UNIVERSITÀ DEGLI STUDI
DI MILANO

Corso di

Visione Artificiale

Laurea Magistrale in Informatica (F94)

Docente:
Raffaella Lanzarotti

Dipartimento di Informatica
Università degli Studi di Milano

Where are we?

First part:
Image formation and Early vision

- Image formation
  - Geometric Camera Models
  - Color spaces
- Image Processing
  - Punctual and spatial processing
  - Feature Extraction
- Reconstruction
  - **Camera calibration**
  - Stereo Vision
  - Structure from Motion and RGB-d Cameras
  - Optical flow and Tracking

Second part:
Machine learning for CV

- Linear Neural Network
- Multi Layer Perceptron
- Convolutional Neural Networks
- Recurrent Neural Networks
- Transformers
- Generative Adversarial Networks
- Graph Neural Networks

1. IMAGE FORMATION

Camera Calibration

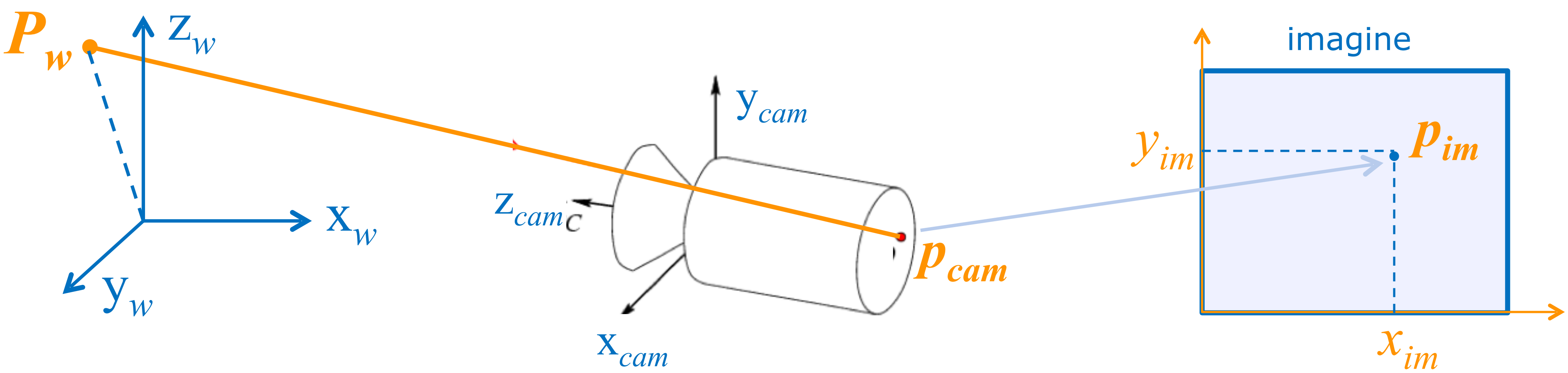
Chapter 1 – Forsyth Ponce

credits, F. Pedersini, S. Nayar

Camera calibration – definition

Camera Calibration:

Process to determine the geometric model of a camera



Projection Matrix: $\mathbf{M}(\xi)$: $\tilde{p}_{im} = \boxed{\mathbf{M}} \bullet \tilde{P}_w$

Calibration: determine **M** (or the camera parameters ξ)

REMARK:
$\tilde{x}$ used to indicate homogeneous coords

RECALL: Complete perspective projection camera model

$$\tilde{\mathbf{p}}_{IM} = \underbrace{\begin{bmatrix} f & 0 & x_C & 0 \\ 0 & f & y_C & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{M_{in}} \cdot \underbrace{\left[\begin{array}{c:c} \mathbf{R} & \mathbf{T} \\ \hdashline \mathbf{0} & 1 \end{array}\right]}_{M_{ext}} \cdot \tilde{\mathbf{P}}_{\mathbf{W}} = \mathbf{M}(\xi) \cdot \tilde{\mathbf{P}}_{\mathbf{W}}$$

- **Linear model in 11 parameters** ($\mathbf{M}_{3x4}$, up to scale)
  - **only 9 params are independent:**

$$\xi = [\mathbf{R}, \quad \mathbf{T}, \quad f, \quad \mathbf{C}] = [\varphi, \vartheta, \rho, \quad t_X, t_Y, t_Z, \quad f, \quad x_C, y_C]$$

