanned docs to text



Digit Recognition, AT&T Labs

License plate recognition

Image Search/Retrieval

Query text/image

Red Flower

27862 0.92 4

27826 0.92 6

27887 0.92 8

27818 0.92 7

27878 0.91 4

52385 0.91 5

3962 0.91 5

27859 0.91 4

52307 0.90 10

27849 0.90 4

27821 0.90 6

15723 0.90 6

Automotive safety

Vision systems are found in many car models,
eg. Subaru, BMW, Volvo, and Lexus.

In supermarkets?

Amazon Go: The Future Store

Vision-based interaction and games

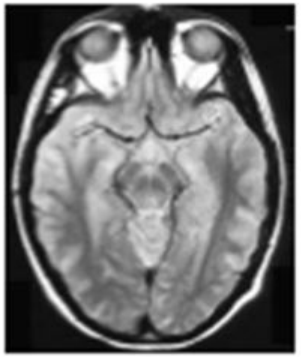
Space exploration

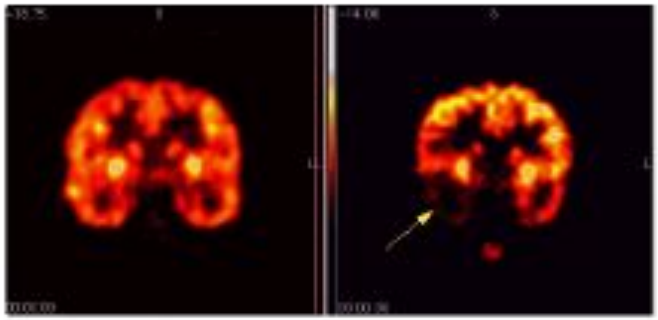
Panoramic view captured by NASA's Mars Exploration Rover Spirit in 2007

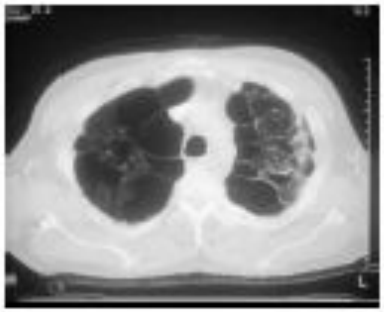
Image Processing (low level manipulation of images) has other applications as well

Medical Image Processing

- Ultra Sound (US)
- Magnetic Resonance Imaging (MRI)
- Positron Emission Tomography (PET)
- Computer Tomography (CT)
- X-Ray

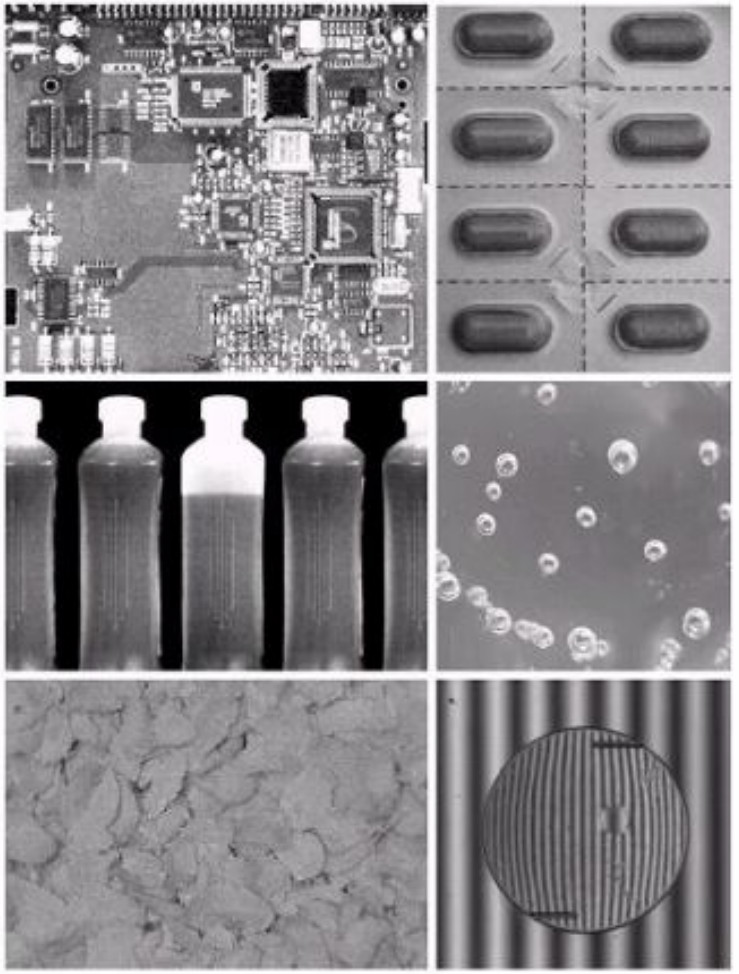






Automated Industrial Inspection

- Circuit board – missing parts
- Pill container – missing pills
- Bottles – filled up levels
- Bubbles in clear-plastic product – detect unacceptable air pockets
- Cereal – inspection for color, presence of burnt flake
- Image of replacement lens for human eye – inspection of damaged implants



Satellite Image Processing

- Remote sensing
- Climate study
- Disaster alert (Flood monitoring, etc.)
- Geology
- Land resource analysis

New York (from Landast-5 TM)

