Semester One 2021 ENGN 6528
Computer Vision

Miaomiao Liu

Overview

- An introductory-level course on computer vision, designed for first-year postgraduate students with electronic engineering, mechatronics, computer science, mathematics, or physics backgrounds.

- We do not assume background knowledge in image and vision processing.

- However, students are required to be familiar with calculus, linear algebra, matrix computation, some probability theory, linear system (signal processing or system and control) theory, and desirable: basic programming skills (in Matlab, Python, or C/C++)

Course Cohort

- Master of Machine Learning and Computer Vision
- Master of Computing / (advanced)
- Master of Engineering
- Master of Applied Data Analytics
- Graduate Certificate of Machine Learning and Computer Vision

CECS Course Representatives

Why become a course representative?

- **Develop skills sought by employers**, including interpersonal, dispute resolution, leadership and communication skills.
- **Become empowered**. Play an active role in determining the direction of your education.
- **Become more aware of issues influencing your University** and current issues in higher education.
- **Ensure students have a voice** to their course convener, lecturer, tutors, and college.

CECS Course Representatives

Roles and responsibilities:

- Act as the official liaison between your peers and convener.
- Be creative, available and proactive in gathering feedback from your classmates.
- Attend regular meetings, and provide reports on course feedback to your course convener and the Associate Director (Education).
- Close the feedback loop by reporting back to the class the outcomes of your meetings.

For more information regarding roles and responsibilities, contact:

ANUSA CECS representatives

Sandy Ma and Swatantra Roy: sa.cecs@anu.edu.au

ANUSA President

Madhumitha Janagaraja: sa.president@anu.edu.au

CECS Course Representatives

Want to be a course representative? Nominate today!

Please nominate yourself via the **CECS Course Representative EOI form** **by midday 1st March 2021**. *You are free to nominate yourself whether you are currently on-campus or overseas.*

Using the following link to nominate yourself

This semester, students are asked to nominate themselves via an eform by midday 1st March 2021.

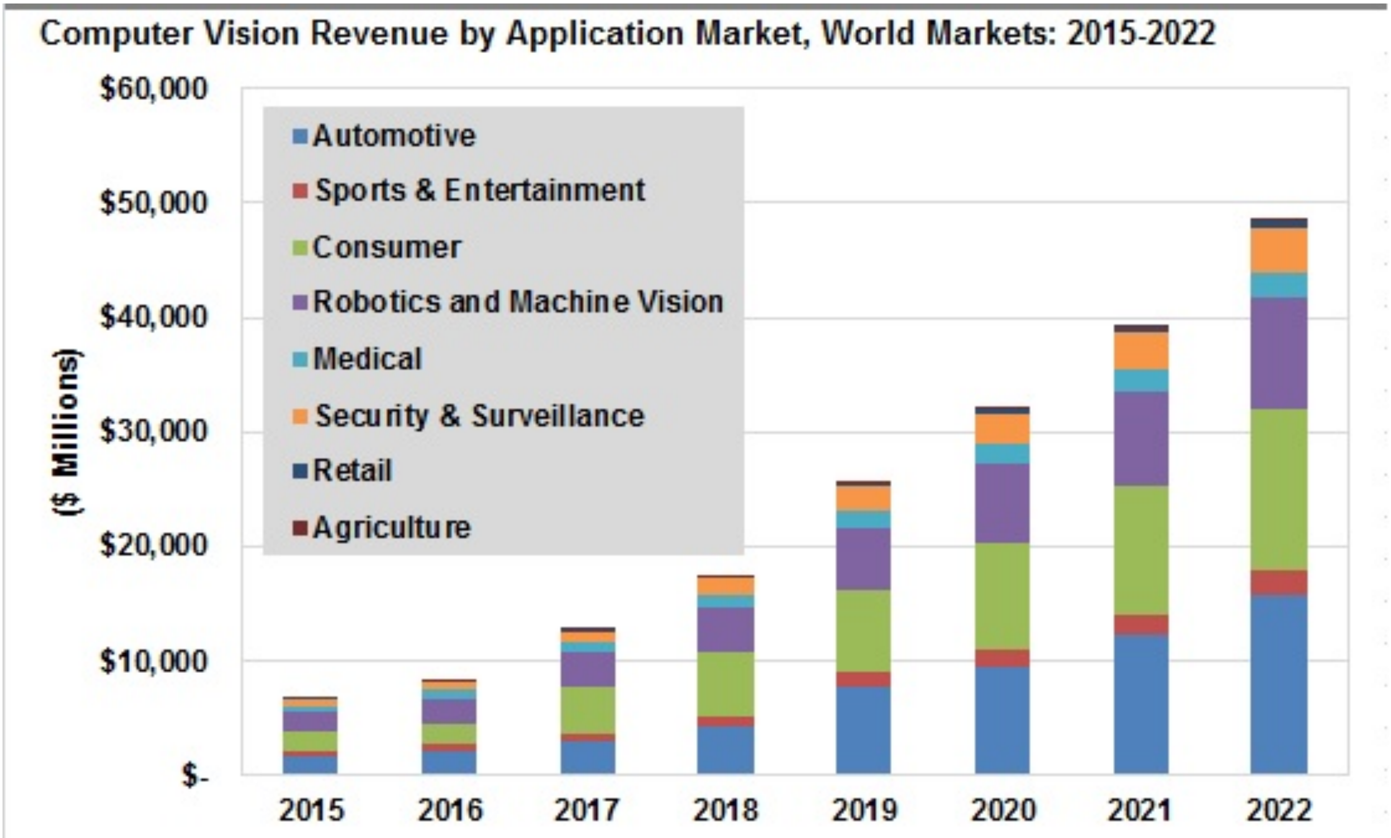
https://anu.au1.qualtrics.com/jfe/form/SV_8H50LYu50DbvXiR

You will be contacted by CECS Student Services, Employability and Experience by 5th March with the outcome of your self-nomination.

All course representative meetings will be held via Zoom in Semester One 2021. There will be three meetings this semester, meeting details will be provided to course representatives shortly.

Best time for CV and ML

- Over 50,000 new jobs tagged “computer vision” listed on LinkedIn at moment.
- Investment has grown 20-fold since 2012



Source: Tractica

Sponsors for Major CV Conferences

- Global tech firms sponsor major computer vision conferences

Today's Lecture

- What is Computer Vision?
- Administrative information about the course

What is Computer Vision?

Why to study Computer Vision?

What can Computer Vision do?

What do you expect to learn from this course?

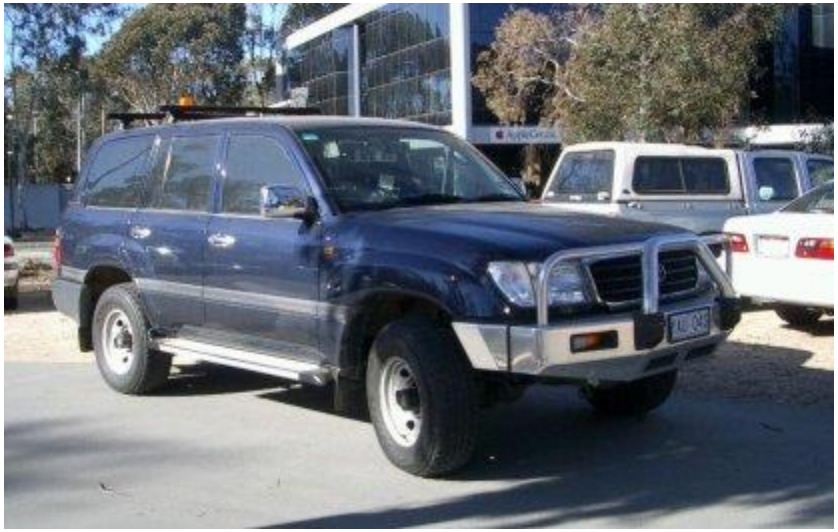
Synonym

Computer Vision

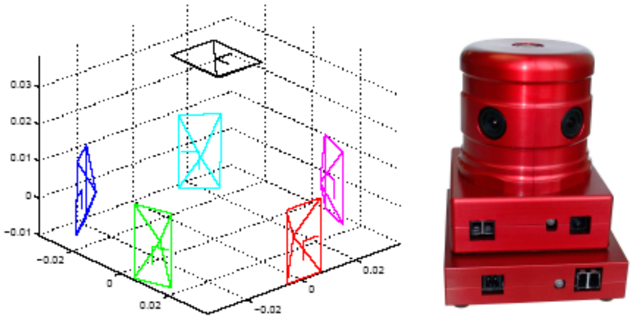
- Image Understanding
- Machine Vision
- Robotic Vision

Previous Computer Vision Projects @ ANU

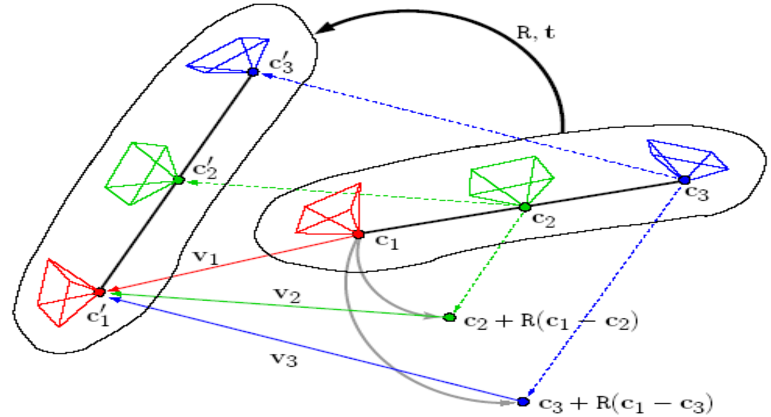
Autonomous vehicles / self-driving cars



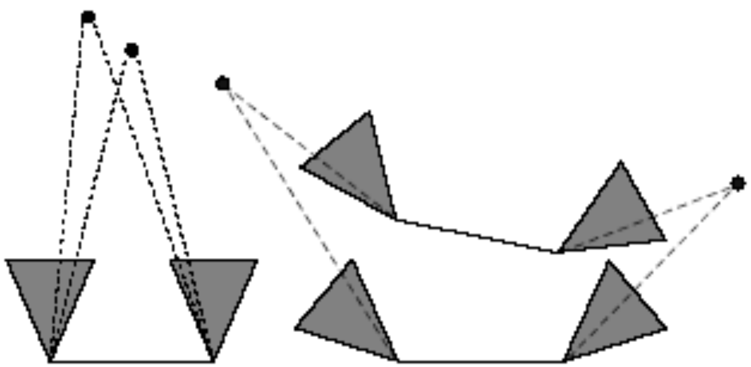
Multiple Camera City Modelling



(a) (b)



Multi-camera system



(a) (b)

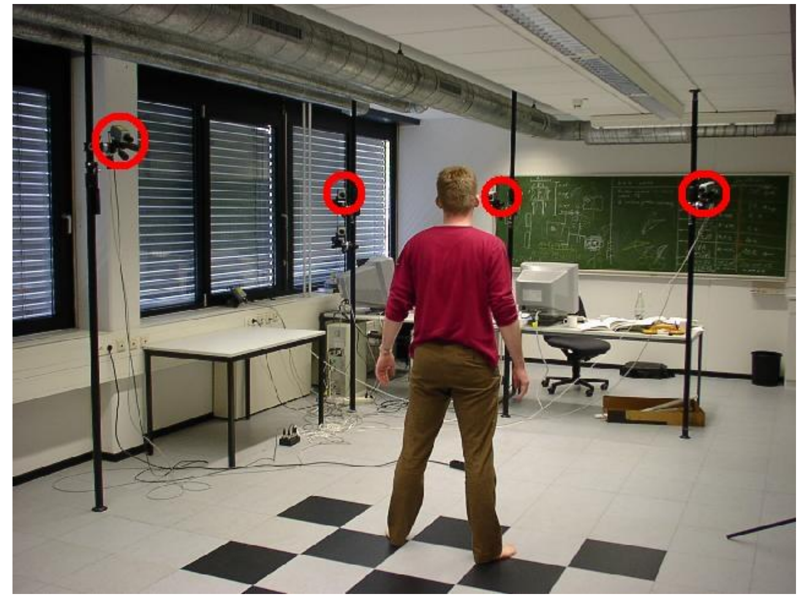
(a) Overlapping vs. (b) non-overlapping Multi-camera systems

Images from one camera cluster

Kim and Li, Hartley, CVPR 08

Human Body Shape and Motion 3D Capture

- Marker-free
- No need to wear a special suit or gloves
- No dedicated hardware
- Standard video cameras
- Cost-effective

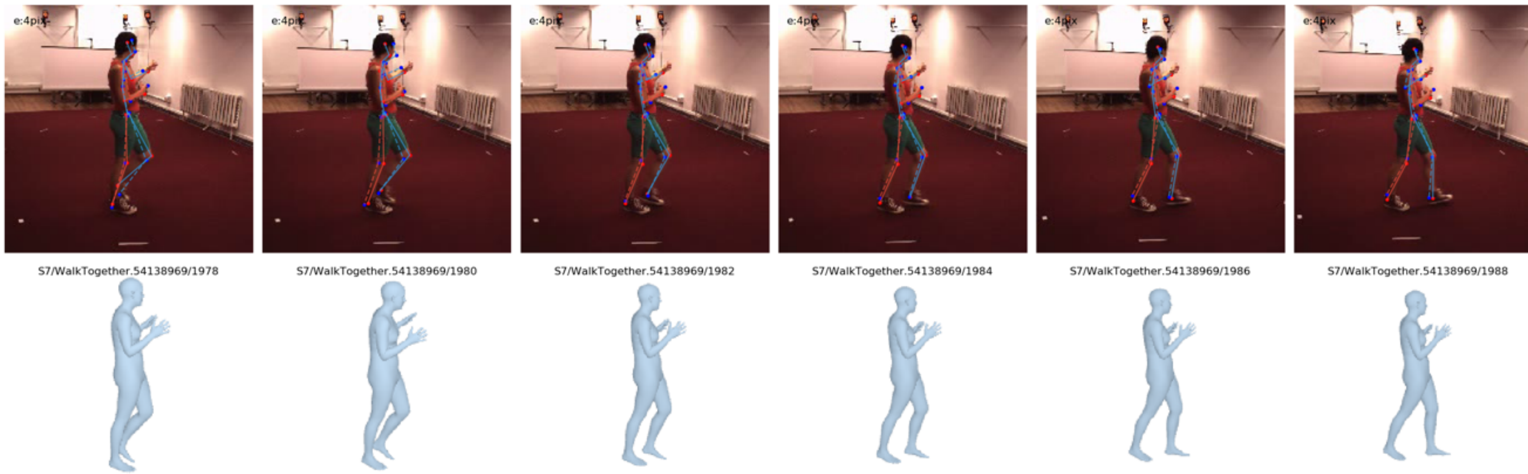


Motion Capture Environment (Image courtesy of Christian Theobalt)

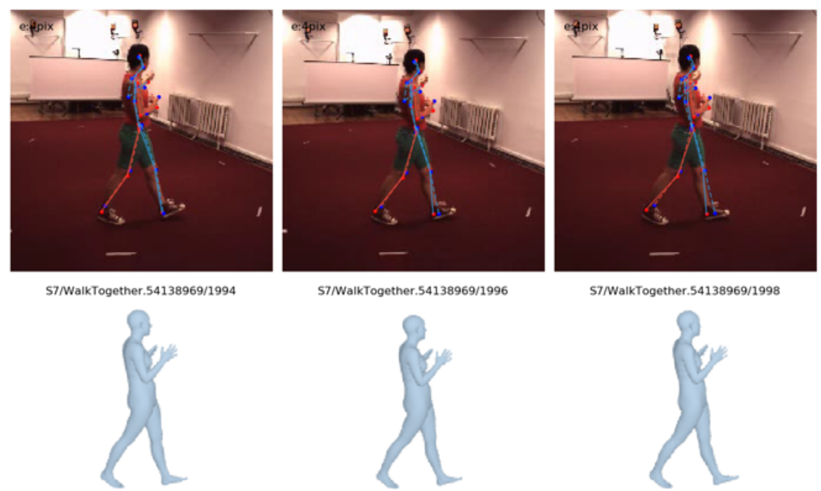
Human Body Shape and Motion Prediction

- Predict Human Motion Dynamics from a Video sequence

Historical Data:



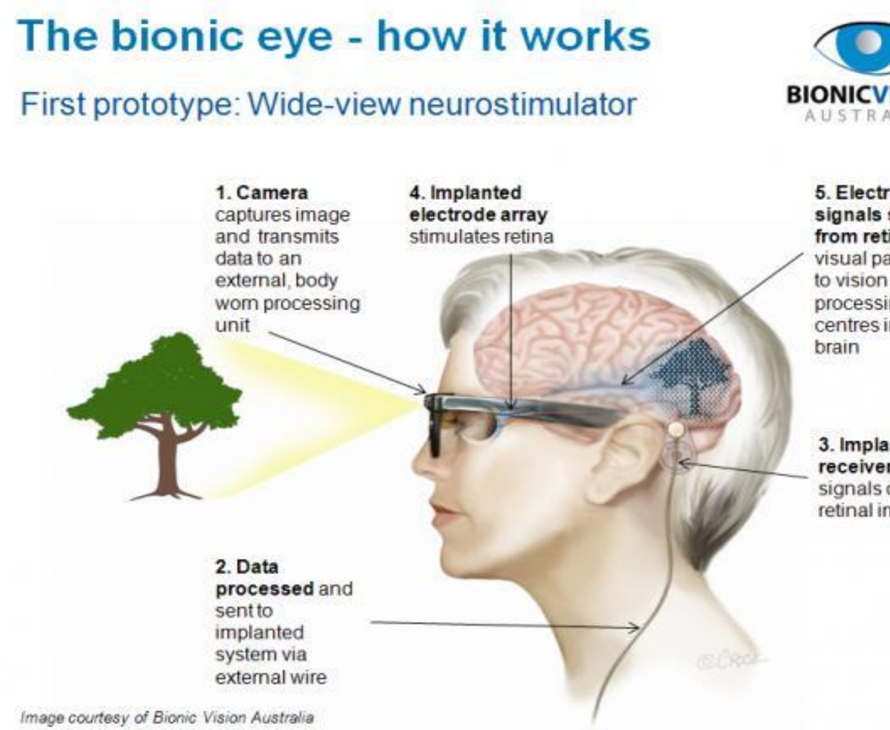
Predict Future Human Motion:



Mao, Liu, Salzmann, Li, ICCV19

Australian Bionic Eye

- 2020 Summit Project
- Budget $50M for first 4 years
- 5 BVA Members and 2 contributing universities
- Officially started July 2010



Patient trial in Canberra 2014

Ms Ashworth had her first 'unplugged' trial in *Canberra*, 2014.

https://www.abc.net.au/news/2014-04-30/bionic-eye-patients-star
first-navigation-tests/5422174

Australian Centre for Robotic Vision

- $25M government funding for 2014- 2020
- 4 Centre Nodes (QUT, ANU, Adelaide, Monash)
- 13 Chief Scientists national-wide

Oxford University, Oxford, UK
Imperial College, London, UK
ETH, Zurich, Switzerland
INRIA, Rennes, France
Georgia Tech, Atlanta, Georgia, USA
QUT, Brisbane
NICTA, Sydney, Brisbane, Canberra, Melbourne
UoA, Adelaide
Monash, Melbourne
ANU, Canberra

What is Computer Vision?

