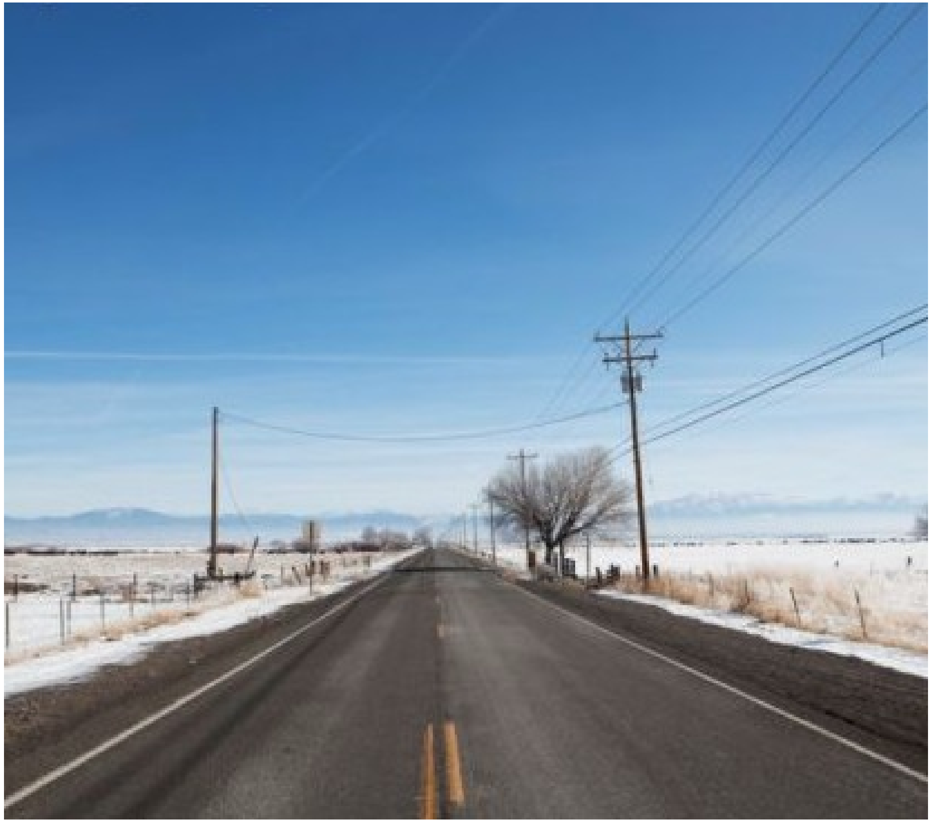
Computer Vision

Class 01

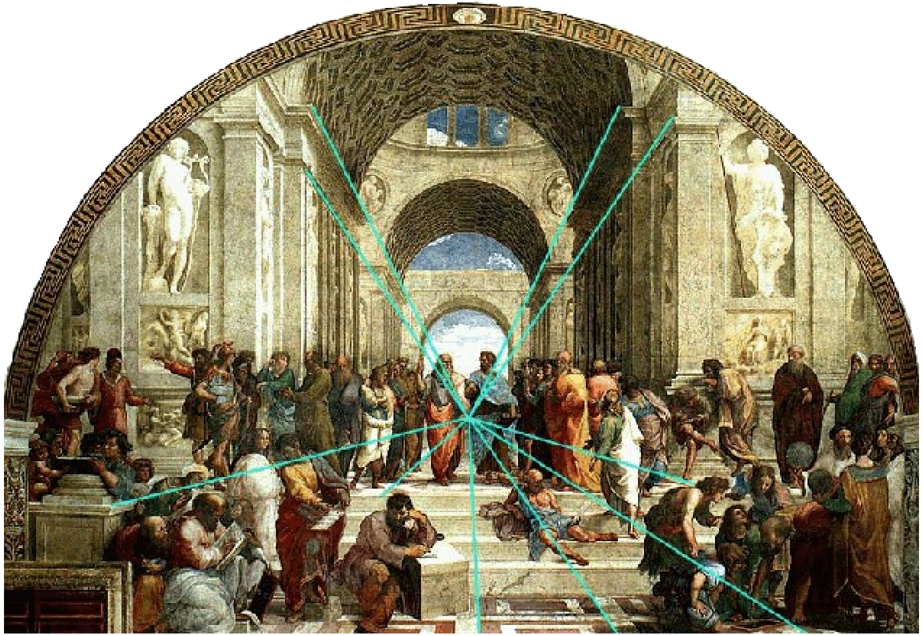
Raquel Frizera Vassallo

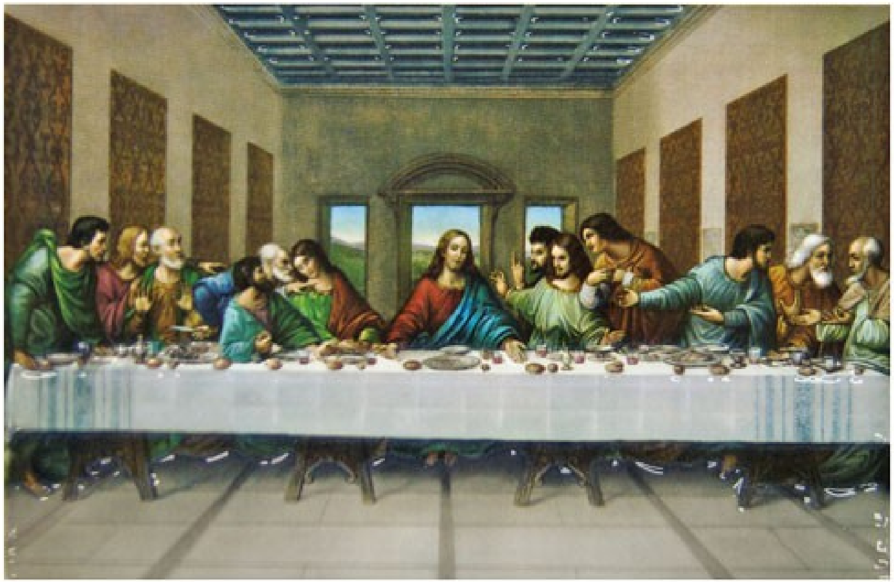
Introduction

Perspective



Perspective





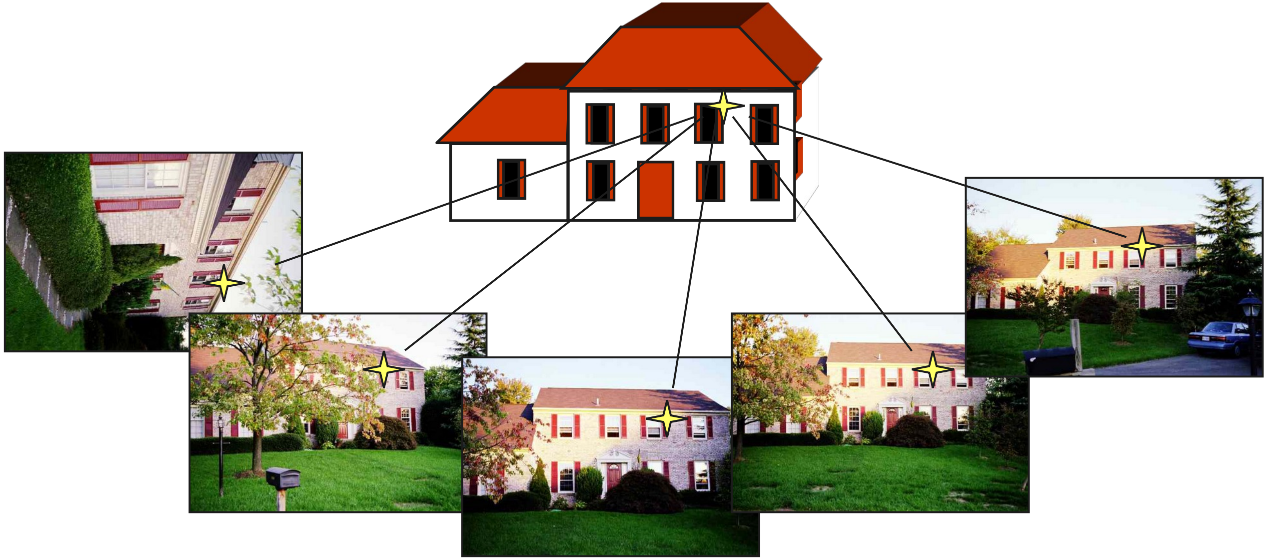
Perspective





3D Reconstruction

- Input: correspondent features on different images.
- Output: camera's position and orientation, camera calibration, 3D structure of the scene.



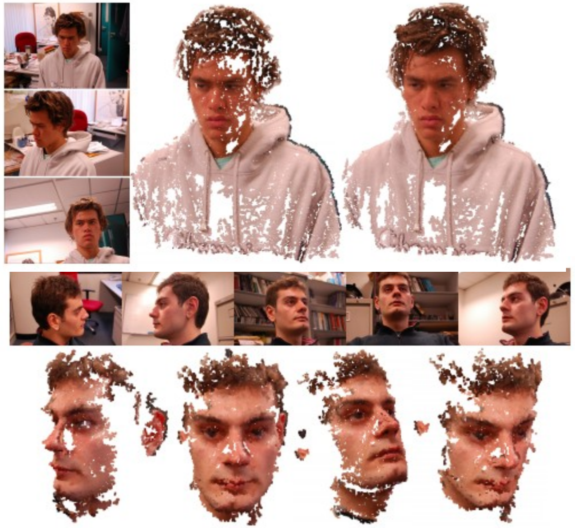
Some interesting videos

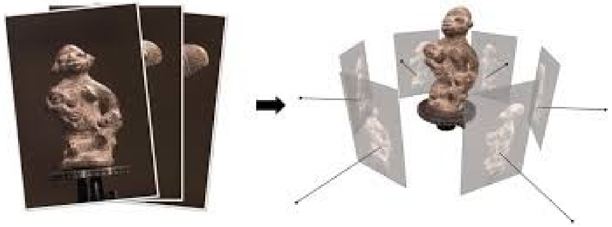
- Stereoscopic Vision - Brain Games
  https://www.youtube.com/watch?v=nKdUD8lIGJY
- What is Stereoscopy?
  https://www.youtube.com/watch?v=AN34QUEayzQ
- How does Stereo Vision work?
  https://www.youtube.com/watch?v=yfjMJfXMBcY

