Università
Ca' Foscari
Venezia

Computer Vision

The image formation process

Filippo Bergamasco (filippo.bergamasco@unive.it)
http://www.dais.unive.it/~bergamasco
DAIS, Ca' Foscari University of Venice
Academic year 2017/2018

Università Ca' Foscari Venezia

The image formation

Almost all the computer vision topics we'll discuss in this course require the concept of "image"

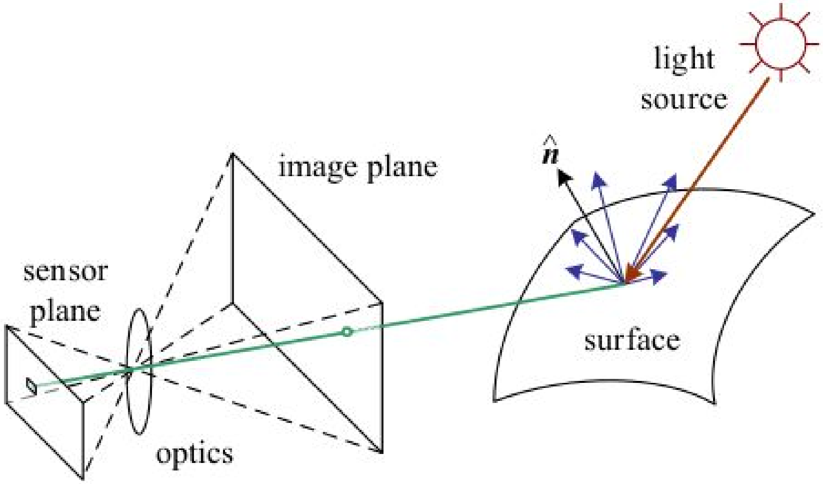
We now discuss how a real world scene is imaged through a digital camera, specifically:

- How it's acquired and stored
  - sampling/quantization
- How digital images are represented
  - 2D Functions, matrices
- Relationship between pixels
  - Neighborhood, Connectivity, etc

All starts with light...

To being able to see any 3D scene, our eyes or a digital camera needs to capture the light radiation reflected from scene surfaces

To produce an image, the scene must be illuminated with one or more light sources

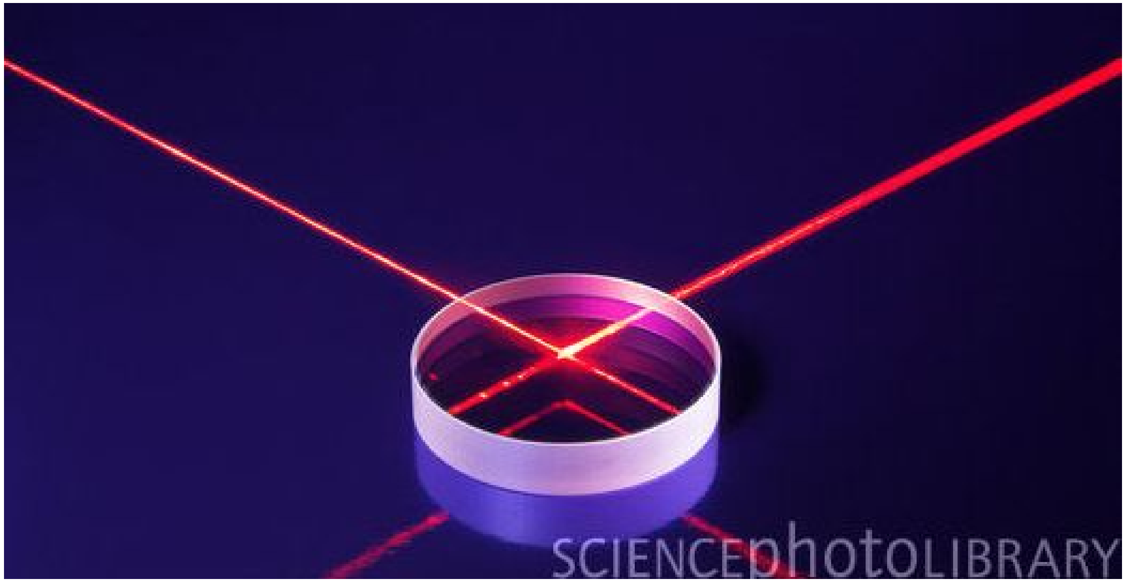


Università Ca' Foscari Venezia

The nature of light

Light has a dual nature:

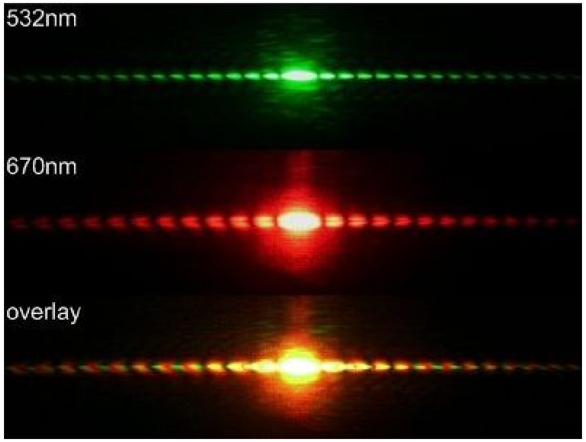
1. Can behave like a particle (photon)
   a. Travels in a straight line
   b. Light can reflect off a mirror and cast shadows
   c. Consists of tiny bits of energy behaving like discrete packets

