COMPUTER VISION

Jorge Silva
Luís Teixeira

AN INTRODUCTION TO COMPUTER VISION

Jorge A. Silva

Summary

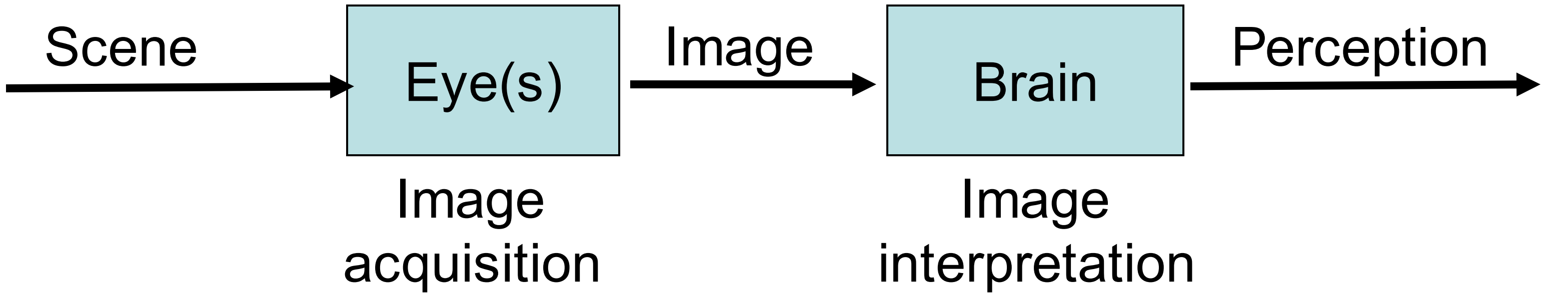
- Introduction
  - Basic concepts, applications, related areas
- Computer vision processes
  - Arquitecture of a computer vision system
- Course overview

What is Computer Vision?

- Goal of Computer Vision
  - Enable a machine to "understand" images and videos
- Automatic understanding
  - Computing properties of the 3D world from visual data (measurement)
  - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities (perception and interpretation)

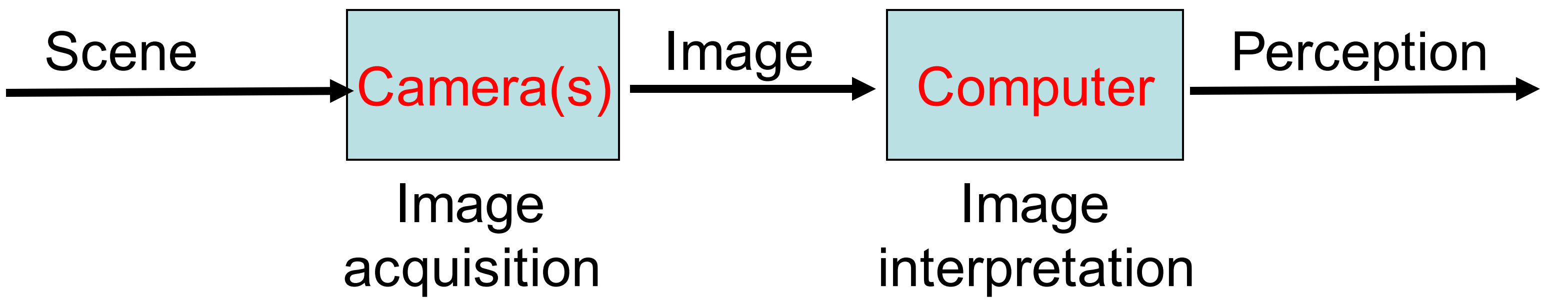
Human vision

- The special sense by which the qualities of an object (as color, luminosity, shape, and size) constituting its appearance are perceived through a process in which light rays entering the eye are transformed by the retina into electrical signals that are transmitted to the brain via the optic nerve
(*Merriam-Webster dictionary*)



Computer Vision

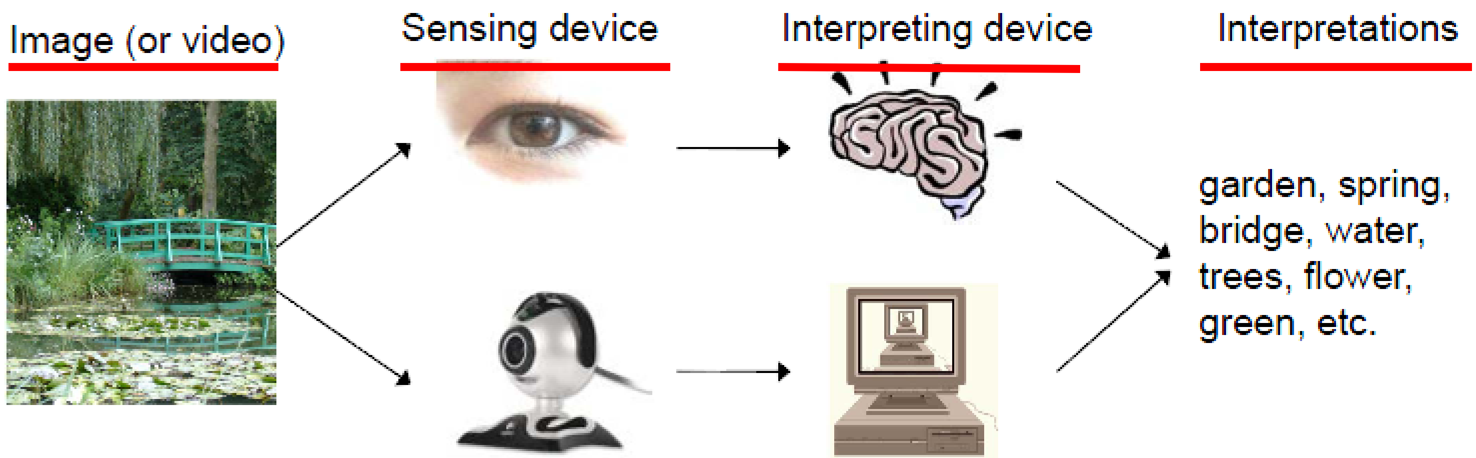
- The objective of Computer Vision is…
  - …obtain useful information about a scene, from images of the scene
  - …obtain explicit and significative descriptions of the scene represented in an image
  - … … … give machines visual capabilities identical to the visual capabilities of an human being …



Computer Vision

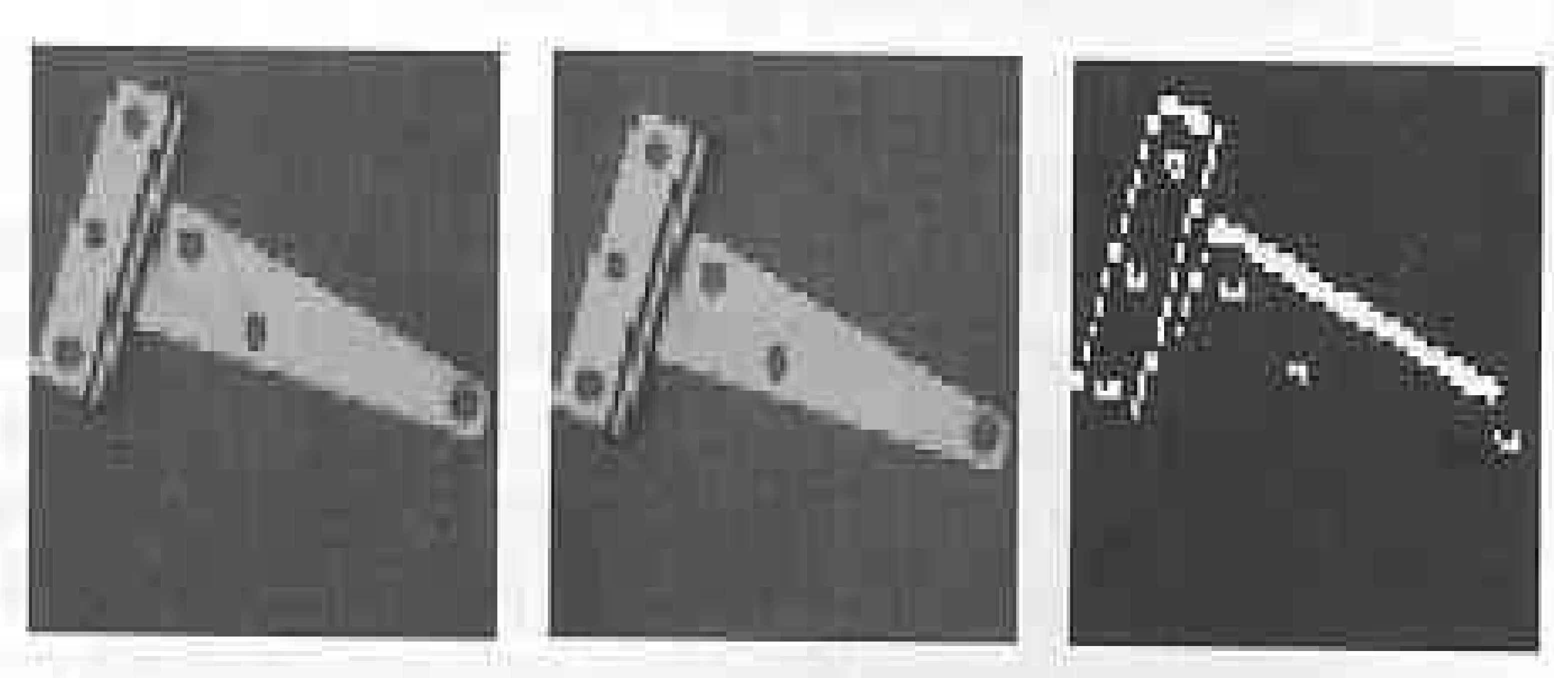
- Some more definitions / objectives :
  - To recover information about a scene from 2D projections of the scene (R. Jain, R. Kasturi, & B. Schunck)
  - To calculate the geometrical and dynamical properties of the 3D world from one or more digital images. (E.Trucco & A.Verri)
  - To build a description of a scene from images.
  - To find out, from images, what is present in the world, where things are located, what actions are being executed (David Marr,1982)
  - To enable a computer to "understand " the environment that surrounds it from visual information.
  - The goal of computer vision is to make useful decisions about objects and real scenes from images acquired. (L.G. Shapiro & G.C. Stockman)

Human / Computer vision



The problems of Computer Vision

- Easy problems:
  - How many holes does the part have ?
  - Is the part complete ?
  - Is it moving ?



Motion detection by image subtraction

The problems of Computer Vision

- Much complex:
  are there people / animals / cars / trees
  in any one of these images ?

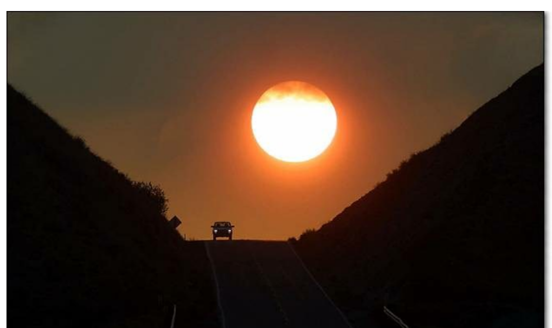
Mark Garkfinkel / The Boston Herald

Phil Noble / AP

Sydney Morning Herald

Jeff J. Mitchell / Reuters

Eric Miller / Reuters



Casey Christie / The Bakersfield Californian

Andy Barron / Reno Gazette-Journal

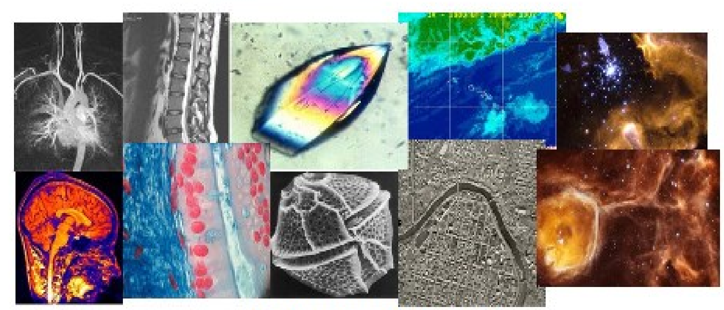
…and now ? …

" Vision is the art of seeing things invisible."
Jonathan Swift (1667-1745), "Thoughts on Various Subjects"

Why study computer vision ?

- Interesting
- Useful
- Difficult
  - half of the cerebral cortex of primates is dedicated to vision
  - to achieve a visual perception similar to that of humans is probably an AI-complete problem (*see Wikipedia*)

Surveillance and security



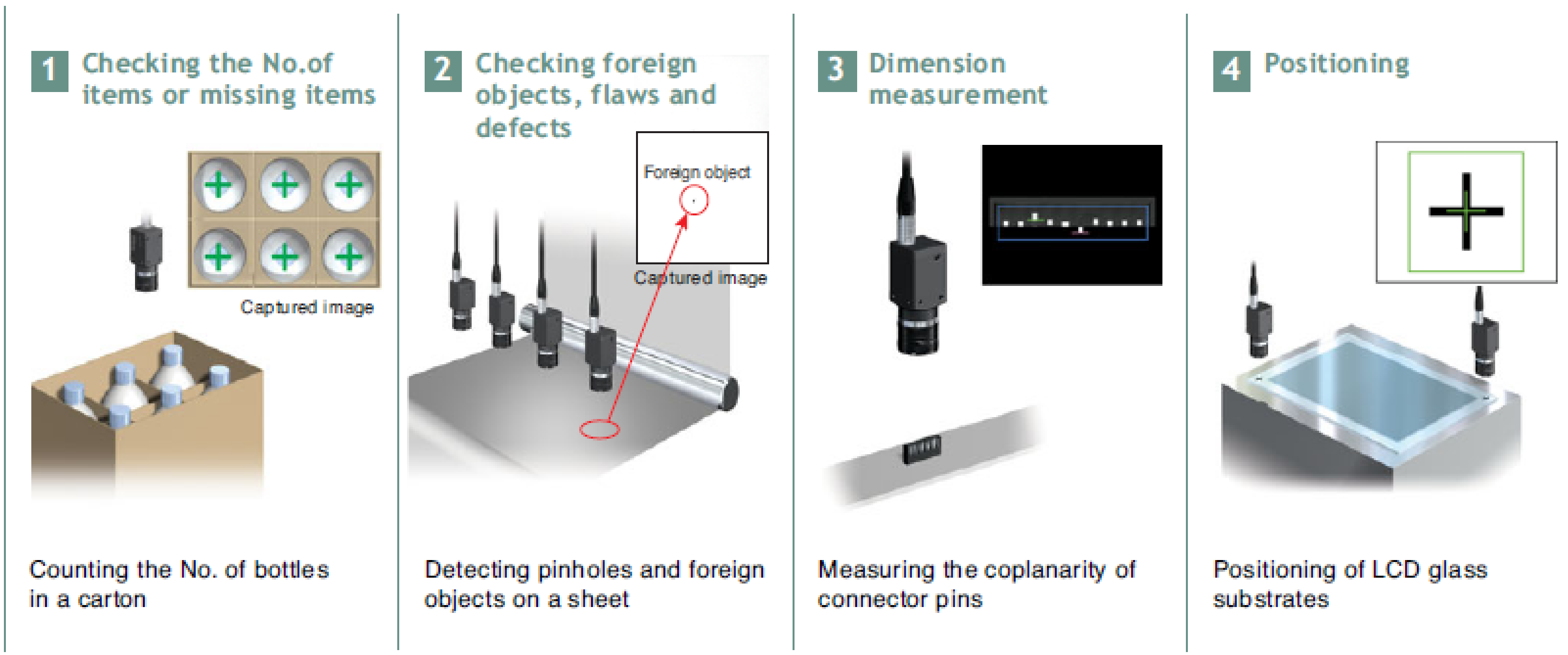
Medical and scientific images

Computer vision applications

- Industrial
- Military / Space
- Surveillance
- Traffic monitoring
- Autonomous vehicles
- Robotics
- Medical
- Face and gesture recognition
- Image-based rendering
- Augmented reality
- *… and many more*

Applications

- Industrial inspection: 4 typical applications



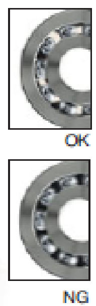
source: Keyence

Applications

- Industrial inspection: other applications

Presence/absence and size detection

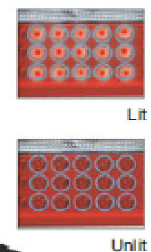
Detecting presence/absence of bearing grease



The color of bearing grease is extracted and the area is measured to detect presence/absence.

Brightness

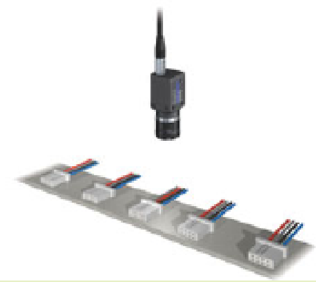
Illumination inspection of LEDs in a tail lamp



Brightness in the inspection area is measured.

Color inspection

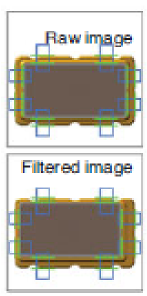
Color identification of flat cables



Items such as hue, saturation and brightness of colors in the inspection area are measured.

Edge measurement

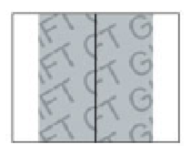
Detecting displaced crystal oscillator covers



Multiple spots are simultaneously measured to detect a displaced cover.

Width measurement

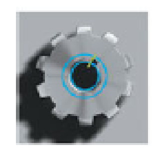
Width measurement of sheeting material



The sides of a sheet are measured with multiple cameras and the results can be combined.

Positioning

Measuring the notch position of a gear



More than two edges are recognized to measure the angle.

