01 Introduction
CS-477 Computer Vision
Enrollment Code: 698125704

Dr. Mohsin Kamal
Associate Professor
dr.mohsinkamal@seecs.edu.pk

School of Electrical Engineering and Computer Science (SEECS)
National University of Sciences and Technology (NUST), Pakistan

"One picture is worth more than ten thousand words."
Arthur Brisbane (1911)

Figure : TESLA's self driving car.

Prerequisites

This course is appropriate as a first course for UG/PG students with a EE/CE/CS background, which should have prepared them with these essential prerequisites:

- Good working knowledge of MATLAB or OpenCV with Python or C/C++ (or willingness and time to learn quickly), data structures
- Linear algebra, calculus and probability

While a background in image processing would certainly be beneficial.

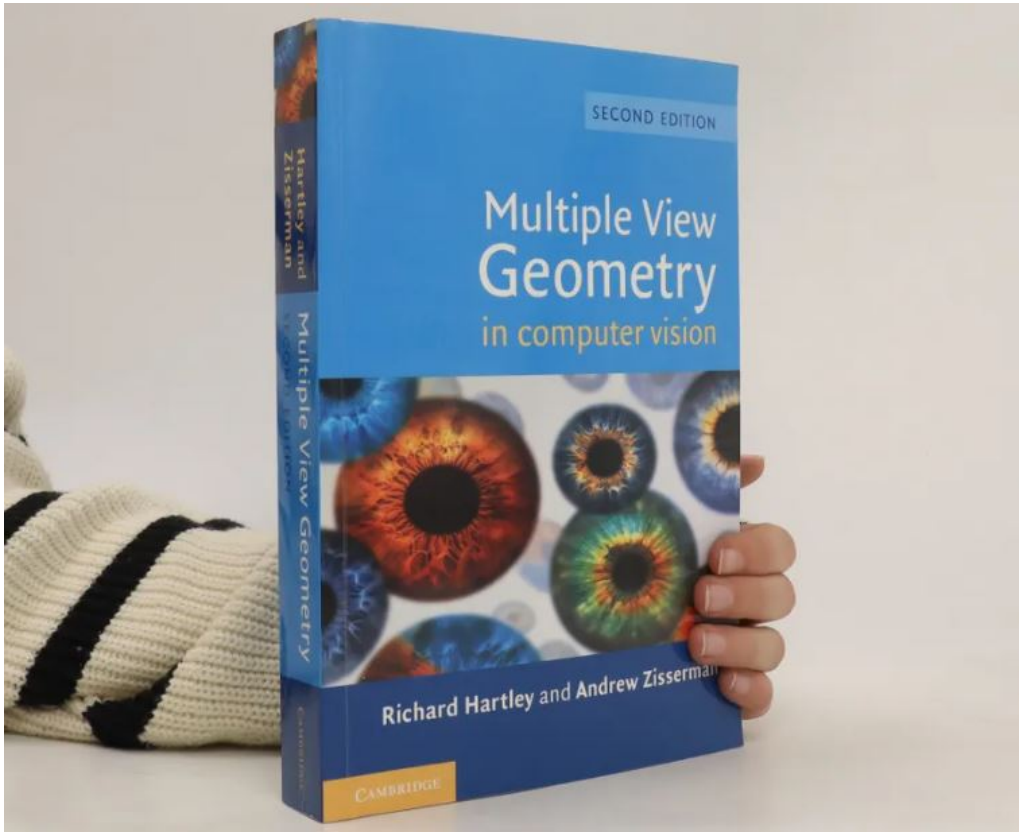
Textbook

Computer Vision: Algorithms and Applications

© 2010 Richard Szeliski, Microsoft Research

http://szeliski.org/Book/

Reference book



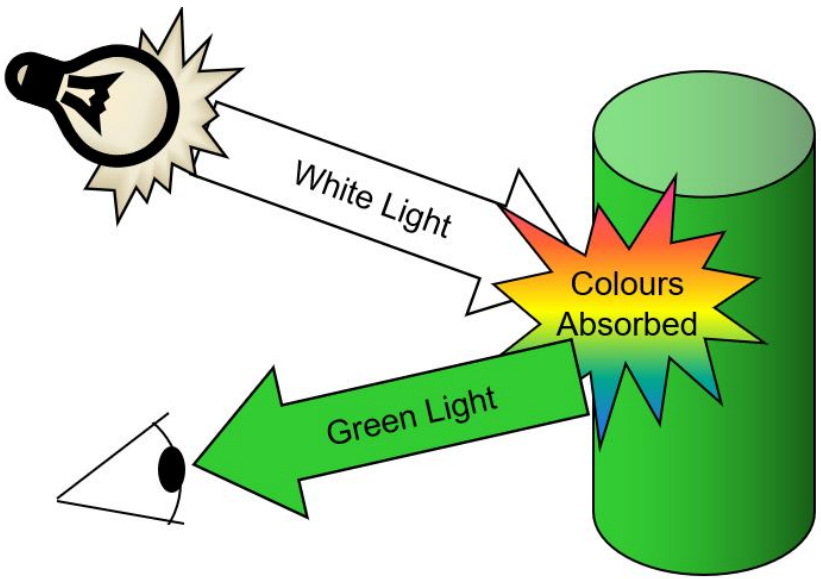
Multiple View Geometry in Computer Vision by Richard Hartley and Andrew Zisserman (latest edition)

Grading policy

Quiz	10% of the theory part
Assignments	10% of the theory part
Lab work	25% of the course
Mid Term Exam	30% of the theory part%
Final Exam	50% of theory part

Image Formation: Reflected Light

- The colors that we perceive are determined by the nature of the light reflected from an object.
- For example, if white light is shone onto a green object most wavelengths are absorbed, while green light is reflected from the object

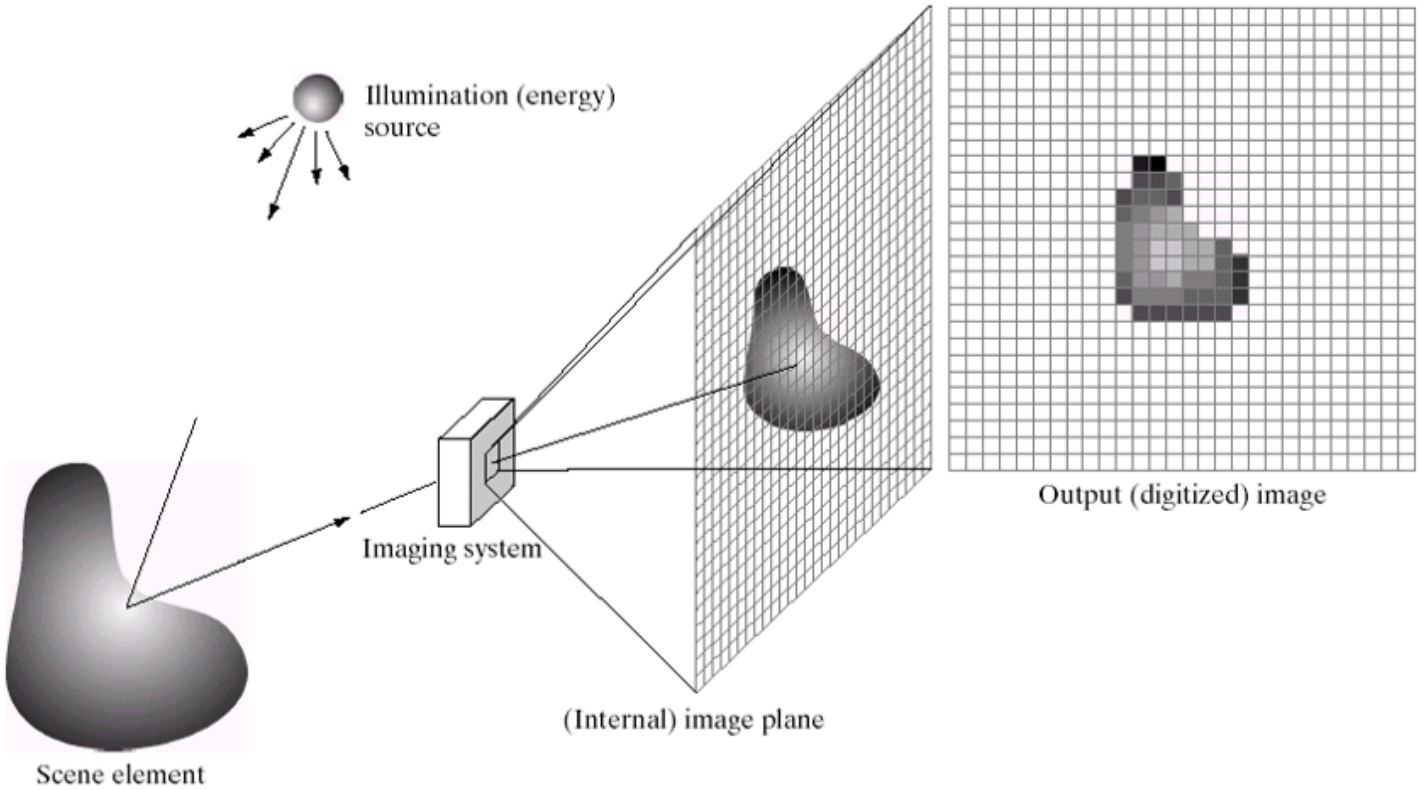


What is a digital image?

- An image is a 2D projection of the 3D scene.
- Mathematically, an image is a 2D function $f(x, y)$, where x and y are spatial (plane) coordinates, and $f(x, y)$ is the amplitude of the function at (x, y) called intensity or gray level.
- A digital image is a $f(x, y)$, where x, y and $f(x, y)$ are all finite and discrete quantities. An image point at (x, y) is called a picture element, image element, pel or pixel.

What is a digital image?

A digital image is a representation of a two-dimensional image as a finite set of digital values, called picture elements or pixels.

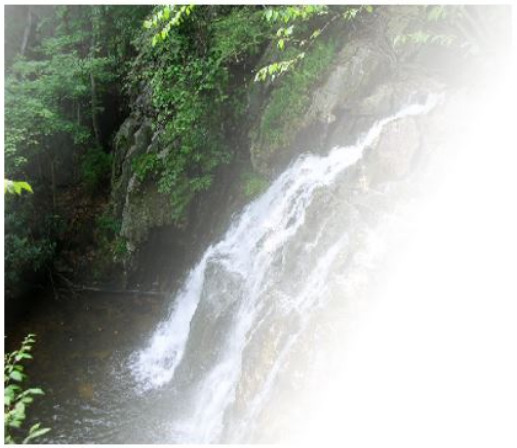


What is a digital image?