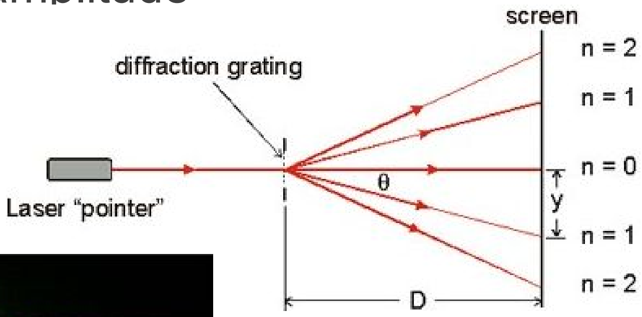


The nature of light

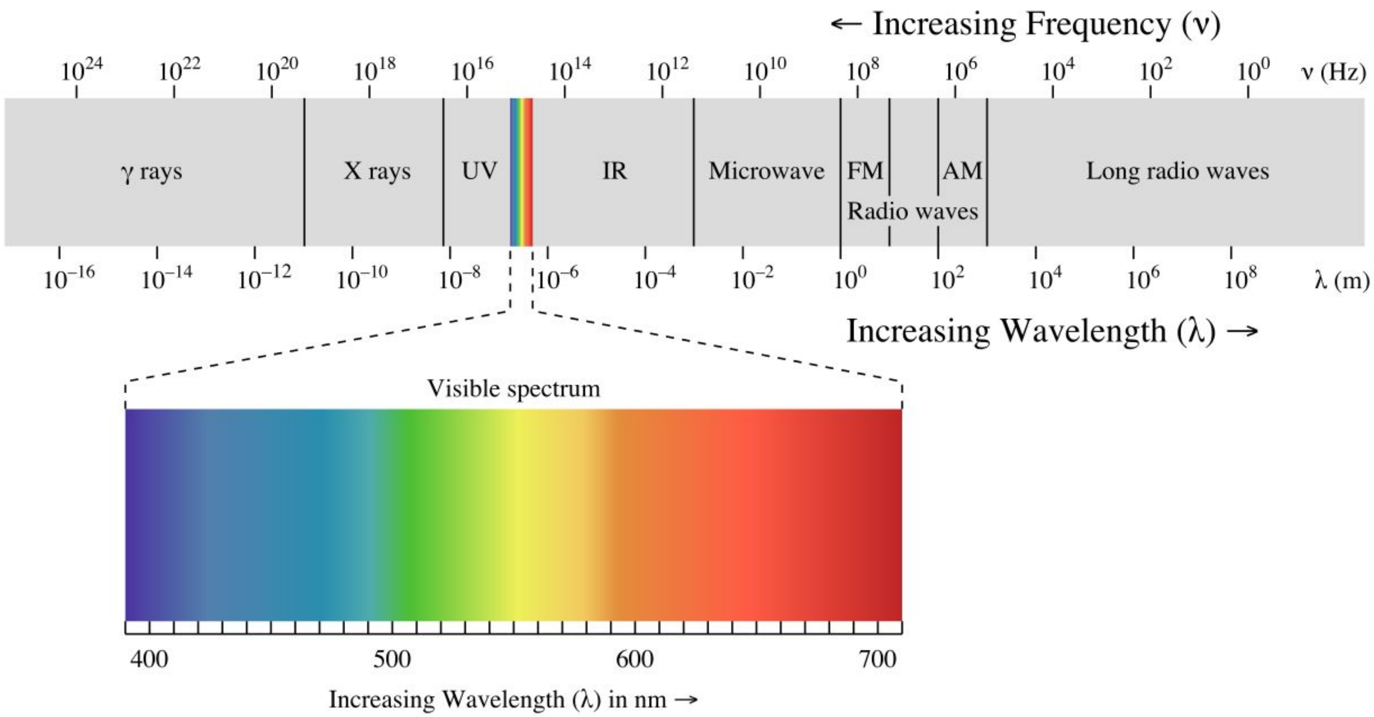
Light has a dual nature:

1. Can behave like a wave
    a. Refraction
    b. Diffraction
    c. Has a wavelength and an amplitude



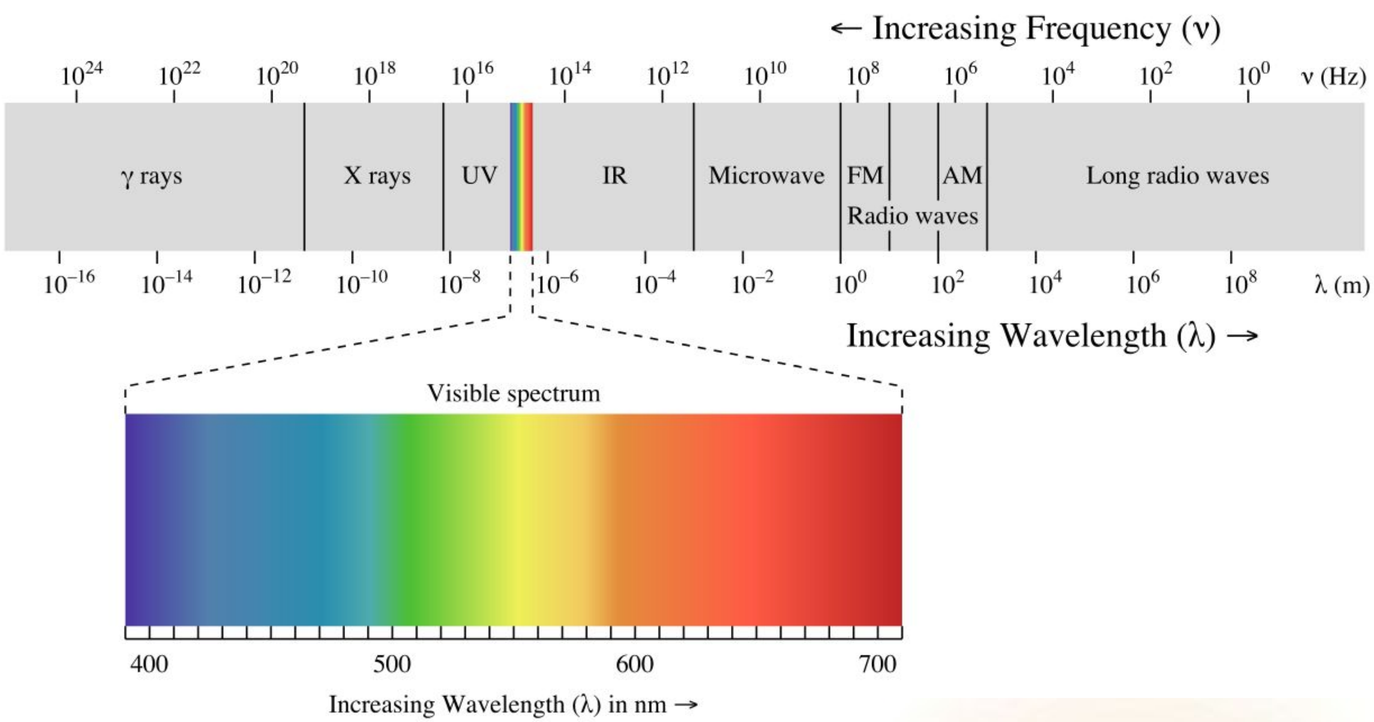


Visible light is part of electromagnetic spectrum

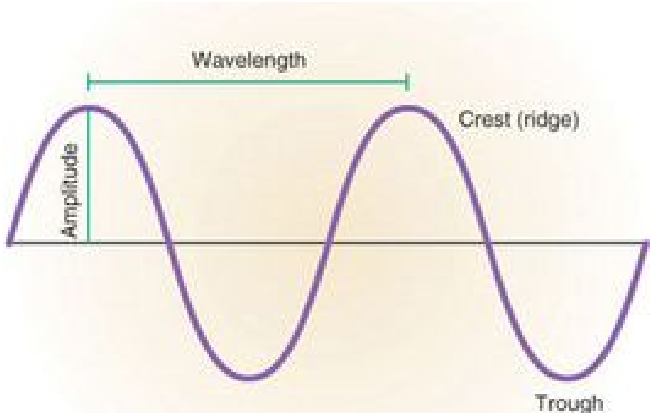


The electromagnetic spectrum can be expressed in terms of wavelength (**λ**) or frequency (**ν**)

Visible light is part of electromagnetic spectrum



$$\lambda = \frac{c}{\nu}$$



Università Ca' Foscari Venezia

Light quantities

The colors that humans perceive in an object are determined by the nature of the light reflected from the object

- A white object reflects light uniformly among visible wavelengths
- Green objects reflect light primarily in the 500-570 nm range