Applications

- Industrial inspection: other applications

General machine vision tasks in industrial automation

- Enhancement
  - filter noise or unwanted features
  - remove distortion
  - calibrate images (*)
- Inspection
  - detect defects
  - check colors
  - check presence / absence
  - verify assembly
- Measurement
  - accurately measure objects or features,
    in 2D or 3D
    - detect edges
    - measure distance
    - calculate geometry

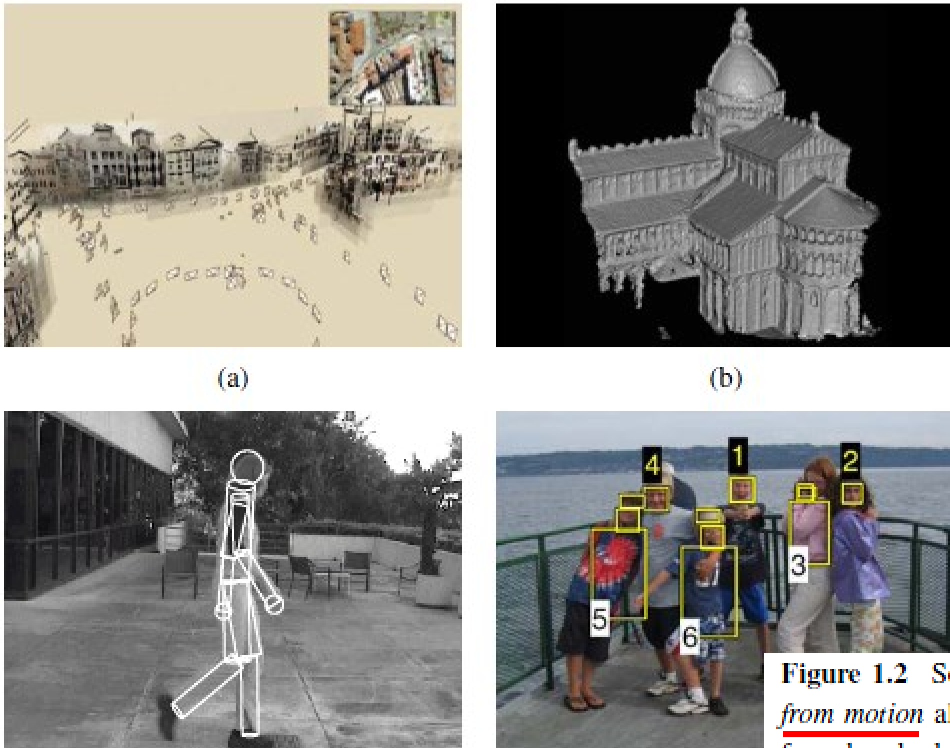
- Location
  - find randomly oriented objects or features,
    in 2D and 3D space,
    for robotic or motion guidance
    - match geometry
    - match patterns
- Identification
  - differentiate objects (**)
  - classify / sort objects (***)

(*) – ex: compensate for changes in the image due to illumination conditions

(**) – ex: read codes (1D or 2D) / characters (OCR)

(***) – Sorting is any process of arranging items systematically,
and has two common, but distinct, meanings:
- ordering: arranging items in a sequence ordered by some criterion;
- categorizing: grouping items with similar properties.

Applications

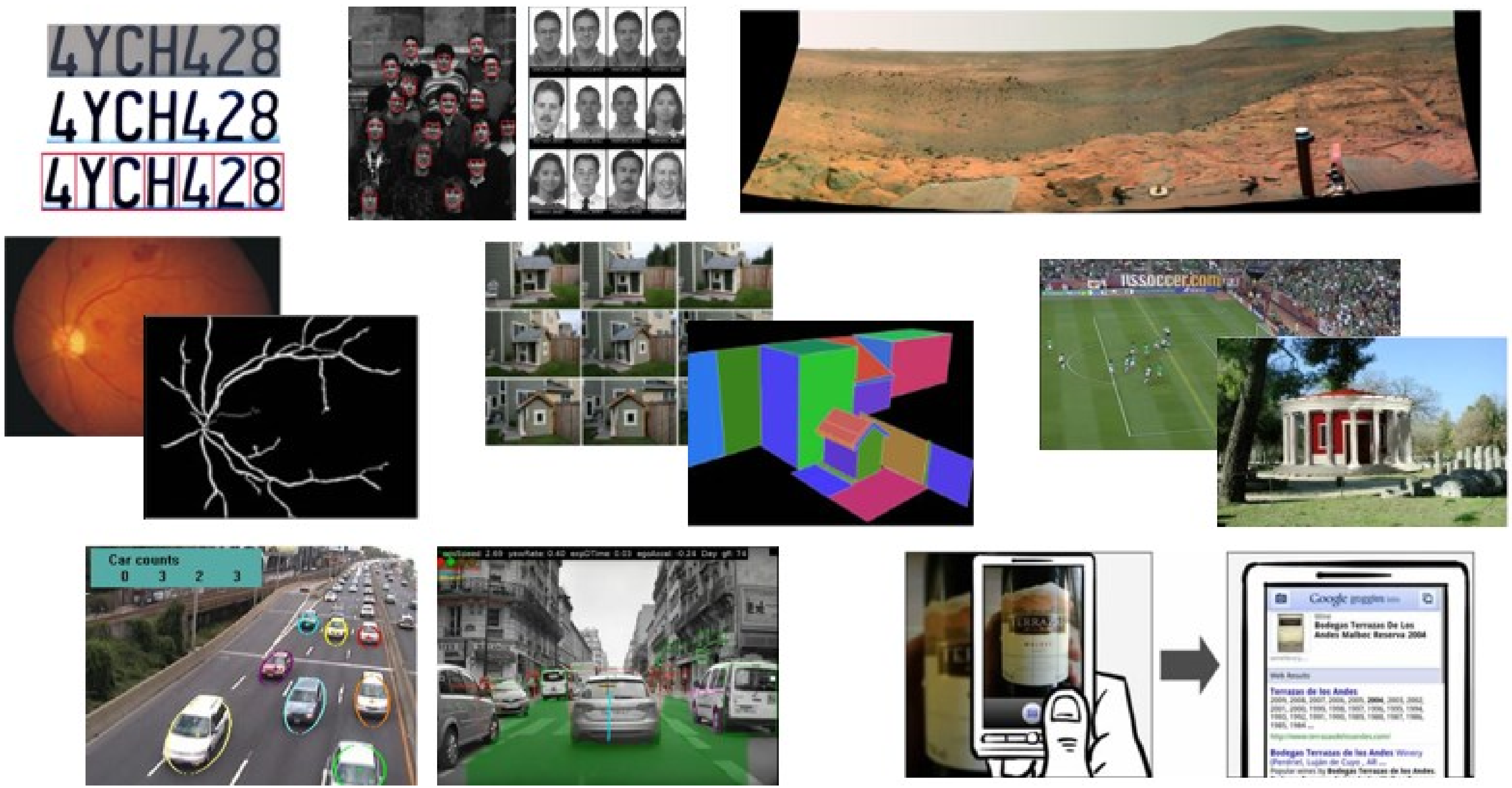


(a) (b) (c) (d)

Figure 1.2 Some examples of computer vision algorithms and applications. (a) *Structure from motion* algorithms can reconstruct a sparse 3D point model of a large complex scene from hundreds of partially overlapping photographs (Snavely, Seitz, and Szeliski 2006) © 2006 ACM. (b) *Stereo matching* algorithms can build a detailed 3D model of a building façade from hundreds of differently exposed photographs taken from the Internet (Goesele, Snavely, Curless *et al.* 2007) © 2007 IEEE. (c) *Person tracking* algorithms can track a person walking in front of a cluttered background (Sidenbladh, Black, and Fleet 2000) © 2000 Springer. (d) *Face detection* algorithms, coupled with color-based clothing and hair detection algorithms, can locate and recognize the individuals in this image (Sivic, Zitnick, and Szeliski 2006) © 2006 Springer.

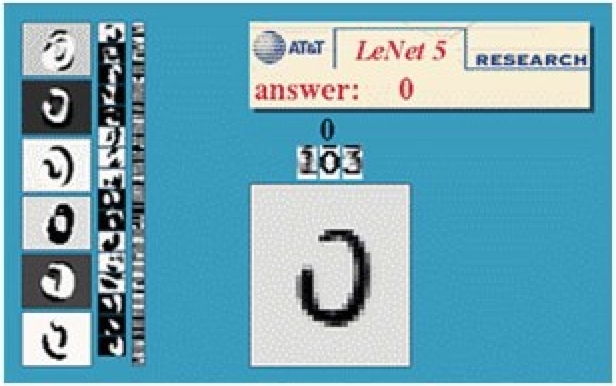
source: Szeliski book

Applications



Applications

- Optical Charater Recognition - OCR



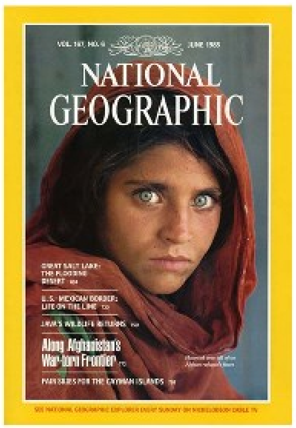
Digit recognition, AT&T labs

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

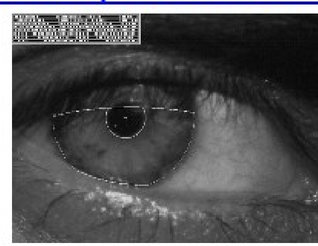
Source: S. Seitz

Applications

- Biometry



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



Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

Source: S. Seitz

Applications

- Face detection

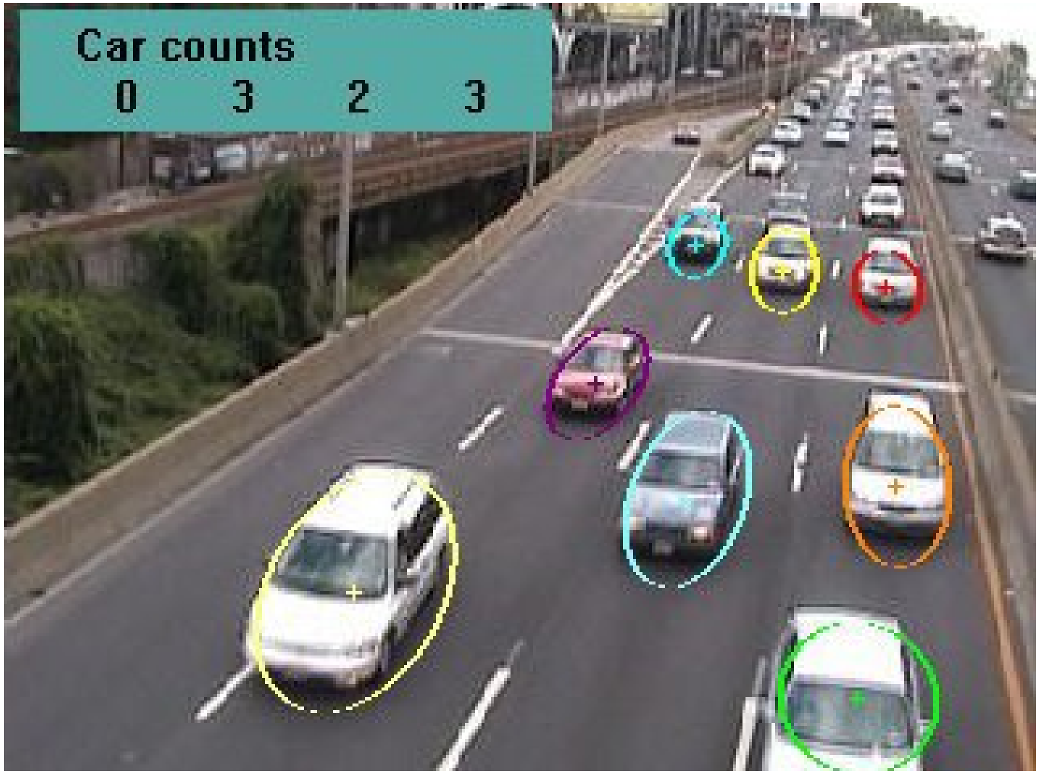
Applications

- Face recognition

Feret database

Applications

- Traffic monitoring



source: Szeliski book

Applications

- Robotic vision (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.

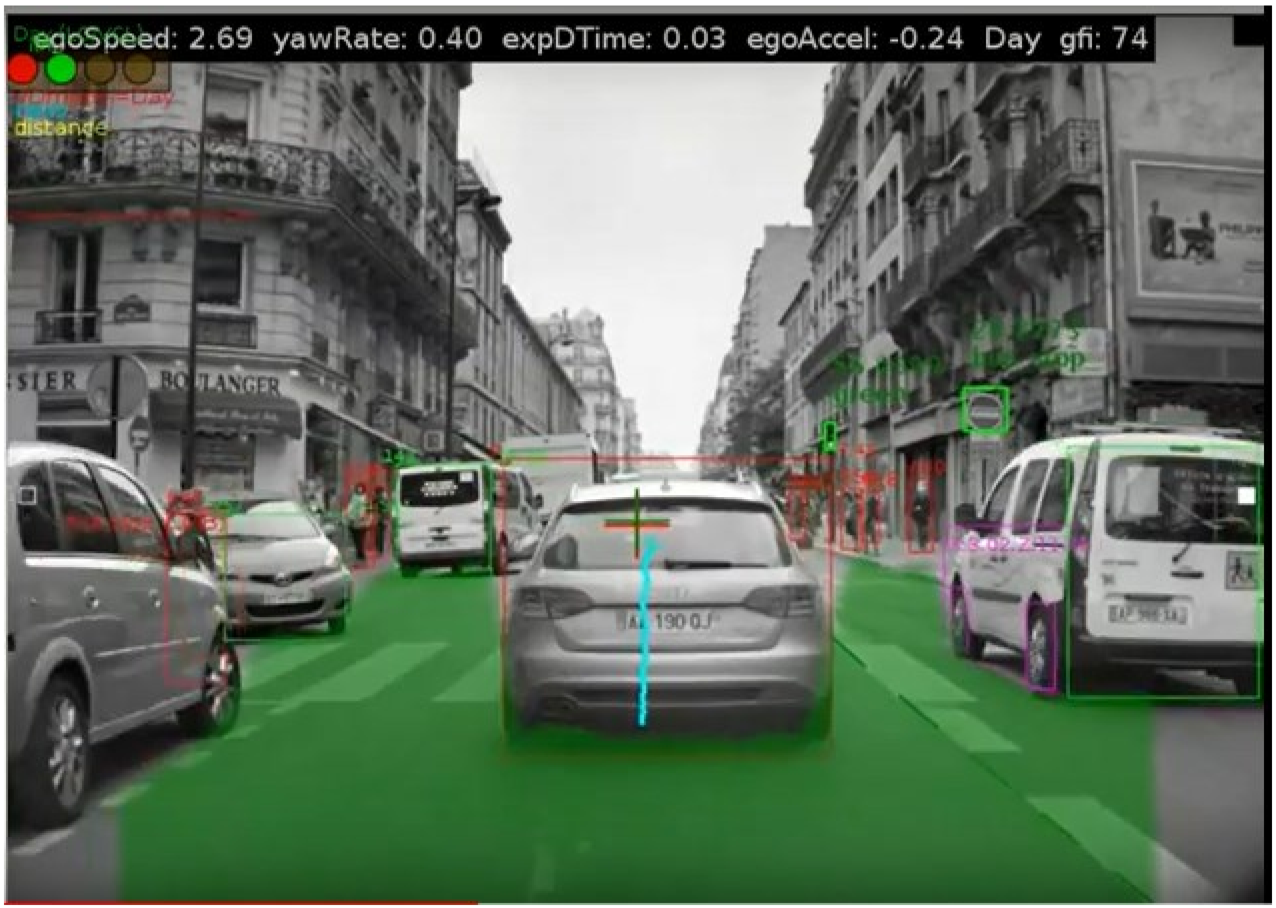
Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “Computer Vision on Mars” by Matthies et al.

Source: S. Seitz

Applications

- Road safety / Autonomous driving



Applications:
1. road sign recognition,
2. lane departure warning,
3. forward collision warning,
4. pedestrian collision warning, and
5. collision mitigation by braking.

The Mobileye system can provide:
- free space (green carpet)
- vehicle and pedestrian detection
- traffic sign recognition
- lane markings
for the vehicle to "understand and negotiate the driving scene"

https://www.mobileye.com/our-technology/

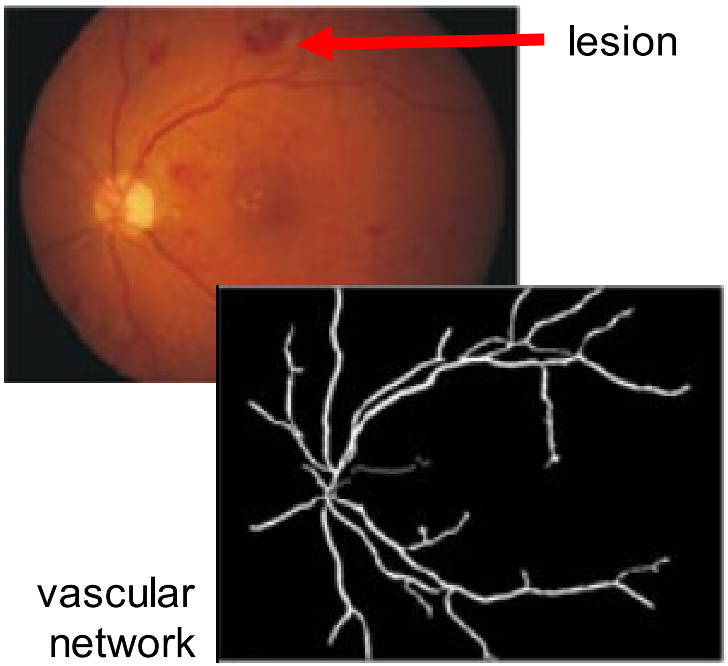
Applications

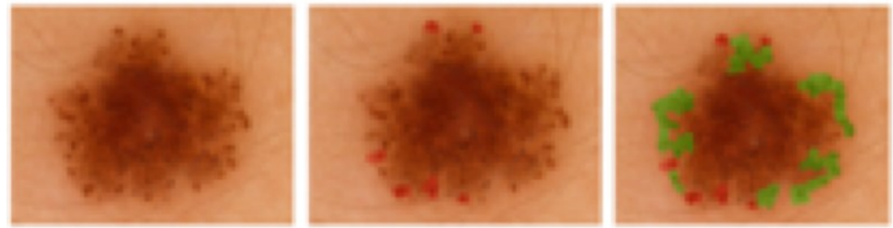
- Television

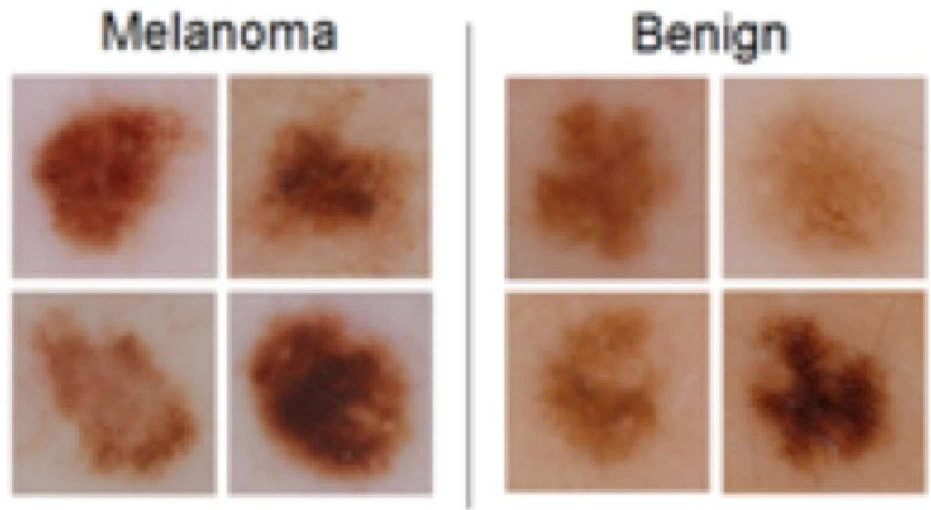
Augmented reality;
the yellow line is inserted in real-time on the TV image
(Azuma, 2001)

