COMPUTER VISION

MEI/1

University of Beira Interior, Department of Informatics

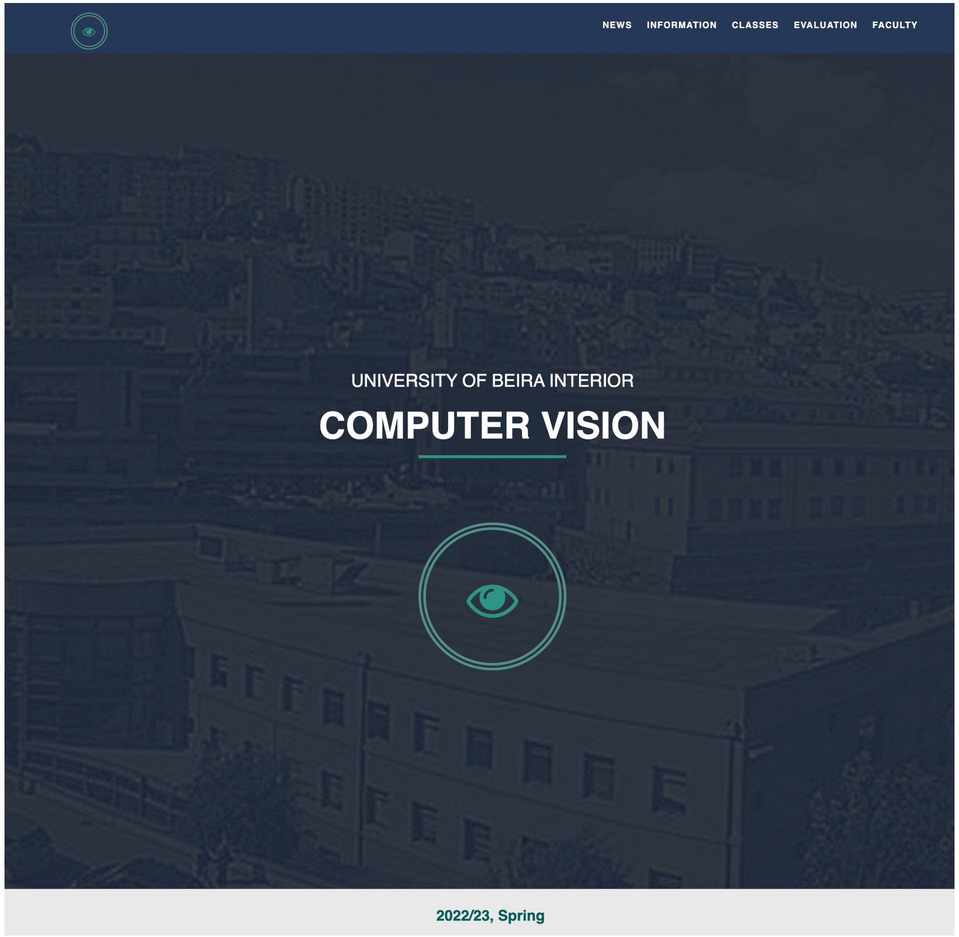
Hugo Pedro Proença

hugomcp@di.ubi.pt, **2024/25**

Computer Vision – Course Main Page

URL: http://di.ubi.pt/~hugomcp/cv/

- □ News
- □ Course Program, Evaluation Criteria, Bibliography
- □ Classes (Theoretical Slides+ Practical Sheets + Exercises)
- □ Evaluation



UNIVERSITY OF BEIRA INTERIOR

COMPUTER VISION

2022/23, Spring

Computer Vision – Evaluation Criteria

- - **Assiduity (A)** To get approved at this course, students should attend to - at least - 80% of the theoretical and practical classes;
- - **Practical Project (P)** The practical projects of this course weights 50% (10/20) of the final mark.
- - To get approved at the course, a minimal mark of 5/20 should be obtained in the practical project part;
- - The pratical project mark is conditioned to an individual presentation and discussion by each student;
- - **Written Test (F)** Monday, June 3rd, 2024, 14:00. Room 6.20
- - **Mark (M)** M = (A >= 0.8) * (P * 10/20 + F * 10/20)
- - **Admission to Exams** Students with M >= 6 are admitted to final exams

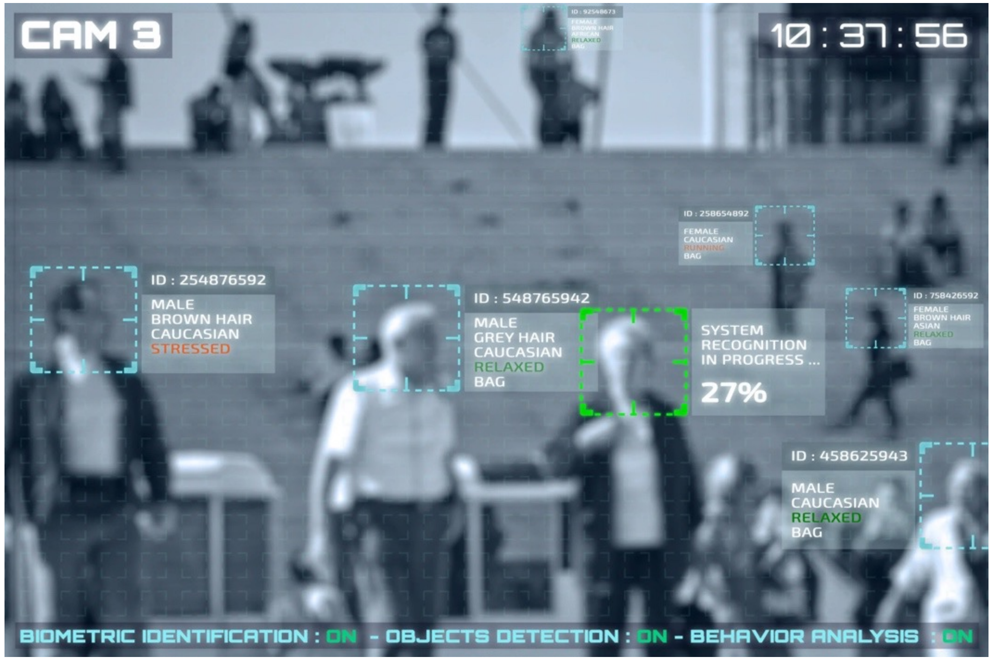
Computer Vision – Practical Project

The first part of the work aims at **collecting and annotating a novel dataset** for research on Human Recognition in Surveillance settings. Students should constitute groups of two elements, and each group will be responsible for collecting synchronized data from (at least) 25 volunteers/participants.

For each volunteer/participant, two synchronized pairs of videos (of at least 1 min) should be collected, simulating:

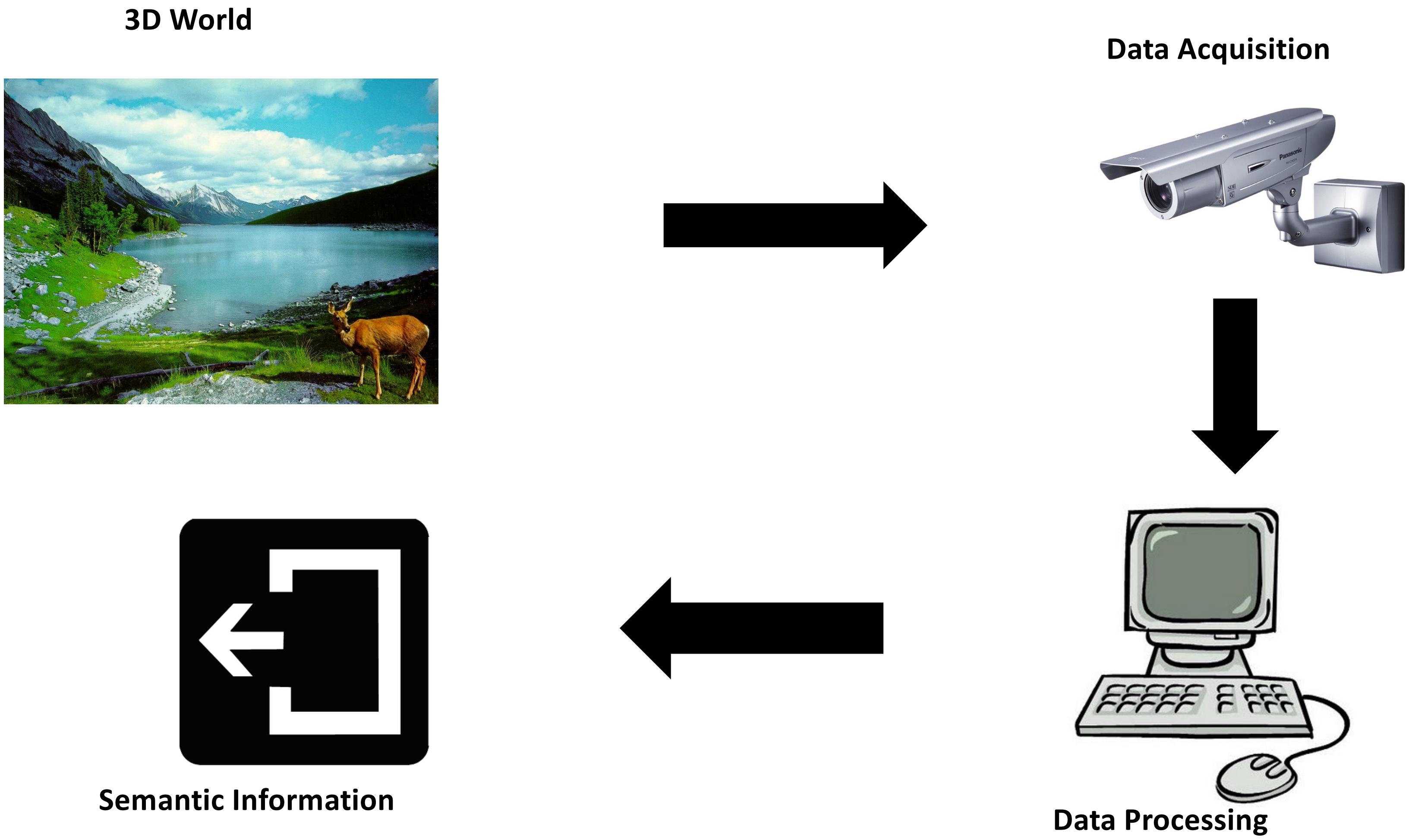
1) the acquisition in totally constrained/controlled scenarios;

2) in surveillance scenarios, with substantial variations in pose, lighting, scale, perspective and heavy degradations in data quality (.e.g., resolution, blur,…).



