Principles of image acquisition and 3D sensing

Sistemi di Elaborazione Accelerata, Modulo 1

A.A. 2025/2026

Stefano Mattoccia

Università di Bologna

Introduction

Cameras are widespread sensing devices used in countless application fields, such as robotics, augmented reality, and autonomous driving, to name a few.

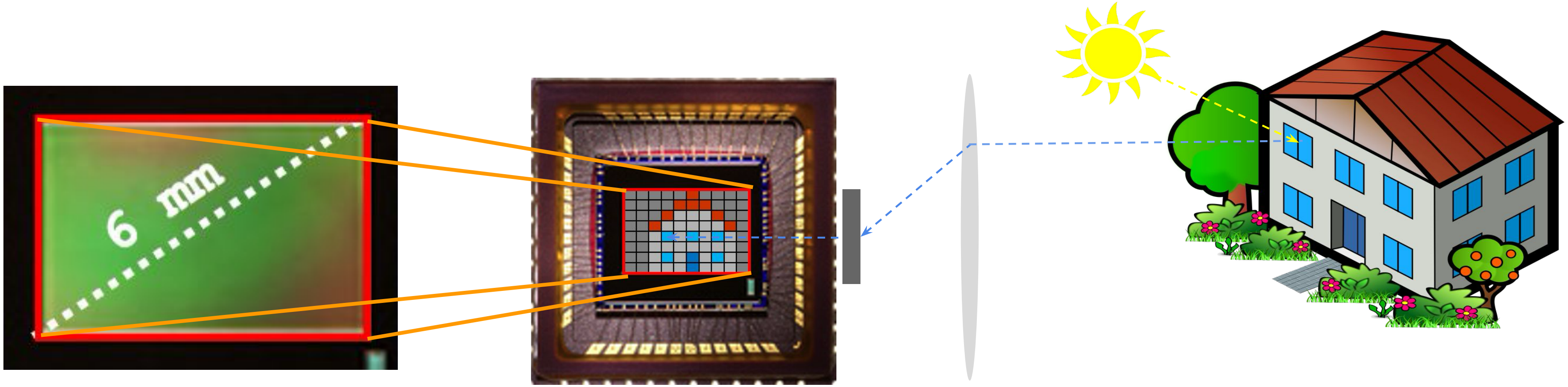
Moreover, the massive processing required by image-based algorithms and networks, in most cases, demands parallel computation techniques.
This issue is further exacerbated when the *framerate* (i.e., the number of images acquired for each second) gets higher (e.g., 30+ frames per second (FPS)).

Considering the broad deployment of cameras as ubiquitous sensing devices, this lecture introduces the basic principles of image acquisition and other emerging devices

Camera: how this sensing device works

Cameras are popular sensing devices:

- It is somehow an *analog* (actual scene) to *digital* (image) converter
- For each pixel, it gathers through the lens the light emitted by each single point in the scene and converts it into a discrete value, typically, one or more values [255,0]
- Each pixel corresponds to a photosensitive area on a tiny image sensing device
- For instance, the Aptina/ON 1/3” below has a diagonal of about ≂ 6 mm



Camera: image and pixels

A **pixel**, at coordinate (x,y), consists of one (*monochrome*), three (*color*) or more (*multi-spectral*) 8 bit values

https://www.pexels.com/photo/red-panda-perching-on-tree-during-daytime-146102/

Camera: Rolling shutter vs Global shutter acquisition

The acquisition phase can occur according to two technologies/approaches:

- *Rolling shutter*: progressively, from the leftmost top pixel to the rightmost bottom one, generating artefacts in the presence of moving objects and/or *ego-motion*
- *Global shutter*; simultaneously for all pixels
- The latter is now mainstream, while rolling shutter is less popular yet still used