Applications

- Medical



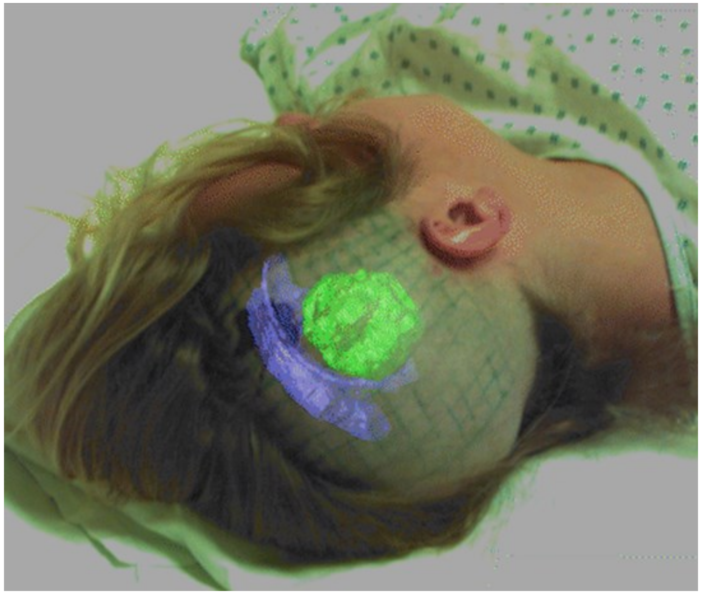




Computer-aided Diagnosis (CAD)

Applications

- Medical



Augmented reality:
superposition of imagens from inside (CT, MRI, …)
to an image from outside the patient

Applications

- Entertainment


Augmented reality:
Reconstruction of the Philippion temple, at Olympia, Greece
(project ARCHEOGUIDE (Augmented Reality-based Cultural Heritage http://archeoguide.intranet.gr)

Applications

- Image stitching

http://labs.live.com/photosynth/

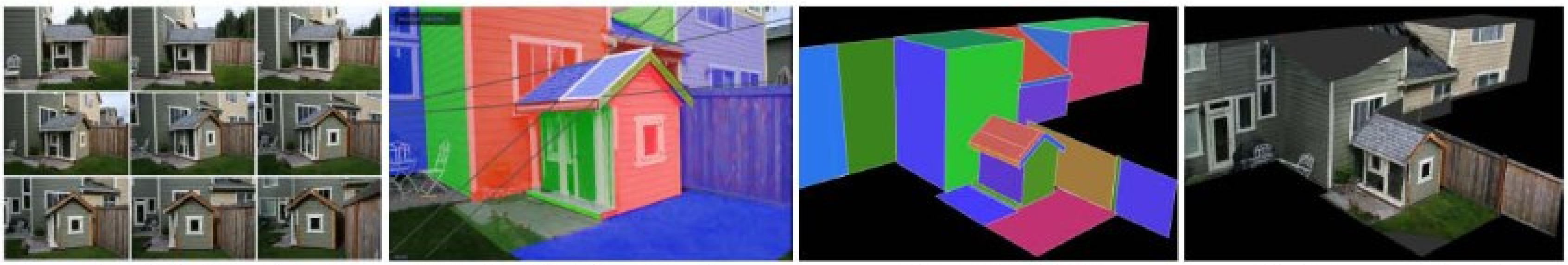
Source: S. Seitz

Applications

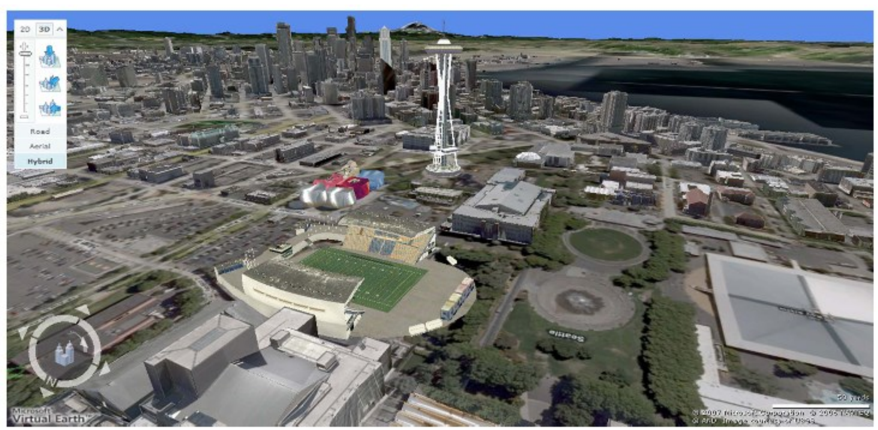
- 3D acquisition / reconstruction (from 2D)

Applications

- 3D urban modelling



Input Photographs | 2D Sketching Interface | Geometric Model | Texture-mapped model



Applications

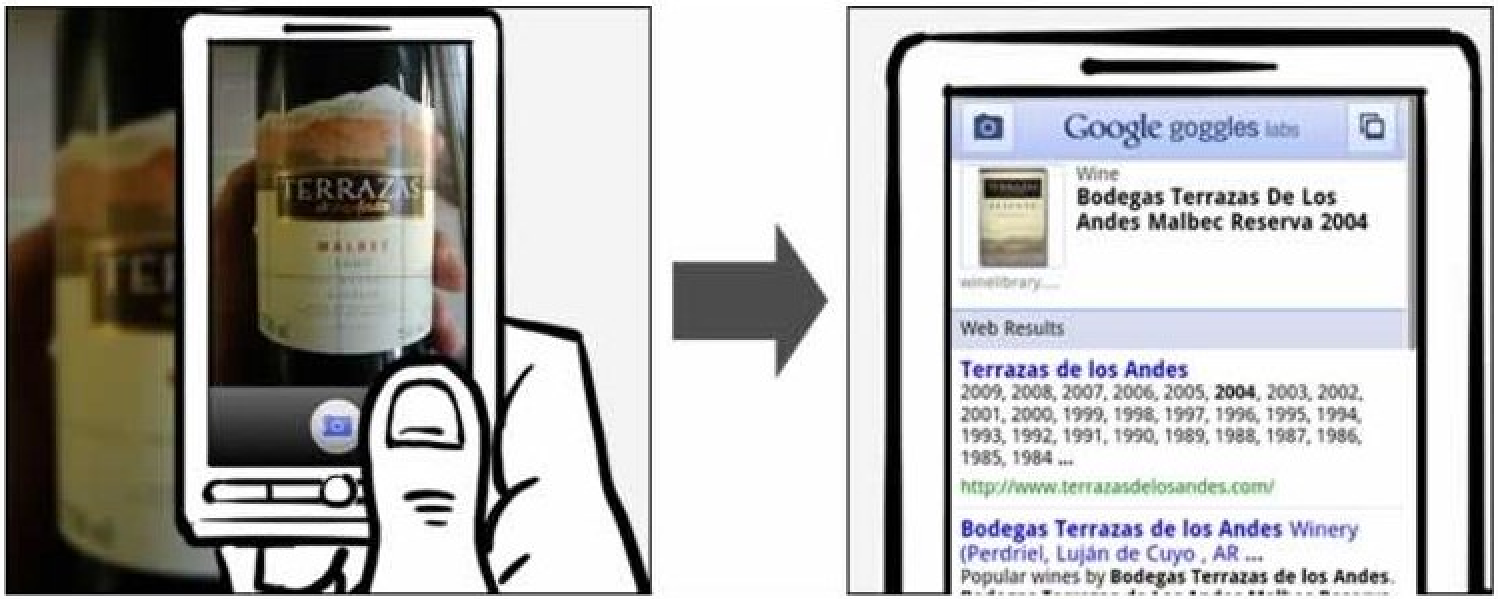
- Games (vision-based interaction)

Sony Eyetoy

Microsoft Kinect

Applications

- Vision for mobile phones



Take photos of objects as queries for visual search

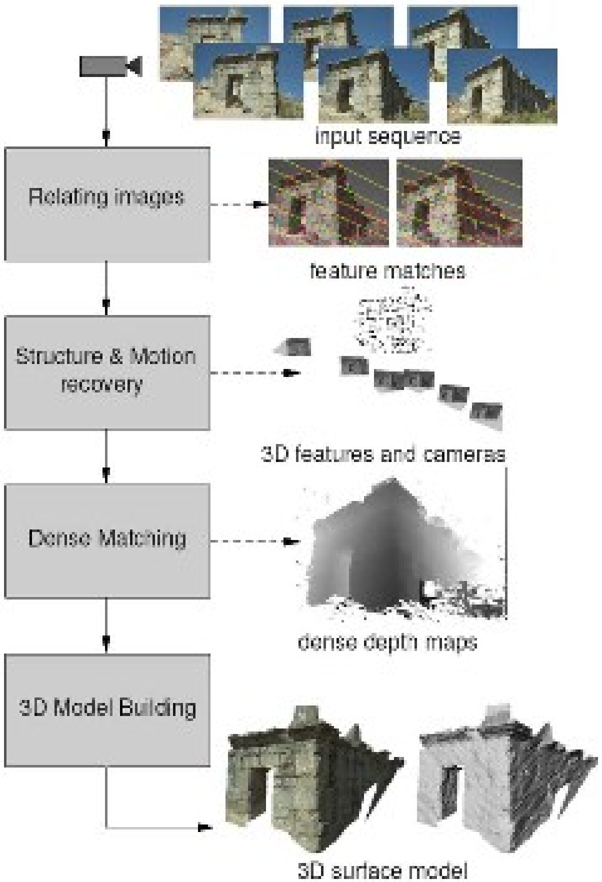
Applications

- Search for
  <u>other</u> computer vision <u>applications</u>

- Try to <u>identify</u> <u>some problems</u>
  that will arise when you want to implement
  a possible a solution

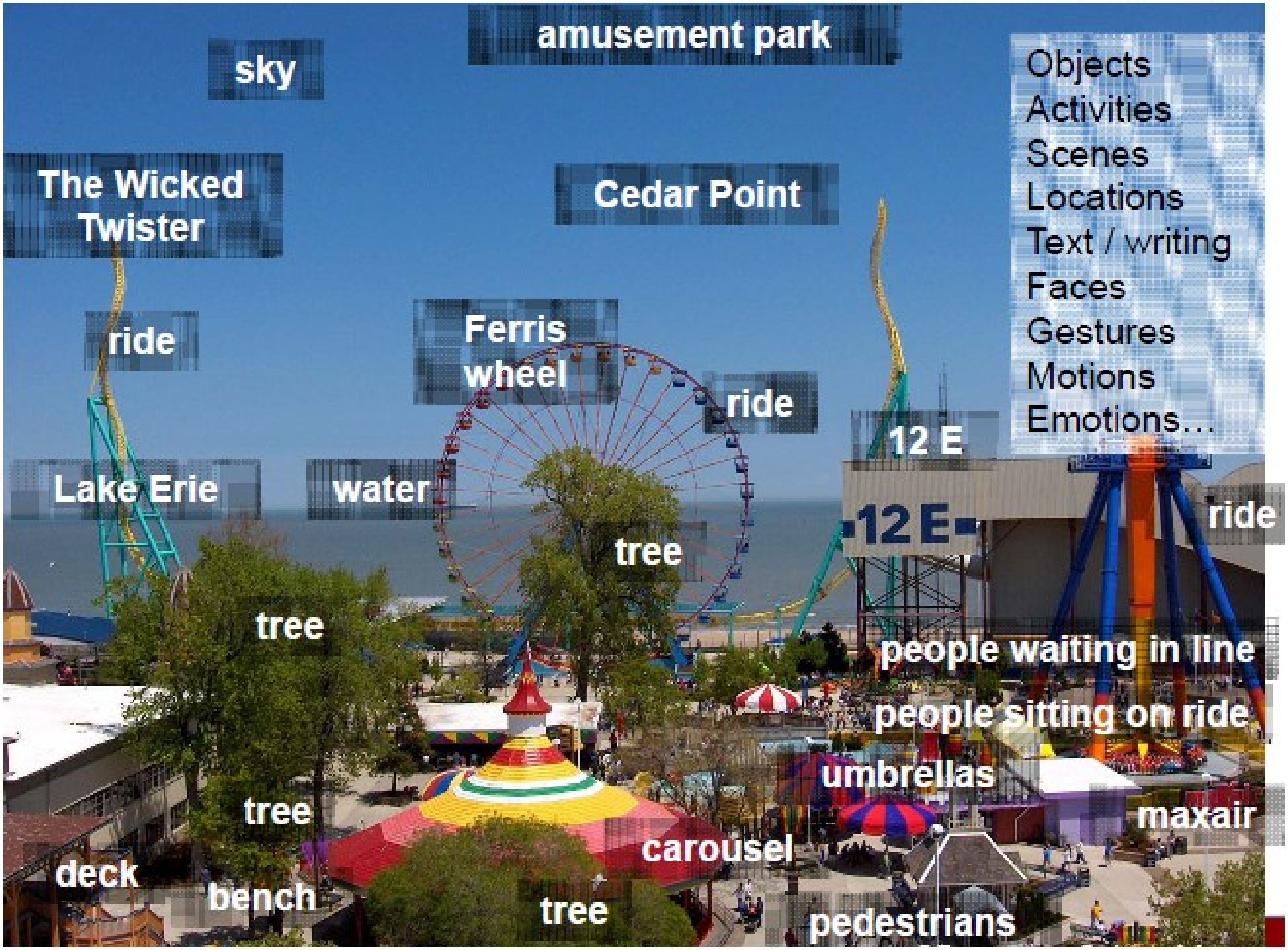
Vision for measurement

source: Keyence



source: Pollefeys at al.

Vision for perception, interpretation



Why is vision difficult ?

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"...

- The objective of computer vision:
  - "fill the gap" between the pixel values and their meaning …

Why is vision difficult ?

- Variations in objects of a same class
  - Color, texture, size, shape, …
- Variations in the acquired images
  - Lighting (brightness, contrast, shadows, ...)
  - Environment: background perturbations
  - Perspective distortion, different viewpoints, occlusion, motion, ...
  - Optical characteristics of the lens
  - Electronic characteristics: sensor resolution, dynamic range, noise, …
- Huge amounts of data
  - 1 minute of color video with 1024x768 resolution = 4.2 gigabyte (without compression)

- About half of the cerebral cortex in primates is devoted to processing visual information
[Felleman and van Essen 1991]

Computer vision challenges

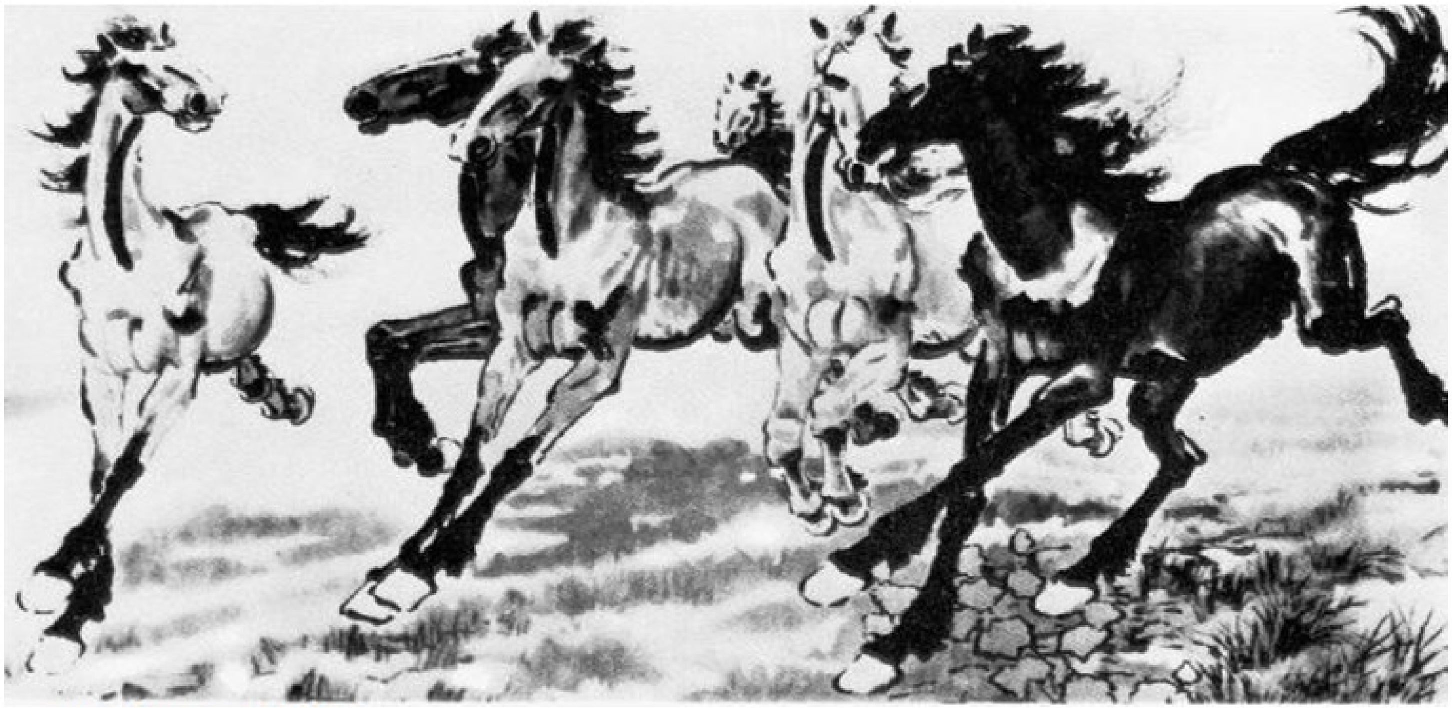
- Intra-class variation






Computer vision challenges

- Shape variations (deformations)



