COMPUTER VISION

MEI/1

University of Beira Interior

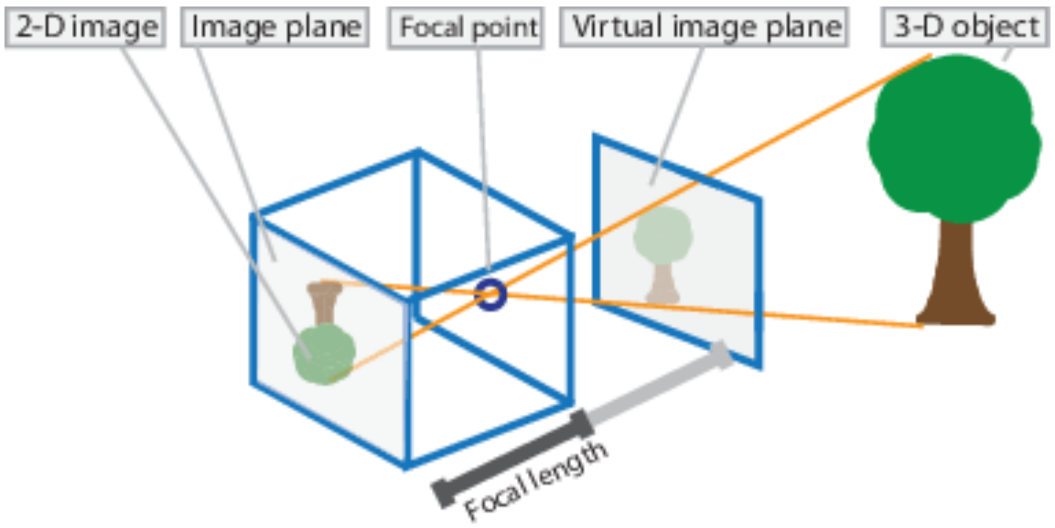
Department of Informatics

Hugo Pedro Proença

hugomcp@di.ubi.pt, **2023/24**

Camera Calibration

- A camera is a device that essentially converts (projects) positions at the 3D world into a 2D image.
- Cameras capture three-dimensional regions of the world and store them in two-dimensional images.
- The fundamental equation for this representation is:



Source: https://www.mathworks.com/help/vision/ug/camera-calibration.html

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{X}$$

Image coordinate

Camera Matrix

World coordinate

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Homogeneous 2D coordinates

Homogeneous 3D coordinates

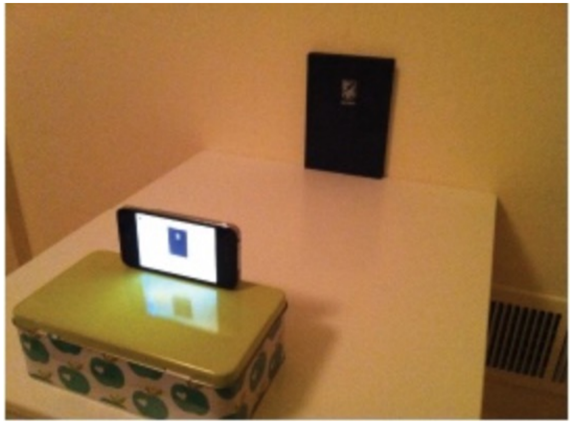
Camera Calibration

- Camera calibration is an important phase in many computer vision applications, in particular when physical information of the scene is required
- The term "**calibration**" (a.k.a. camera **resectioning**) refers to the estimation of two types of parameters:
  - **Intrinsic** (a.k.a. internal) **parameters**
    - Allows the mapping between pixel coordinates and camera coordinates in the image frame
    - Optical center, focal length and radial distortion of the lens
  - **Extrinsic** (a.k.a. external) **parameters**
    - Describe the orientation and location of the camera in the world
    - Refers to the rotation and translation information with respect to some world coordinate system.
- These parameters can be used to:
  - Correct for lens distortion, measure the size of an object in world units, or determine the location of the camera in the scene.
  - This kind of tasks is important to applications such as **object detection and measurement**. They are also used in robotics, for **navigation systems**, and 3-D scene reconstruction.