Common image formats include:

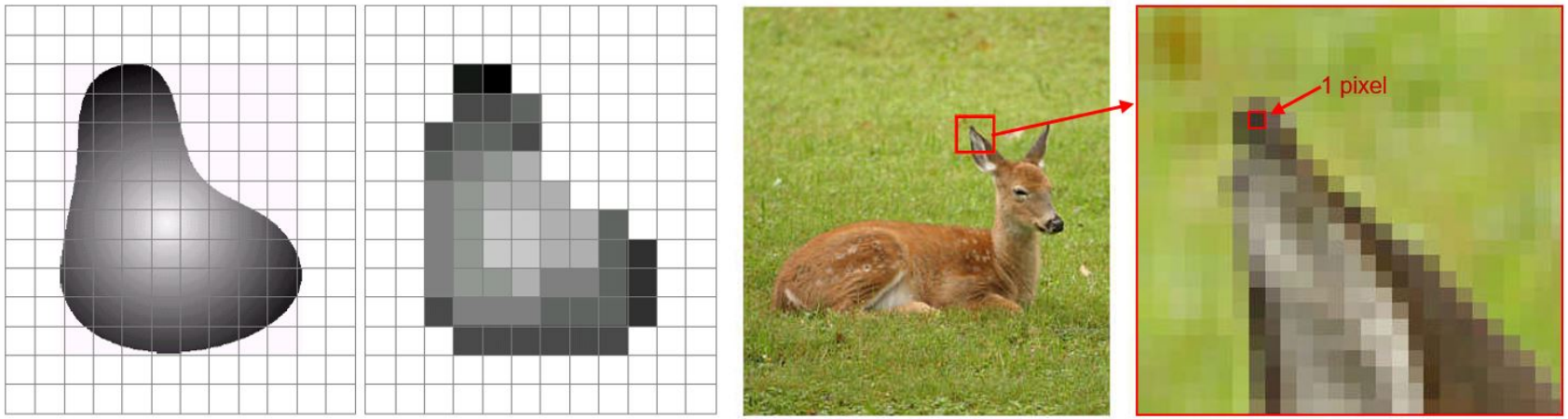
- 1 sample per point (B&W or Grayscale)
- 3 samples per point (Red, Green, and Blue)
- 4 samples per point (Red, Green, Blue, and "Alpha", a.k.a. Opacity or Transparency)





What is a digital image?

- Pixel values typically represent gray levels, colors, heights, opacities etc.
- **Remember** digitization implies that a digital image is an *approximation* of a real scene



What is Image Processing?

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it.

What is Digital Image Processing?

- Digital image processing is the transformation of a digital image by means of a digital computer.
- Image processing starts with an image and (usually) produces a modified version of that image - e.g., enhancement, restoration, compression, i.e., the output is also an image.
- Machine vision / computer vision aims to understand the content of images at a higher level so as to take decisions based on visual input.

Image Acquisition

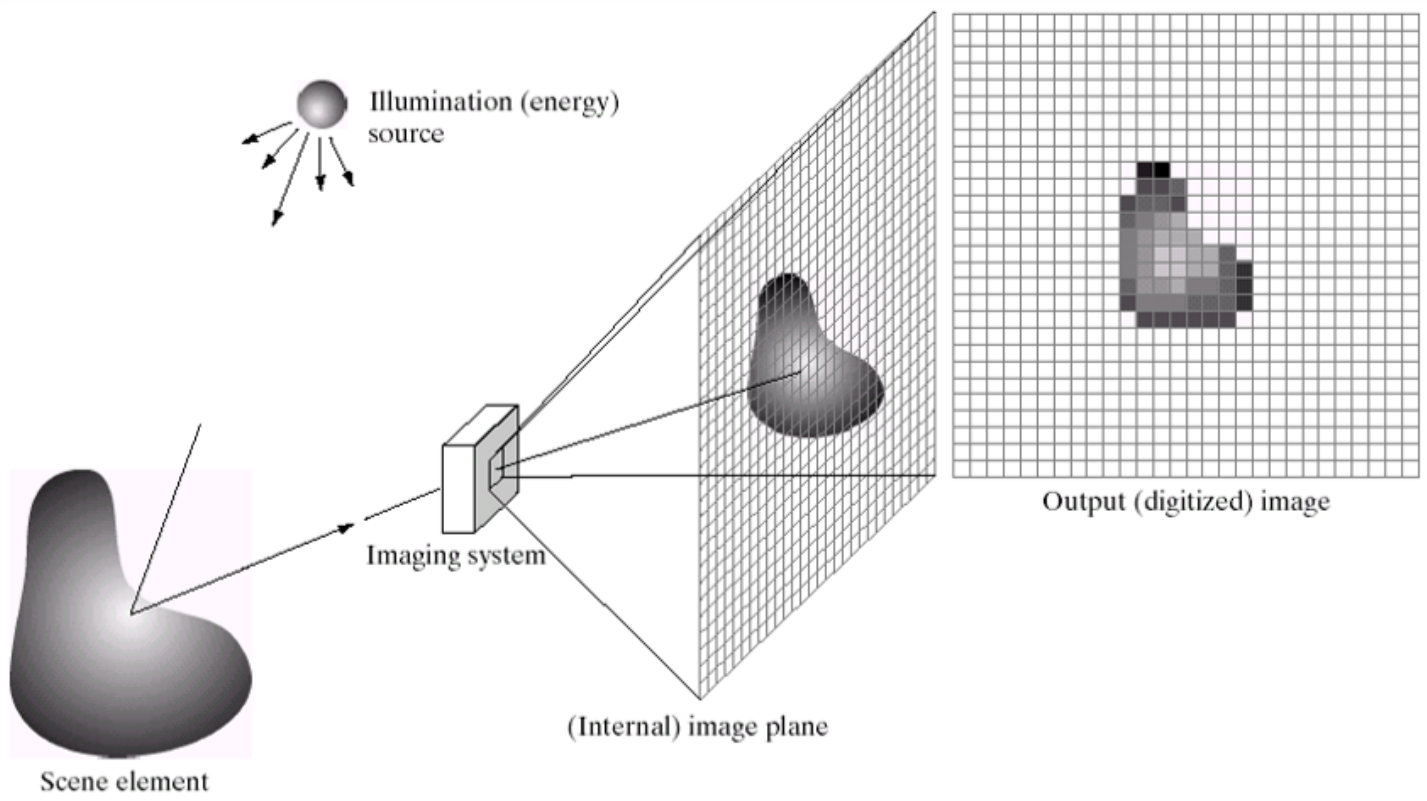


Figure : An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

Sampling and Quantization

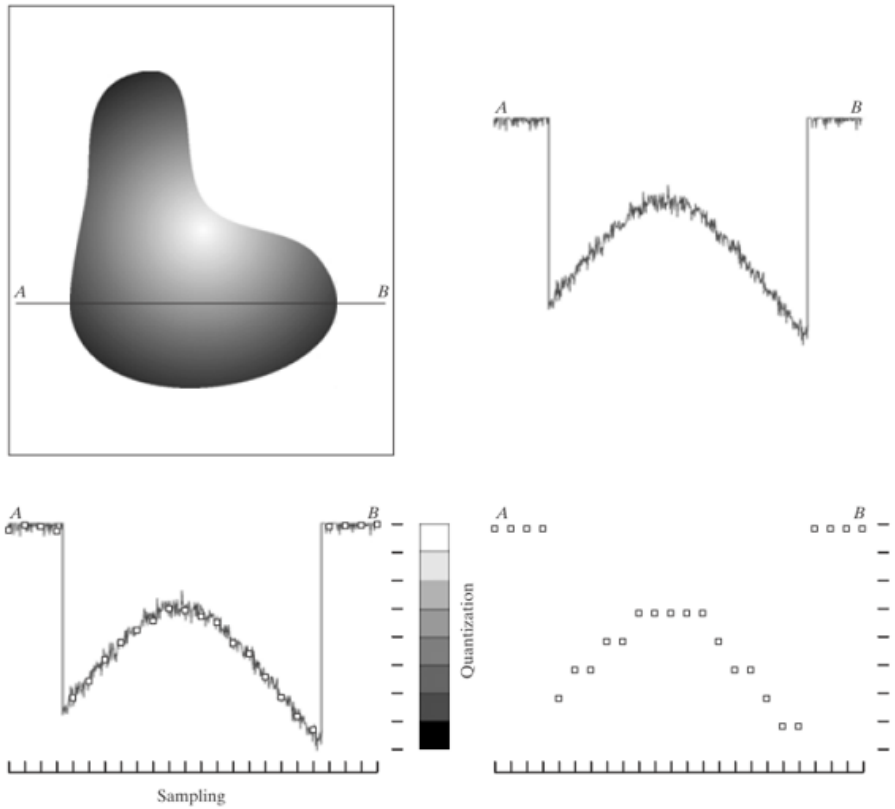
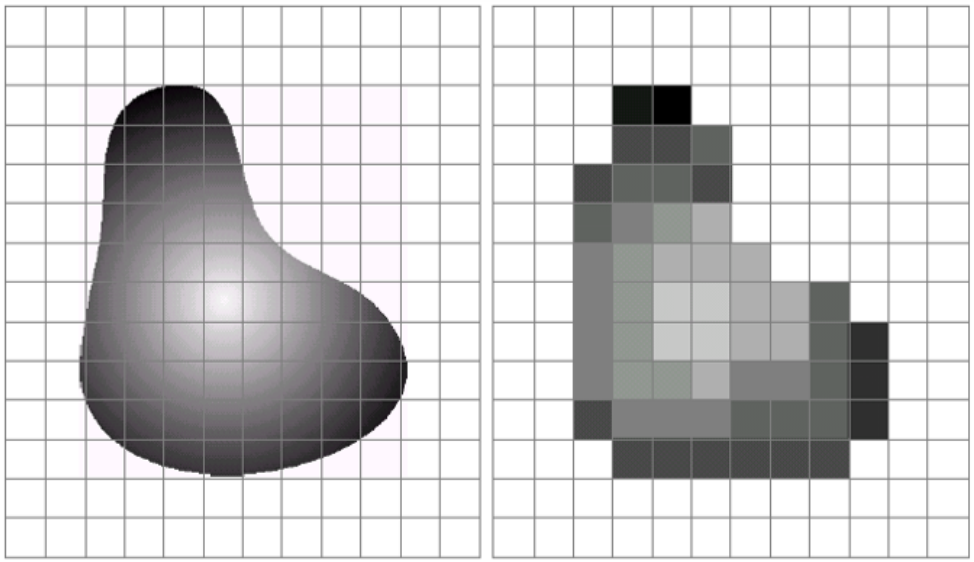


Figure : Generation of a digital image. (a) Continuous image. (b) A scan line from A to B in the continuous image, used to illustrate the concepts of sampling and quantization. (c) Sampling and quantization. (d) Digital scan line.