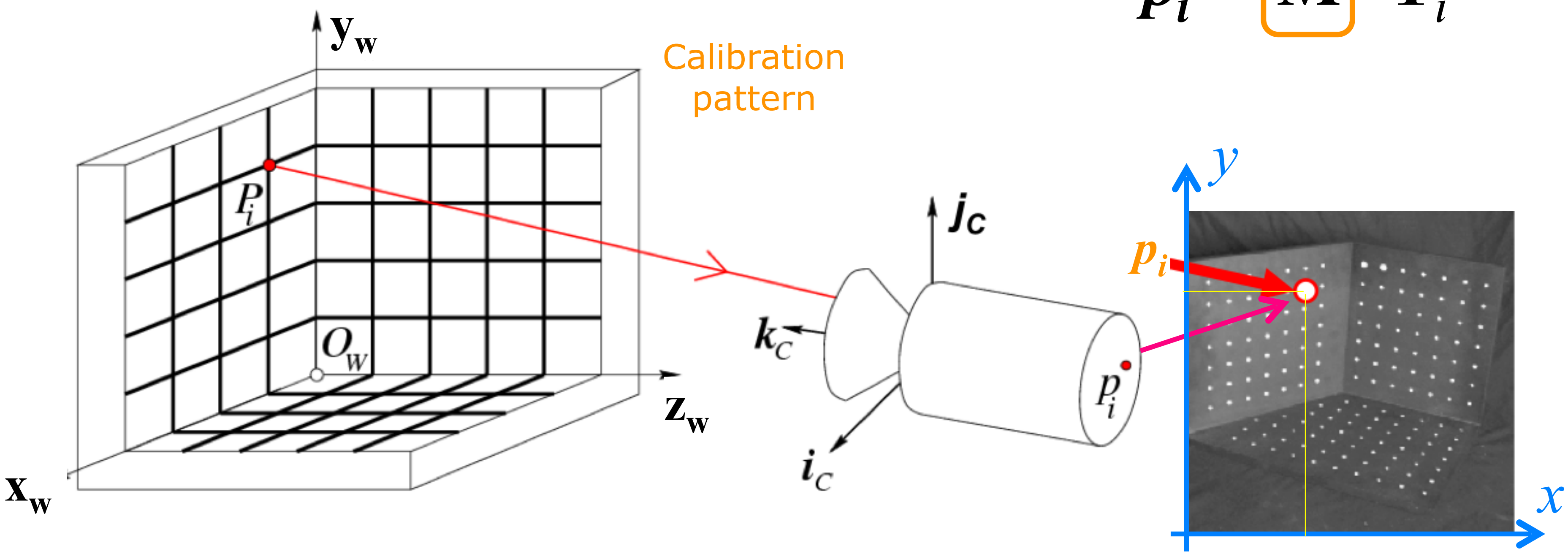
- **Extrinsic Parameters:** depend on the relative position <u>camera-scene</u>
  - **Rotation**: Euler angles: $\mathbf{R} = [\varphi, \theta, \rho]$
  - **Translation**: translation vector: $\mathbf{T} = [t_X, t_Y, t_Z]$
- **Intrinsic Parameters:** depend on the camera characteristics
  - **Focal length**: f
  - **Optical Centre position**: $C = \langle x_C, y_C \rangle$

Camera calibration – Setup

- **Calibration pattern**: set of fiducial points (easy to be accurately located)
- $\boldsymbol{P_i}$: World coordinates of the f.p.
  - *Expressed wrt a reference system integral with the calibration pattern*
  - *$\boldsymbol{P_i}$ a priori known*
- $\mathbf{p_i}$: Image coordinates of the f.p.
  - *$\boldsymbol{p_i}$ determined by analyzing the image*