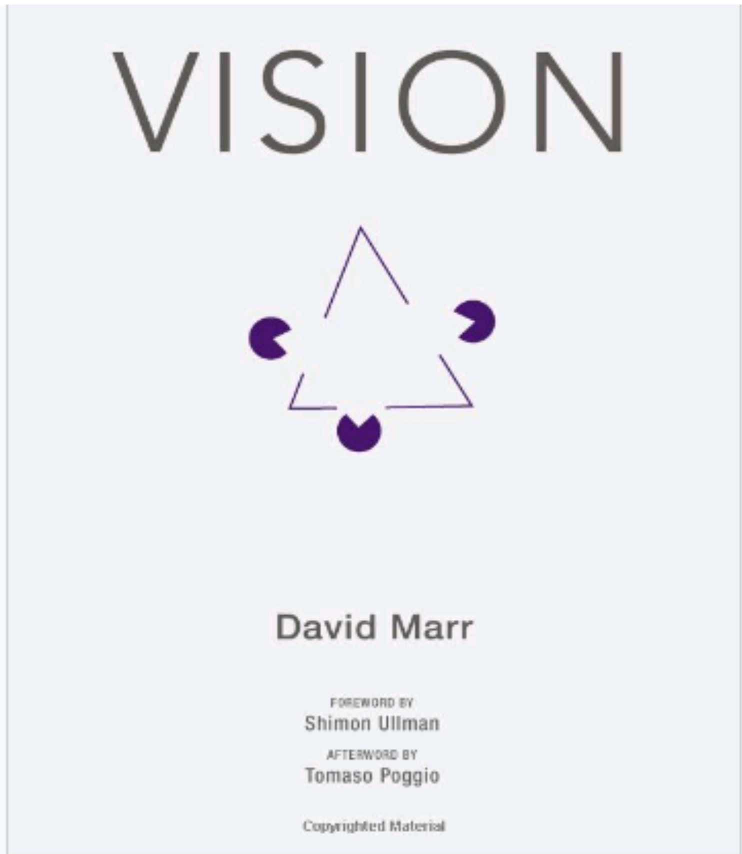
- (aka machine vision, robot vision, image understanding, scene recognition)
- and **applications of Computer Vision**

What is Computer Vision?

- Humans see things very easily.
- Can a computer do the same task?
- **How to teach a computer (or robot) to see (and understand what it sees)?**

What is Vision?

Is [the art] ``*to know what is where, by looking*.''

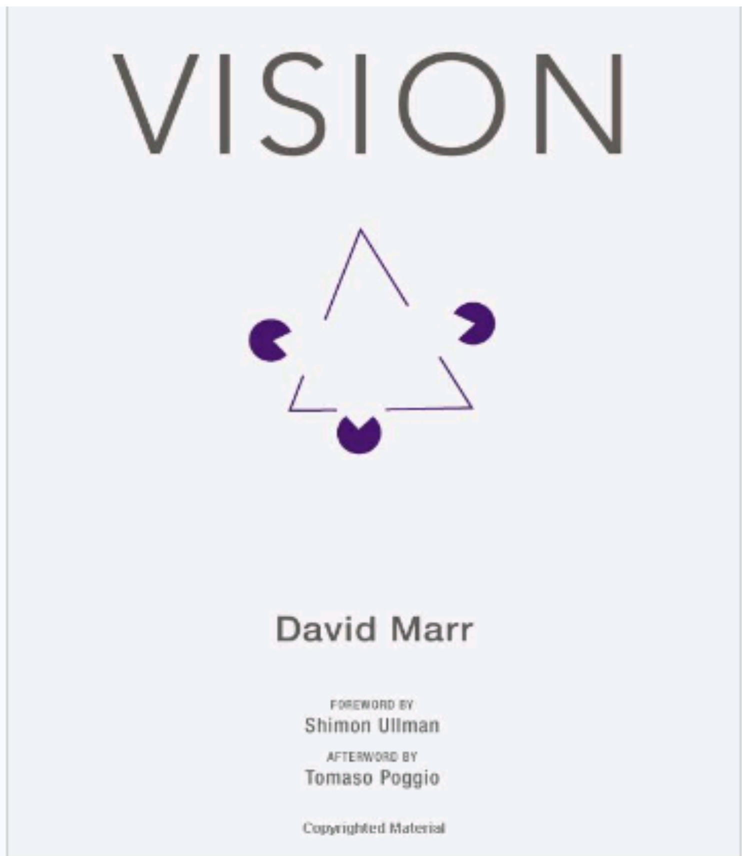


Published in 1982

David Marr (1945 --1980)

What is Vision?

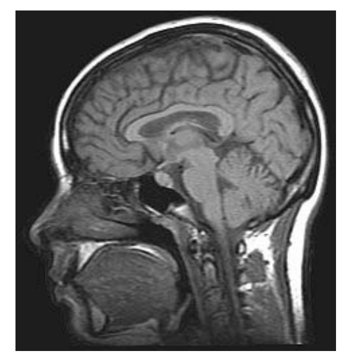
The process of discovering from images what is present in the world and where it is.



David Marr (1945 --1980)

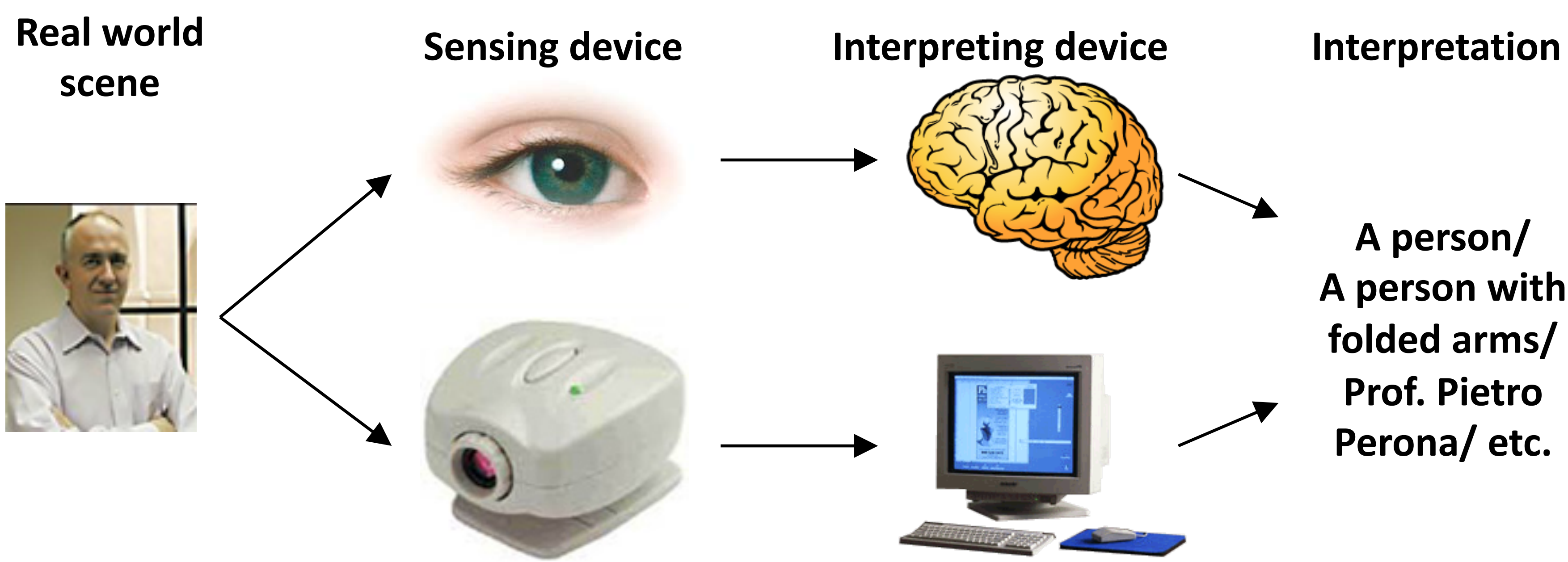
What is Computer Vision?

- Computer Vision is concerned with ***<u>the theory</u>*** of building artificial vision systems that obtain useful information from images.
- Image data taken by many forms, e.g. *video sequence, depth images, multi-dimensional data from a medical scanner etc.*



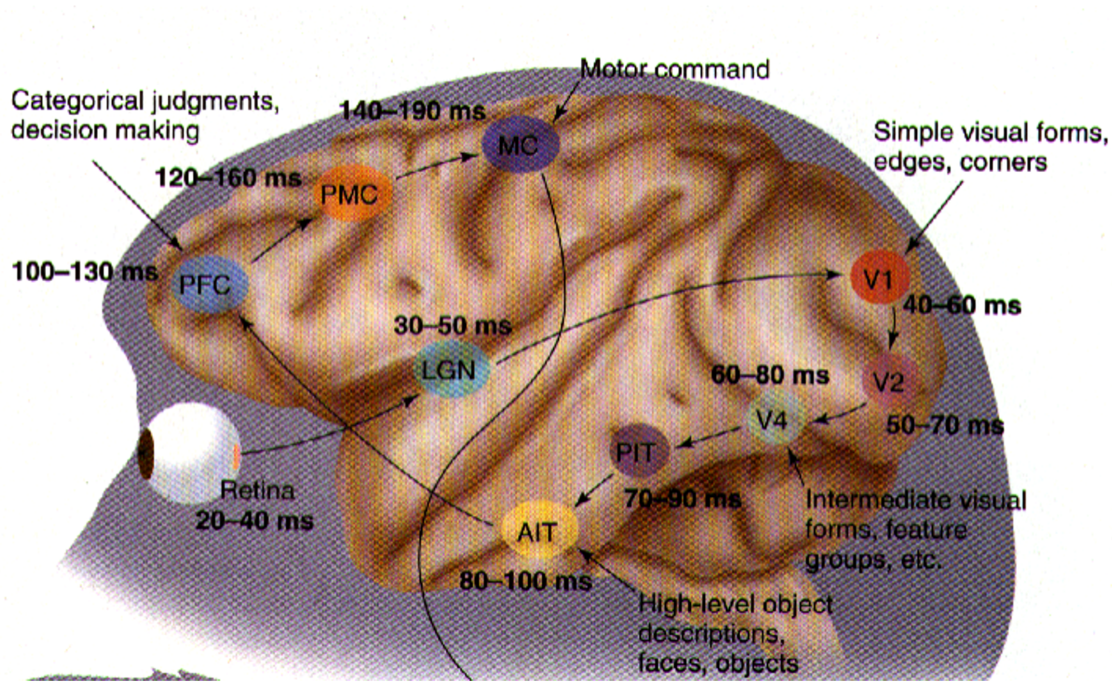
Computer Vision Problems

- Make a computer see and understand images.
- We know it is physically possible – we do it every day and effortlessly!



Human Visual System

- Vision is the most powerful sense.

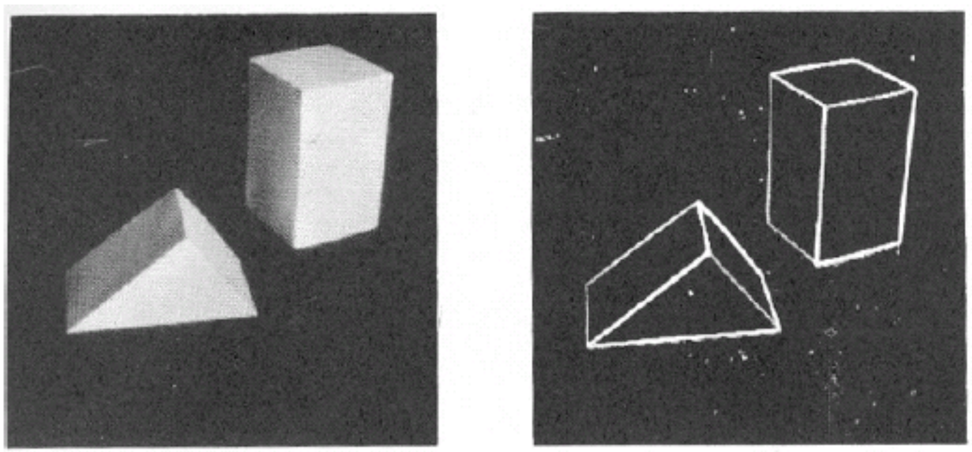


[Thorpe et. al.]

- Over 70% information are from eyes.
- Around 2/3 of our brain cortex area are devoted to processing the signals captured by our eyes.
- The visual cortex alone has around $O(10^{11})$ neurons!

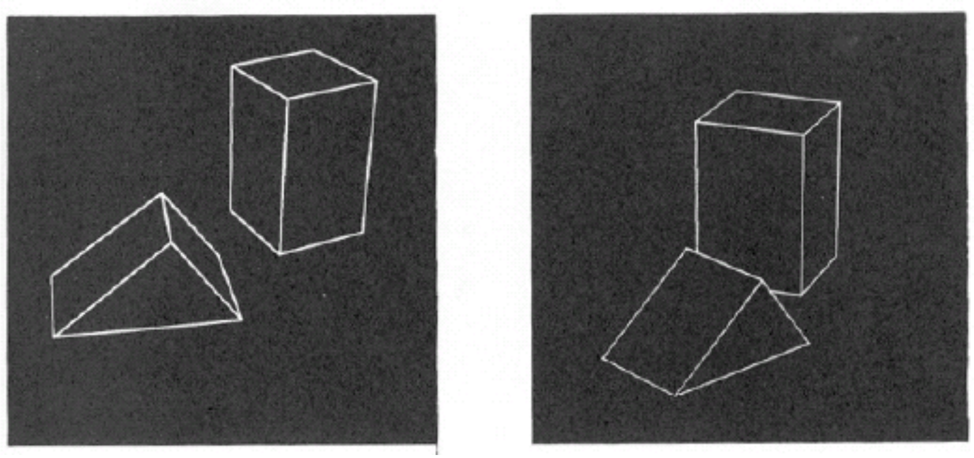
Brief history of computer vision research

Origins of Computer Vision



(a) Original picture.

(b) Differentiated picture.



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

An (overly) optimistic start

In 1966, Marvin Minsky at MIT asked his undergraduate student Gerald Jay Sussman to "***spend 3 months in this summer linking a camera to a computer and getting the computer to describe what it saw***".

Now, more than fifty years, we know the problem is significantly more difficult than a 3-month student project.

Brief history of computer vision

1963: Robert's thesis

1966: Minsky assigns computer vision as an undergrad summer project

1960's: interpretation of synthetic worlds

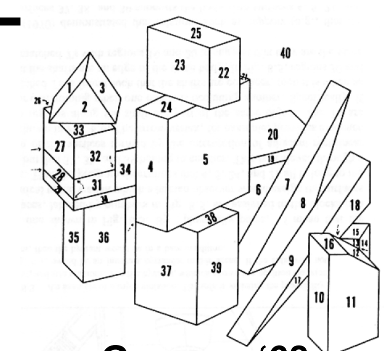
1970's: some progress on interpreting selected images

1980's: ANNs come and go; shift toward geometry and increased mathematical rigor, Marr's book published (after he died).

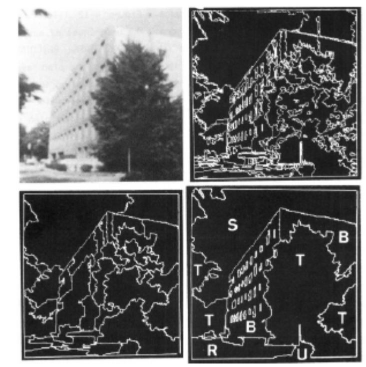
1990's: face recognition; statistical analysis in vogue

2000's: broader recognition; large annotated datasets available; video processing starts; vision & graphics; vision for HCI; internet vision, etc.