Camera Calibration: Simplest Method

- In this simplest camera calibration procedure we just need a flat object with a simple form (e.g., square or rectangle) and a ruler to obtain its physical dimensions.
- The key factor to be determined is the **focal length** because most of the remaining parameters can be estimated using reasonable assumptions
  - Square straight pixels, optical center at the image center
- The setup is simple. We **place the camera parallel to the object**, with the **center of mass** of the **object** appearing near the **center** of the **image**.

a) Scene



b) Image

Camera Calibration: Simplest Method

1. Using the ruler, take a measurement of the width and height of the object (e.g., book).
   1. Let these values be denoted by **w_w** and **h_w**
2. Using the ruler, measure the distance between the camera and the center of mass of the object.
   1. Let that value be denoted by **d**
3. Capture an image of the object, using the camera (similar to the “b)” image in the previous slide).
4. Obtain the width and height of the object in the image, measured in pixels
   1. Let these values be denoted by **w_i** and **h_i**
5. The camera focal length is given by:

$$\boldsymbol{f_x = \frac{w_i}{w_w} d} \text{ and } \boldsymbol{f_y = \frac{h_i}{h_w} d}$$

Camera Calibration: Simplest Method

- **Of course, the obtained values fully depend on the image resolution used.**
  - **Typically, the focal length and the optical center are measured in pixels and scale with respect to image resolution.**
- **The full calibration matrix will be given by:**

$$\begin{bmatrix} f_x & s & c_x \\ 0 & af_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

where (c_x, c_y) denote the optical center, "s" is the skew factor and "a" is the aspect ratio (respectively, "0" and "1" for practical purposes and most cases).

- **Also, typically, (c_x, c_y) = image resolution / 2 is a good approximation for most commercial cameras.**

Camera Calibration: Simplest Method

This matrix (M) is used to convert between world coordinates w.r.t. the camera and image coordinates

Position in Image coordinates

$$\vec{i} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \vec{w}$$

Position in World coordinates

The image coordinates are expressed in homogenous form. To convert to standard (Euclidean) form, we just need to divide by the last coordinate and discard the last dimension.

[3, 2, 4] → [0.75, 0.5, 1] → [0.75, 0.5]

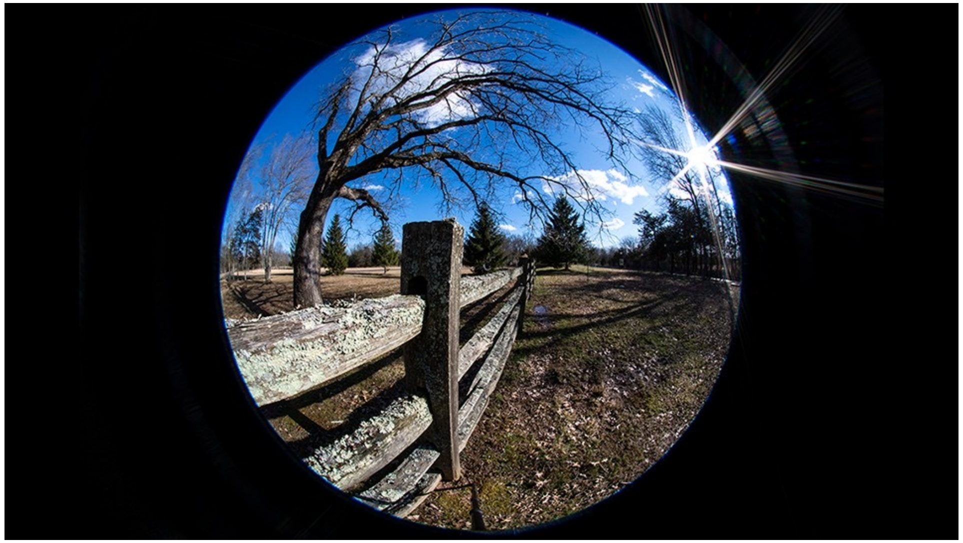
This matrix can also useful for augmented reality purposes, by obtaining: $\vec{w} = M^{-1}\vec{i}$

"Adding some information at pixel (a,b) simulates a physical object at position (p,q,j)"

Camera Models

- In most cases, camera calibration assume one of two models: Pinhole and Fisheye.
  - Pinhole is the most used and it regards a basic camera model without even considering a lens.
  - As the pinhole camera model cannot be used in Fisheye cameras (due to the high distortions produced by fisheye lenses), an alternative model had to be considered.