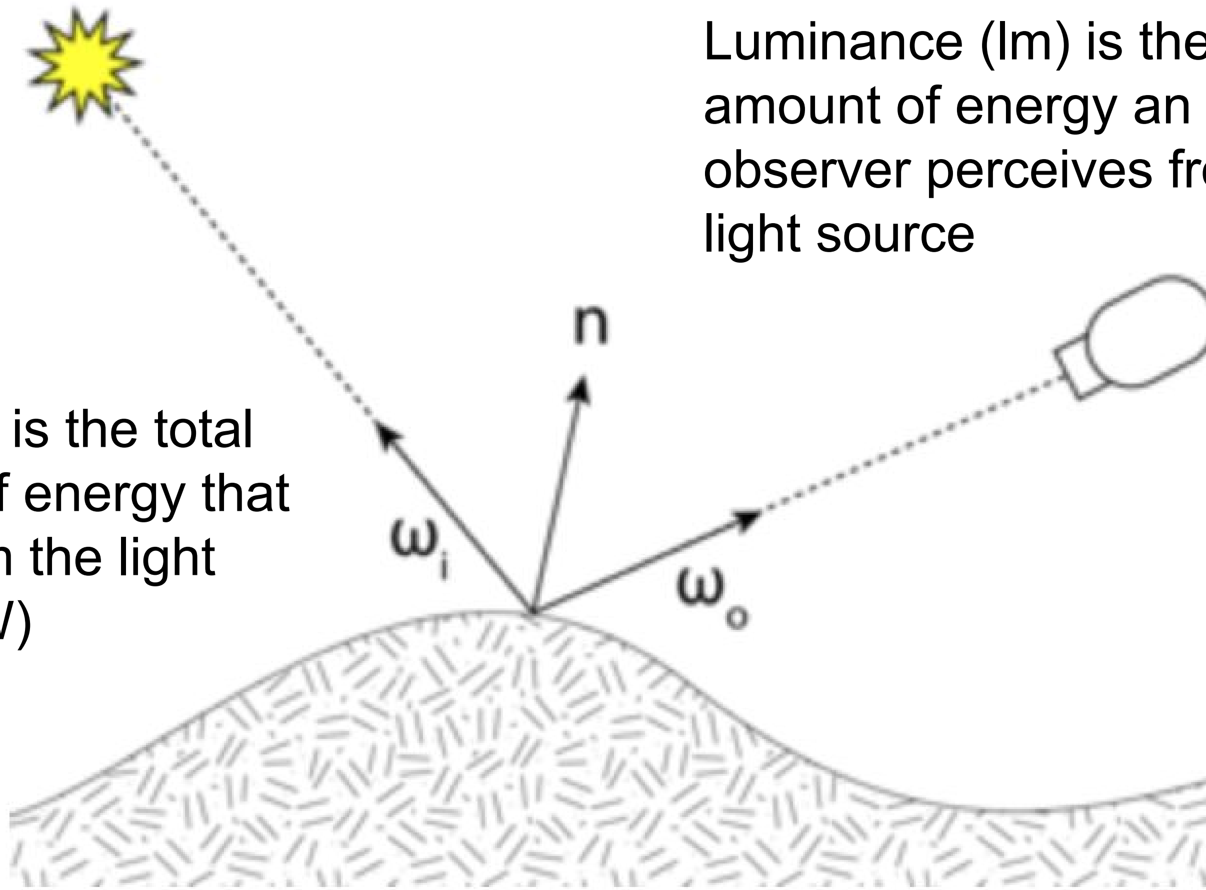
Quantities that are usually used to describe a light source:

1. Radiance
2. Luminance
3. Brightness

Light quantities

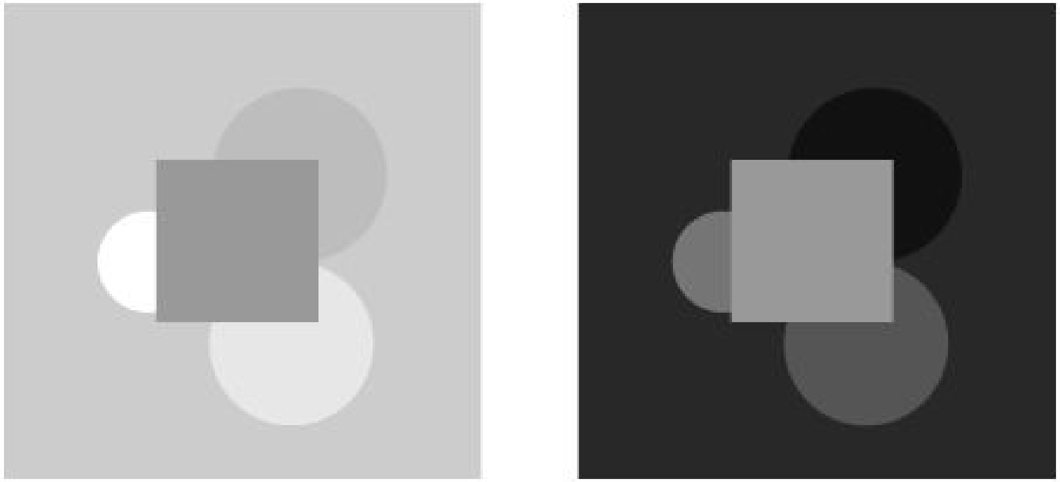
Luminance (lm) is the amount of energy an observer perceives from a light source

Radiance is the total amount of energy that flows from the light source (W)



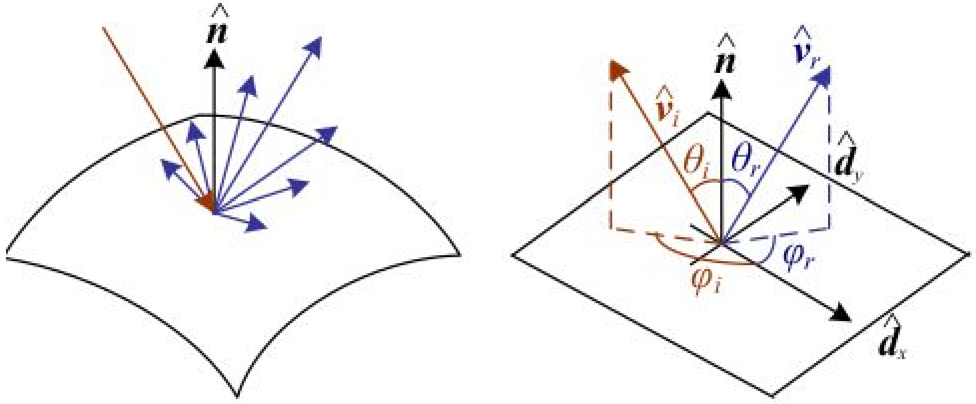
Brightness

Brightness is a subjective descriptor of light "intensity" and is one of the key factors in describing color sensation



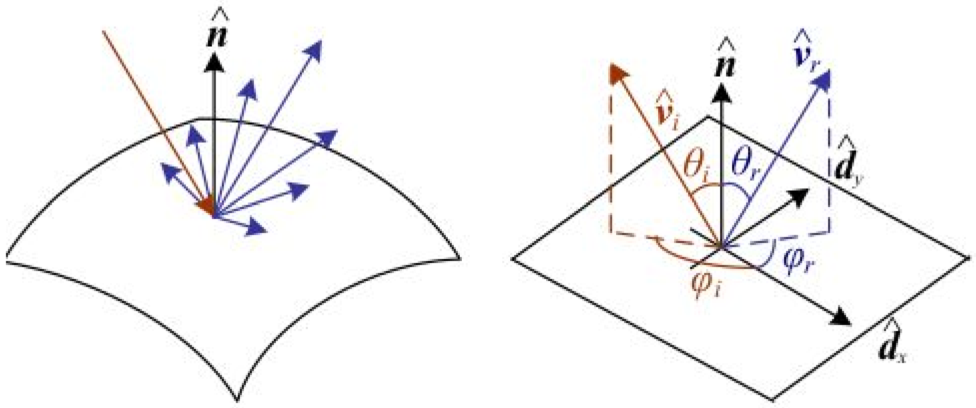
Luminance of the inner square is exactly the same. Perceived brightness is different

The BRDF



When light hits a surface it is scattered and reflected. The most general way to model this interaction is through a 5-dimensional function called **BRDF** (Bidirectional Reflectance Distribution Function)