Calibration: exploits the association $P_i \rightarrow p_i$ to determine M

$$\tilde{p}_i = \boxed{\mathbf{M}} \bullet \tilde{P}_i$$

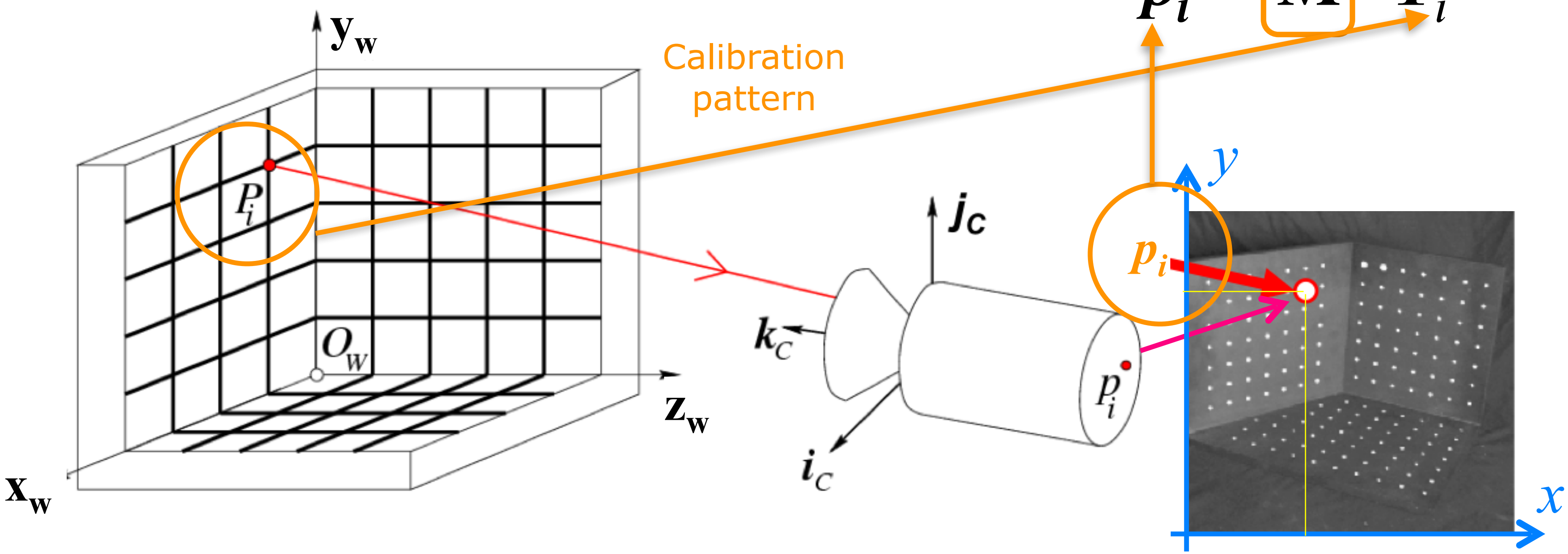


Camera calibration – Setup

- **Calibration pattern**: set of fiducial points (easy to be accurately located)
- ***P*$_i$** : World coordinates of the f.p.
  - *Expressed wrt a reference system integral with the calibration pattern*
  - ***P*$_i$** *a priori known*
- **p$_i$** : Image coordinates of the f.p.
  - ***p*$_i$** *determined by analyzing the image*

Calibration: exploits the association $P_i \rightarrow p_i$ to determine M

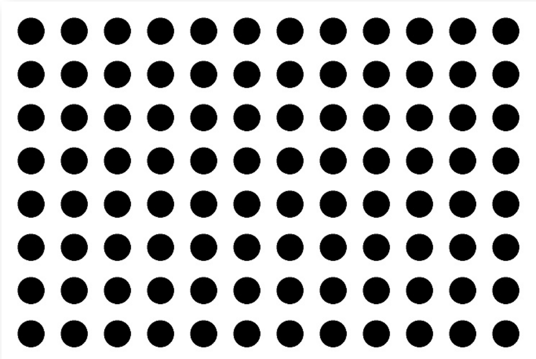
$$\tilde{p}_i = \boxed{\mathbf{M}} \bullet \tilde{P}_i$$



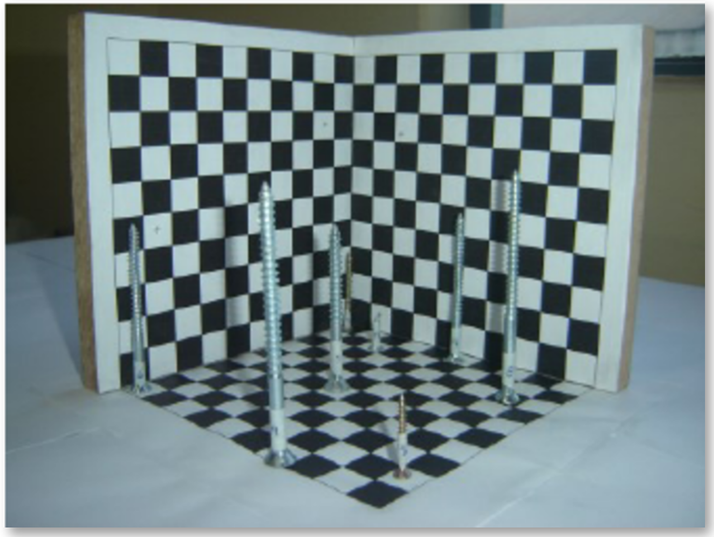
Camera calibration – Calibration pattern