Computer Vision: Sample Problem

Let's consider a simple CV problem: “*Identify the species of the animal(s) in the scene*”. This is an obvious **image classification problem**.



3D World

Data Acquisition

Data Processing

Semantic Information

Computer Vision: Preprocessing

This is the first phase of the pipeline and typically involves: a) to normalize the information with respect to the expected **data variation factors**; b) **reduce the amount of information** available.

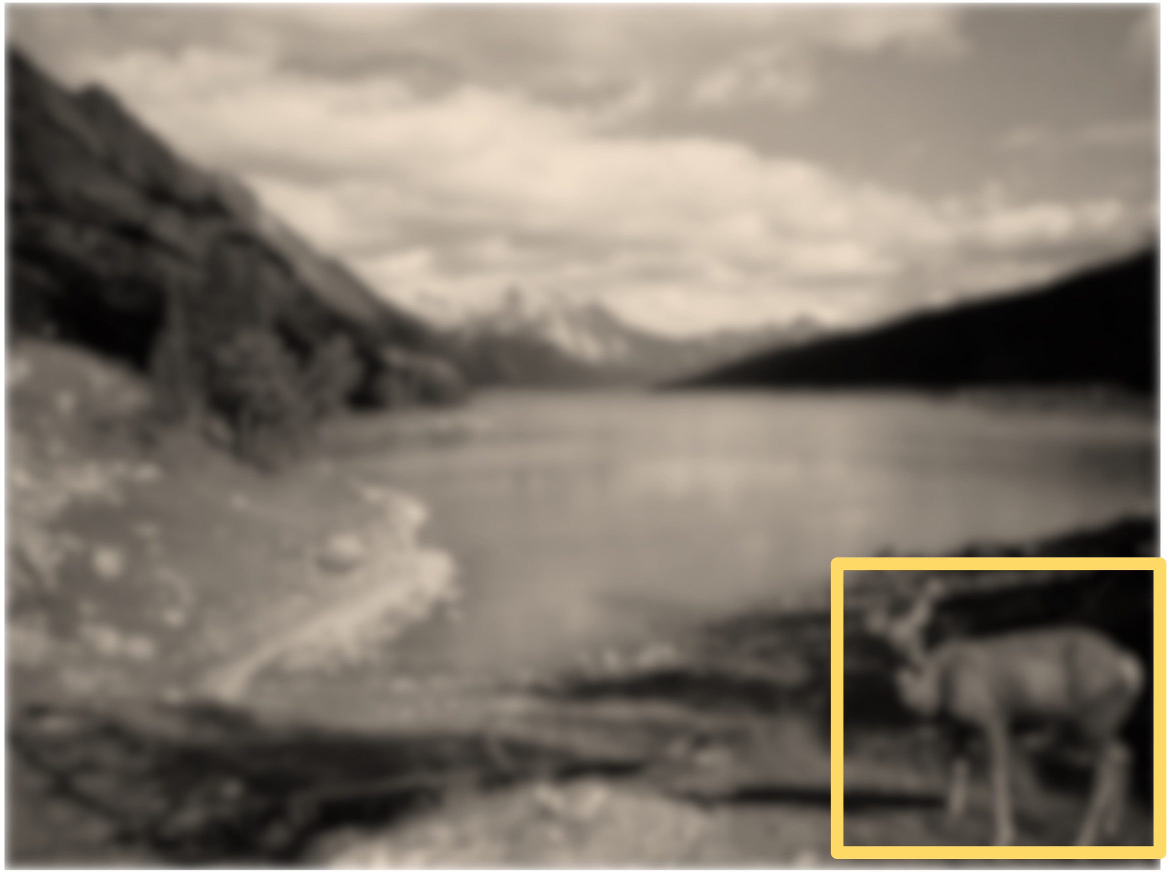
CV is **all about obtaining invariances** with respect to the most factors possible (e.g., scale, translation, rotation, pose, occlusion,…)

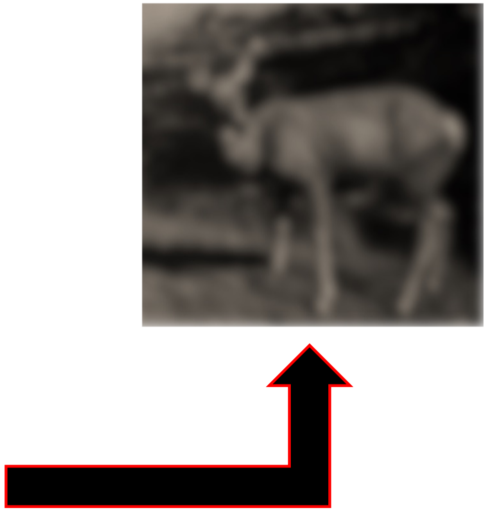
Computer Vision: Detection

This is the phase where a **ROI : Region of Interest** should be obtained. This phase is about discarding a substantial partn of the input, considering it irrelevant for the purposes of our problem.

This phase is *per se* often regarded as a full CV problem, involving feature extraction, classification.

- There are lots of M.Sc./Ph.D. works only about detection;
- There are hundreds of CV scientists exclusively concerned about this problem;

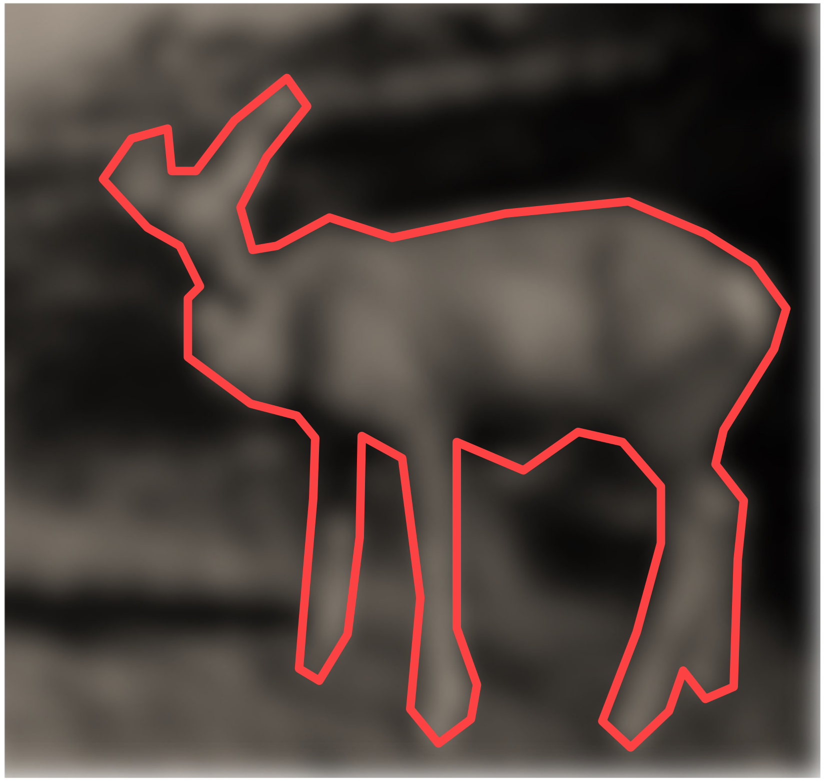




Computer Vision: Segmentation

"Segmenting" refers to obtaining **full parameterizations ($\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\theta_3}, \boldsymbol{\theta_4}, .. \boldsymbol{\theta_n}$)** of the object boundaries, from where all its features will be extracted. At the bottom level, it can be seen as a system that a) receives a ROI; and b) returns a **parameterization vector**.

Parameterizing a complex shape is particularly hard from the computational cost perspective (e.g., a circle has 3 parameters (x,y,r); an ellipse has 5 parameters,…)

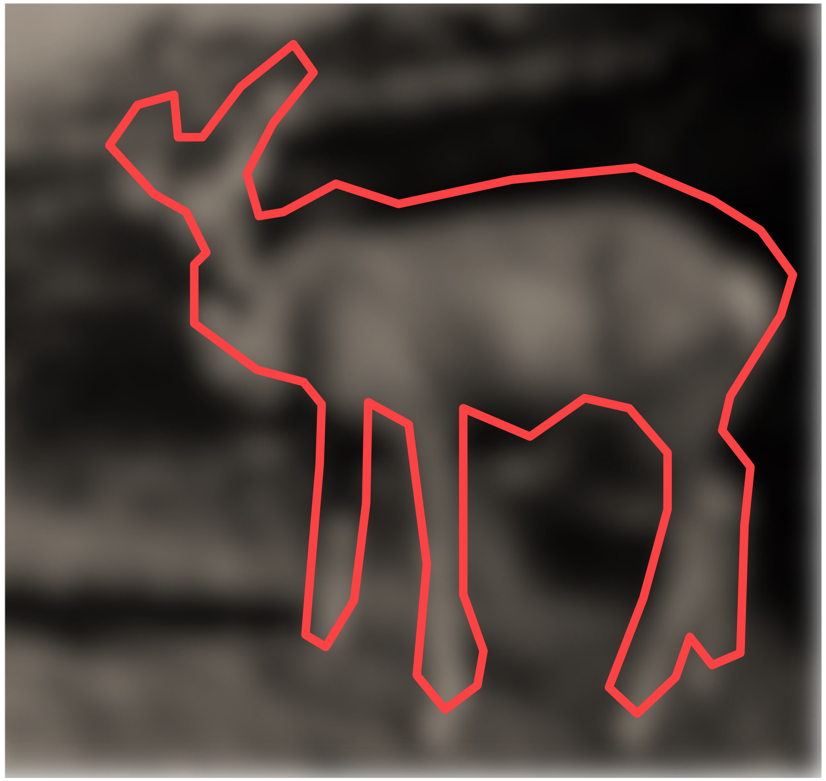


The segmentation step is often regarded as the **most difficult phase** of a CV processing chain. From one perspective, it is still one of earliest processing phases, where the variability/dynamics of the world should have more impact. Also, being at the basis of that "processing pyramid", failures in this step compromise the whole process.
As in the previous phase, there are hundreds of labs. / researchers exclusively concerned about the segmentation problem (particularly in the **medical informatics** topic)

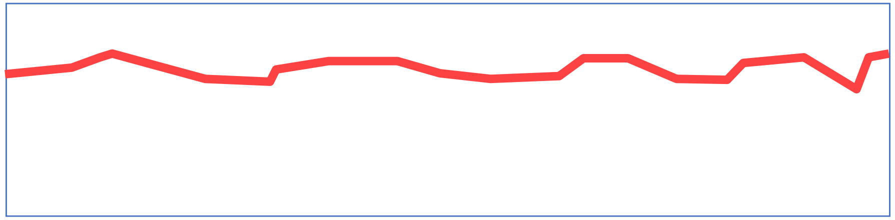
Computer Vision: Normalization

The normalization phase is often included in the feature extraction step and currently ignored in many deep learning-based approaches. The key in this phase is to obtain **representations that are invariant** – as much as possible – to the maximum number of data variability factors.