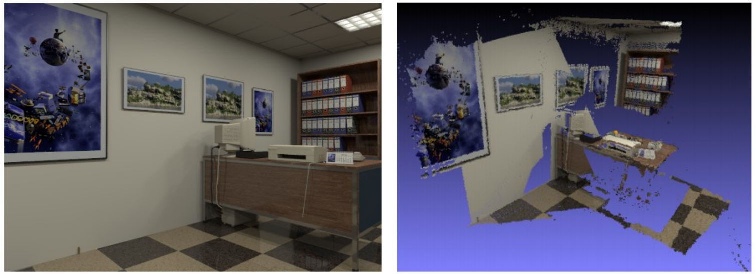
Some Applications

Aplications

- 3D Reconstruction

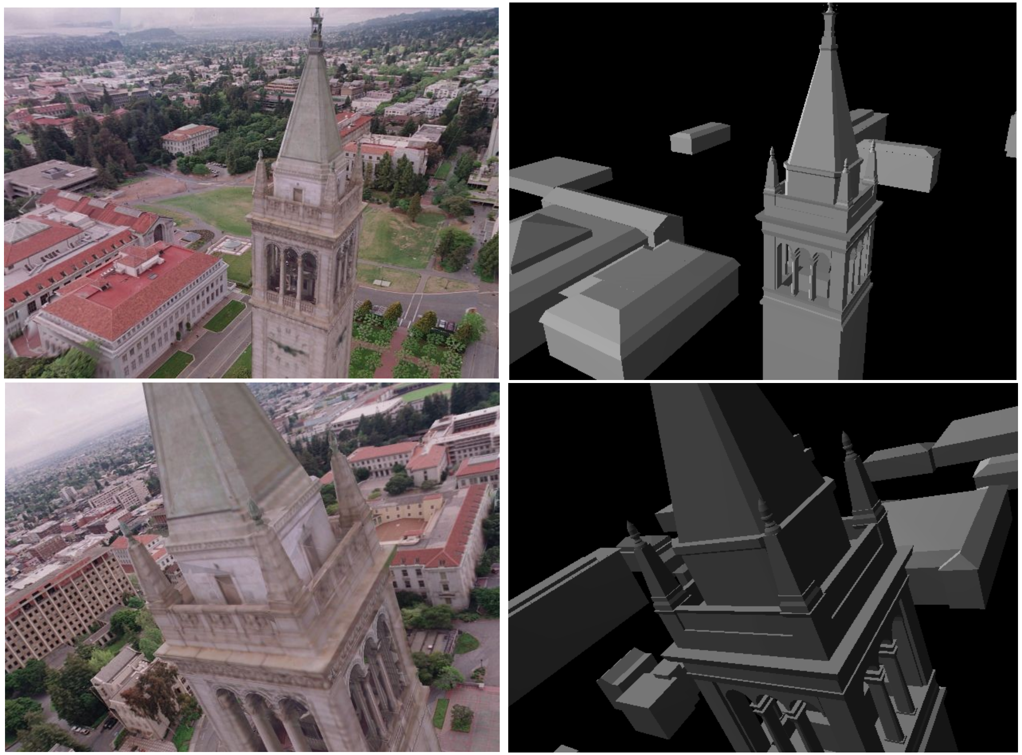






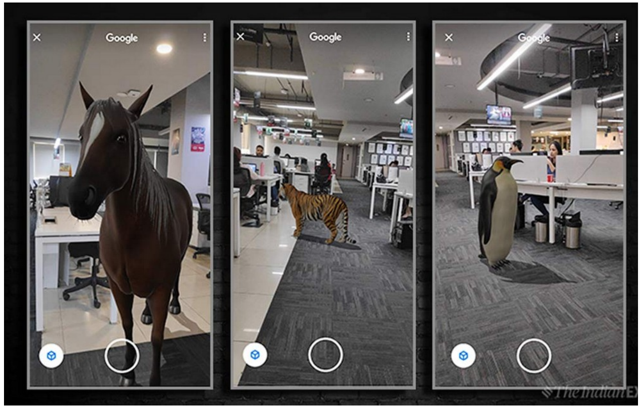
Aplications

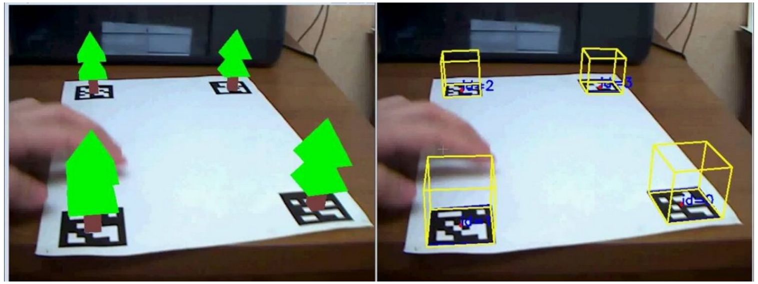
- Creation of 3D Models

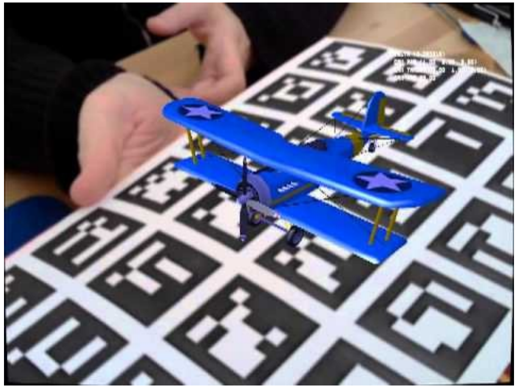


Aplications

- Inserting virtual objects, augmented reality

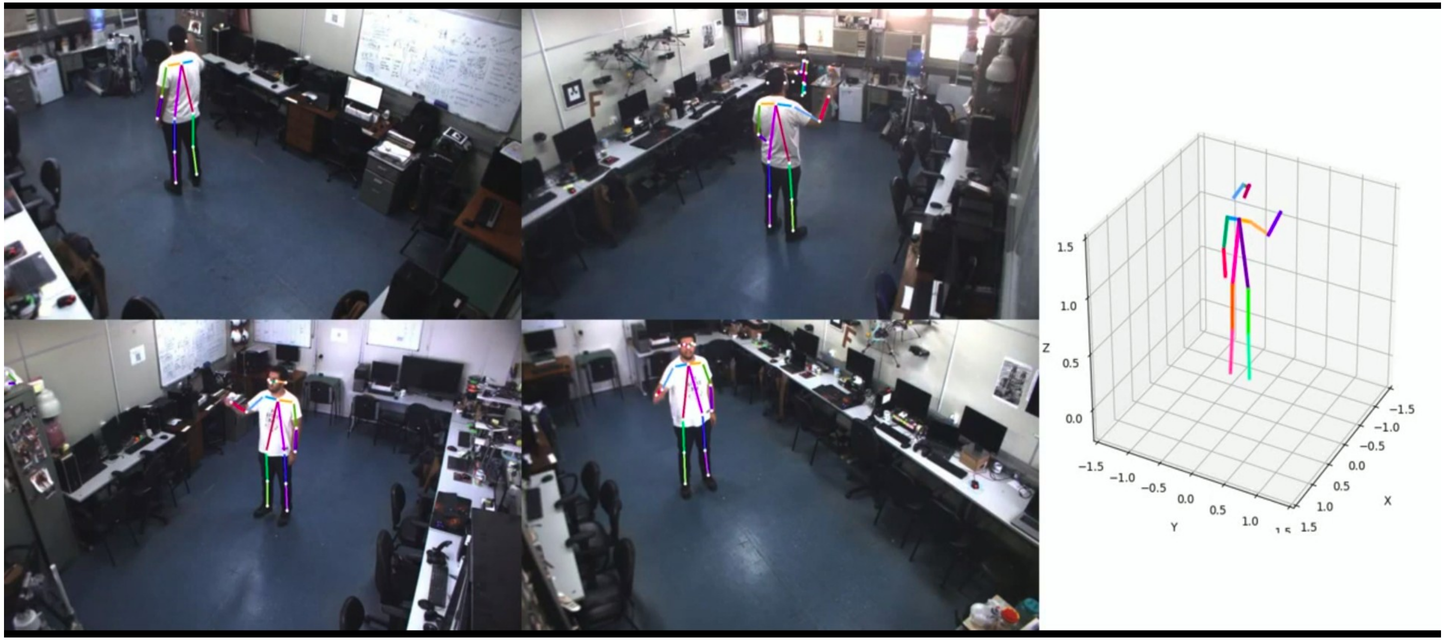