The BRDF

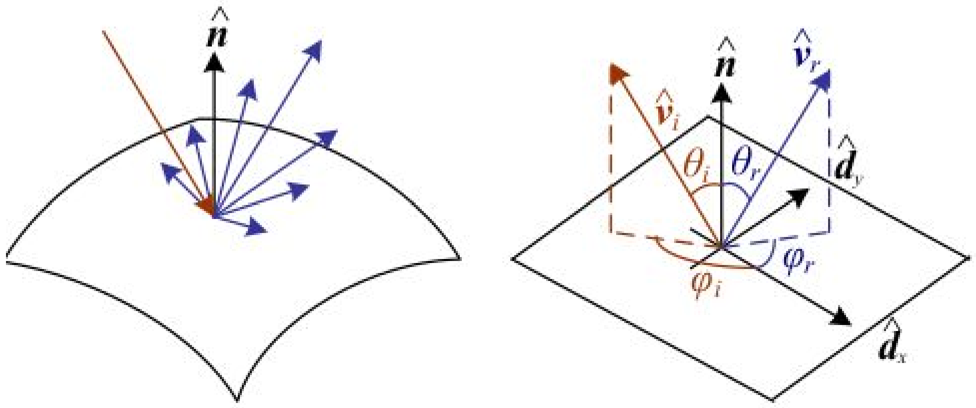


$$f_r(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)$$

Properties:

- Always positive
- Helmholtz reciprocity (can exchange i with r)
- Conserving energy

The BRDF

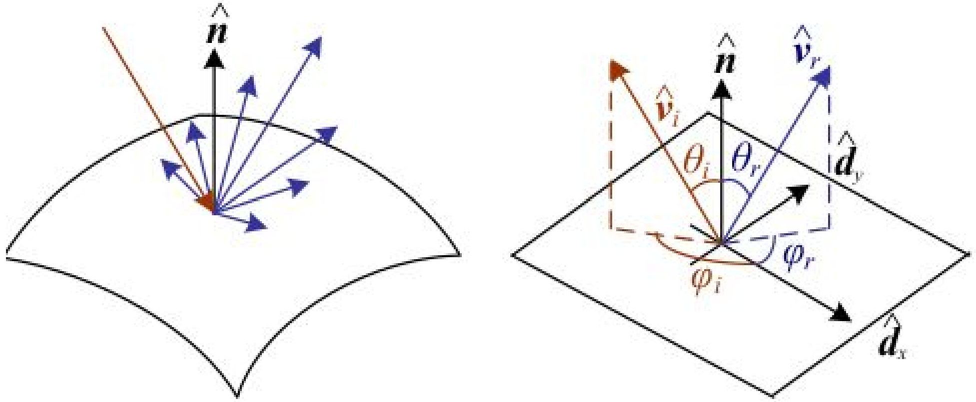


$$f_r(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)$$

BRDF of a given surface can be obtained through:

- Physical modeling
- Heuristic modeling
- Empirical observation

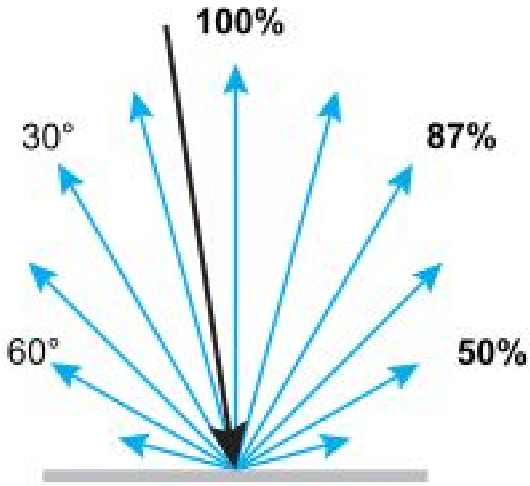
The BRDF

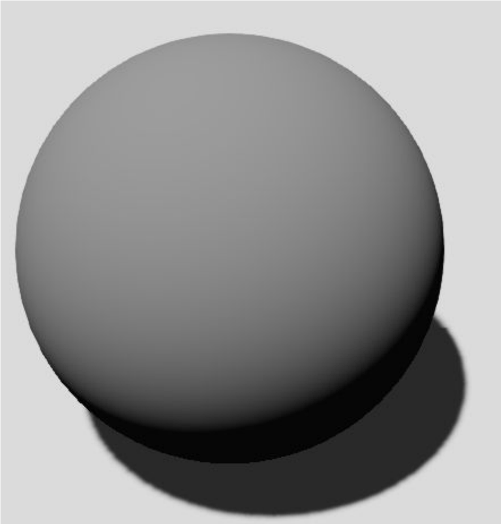


$$f_r(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)$$

In practice, most of the time we do not account the full BRDF but a simple combination of diffuse and specular reflection models

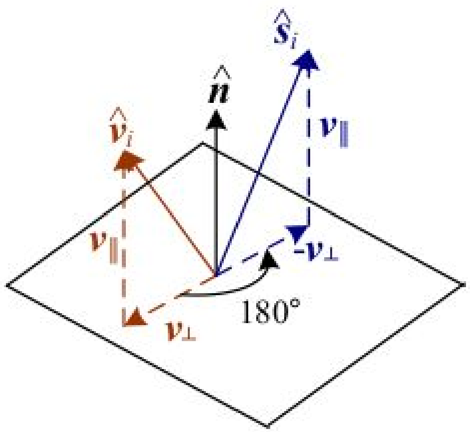
Diffuse reflection





Diffuse reflection scatters light uniformly in all directions. The amount of light observed depends on the angle between the incident light direction and surface normal (surface area exposed to a given amount of light become larger at oblique angles, so less light is observed at each point)

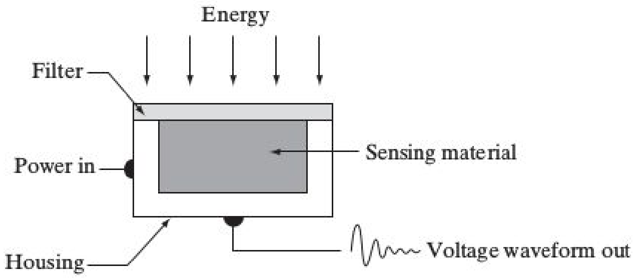
Specular reflection

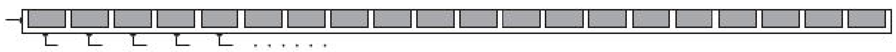


Specular reflection depends strongly on the direction of the outgoing light (relative angle between the view direction and the specular direction $\hat{s}_i$)

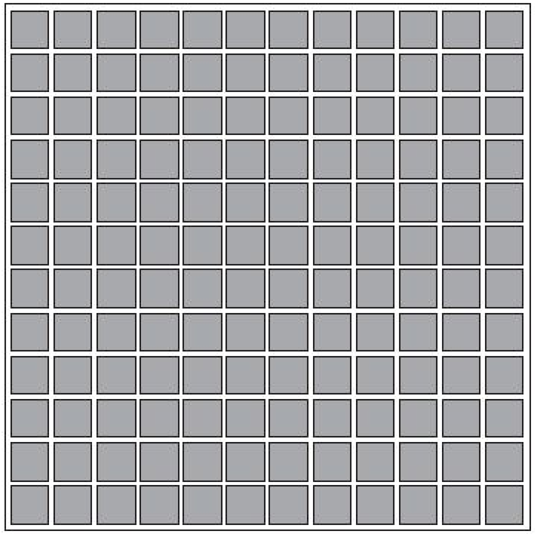
Image sensing

Incoming light radiation reflected from a 3D scene is transformed to a voltage by an imaging sensor that is sensible to a specific type of energy (wavelength).