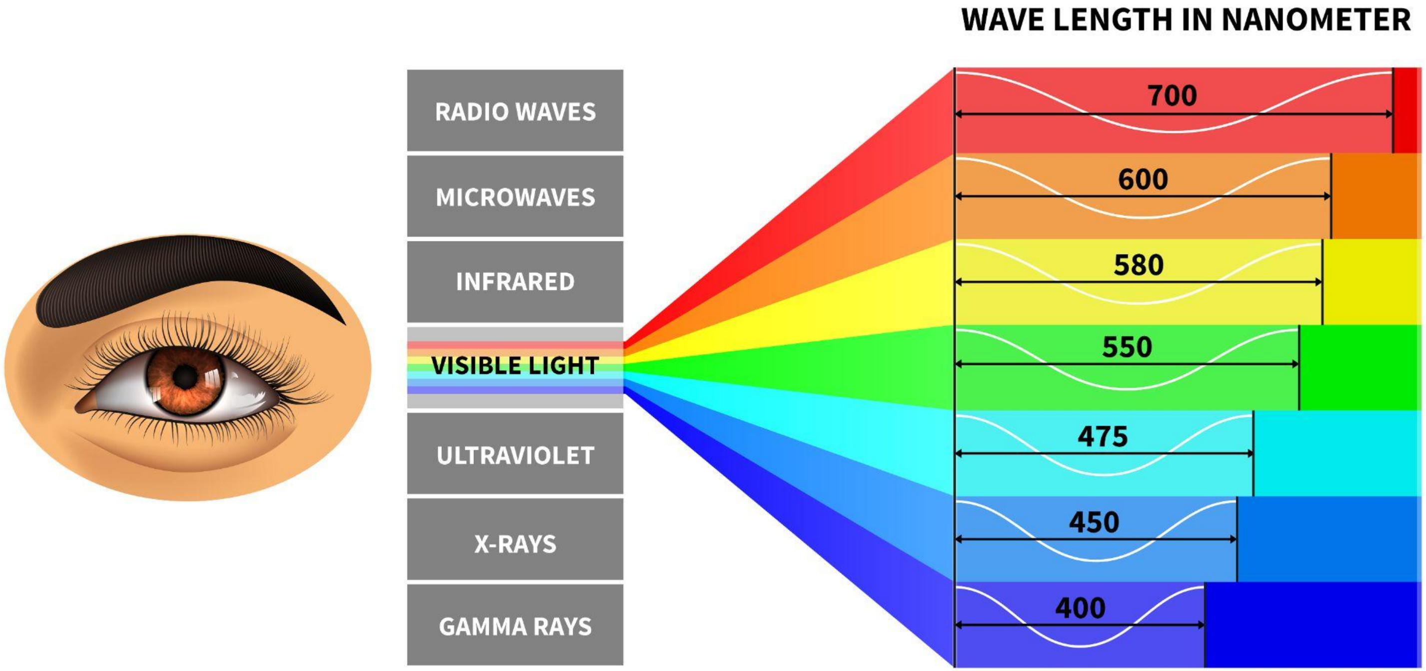
Rolling shutter

Global shutter

https://www.flir.com/

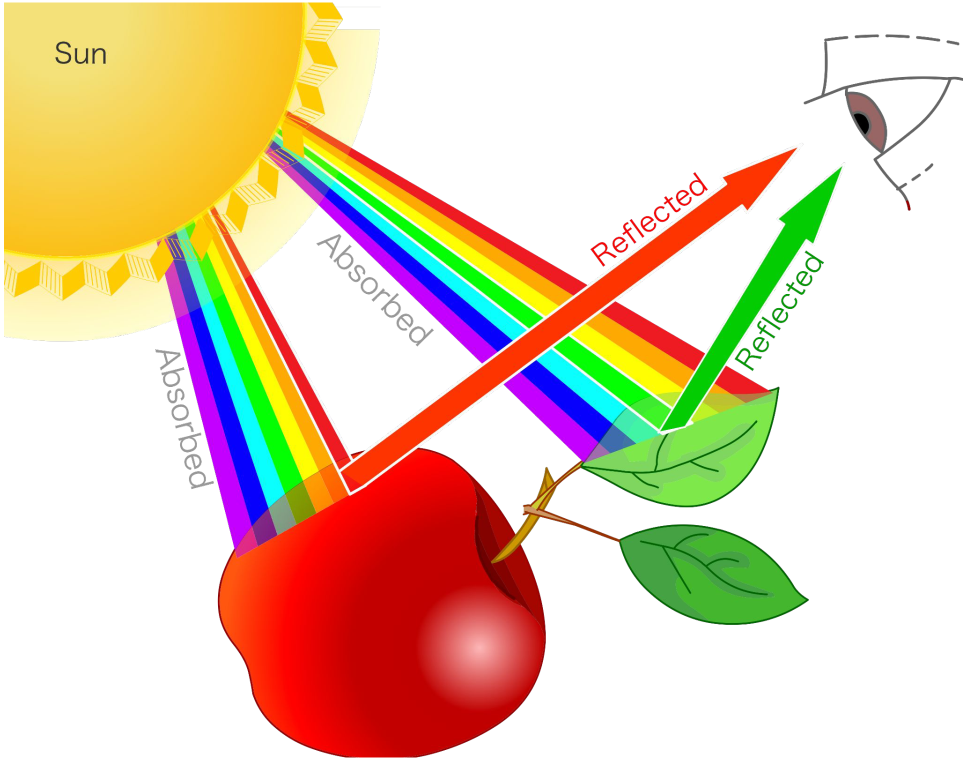
Camera: the visible spectrum

- Electromagnetic signals have a broad spectrum
- However, our eyes can perceive only a tiny subpart of it called the visible spectrum
- Specifically, the visible spectrum ranges from 400 to 700 nanometer



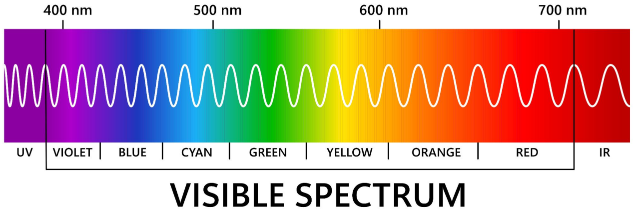
Camera: Perception in the visible spectrum

- The color perceived by our eyes is only the component of the visible spectrum reflected by a specific material (eg, red for the apple, green for the leaf)



Camera: Greyscale and Color image formation

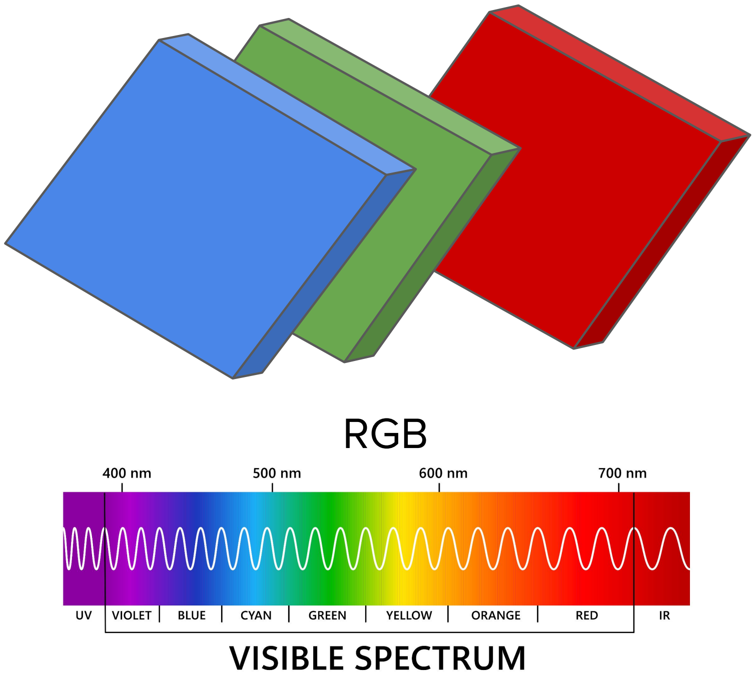
Image sensors can gather through their photosensitive devices all light components (*greyscale*, or *monochrome*) or some of its bands (3 for color, 4+ for *multispectral* devices)

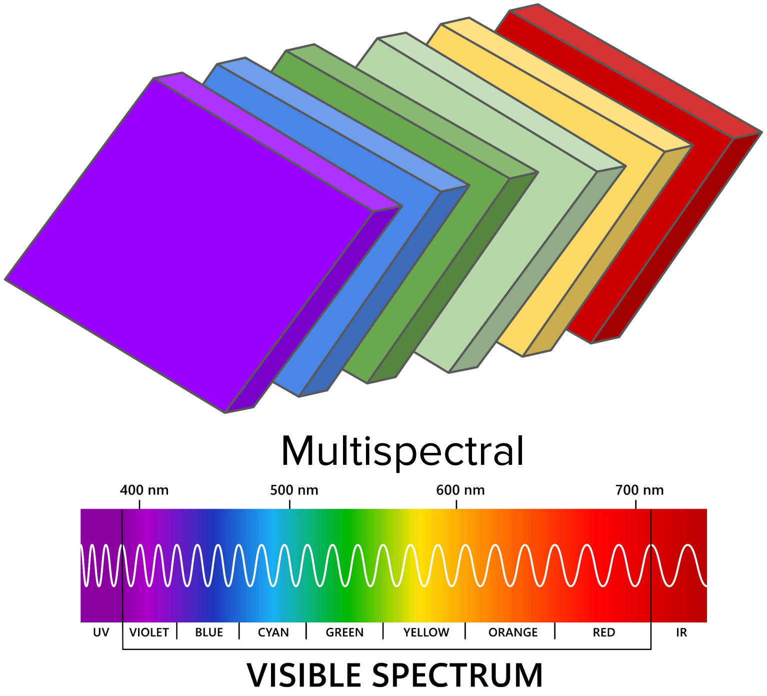


http://digital.ni.com/public.nsf/allkb/8AAAE9435D30CEB886256F80007A39AD

Camera: Color and Multispectral image formation

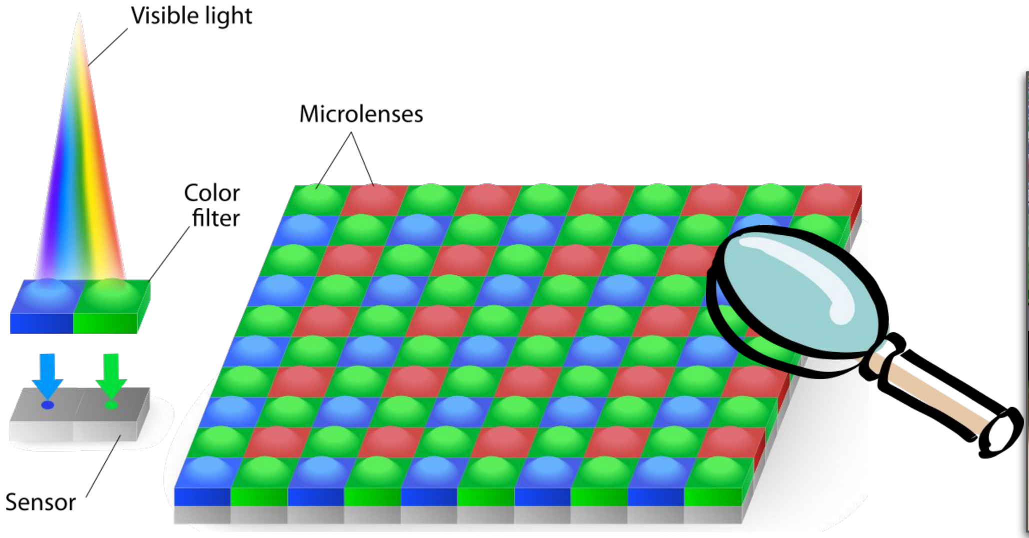
- A multispectral sensing device can gather much more information from the sensed scene than a conventional color camera
- Helpful in many fields such as agriculture, medical imaging, quality inspection, etc