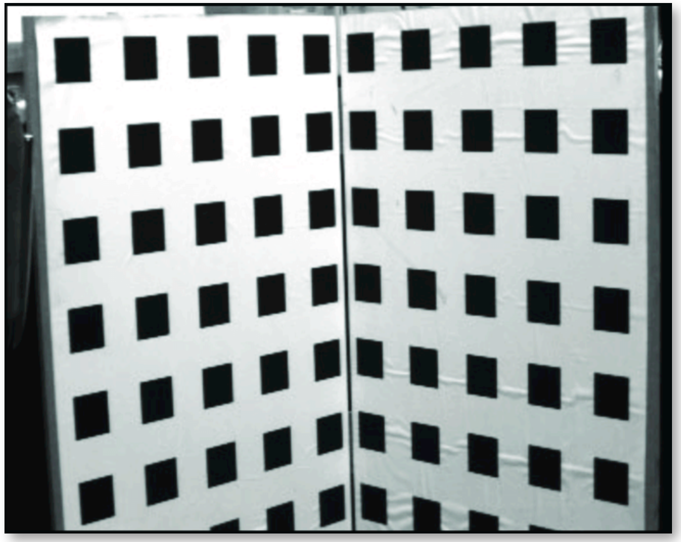
Set of fiducial points: easy to locate, with high precision

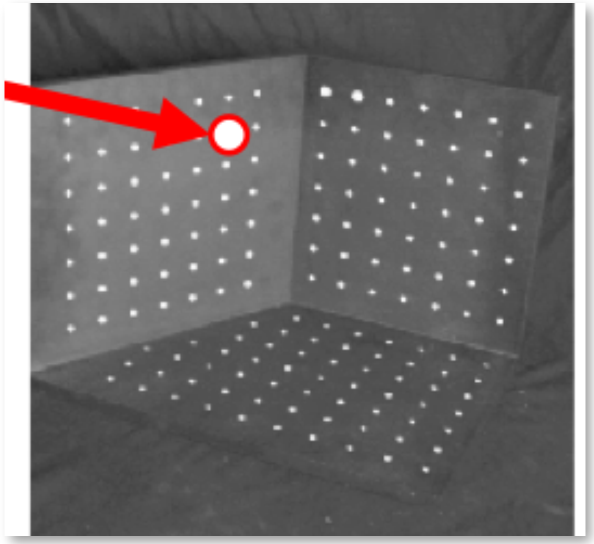
PATTERN:		***Fiducial Point:***
➢ *spheres*	→	*sphere centre*
➢ *circles*	→	*circle centre*
➢ *chessboard*	→	*square vertices*



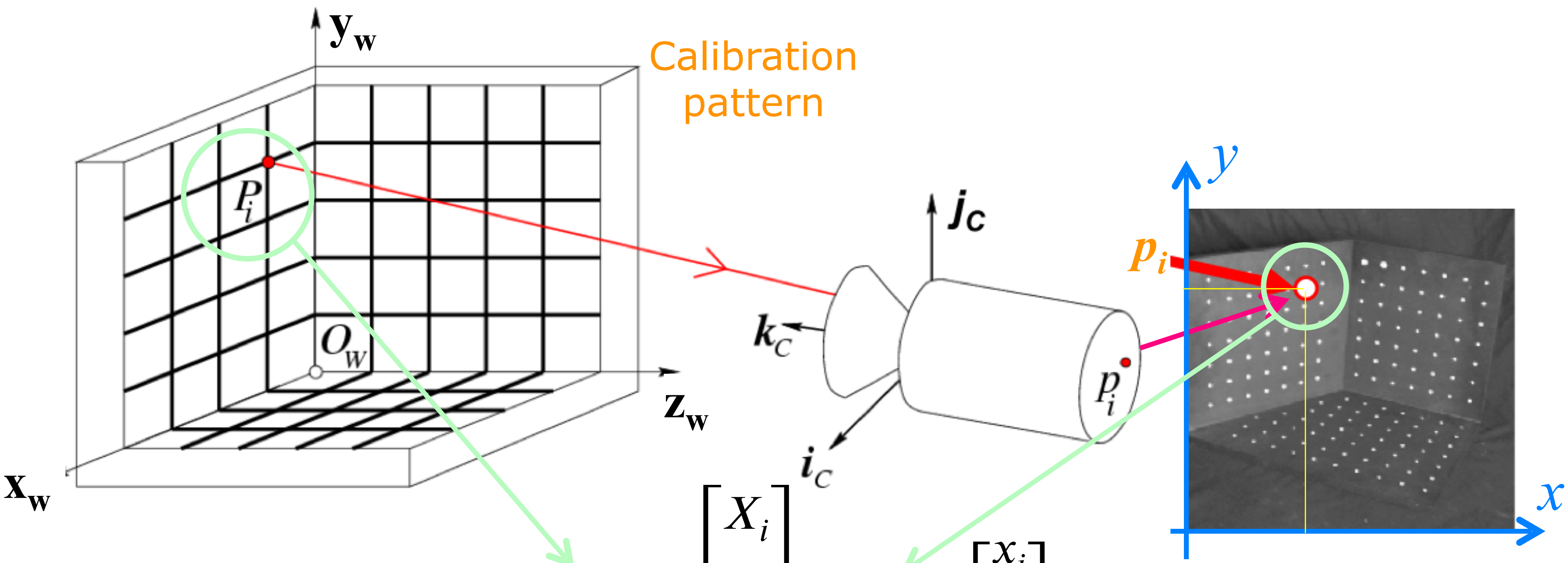
- Fiducial points over a non-degenerate 3D-space
  - ➢ If I use planar patterns, I need at least 2 images, on different planes