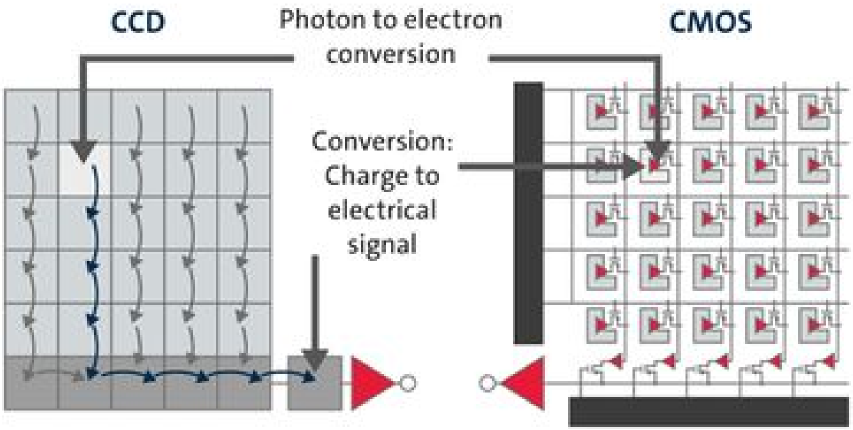


Usually, sensors are arranged in linear or 2D arrays



Università Ca' Foscari Venezia

CCD vs CMOS

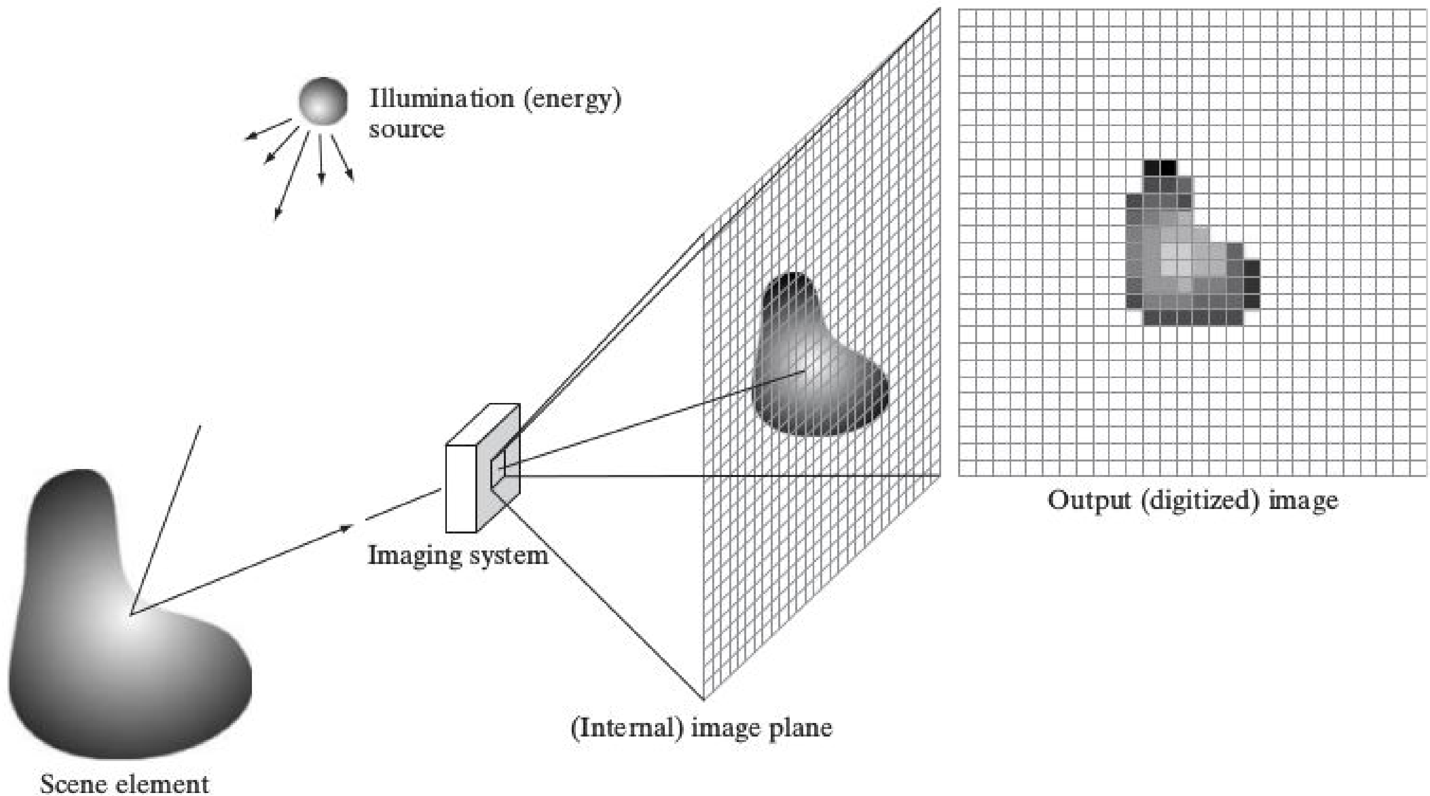


Two different principles: CCD vs CMOS

A/D conversion from voltage to a digital signal can happen in 2 ways: At the end of each row/column (CCD) or directly at each sensing cell (CMOS).

As a rule of thumb CMOS is in general better but CCD is faster and works better in low light

Imaging process



The sensor array, coincident with the focal plane of the lens, produces outputs proportional to the integral of the light received at each sensor for a certain amount of time

Imaging function

An image can be modelled as a function:

$$I : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$$

The domain is a (usually rectangular) subset of the real image plane

$$I(x, y)$$

Is proportional to the amount of light energy that is collected at the image plane at coordinates x,y. The amount of energy is called **intensity.**

Sampling and Quantization

A continuous real image is converted into a digital one through a process of **sampling** and **quantization**

Sampling:

Reduces the image domain to a finite set of spatial coordinates

$$\mathbb{R}^2 \rightarrow M \times N$$

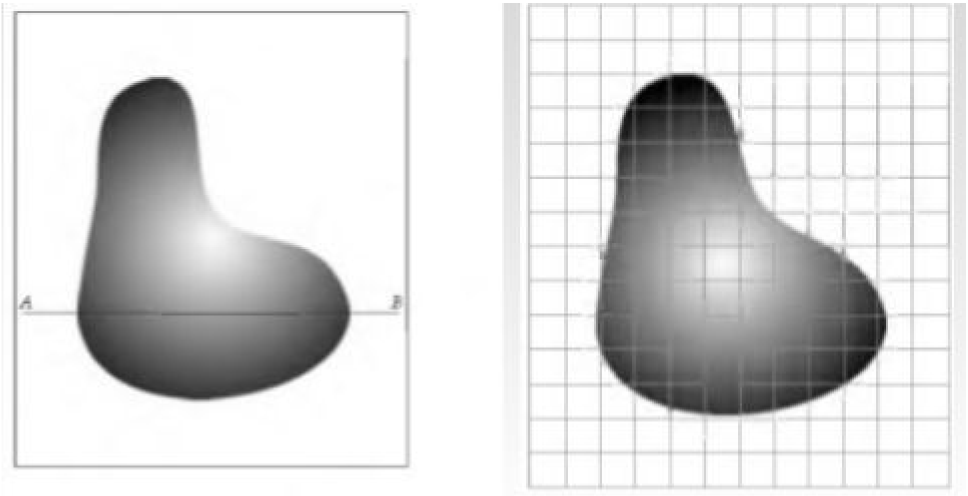
Quantization:

Reduces the sensor response (function codomain) to a finite set of values

$$\mathbb{R} \rightarrow [0, 1 \ldots 2^b]$$

Università
Ca’ Foscari
Venezia

Sampling



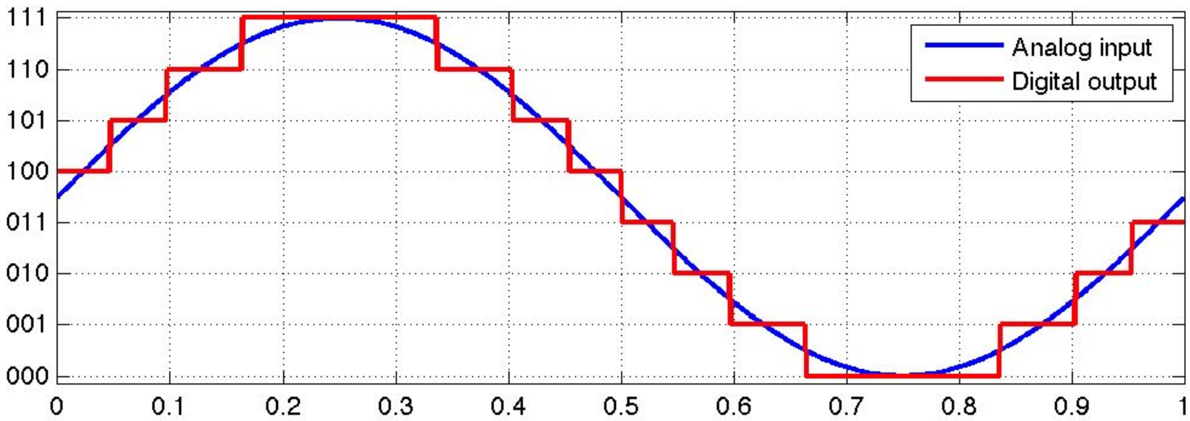
To sampling the continuous image domain into a finite set of values we usually consider an equispaced grid of values in a given area (matrix). This reflect the regular arrangement of cells in a CMOS or CCD sensor.

Each sample is called a **pixel**

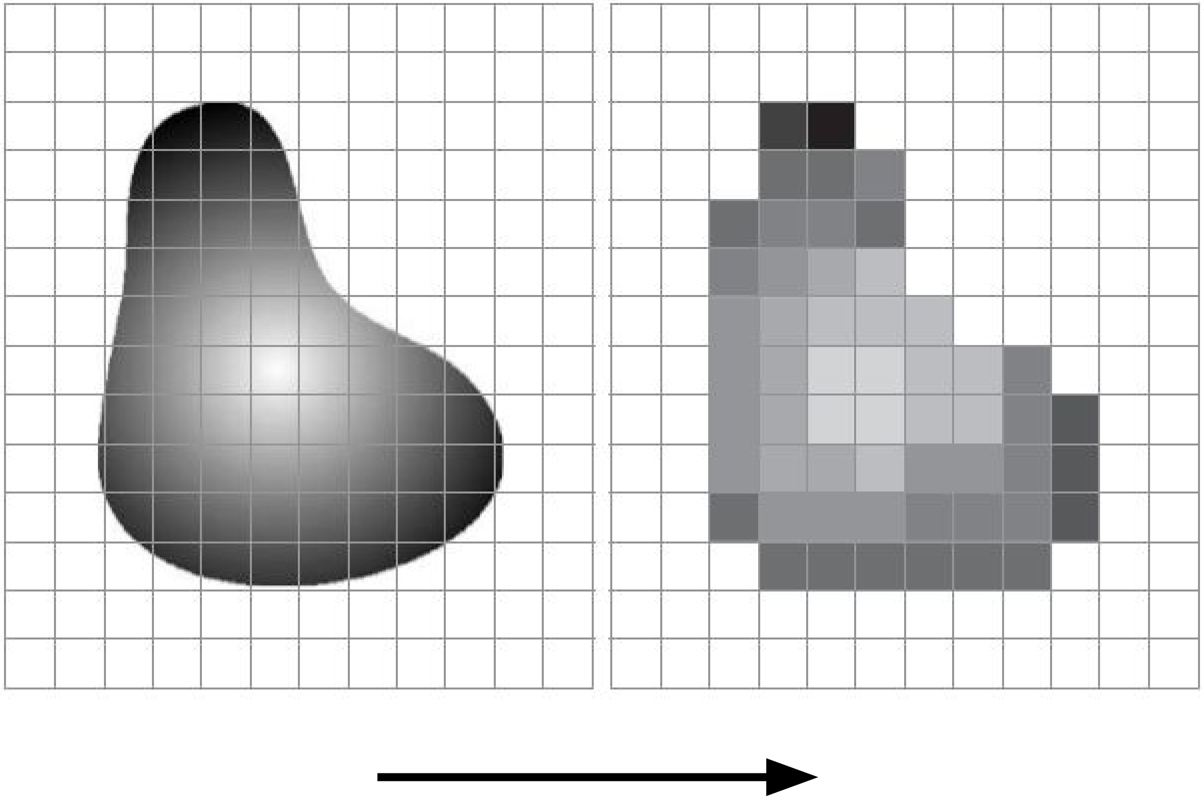
Quantization

The intensity (output) of the function must also be discretized (quantized) in a finite set of values to be digitized and used in a computer.

Image codomain is divided into a set of values and each f(x,y) is rounded to the closest one



Sampling and Quantization



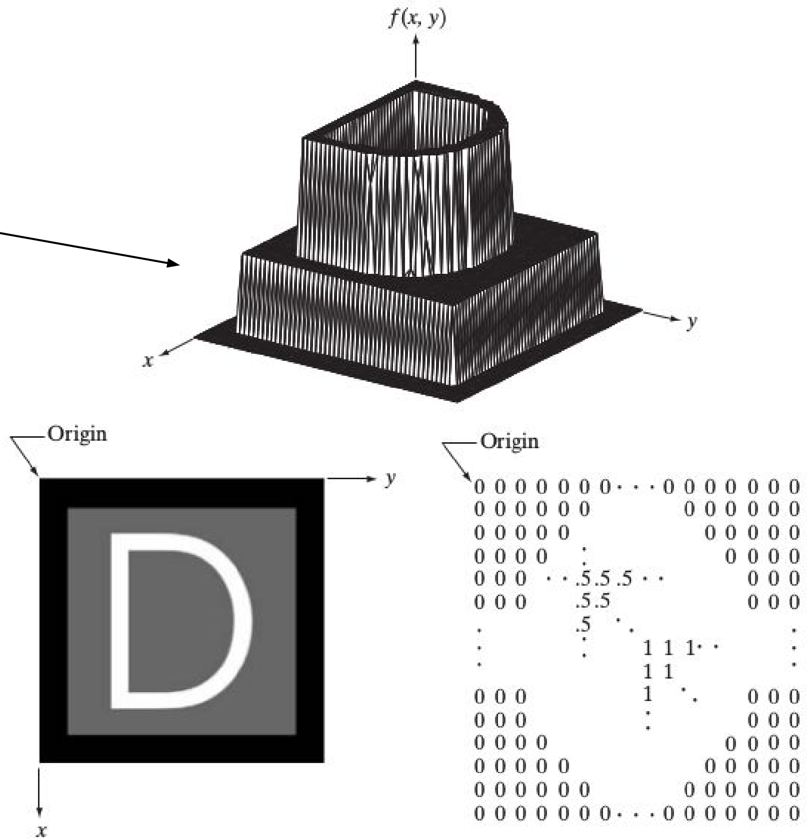
Digital images

The continuous image function is converted to a digital image after sampling and quantization. We can represent a digital image in the following ways:

As a surface

As a visual intensity array

As a Matrix



Digital Images

Intensity and Matrix representation are the most common to be used in practice.

The digitization process requires that decisions be made regarding the matrix size M,N and the number L (usually a power of 2) of quantized intensity levels.

M,N are related to the **spatial resolution** of an image

L is related to the **intensity resolution**

Spatial resolution

Spatial resolution is a measure of the smallest discernible detail in an image

Usually defined as dots (pixels) per unit distance

NOTE: to be meaningful, measures of spatial resolution must be stated with respect to spatial units.