Result of Sampling and Quantization



A digital image is an approximation of a real world scene.

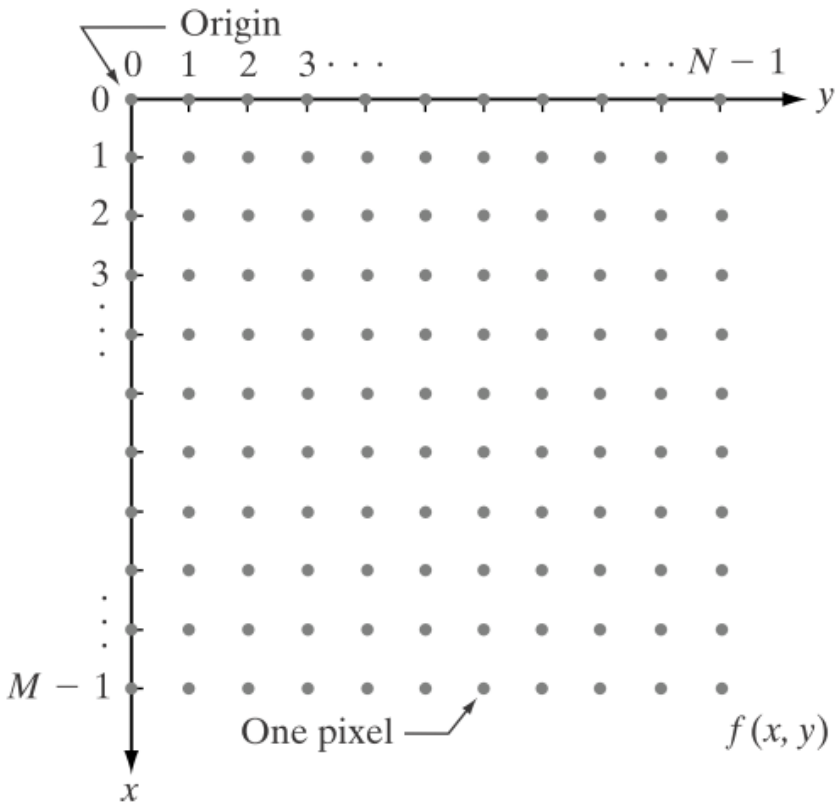
Digital Image Representation

- Number of bits required to store an image are:

$$b = M \times N \times k$$

- Image having 2^k intensity levels
  - $k =$ bit depth (bits per pixel) of an image.
  - For a grey scale image, there are 256 intensity levels (0-255), hence, $k = 8$.
  - 0 represents black and 255 the white.

Coordinate Convention



Standard convention in programming is to place the image origin at top, left corner

From Image Processing to Computer Vision

Low, mid and high-level processes

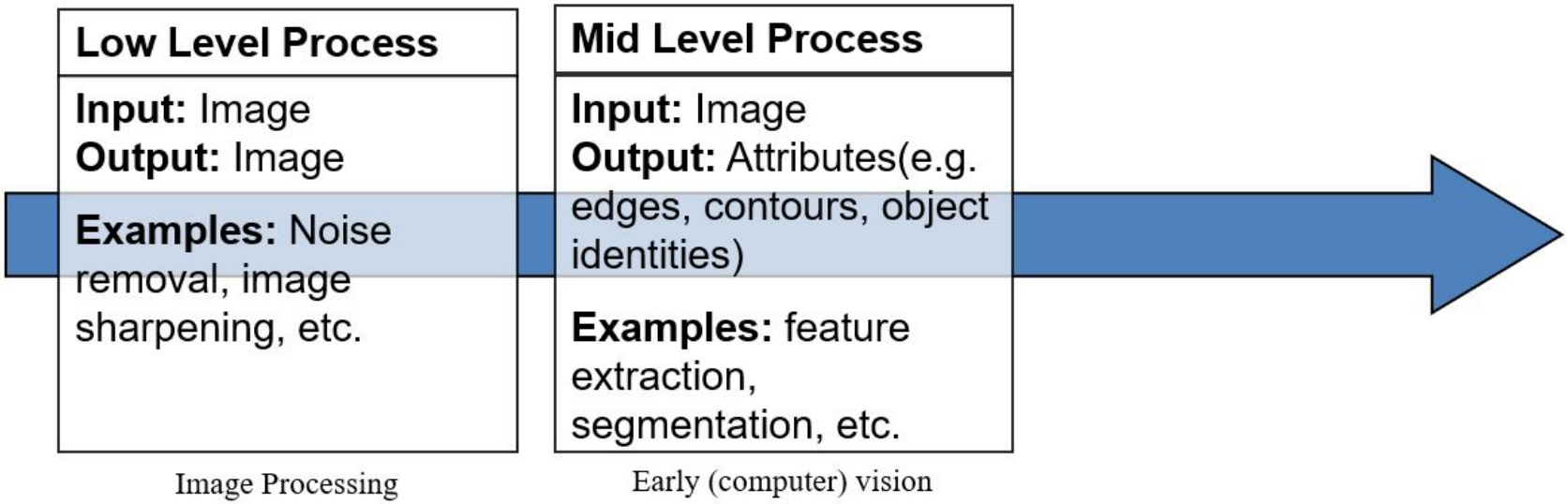
Image Processing

From Image Processing to Computer Vision

Figure : Example of Low Level Processing

From Image Processing to Computer Vision

Low, mid and high-level processes



From Image Processing to Computer Vision

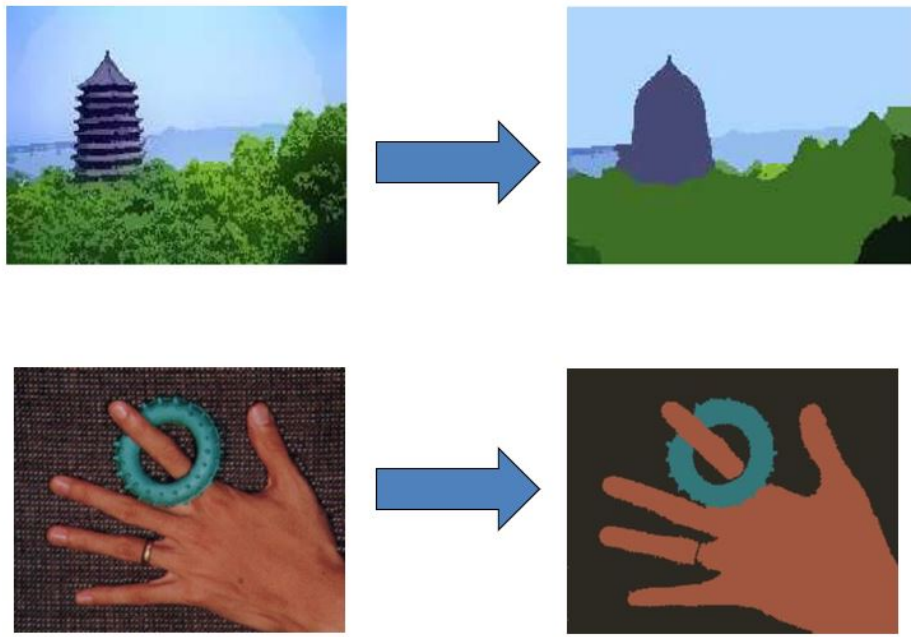
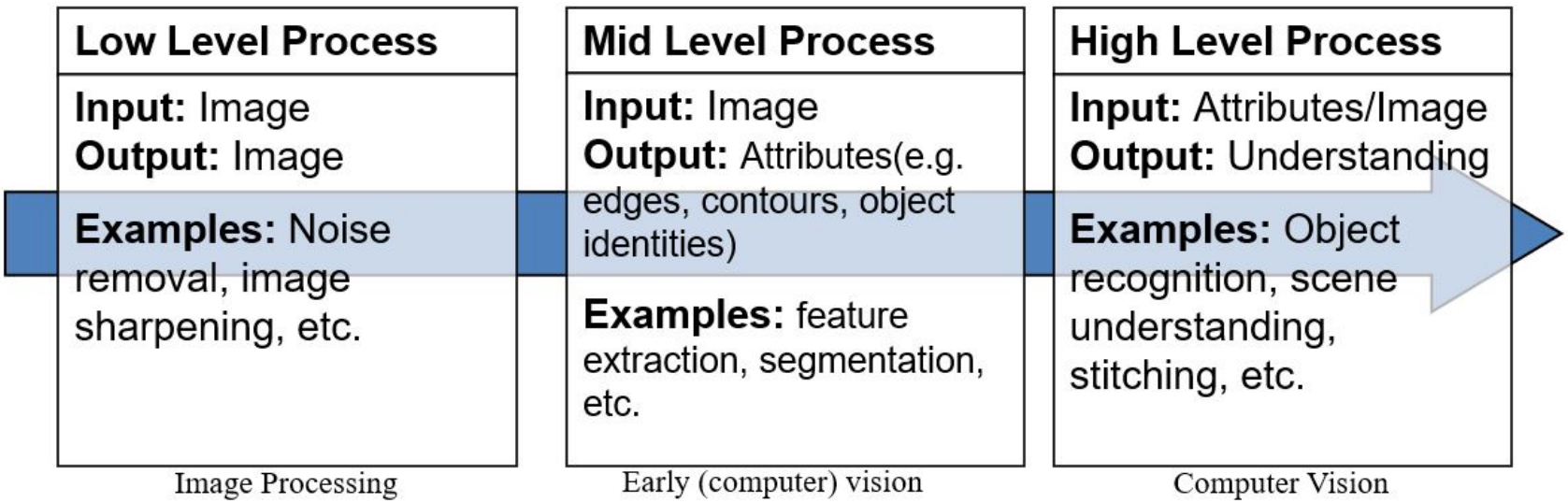


Figure : Example of Mid Level Processing - Segmentation[2] of an image into regions

[2] Image segmentation is a commonly used technique in digital image processing and analysis to partition an image into multiple parts or regions, often based on the characteristics of the pixels in the image.