Camera calibration – Problem Definition



$$\tilde{P}_i = \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}, \quad \tilde{p}_i = \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

Problem definition:

- Given a set of ***N*** fiducial points ***P_i***, of known 3D world position
- and given the corresponding image-coordinates ***p_i***

➔ determine the camera model **M** (function of ξ) such that:

$$\tilde{p}_i = \mathbf{M} \bullet \tilde{P}_i \,, \quad i = 1..N$$

Camera calibration - Linear approach

Determine the matrix $\mathbf{M}$ [3 x 4] of the <u>linear model</u> $\tilde{p}_i = \mathbf{M} \bullet \tilde{\mathbf{P}}_i$ given:

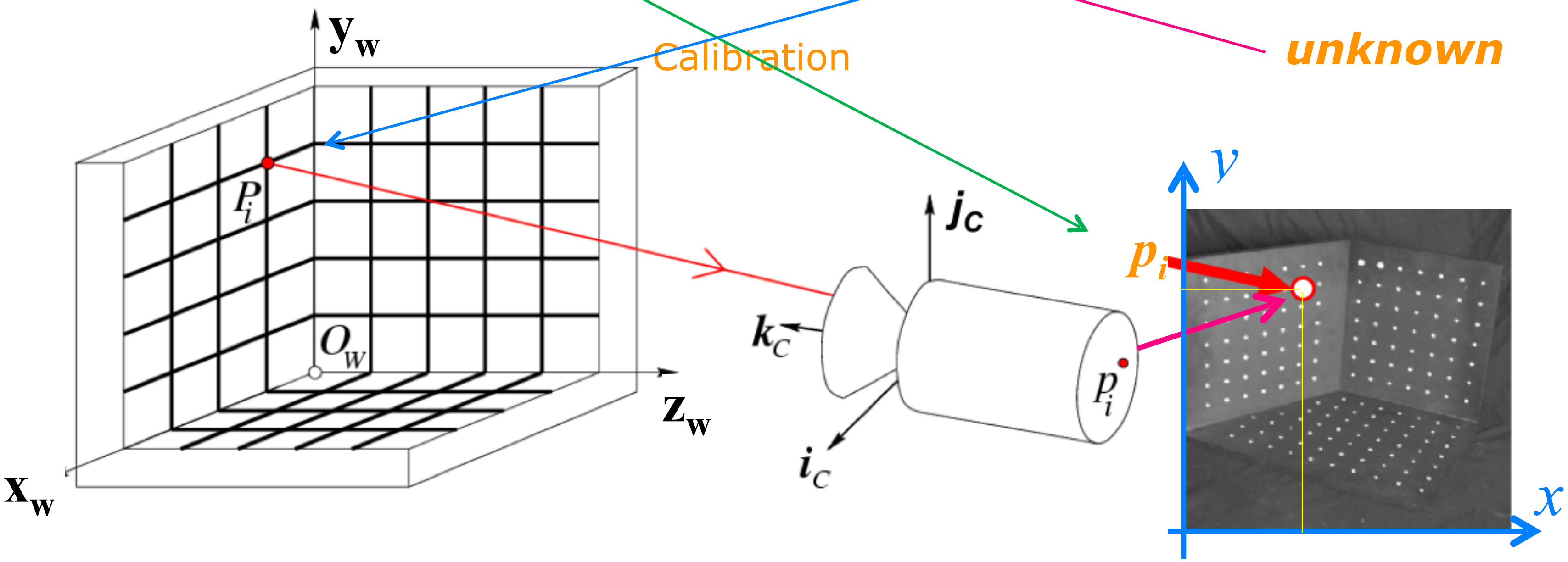
- the coords-World : $P_i, \quad i = 1..N$
- the coords-image: $p_i, \quad i = 1..N$