DL-based approaches are theoretically able to extract the "optimal" features, regardless of the input representations. Therefore, some of such approaches are *segmentationless* and also skip normalization.
The main problem – however – is that they only extract such optimal feature sets when receive **enough** learning data (i.e., millions?... trillions of samples?).



Again, normalization can be seen as a full system that receives a) a ROI and b) a parameterization of one object; returning a "*more friendly*" version of the data.

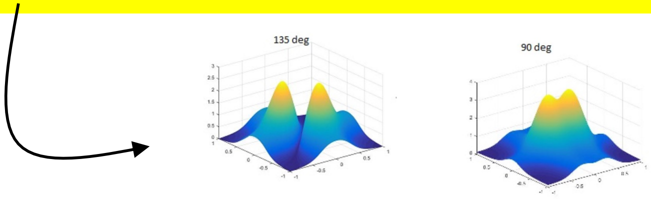


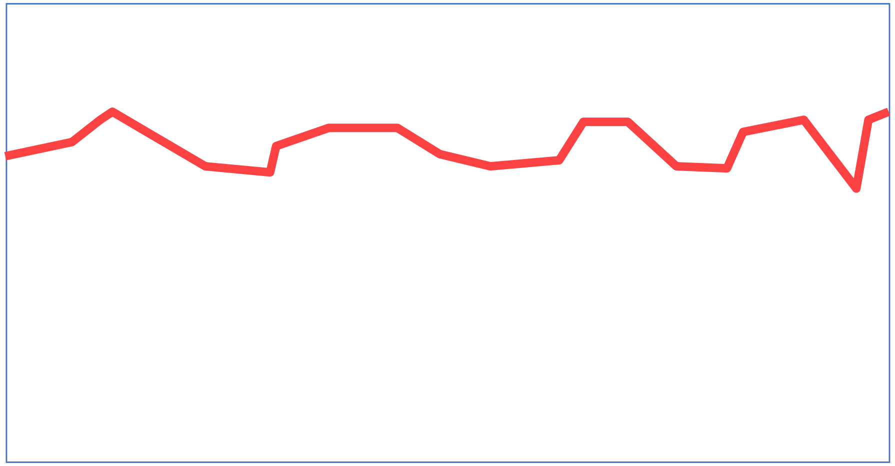
Computer Vision: Encoding (Feature Extraction)

This used to be the "core" of a CV problem, i.e., the phase where appropriate (and **invariant**) representations are extracted. Recently. DL-based methods replaced the human experts in this step and state-of-the-art techniques now extract the optimal feature sets in a fully autonomous way.

DL-based methods are **extremely data driven**. We should not expect reasonable results based in dozens, hundreds of even few thousands of learning samples. Even the best DL-based framework for a specific problem will **fail catastrophically** if it does not receive enough learning data.

For other problems where collecting and annotating enough learning data is not feasible, the "old-fashion" handcrafted features are still of **maximum importance**





01101000101010101001010101010
01010100010101010101010101010
10101010101001010101010101010
00010010101010101010100010011
01010001010101001010101010101
11010100101010010101001010101

Computer Vision: Matching

Matching is closely related to the previous phase (feature extraction) and involves to **find** the best possible **metric function** to compare representations of two different objects.

In such metric spaces, **similar** objects should correspond to **small distance values**, whereas **very different** objects should correspond to **large distances**.

The chosen function can be particularly suited for a **classification problem** (e.g., “yes”/”no”), or for a **regression problem** (e.g., value estimation).

```
011010001010101010010101010
010101000101010101010101010
101010101010010101010101010
000100101010101010100010011
010100010101010010101010101
110101001010100101010010101
```

```
11010101010101001010101010100
00010101010101010101010101010
01010101000101010101010101010
01010101000001011110001010101
01010000101010010101001010100
10101010010101010010101010100
```

E.g., Hamming distance, Euclidean distance, …

✗ ? ✓

In mathematics, a metric or distance function is a function that gives a distance between each pair of point elements of a set. A set with a metric is called a **metric space**. Three properties should be satisfied: **i)** d(x,y)=0 <-> x=y; **ii)** d(x,y) = d(y,x) and **iii)** d(x,y) <= d(x,z) + d(z,y)

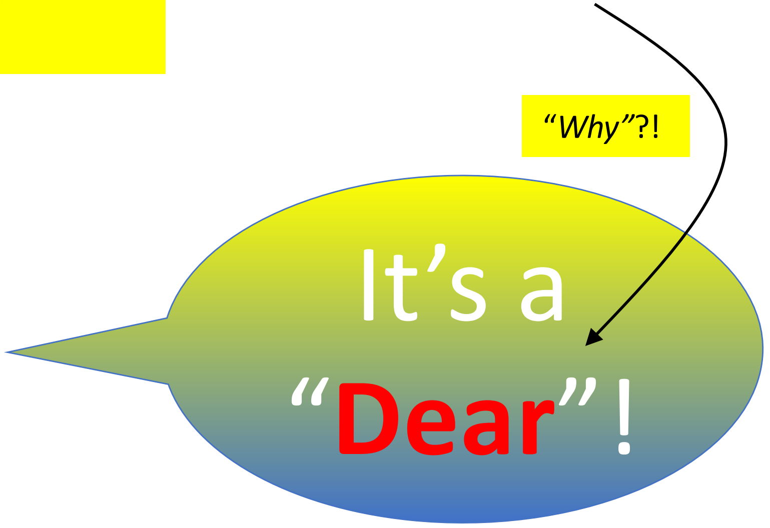
Computer Vision: Classification (Recognition)

The classification (or recognition) step is often integrated in the matching step. It involves to scan a database of **previously known obje**cts and – after applying the matching function to each pair – provide the final response (e.g., class or value) provided by the system.

Even not being considered the most challenging step from the technical/conceptual perspective, there are several subtleties in this step that should be considered. For example: are we working in the **closed or open** world paradigm? (i.e., do we know all the possible responses that should be provided by the system? Or can the system respond "new element"? Also, what are the **prior probabilities** of each class/family/value? How should **outlier values** be handled?

Recently, substantial attention has been paid to the development of **interpretable/explainable** recognizers/classifiers, i.e., that are able to justify each response provided. This is one of the **hot topics** in CV/Machine learning nowadays.

```
11010101010101001010101010100
00010101010101010101010101010
01010101000101010101010101010
01010101000001011110001010101
01010000101010010101001010100
10101010010101010010101010100
```

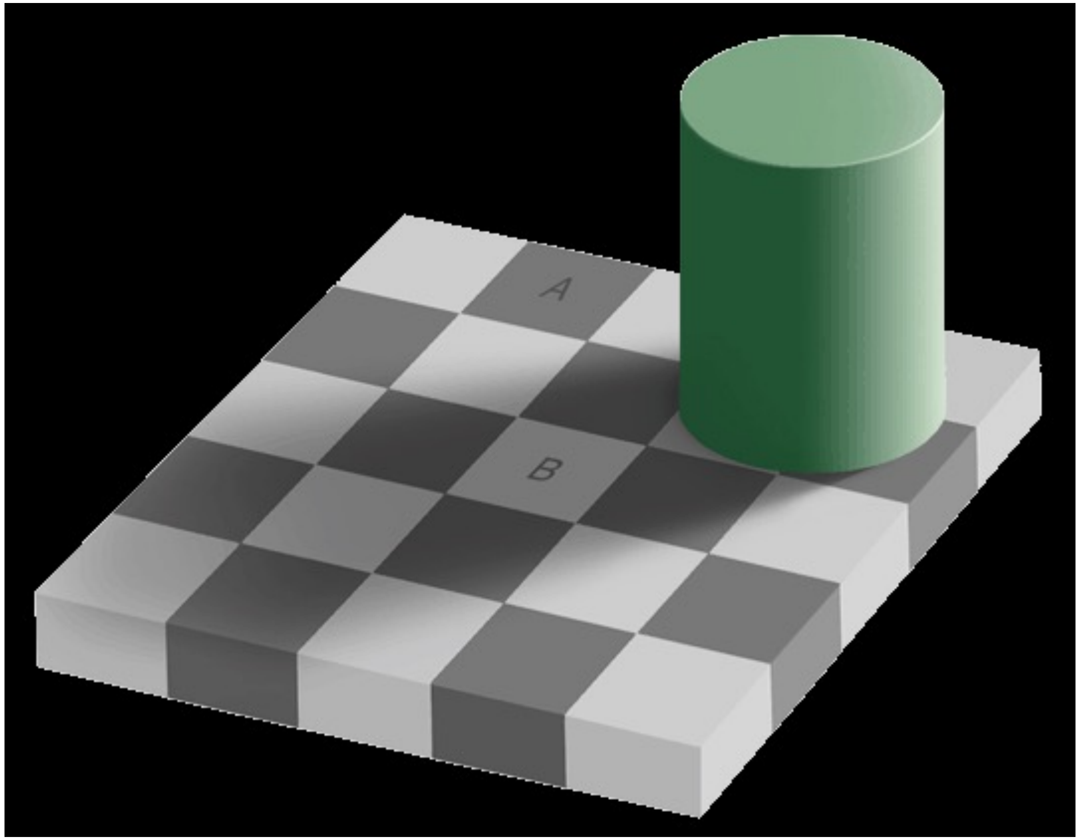


"Why"?!