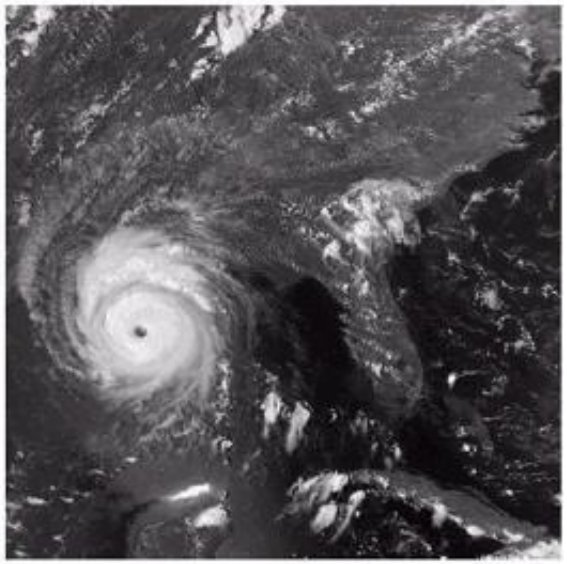
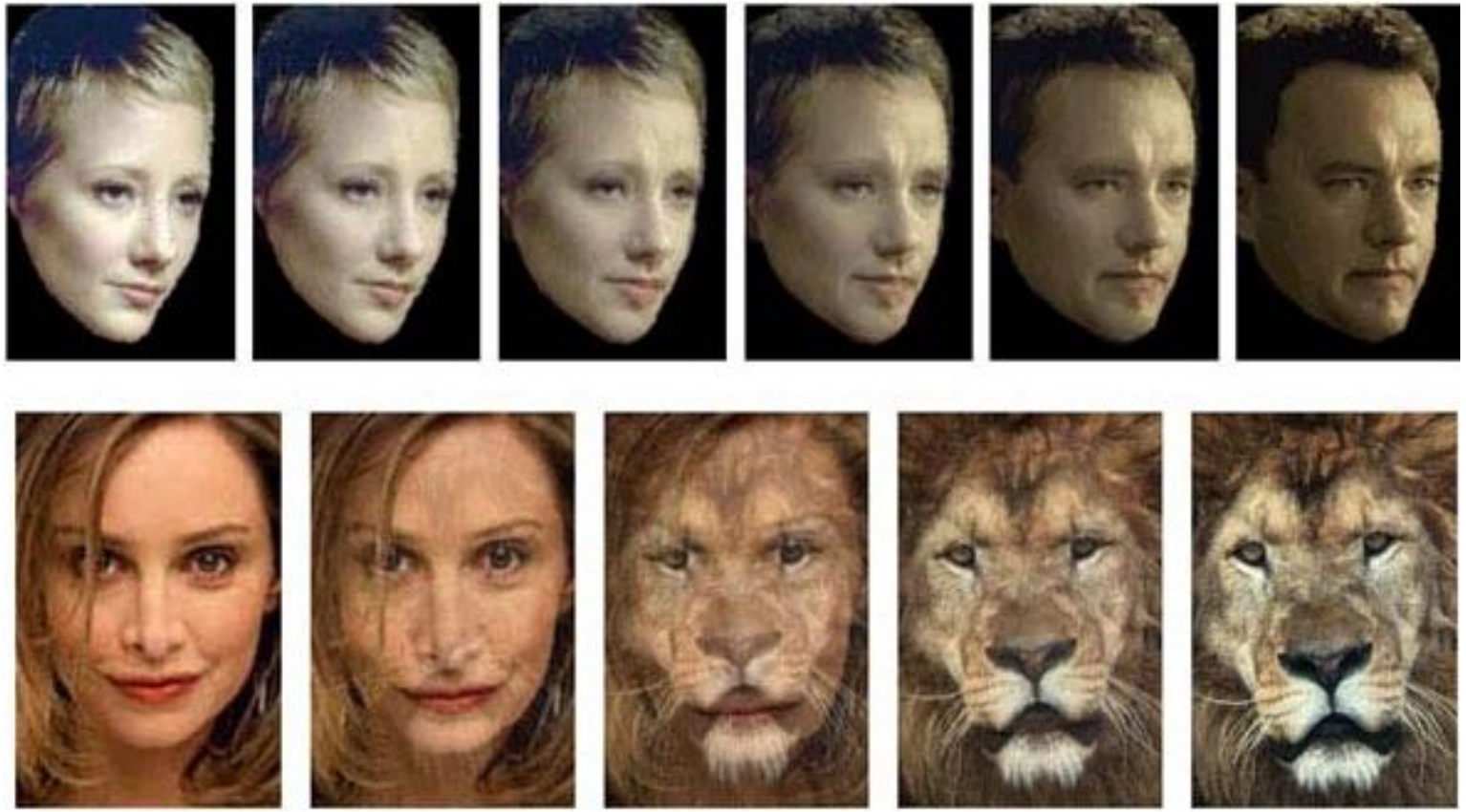