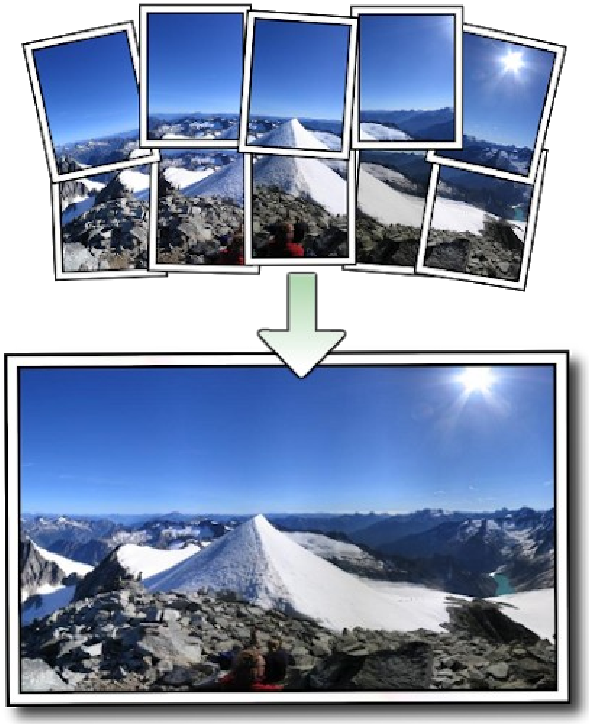
Aplications

- Recovering pose and orientation of objects or people



Aplications

- Image stitching and mosaics

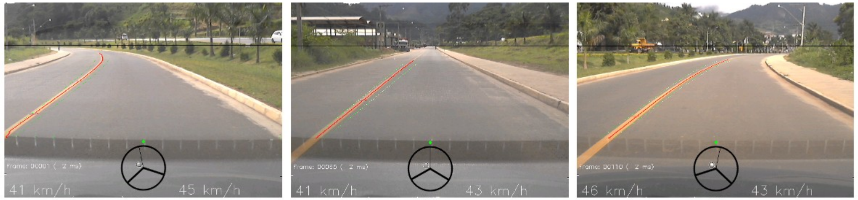


Aplications

- Virtual publicity and virtual checking

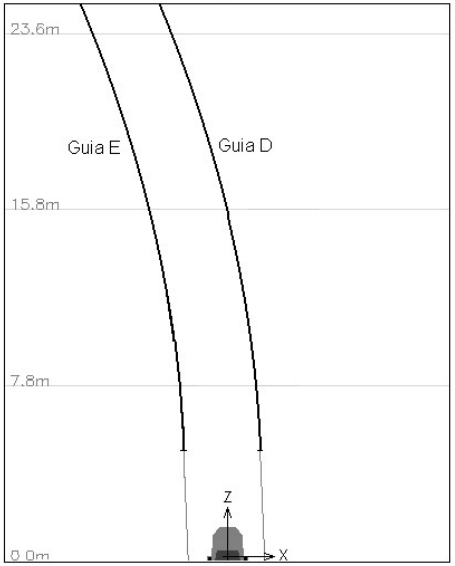
Aplications

- Autonomous vehicles

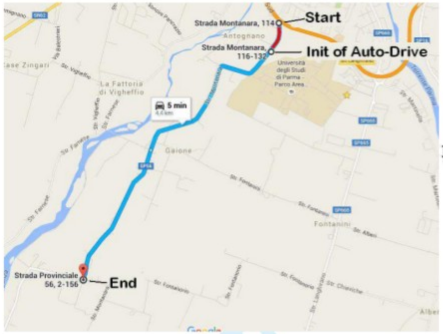


(A) (B) (C)