It's a "**Dear**"!

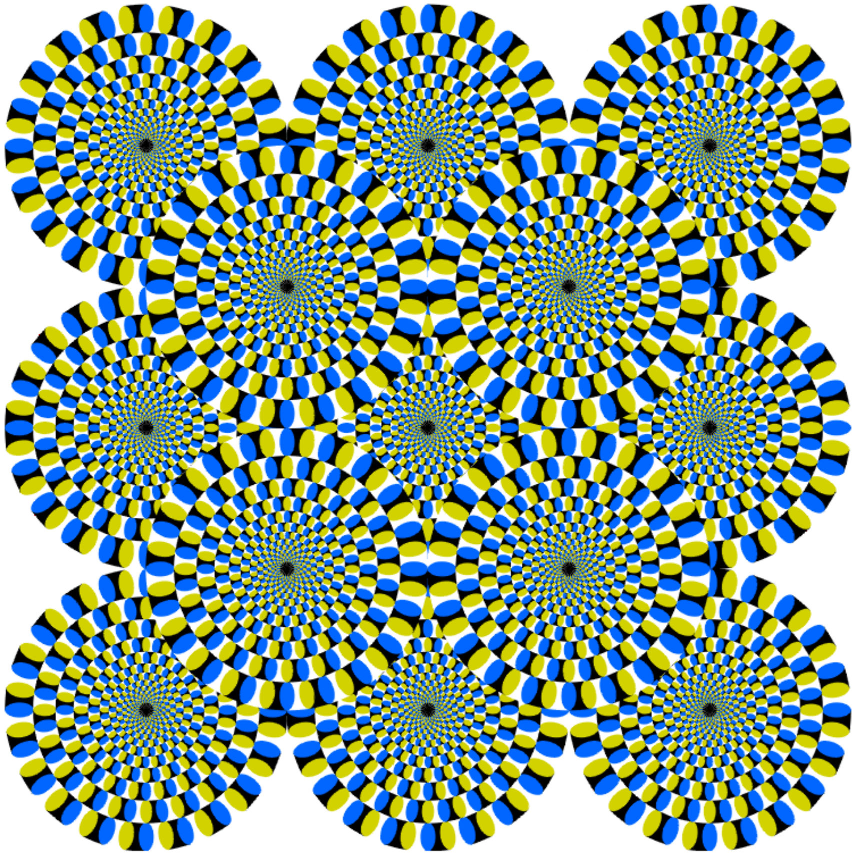
Vision, Why Is It Hard? Illusions

- What is the relationship between the color of regions “**A**” and “**B**” in the image below?
  - Which one is darker?



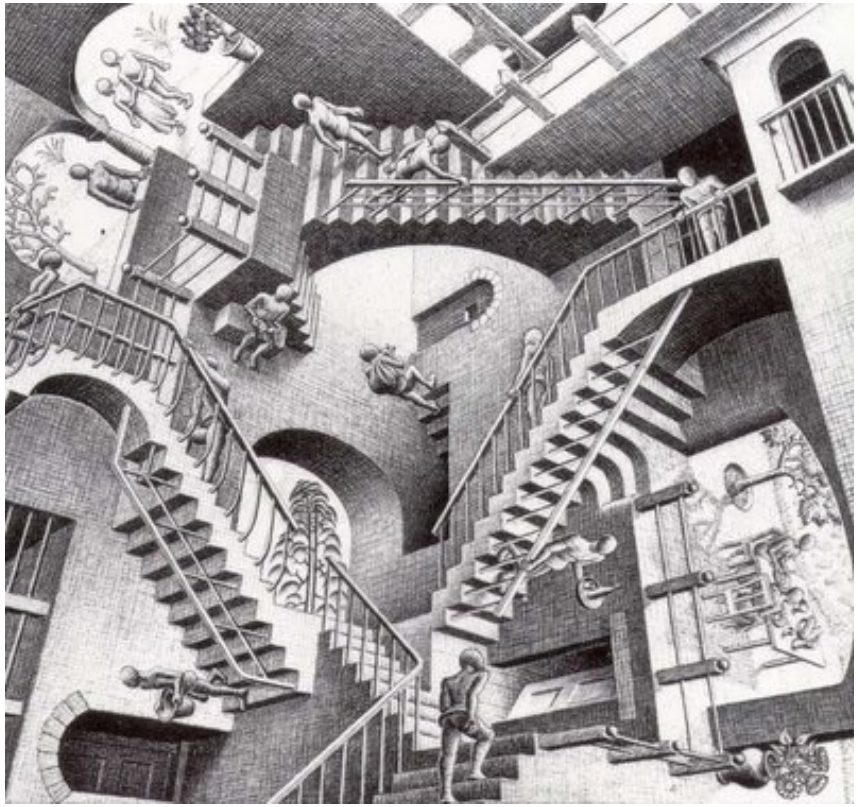
Vision, Why Is It Hard? Illusions

□ Do we have anything moving in this image?



Vision, Why Is It Hard? Illusions

- □Classical M.C. Escher illusions describe ***impossible worlds,*** *where stairs continously go up/down.*

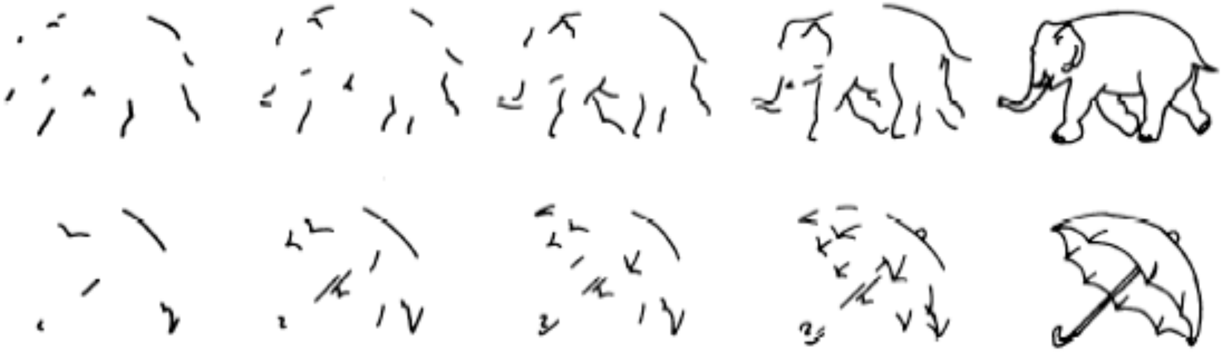


Vision: Why is it so Hard?

- □ Most of the vision problems are **ill-posed,** in opposition to well-posed problems
- □ Hadamard defined ***well-posed* mathematic models**, as those that meet the following properties:
  1) There is a solution;
  2) The solution is unique;
  3) Solution depends exclusively on the data.
- □ Not even considering other factors:
  - □ Ambiguity resulting from the representation of 3D worlds by 2D data
  - □ The variation/noise associated with data acquisition turns most vision problems ill-posed.
- □ As such, **errors are expected!!** (they should simply be minimized)

Vision: Why is it so Hard?

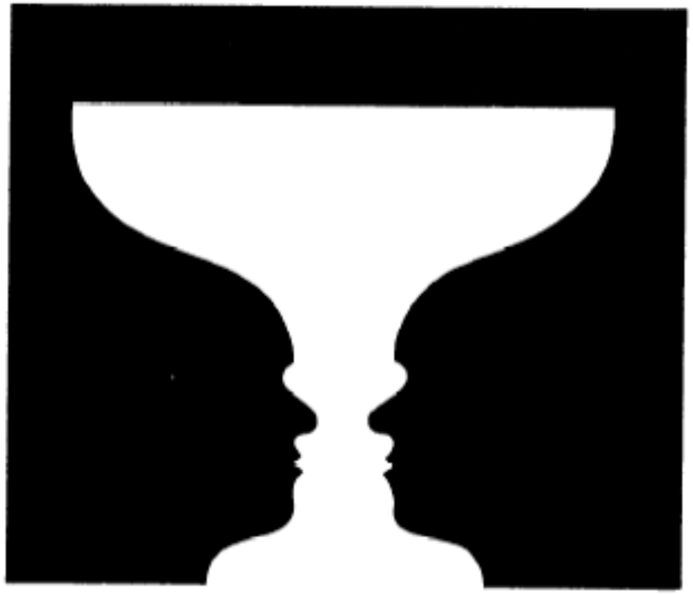
- □ The process of representing 3D data in two dimensions brings ambiguity to the represented data:
  - □ As na exemple, at what level of detail an elephant and an umbrella look-alike?
    - □ For sure not at the finest level. But what about the 1st, 2nd and 3rd coarsest levels?

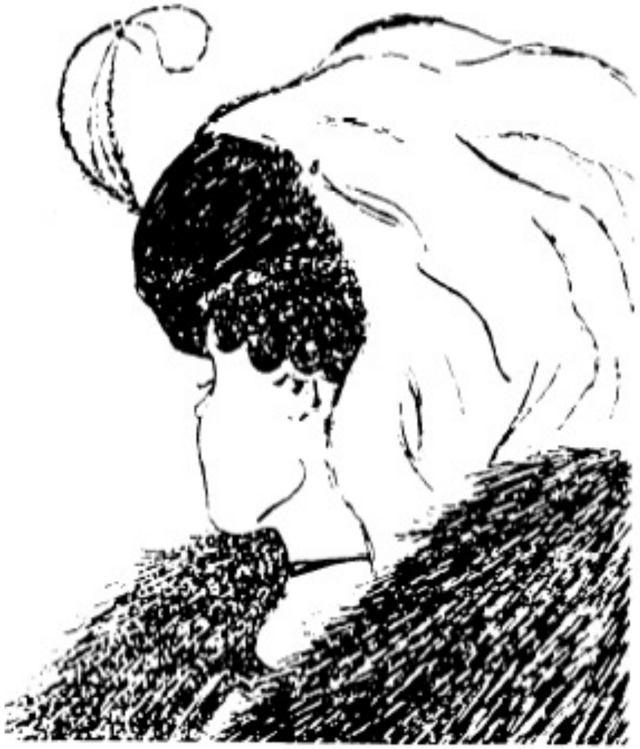


Vision: Why is it so Hard?