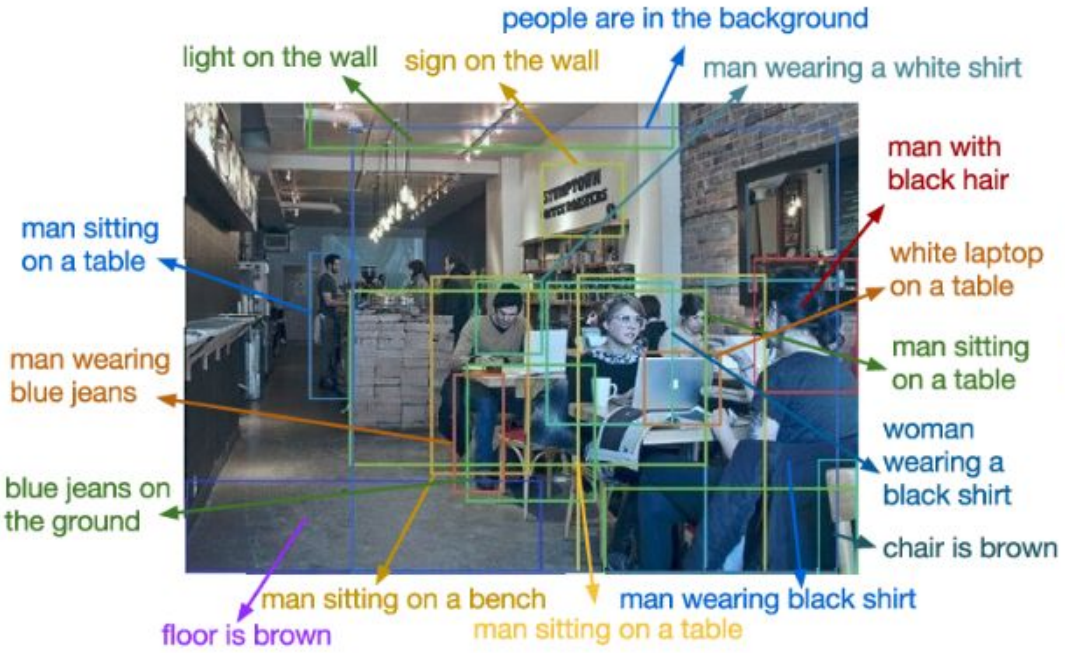
Image Morphing

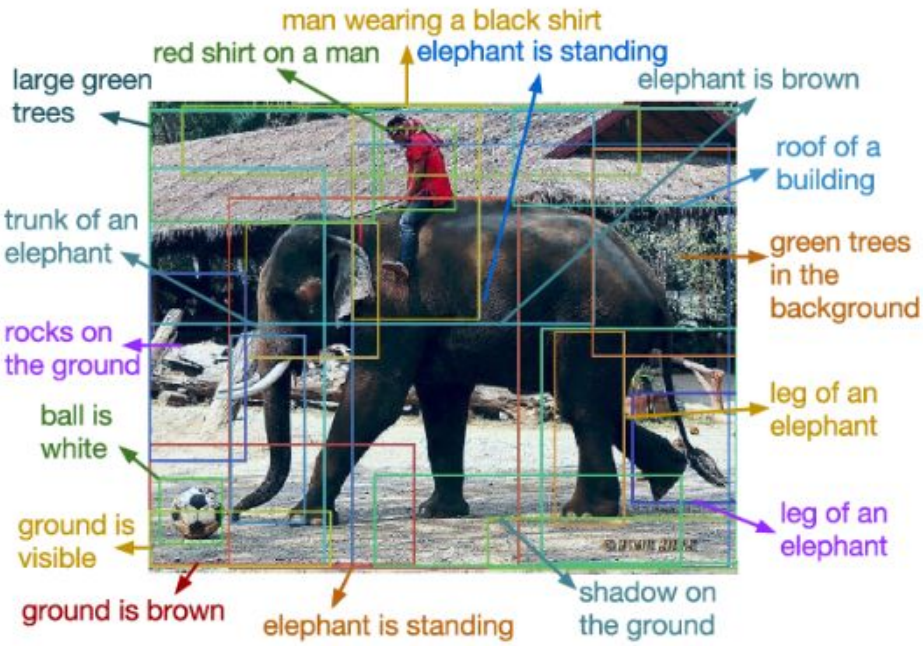


Morphing Video: From Washington to Trump

New Horizons

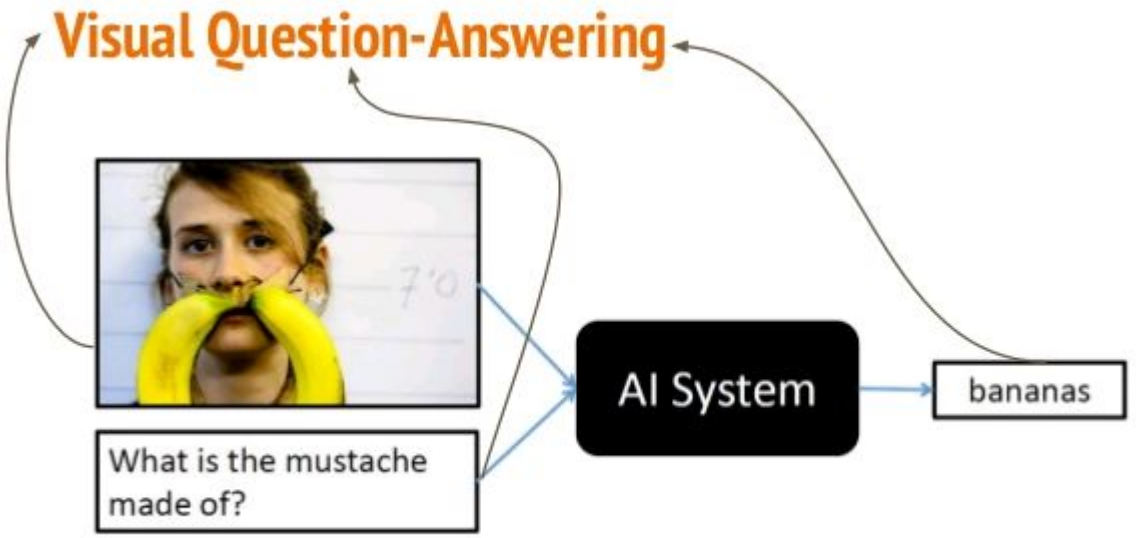
- Visual Captioning





Johnson, J., Karpathy, A., & Fei-Fei, L. (2016). Densecap: Fully convolutional localization networks for dense captioning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4565-4574).

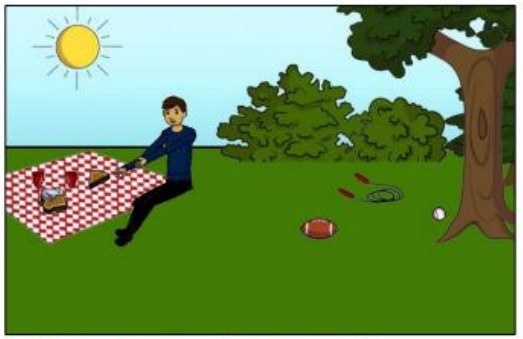
New Horizons



Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Lawrence Zitnick, C., & Parikh, D. (2015). Vqa: Visual question answering. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2425-2433).