Example of an image obtained by a Fisheye camera

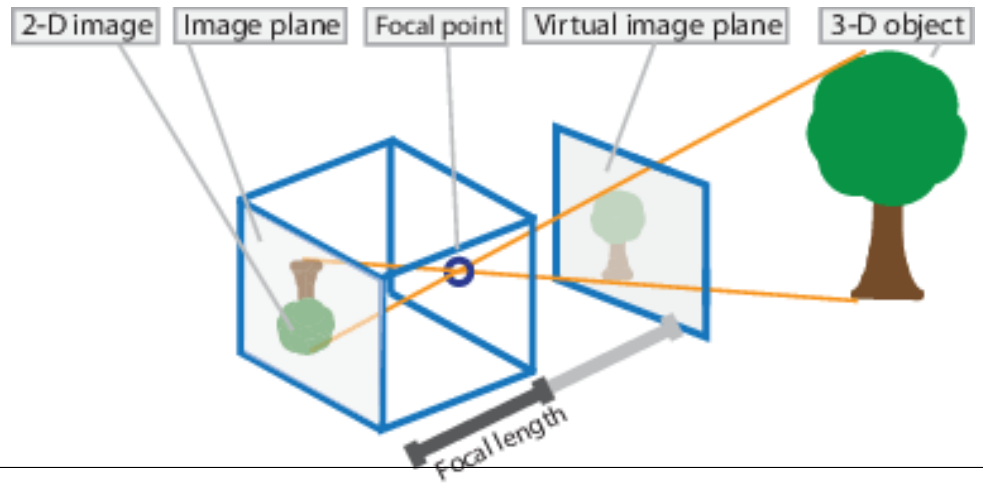


Source: https://www.bhphotovideo.com/explora/photography/

Pinhole Camera Model

- The pinhole model assume a *perfect lens* , in the sense that light rays passing through the aperture are projected at the precise opposite side of the camera (Image Plane)
  - The distance between the Image plane and the aperture (Focal Point) is name “Focal Length”.
- At the same distance between the image plane and focal point (i.e., the focal length) and at an opposite direction, we have the Virtual Image Plane.

Source: https://www.mathworks.com/help/vision/ug/camera-calibration.html



Pinhole Camera Model

- **The pinhole camera parameters are represented by a 4 x 3 matrix, named the "Camera Matrix"**
  - **This matrix maps the 3-D world scene into the image plane. The calibration process calculates the camera matrix using the extrinsic and intrinsic parameters.**
  - **The extrinsic parameters represent the location of the camera in the 3-D scene.**
  - **The intrinsic parameters represent the optical center and focal length of the camera.**

Image coordinates

World coordinates

$$\boldsymbol{w}\,[\boldsymbol{x}, \boldsymbol{y}, \mathbf{1}] = [\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}, \mathbf{1}]\,\boldsymbol{P}$$

Scale factor

Intrinsic Matrix

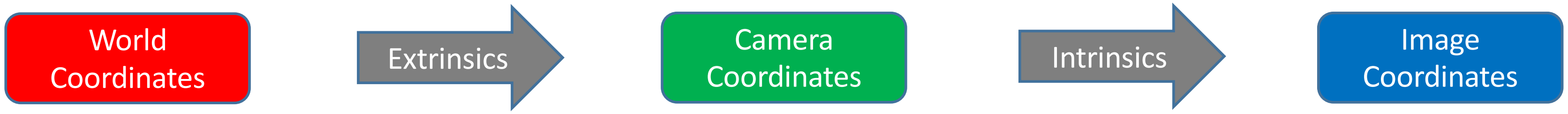
$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{t} \end{bmatrix} \boldsymbol{K}$$

Camera Matrix

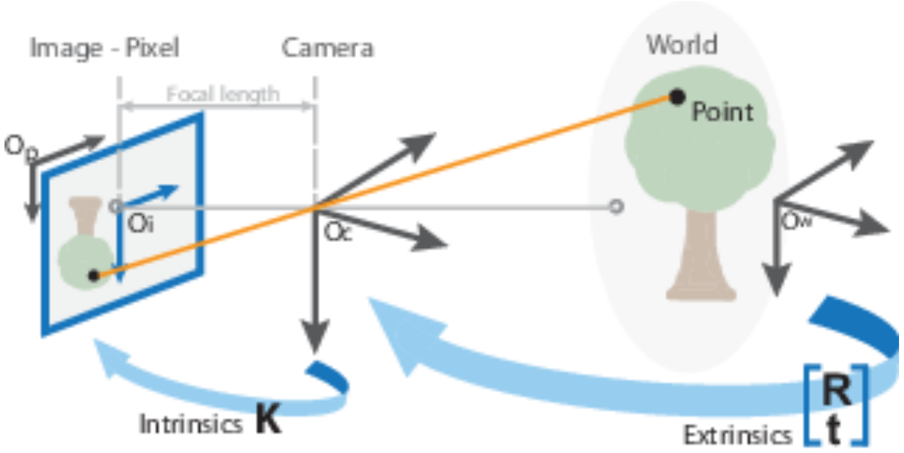
Extrinsic Matrix

Pinhole Camera Model

- **The world points are transformed to camera coordinates using the extrinsics parameters**
- **The camera coordinates are converted into image coordinates using the intrinsics parameters**



World Coordinates → Extrinsics → Camera Coordinates → Intrinsics → Image Coordinates



Source: https://www.mathworks.com/help/vision/ug/camera-calibration.html

Camera Parameters

- The Extrinsic parameters are the 3 x 3 Rotation matrix (R) and the 3 x 1 translation vector (t)
- These elements describe the absolute position of the Focal Plane in the world coordinate system.
- By composition of transformations (Rotation + Translation), the final Projection Matrix is given by:

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix}, \boldsymbol{R} = \begin{bmatrix} \boldsymbol{r}_{11} & \boldsymbol{r}_{12} & \boldsymbol{r}_{13} \\ \boldsymbol{r}_{21} & \boldsymbol{r}_{22} & \boldsymbol{r}_{23} \\ \boldsymbol{r}_{31} & \boldsymbol{r}_{32} & \boldsymbol{r}_{33} \end{bmatrix}, \boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3]$$

where 0= [0, 0, 0], and 1 is a scalar.