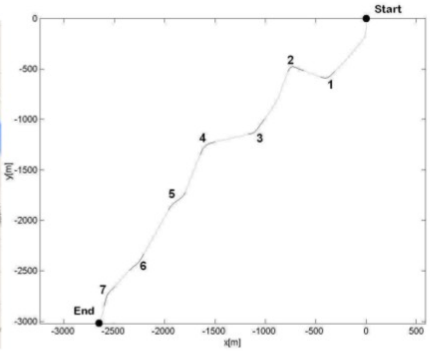

(A)



(B)



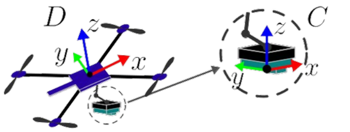
(a)

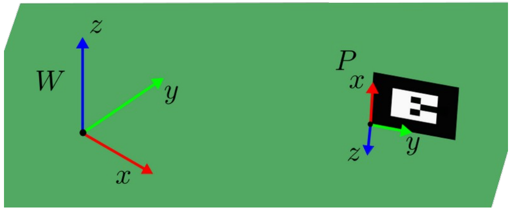


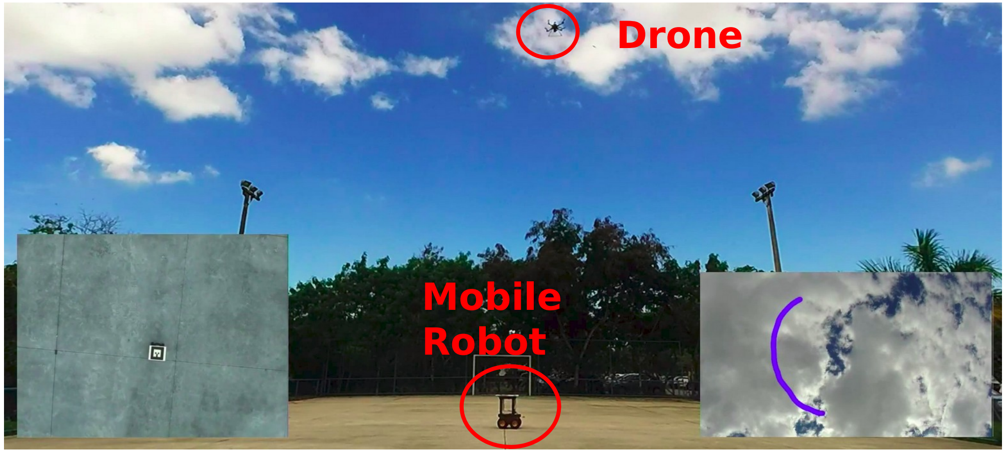
(b)