□Ambiguities

- □ The image at the left side shows 2 faces? Or a cup?
- □ And at the right side, do we have a yung girl? Or an old lady?

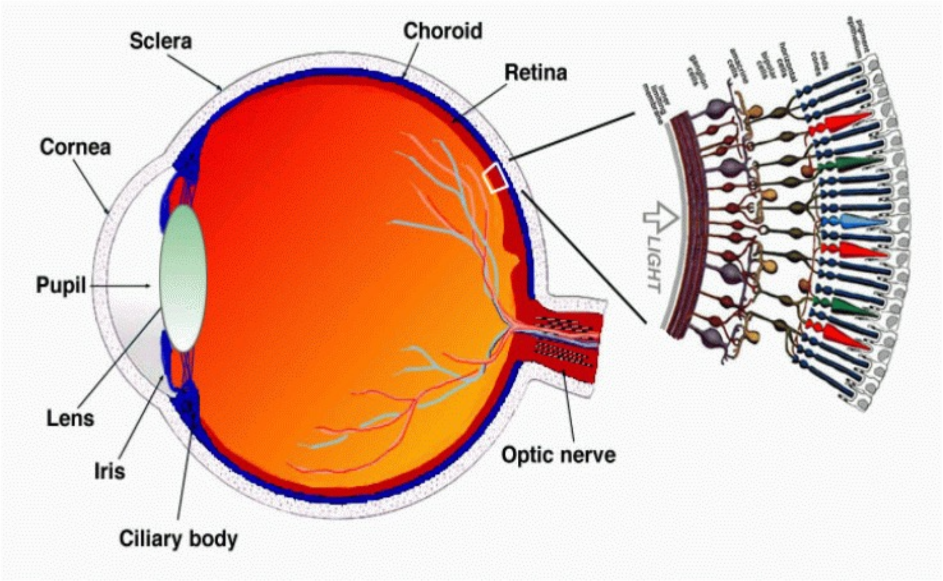




CV Biological Roots

- The human retina has around **120M** (120,000,000) **photo receptors**.
  - Of those, only around **6M are sensitive** to specific wavelengths (colors). These are called "**cone cells**".

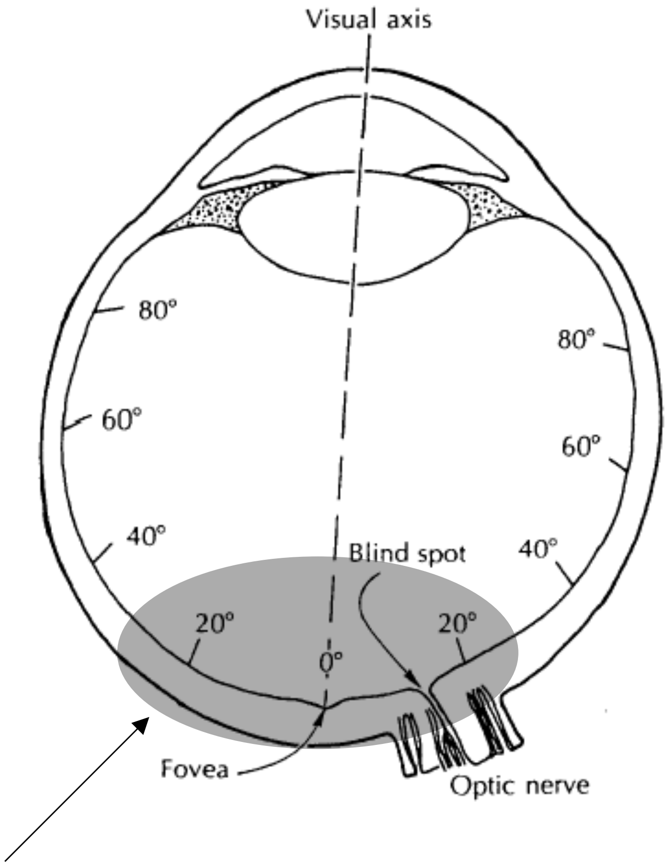
  - The remaining **119M** (around 95%) can only perceive "energy". They are called "**rod cells**".



CV Biological Roots

- **Cones** are located predominantly around the optical nerve (+/- 20º), having much higher spatial resolution than rods.
  - Cones have a strong connection to the brain.
  - The researchers examined the eyes of rhesus monkeys to determine the combinations of cone and rod "coupling."
    - They found the rods and cones were couple with one another and the red cones couple with the green cones, but the blue cones rarely connected with the other cones. The researchers hypothesized these couplings amplify specific electrical signals and help animals to see more details.

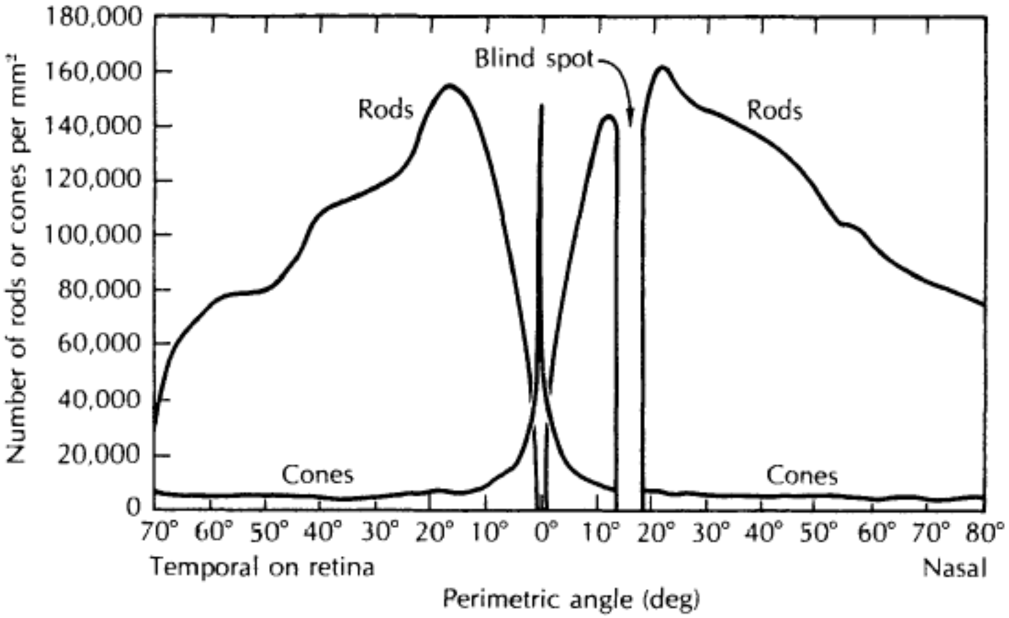
- Muscles that control the eye movements change the position of the optical globe up to the moment where image falls inside the central part of the retina (designated as **fovea**).



Cones region

CV Biological Roots

- □ Cones and rods **are distributed in local structures** around the retina, in direct proportion to the angular distance to the optic nerve.
- □ The combined stimulus from the three different types of rods enables the perception of millions of different colors by the human brain.
- □ *Daltonism* is a disease related with the absence of rods, or with deficiencies in their connections to the optical nerve.



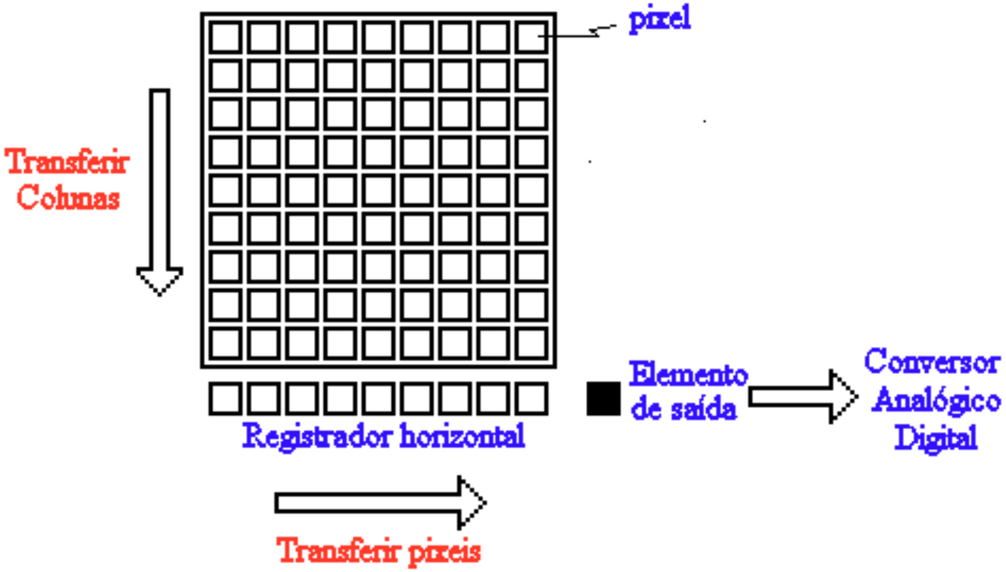
CV Biological Roots

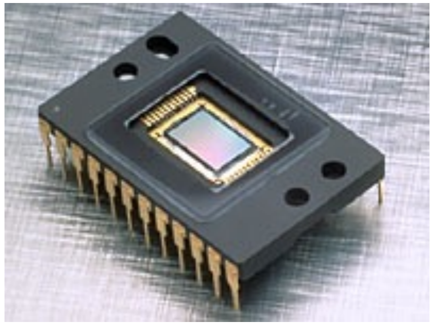
- However, the human retina cannot be considered a simple “input device” for the brain.
- Note that the retina:
  - **Has over 120M input channels** (cones and rods photoreceptors).
  - However, **it has only 1M output channels** to the brain.
  - It is known that **input** and **output** channels operate at **similar frequencies**.
- For sure, an important part of the vision task might have been already done when information reaches the brain.
  - Hence, only the most relevant visual information arrives to the brain.
- In conclusion, the retina cannot be considered *adjacent to the brain*

Instead, the retina is (part of) the brain!!

Image Basics: Charged Couple Devices (CCDs)

- At the bottom line, CCDs are devices (arrays) with the **ability to convert light (photons) into electrons (electrical charge).**
- The smallest conversion units are called **pixels** (picture elements), with dimensions smaller than 0.013 mm^2 each.
- Upon exposition to light, the amount of electrical charge in each pixel is transferred iteratively to its neighbor (in vertical direction). The last sensorial unit transfers its content into an amplifier, from where an *analogic-to-digital* converter operates.