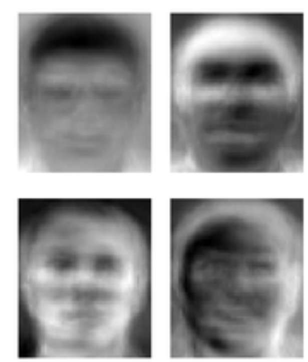
2012: Kinect, big-data, Google Car, … Deep Neural network, Deep learning



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

Current State of the art of CV

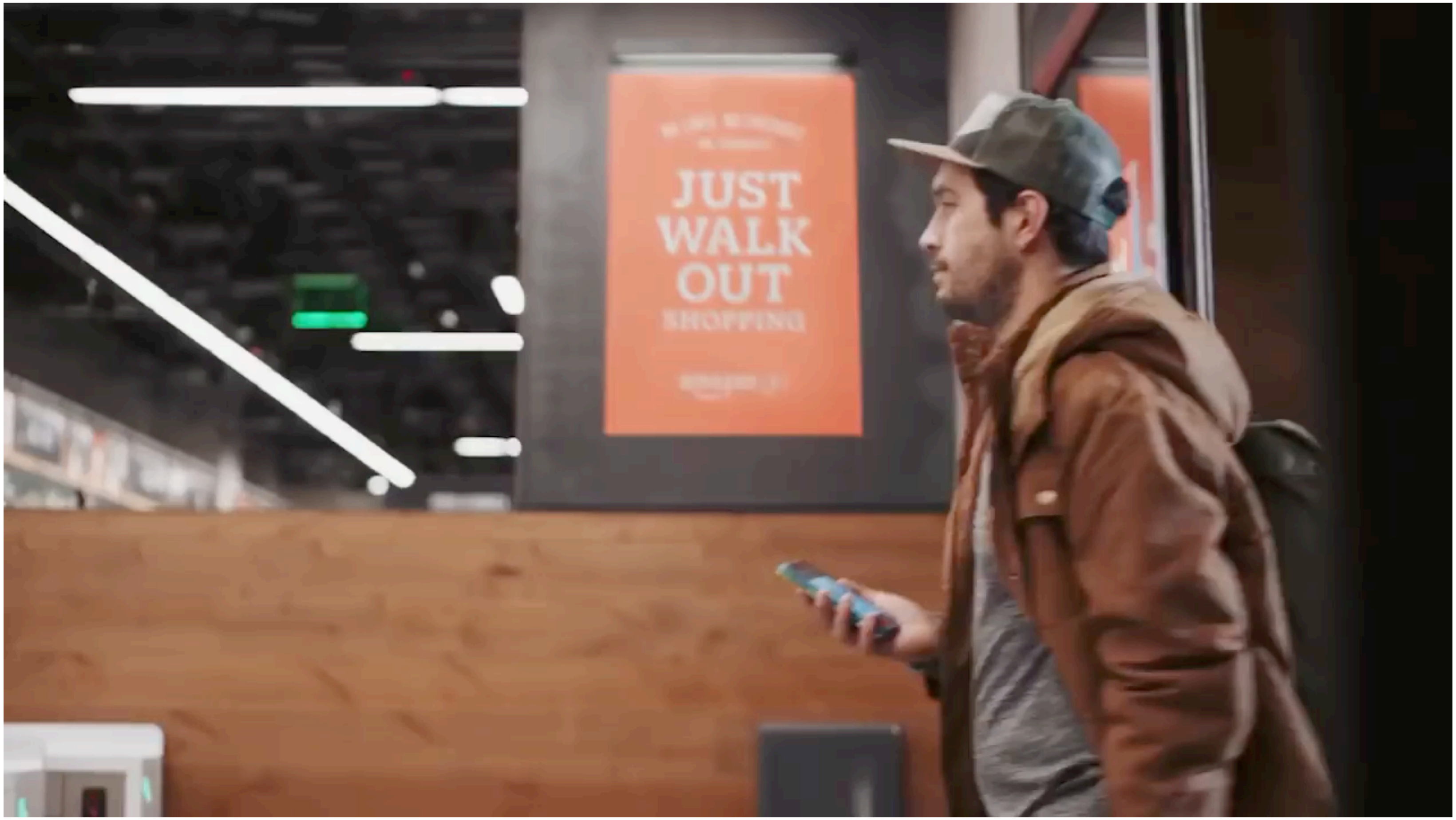
Video Demo

- Google Car
- Amazon Go
- HoloLens- Microsoft

State of the art google self-driving car (GOOGLe)

https://www.youtube.com/watch?v=B8R148hFxPw

Amazon Go



https://www.youtube.com/watch?v=NrmMk1Myrxc

HoloLens (Microsoft)

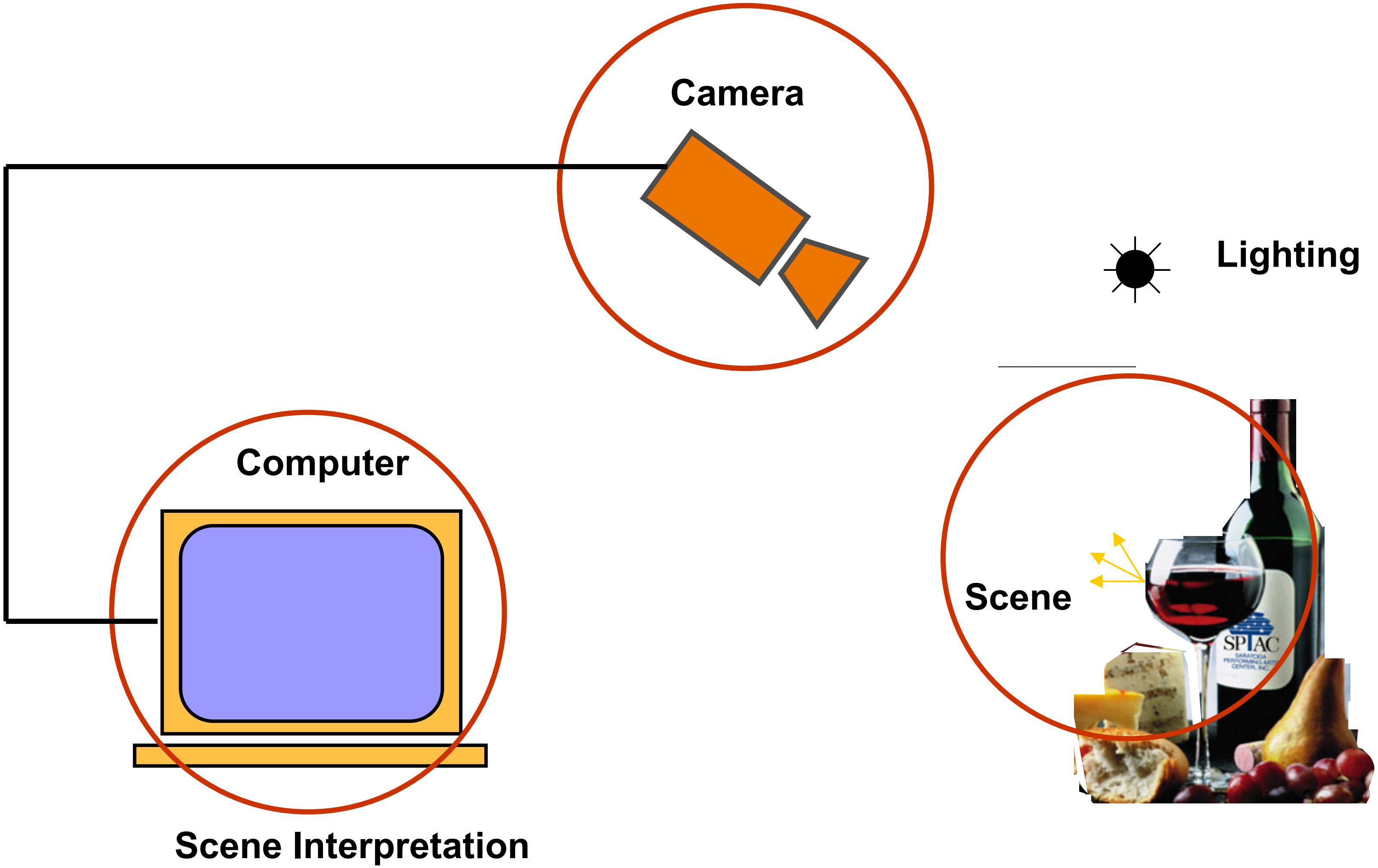
https://www.youtube.com/watch?v=waNNsTI-56k

What makes a **Computer Vision system** ?

- In a **Computer Vision System**, a camera (or several cameras) is linked to a computer.

- The computer interprets images of a real scene to obtain information useful for tasks such as navigation, manipulation and recognition.

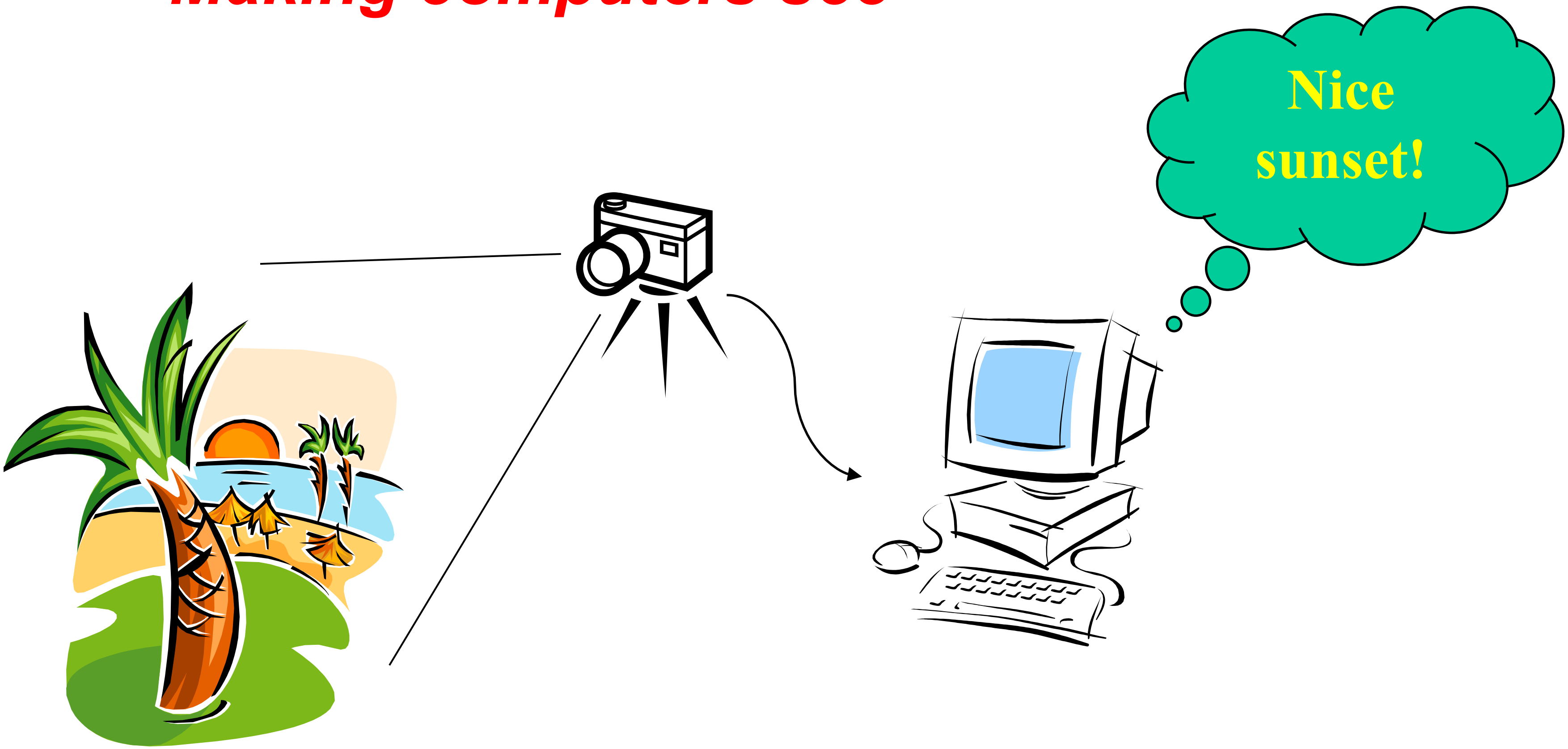
Computer Vision System Example



Srinivasa Narasimhan's slide

- Computer Vision is also known as *machine vision* or *robot vision*

  – ***“Making computers see”***



– ***To ‘see’ is not only to take image, but also ‘to understand’ it.***

Understand the contents of images and videos.

What kind of scene?

Where are the cars?

How far is the building?