Xu, Beihong 1943

slide credit: Fei-Fei, Fergus & Torralba

Computer vision challenges

- Scale variations

slide credit: Fei-Fei, Fergus & Torralba

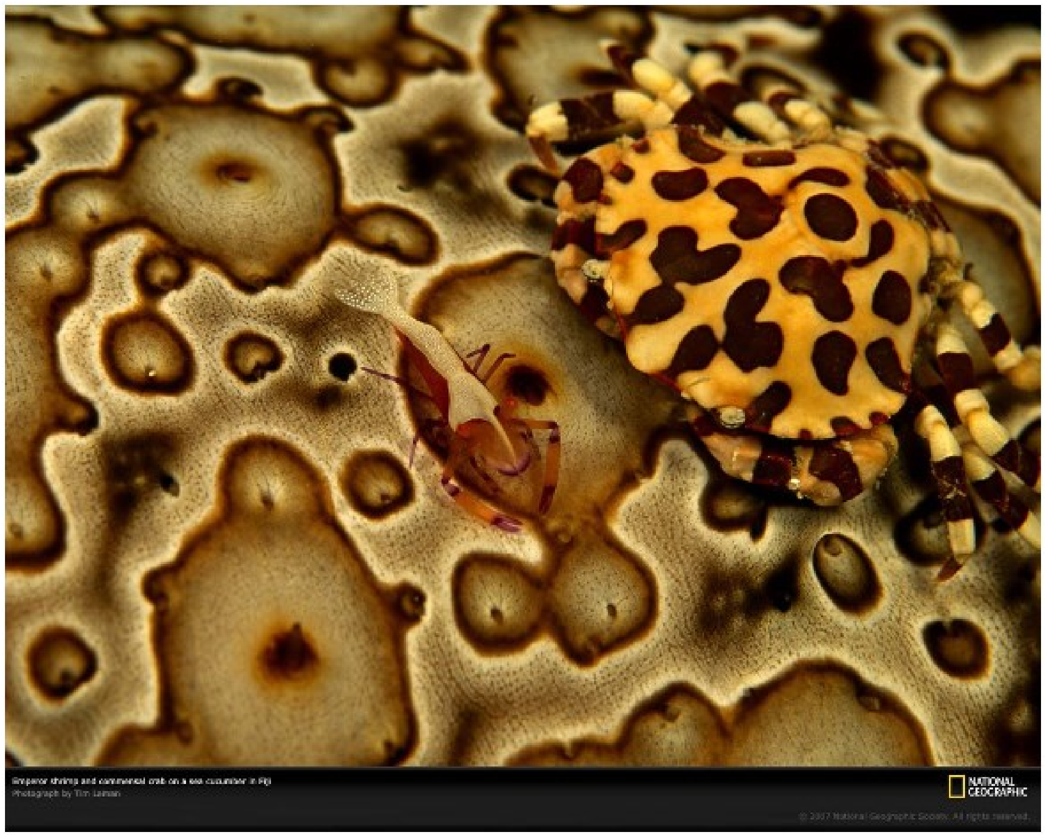
Computer vision challenges

- Occlusion (?!!!)

Magritte, 1957

Computer vision challenges

- Background perturbations



slide credit: Svetlana Lazebnik

Computer vision challenges

- Lighting variations

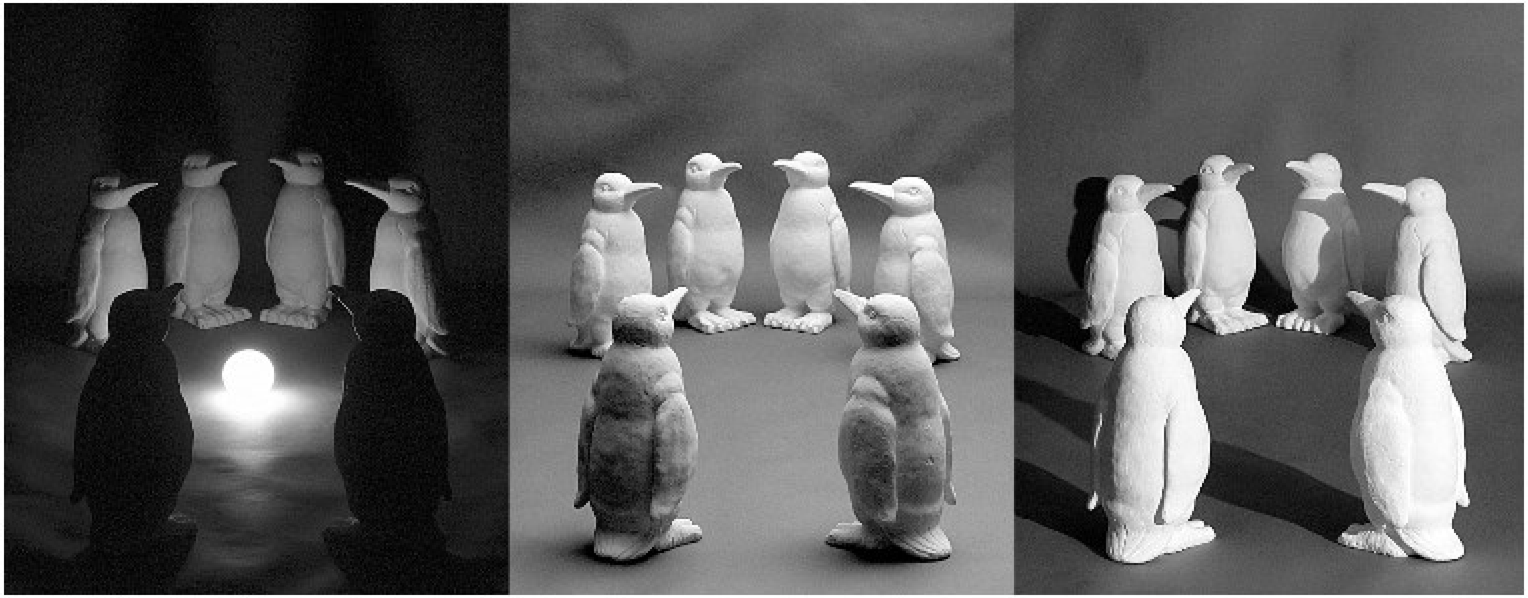
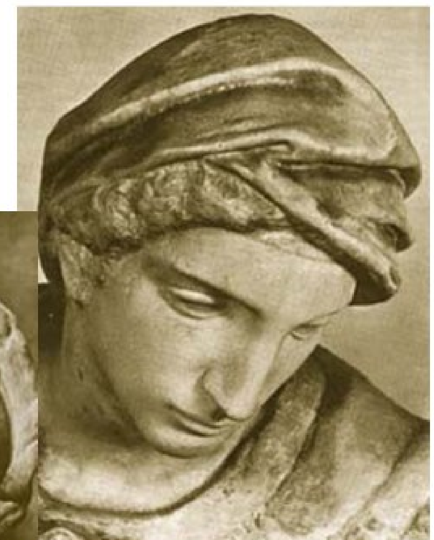


image credit: J. Koenderink

Computer vision challenges

- Viewpoint variations

Michelangelo 1475-1564



slide credit: Fei-Fei, Fergus & Torralba

Computer vision challenges

- Motion

slide credit: Svetlana Lazebnik

Why is vision difficult ?

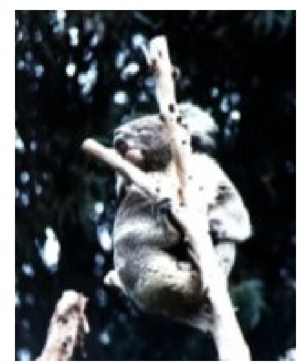
Illumination

Object pose

Clutter

Occlusions

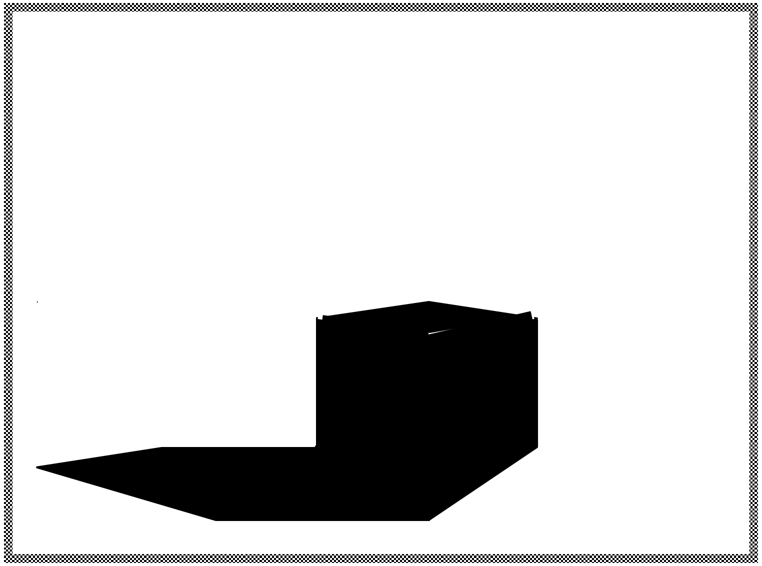
Intra-class appearance

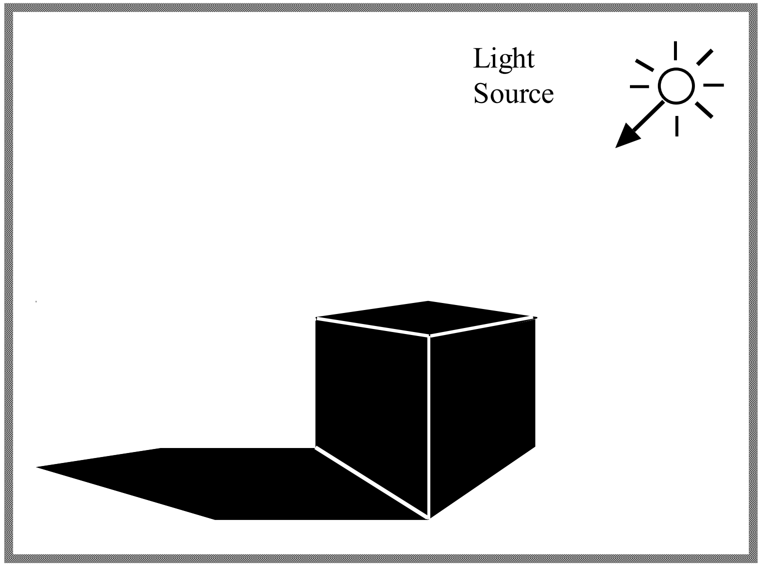


Viewpoint

Why is vision difficult ?

The need of previous knowledge and the importance of the context





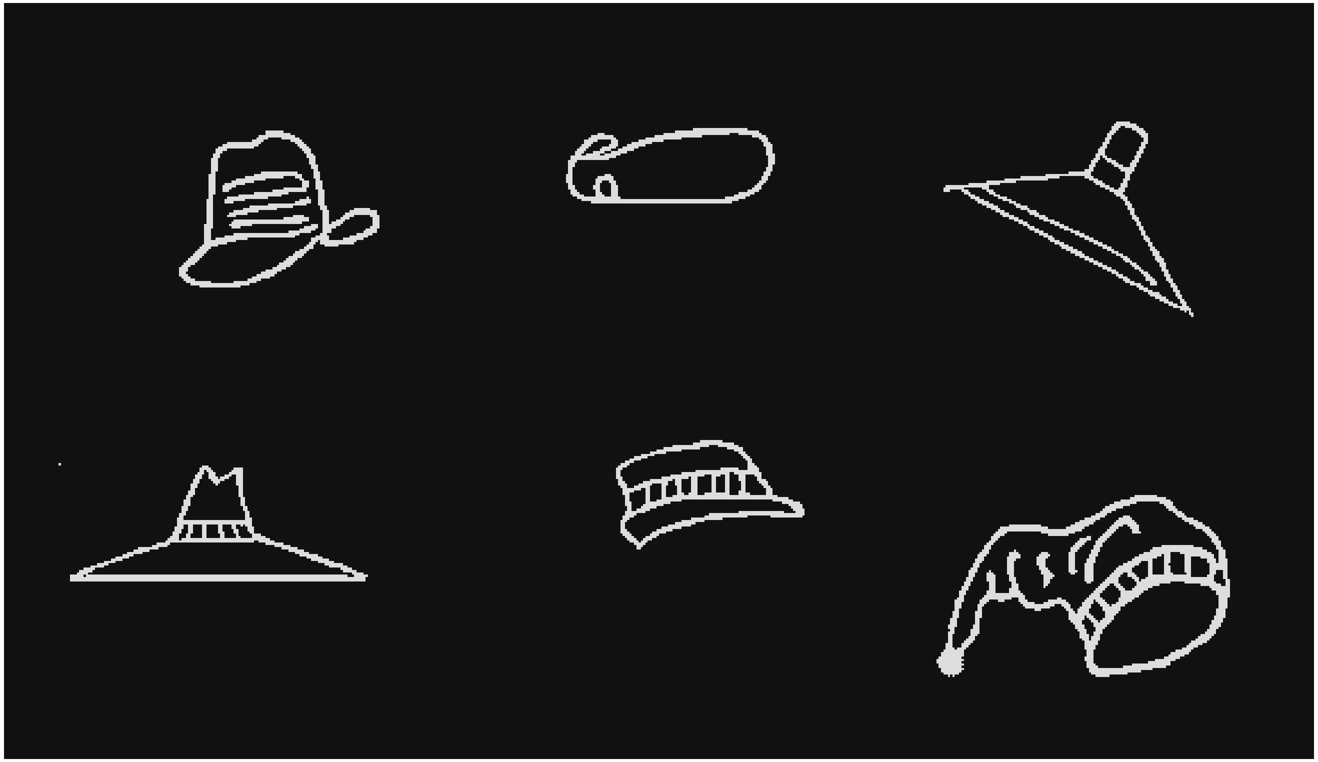
The need of previous knowledge and the importance of the context

THE

CAT

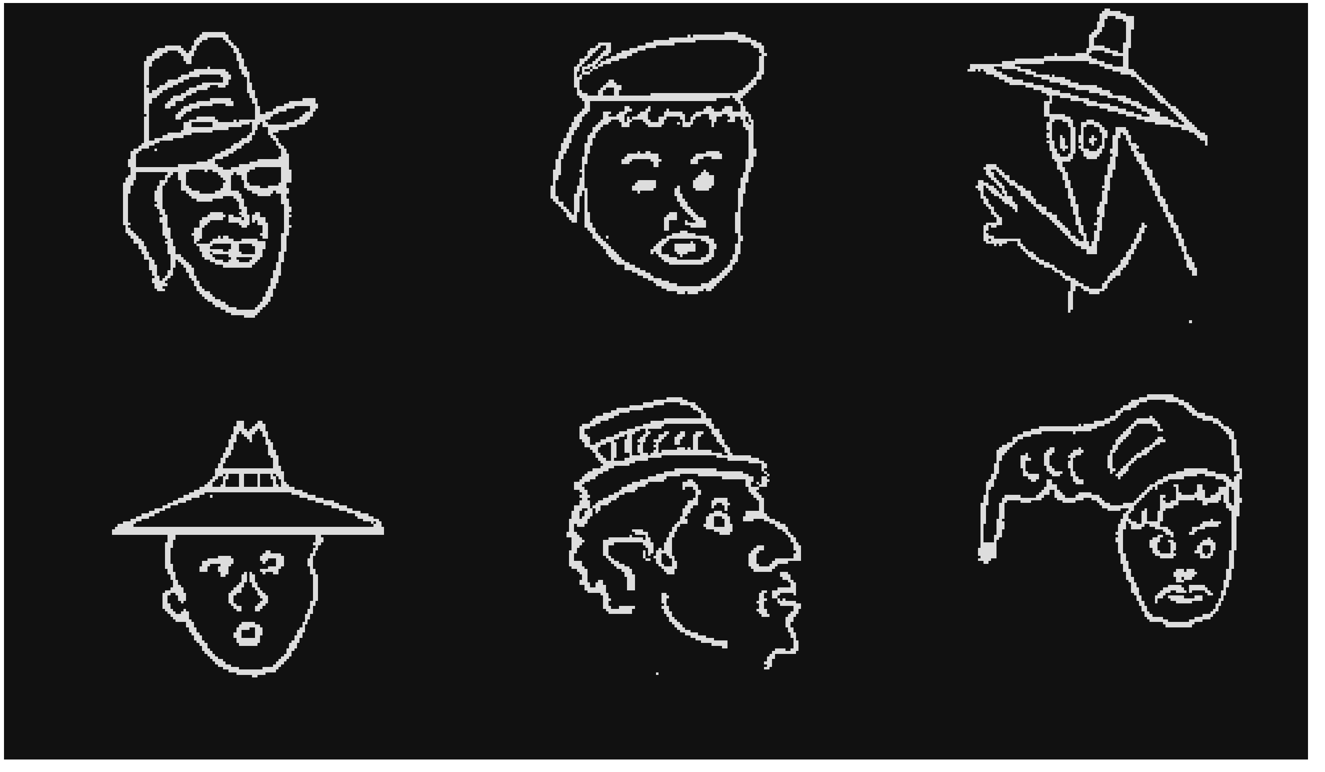
The need of previous knowledge and the importance of the context

- Different "objects" ...