From Image Processing to Computer Vision

Low, mid and high-level processes



From Image Processing to Computer Vision

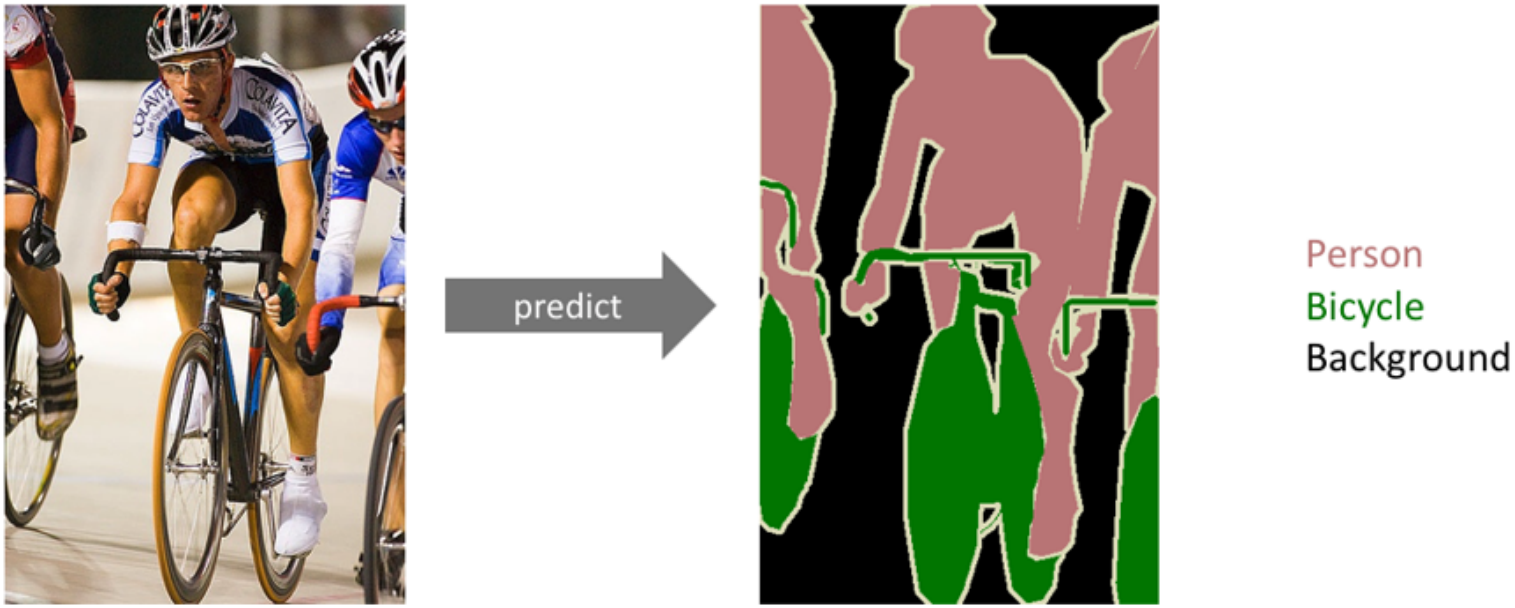
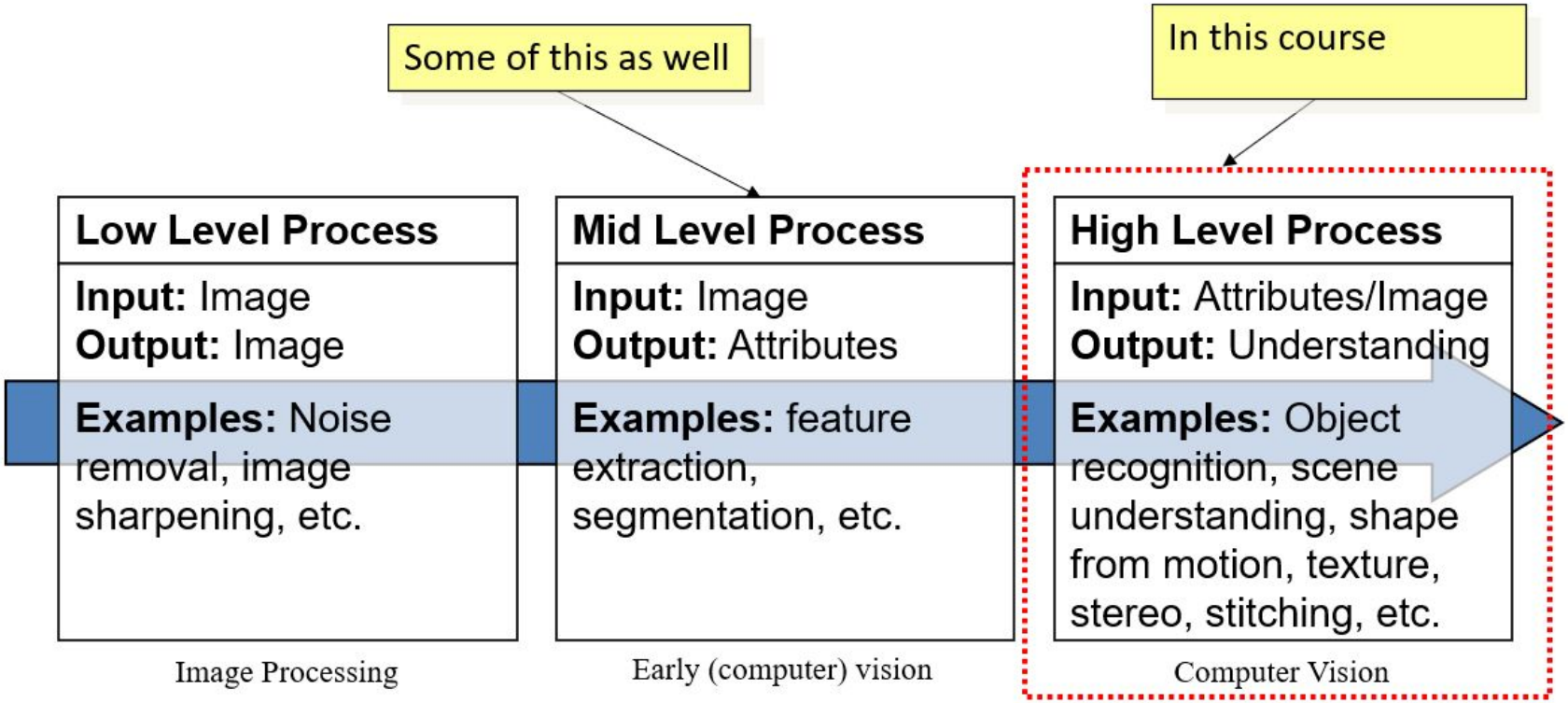
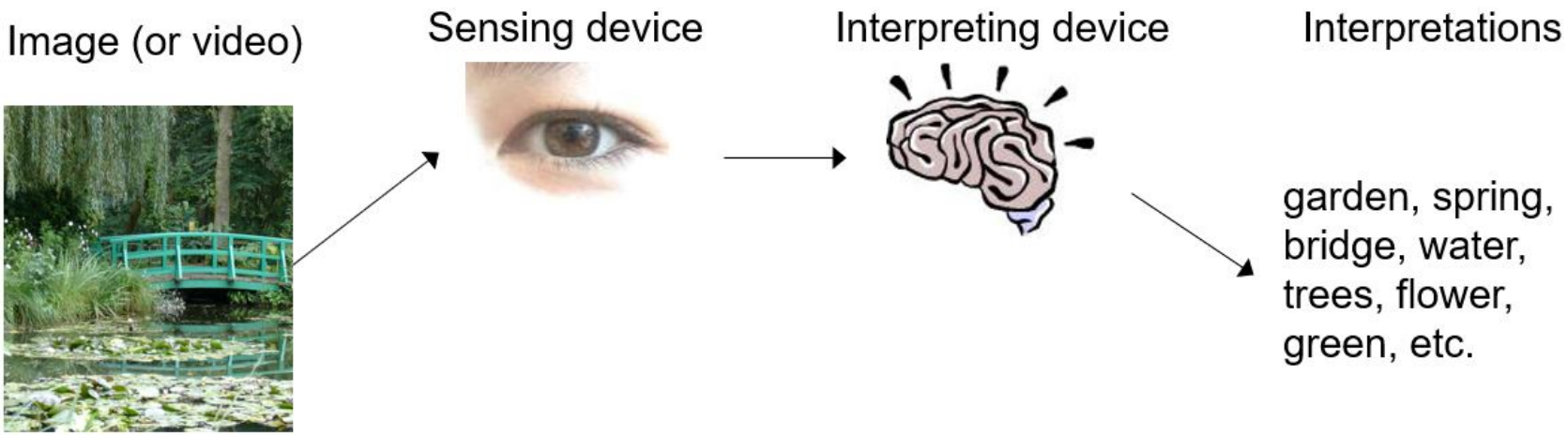