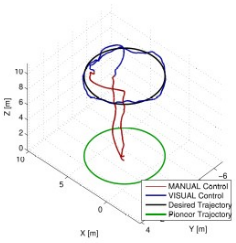
Aplications

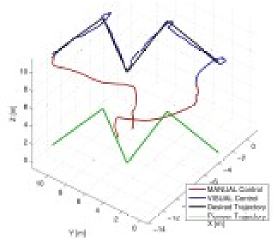
- Aerial vehicle control

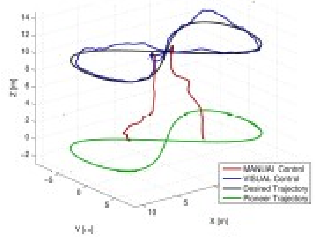






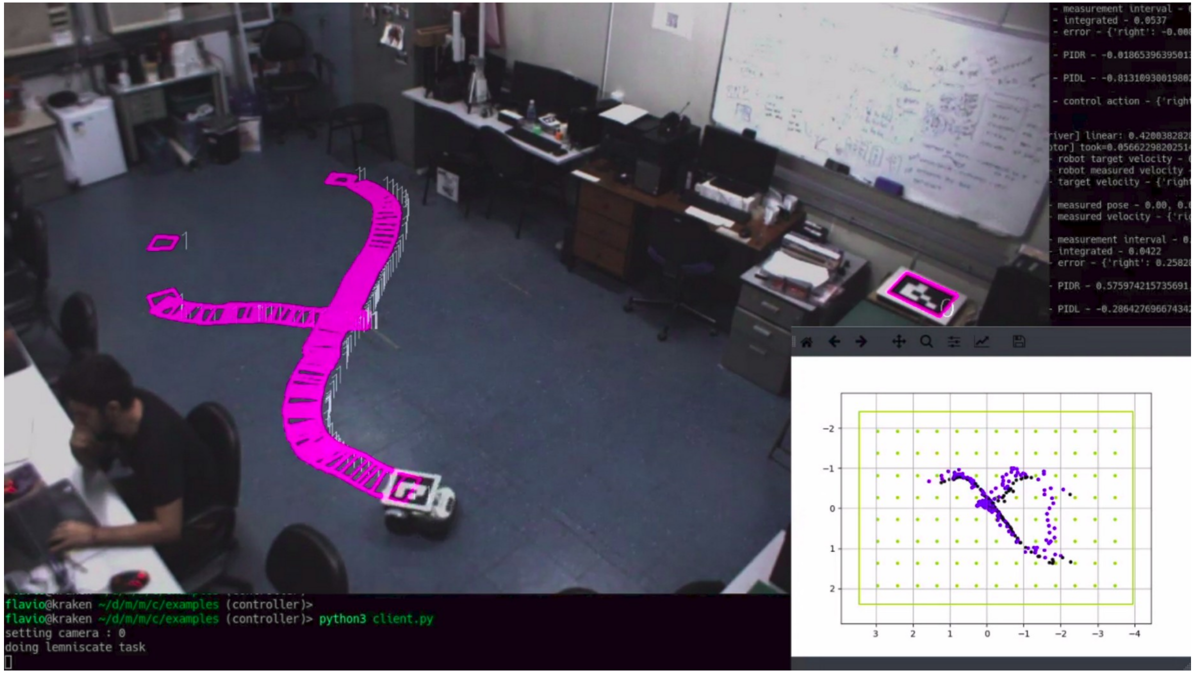






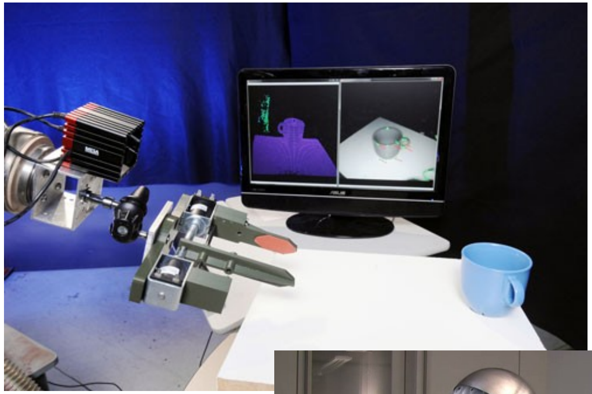
Aplications

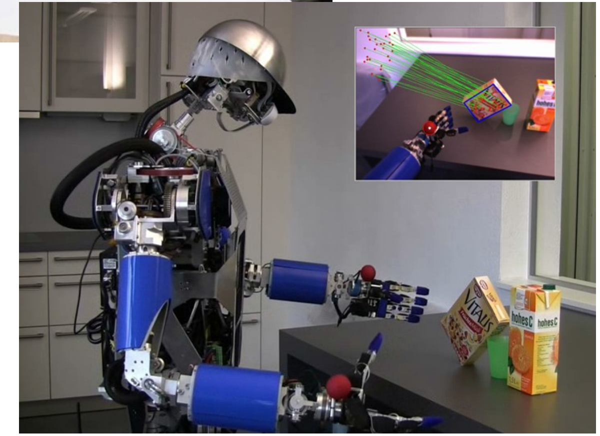
- Visual robot control



Aplications

- Visual servoing

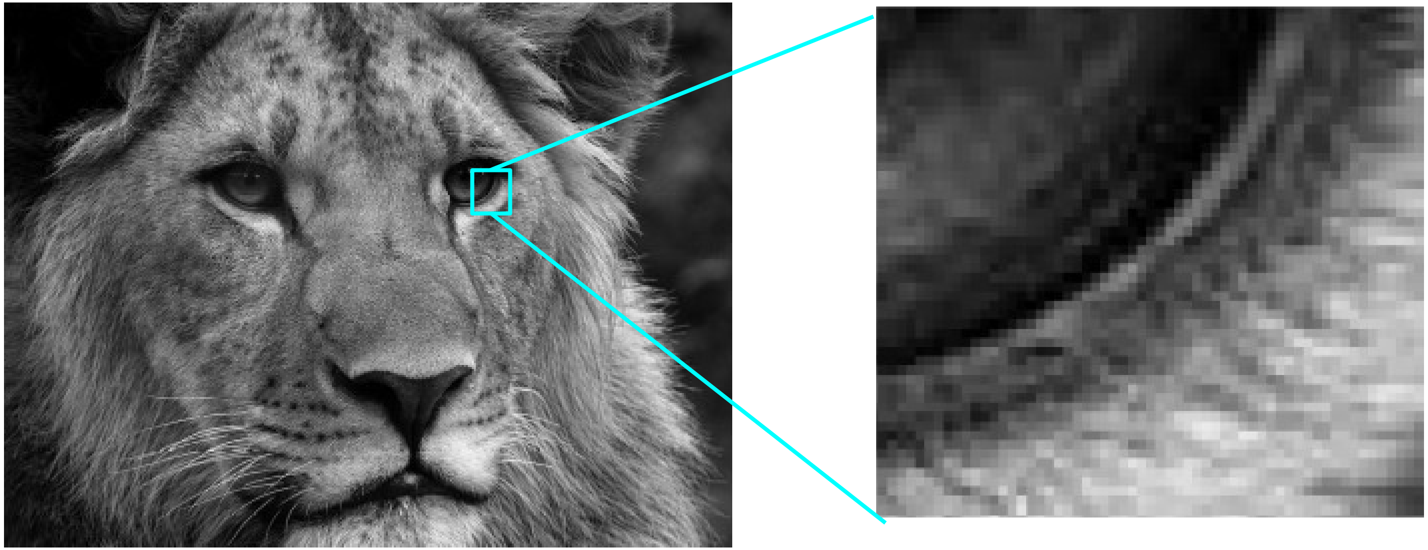




Basic Concepts

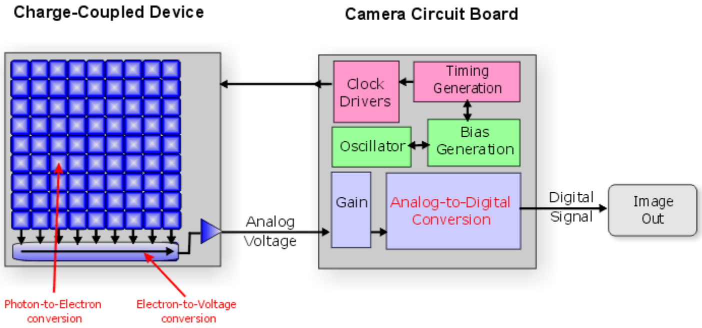
Gray levels and pixel

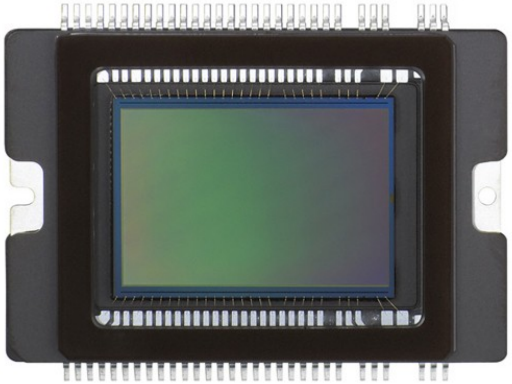
- Gray levels: 0 is "Black", 255 is "White", numbers in between are levels of gray
- Pixel: each number is shown as a shaded small square



Digital Images

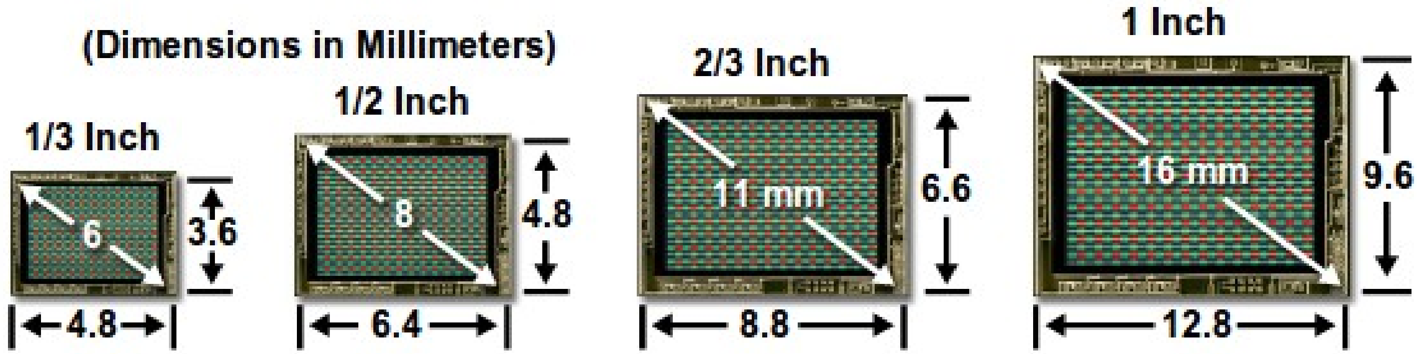
- A digital image: sampling and integrating continuous (analog) data in a spatial domain
- Uses one or several matrix sensors for recording an image
- Sensor matrix: $N_{cols} \times N_{rows}$ array of sensor elements (phototransistors)