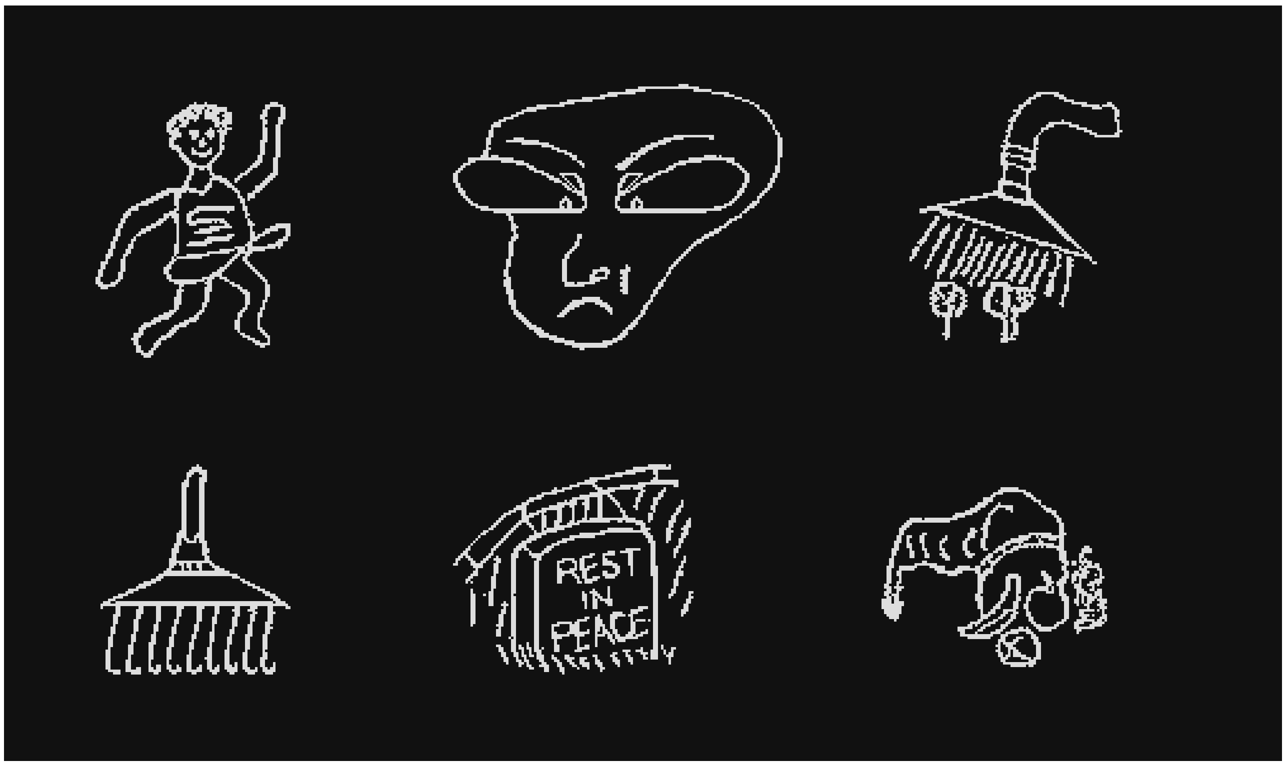
The need of previous knowledge and the importance of the context

- ... can have the same interpretation !!!

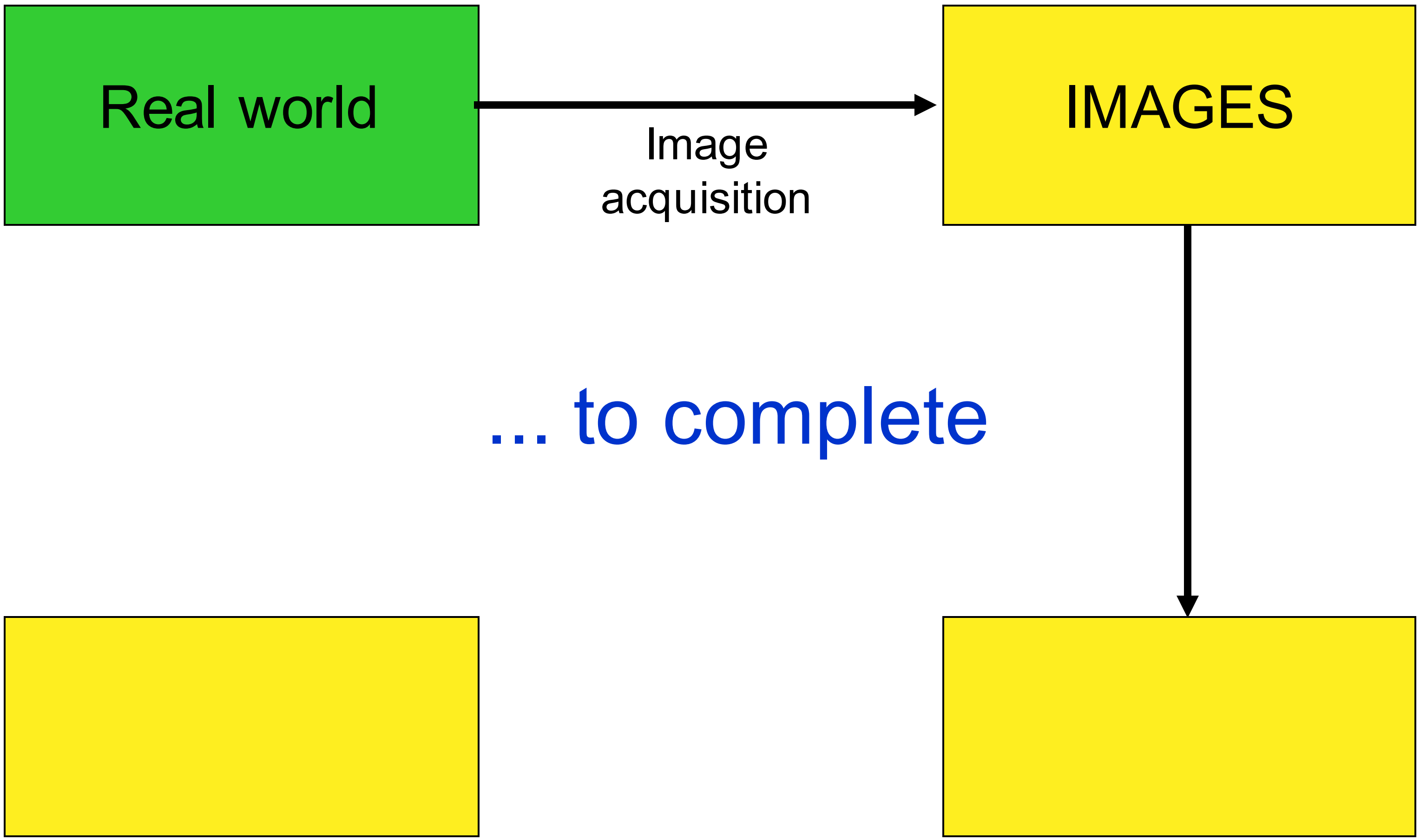


The need of previous knowledge and the importance of the context

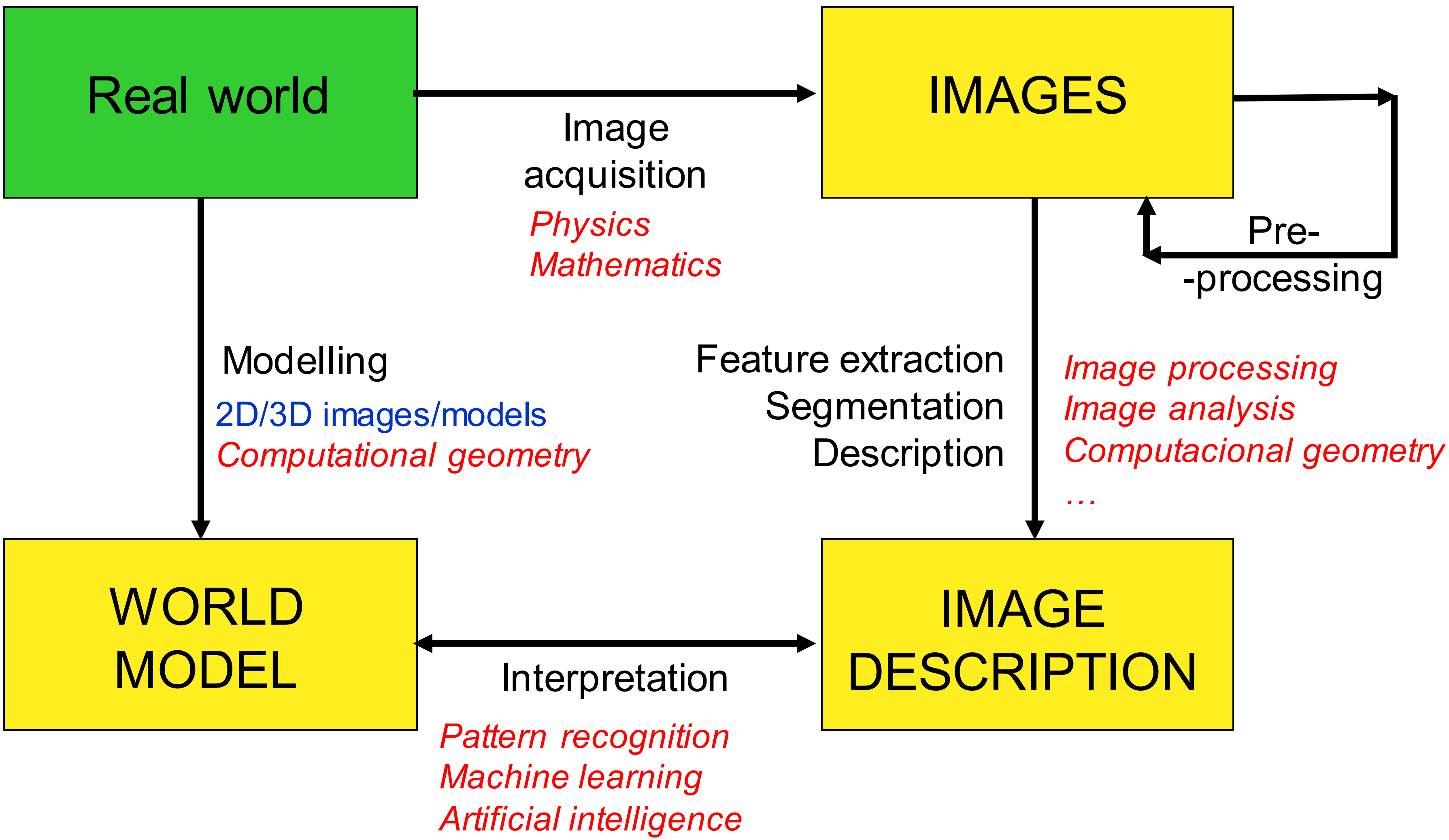
- ... or many different interpretations !



General structure of a classical computer vision system



General structure of a classical computer vision system



General structure of a classical computer vision system

- but … the paradigm is changing
  - Traditional architectures for solving computer vision problems … have been heavily reliant on hand-crafted features.
  - … deep learning techniques have offered an alternative -- -- that of automatically learning problem-specific features.
  - With this new paradigm, some problems in computer vision are now being re-examined from a deep learning perspective.
  - … one form of deep networks widely used in computer vision - convolutional neural networks (CNNs)."

Suraj Srinivas et al., Frontiers in Robotics and AI 2(36), January 2016

General structure of a computer vision system

- One of the big challenges with traditional machine learning models is feature extraction.
  - The programmer needs to tell the computer what kinds of things it should be looking for that will be informative in making a decision.
  - This places a huge burden on the programmer, and the algorithm's effectiveness relies heavily on how insightful the programmer is.
  - For complex problems such as object recognition or handwriting recognition, this is a huge challenge.
- In the last years, some problems in computer vision are being re-examined from a deep learning perspective.
  - Deep learning is a branch of machine learning based on a set of algorithms that attempt to model high level abstractions in data by using a deep graph with multiple processing layers, composed of multiple linear and non-linear transformations.
  - Deep learning aims to overcome those challenges (of feature extraction) by automatically learning hierarchies of visual features in both unsupervised and supervised manners directly from data. This is because deep learning models are capable of learning to focus on the right features by themselves, requiring little guidance from the programmer.
  - However, lots of data and computing power are needed.
  - A main criticism of deep learning concerns the lack of theory surrounding many of the methods.

Scientific areas related to Computer Vision

- COMPUTER VISION
- Image processing and analysis
- Pattern recognition
- Machine learning
- Artificial intelligence
- Photogrammetry
- Computer graphics
- Computational geometry
- Robotics
- Applied mathematics
- Physics
- Neuroscience
- Cognitive science
- …

Scientific areas related to Computer Vision

- Image processing
  - 2D low level image transformation and processing.
  - Used in low level vision for enhancing and extracting features such as points, lines, contours, and regions.
- Computer graphics
  - Synthesize images using geometric primitives, physical properties of objects, and illumination conditions.
  - It is the inverse of computer vision. Vision-based graphics is becoming popular.
- Photogrammetry
  - Study the geometric relationship between 3D scenes and their 2D projections to obtain accurate measurements from noncontact imaging.
- Pattern recognition and machine learning
  - Statistical and syntactical techniques for classifying patterns.
  - The techniques are widely used in computer vision, especially in object detection and recognition.
- Artificial Intelligence
  - Computational intelligence that includes perception, cognition, and action
  - Computer vision can be viewed as a subfield of AI

Computer Vision related scientific areas

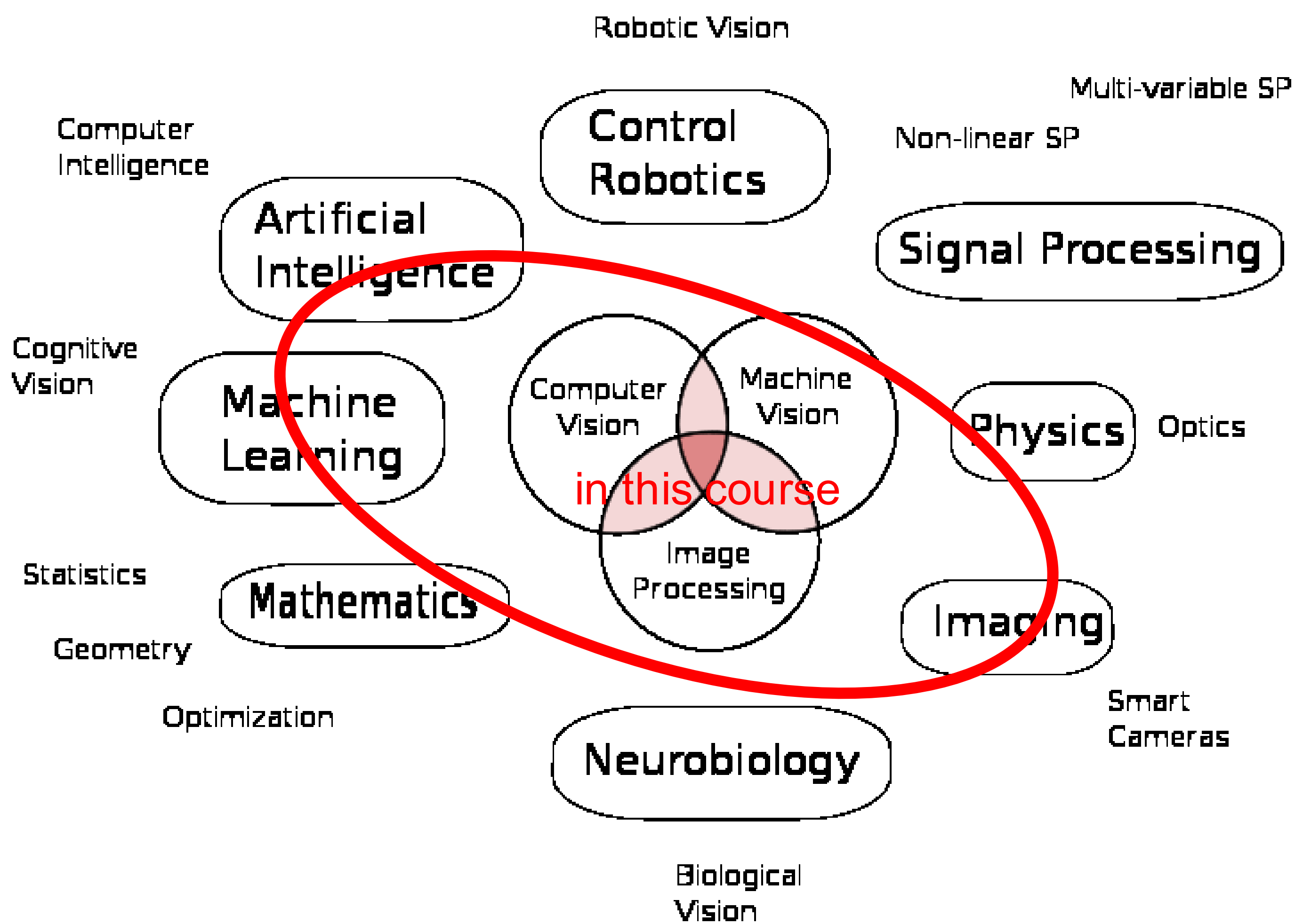


Image processing and analysis: some tasks

- "Image → Image" transformations
- "Image → Description" transformations

- Image (pre)processing (essentially 2D)
  - Image restoration
    - The operation of taking a corrupt/noisy image and estimating the clean, original image.
    - Corruption may come in many forms such as motion blur, noise and camera mis-focus.
    - Image restoration is performed by reversing the process that corrupted the image
  - Image enhancement
    - is the process of adjusting digital images so that the results are more suitable for display or further image analysis.
    - remove noise, sharpen, or brighten an image, making it easier to identify key features.
- Image analysis
  - is the extraction of meaningful information from images; mainly from digital images by means of digital image processing technique
  - Segmentation and Feature Detection
    - split the image into its elements; a feature is an "interesting" part of an image (no exact definition)
  - Description (ex: size of a tumor, orientation of an object, distance between objects, …)