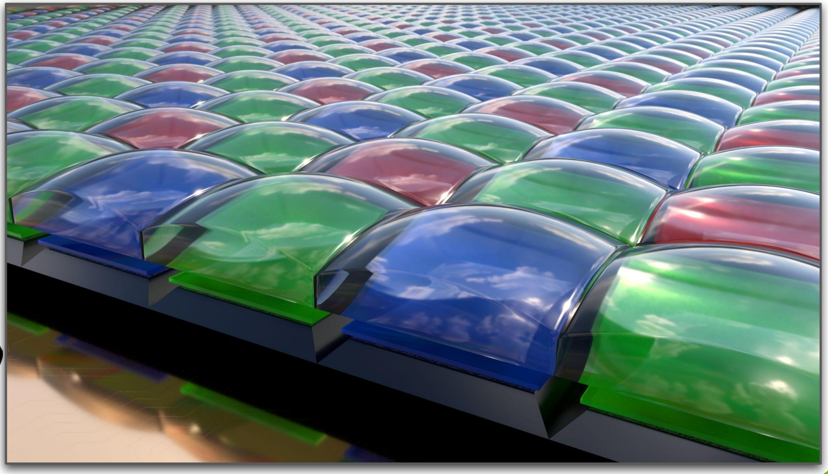




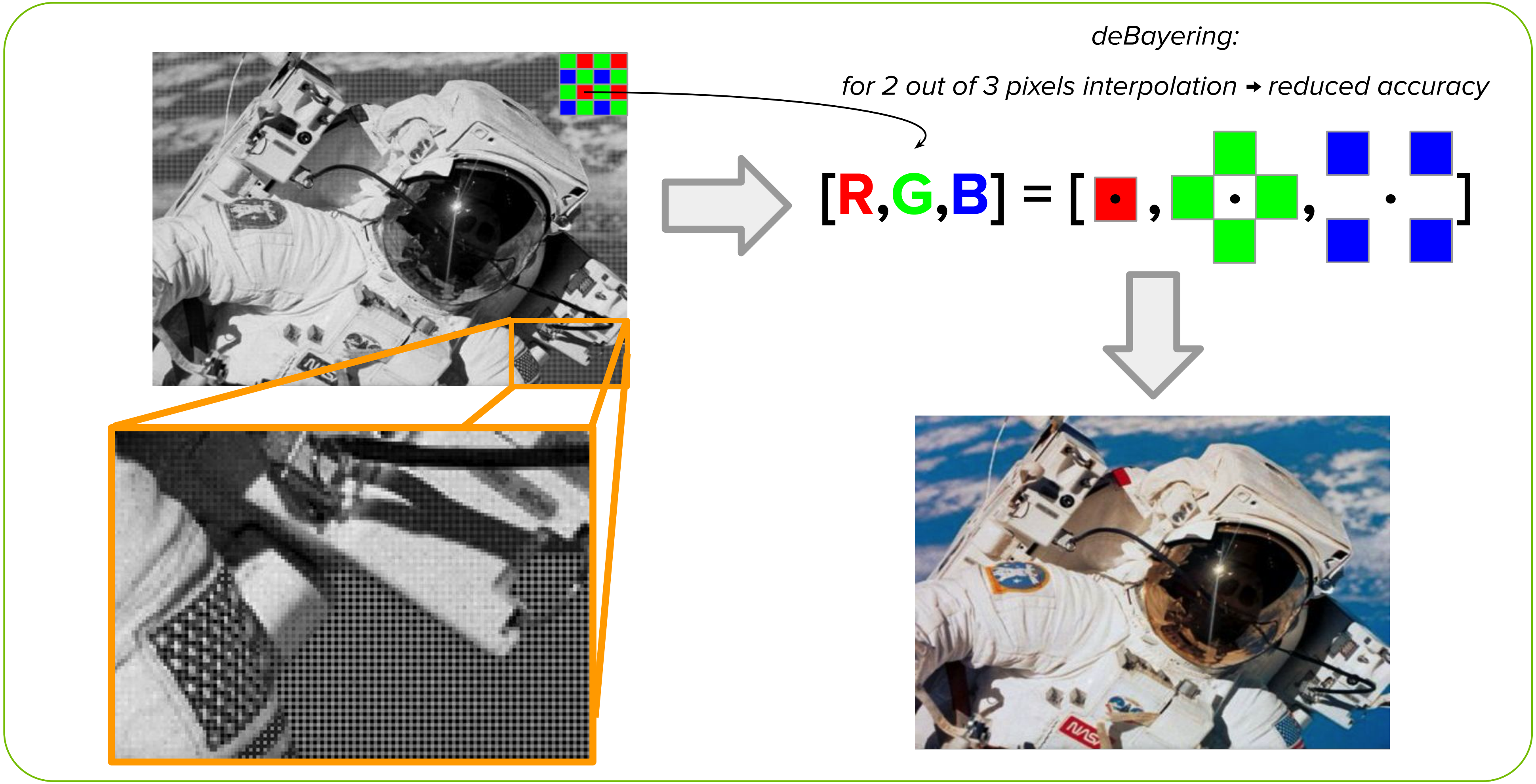
Camera: Bayer filter to sense RGB bands

- Rather than using a single sensing device for each band, in most cases, a single image sensor with microlens is used to gather (approximate) information from multiple portions of the spectrum using color filters for Red, Green and Blue
- The Bayer filter relies on this strategy; hence, 2 out of 3 components of each pixel are always obtained through interpolation of neighbours points with a cyclic scheme