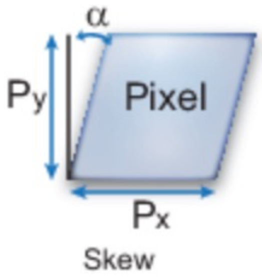
Camera Parameters: Intrinsic

- **The** intrinsic parameters **include the focal length, the optical center (a.k.a. the principal point) and the skew coefficient. The intrinsic camera matrix is given by:**

$$\begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

(c_x, c_y): optical center, (f_x, f_y): focal length (in pixels), where $f_x = F/p_x$, $f_y = F/p_y$, F is the focal length expressed in world units (typically millimeters) and (p_x, p_y) is the size of one pixel in world units. s is the skew coefficient, which is non-zero when the image axes are not perpendicular ($s = fx\tan(\alpha)$)

The camera matrix **does not account for lens distortion** because an ideal pinhole camera does not have a lens. To accurately represent a real camera, the camera model **must consider** the **radial** and **tangential lens distortion**.



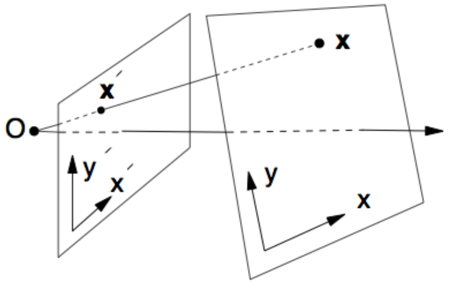
Camera Parameters: Extrinsic

- As previously seen, the extrinsic parameters enable to obtain the "pose" of the camera in a 3D world.
- The idea is to get a mapping between the image plane and the camera plane.
- Such mapping is expressed by means of a Projective Matrix, that can be obtained by classical "least squares" methods, upon groups of correspondences between points in pairs of images.

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} \doteq \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix} x & y & 1 & 0 & 0 & 0 & -x'x & -x'y & -x' \\ 0 & 0 & 0 & x & y & 1 & -y'x & -y'y & -y' \end{bmatrix} \mathbf{h} = \mathbf{0}$$



**The solution "h" corresponds to the one-dimensional null space of "A".
With more than 4 correspondences, a least squares solution can be found.**

Camera Parameters: Extrinsic

- There are different packages/tutorials to obtain the intrinsic + extrinsic camera calibration parameters in a semi-automated way.

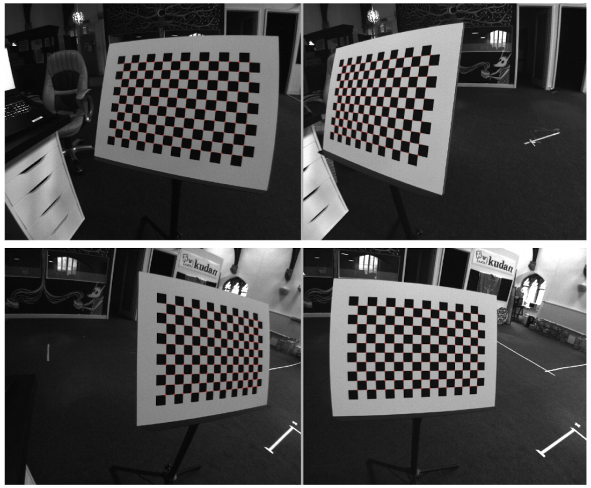
- (MATLAB)

https://www.mathworks.com/help/vision/ug/camera-calibration.html

- (OpenCV)

https://opencv24-pythontutorials.readthedocs.io/en/latest/py_tutorials/py_calib3d/py_calibration/py_calibration.html

- Most of them use the classical "checkboards", to obtain point correspondences between Images.



Camera Parameters: Extrinsic

- Today's cheap pinhole cameras introduces a lot of distortion to images. Two major distortions are radial distortion and tangential distortion.
    - Due to radial distortion, straight lines will appear curved. Its effect is more as we move away from the center of image.
    - Tangential distortion occurs because the device lense is not aligned perfectly parallel to the imaging plane. As result, some areas in the image may look nearer than expected.