Figure : Example of High Level Processing

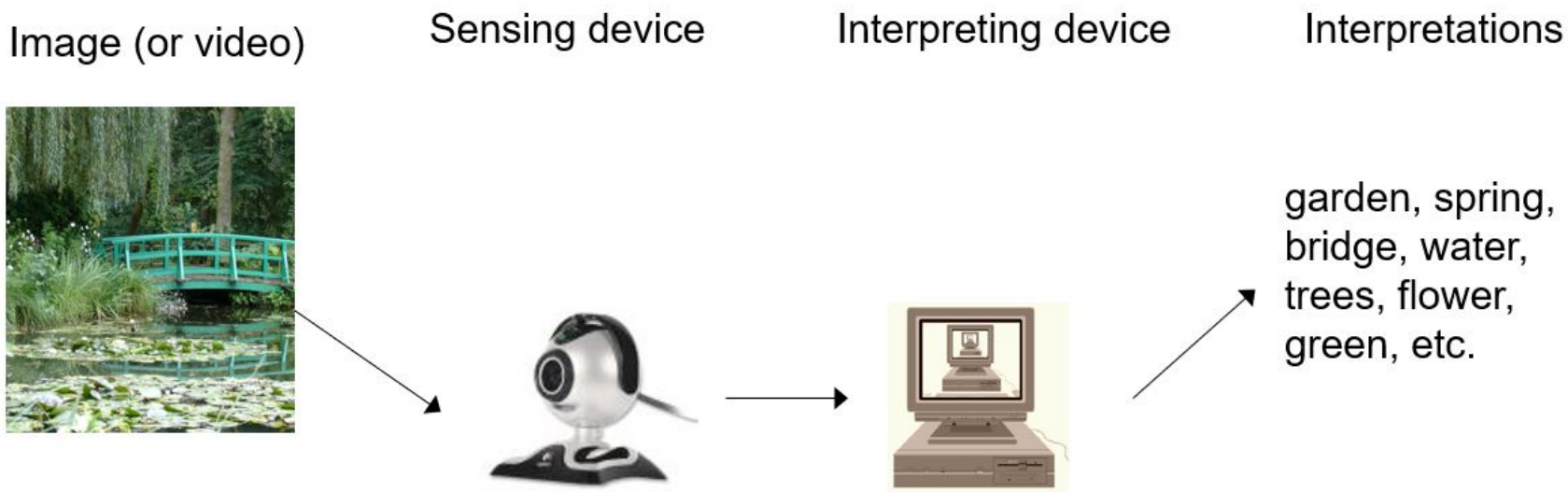
From Image Processing to Computer Vision



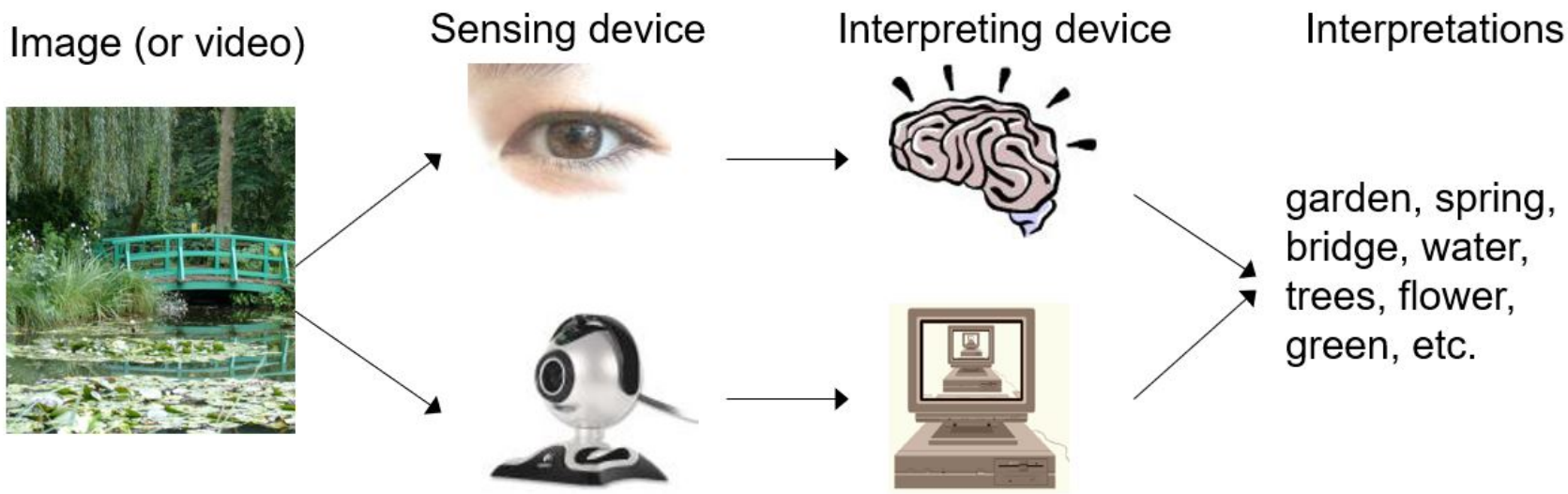
What is vision?



What is (computer) vision?



What is (computer) vision?



What is computer vision?

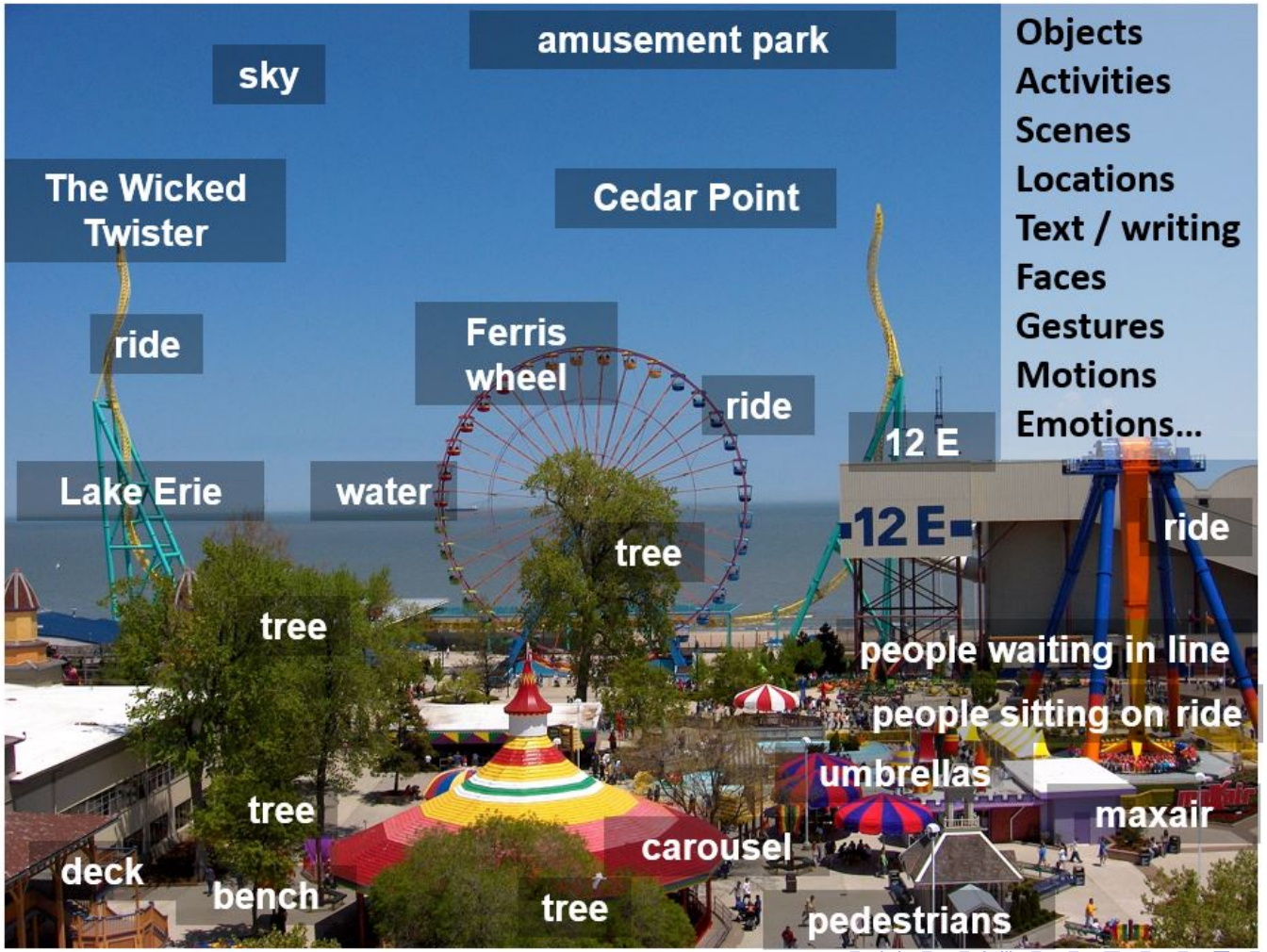
The ability of computers to see.

- Extract descriptions of the world from pictures or sequences of pictures (videos) [Forsyth & Ponce, 2003].
- "Vision is the process of discovering from images what is present in the world, and where it is" (David Marr).
- Computer vision is the study of the analysis of pictures and videos in order to achieve results similar to those done by human.

Computer Vision

- Make computers understand images and videos.
- What kind of scene?
- Where are the cars?
- How far is the building?