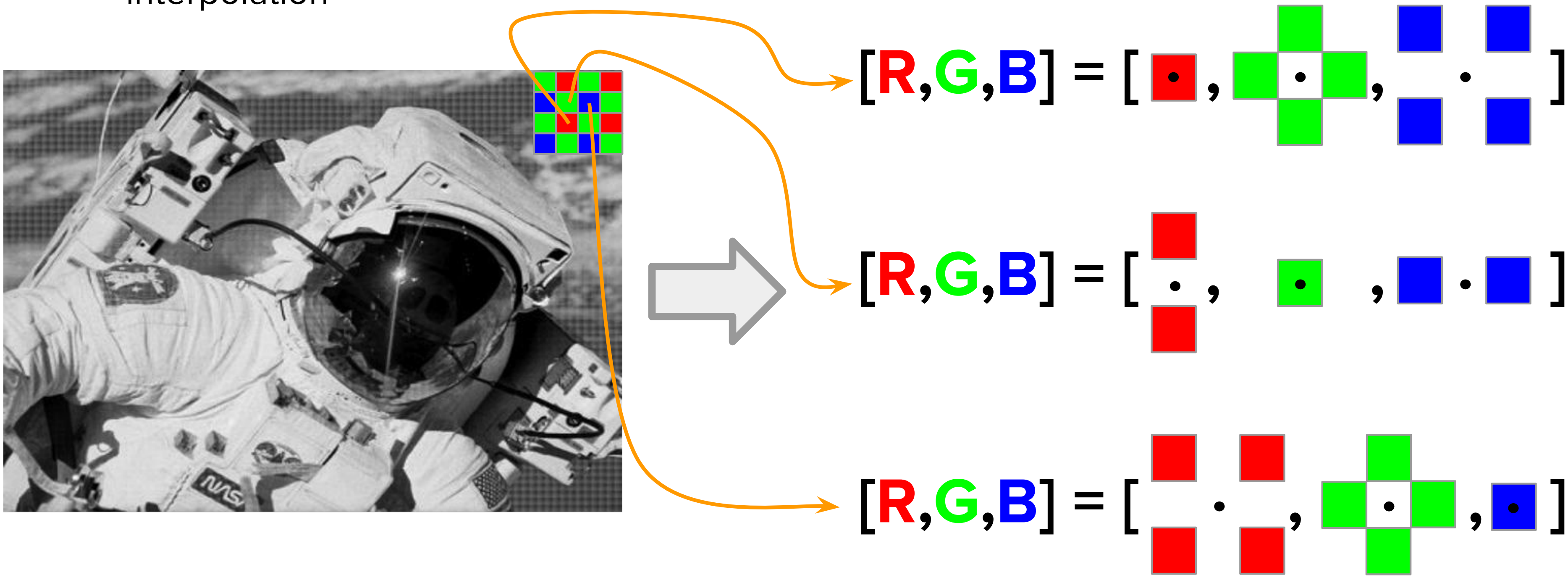


Camera: Bayer filter and conversion from raw data to RGB



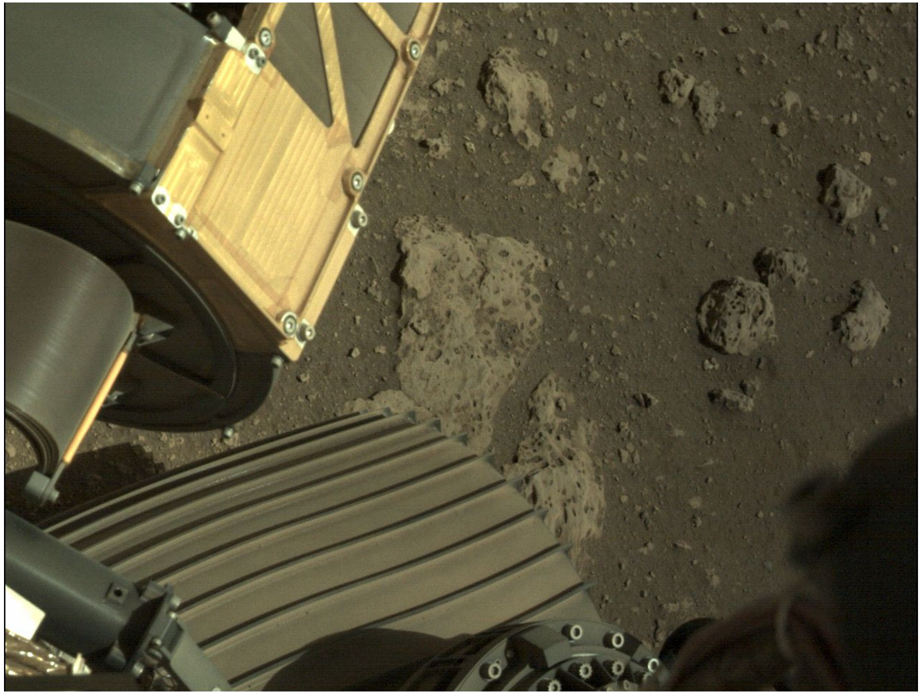
Camera: Bayer filter and interpolation (ie, deBayering)

- Below, the three recurrent schemes to infer RGB for each pixel through deBayering
- One of the three RGB values is actually sensed, the other two are obtained through interpolation



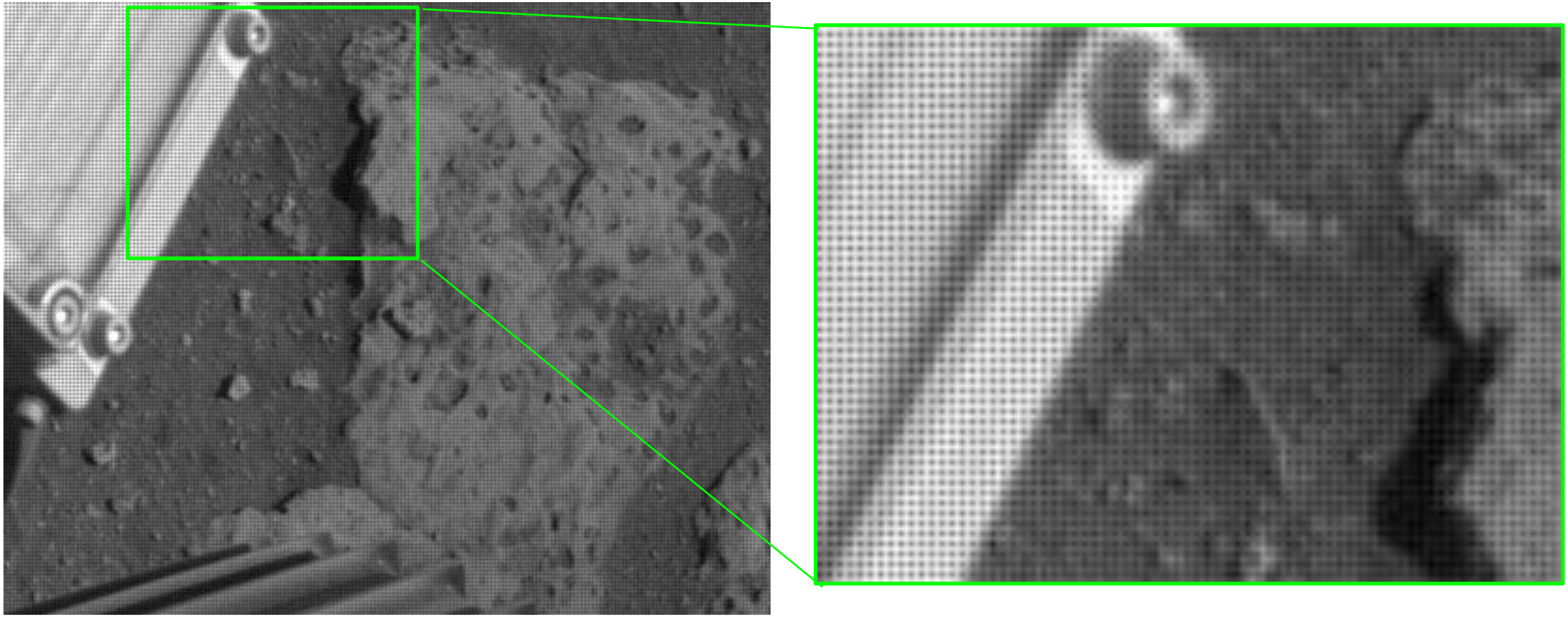
Camera: Bayer image example 1/2

- Example of images acquired by rover NASA Perseverance on Mars
- Many images available at: https://mars.nasa.gov/mars2020/multimedia/raw-images/