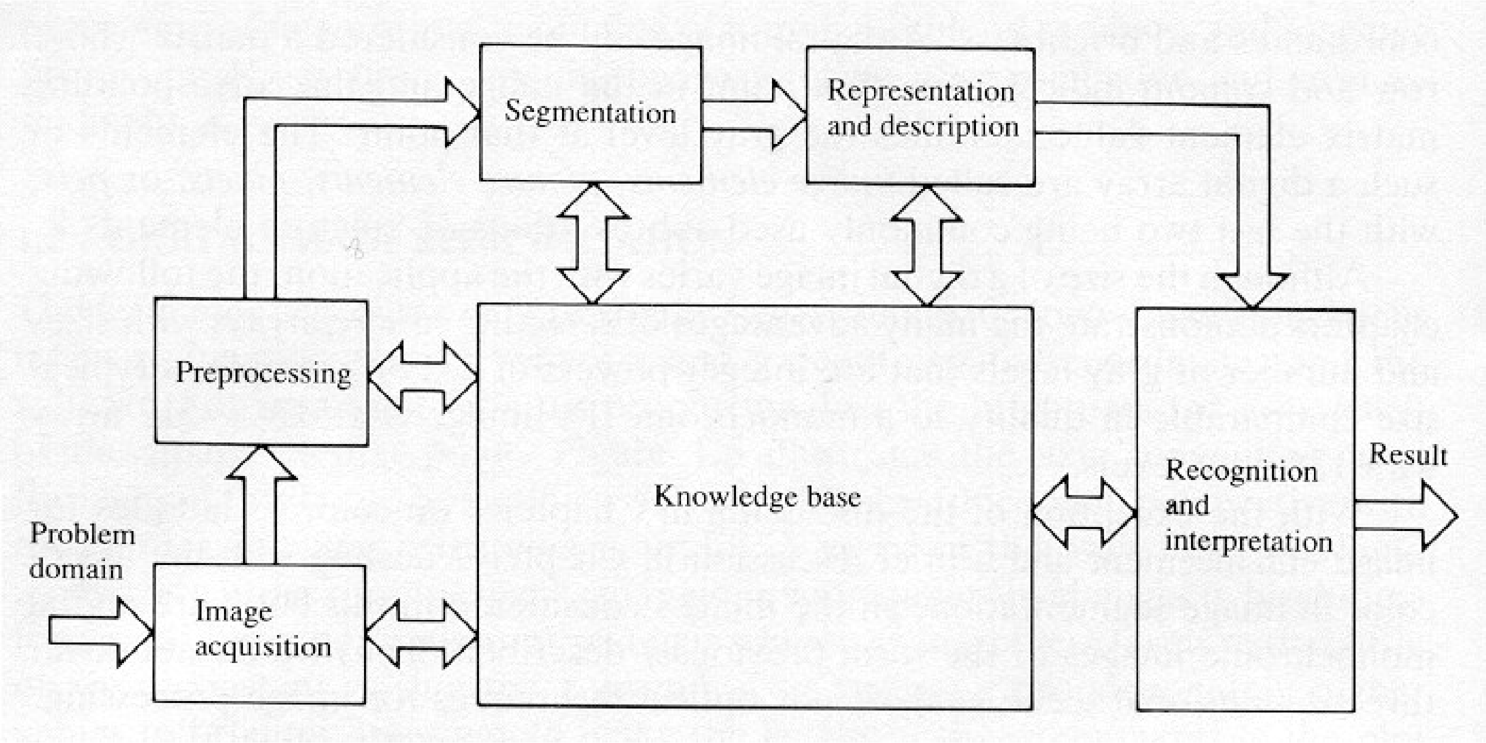
Image Processing *versus* Computer Vision

- Some **Image Processing** tasks:
  - Reduce image noise.
  - Enhance contrast.
  - Detect feature points / edges.
  - Segment images.
- Some **Computer Vision** tasks:
  - Classify objects.
  - Detect moving objects.
  - Build object models from images.
  - Recognize faces.
  - Interpret scenes.
  - …

Classical computer vision processes

- Image acquisition
- Preprocessing
- Processing
- Analysis
- Recognition
- Interpretation

Classical computer vision processes



Example: visual inspection

- **Parts of an assembly line:**
  - Acquire image
  - Preprocess it
    to reduce noise, enhance contrast …
    - filtering, histogram equalization, …
  - Which *pixels* constitute the part ?
    - segmentation
  - Which is the
    position and orientation of the part ?
    - description, representation
  - Which part is it ?
    - feature extraction (morphologic, …), classification (type)
  - Is the part in good condition or not ?
    - matching, shape analysis, classification (quality)

PROGRAM

(samples)

Modeling the signal path

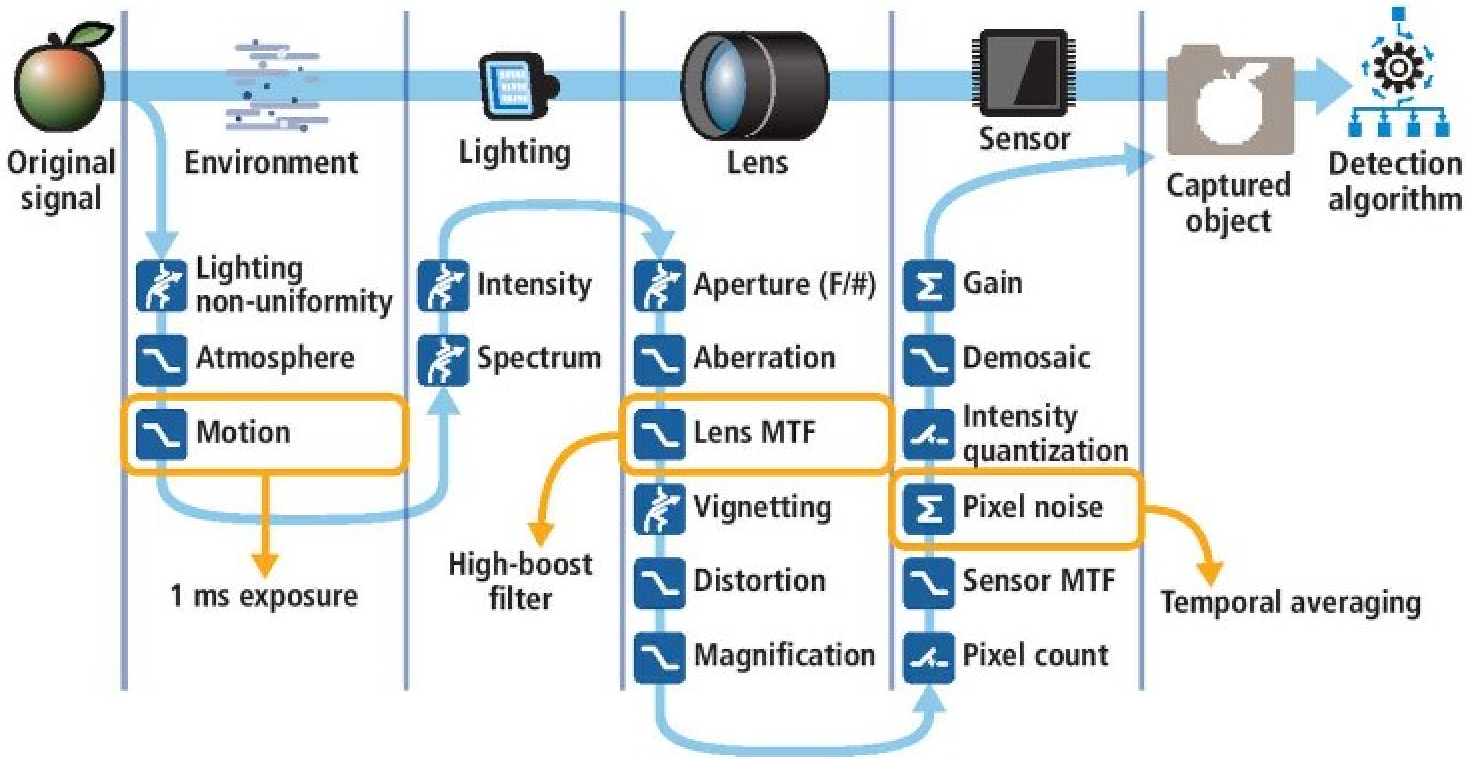
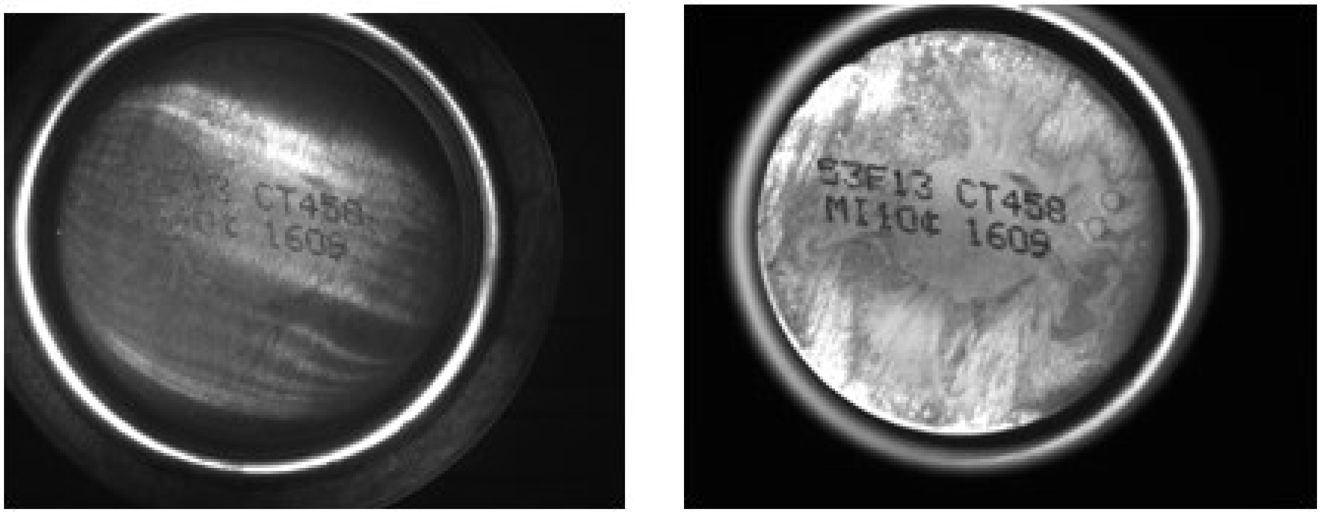
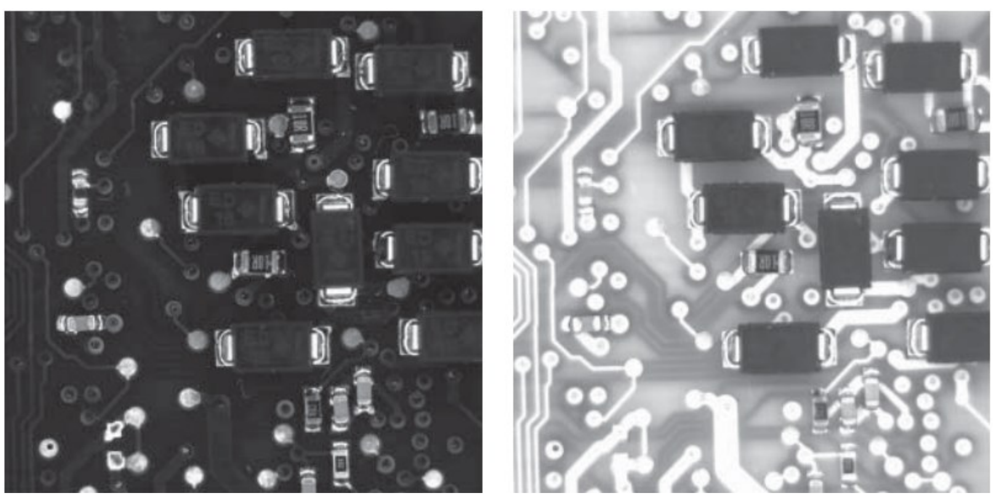


Image acquisition: Illumination



On the <u>right</u>, the soda can is imaged with <u>diffuse light</u>



On the <u>left</u>: circuit board with dark green solder resist, illuminated in red
On the <u>right</u>: same board illuminated with <u>infrared (IR) light</u>; The IR light passes the solder resist almost without loss and is reflected on the surface of the supporting material of the circuit board.

Image acquisition: Filters (polarizing)

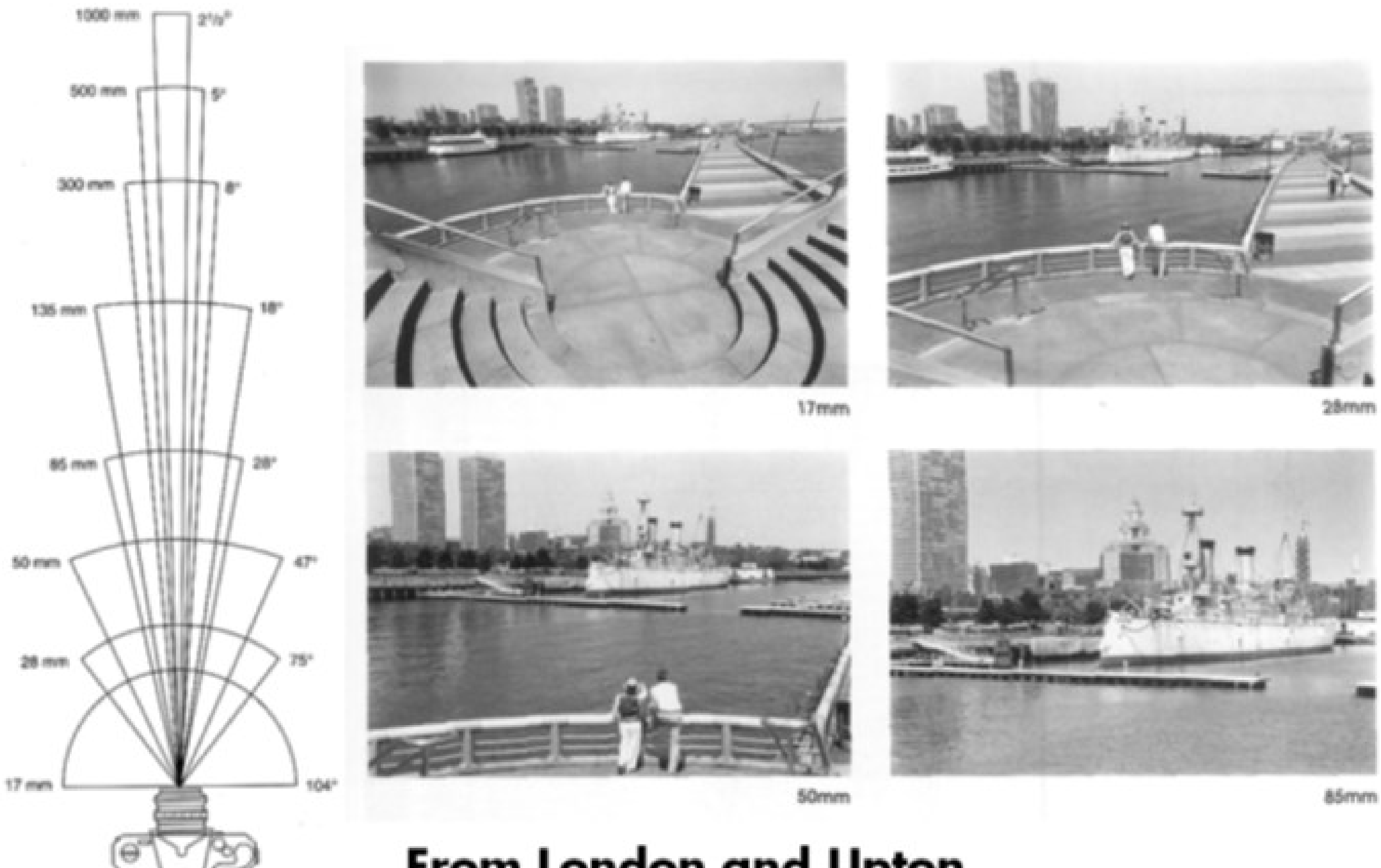
Removing specularities

Separating reflected and transmitted scenes

Removing haze and underwater scattering effects

Image acquisition: Lenses

- Field of view (zoom, focal length)

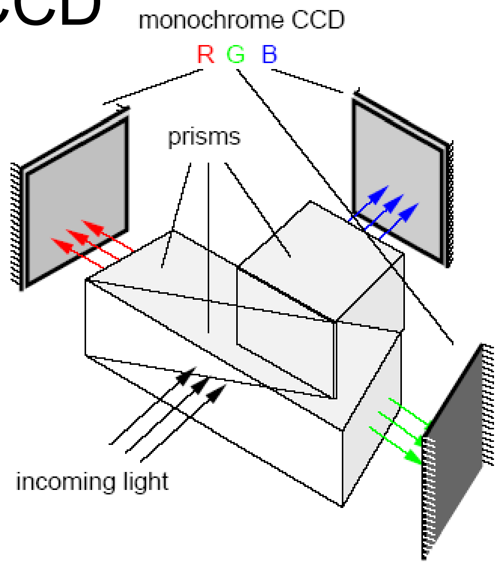


From London and Upton

Image acquisition: Sensors

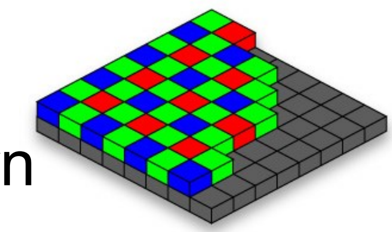
- Sensing color

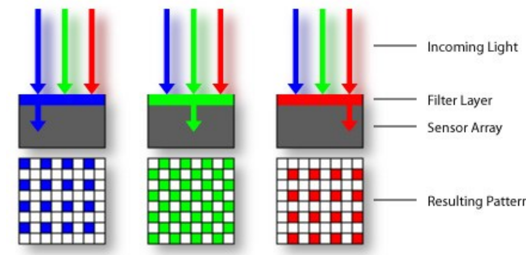
3 CCD

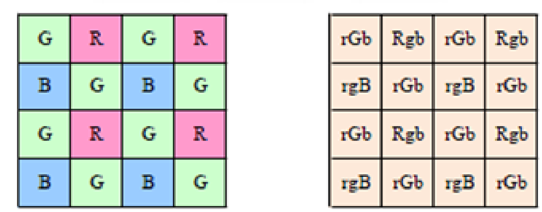


The incoming light is divided by prisms into its basic components: R,G, B

Bayer pattern







(a) (b)