Vision for perception, interpretation



The goal of computer vision

To bridge the 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

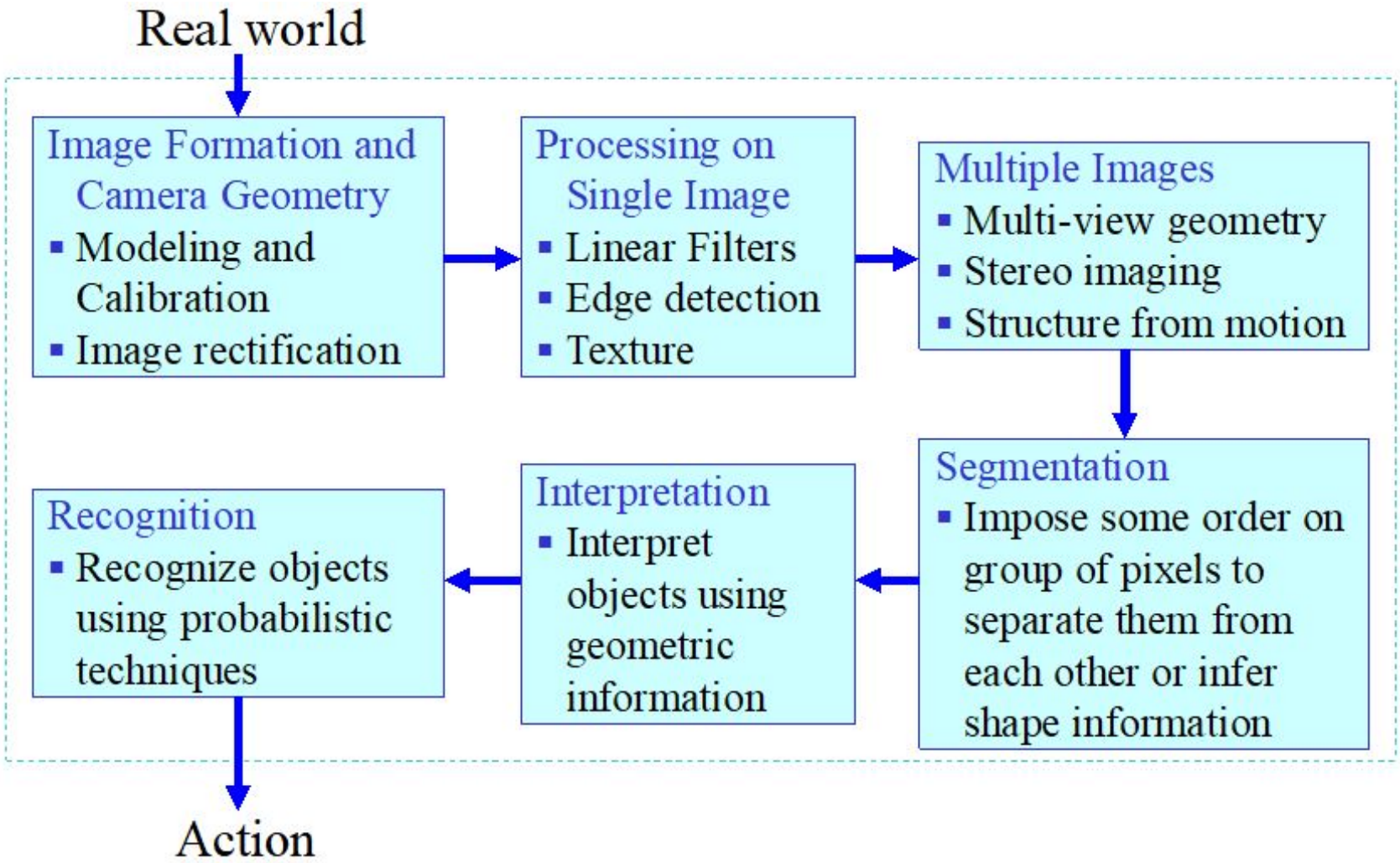
What a computer sees

Focus of computer vision

Computer Vision focuses on the following problems:

- What information should be extracted from the visual sensors?
- How is this information extracted?
- How should it be represented?
- How the information be used to allow a robotics system to perform its tasks?

Overview of classical computer vision



Why vision?

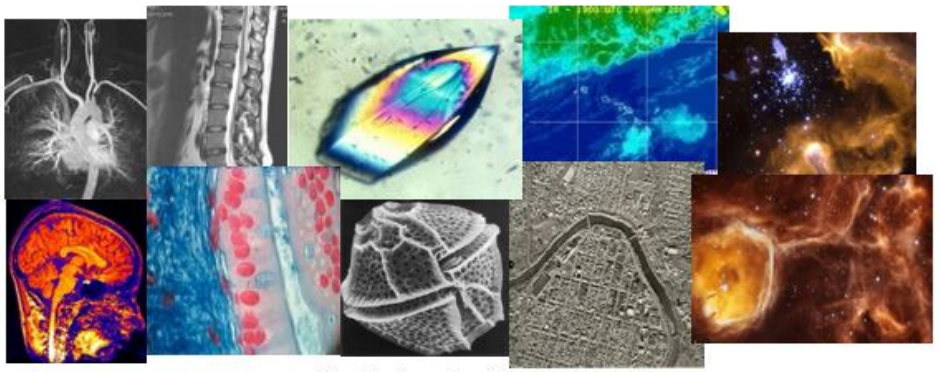
Images are everywhere

Personal photo albums

Movies, news, sports

Google Image Search Picasa flickr webshots picsearch YouTube Broadcast Yourself

Surveillance and security



Medical and scientific images

Why computer vision matters?

Safety

Health

Security

Comfort

Fun

Access

A story about computer vision

In 1966, Marvin Minsky[3] at MIT asked his undergraduate student Gerald Jay Sussman to "spend the summer linking a camera to a computer and getting the computer to describe[4] what it saw". We now know that the problem is slightly more difficult than that. (Szeliski 2009, Computer Vision)

[3] Founder, MIT AI project

[4] Image understanding

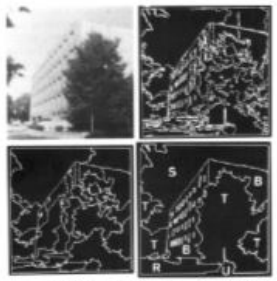
A story about computer vision

In 1966, Marvin Minsky at MIT asked his undergraduate student Gerald Jay Sussman to "spend the summer linking a camera to a computer and getting the computer to describe what it saw". We now know that the problem is slightly more difficult than that. "Needless to say, Sussman didn't make the deadline,". "Vision turned out to be one of the most difficult and frustrating challenges in AI over the next four decades. As machine vision expert Berthold Horn once pointed out, Sussman opted never to work in vision again."

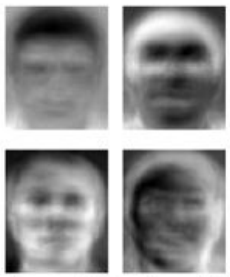
Ridiculously brief history of computer vision

- 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
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts; vision & graphis; vision for HCI; internet vision, etc.

Guzman '68



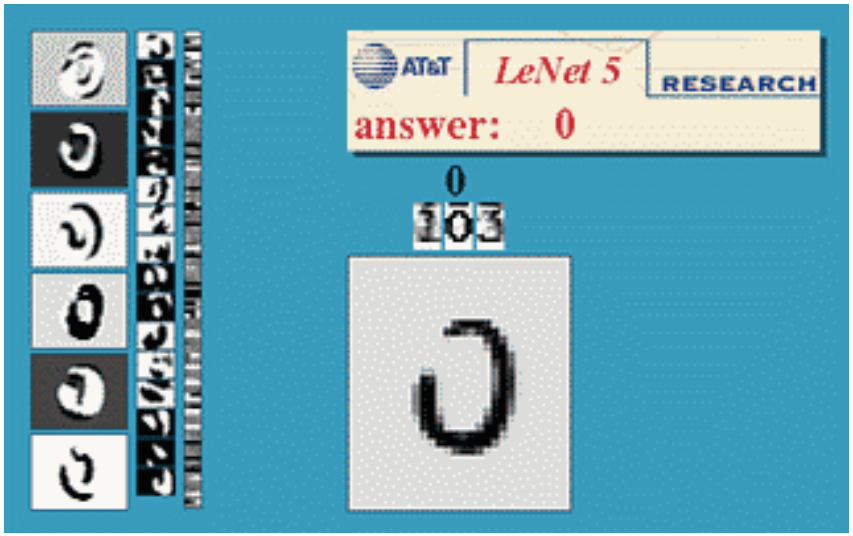
Ohta Kanade '78



Turk and Pentland '91

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

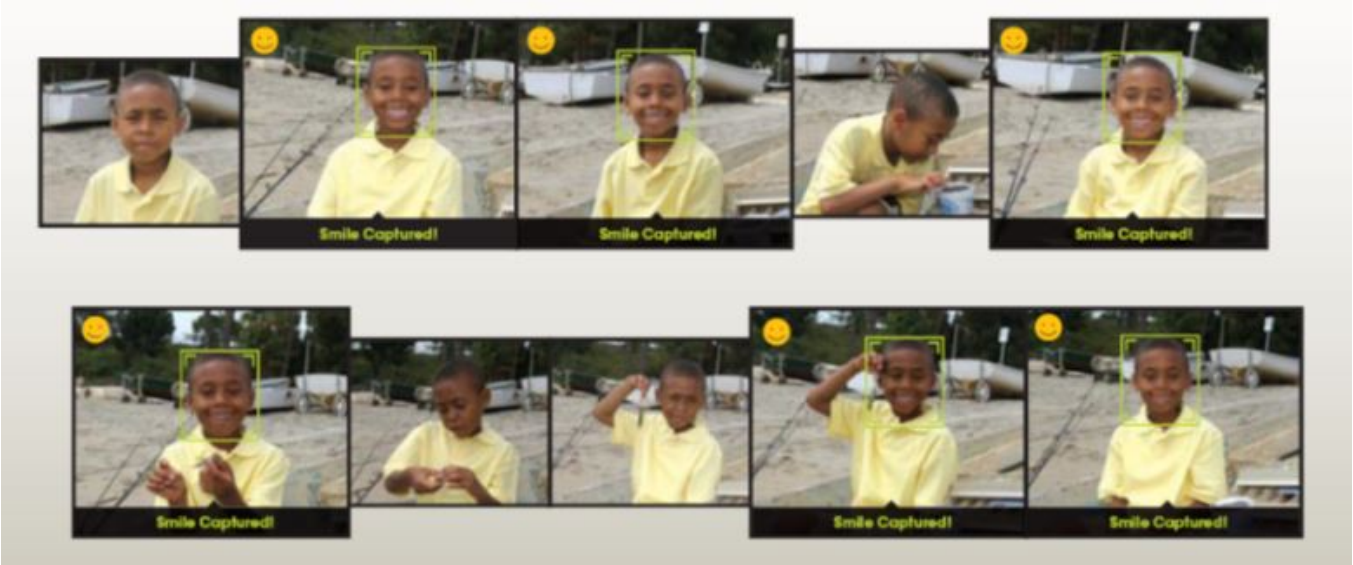
Face detection

Many new digital cameras now detect faces e.g., Canon, Sony, Fuji, ...

Smile detection

The smile shutter flow

- A camera is smart enough to catch every smile.
- In Smile Shutter Mode, camera can automatically trip the shutter at just the right instant to catch the perfect expression.