…

What we could see

What computer sees

```
01 00 05 00 03 00 02 00 00 03 01 01 01 01 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 02 00 01
03 30 3A 38 39 2D 1D 15 10 0E 0C 0A 0A 0A 09 06 08 07 06 06 05 05 07 07 04 05 04 04 06 02 01 02 02 02 02 07 01 02 02 03
03 22 1B 16 14 0A 08 0B 0A 0D 0B 0B 0C 06 07 05 05 06 06 06 03 07 04 06 05 09 05 04 05 01 04 04 02 03 03 04 02 04 03 02
00 0F 0B 04 10 07 09 07 08 09 09 08 05 08 08 05 09 03 08 05 02 08 08 06 06 04 02 05 03 02 05 05 00 02 02 04 04 00 00 03
00 07 09 0E 0C 07 08 0A 0A 0B 0F 0A 0C 07 06 0B 07 0B 05 0B 08 09 07 03 08 04 04 02 00 04 02 04 00 04 03 08 00 06 09 04
00 0E 0C 09 09 08 08 07 08 09 09 0A 05 08 07 07 07 09 08 0A 08 09 06 0A 03 09 07 06 06 03 05 03 01 06 02 03 07 01 04 04
02 0C 0B 0A 05 08 09 0A 0C 0A 0A 08 0A 0A 06 08 06 06 04 06 02 06 07 04 04 04 06 09 05 05 08 06 04 05 04 06 01 0A 03 02
02 0B 14 0F 0F 0D 0A 0E 0A 0C 0C 0E 0A 0C 0B 09 0A 09 0A 0A 09 0B 0B 05 0C 0C 0A 04 07 06 03 05 07 04 05 03 02 01 06 03
02 10 12 0B 10 0A 0D 0D 0B 0D 0C 0B 0B 0C 0D 0B 0B 0A 0A 0A 0B 0C 17 15 1C 15 0D 08 09 08 05 05 05 04 02 05 04 04 00 04
01 15 0E 10 12 0C 0D 0C 0C 0A 0B 0B 09 0C 0F 09 09 0D 07 0B 08 15 60 5D 61 59 33 0D 0A 07 08 08 05 03 06 07 01 03 05 02
02 12 10 0F 0E 10 10 0B 0C 0F 0F 0E 0C 10 0D 15 10 09 12 11 12 50 68 66 89 71 5E 3F 08 09 0A 09 0A 03 03 02 05 05 04 02
01 11 12 0C 11 13 10 10 0B 10 0F 0C 11 11 13 0D 0F 0D 0D 0B 25 7A 7F 79 6D 80 6E 54 0C 0D 09 0A 06 04 02 05 00 05 04 03
01 10 0F 0D 12 0E 10 0E 0F 13 13 11 13 17 11 0F 14 11 11 14 39 84 88 7E 8C 73 7A 5C 1E 05 0A 0F 0E 0C 05 02 04 03 06 05
02 0F 15 0D 18 11 0D 11 14 10 12 12 14 19 13 17 13 16 16 20 73 68 87 89 93 8B 83 69 43 07 0A 12 0A 0B 06 06 03 04 05 03
02 13 14 14 16 11 13 13 17 12 17 17 28 1E 1A 17 19 14 12 4F 7D 74 85 91 93 8C 7F 6F 5F 0B 09 12 0D 0C 02 04 07 04 05 04
00 0F 16 0F 13 12 10 1D 12 21 15 1E 21 1F 1C 1D 2D 1A 2D 7C 7A 95 6B 30 48 62 87 71 5C 0A 08 11 0C 09 04 04 02 06 04 03
00 10 1C 10 11 1A 0D 1A 1A 25 28 33 30 26 2B 3E 29 35 6C 83 5E 7B 94 8A 5A 3D 42 76 5C 13 08 13 0F 0C 04 04 01 05 05 03
01 12 17 1A 19 18 15 20 29 20 3F 1F 37 29 39 49 24 33 8F 93 B4 AE 79 42 39 73 7D 89 46 12 06 12 12 0F 08 03 03 03 04 03
01 13 20 0F 14 26 1B 18 20 2F 3D 3E 42 3B 45 2E 48 70 96 9F 96 6B 24 0F 22 4B C3 A4 3F 4F 0C 18 16 0F 05 05 08 05 05 04
00 19 1C 13 13 21 1D 12 18 47 3D 47 45 3A 27 3B 33 A8 A6 91 81 4B A1 75 4B AC A1 B5 79 0C 0B 13 0F 0B 02 03 06 07 07 04
00 1B 1D 1C 1C 1C 1B 1B 1E 55 49 49 36 28 2A 24 9F AD AC AA B1 9C 8D 5F 3E 98 B7 B7 A3 31 11 14 0A 0D 04 08 07 07 07 06
02 21 18 15 16 1D 15 18 1E 36 5B 29 2C 19 29 4F AF BC AF AB 9E A1 97 82 70 9F AE AD A5 92 16 10 07 0E 0A 0C 08 05 0B 05
01 17 1B 1A 1A 2B 1B 2A 32 34 46 2C 1B 26 4C 40 BA BB B5 AE 95 94 84 7A 8A 9A B9 BB AD 9C 8A 15 09 09 05 0B 0D 0F 0B 07
00 1A 18 1C 1E 27 21 1D 3F 4E 32 25 1B 1B 93 46 AF AB B1 AC A4 93 89 91 86 90 AA 9F 91 97 AD 7F 0C 0B 0E 0B 0C 0C 09 05
00 15 1A 21 1E 2E 1B 23 47 4E 23 21 19 49 99 5B AA AC B7 AF A6 9A 93 8F 85 7F A0 A4 C2 9F 99 4E 09 08 0A 0D 0C 0A 0C 07
00 13 18 21 26 31 28 25 34 4C 1F 2B 1C 8B 9B 42 9B A7 A1 B4 B0 AA A0 9D 92 72 8E 97 71 A7 32 04 0A 0A 0D 0D 09 0D 0C 07
00 1A 1C 21 28 3A 30 26 40 4C 26 18 2C 90 A1 39 A0 97 B8 AA B2 A5 A6 A3 98 76 92 96 98 6D 08 0D 07 08 0C 0B 0E 0D 0D 0A
04 1E 29 1F 27 32 26 2E 41 4A 2C 34 46 8A A5 89 9E A3 B0 B7 AF AB AB 99 97 90 A4 94 85 7C 08 07 07 08 09 09 08 0C 0D 0B
01 1F 29 27 27 2A 2C 36 4D 50 34 42 45 95 9B AA 7E AD B3 AA B2 A8 B2 92 98 8E 9E 8E 44 34 18 05 06 0A 0D 0D 0D 0F 0C 08
00 21 2E 23 29 2C 2A 34 44 5A 39 4F 29 90 9B A5 86 AA B2 B3 AE A0 A3 9C 94 79 43 2B 25 2D 07 0E 05 06 0C 0A 0F 0D 09 0C
00 21 27 20 28 29 2F 2A 44 57 42 31 28 8C 93 A3 AC 60 BA BD B4 AE A8 A2 62 91 5F 52 4F 3F 09 0D 0D 09 0E 0E 0B 12 0B 0B
03 30 2E 2C 29 2A 3B 30 4E 3C 40 40 49 5E AE 9F A4 B1 4E AA AA A0 A4 9C 94 A2 AB A8 93 52 0E 0E 09 0B 0D 10 0C 0C 10 09
00 30 32 2E 36 39 36 24 2D 5A 46 46 68 30 8B 8C A3 AC A5 3E A1 AF A8 82 A4 AC A2 96 71 73 08 10 0B 0B 0B 0E 0F 10 11 0A
00 54 34 1E 3C 3F 3E 29 27 56 38 4C 5C 44 26 94 9A A2 A2 A6 8E 4E 70 99 AC A6 A2 89 7E 5B 11 0E 10 10 17 12 0D 0C 0D 0C
00 4B 30 23 36 44 48 3C 2E 2D 34 35 29 58 5B 0D 36 50 34 52 9C A8 B5 AA B3 AE A0 9C 8C 62 0A 12 14 0D 16 14 11 10 0E 0D
01 38 2C 24 2E 51 59 4B 30 27 39 2B 2B 24 29 69 37 25 29 82 97 A1 AB AC B2 A6 A6 A0 89 69 0F 10 1C 18 14 10 10 0F 0C 0F
03 21 2A 27 22 5C 44 31 3F 33 1F 37 24 23 36 27 24 2B 4D 50 85 90 96 86 A3 A5 99 8D 7A 4E 0E 1B 15 20 0F 0F 16 12 13 0B
01 1D 1F 2B 20 21 48 2F 40 2F 2D 2A 25 2B 2C 20 25 25 26 3E 55 5E 62 6D 6D 6E 68 5E 43 0D 10 21 18 32 1A 13 10 13 15 10
04 27 2F 2A 28 21 3B 45 2E 3A 40 33 2D 2F 1F 1E 1B 20 37 3C 3F 3C 34 30 24 17 0D 0B 0E 11 1E 23 1B 25 14 0D 10 0F 12 0F
04 22 27 37 33 1A 1B 35 4A 1D 20 2C 2F 1F 1F 3B 34 1A 2A 38 44 1E 0C 0C 06 0C 10 12 1B 21 21 34 32 20 0B 0E 10 0D 0D 0F
02 32 22 33 29 20 22 19 30 35 1D 1E 16 19 18 1C 16 18 23 39 10 13 0E 0E 1A 15 15 13 1A 18 2C 2E 19 0F 0D 10 0E 0E 14 0D
01 33 36 23 31 29 20 19 1B 1E 17 1C 1F 1F 1F 1C 31 23 1C 2F 13 11 16 10 12 16 13 19 1B 17 19 1D 13 14 10 10 12 11 12 0D
01 28 31 34 24 30 23 19 18 28 2A 1D 1F 1D 1B 1E 1B 26 31 39 16 14 13 14 13 15 1B 22 1A 1E 1B 15 13 16 0C 0D 11 0E 12 0D
00 29 20 1C 2E 25 28 28 22 1E 20 1F 1F 1D 1B 1C 29 22 43 37 17 10 15 15 12 10 14 15 1B 1E 15 1A 11 10 14 13 14 17 12 11
01 25 28 2A 23 23 29 26 1E 1D 34 38 1B 1B 22 26 18 1A 4C 33 1C 11 14 14 14 10 10 18 17 1E 29 20 1A 15 12 17 0E 14 12 12
```


01 00 05 00 03 00 02 00 00 03 01 01 01 01 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 02 00 01
03 30 3A 38 39 2D 1D 15 10 0E 0C 0A 0A 0A 09 06 08 07 06 06 05 05 07 07 04 05 04 04 06 02 01 02 02 02 02 07
01 02 02 03
03 22 1B 16 14 0A 08 0B 0A 0D 0B 0B 0C 06 07 05 05 06 06 06 03 07 04 06 05 09 05 04 05 01 04 04 02 03 03 04
02 04 03 02
00 0F 0B 04 10 07 09 07 08 09 09 08 05 08 08 05 09 03 08 05 02 08 08 06 06 04 02 05 03 02 05 05 00 02 02
0404 00 00 03
00 07 09 0E 0C 07 08 0A 0A 0B 0F 0A 0C 07 06 0B 07 0B 05 0B 08 09 07 03 08 04 04 02 00 04 02 04 00 04 03 08
00 06 09 04
00 0E 0C 09 09 08 08 07 08 09 09 0A 05 08 07 07 07 09 08 0A 08 09 06 0A 03 09 07 06 06 03 05 03 01 06 02 03
07 01 04 04
02 0C 0B 0A 05 08 09 0A 0C 0A 0A 08 0A 0A 06 08 06 06 04 06 02 06 07 04 04 04 06 09 05 05 08 06 04 05 04 06
01 0A 03 02
02 0B 14 0F 0F 0D 0A 0E 0A 0C 0C 0E 0A 0C 0B 09 0A 09 0A 0A 09 0B 0B 05 0C 0C 0A 04 07 06 03 05 07 04 05 03
02 01 06 03
02 10 12 0B 10 0A 0D 0D 0B 0D 0C 0B 0B 0C 0D 0B 0B 0A 0A 0A 0B 0C 17 15 1C 15 0D 08 09 08 05 05 05 04 02 05
04 04 00 04
01 15 0E 10 12 0C 0D 0C 0C 0A 0B 0B 09 0C 0F 09 09 0D 07 0B 08 15 60 5D 61 59 33 0D 0A 07 08 08 05 03 06 07
01 03 05 02
02 12 10 0F 0E 10 10 0B 0C 0F 0F 0E 0C 10 0D 15 10 09 12 11 12 50 68 66 89 71 5E 3F 08 09 0A 09 0A 03 03 02
05 05 04 02
01 11 12 0C 11 13 10 10 0B 10 0F 0C 11 11 13 0D 0F 0D 0D 0B 25 7A 7F 79 6D 80 6E 54 0C 0D 09 0A 06 04 02 05
00 05 04 03
01 10 0F 0D 12 0E 10 0E 0F 13 13 11 13 17 11 0F 14 11 11 14 39 84 88 7E 8C 73 7A 5C 1E 05 0A 0F 0E 0C 05 02
04 03 06 05
02 0F 15 0D 18 11 0D 11 14 10 12 12 14 19 13 17 13 16 16 20 73 68 87 89 93 8B 83 69 43 07 0A 12 0A 0B 06 06
03 04 05 03
02 13 14 14 16 11 13 13 17 12 17 17 28 1E 1A 17 19 14 12 4F 7D 74 85 91 93 8C 7F 6F 5F 0B 09 12 0D 0C 02 04
07 04 05 04
00 0F 16 0F 13 12 10 1D 12 21 15 1E 21 1F 1C 1D 2D 1A 2D 7C 7A 95 6B 30 48 62 87 71 5C 0A 08 11 0C 09 04 04
02 06 04 03
00 10 1C 10 11 1A 0D 1A 1A 25 28 33 30 26 2B 3E 29 35 6C 83 5E 7B 94 8A 5A 3D 42 76 5C 13 08 13 0F 0C 04 04
01 05 05 03
01 12 17 1A 19 18 15 20 29 20 3F 1F 37 29 39 49 24 33 8F 93 B4 AE 79 42 39 73 7D 89 46 12 06 12 12 0F 08 03
03 03 04 03
01 13 20 0F 14 26 1B 18 20 2F 3D 3E 42 3B 45 2E 48 70 96 9F 96 6B 24 0F 22 4B C3 A4 3F 4F 0C 18 16 0F 05 05
08 05 05 04

…and the output is

A harbor…

… with many dozens of boats;

… water is calm and glassy;

… vertical masts;

… mountains in background,

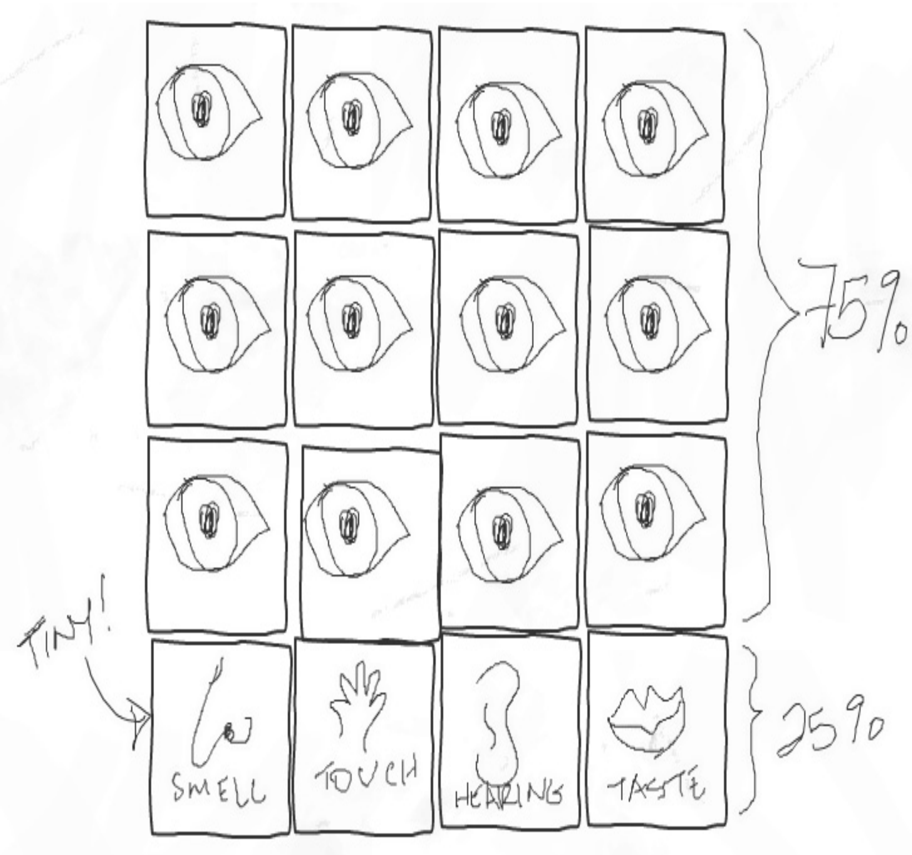
… blue sky with a touch of clouds…

Interpret images.

Image → Symbols, Semantics, Meanings,..

Why Vision ?

- Vision is the single most important sensing and perception for animals, for human beings, and for robots.
    - About 70% of daily information is captured by our eyes.
    - On the brain, two third (2/3) of the cortical areas are devoted to vision related processing.
- An image is worth of thousands of words.
- Camera is cheap, passive and energy-efficient, but very powerful.



Concept: Sunni Brown via Dan Roam

A picture is worth a 1,000 words

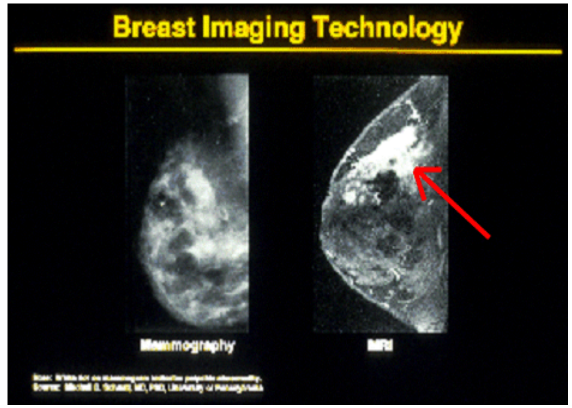
A picture is worth 100million dollars

100 million $

"Dora Maar au Chat"
Pablo Picasso, 1941

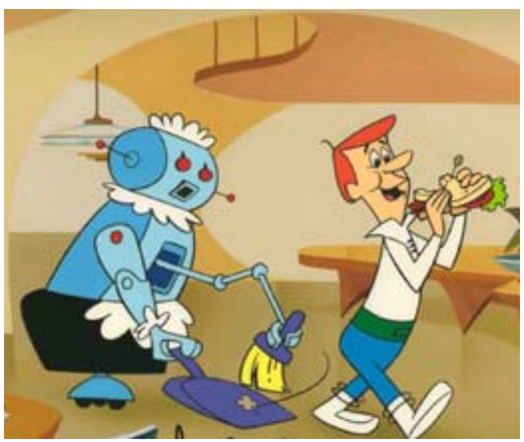
Why computer vision matters?


Safety

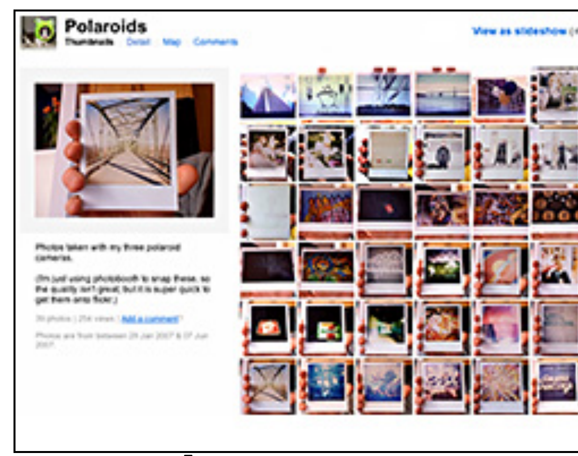


Health


Security

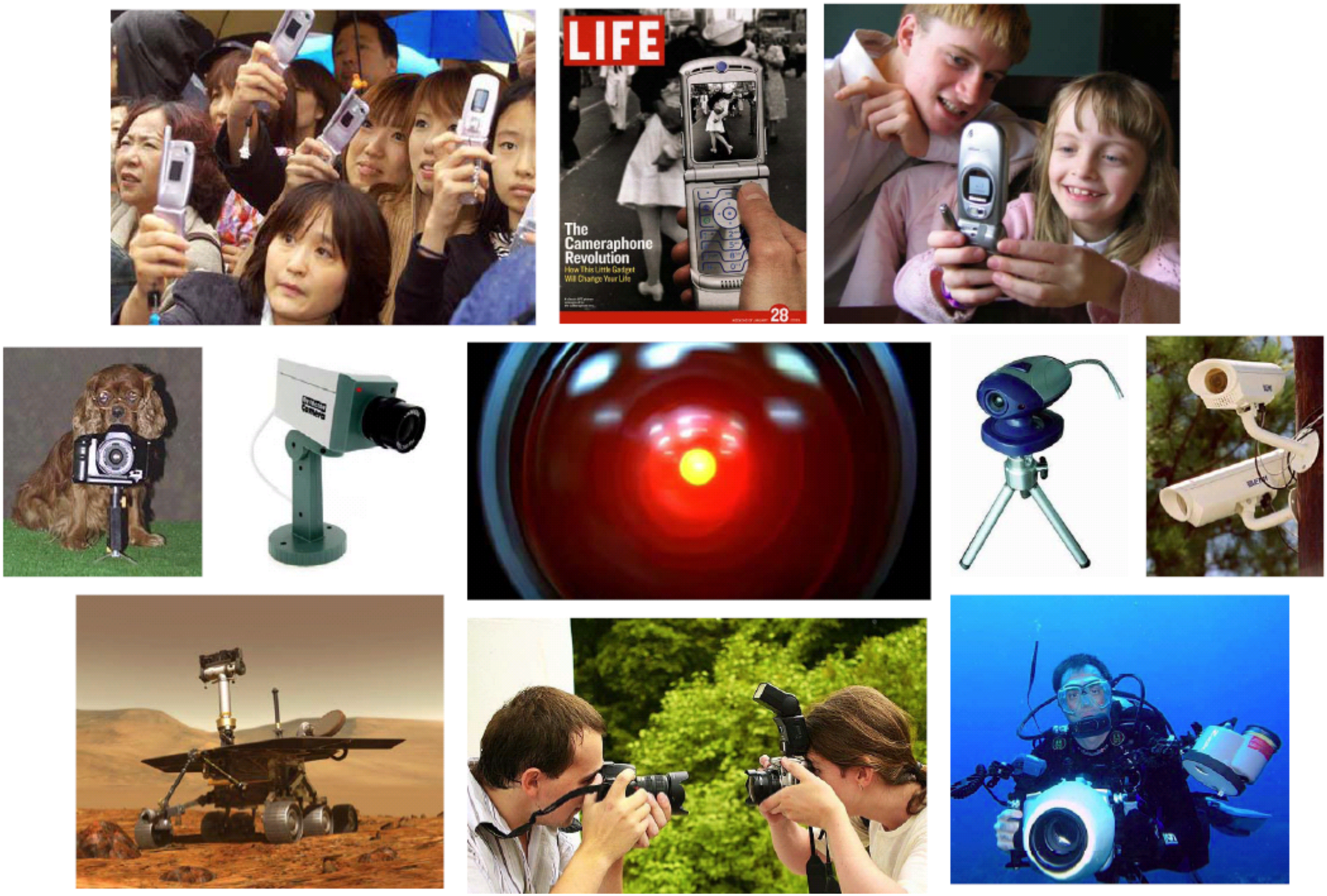

Comfort


Entertainment

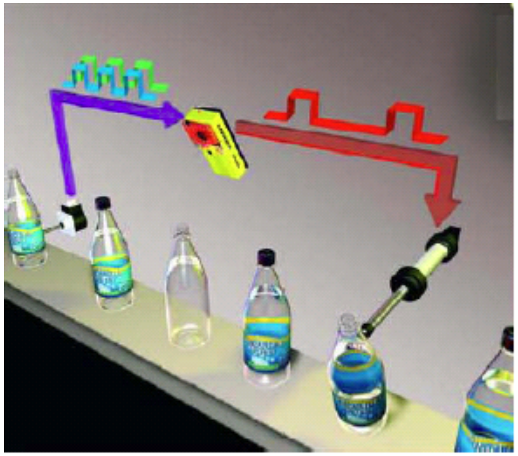


Access

Cameras are everywhere!