For each i=1..N:

***N* equations:** $$\tilde{p}_i = \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \mathbf{M} \bullet \tilde{\mathbf{P}}_i = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix} \bullet \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}, \qquad i = 1 \cdots N$$

Calibration

unknown



Camera calibration - Linear approach

For a pair $< P_i, p_i >$:

$$\tilde{\mathbf{p}}_i = \begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \\ \tilde{z}_i \end{bmatrix} = \mathbf{M} \cdot \tilde{\mathbf{P}}_\mathbf{W} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix} \cdot \tilde{\mathbf{P}}_\mathbf{i} = \begin{bmatrix} \mathbf{m}_1 \cdot \tilde{\mathbf{P}}_\mathbf{i} \\ \mathbf{m}_2 \cdot \tilde{\mathbf{P}}_\mathbf{i} \\ \mathbf{m}_3 \cdot \tilde{\mathbf{P}}_\mathbf{i} \end{bmatrix} \qquad \text{(eq. 1)}$$

- Remember Euclidean vs Homogeneous coords:

$$\mathbf{p}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \dfrac{\tilde{x}_i}{\tilde{z}_i} \\ \dfrac{\tilde{y}_i}{\tilde{z}_i} \end{bmatrix} \qquad \text{(eq. 2)}$$

- Considering $\mathbf{p}_i$ (in Euclidean coords) and combining (eq. 1) and (eq. 2) we obtain:

$$\mathbf{p}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \dfrac{\tilde{x}_i}{\tilde{z}_i} \\ \dfrac{\tilde{y}_i}{\tilde{z}_i} \end{bmatrix} = \begin{bmatrix} \dfrac{\mathbf{m}_1\tilde{\mathbf{P}}_i}{\mathbf{m}_3\tilde{\mathbf{P}}_i} \\ \dfrac{\mathbf{m}_2\tilde{\mathbf{P}}_i}{\mathbf{m}_3\tilde{\mathbf{P}}_i} \end{bmatrix} \implies \boxed{\begin{cases} \mathbf{m}_1\tilde{\mathbf{P}}_i - x_i\,\mathbf{m}_3\tilde{\mathbf{P}}_i = 0 \\ \mathbf{m}_2\tilde{\mathbf{P}}_i - y_i\,\mathbf{m}_3\tilde{\mathbf{P}}_i = 0 \end{cases},}$$

Tablet

Camera calibration - linear approach

For a pair $< P_i, p_i >$:

$$p_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \frac{\tilde{x}_i}{\tilde{z}_i} \\ \frac{\tilde{y}_i}{\tilde{z}_i} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{m}_1 \tilde{\mathbf{P}}_i}{\mathbf{m}_3 \tilde{\mathbf{P}}_i} \\ \frac{\mathbf{m}_2 \tilde{\mathbf{P}}_i}{\mathbf{m}_3 \tilde{\mathbf{P}}_i} \end{bmatrix} \Longrightarrow$$

$$\begin{cases} \mathbf{m}_1 \tilde{\mathbf{P}}_i - x_i \, \mathbf{m}_3 \tilde{\mathbf{P}}_i = 0 \\ \mathbf{m}_2 \tilde{\mathbf{P}}_i - y_i \, \mathbf{m}_3 \tilde{\mathbf{P}}_i = 0 \end{cases}, \quad i = 1..N$$

2 equations, **12** unknown m_{ij}

- In matricial form:

$$\begin{bmatrix} P_{1x} & P_{1y} & P_{1z} & 1 & 0 & 0 & 0 & 0 & -x_1 P_{1x} & -x_1 P_{1y} & -x_1 P_{1z} & -x_1 \\ 0 & 0 & 0 & 0 & P_{1x} & P_{1y} & P_{1z} & 1 & -y_1 P_{1x} & -y_1 P_{1y} & -y_1 P_{1z} & -y_1 \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Camera calibration - Linear approach

For N pair $< P_i, p_i > \quad (i = 1..N):$

- write the **2N** eq as a linear system in the 12 unknowns m_{ij}

$$\begin{bmatrix} P_{1x} & P_{1y} & P_{1z} & 1 & 0 & 0 & 0 & 0 & -x_1 P_{1x} & -x_1 P_{1y} & -x_1 P_{1z} & -x_1 \\ 0 & 0 & 0 & 0 & P_{1x} & P_{1y} & P_{1z} & 1 & -y_1 P_{1x} & -y_1 P_{1y} & -y_1 P_{1z} & -y_1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & & \cdots & \cdots & \cdots & \cdots & \cdots \\ P_{Nx} & P_{Ny} & P_{Nz} & 1 & 0 & 0 & 0 & 0 & -x_N P_{Nx} & -x_N P_{Ny} & -x_N P_{Nz} & -x_N \\ 0 & 0 & 0 & 0 & P_{Nx} & P_{Ny} & P_{Nz} & 1 & -y_N P_{Nx} & -y_N P_{Ny} & -y_N P_{Nz} & -y_N \end{bmatrix} \bullet \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}$$

$$\mathbf{P}_{[2N\times 12]} \bullet \mathbf{m}_{[12\times 1]} = \mathbf{0}_{[2N\times 1]}$$