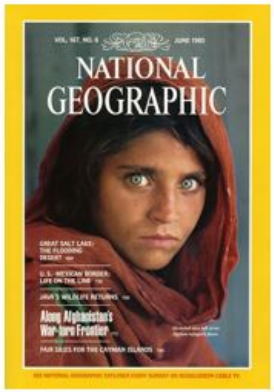


Object recognition (in supermarkets)

LaneHawk by Evolution Robotics

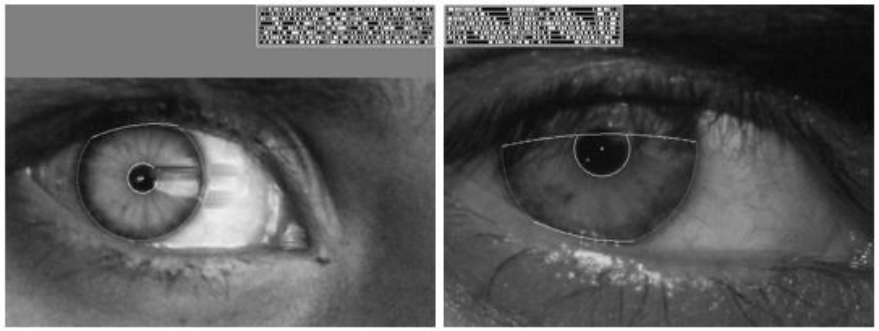
A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you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

Vision-based biometrics



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

`https://www.cl.cam.ac.uk/~jgd1000/afghan.html`



The remarkable story of Sharbat Gula, first photographed in 1984 aged 12 in a refugee camp in Pakistan by National Geographic photographer Steve McCurry, and traced 18 years later to a remote part of Afghanistan where she was again photographed by McCurry. National Geographic accepted and published this conclusion in a second cover issue featuring Sharbat Gula, 18 years after the first, and the Society launched their "Afghan Girl's Fund" to assist the education of Muslim girls in cultures that discourage or prohibit female education.

Login without a password

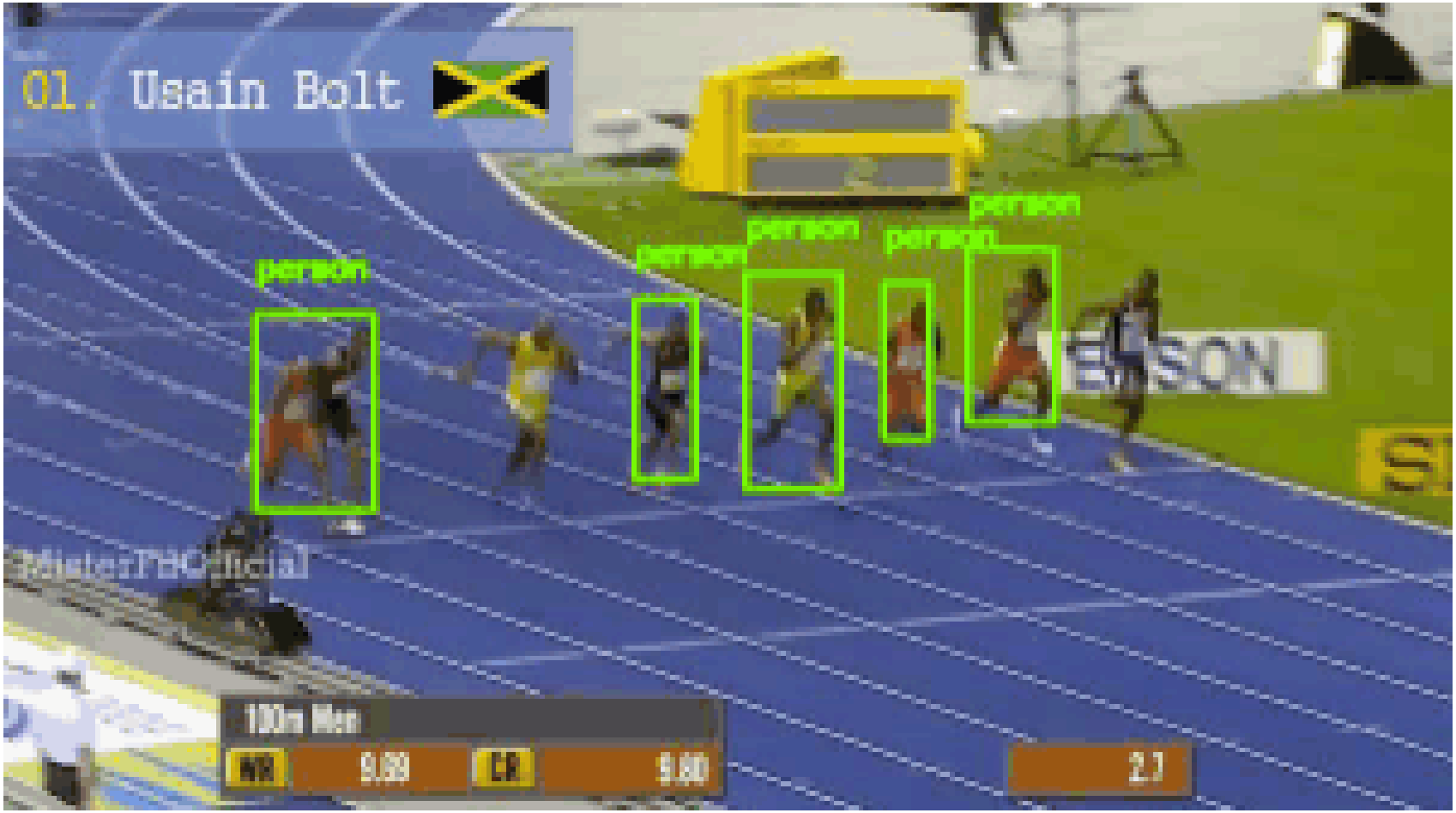
Face recognition systems now beginning to appear more widely

`http://www.sensiblevision.com/`

Object recognition (in mobile phones)

Special effects: motion capture

Sports


01. Usain Bolt
person
person
person
person
person
100m Men
WR
9.69
CR
9.80
2.7

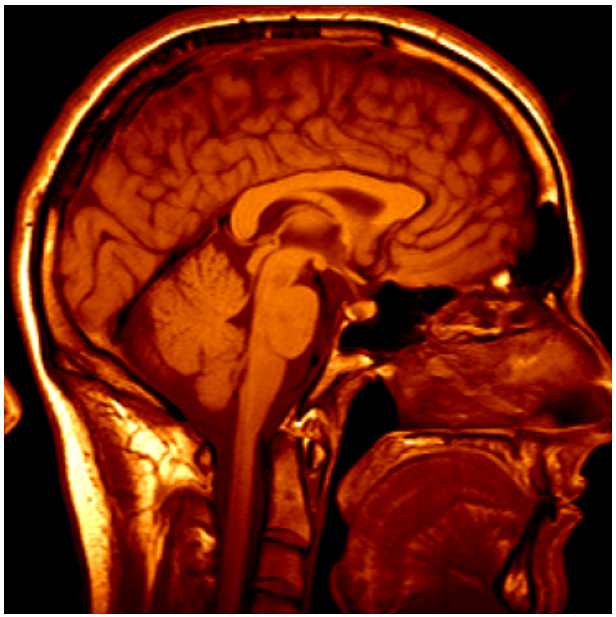
Smart cars

MobileEye (`https://www.mobileye.com/`)
Vision systems currently in many car models