A 1024x1024 pixel image is not meaningful without stating the spatial dimensions encompassed by the image (the size of each pixel in the 3D world)

Università
Ca' Foscari
Venezia

Spatial resolution

a. Landsat ETM+ b. ATLAS c. QuickBird

The number of pixels of an image directly affect spatial resolution only with comparable lenses and if the subjects are taken at the same distance

Spatial resolution

1250 dpi

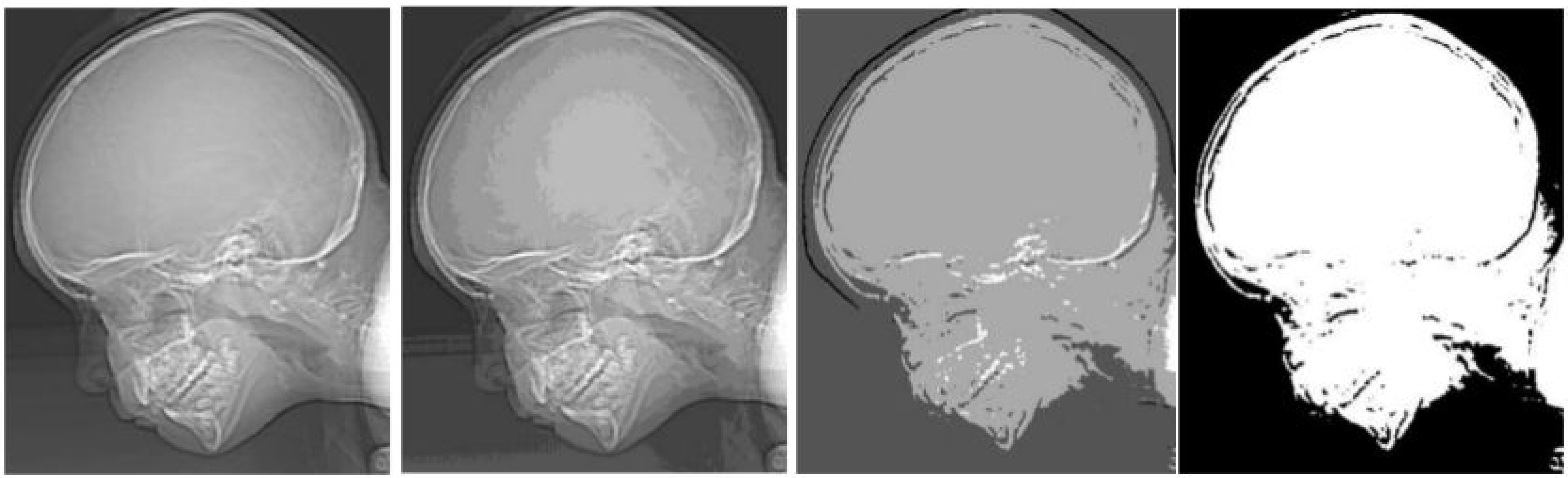
300 dpi

72 dpi

Intensity resolution

Intensity resolution refers to the smallest discernible change in intensity level.

Usually it is a power of 2 (8 bits most common)



256 levels 16 levels 4 levels 2 levels

Università Ca' Foscari Venezia

Dynamic Range/Contrast

We define the **dynamic range** of an imaging system as the *ratio* of the maximum measurable intensity to the minimum detectable intensity level.

- Depends by the number of bits we use and establish the lowest and highest intensity levels that a system can represent

Similarly, with **contrast** we define the *difference* between the highest and the lowest intensity level of an image.

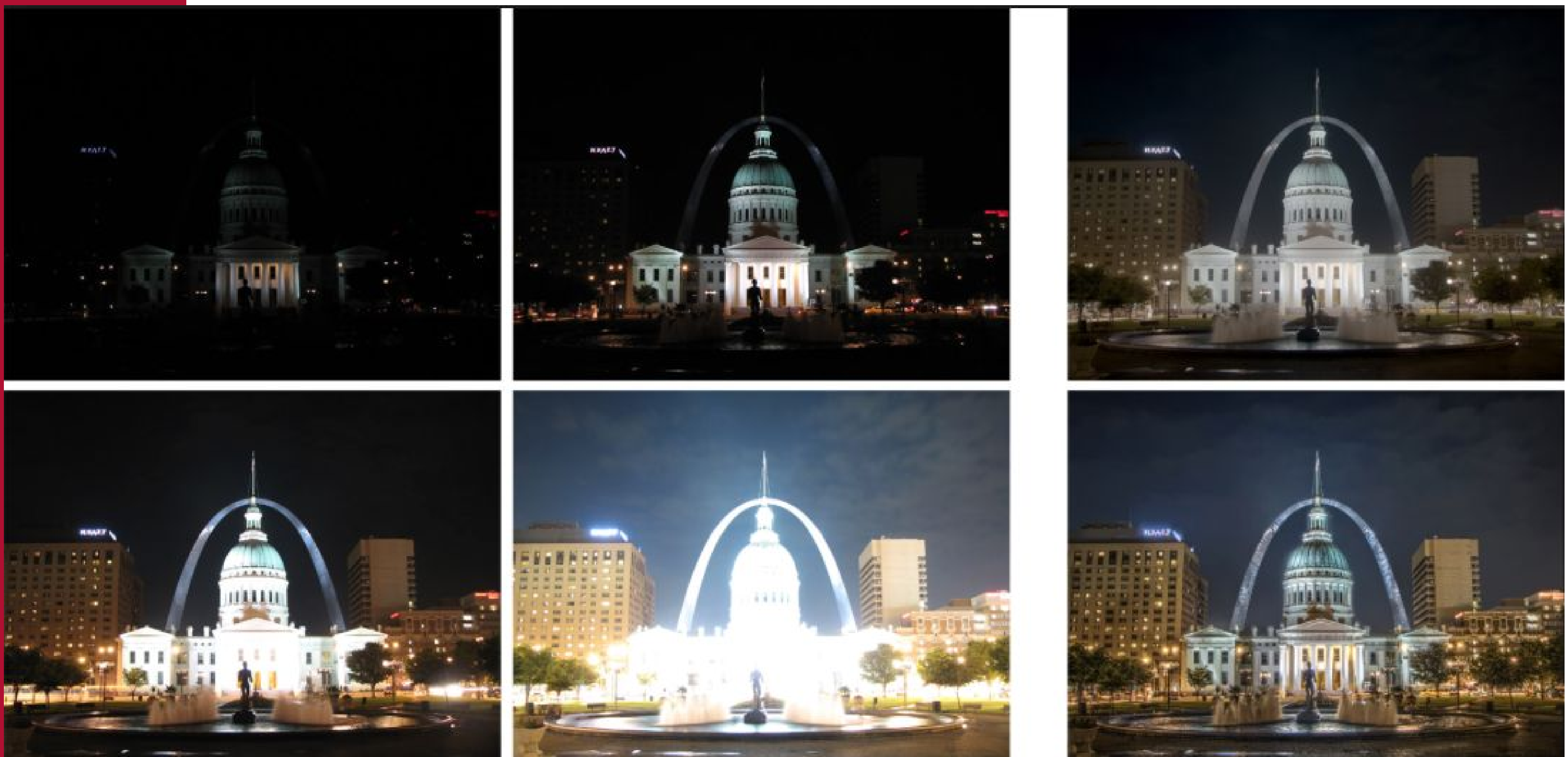
Dynamic Range/Contrast

Human eye has a dynamic range of about 10^9!