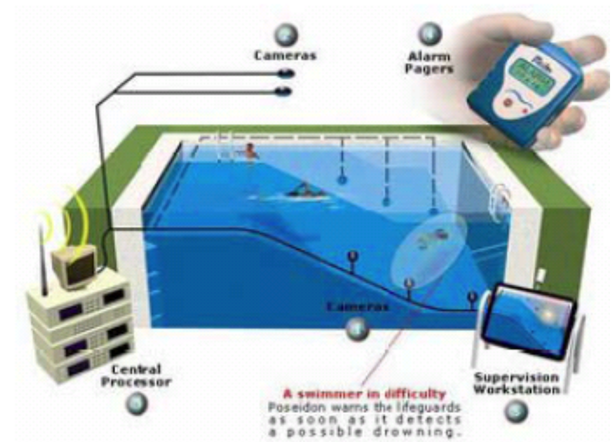


Sample Applications



Factory inspection

Reading license plates, checks, ZIP codes



Monitoring for safety (Poseidon)

Surveillance

Autonomous driving, robot navigation

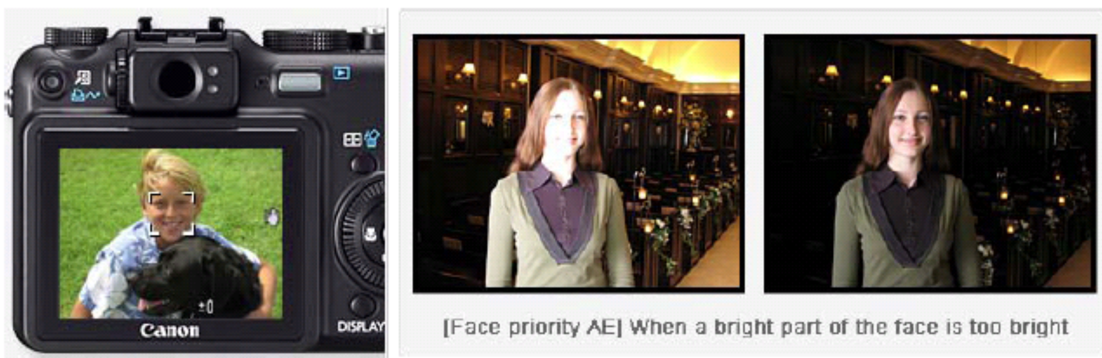
Driver assistance (collision warning, lane departure warning, rear object detection)

Sample Applications

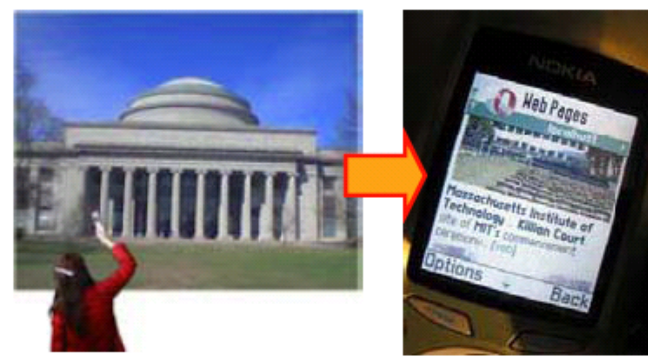
Assistive technologies

Entertainment
(Sony EyeToy)

Movie special effects



Digital cameras (face detection for setting focus, exposure)

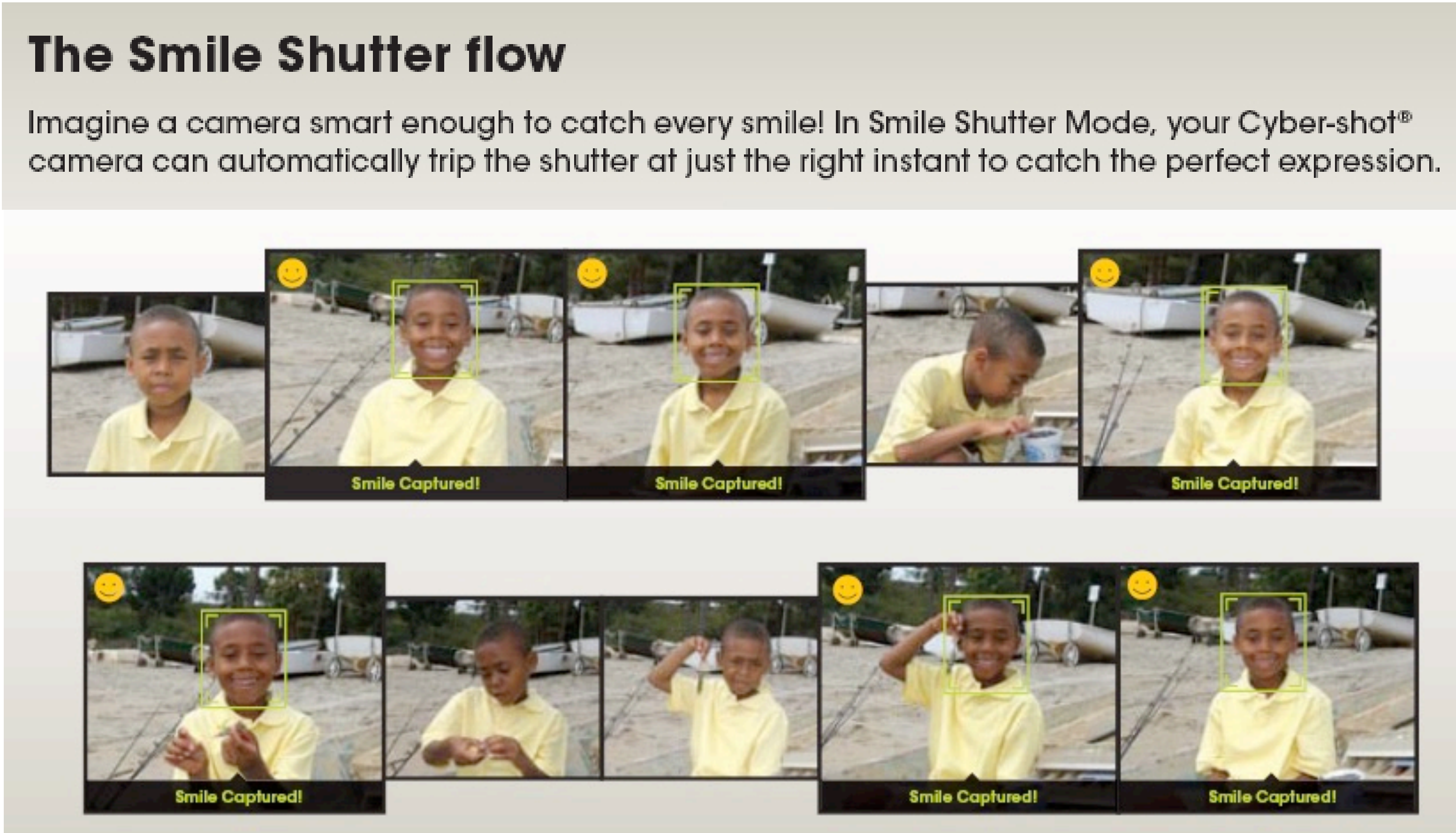


Visual search
http://www.kooaba.com/

Face Detection in Cameras

[Face priority AE] When a bright part of the face is too bright

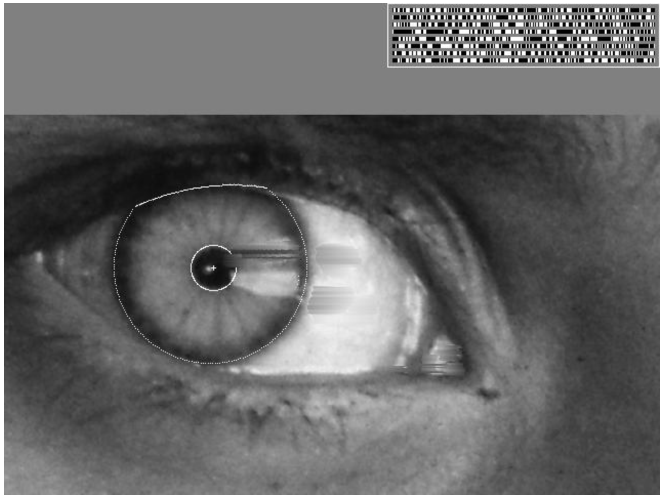
Smiling face detection?

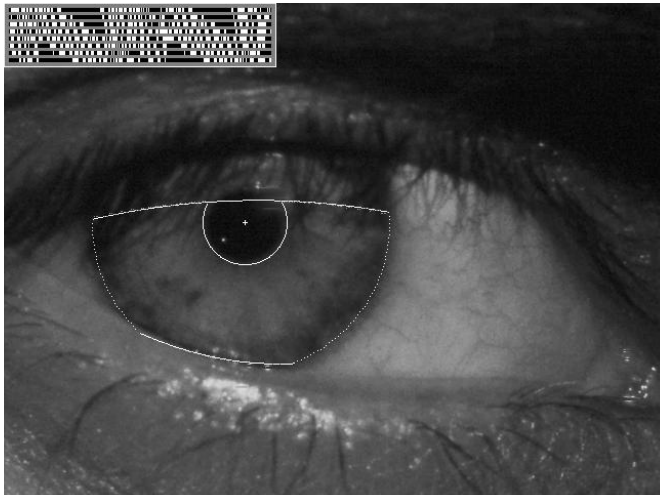


[Sony Cyber-shot® T70 Digital Still Camera](#)

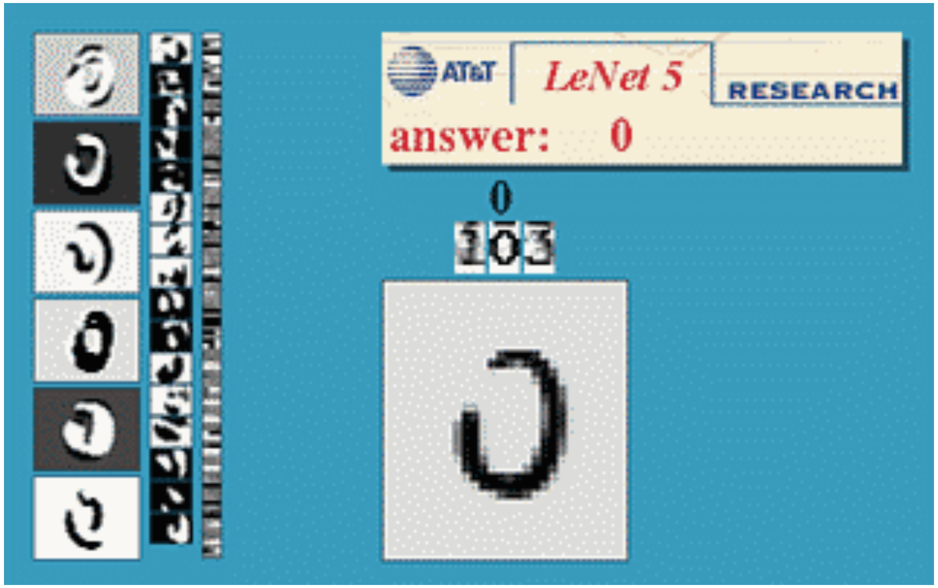
Vision-based biometrics

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





Handwritten Digit Recognition



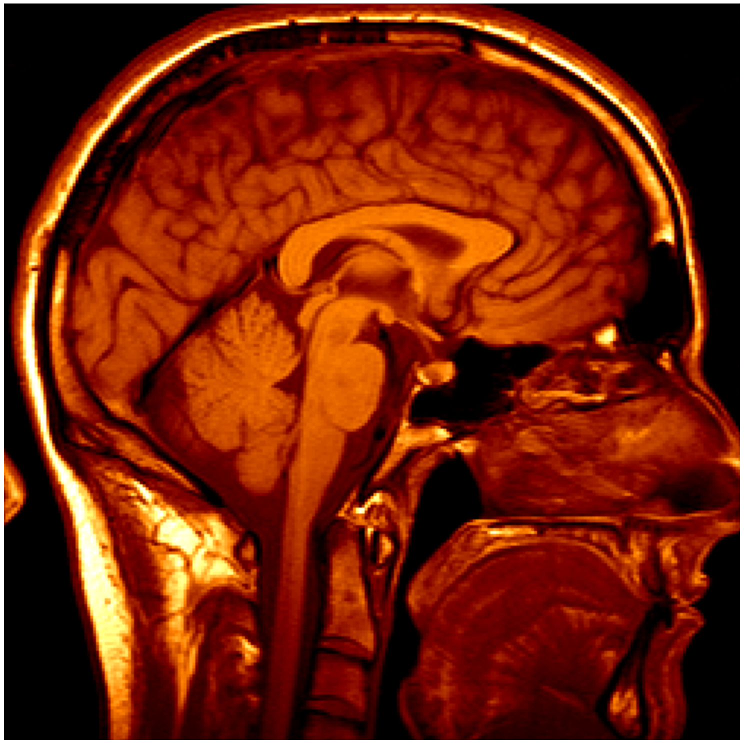
Digit recognition, AT&T labs
Prof. Yann LeCun (NYU)

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

1/3 of all checks written in US use this system

Source: S. Seitz

Medical Imaging



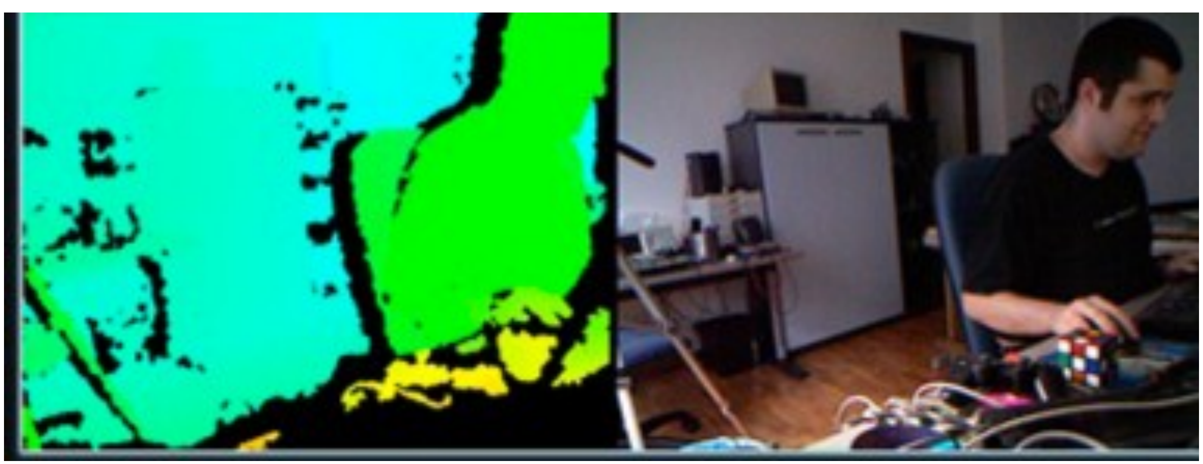
3D imaging
MRI, CT

Image guided surgery
[Grimson et al., MIT]

Vision-based human computer interface

Microsoft Kinect

KINECT
for XBOX 360.