Digital Images

- Produced either in charge-coupled device (CCD) or complementary metal-oxide semiconductor (CMOS) technology
- Sensor cell converts measured light into an electric charge, represented by a single number (the sample) in the created image.
- Typically been configured with square pixels assembled into rectangular area arrays, with an aspect ratio of 4:3 being most common

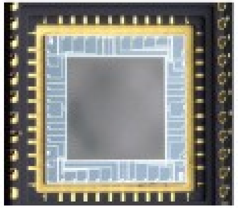


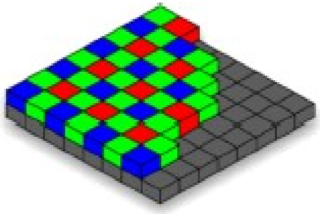
CCD x CMOS

- CCD Sensor Advantage
  - High quality
  - High sensitivity
  - Good for professional photographers and image scientists where the highest quality is needed
- CMOS Sensor Advantage
  - Less power consumption
  - Cheaper cost
  - Used for camera phones, tablets and other portable devices where power is very important and quality is not essential

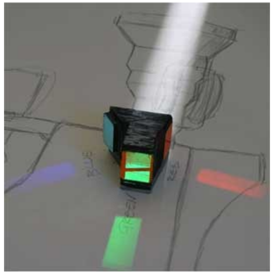
Color Digital Image

- Bayer Pattern
  - One colour pixel: four sensor elements (2x2)
  - Two for Green and one for Red and Blue each
  - A sensor array of size $N_{cols} \times N_{rows}$ generates colour images of $N_{cols}/2 \times N_{rows}/2$





- Beam Splitting
  - High quality digital colour cameras: a beam splitter (e.g. using two dichroic prisms) is used to split light into three beams: Red, Green, and Blue.
  - $N_{cols} \times N_{rows}$ sensor arrays for R, G, and B for one pixel
  - Images of $N_{cols} \times N_{rows}$



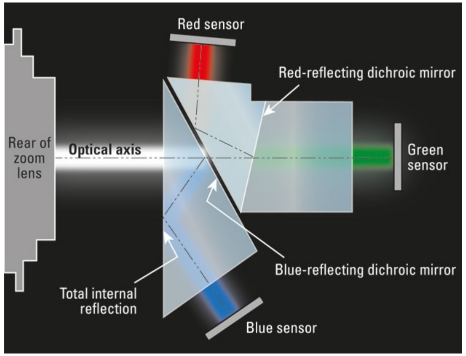
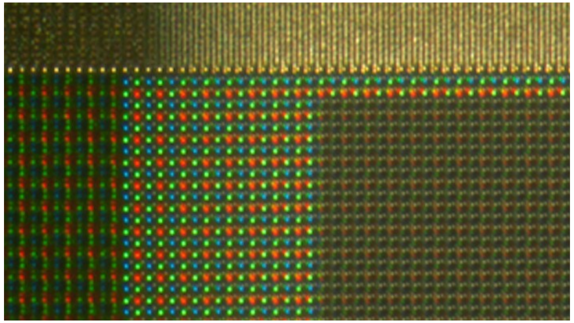


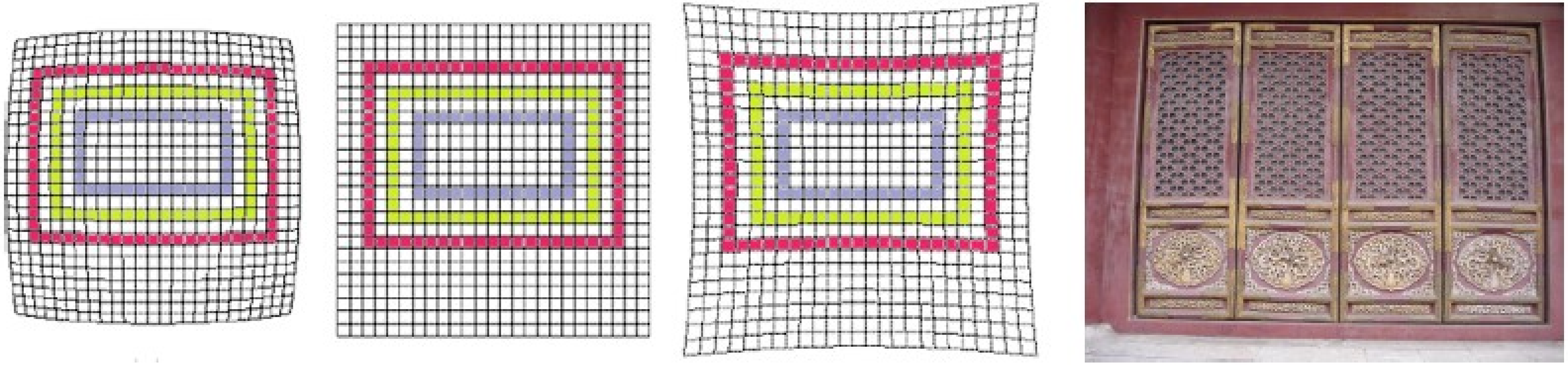
Image Resolution and Bit Depth

- **Aspect Ratio**. Each phototransistor is a rectangular cell. Ideally, the aspect ratio should be equal to 1 (i.e. square cells)
- **Resolution.** Number of sensor elements. Example: 4 Mpixel camera (4.000.000 pixel) in some image format. Without further mentioning, the number of pixels means "color pixel"
- **Sensor Noise**. More pixels: Smaller sensor area per pixel. Thus less light per sensor area and a worse signal-to-noise ratio (SNR)
- **Bit Depth.** Number of bits per pixel. Common goal: more than just 8 bits per pixel value in one channel.
  E.g. 16 bits per pixel in a gray-level image for motion or stereo analysis.



Lens Distortion

- Optic lenses contribute to radial lens distortion on the projection process
- Barrel transform or pincushion transform



Left to right: Barrel transform, ideal rectangular image, pincushion transform, and projective and lens distortion combined in one image

Image

An image I is a rectangular array of pixels (x,y,u).
A pixel combines location $p = (x,y) \in \mathbb{Z}^2$ and sample u at p.

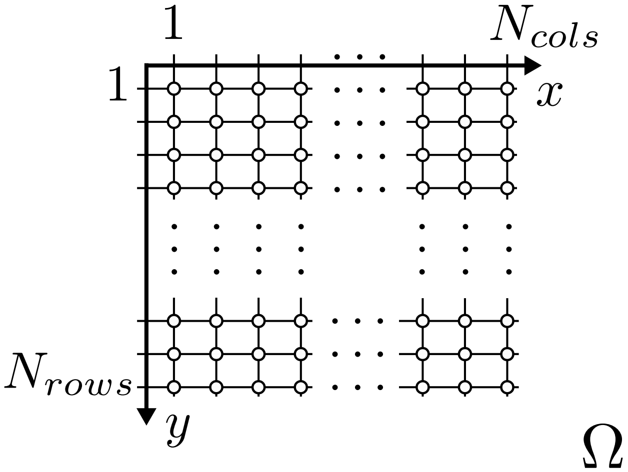
$\mathbb{Z}$ is the set of all integers; points $(x,y) \in \mathbb{Z}^2$ form a regular grid.