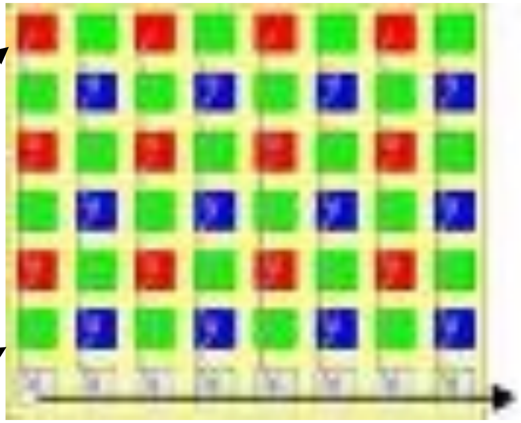




Charged Couple Devices and Color

- Color information results (at least conceptually) from three different arrays, each one with different color filters.
- However, in practical terms, a single CCD contains different types of photo receptors, so that each one handles different wavelengths.
- This reduces a bit the spatial resolution (i.e., it is not possible to perceive a specific wavelength at a specific position.
  - On the other way, it makes possible to

Produce CCDs in a much cheaper way.



Cannot *"see"* blue/green here!!

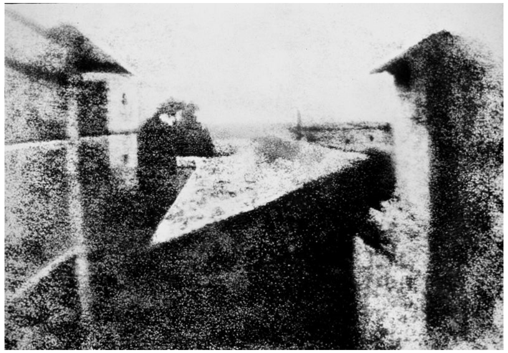
Cannot *"see"* red/blue here!!

Charged Couple Devices

- □ There are at least two quantization processes evolved in the acquisition of pictorial data:
  - □ There is a finite number of available pixels. (this determines the spatial resolution of the sensor)
  - □ Each pixel is represented by a finite number of bits, in terms of the amount of electrical charge. (this determines the luminance resolution, i.e., the number of colors/intensities perceived)
- □ **Spatial resolution**: it depends on the density of photoreceptors inside the sensor and by the properties of the used lens
- □ **Luminance resolution**: it depends on the analog-to-digital converter properties and by the number of bits used to store information.

Image Acquisition: Brief History

- The first photography was due to French inventor *Joseph Nicéphore Niépce*, in 1816.

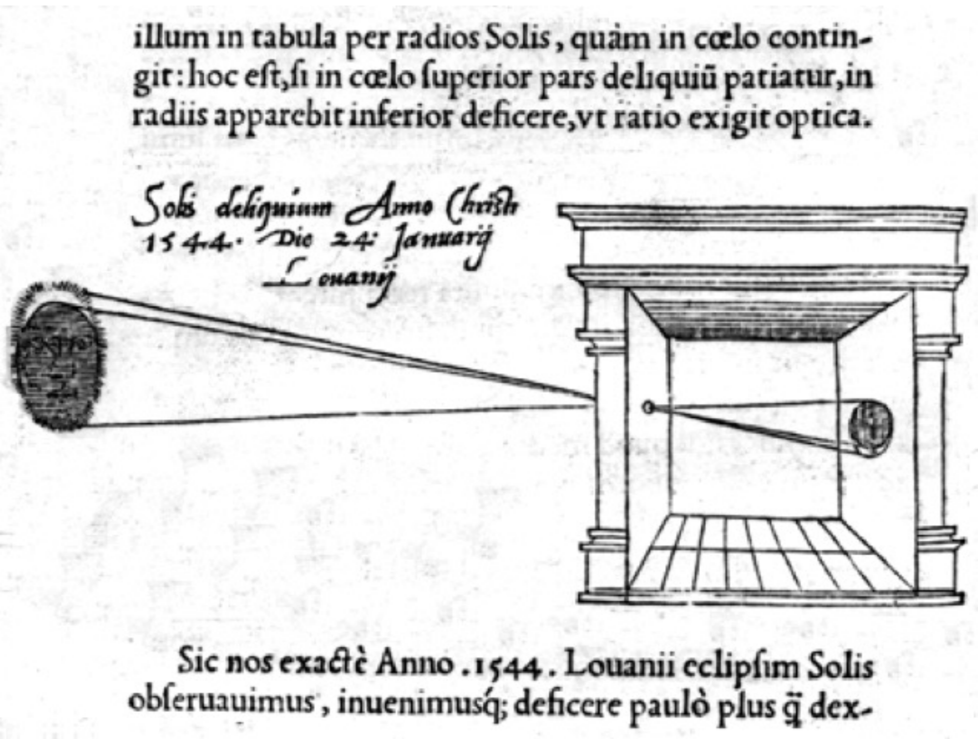


- Experiments started in 1793, but images were not persistent.

- In the image at the right, the process was known as "**heliography**" and – for this pioneer – experiment, it took about eight hours to collect the enough amounts of light to produce the image.

Image Acquisition: Brief History

- Leonardo da Vinci (1452-1519) first formalized the concept of ***camera obscura***.
- *(sic) "When images of illuminated objects penetrate through a small hole into a very dark room you will see [on the opposite wall] these objects in their proper form and color, reduced in size in a reversed position, owing to the intersection of the rays".*



illum in tabula per radios Solis, quàm in cœlo contingit: hoc eſt, ſi in cœlo ſuperior pars deliquiũ patiatur, in radiis apparebit inferior deficere, vt ratio exigit optica.

Solis deliquium Anno Christi
1544. Die 24: Januarij
Louanij

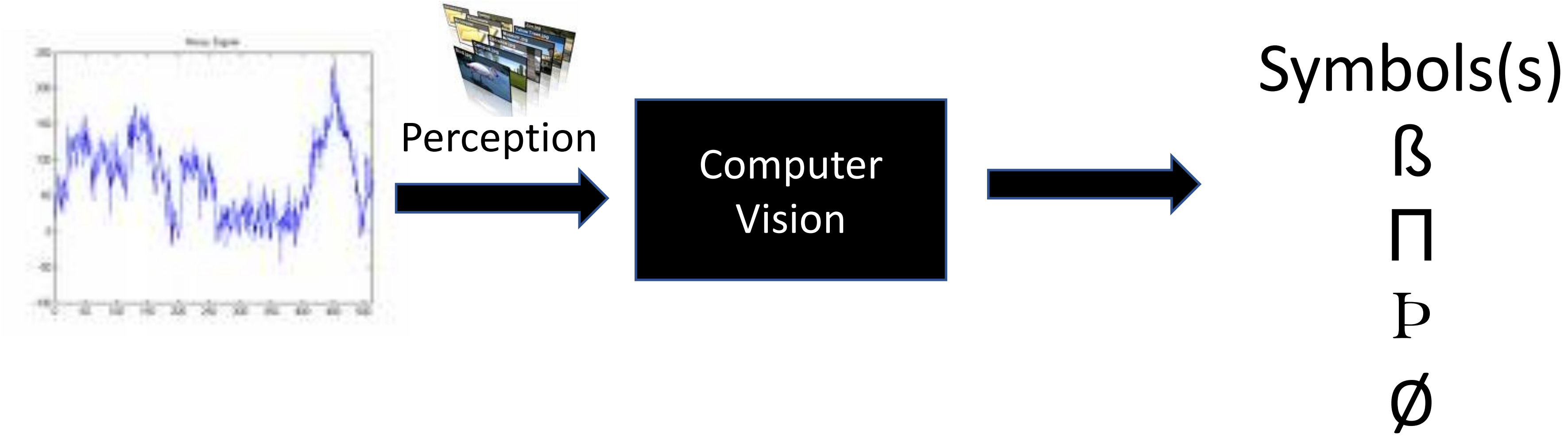
Sic nos exactè Anno .1544. Louanii eclipſim Solis obſeruauimus, inuenimusq; deficere paulò plus q̃ dex-

Computer Vision? What is It?

- □ *Trucco and Verri*: "**Computing properties** of the 3-D world from one or more digital images".
- □ *Sockman and Shapiro*: "To **make useful decisions** about real physical objects and scenes based on sensed images".
- □ *Ballard and Brown*: "The construction of explicit, **meaningful descriptions** of physical objects from images".
- □ *Forsyth and Ponce*: "**Extracting descriptions** of the world from pictures or sequences of pictures".
- □ *English Dictionary*: "The use of digital computer techniques to **extract, characterize, and interpret information** in visual images of a three-dimensional world".
- □ *Wikipedia*: "Computer vision is the science and technology of machines that see. As a scientific discipline, computer vision is concerned with the theory for building artificial systems that obtain **information from images**".

Computer Vision? What is It?