Vision for robotics, space exploration

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

Panorama stitching

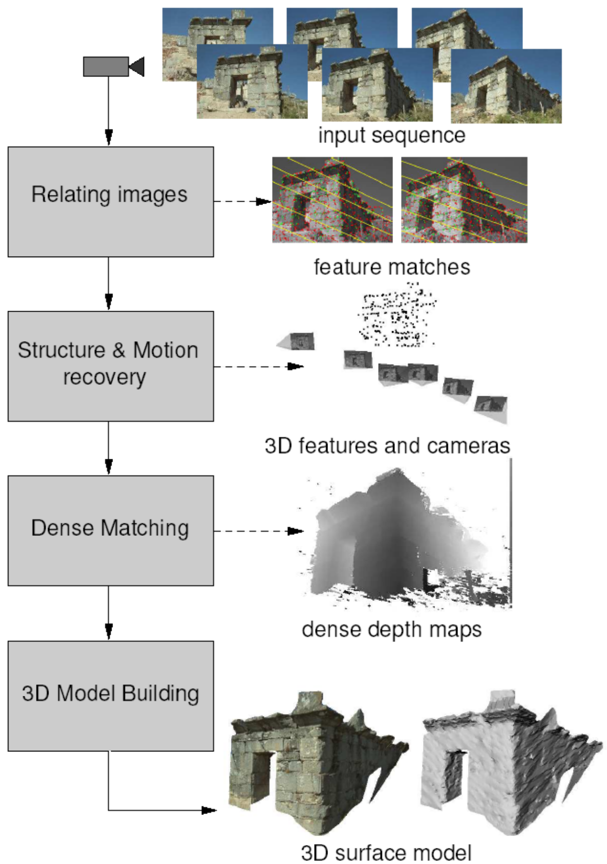
3D terrain modeling

Obstacle detection, position tracking

For more, read "Computer Vision on Mars" by Matthies et al.

Source: S. Seitz

3D Reconstruction

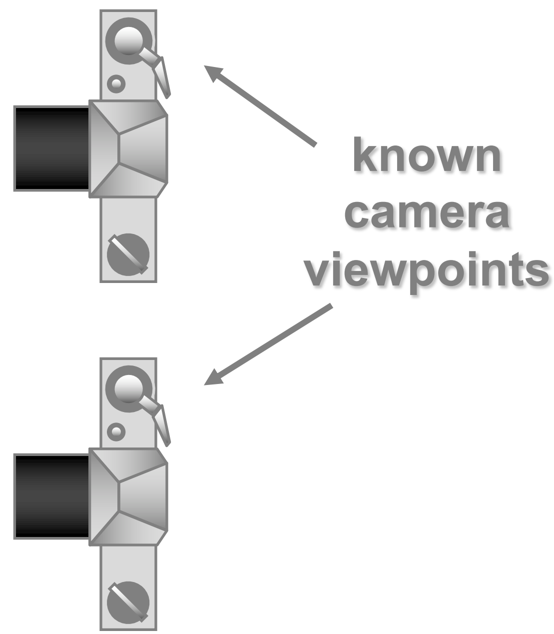


Pollefeys et al.

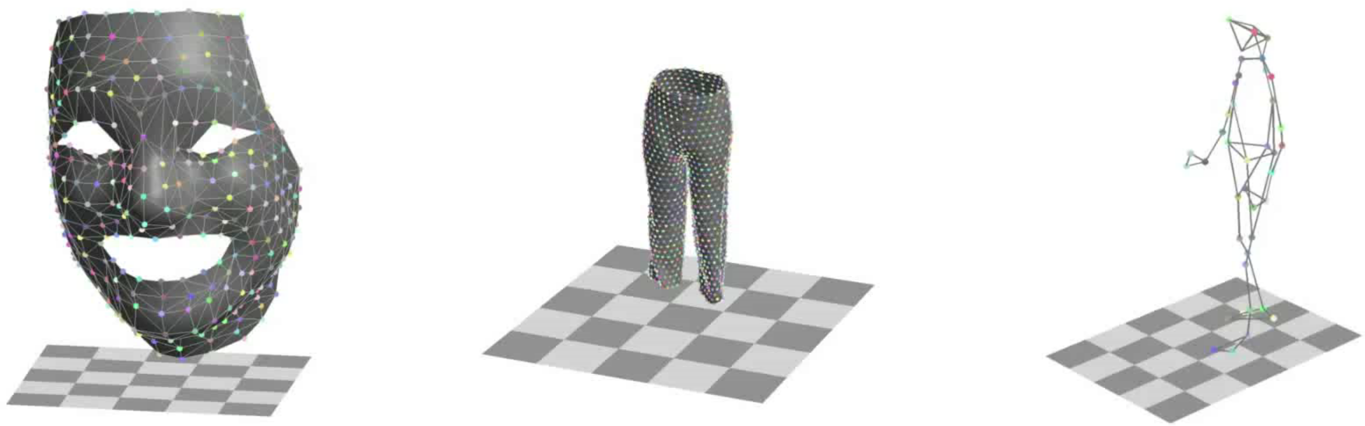
Goesele et al.

Structure from motion

Given two or more images of the same scene, compute the 3D structure



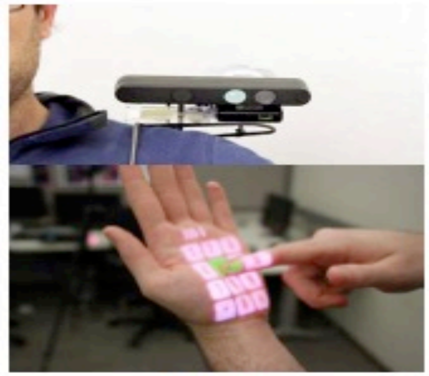
Non-rigid Structure form motion



Non-rigid applications : Applications are everywhere

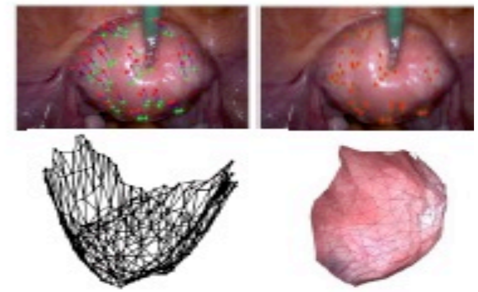
Motion capture/animation

Andy Serkins - Rise of the Planet of the Apes



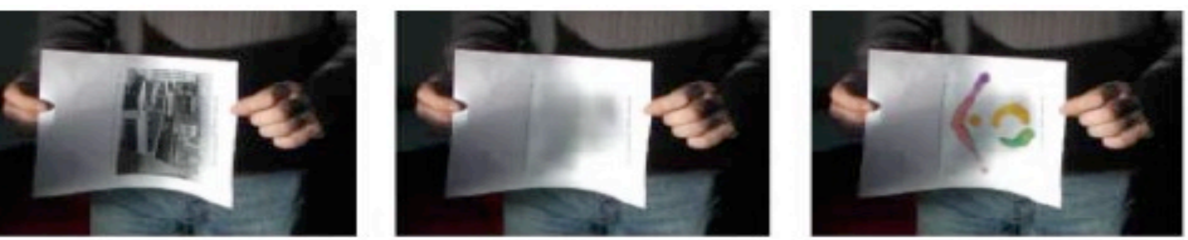
HCI

OmniTouch – C. Harrison et al. 2011



Medical Imaging

A. Malti et al., MIUA 2011

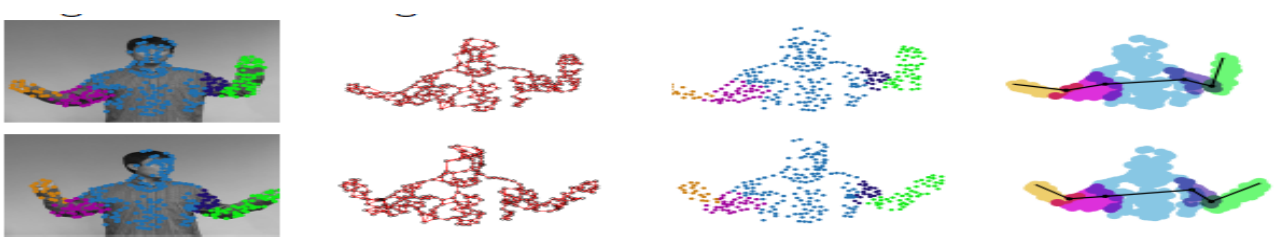


Augmented reality

Pilet et al. 2008

Cloth animation

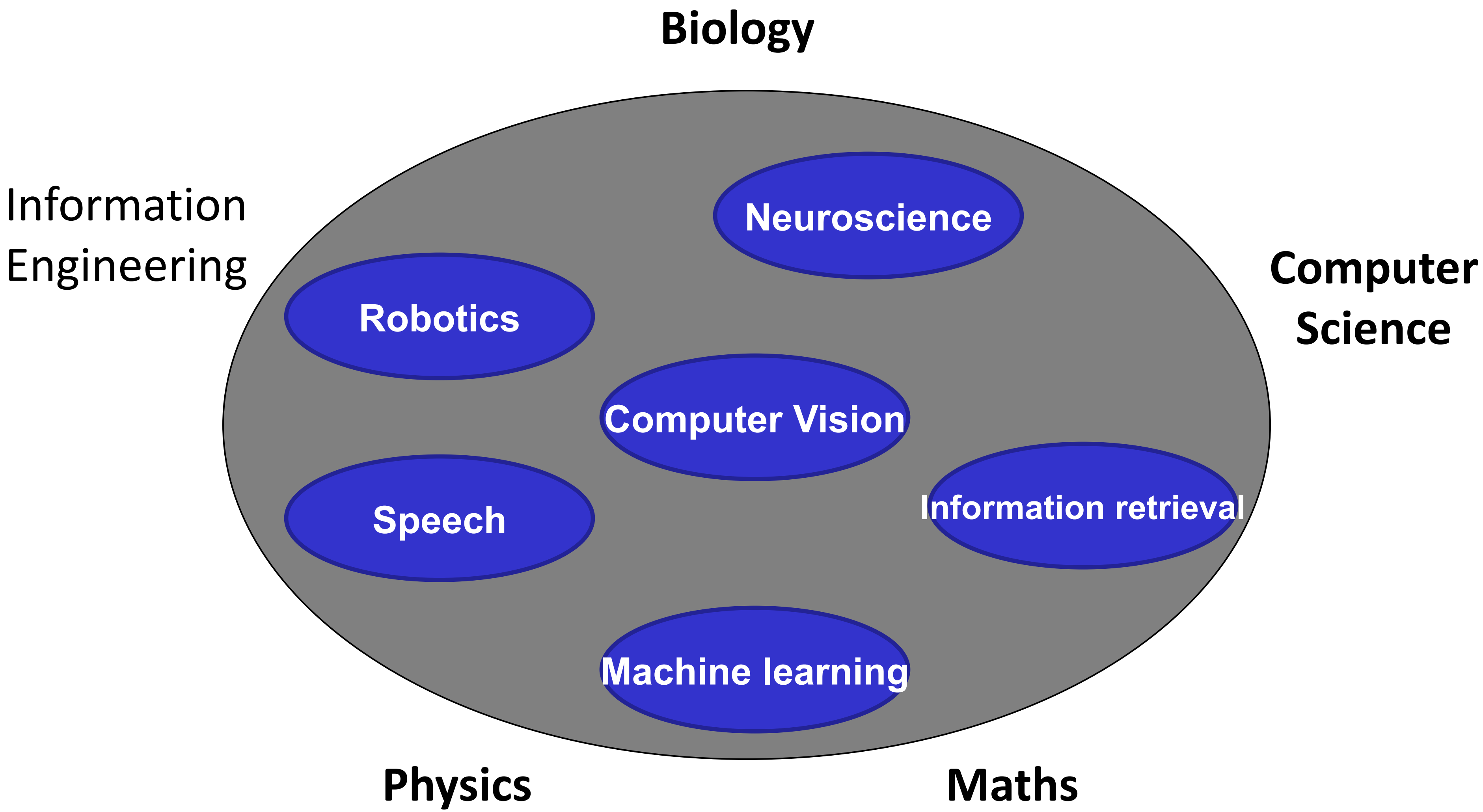
R. White et al., SIGGRAPH 2007



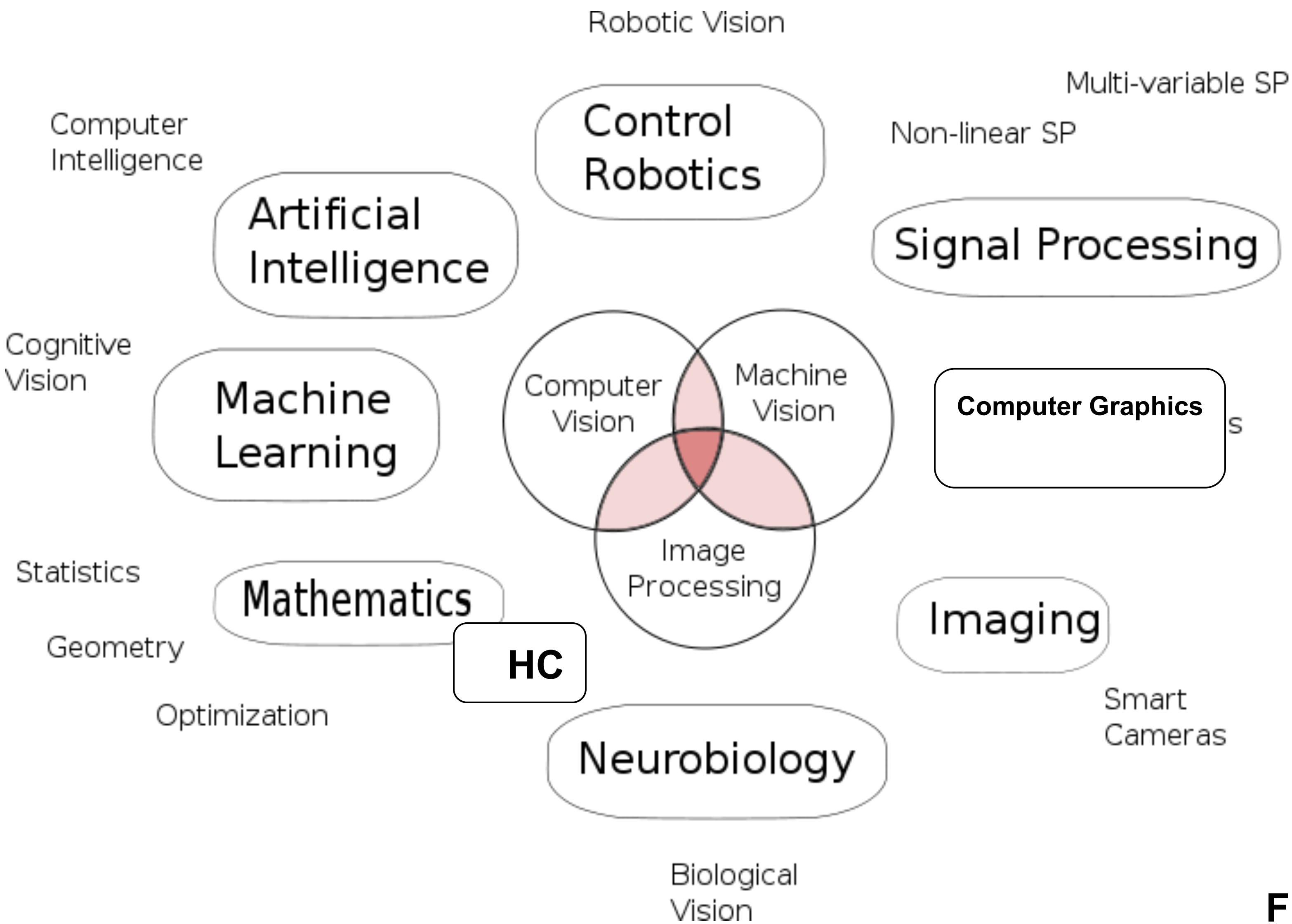
This slide is taken from Agapito et al's Tutorial at

Non-rigid application: movie special effects

Computer vision - multidisciplinary research field



See also Wikipedia: computer vision



From wiki

Administrative Information ENGN6528

Disclaimer

- Many of the slides used here are adapted from online resources (including many open lecture materials) without proper acknowledgement.

- They are used here for the sole purpose of classroom teaching. All credit and all their copy-rights belong to the original authors.

- You should *not copy it, redistribute it, put it online, or use it for any other purposes than for this course of ENGN4528/6528.*

General course information

- An introductory-level (but research-oriented) course for computer vision.
- ***Postgraduate students (Master/PhD).***
- 12 weeks of 2~3-hour lectures, 3 computer lab assignments (Matlab/python-based), 1 mid-term exam, and 1 final-exam.

Assessment Components

	Lectures	Tutorials (TBA: on demand)	Computer Lab x3	One quiz	Final-exam
6528			15%x2+20% = 50%	20%	30%

Computer Lab group sign in

- You can do it on Wattle from Week-2.
- Each group is capped at 25 students at most.
- Lab sessions
- You need to use your own laptop with Matlab/Python and image processing toolbox.

Tutors

Ryan Pike

Yujiao Shi

Wei Mao

Jiayu Yang

Lecture Attendance

- You are strongly encouraged to attend as many lectures (live session) as you can, as you will benefit significantly from participating lectures, for this open and research oriented course.