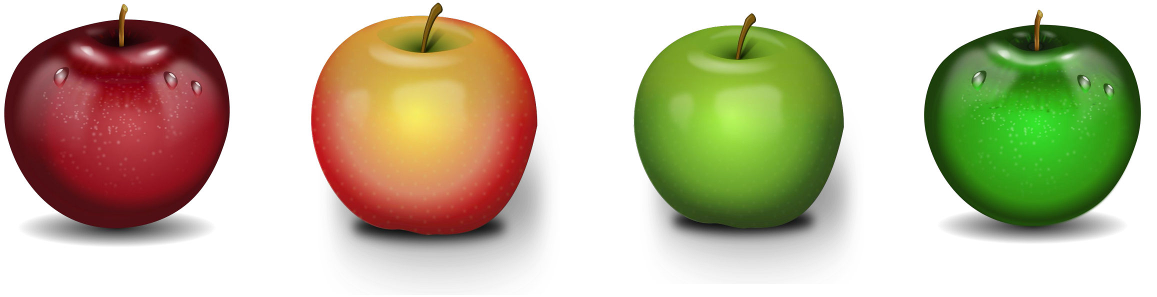
Camera: Bayer image example 2/2

- Zooming allows to better perceive the effect of the Bayer pattern onto the raw image

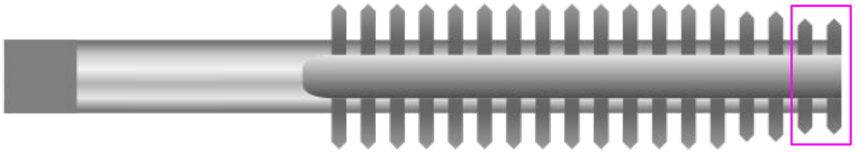


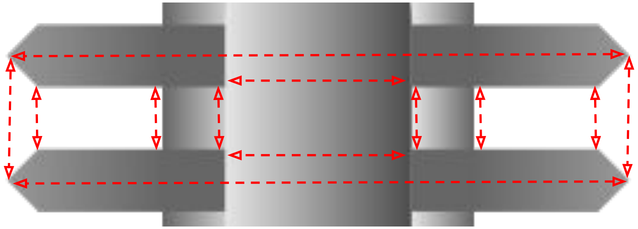
Camera: color or monochrome images?

In many applications, the color cue is crucial, for instance, when sorting items like fruit or vegetables according to their ripeness



On the other hand, color is obtained through interpolation; hence, for applications like *metrology,* a monochrome camera might be a more appropriate choice

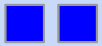


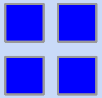


Exercise: DeBayering

Exercise: describe how to perform image deBayering (demosaicing) using only integer data and arithmetic/logic operations.

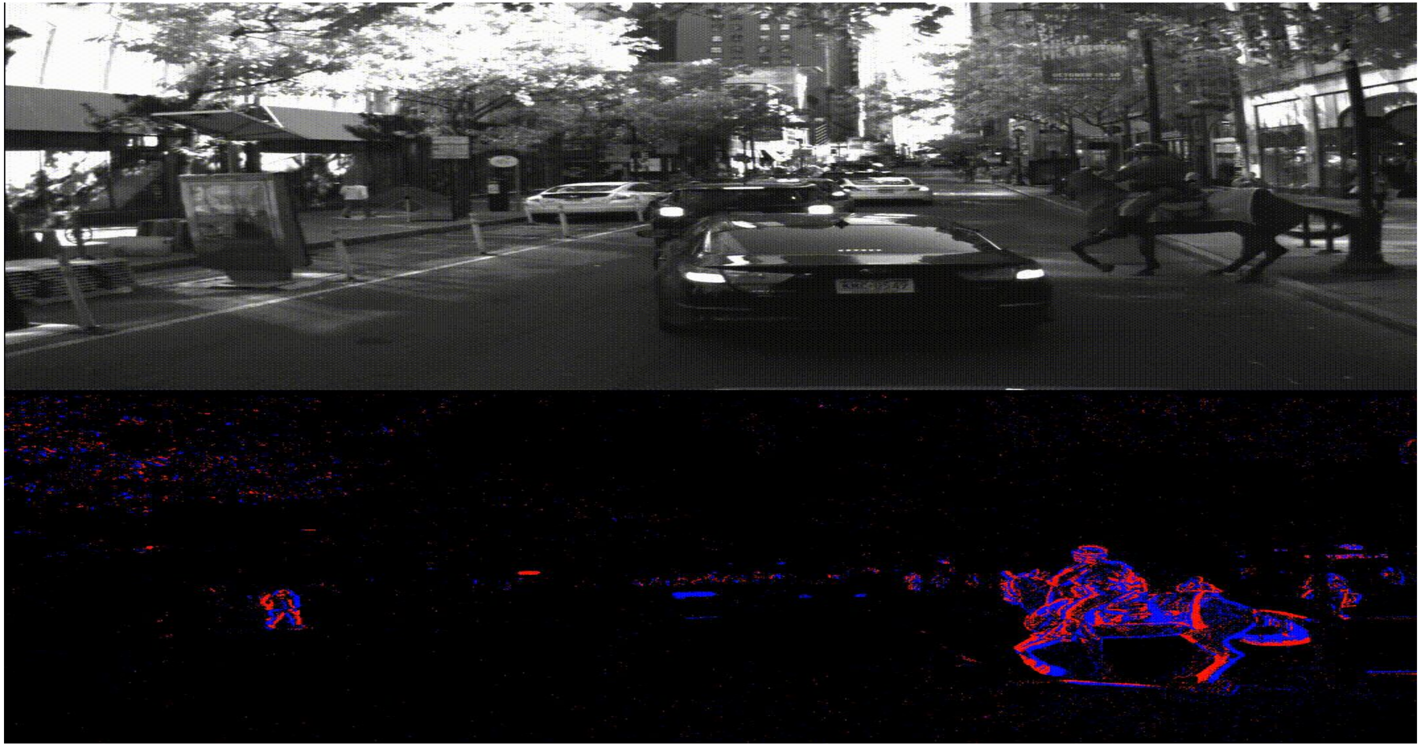
For this purpose, we need to average two or four pixels, each in the range [255,0]:

i) two pixels: 67, 188 

ii) four pixels: 67, 188, 245, 177 

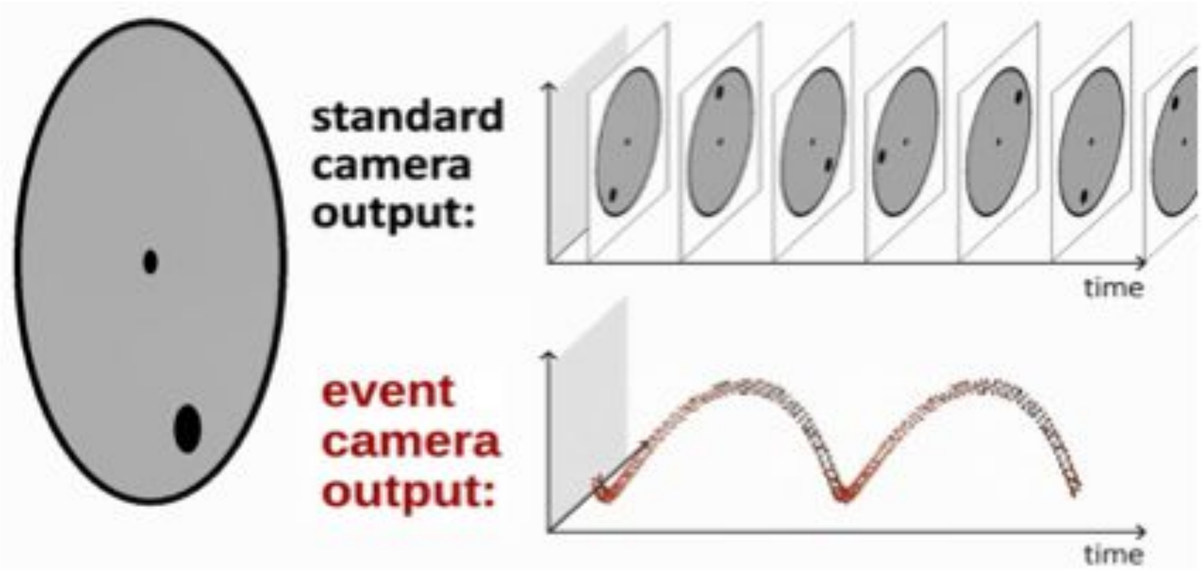
Event cameras 1/4

- In recent years, *event cameras* emerged as a novel image sensing paradigm
- These devices continuously capture brightness change in the sensed scene