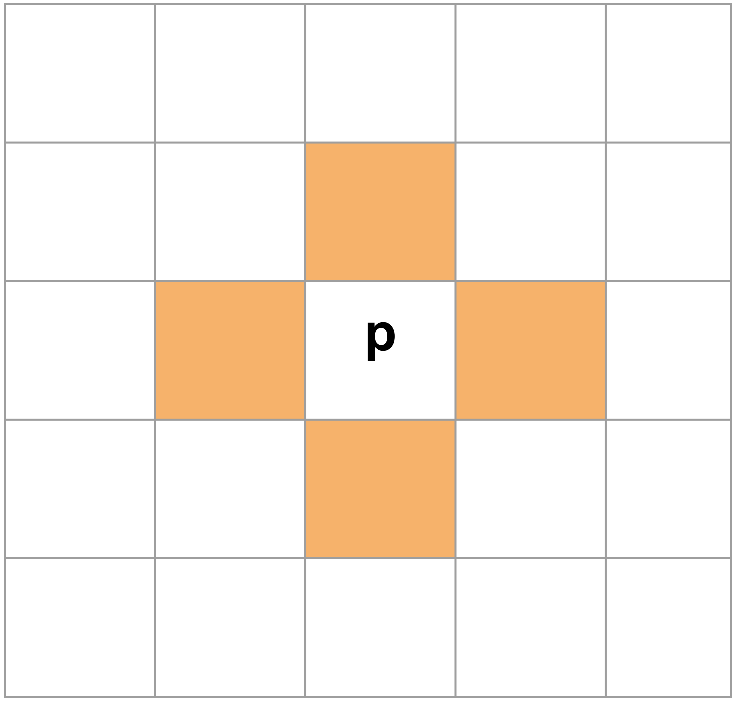
Even if we use techniques to acquire images at high dynamic range we will probably won't be able to display it on a computer screen...

Compromises have to be done in the acquisition process to either discard dark details or saturate heavily illuminated areas

Dynamic Range/Contrast



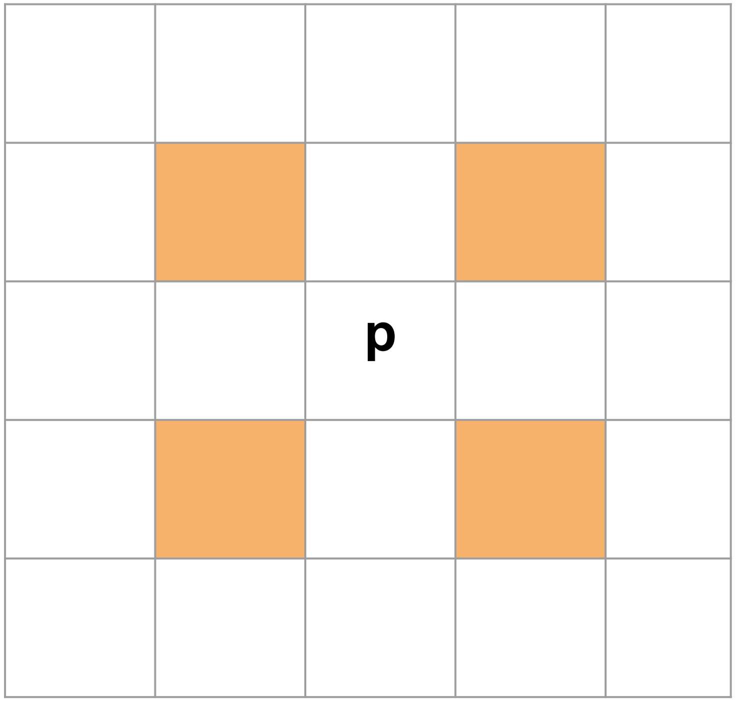
Relationships between pixels



A pixel p at coordinates (x,y) has 2 horizontal and 2 vertical neighbours called *4-neighbours of p*.

$$N_4(p) = \{(x+1, y), (x-1, y), (x, y+1), (x, y-1)\}$$

Relationships between pixels

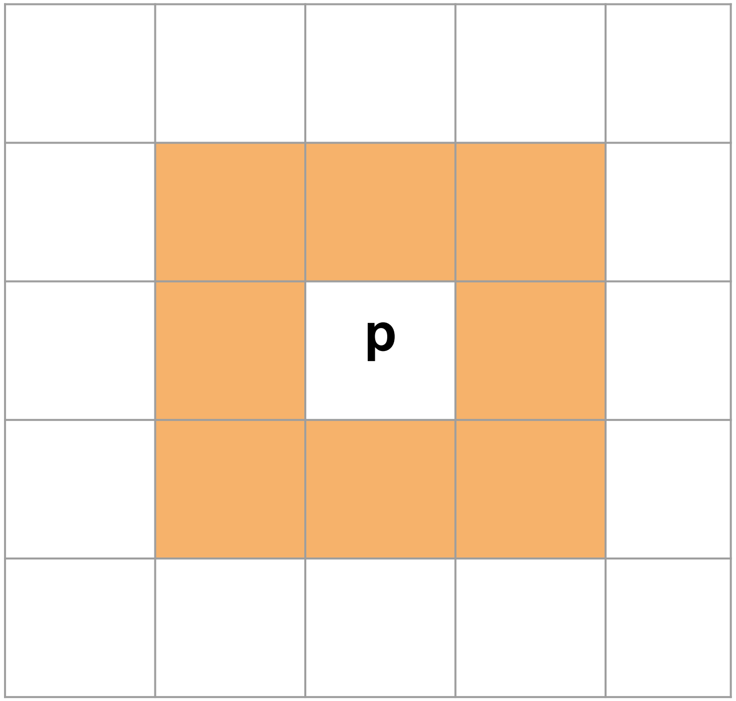


A pixel p at coordinates (x,y) has also 4 diagonal neighbours:

$$N_d(p) = \{(x+1, y+1), (x-1, y-1), (x-1, y+1), (x+1, y-1)\}$$

Università Ca' Foscari Venezia

Relationships between pixels



All the neighbours of p are called 8-neighbours

$$N_4(p) \cup N_d(p) = N_8(p)$$

Adjacency

Let V be the set of intensity values used to define adjacency

4-adjacency: p,q with values from V are 4-adjacent if

$$q \in N_4(p)$$

8-adjacency: p,q with values from V are 8-adjacent if

$$q \in N_8(p)$$

M-adjacency: p,q with values from V are m-adjacent if:

$$q \in N_4(p)$$

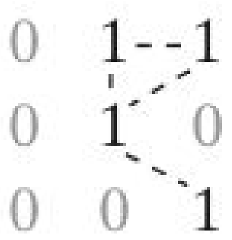
Or: $q \in N_D(p)$ and $N_4(p) \cap N_4(q)$ has no pixels with values from V

Università
Ca’ Foscari
Venezia

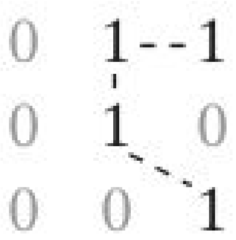
Adjacency

$$V = \{1\}$$

$$\begin{array}{ccc} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}$$



8-adjacency



m-adjacency

Università Ca' Foscari Venezia

Path

A path from a pixel p=(x,y) to q=(s,t) is a sequence of pixels with coordinates

$$(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$$

Where

$$(x_0, y_0) = (x, y), (x_n, y_n) = (s, t)$$

$$\forall 1 \leq i \leq n, \text{ pixels } (x_i, y_i) \text{ and } (x_{i-1}, y_{i-1}) \text{ are adjacent}$$

Connected Component

Two pixels p and q are **connected** if there exist a path between them.

Given a pixel p, the set of all pixels connected to p is called **connected component**

Let R be a subset of pixels in an image. R is called a **region** if it contains only one connected component