Lectures (12 weeks x (2~3) hours)

Tentative Plan, Subject to changes

1. Introduction & Image processing basics
2. Spatial transformation.
3. Image filtering.

4. Feature extraction; Image segmentation
5. PCA and eigen-Face,
6. 3D vision: Multi-view geometry-I

7. 3D Vision: Multi-view geometry-II
8. Stereo and optical flow, Shape from X and other 3D vision.
9. High level visual recognition

10. Deep Learning-1
11. Deep Learning-2
12. Review

Work-load and study-plan

- 2 Hour Lectures per week
- 1-Hour Tutorial [not every week]
- 3 hour Computer Lab time [2-hour in-class +1-hour after-hours]

Classroom rule

- **Late arrival:** please make all effort to come to the lecture theatre in time. If you are late, please entre the room as quietly as you can. Otherwise you will distract your classmates.
- **Late arrival via zoom:** Please mute yourself.
- **No small talks:**
- Asking class-related questions are encouraged, however,
- "small talks" (any private conversation with your neighbours) are not appropriate during lecture time, as this may affect other students' concentration, and may upset the lecturers.
- Use your courtesy.
- **No mobile phone, no earphone.** This is a basic manner for any one during any public meeting session.
- **Academic integrity !!!**

Tutorials (TBA)

- On-demand, not fixed. (typically will use the hour of Wed. lecture time 4:00PM--5:00 PM).

- Some possible topics for tutorial (TBA):
  - Matlab prime, Python, Explain the assignments,

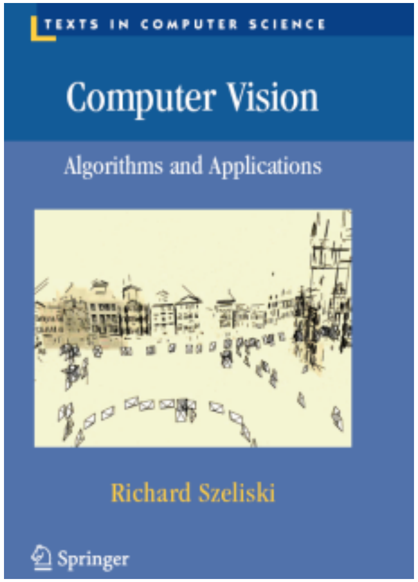
- You are strongly encouraged to contact tutors or me, for any questions/inquiry or seeking feedbacks.

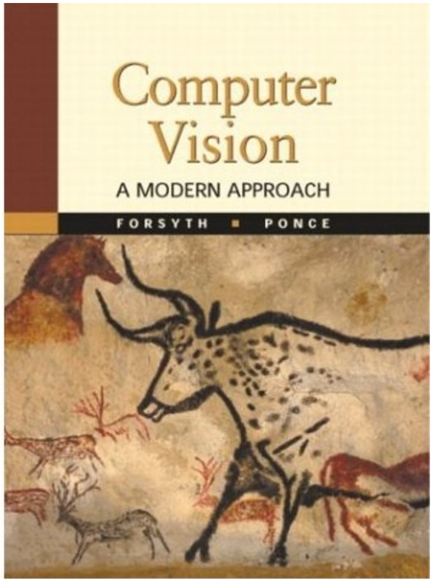
Recommended textbooks

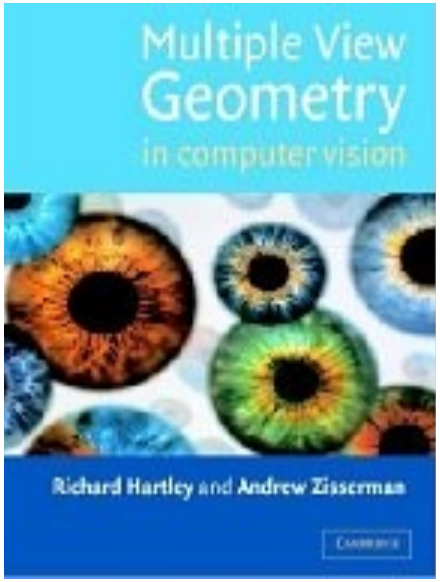
1. Computer Vision: Algorithms and Applications by Rick Szeliski 2021 (2010) (online, or can be purchased in hardcopy)

2. Computer Vision: A Modern Approach, Forsyth and Ponce 2002

3. Multiple View Geometry in Computer Vision 2rd Edition, by Hartley and Zisserman 2004.







Final Exam (30%)

- Basic concepts, basic knowledge points.
- Basic problem-solving-skills, calculations.
- Design problems: algorithm/application system design.
- We will provide sample questions before the mid-term and final exam.

Subjects of Computer Vision

-A quick preview of this semester's lectures

Three levels of vision processing

- **Low-level vision**:

– image processing, denoise,
filtering, image restoration,
low level feature extraction.

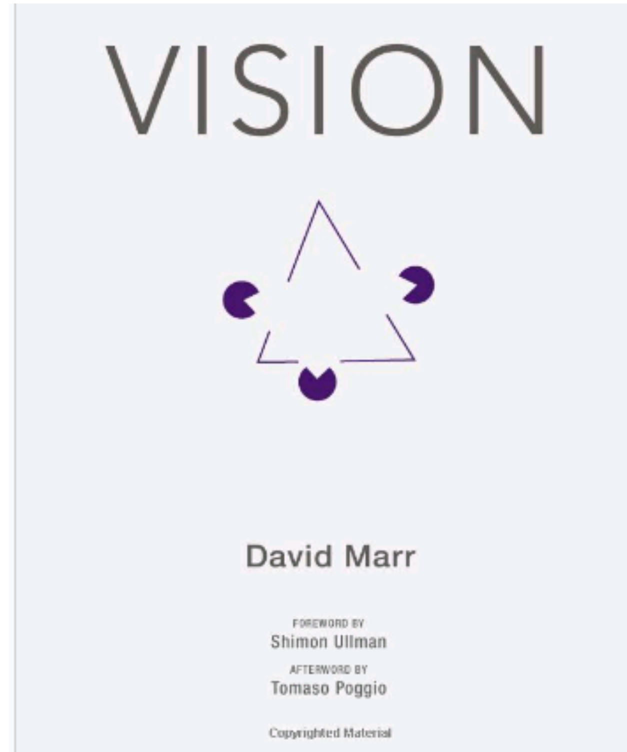
- **Mid-level vision**:

– image analysis, image segmentation,
contour extraction, perceptual organization

- Inferring 3D geometry

- **High level vision**:

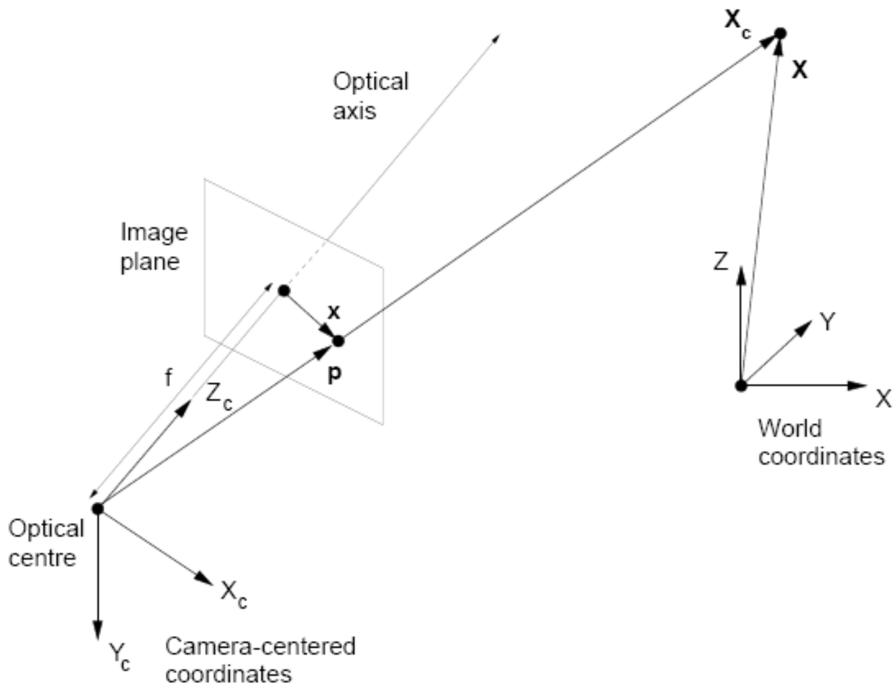
– visual detection, recognition, understanding,
semantic labelling, activity, event detection.



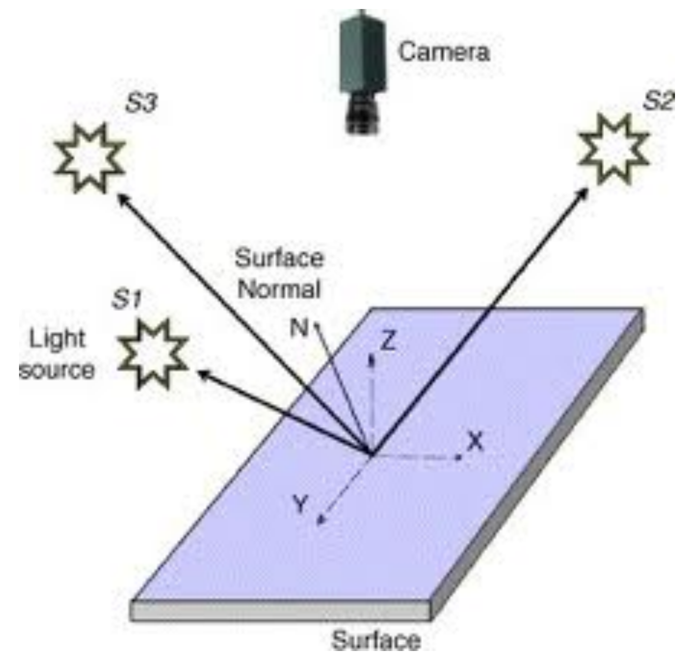
Four Modules

Low level Vision	Mid-Level Vision
Multi-View Geometry	High Level Vision

Image Formation



Geometry



Photometry (Radiometry)

Image Representation

Binary image,

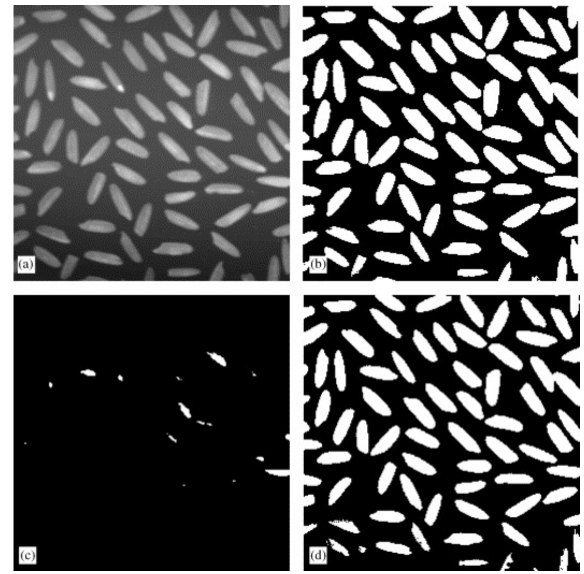
Grey scale,

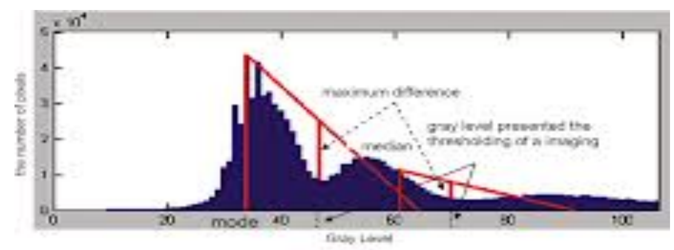
Colour space, …

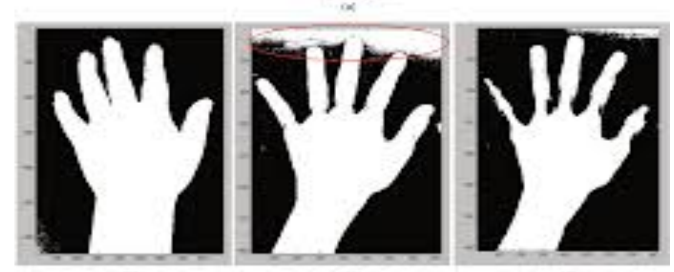
Image processing (deblur, denoise)

Input blurred image Deblurring result Magnified views

Binary image analysis







Feature extraction

Edges, Corners

Local regions

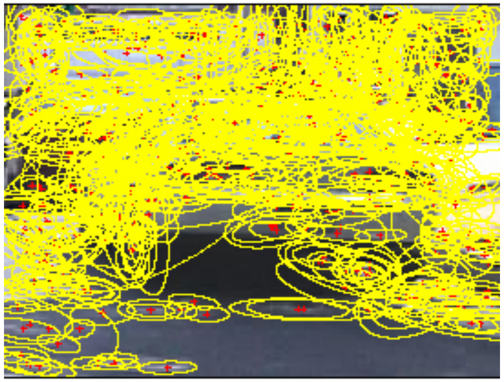


Image segmentation

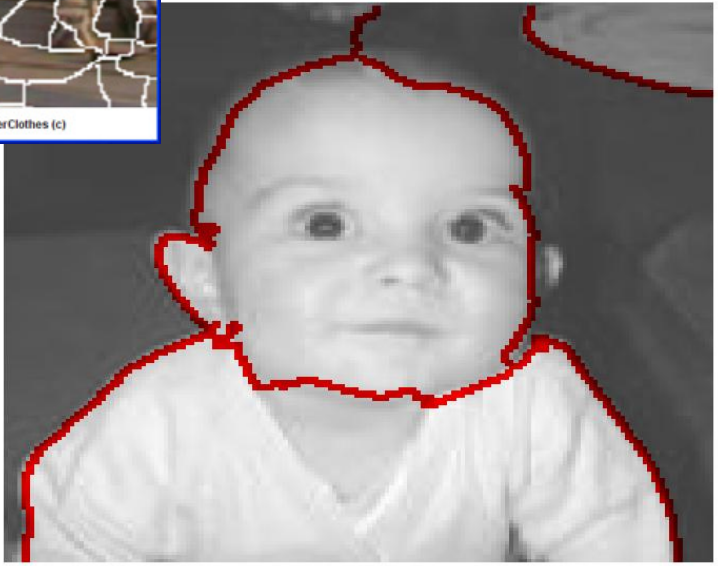
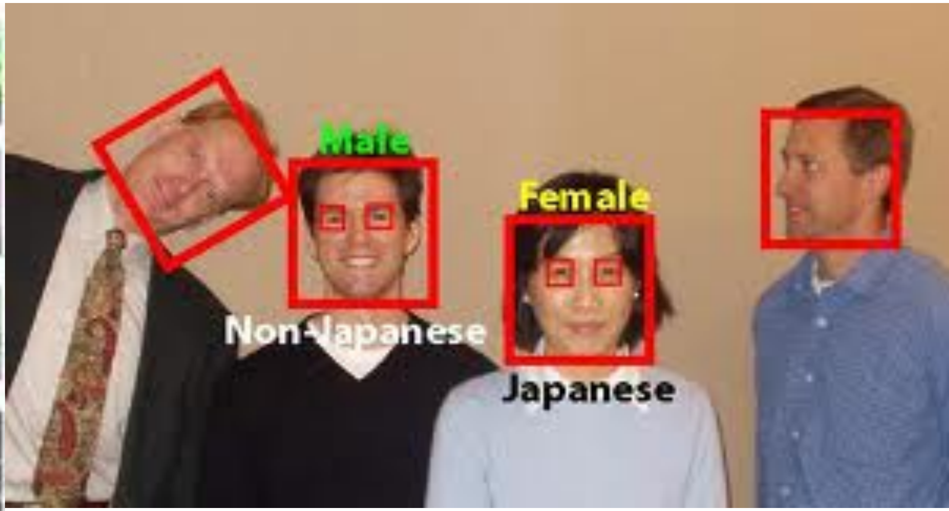