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

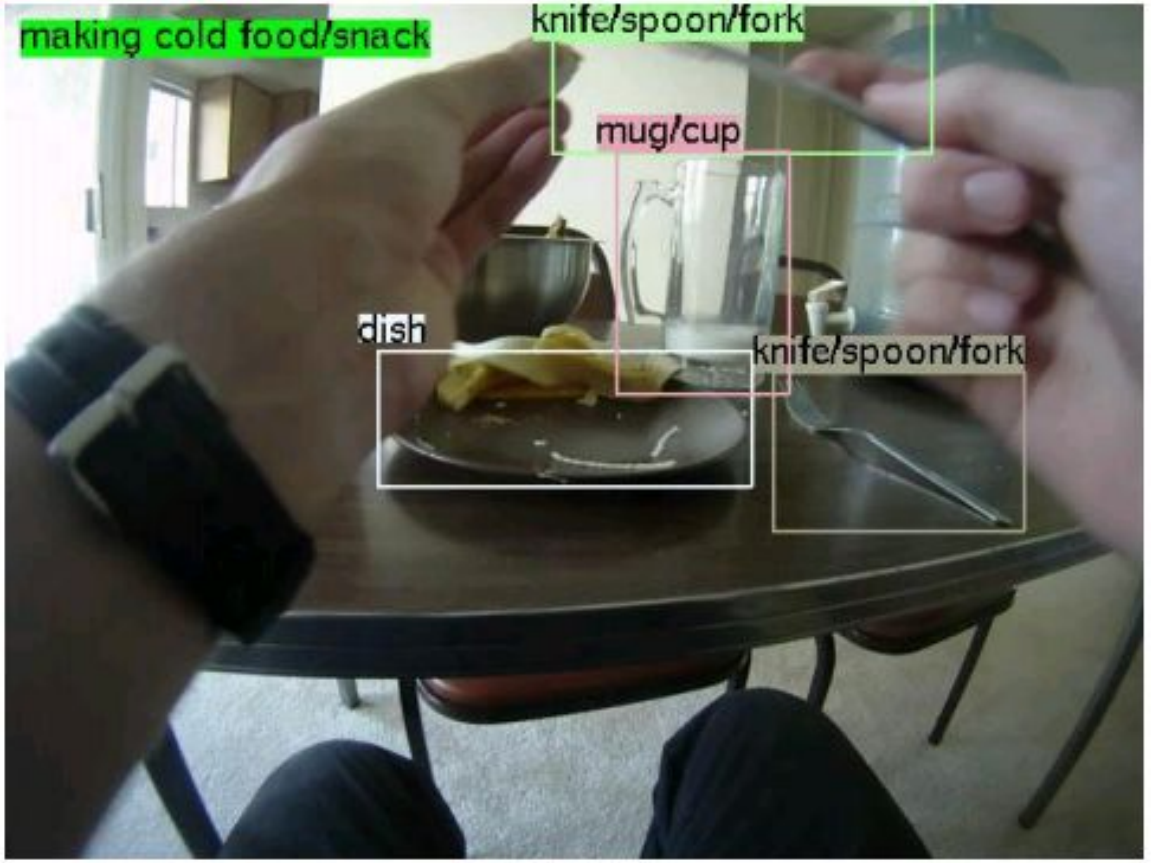


Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

New Horizons

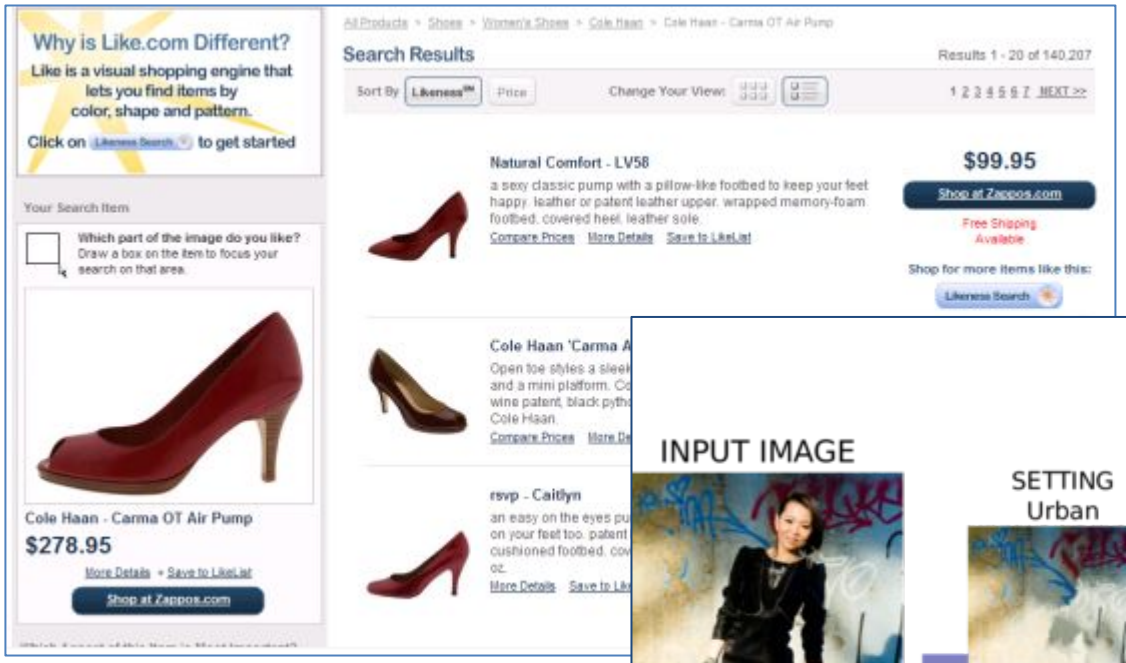
- Egocentric (“first person”) vision



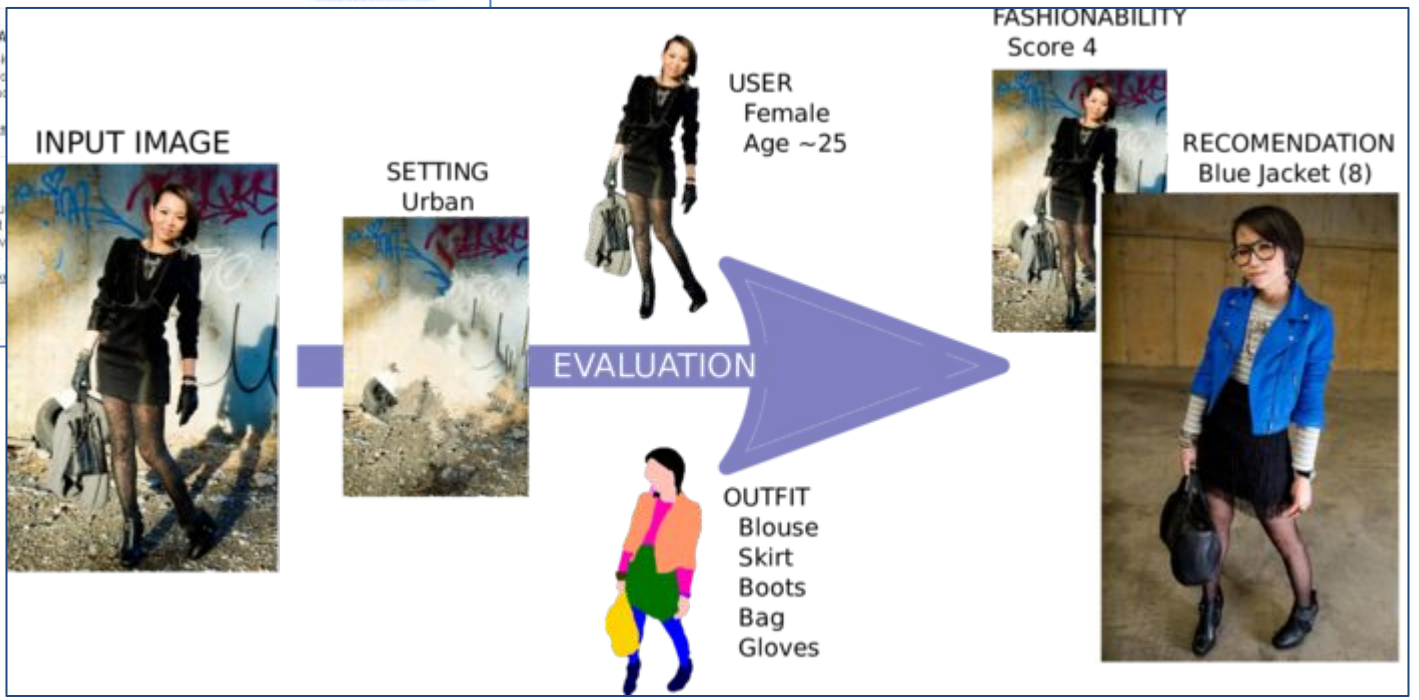
Activities of Daily Living – Hamed Pirsiavash et al.

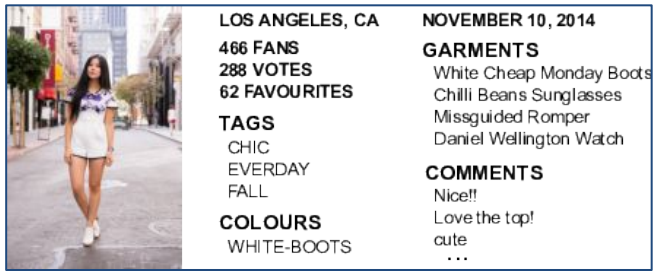
New Horizons

- Fashion Mining and Recommendation



Like.com





Simo-Serra et al. (CVPR2015)

Why is it challenging to understand visual data?

- **Ill-posed problem**: Real world much more complex than what we can measure in images
  - 3D ⇒ 2D
- Can "reconstruct" but **impossible** to "invert" image formation process