Industrial robots

Vision-guided robots position nut runners on wheels

Medical imaging



3D imaging MRI, CT

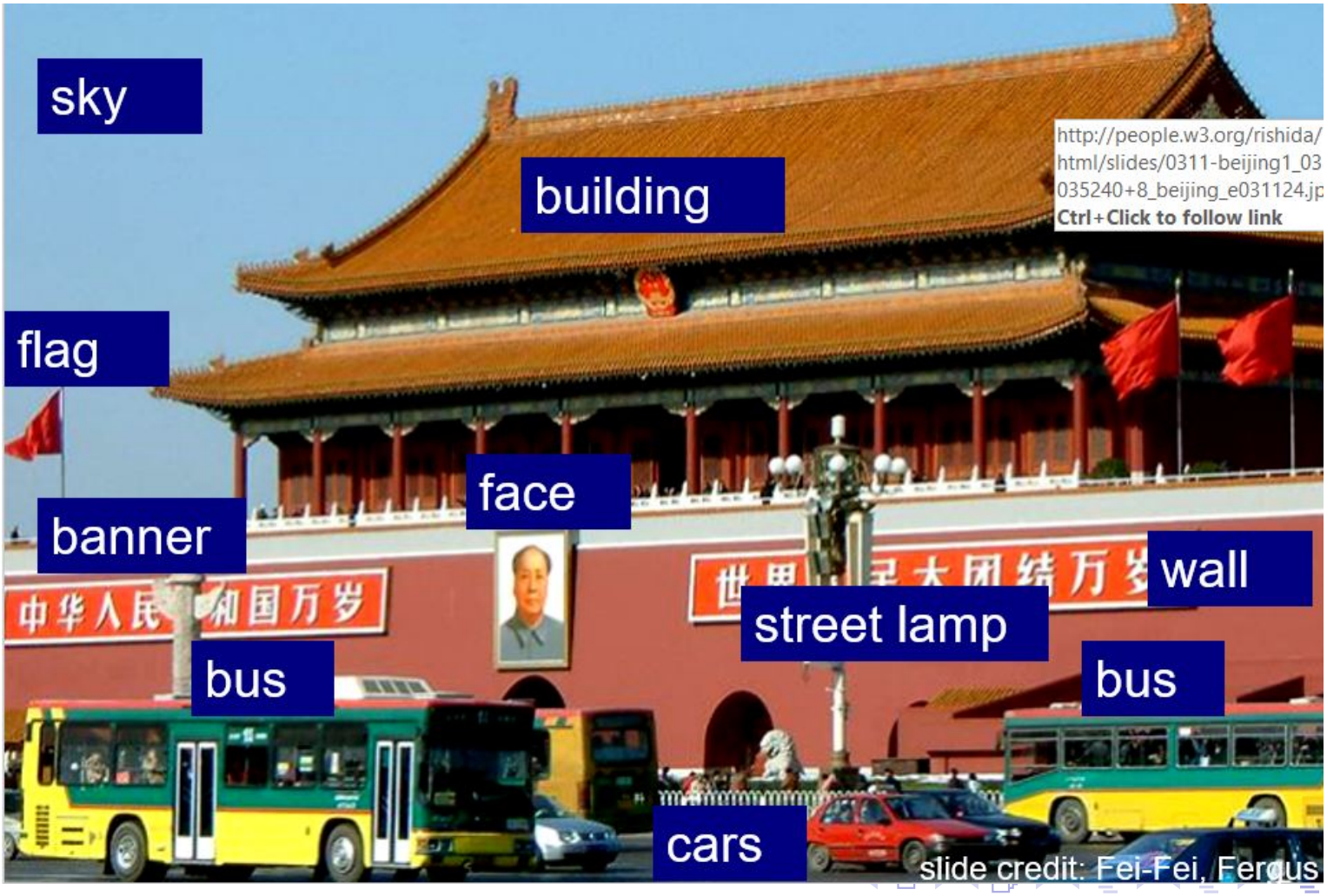
Image guided surgery

https://www.csail.mit.edu/research/medical-vision-group

Vision as a source of semantic information

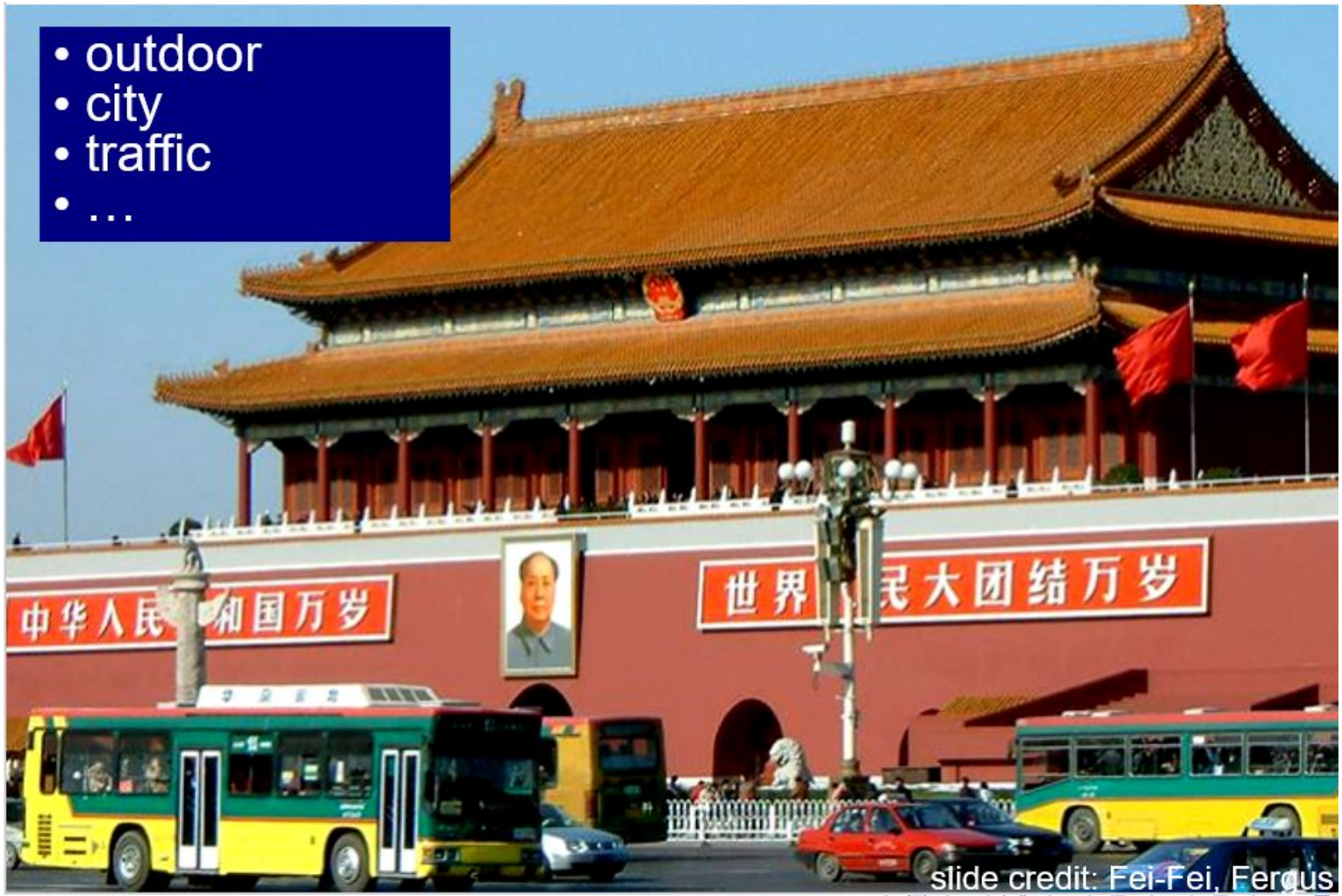
Vision as a source of semantic information

Object categorization



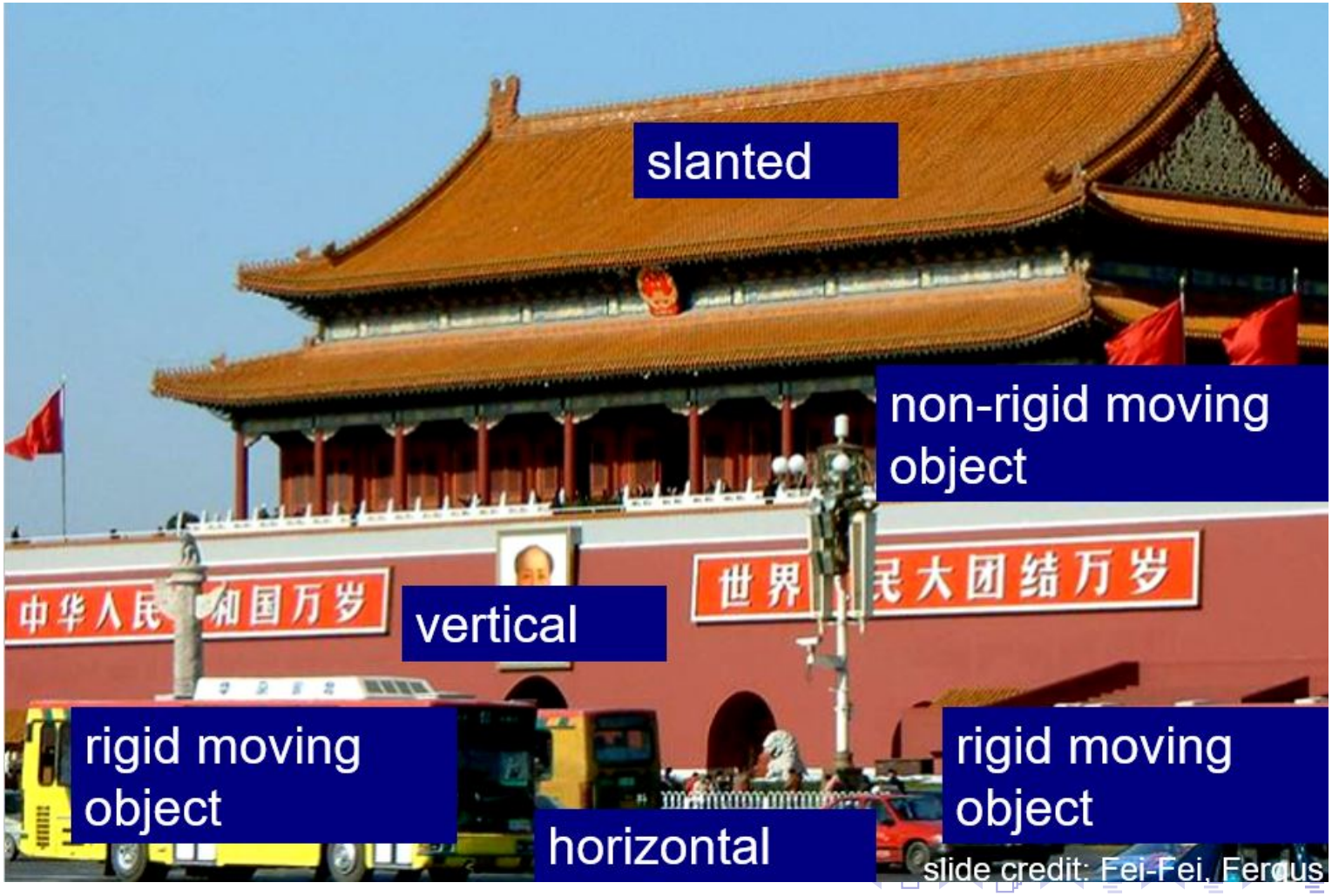
Vision as a source of semantic information

Scene and context categorization



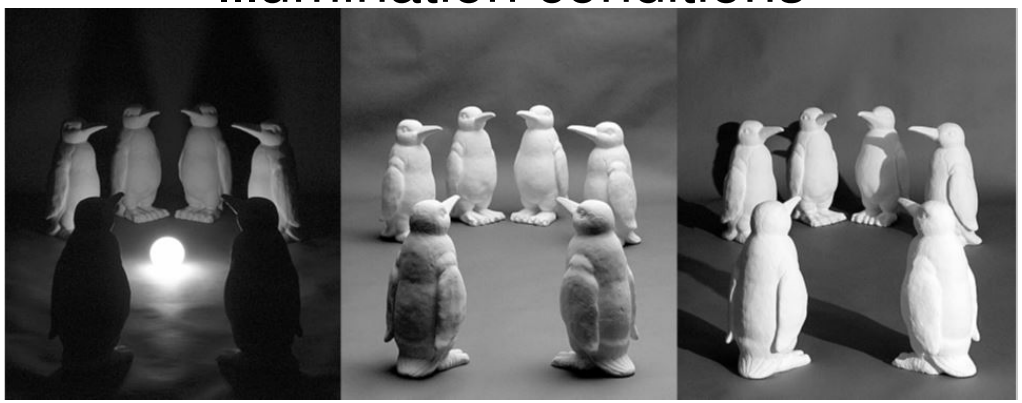
Vision as a source of semantic information

Qualitative spatial information



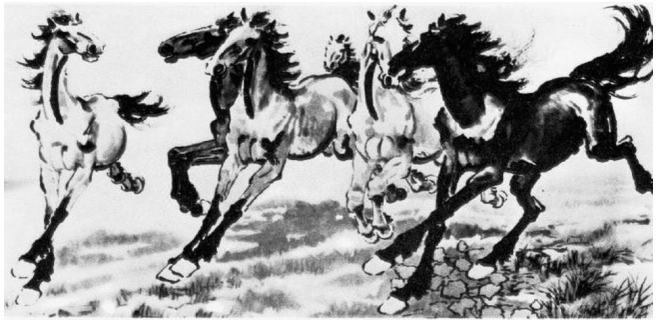
Viewpoint variation

Illumination conditions



Scale variation

Deformation



Background clutter

Occlusion

Intra-class variation