An image I is defined on a carrier:

$$\Omega = (x, y) : 1 \leq x \leq N_{cols} \wedge 1 \leq y \leq N_{rows} \subset \mathbb{Z}^2$$

of N_{cols} x N_{rows} pixel locations (*grid points*)

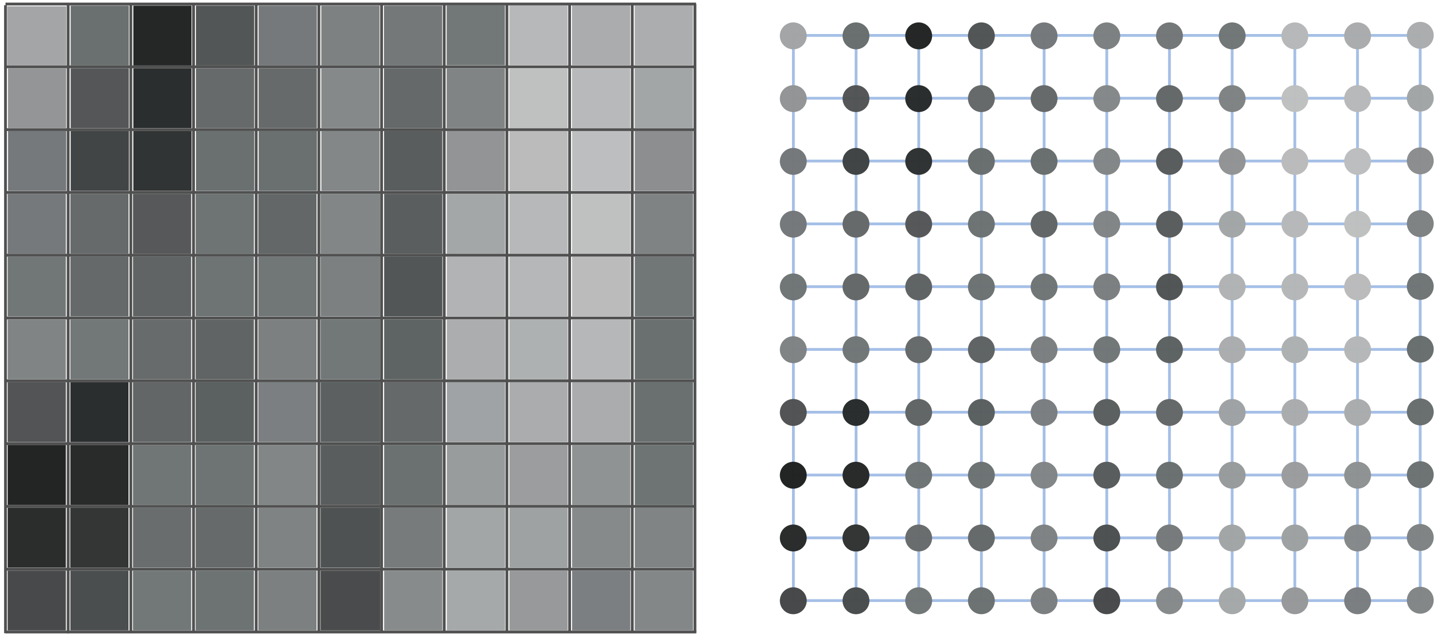
Image Coordinate System



Row y contains *grid points* $\{(1, y), (2, y), \ldots, (N_{cols}, y)\}$, for $1 \leq y \leq N_{rows}$

Column x contains *grid points* $\{(x, 1), (x, 2), \ldots, (x, N_{rows})\}$, for $1 \leq x \leq N_{cols}$

Pixels

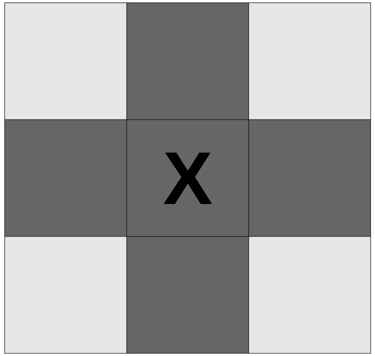


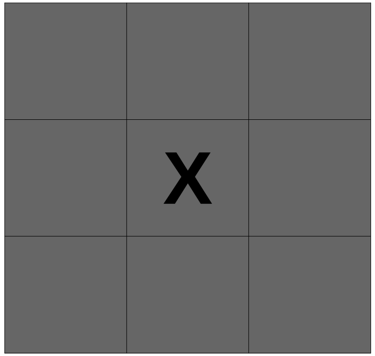
Left: Image values as shades in *gridsquares* (grid cells)

Right: Image values as labels at *gridpoints* (center of grid squares)

Adjacency

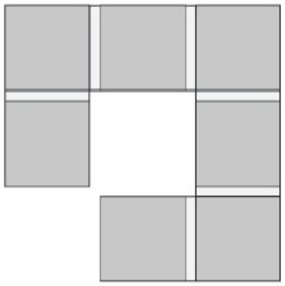
- Two types: 4 or 8-adjacency.
- Two pixel locations p and q in *grid cell* model are adjacent if:
  - **4-adjacency:** their tiny shaded squares share an edge;
  - **8-adjacency:** their tiny shaded squares share an edge or corner.

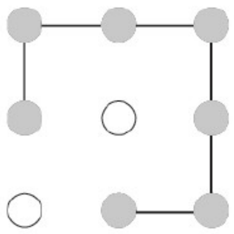




Connected or not connected sets of pixels

- Depends on adjacency
- Left and middle: Connectedness of gray pixels due to 4-adjacency; the two white pixels are not connected. We have an open blob
- Right: Connectedness of gray pixels due to 8-adjacency. The two white pixels are "crossing" this loop. We have a closed blob.





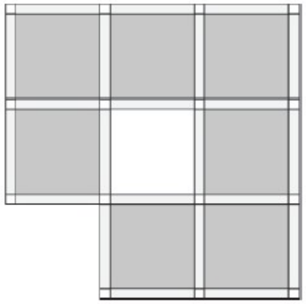


Image Window

- Image processing is often local (just addressing data in a window).
- The reference point of a window is usually at the window's centre.