Many problems…

Illumination

Object pose

Clutter

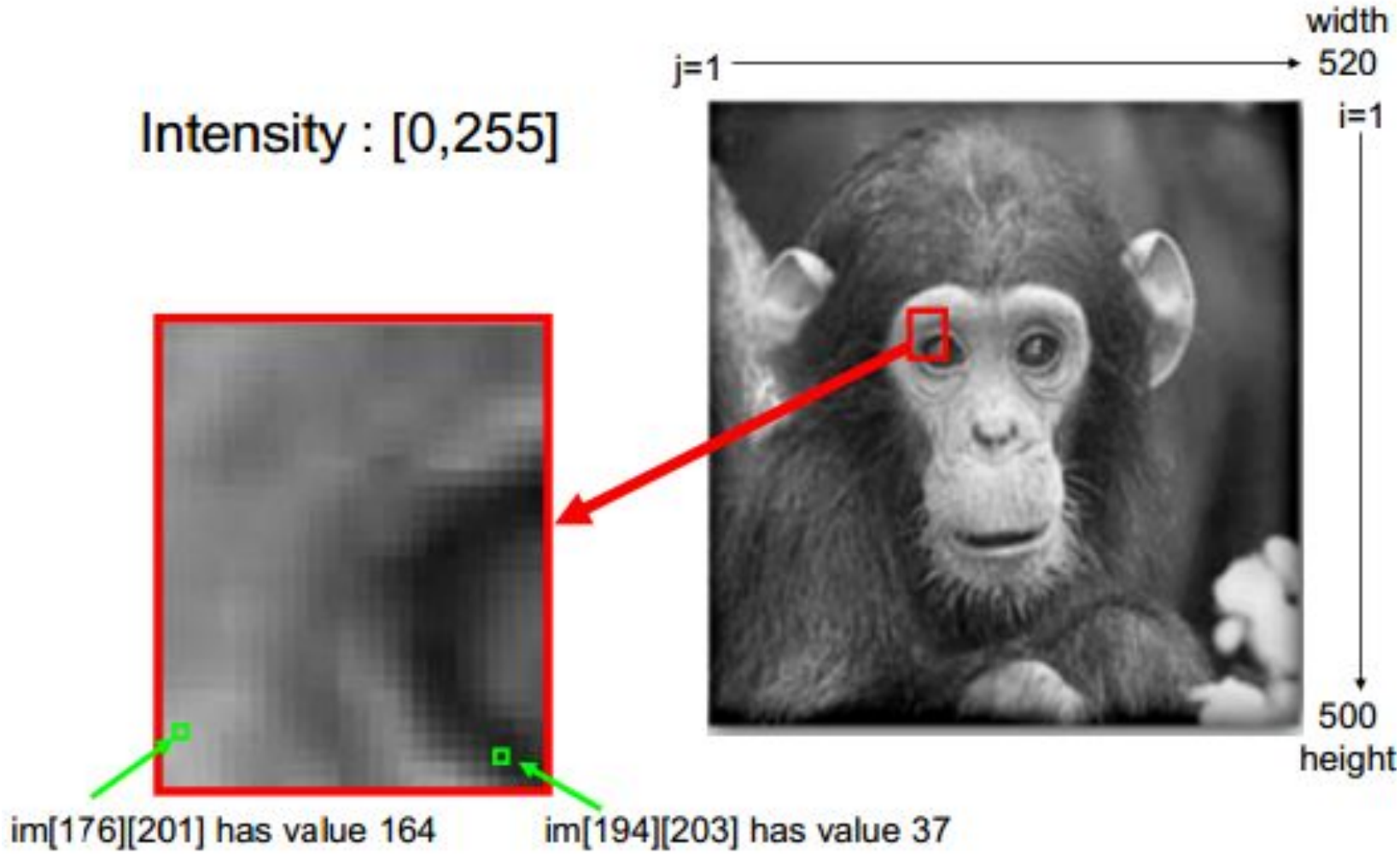
Occlusions

Intra-class appearance

Viewpoint

Topics

Digital images



Digital images

Color images, RGB color space

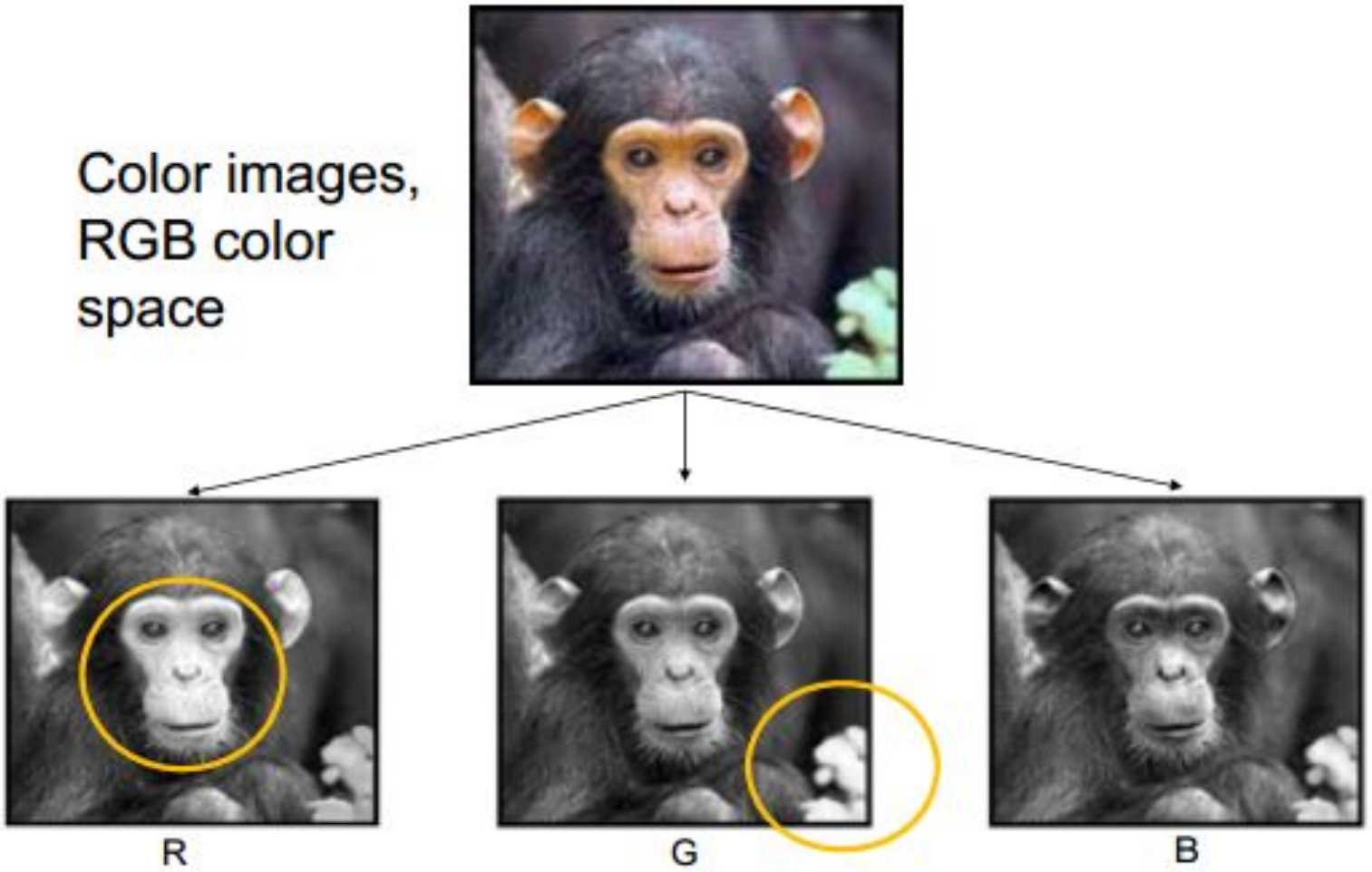


Image Processing Tasks

- Filtering

Observed Image

Enhanced Image
(sharpening)