Event cameras 2/4

- The different behaviour of conventional and event cameras is described by this video:



https://www.youtube.com/watch?v=LauQ6LWTkxM&t=30s

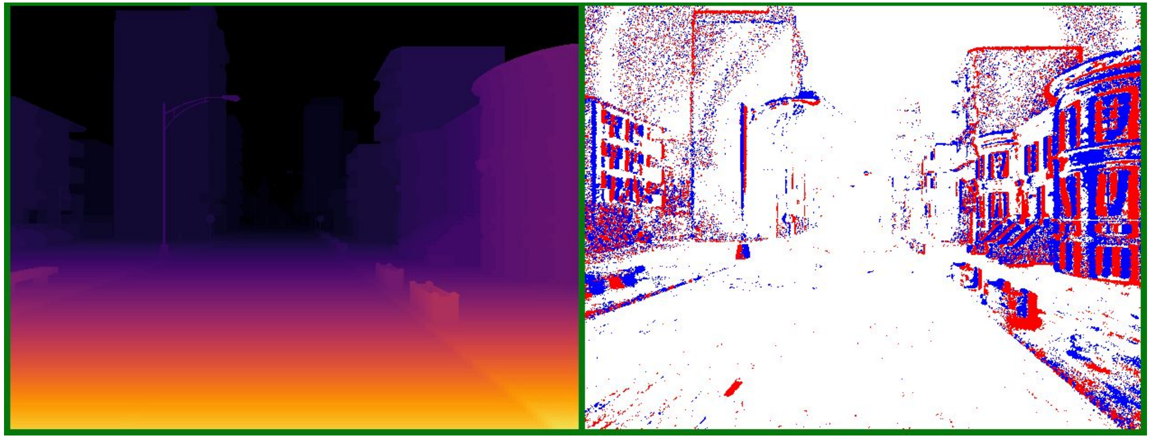
Event cameras 3/4

- The event stream consists of a binary information: brightness increase and decrease
- Compared to conventional imaging devices:
    - information is gathered and delivered at microsecond resolution
    - not affected by motion blur
    - much higher dynamic range
    - low power

https://rpg.ifi.uzh.ch/docs/scaramuzza/2019.07.11_Scaramuzza_Event_Cameras_Tutorial.pdf

Event cameras 4/4

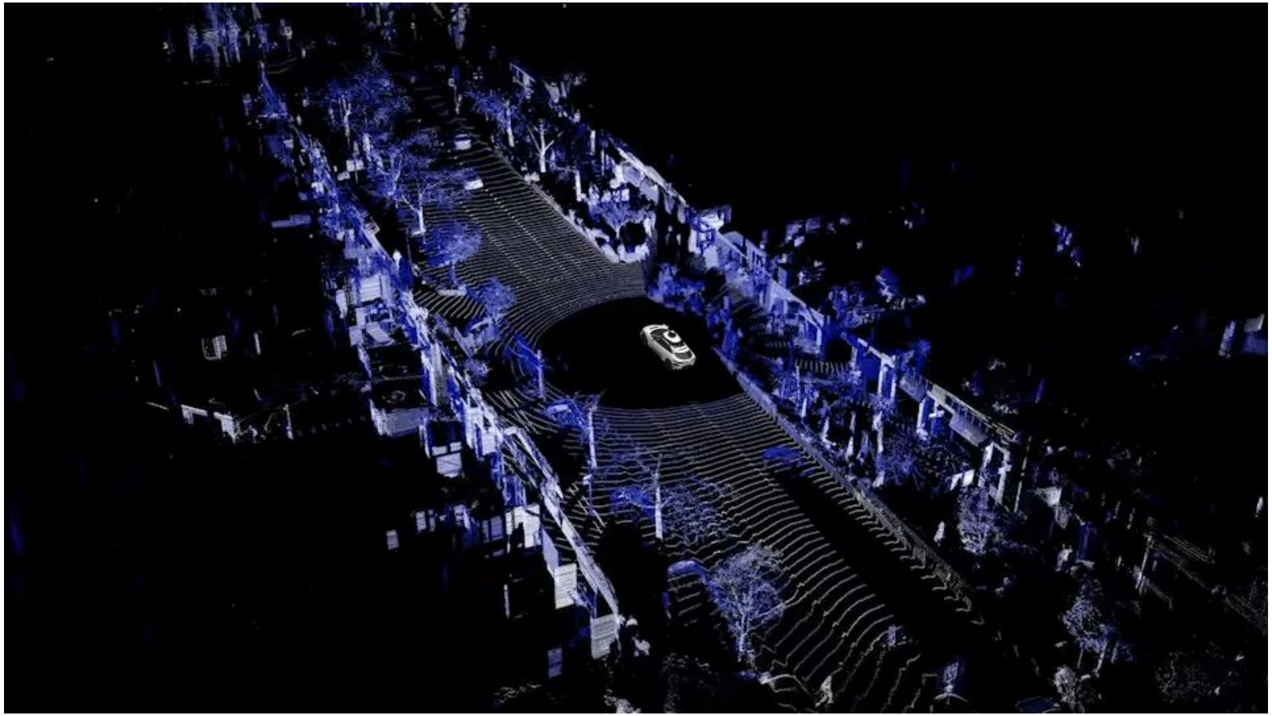
- Despite the much lower data density, learning based methods enabled to face several downstream tasks, such as depth perception and pose estimation, potentially at microsecond resolution



- The only constraints are:
  - the sparse data available vs traditional cameras
  - the computational requirements

3D sensing

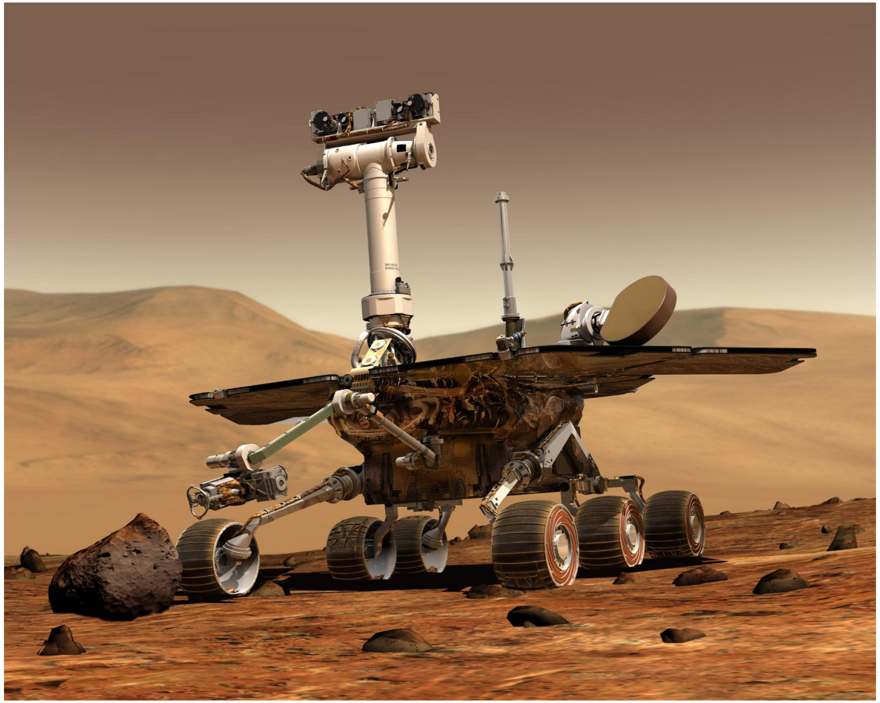
- 3D sensing aims at inferring the geometry of the sensed scene to obtain a *point cloud*
- It is pivotal for autonomous navigation, augmented reality and robotics, etc