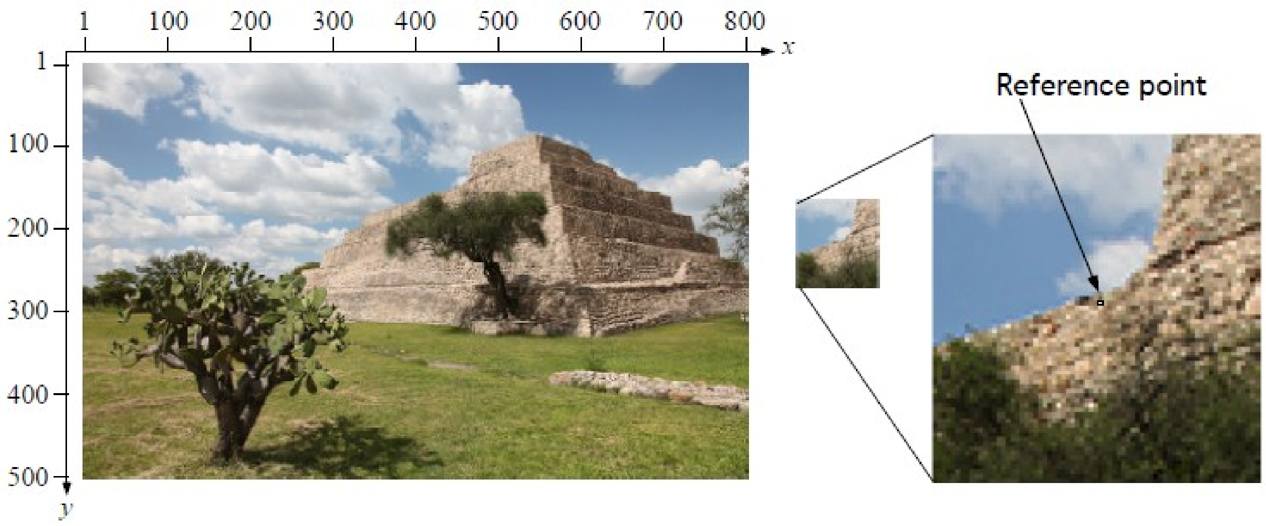


Image Window

A *window* $W_p^{m,n}(I)$ is a subimage of image I of size m x n positioned with respect to a *reference point* p.

Default: odd number $m = n$ and p at the center of the window.

Usually: Simplified notation W_p
(because image and size of window are known by given context)

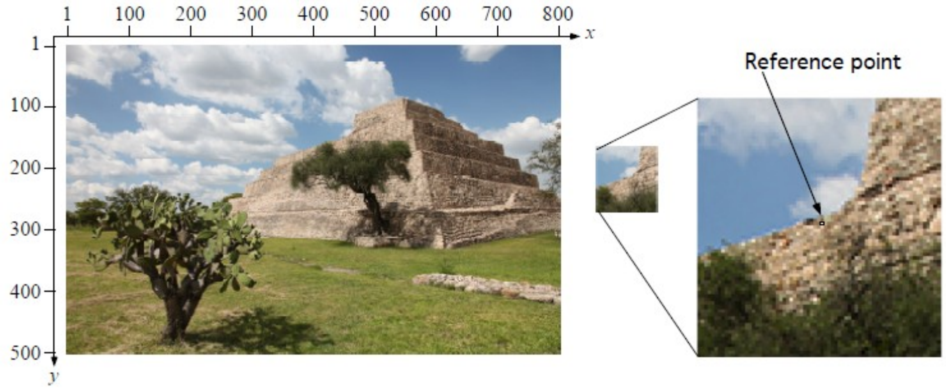


Image Values

Scalar image: Values are integers $u \in \{0, 1, \ldots, 2^a - 1\}$.

Common: $a = 8$ (i.e. one byte) or $a = 16$.
Let $G_{max} = 2^a - 1$ be the general maximum image value.

Example
A *gray-level* image is a scalar image where scalar values represent gray levels, with $0 =$ black and $2^a =$ white; all other gray-levels are linearly interpolated between black and white.

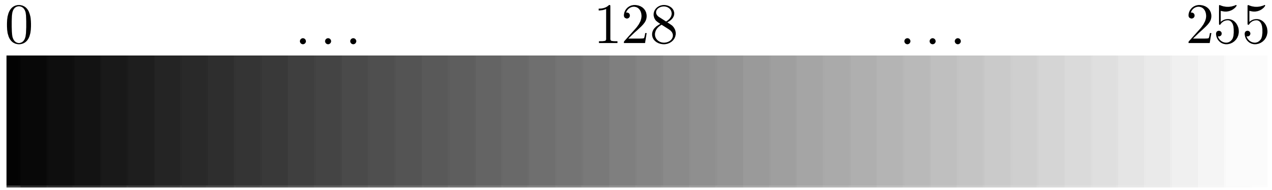
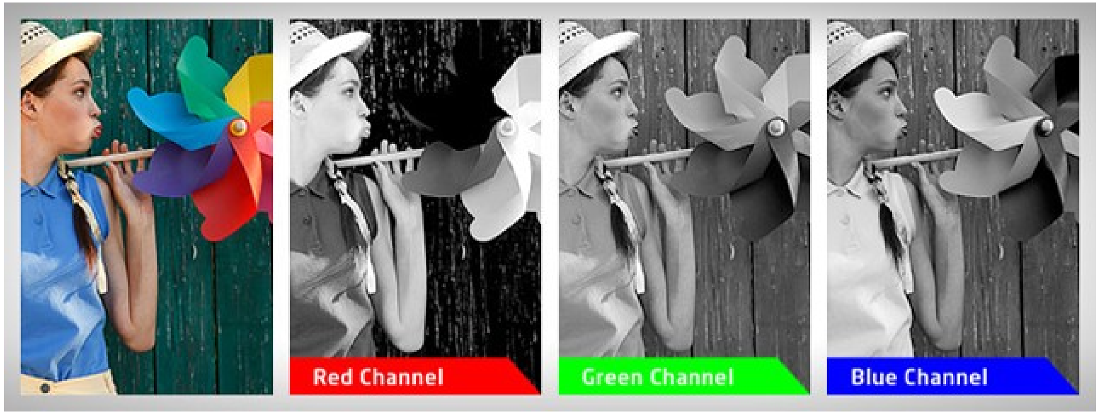


Image Values

- A binary image has two values, traditionally 0 = white and 1 = black (black objects on white background).
- A vector-valued image has more than one channel. Image values are represented as $(u_1; \ldots; u_{Nchannels})$.
- Color images in the RGB color model have 3 channels: red, green, and blue.
- Values in each channel are in the set $\{0; 1; \ldots; G_{max}\}$ (like a gray-value image).



Mean

Given: N_{cols} x N_{rows} scalar image I

Mean (i.e. the “average gray level”) of image I

$$\mu_I = \frac{1}{N_{cols} \cdot N_{rows}} \sum_{x=1}^{N_{cols}} \sum_{y=1}^{N_{rows}} I(x,y) \; = \frac{1}{|\Omega|} \sum_{(x,y)\in\Omega} I(x,y) \; = \frac{1}{|\Omega|} \sum_{p\in\Omega} I(p)$$

$|\Omega| = N_{cols} \,.\, N_{rows}$ is the cardinality of the carrier Ω

Variance and Standard Deviation

Variance of image *I:*

$$\sigma_I^2 = \frac{1}{|\Omega|} \sum_{(x,y)\in\Omega} [I(x,y) - \mu_I]^2$$

Root σ_I is the *standard deviation* of image *I.*

Alternative formula*:*

$$\sigma_I^2 = \left[\frac{1}{|\Omega|} \sum_{(x,y)\in\Omega} I(x,y)^2\right] - \mu_I^2$$

Credits

- R. Klette. **Concise Computer Vision.**
  Springer-Verlag, London, 2014.
- Yi Ma, Stefano Soatto, Jana Kosecka e S. Shankar Sastry. **An Invitation to 3D Vision: From Images to Geometric Models.**
  Springer, ISBN 0387008934