- □ It can be considered a full **Artificial Intelligence** problem
- □ As such, at the bottom line, it can be simply regarded as a "signal-to-symbol" converter.
  - □ In complete opposition to "Computer Graphics", which can be regarded as a "symbol-to-signal" converter (i.e., signal synthesis)


Perception

Computer Vision

Symbols(s)

ß

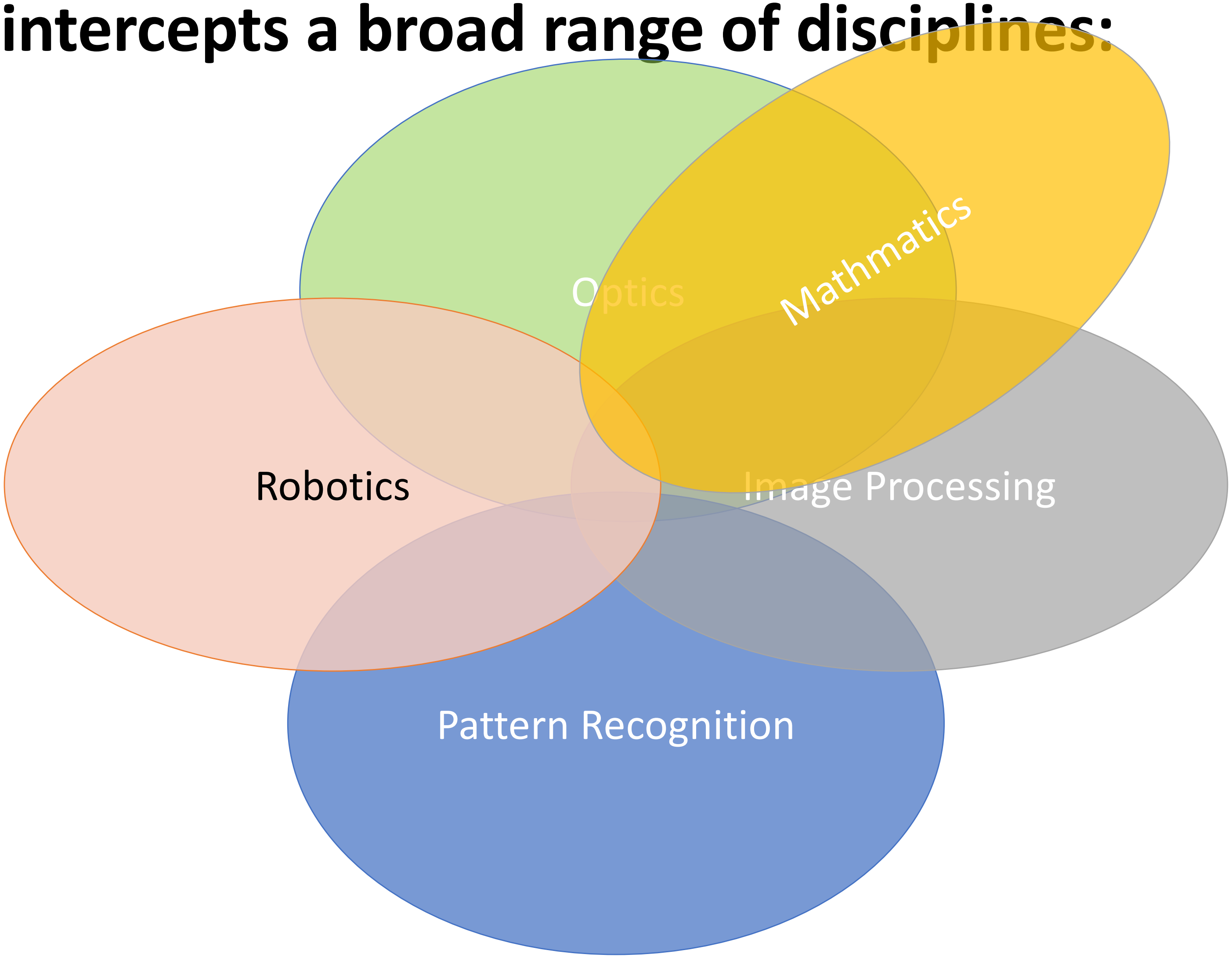
∏

Þ

Ø

Computer Vision? What is It?

- **CV intercepts a broad range of disciplines:**



Optics

- □ Is related to the biological sensing process of vision (the way light is handled by human brain).
  - ☐ Describes the behavior of light and its interaction with matter.
  - ☐ Three main types of light can be identified, using as reference visible wavelength: **infra-red, visible and ultra-violet**.
  - ☐ However, being a radiation, similar fenomena also occur in x-rays, micro-waves, radio waves or any other type of radiation (interaction between charged particles)

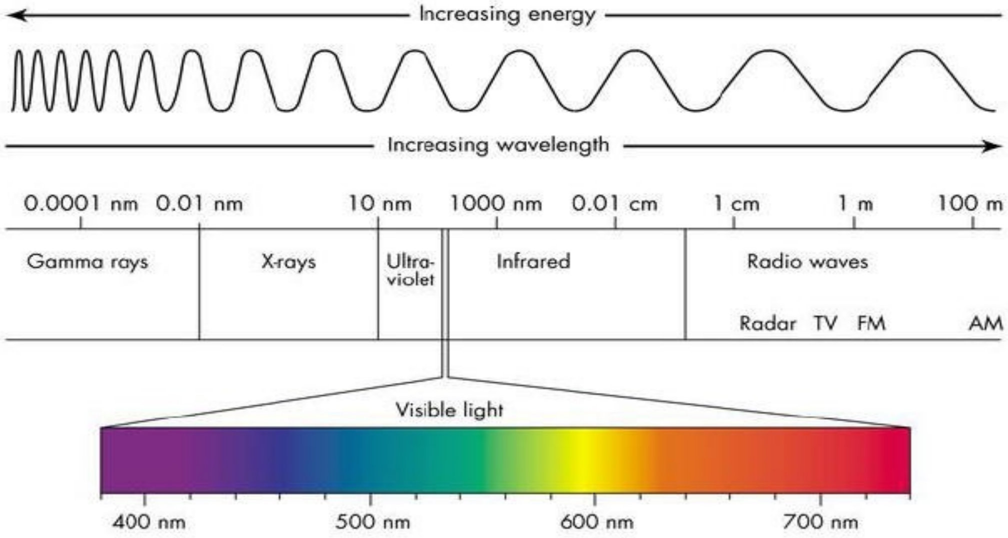
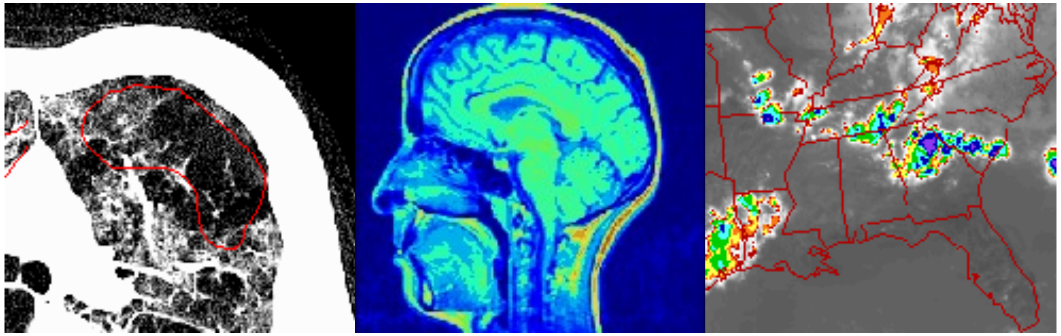
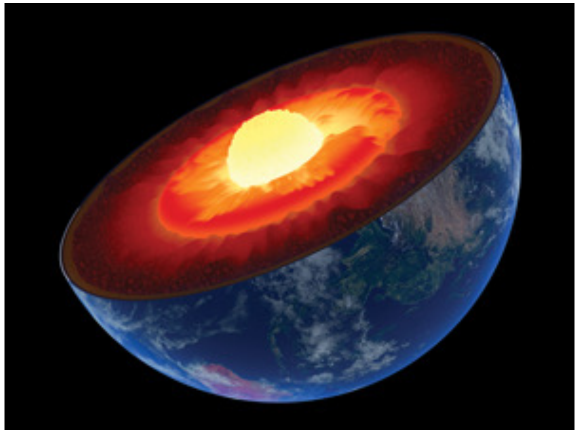


Image Processing

- Discipline that **processes signals**, having as input **bidimensional data**.
- Transformation of the original data, in order to make easier further interpretation phases.
  - Geometric transforms (scale, rotation, translation, affine and projective transforms).
  - Color or intensity adjustement
  - Data / region reconstruction
  - Data registration
  - Detection
  - Segmentation
  - recognition

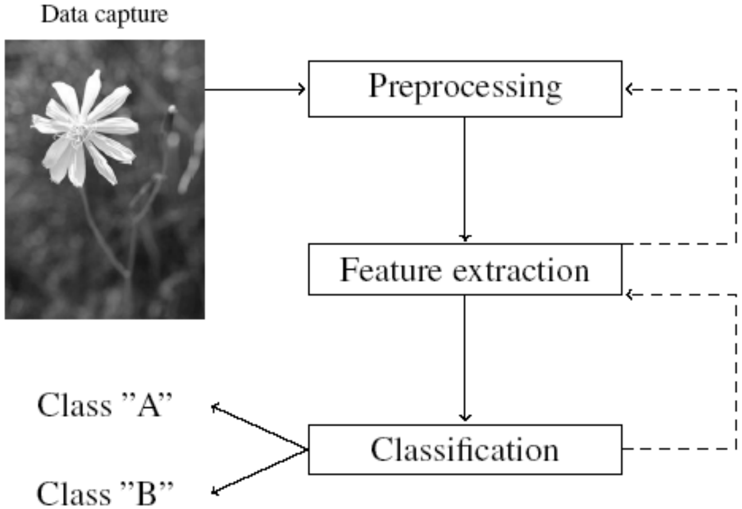


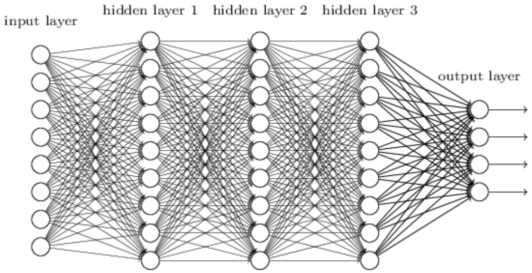
Pattern Recognition



- It is often considered the **core** of a vision system
- Pattern Recognition aims at **classify/labelling** the input data
- There are, typically, three variants:
  - Statistics
  - Structural
  - Neural

Hot topic nowadays!!!





Robotics

- Domain of knowledge that evolves planning and development of phisical automata (robots)
- It intersects electronical engineering, mechanics, computer science and cybernetics areas.
- Even though a precise definition is hard to find, a robot is a machine that:
  - Has **sensorial** abbilities
  - It has the ability to **actuate** in the environment, and change its state.