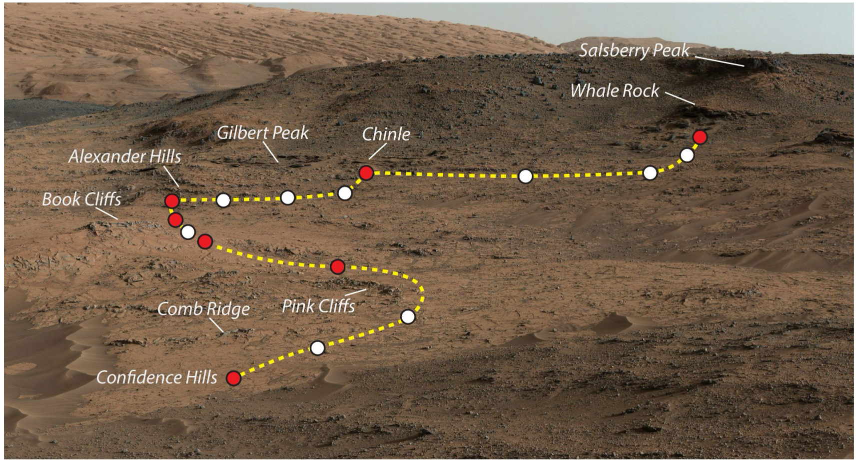


https://waymo.com/

3D sensing with cameras (standard and event)

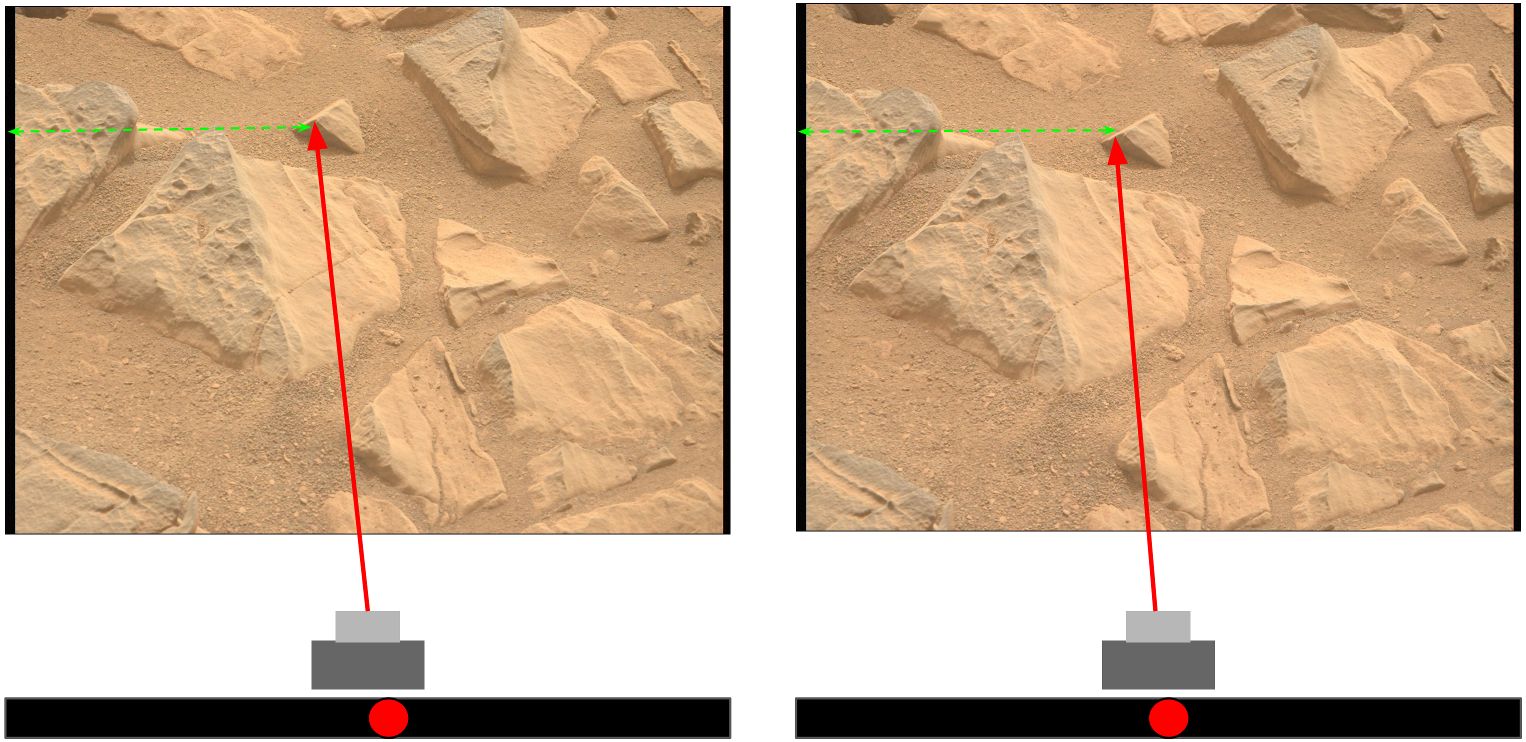
Conventional and RGB cameras enable 3D reconstruction through different strategies (eg, stereo vision, monocular depth perception, etc)





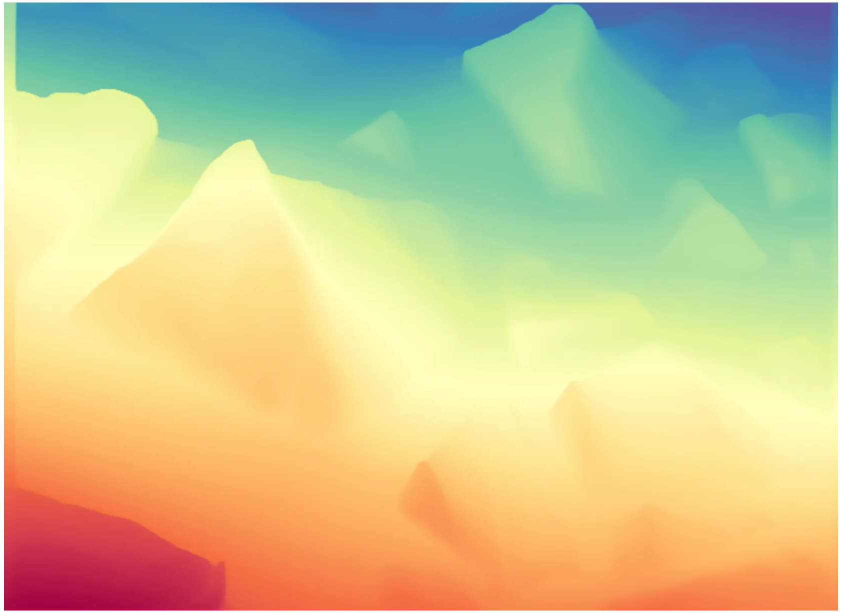
3D sensing through stereo vision 1/2

- Cameras enable 3D reconstruction through different strategies (eg, stereo vision)