a) Filter array layout; b) interpolated pixel values, in lowercase

Foveon X3™

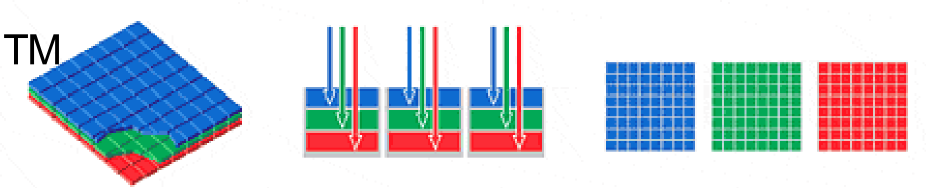
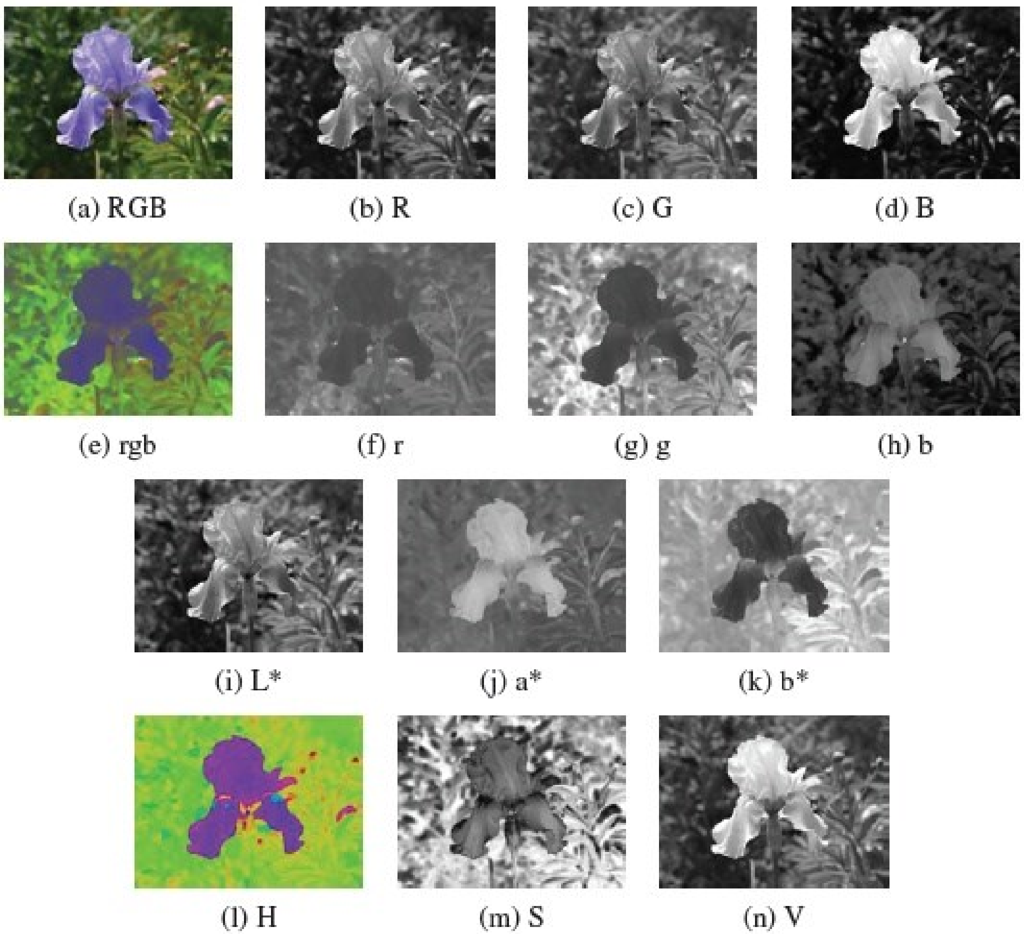


Image representation: Color spaces



$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B}$$

source: Szeliski's book

Figure 2.32 Color space transformations: (a–d) RGB; (e–h) rgb. (i–k) L*a*b*; (l–n) HSV. Note that the rgb, L*a*b*, and HSV values are all re-scaled to fit the dynamic range of the printed page.

Image preprocessing: examples of operations

- Noise reduction / smoothing

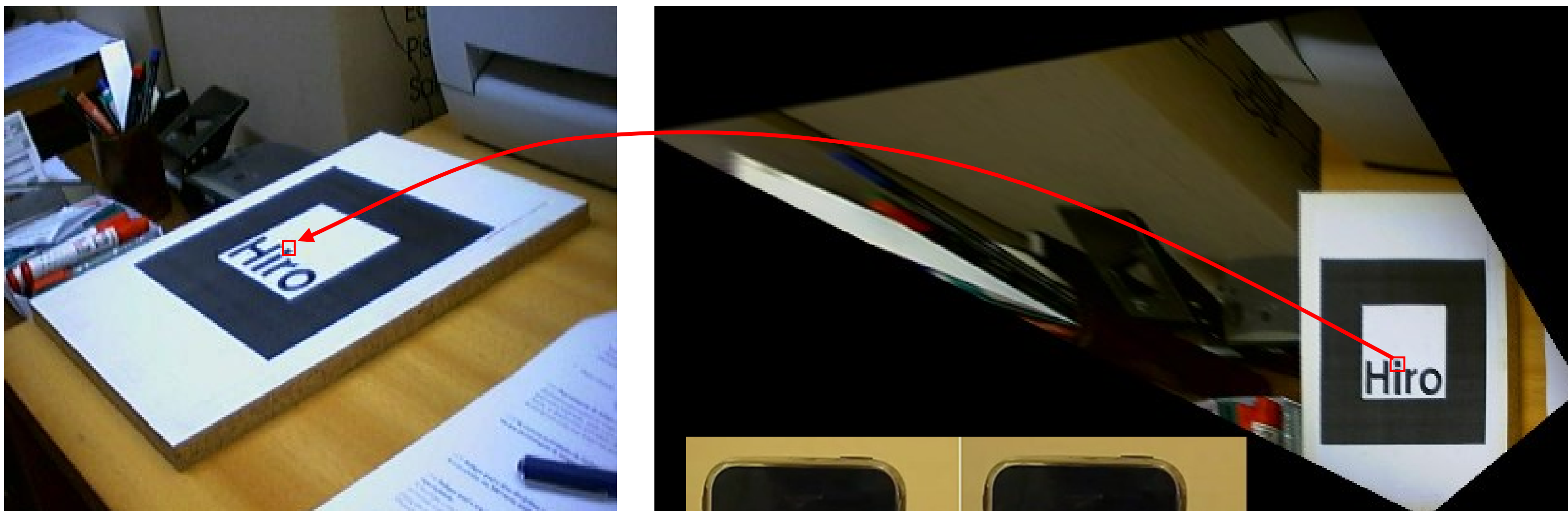
Image preprocessing: examples of operations

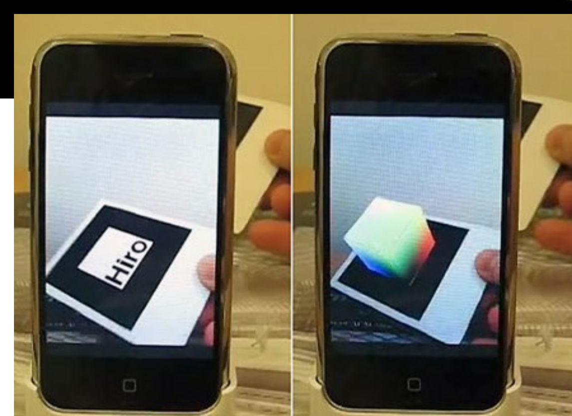
- Image enhancement

histogram equalization

Image preprocessing: examples of operations

- Geometric transformations

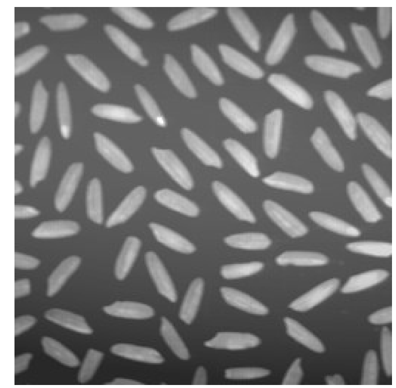


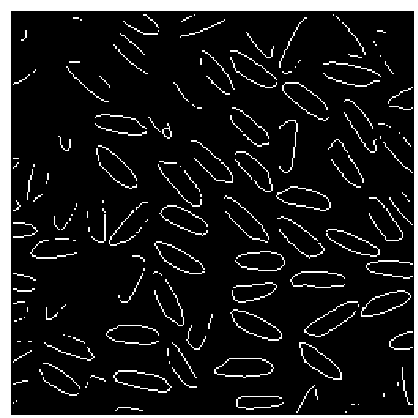


Augmented reality application

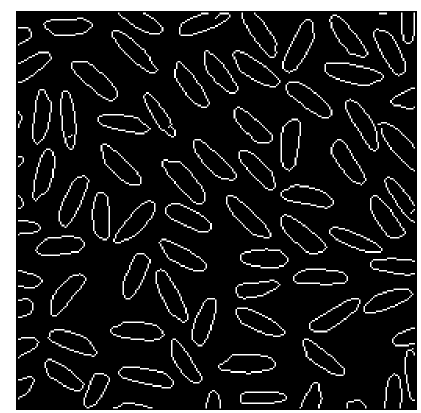
Image processing: examples of operations

- Edge detection





Sobel Filter



Canny Filter

Image processing: examples of operations

- Line detection - Hough transform

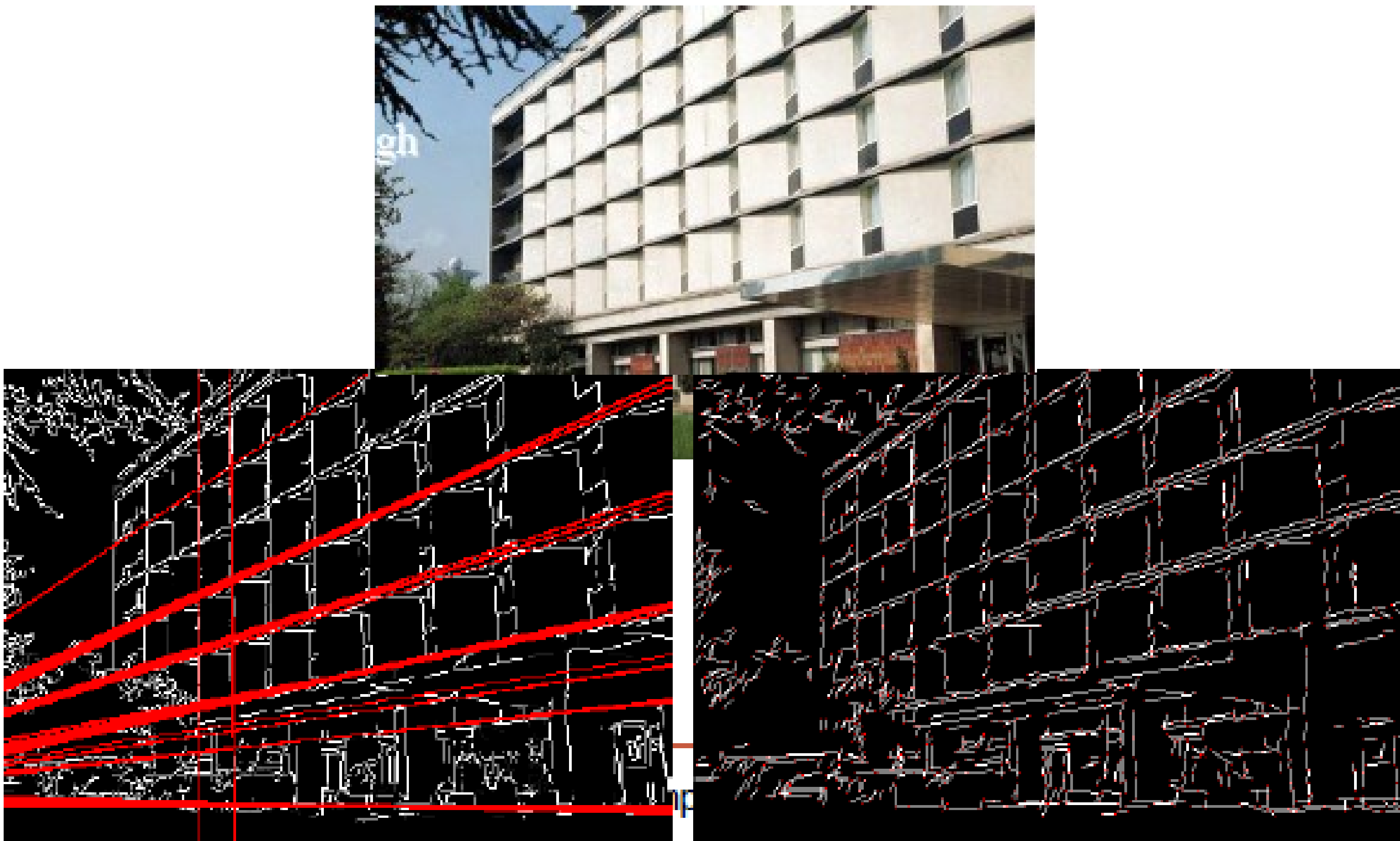
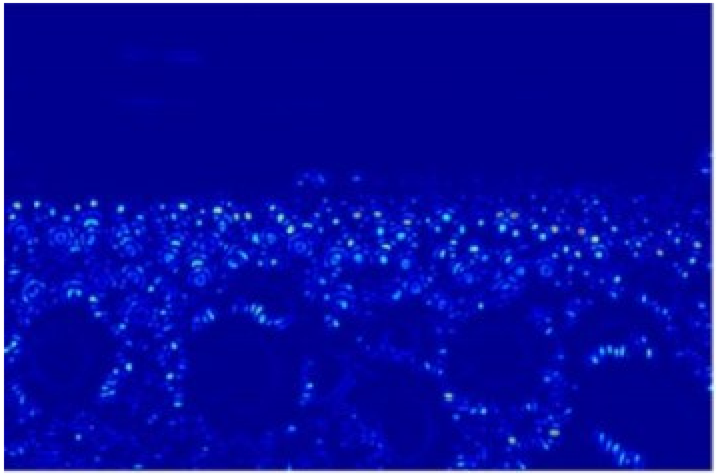


Image processing: examples of operations

- Corner detection

Image processing: examples of operations

- Scale invariant feature detection



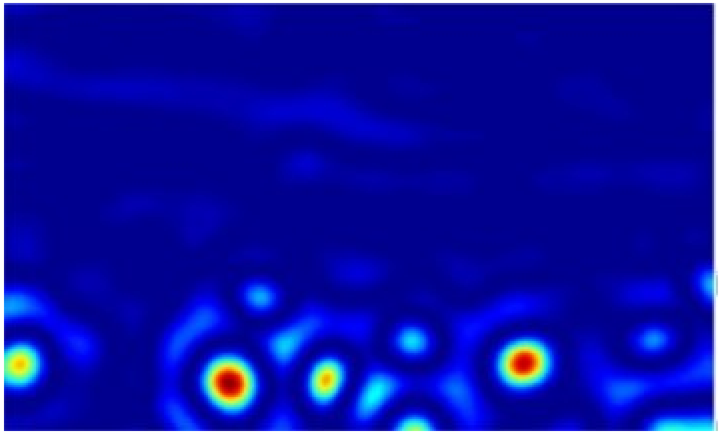


Image processing: examples of operations

- Segmentation



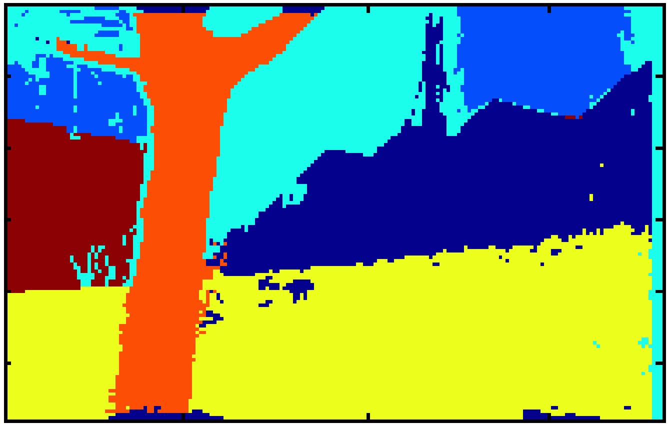
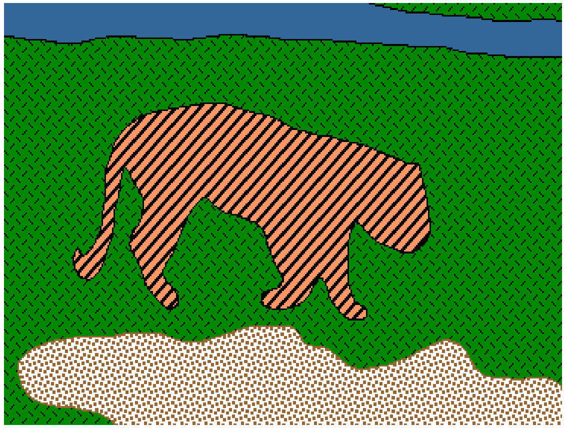



Image processing: examples of operations

- Segmentation

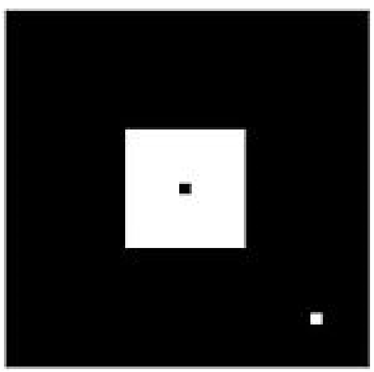
(a) (b) (c)

(a) the user draws a bounding box in red;
(b) the algorithm guesses color distributions for the object and background and performs a binary segmentation;
(c) the process is repeated with better region statistics.