$\rightarrow \quad \mathbf{P} \cdot \mathbf{m} = \mathbf{0}$

Homogeneous linear system, in 11 unknowns (12, up to a scale factor) and 2N equations resolvable for N = 6 (at least 6 fiducial points)

Scale of Projection Matrix

- REMEMBER:
  - Projection Matrix M acts on homogeneous coords, i.e.:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} \equiv k \cdot \begin{bmatrix} u \\ v \\ w \end{bmatrix} \quad (k \neq 0 \text{ is any constant})$$

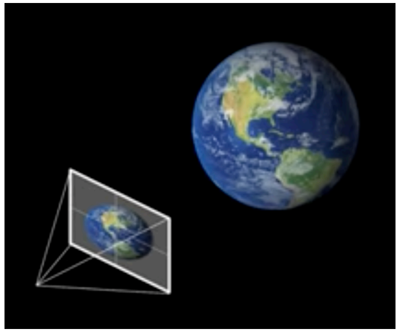
  - That is:

$$\tilde{P} \cdot m = \tilde{P} \cdot (k \cdot m)$$

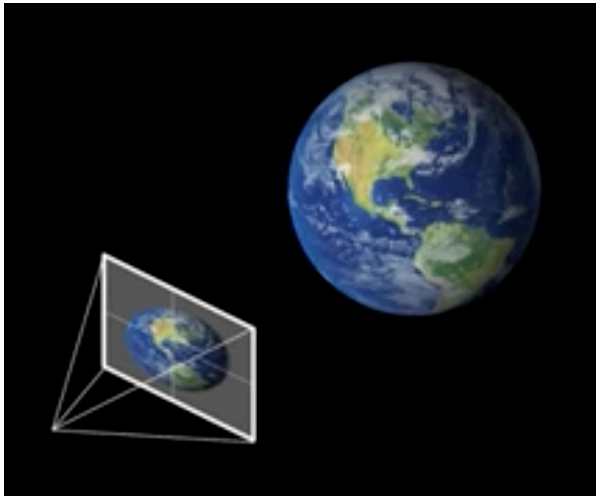
- so that: the projection matrices M and $(k \cdot M)$ produce the same homogeneous pixel coordinates

Projection matrix M is defined up to a scale factor

Scale Projection Matrix



Scale k_1



Scale k_2

Scaling projection matrix, implies simultaneously scaling the world and camera, which does not change the image

we can set projection matrix to some **arbitrary scale**

Least squares solution for m

- Option 1: set scale so that $m_{34} = 1$
- Option 2: set scale so that $\boxed{\|m\|^2 = 1}$
- We want: $\tilde{P} \cdot m = 0$, <u>and</u> $\|m\|^2 = 1$
- Formulated with the constrained least squares problem:

$$\min_m \|\tilde{P}m\|^2, \quad s.t. \quad \|m\|^2 = 1,$$

$$\min_m \|m^T \tilde{P}^T \tilde{P} m\|, \quad s.t. \quad \|m^T m\| = 1$$

- Let's define the Loss function $L(m, \lambda)$:

$$L(m, \lambda) = m^T \tilde{P}^T \tilde{P} m - \lambda(m^T m - 1)$$

Tablet

- We want to minimize L wrt m

Constrained Least squares solution

- Let's take the derivatives of $L(m,\lambda)$ wrt m and set it to 0:

$$2\tilde{P}^T\tilde{P}m - 2\lambda m = 0$$

- equivalent to solve the **eigenvalue problem**:

$$\boxed{\tilde{P}^T\tilde{P}m = \lambda m}$$

➡ **Eigenvector** m corresponding to the smallest eigenvalue λ of the matrix $\tilde{P}^T\tilde{P}$ minimizes the loss function $L(m,\lambda)$

➡ or equivalently, m is the **singular vector** corresponding to the minimum singular value (not null) using the Singular Value Decomposition **(SVD)** of $\tilde{P}$

Camera calibration - Linear approach

Given the vector m we have to:

- rearrange it to form the projection matrix M (4×4)
- it remains to determine the intrinsic and extrinsic matrices:

$$M = \begin{bmatrix} \boxed{\begin{matrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{matrix}} & \begin{matrix} m_{14} \\ m_{24} \\ m_{34} \end{matrix} \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{M_{int}} \cdot \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{M_{ext}}$$

- it holds that:

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = KR$$

- Given K: upper triangular matrix, R is an orthonormal matrix, it is possible to <u>decouple K and R from their product</u> using the **QR factorization** method from linear algebra

Camera calibration - Linear approach

- It remains to determine the translation vector $\mathbf{t}$ of the extrinsic matrix
- Given:

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{M_{int}} \cdot \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & \boxed{t_x} \\ r_{21} & r_{22} & r_{23} & \boxed{t_y} \\ r_{31} & r_{32} & r_{33} & \boxed{t_z} \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{M_{ext}}$$

- it holds that:

$$\begin{bmatrix} m_{14} \\ m_{24} \\ m_{34} \end{bmatrix} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = K\mathbf{t}$$

- and inverting we have:

$$\mathbf{t} = K^{-1} \cdot \begin{bmatrix} m_{14} \\ m_{24} \\ m_{34} \end{bmatrix}$$

Camera Calibration: Whang algorithm (2000)

"A_flexible_new_technique_for_camera_calibration" in ***IEEE Transaction on Pattern Analysis and Machine Intelligence***, vol. 22, no. 11, pp. 1330-1334, 2000.

- Planar pattern in at least 2 views (chessboard)

- **Hypothesis**:
  - *The pattern dimensions are known*
  - *Trick: the scene reference system is joint to the chessboard (different extrinsic parameters for each photo, while shared intrinsic parameters)*

Camera calibration - Non Linear approach

- To be used in presence of radial distortion

$$\tilde{\mathbf{p}}_i = \begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \\ \tilde{z}_i \end{bmatrix} = \mathbf{M}(\xi) \cdot \tilde{\mathbf{P}}_\mathbf{i} \quad \rightarrow \quad \tilde{\mathbf{p}}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = f(\xi, \mathbf{P}_\mathbf{i})$$

linear *non linear*

Algorithm:

1. Solve the linear model

➔ first estimate of the linear parameters: ξ_0

2. Non linear optimization starting from ξ_0 to minimize the reprojection error E

➔ Final model: $\hat{\xi}$

$$E = \sum_i \left\| \mathbf{p}_\mathbf{i}^{MEAS} - \mathbf{p}_\mathbf{i}^{EST} \right\|^2 = \sum_i \left\| \mathbf{p}_\mathbf{i}^{MEAS} - f(\xi, \mathbf{P}_\mathbf{i}) \right\|^2 \quad \rightarrow \quad \hat{\xi} \quad t.c. \quad \hat{\xi} = \underset{\xi}{\operatorname{argmin}}(E)$$

Camera calibration - Non Linear approach

More specifically:

$$E = \sum_i \left\| \mathbf{p}_i^{MEAS} - \mathbf{p}_i^{EST} \right\|^2 = \sum_i \left\| \mathbf{p}_i^{MEAS} - f\left(\xi, \mathbf{P}_i\right) \right\|^2 \rightarrow \hat{\xi} \quad t.c. \quad \hat{\xi} = \underset{\xi}{\operatorname{argmin}}\left(E\right)$$

$$Given\ the\ initial\ model: \ \xi_0 = \left\{\mathbf{R}, \mathbf{t}, f, x_C, y_C, k_D\right\} \quad and \quad N\ fiducial\ points\ P_i :$$

$$\xi = \xi_0$$

$$p_i^{EST} = f(\xi\ , \ P_i) :$$

$$\begin{cases} P_i \longmapsto P_{cam,i} = \mathbf{R} \bullet P_i + \mathbf{t}\ , \quad i = 1..N \\ P_{cam,i} \longmapsto p_{im,i} = \begin{bmatrix} x_C \\ y_C \end{bmatrix} + \dfrac{f}{z_{cam,i}} \begin{bmatrix} x_{cam,i} \\ y_{cam,i} \end{bmatrix}, \quad i = 1..N \\ p_{im,i} \longmapsto p_i^{EST} = p_{im,i}\left(1 + k_{D1} r^2 + k_{D2} r^4\right), \quad i = 1..N \end{cases}$$

$$error: \quad E(\xi) = \sum_{i=1}^{N} \left\| p_i^{MEAS} - p_i^{EST} \right\|^2 \longrightarrow \boldsymbol{new\ estimate} \quad \xi$$

Lab time

- `syntLinearCalibration`
- `RealCalibration`