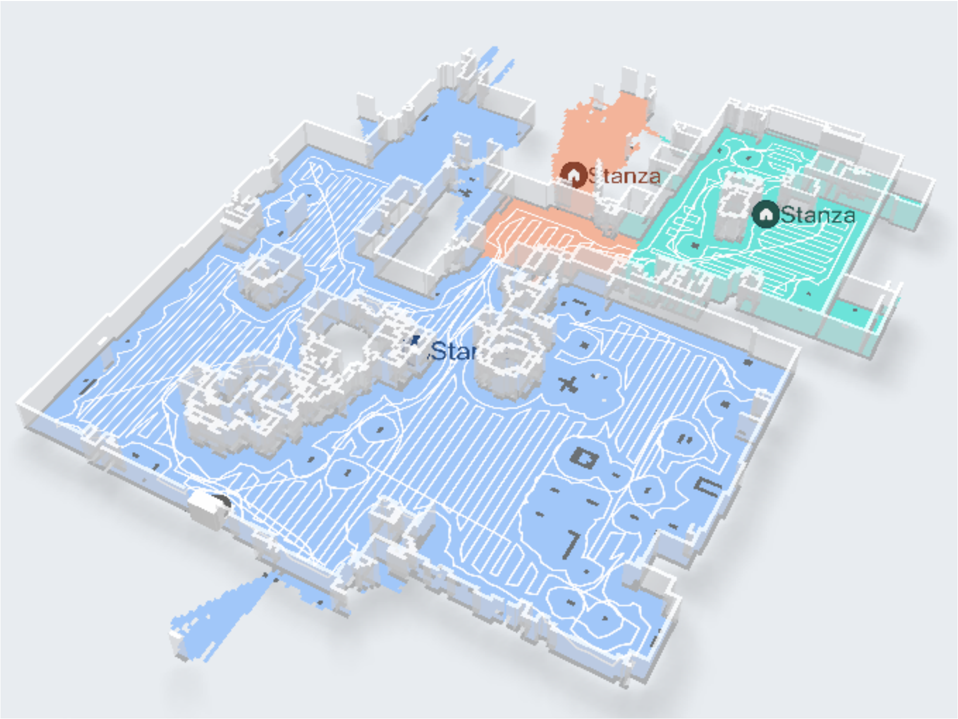
3D sensing through stereo vision 2/2

Once located corresponding points, a point cloud can be obtained from the disparity map shown at the right



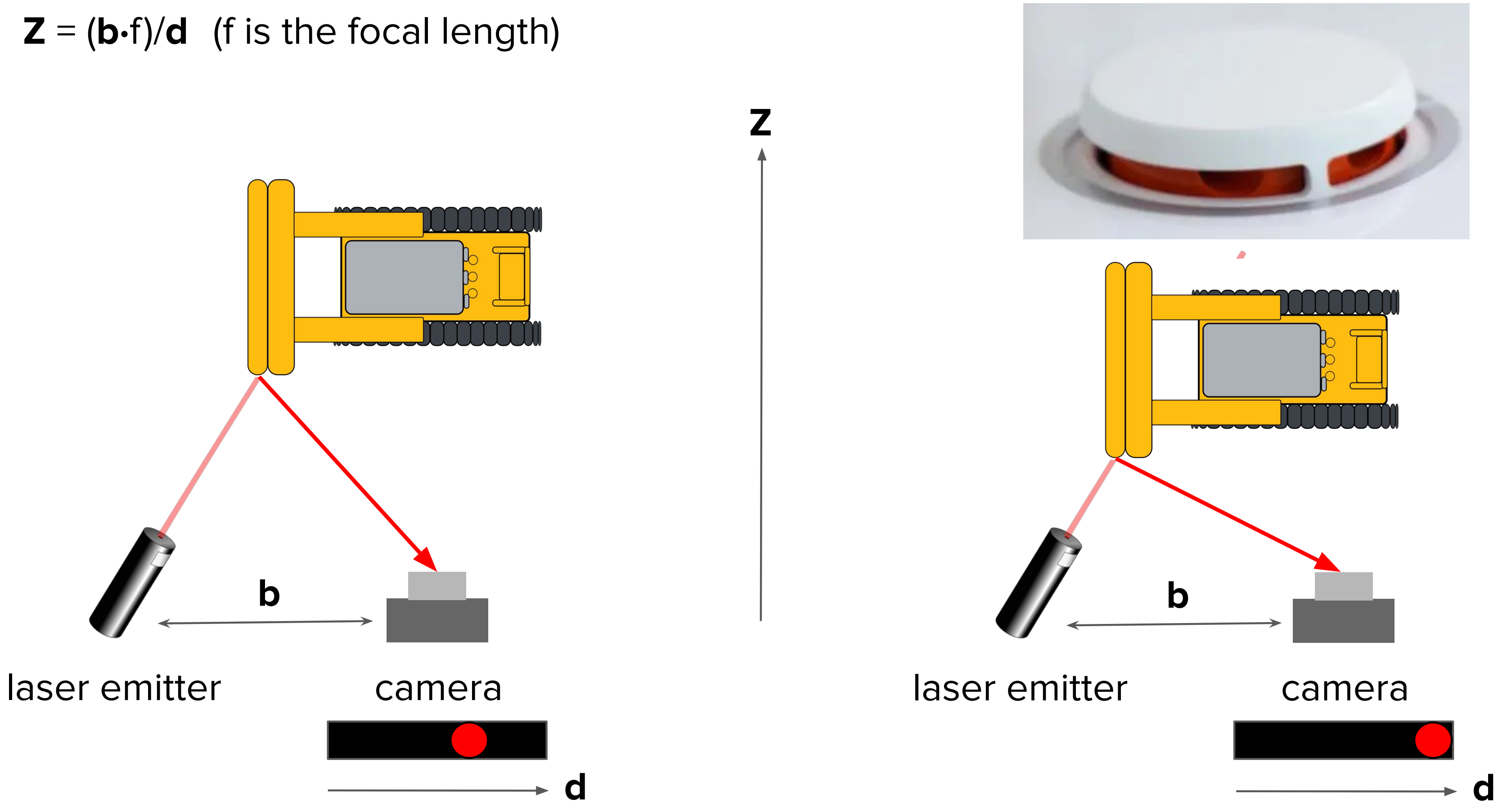
3D sensing in consumer devices

A well known deployment of 3D sensing in consumer devices: vacuum cleaners



Depth perception through active stereo triangulation

- $\mathbf{Z} = (\mathbf{b} \cdot f)/\mathbf{d}$ (f is the focal length)



Depth sensors

- Although cameras are cheap and have high resolution, they have some limitations
- 3D geometry of a scene can be obtained even deploying specialized sensors
- Most of them rely on the time-of-flight principle
- Used in notable application fields

https://apple.com/

https://waymo.com/

Depth perception through time-of-flight

- The speed of light is about 300000 m/s and it takes about 3.34 ns to travel for 1 meter
- Distance can be estimated measuring the time to bounce from emitter to receiver
- Distance **Z = v·t**
- Emitter and receiver can be designed according to different technologies
- For instance, laser emitter and receiver