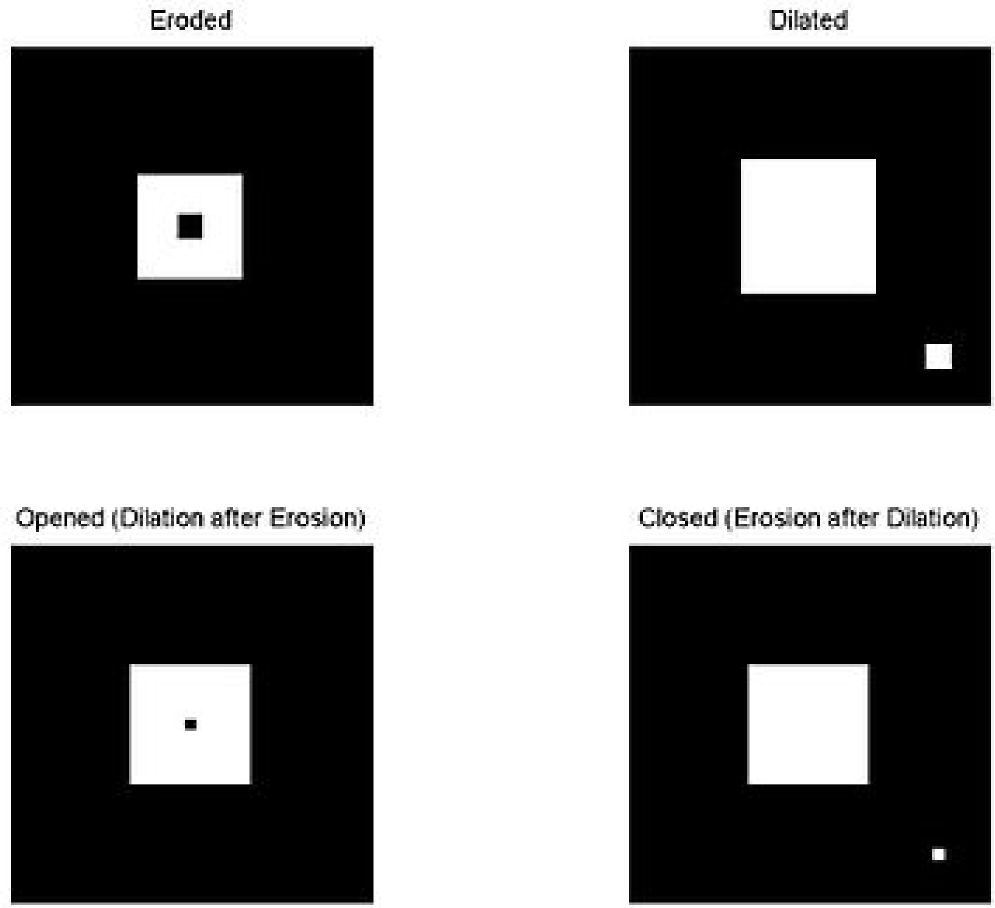
source: Szeliski book

Image processing: examples of operations

- Morphologic operations

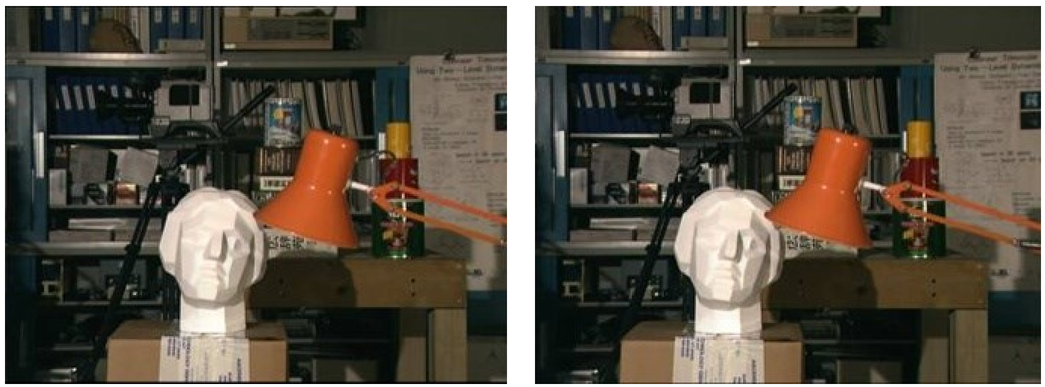


original image

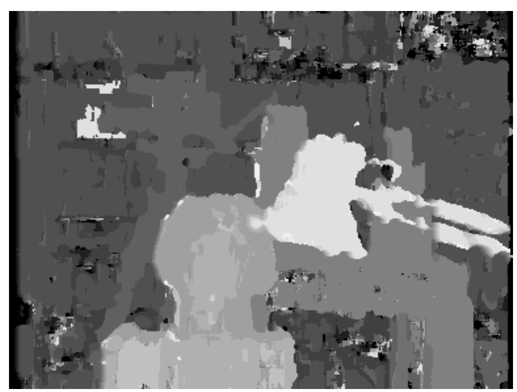


3D data acquisition - stereo

- Stereoscopic pair & disparity map



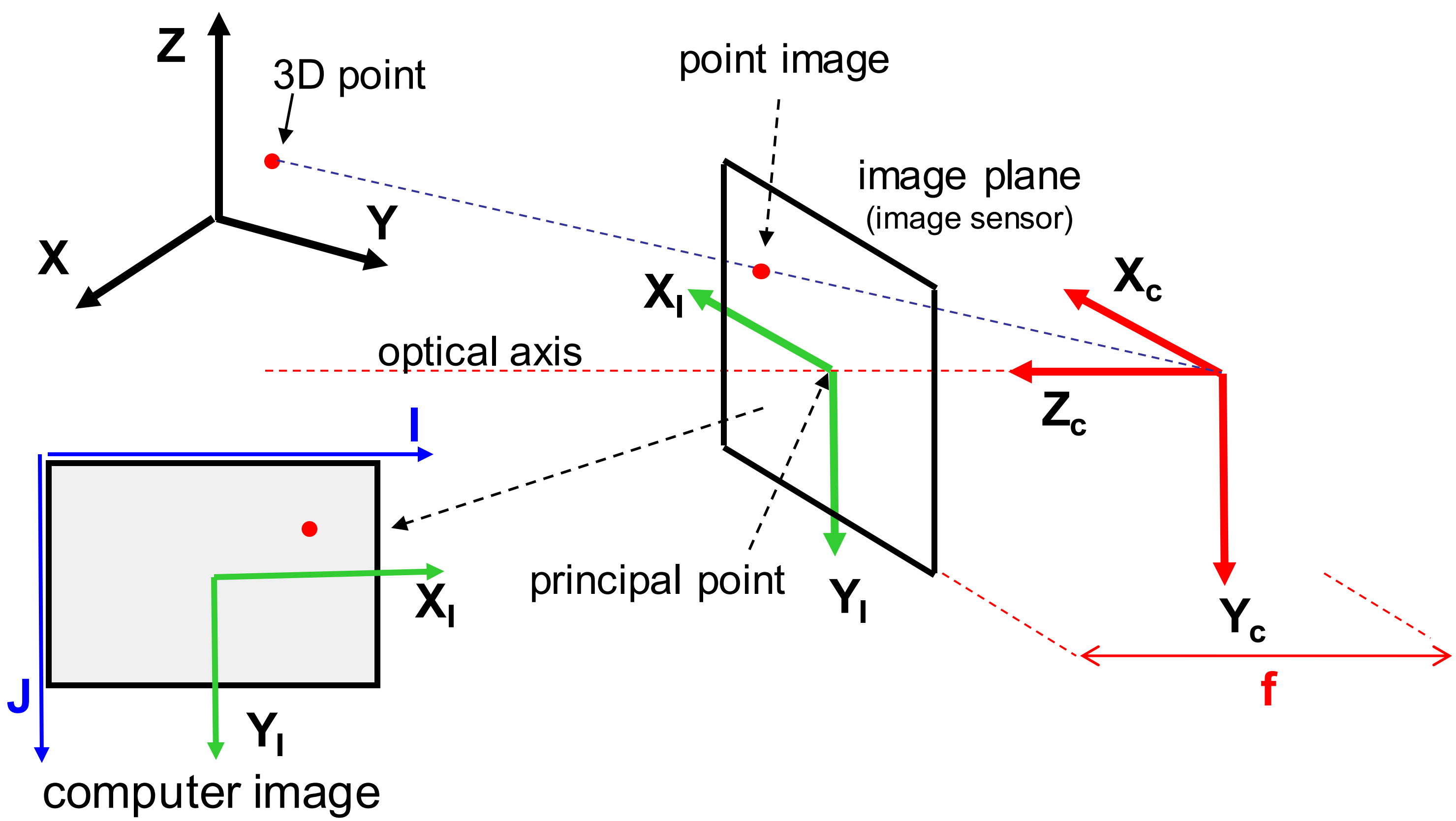
Stereoscopic pair(Tsukuba)



Each point
represents the disparity
(difference in position in the images)
between corresponding points
in the 2 images of the stereoscopic pair

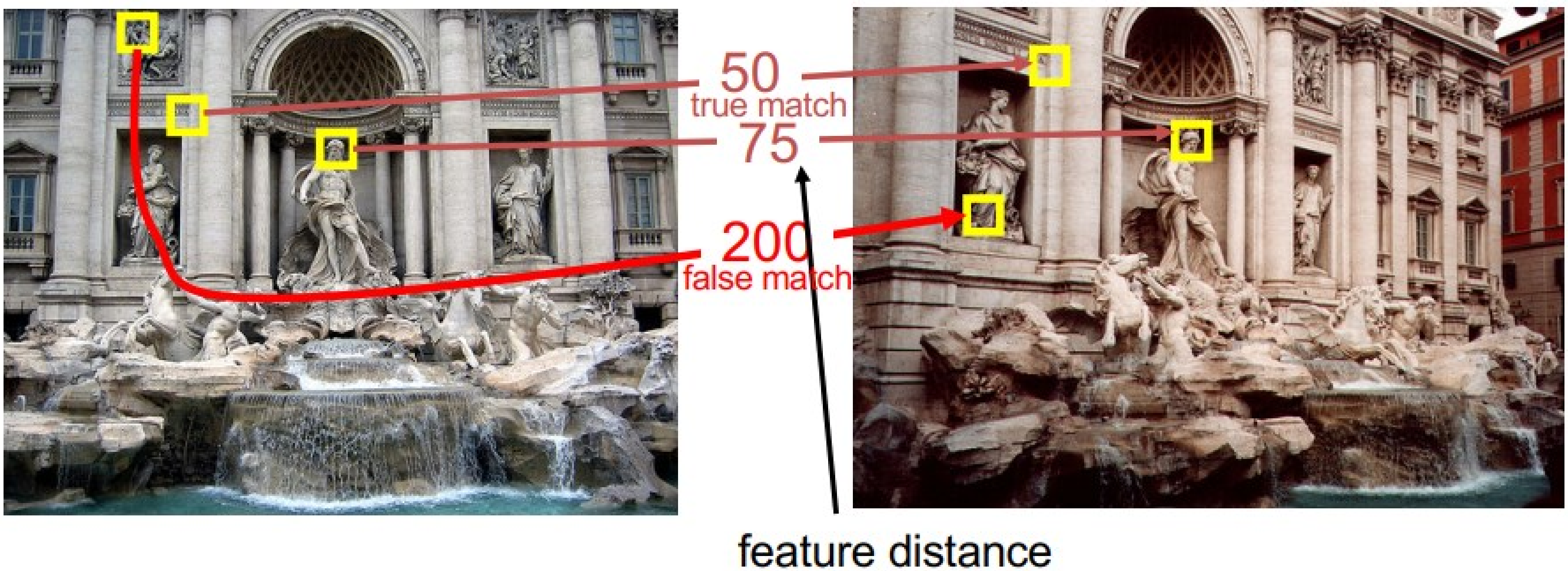
Disparity map

3D data acquisition: Camera model



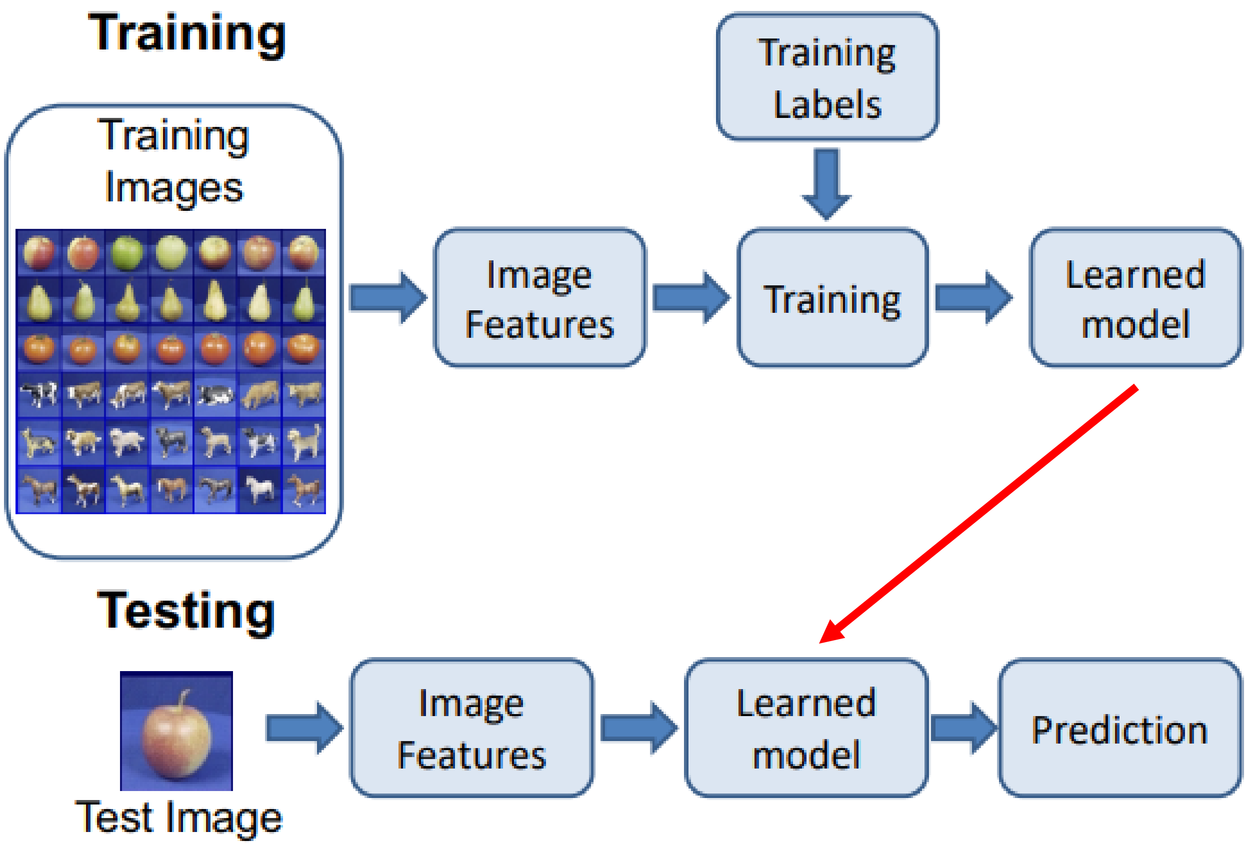
Recognition

- Feature matching



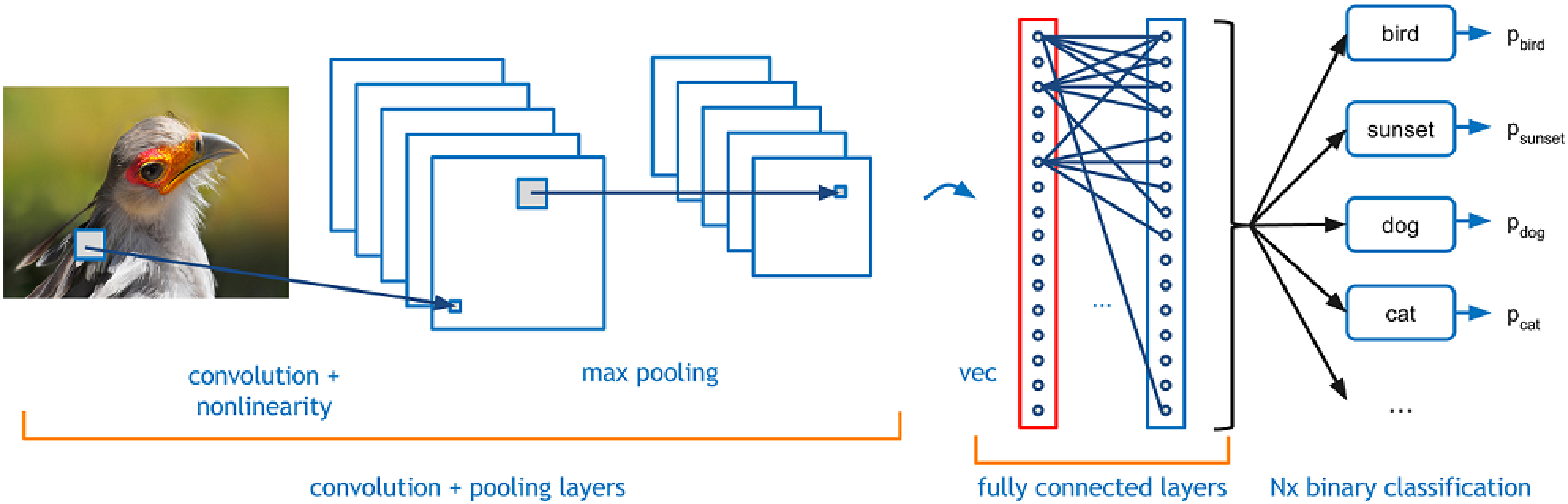
Machine learning

- Classification



Convolutional Neural Networks

- CNN arquitecture

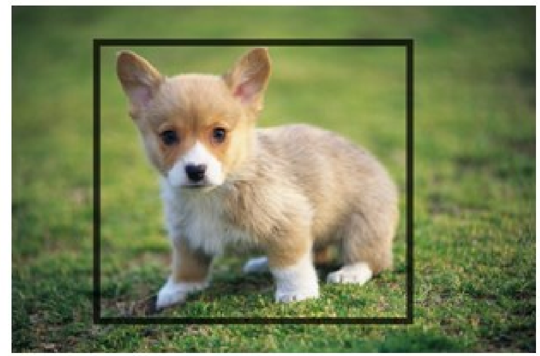


https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/

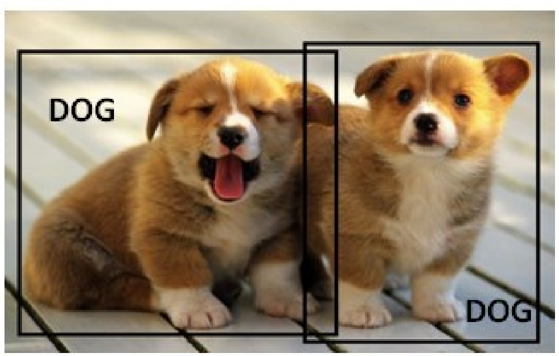
Convolutional Neural Networks

- Localization, detection, segmentation, classification

Object Classification is the task of identifying that picture is a dog



Object Localization involves the class label as well as a bounding box to show where the object is located.



Object Detection involves localization of multiple objects (doesn't have to be the same class).

Object Segmentation involves the class label as well as an outline of the object in interest.

https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks-Part-2/

Motion

- Optical flow