Image Segmentation

Image

Segmentation

Face detection and Face recognition



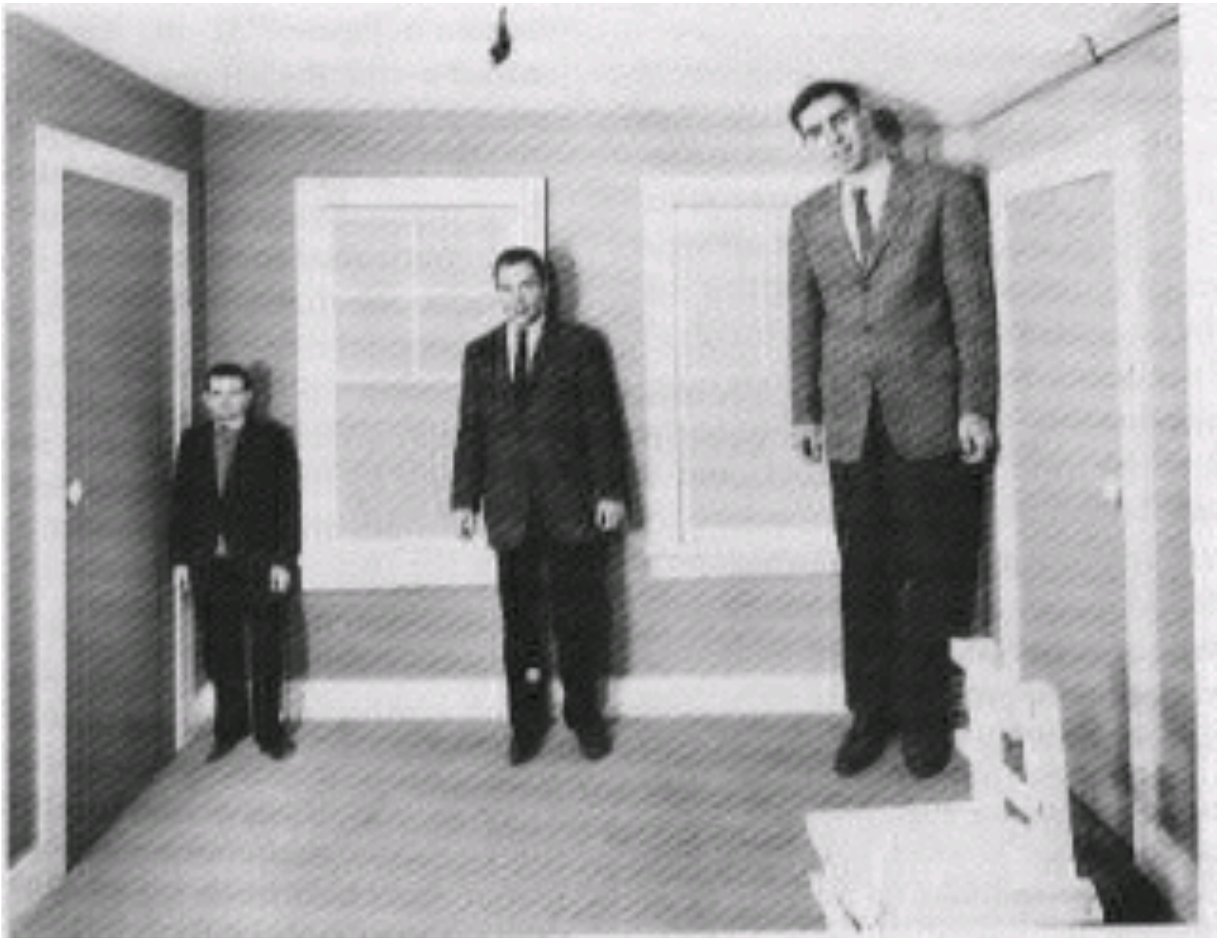
Object Detection: localize the street-lights

Object category recognition

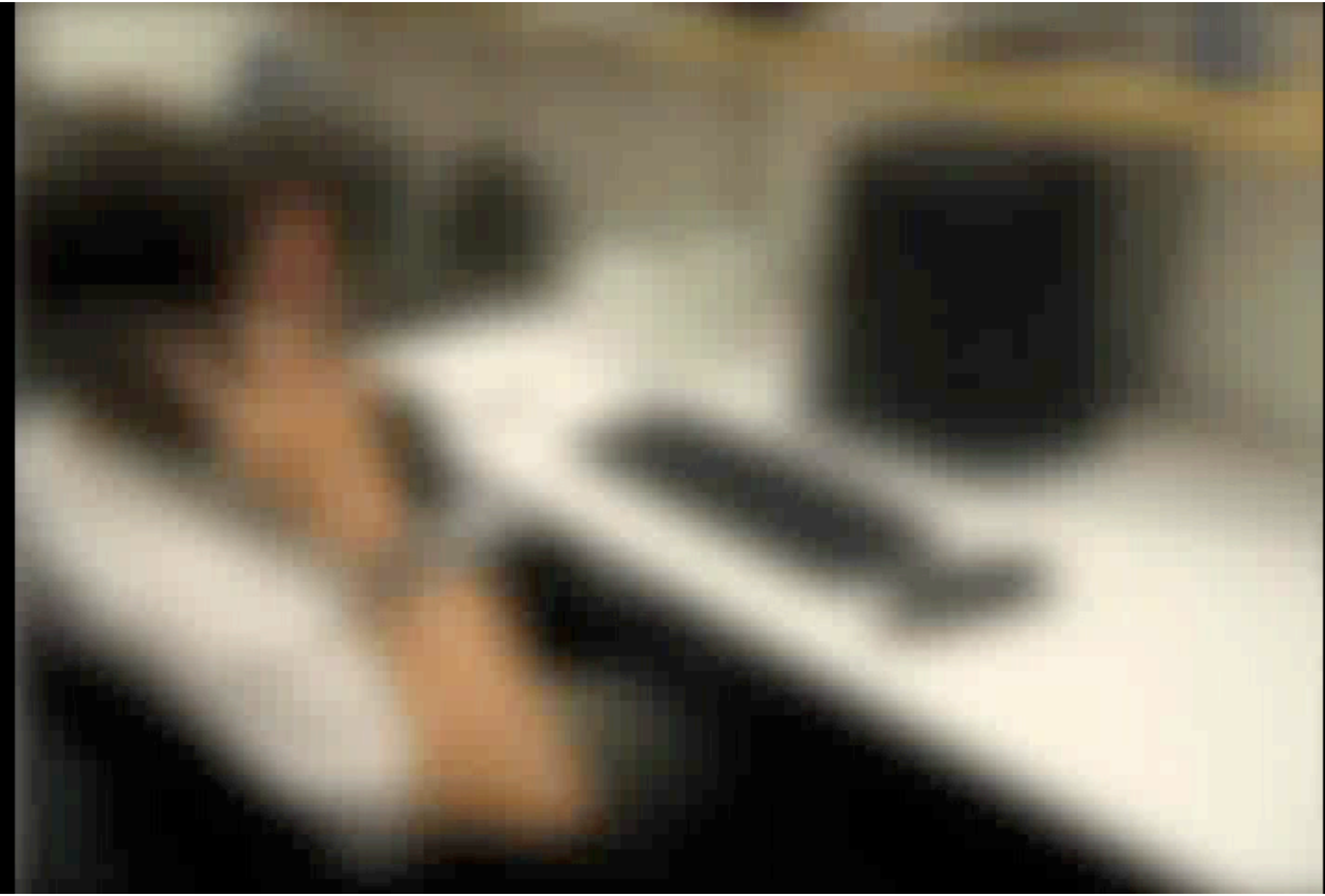
Human motion detection

Vision is not an easy task; even humans sometimes may be confused

Prior assumption sometimes can be wrong or misleading



Ames room



Credit: Antonio T. (MIT)

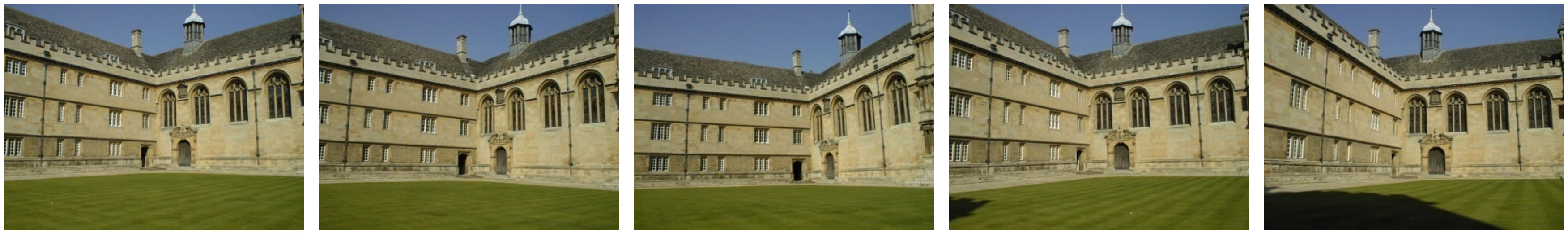
How to develop a vision-based chair recognizer ?

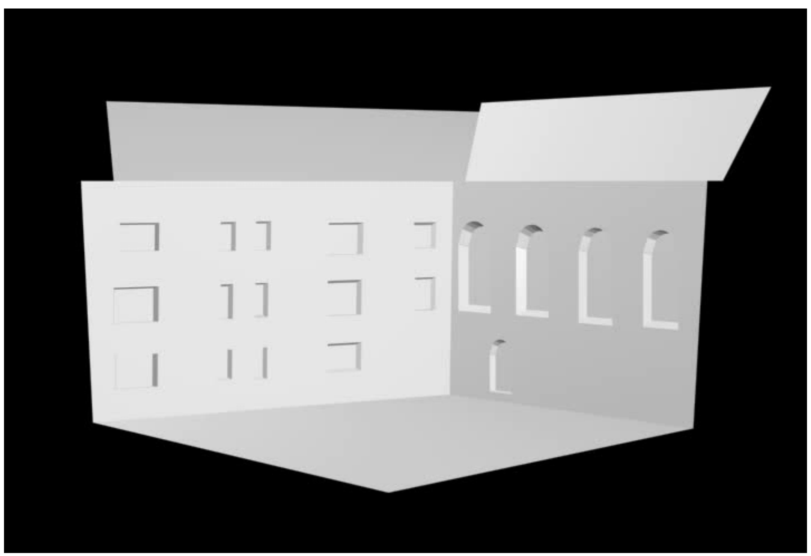
Design a chair recognizer ?

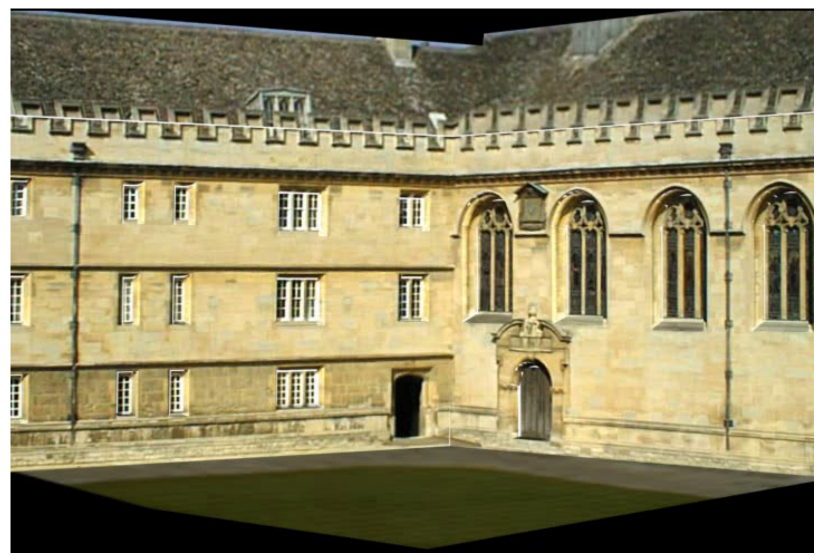





Multiview Geometry

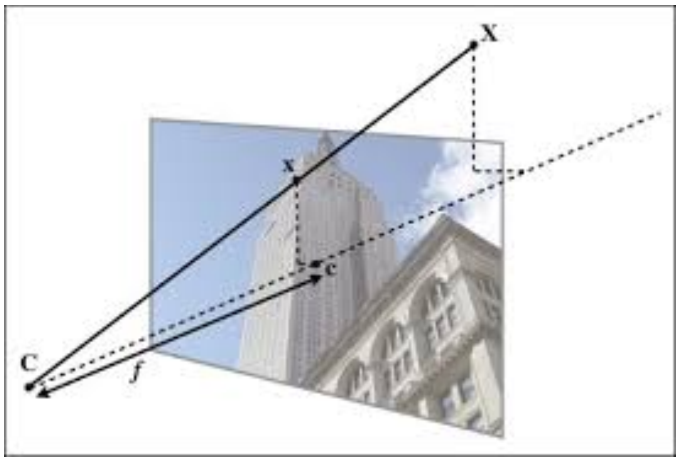


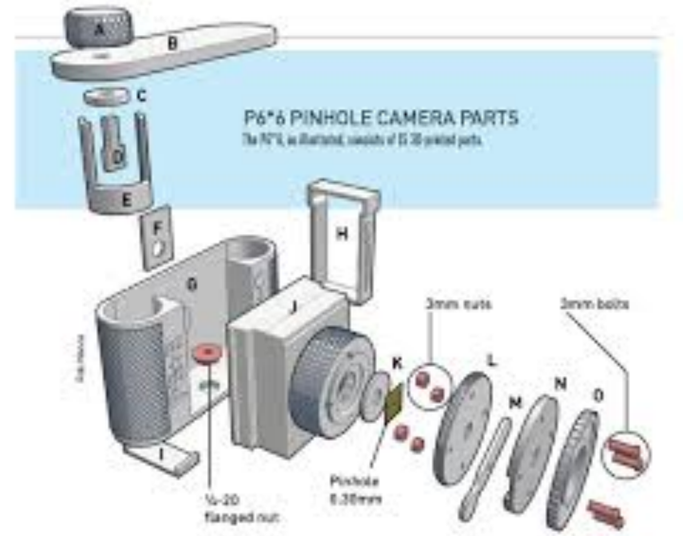


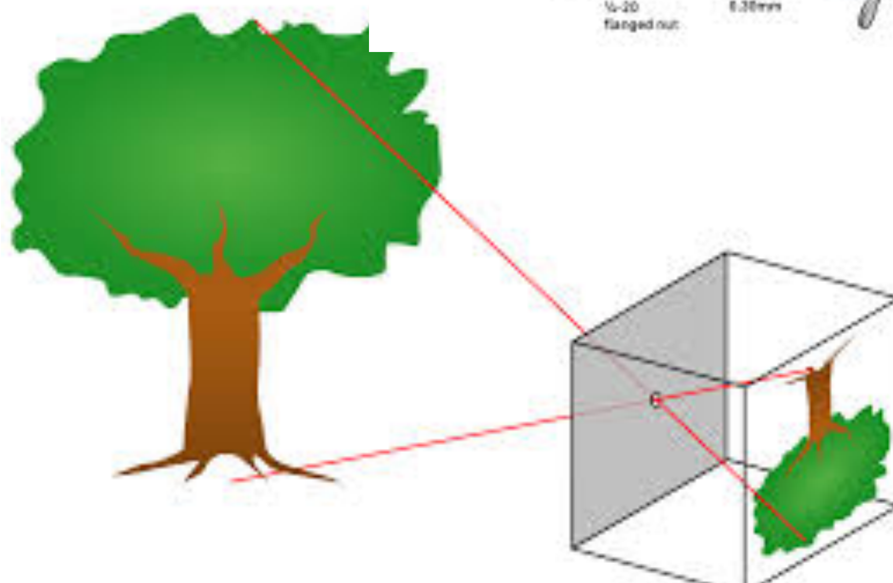


[Fitzgibbon et. al]
[Zisserman et. al.]

Camera model and camera calibration







3D Reconstruction: Structure from Motion

Objective: given a set of images …

- aim to compute where the camera is for each image and the 3D scene structure:

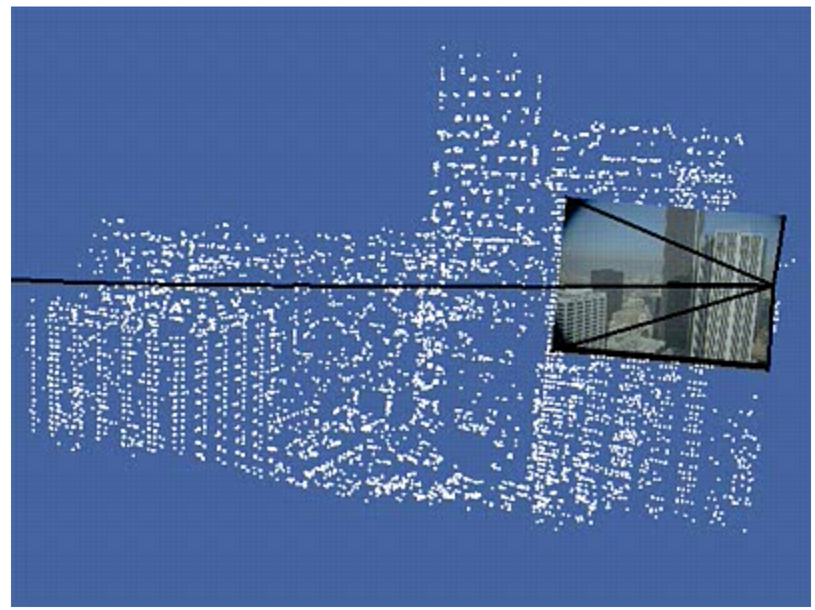
  - Uncalibrated cameras

  - Automatic estimation from images (no manual clicking)

Example

Image sequence

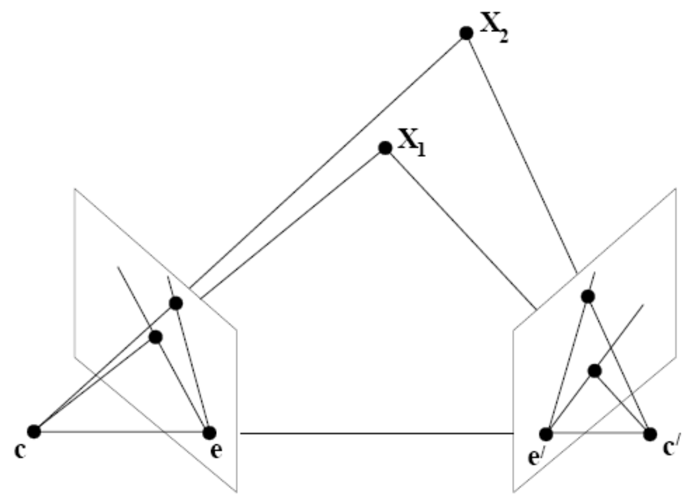
Camera path and points



[Fitzgibbon et. al]
[et. al. Zisserman]

Stereo Vision

- By having two cameras, we can triangulate features in the left and right images to obtain depth.
- Need to match features between the two images:
  - Correspondence Problem



Application: Augmented reality

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

Readings

Download and read Ric’s book, Chapter One.

Computer Lab group sign-up will open from Week-2.

Any Question ?

Tomorrow (one hour only)

Image Formation

Image Representation