Image Processing Tasks

- Restoration

Degraded Image

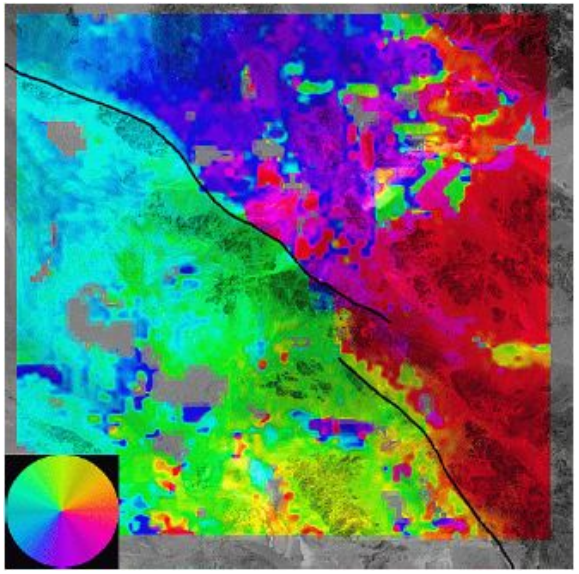
Restored Image

Degraded Image
(salt and pepper noise)

Noise reduced Image
(after median filtering)

Image Processing Tasks

- Segmentation



Ground replacement due to earthquake in California, 1992

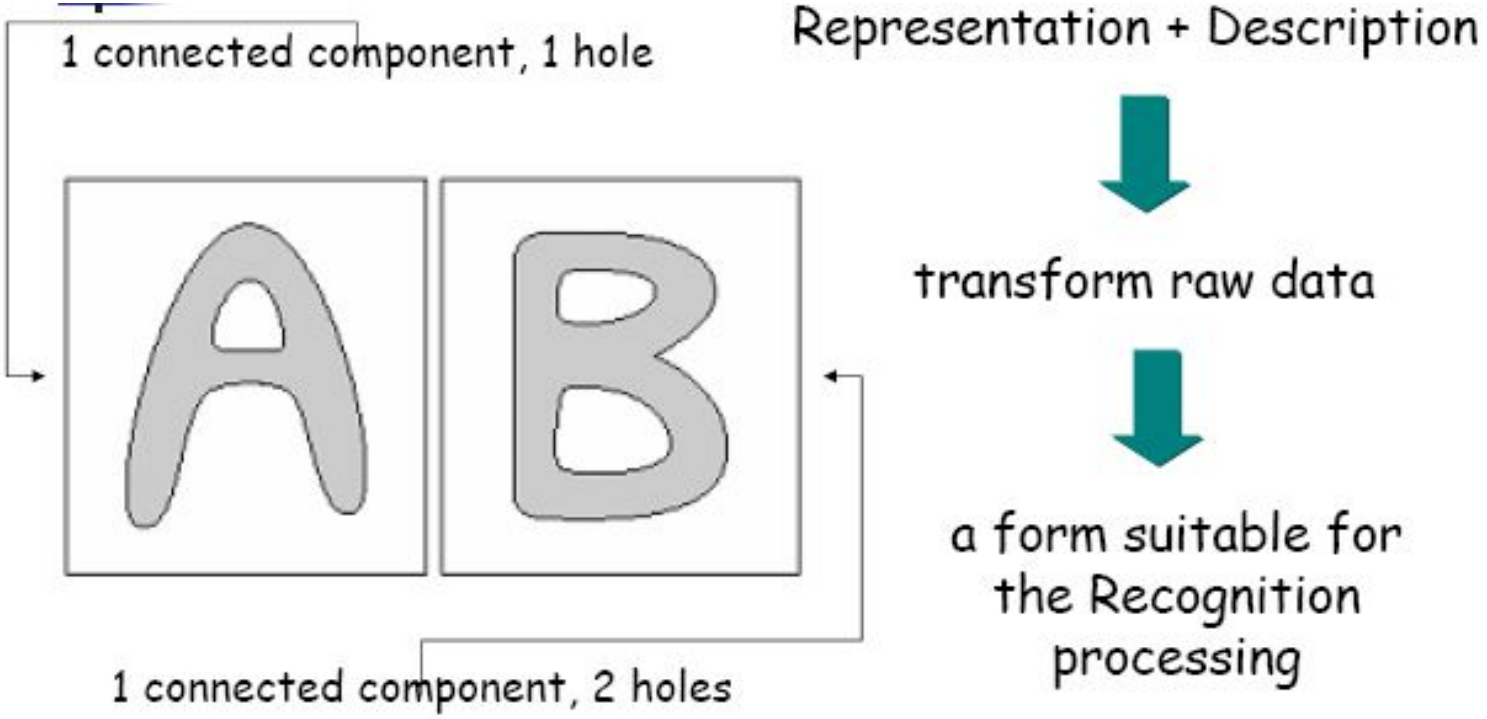
Colour Image Processing

- Color Processing

Humans are “still” smarter than computers

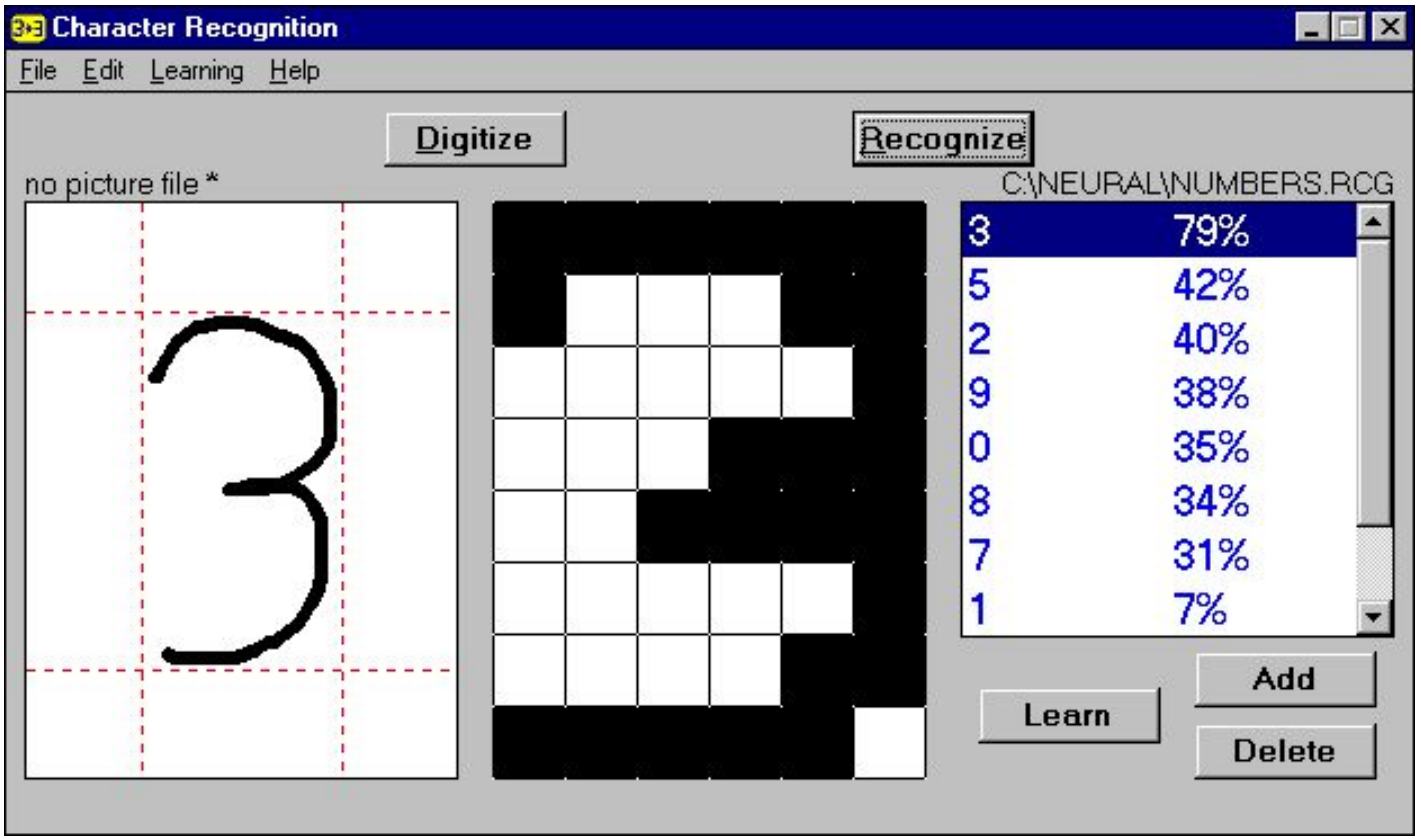
Moving up a notch…

- Representation & Description



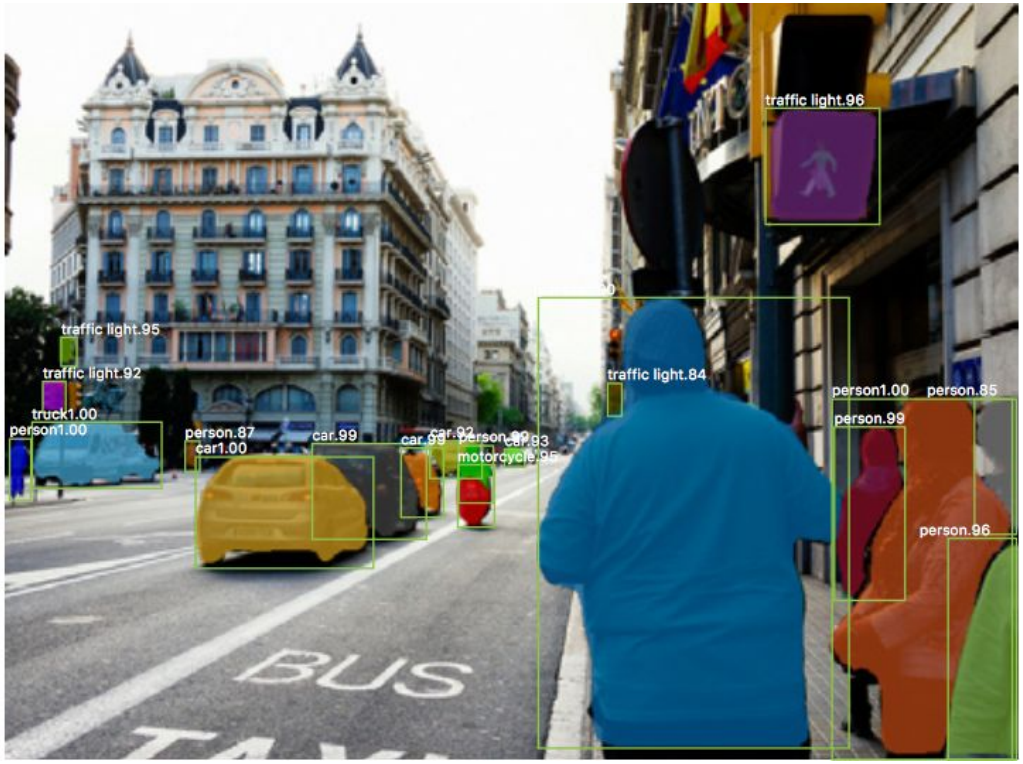
…to “higher level” tasks

- Recognition & Interpretation



…to “higher level” tasks

- Recognition & Interpretation



He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). **Mask r-cnn**. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

…to “higher level” tasks

- Recognition & Interpretation

Zhao, H., Qi, X., Shen, X., Shi, J., & Jia, J. (2018). Icnet for real-time semantic segmentation on high-resolution images. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 405-420).

https://www.youtube.com/watch?v=qWl9idsCuLQ

Summary

- Be aware of subject matters
- Prepare software tools for lab
- Image/video processing deals with manipulation
- Computer vision analyses and derive understanding
- Areas include “measurement”, “perception and interpretation”, “search and organization”