Computer Vision: Applications

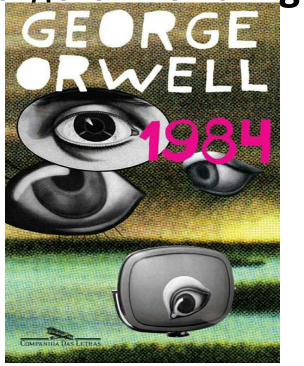
□Biometric Recognition:

- □ It is perhaps **the most studied** application of computer vision/artificial intelligence domains;
- □ An attempt to develop **machines able to recognize human beings**;
- □ Can use many different **traits**: Iris, face, gait, …

Computer Vision: Applications

- □ **Surveillance / Security Systems**
  - □ At the current level, in opposition to common beliefs, most of the analysis is still made by humans in the loop.
    - □ **Tedious task**
    - □ **Prune to errors**

- □ One of the main challenges in CV is to develop automata able to **continuously identify humans** “**in-the-wild**”, in a way like George Orwell’s **Big Brother.**

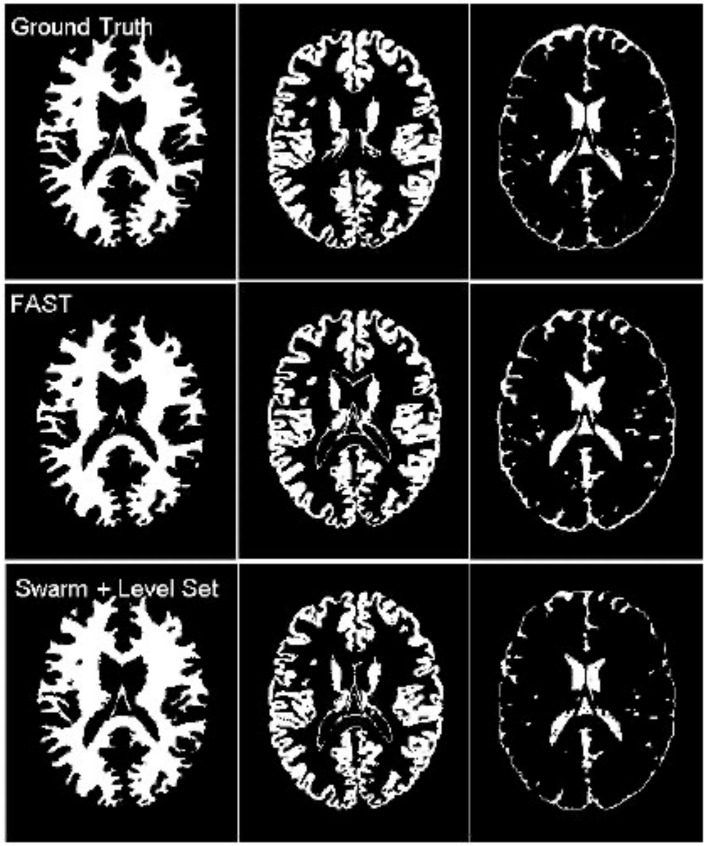


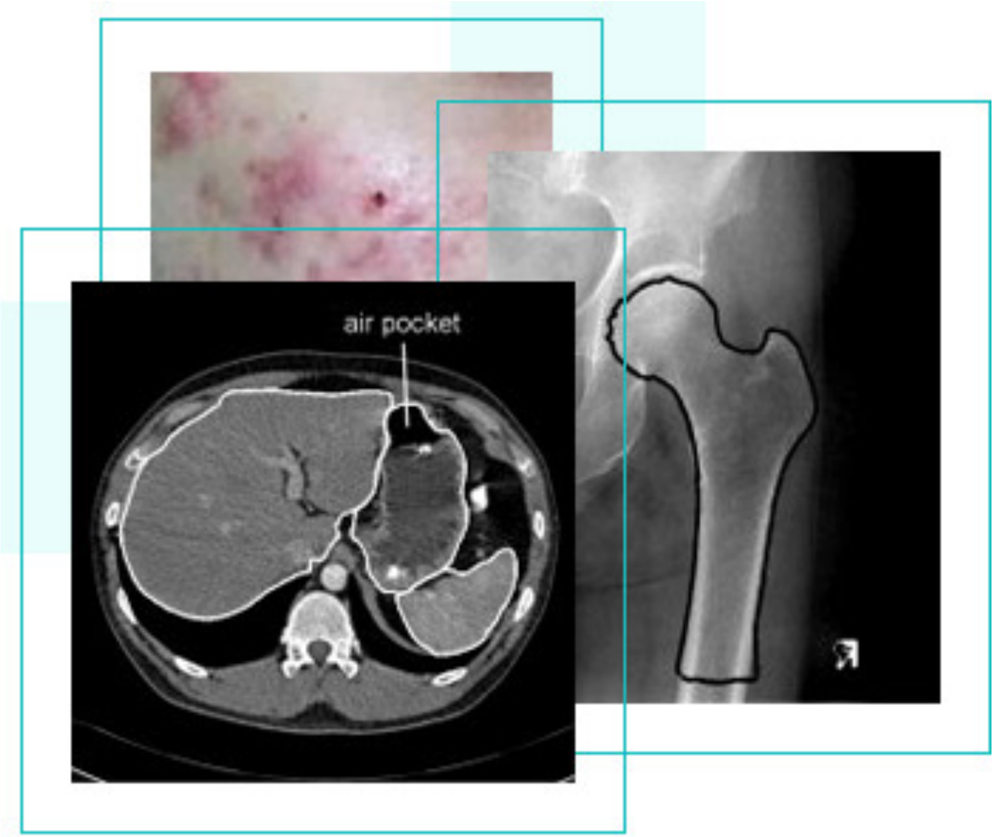
Computer Vision: Applications

- □ Autonomous Driving, Navigation
  - □ **The "Google car" is the most well known (even though Stanford's might be more appealing…)**

Computer Vision: Applications

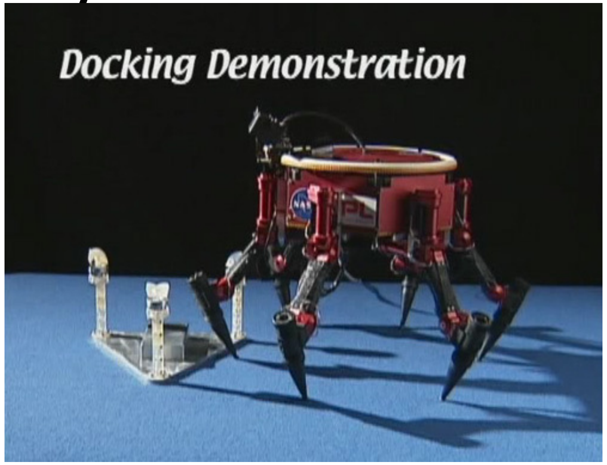
□Medical Image Analysis





Computer Vision: Applications

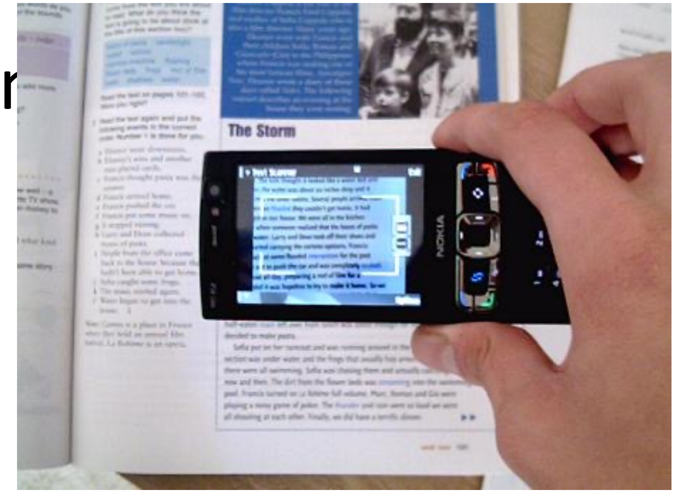
- □Robotic Production/ Inspection Systems
  - □ **NASA's autonomous walker:**



  - **Subsea 7 inspector**

Computer Vision: Applications

□Automatic Character recognition



- □ **NOKIA multi-scanner**

□Defense Systems (Ballistics)

- □ **China Dongfeng 21D with infligh**

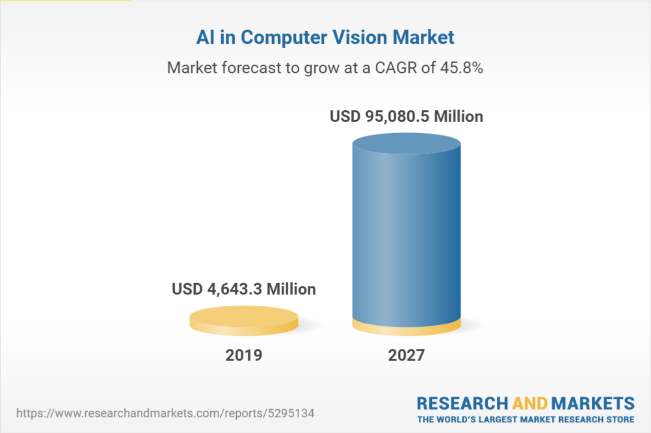
Computer Vision: Research

- □ **CV is one of the more active disciplines** in the Computer Science domain.
- □ However, **it is still in the earliest development stage**, as there are not autonomous and generic systems with vision abilities close to the human being.
  - □ This was evidently true up to a couple of year ago, when “deep learning” frameworks provided the last breakthrough
- □ We know that this type of vision-problems can be solved, as humans do it for thousands of years. However:
  - □ How to represent knowledge?
  - □ What inference mechanisms should be created?
  - □ What is intelligence?
- □ International Conferences
  - □ **ICCV**: International Conference on Computer Vision
  - □ **CVPR**: Computer Vision and Pattern Recognition International Conference
  - □ **ICPR**: International Conference on Pattern Recognition
- □ International Journals
  - □ **IVC**: Elsevier *Image and Vision Computing*
  - □ **CVIU**: Elsevier *Computer Vision and Image Understanding*
  - □ **TIP**: IEEE Transactions on Image Processing
- □ Hundreds of Research Groups
  - □ Académicos: MIT, Stanford, Cambridge, UCLA, ...
  - □ Comerciais: Microsoft, IBM, Honda, Sarnoff, Panasonic, ...
  - □ http://www.cs.cmu.edu/~cil/v-groups.html

Computer Vision Market

□CV is surely the sub-topic of AI where **the most evident increase** in the global market value is expected.

□The computer vision market was valued at **US$ 4,643.3 million** in 2019 and is projected to reach **US$ 95,080.5 million** by 2027; it is expected to **grow at a CAGR of 45.8%** from 2019 to 2027.



Source: https://www.researchandmarkets.com/reports/5295134/ai-in-computer-vision-market-forecast-to-2027