SINGLE VIEW GEOMETRY

computer vision

ABOUT THIS CLASS

- very basic elements of projective geometry :
  - the projective plane and homogeneous coordinates
  - points, lines in homogeneous coordinates
- the geometry of image acquisition
  - single view (finite) camera models
  - camera calibration

(A FEW SLIDES ON) PROJECTIVE GEOMETRY

computer vision

POINTS

- we first consider points on a plane $(x, y) \in IR^2$
- considering the plane as a vector space a point is identified by a vector
- Convention: points will be represented as column vectors $\boldsymbol{x} = (x,y)^\top$

LINES

$$ax + by + c = 0$$

- A line may also be represented as a vector $(a, b, c)^\top$
- The line $(ka)x + (kb)y + kc = 0$ is the same line, for each $k \neq 0$
- then vectors $(ka, kb, kc)^\top \quad \forall k \neq 0$ can be seen as equivalent (belong to the same **equivalence class**)
- such equivalence class is called a **homogeneous vector**

PROJECTIVE PLANE

- the projective plane $\mathcal{P}^2$ or 2D projective space is formed by any vector $(a, b, c)^\top$ representative of an equivalence class:

$$IR^3 - (0, 0, 0)^\top$$

-

LINES AND POINTS

- A point $(x, y)^\top$ lies on the line $\mathbf{l} = (a, b, c)^\top$ iff $ax + by + c = 0$
- This may be written in terms of an inner product:

$$(x, y, 1)(a, b, c)^\top = \mathbf{x}^\top \mathbf{l} = 0$$

- A point on the Euclidean plane may be represented as a point in the projective space (a 3D vector with a final 1).
- Notice that here again points define equivalence classes

LINES AND POINTS

- The point **x** lies on the line **l** iff $\mathbf{x}^\top \mathbf{l} = 0$
- <u>Intersection of lines:</u> $\mathbf{x} = \mathbf{l} \times \mathbf{l}'$ (since the intersection is a point lying on both lines, thus a point whose inner product with both lines is 0...)
- <u>Lines joining points:</u> $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$ (as above)

IDEAL POINTS AND LINES AT INFINITY

- What is the intersection of parallel lines?

$$
\begin{array}{ll}
ax + by + c = 0 & \mathbf{l} = (a, b, c)^\top \\
ax + by + c' = 0 & \mathbf{l}' = (a, b, c')^\top
\end{array}
$$

$$\mathbf{x} = \mathbf{l} \times \mathbf{l}' = (b, -a, 0)^\top$$

SUMMING UP

- Points $\mathbf{x} = (x_1, x_2, x_3)^\top$ so that $x_3 \neq 0$ correspond to finite points on the Euclidean plane $(x_1/x_3, x_2/x_3)^\top$
- Augmenting the Euclidean plane by adding points with $x_3 = 0$ produces the set of all homogeneous vectors
- Points with $x_3 = 0$ are known as ideal points or points at infinity
- All points at infinity lie on a single line $\mathbf{l}_\infty = (0, 0, 1)^\top$

SUMMING UP

- a line $(a, b, c)^\top$ intersects $\mathbf{l}_\infty$ in the ideal point $(b, -a, 0)^\top$
- the line $(a, b, c')^\top$ intersects $\mathbf{l}_\infty$ in the same point
- in inhomogeneous notation $(b, -a)^\top$ is a vector tangent to the line representing the line's direction
- for these reasons the line at infinity can be thought of as the set of directions of lines in the plane

PROJECTIVE GEOMETRY

- points at infinity simplify intersection properties
- in the projective plane
  - two distinct lines meet at a single point
  - two distinct points lie on the same line
- the geometry of the projective plane is known as **projective geometry**

THE DUALITY PRINCIPLE

- to any theorem of 2D projective geometry there corresponds a dual theorem, which may be derived by interchanging the roles of lines and points in the original theorem

PROJECTIVE TRANSFORMATIONS

(an anticipation…)

- A homography is a linear transformation of homogeneous vectors represented by a non-singular 3x3 matrix H

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

$$\mathbf{x}' = H\mathbf{x}$$

A HIERARCHY OF TRANSFORMATIONS

all linear in homogeneous coordinates!

- 1. isometries
- II. similarities
- III. affinities
- IV. projectivities

PROJECTIVE SPACE

- A point X in 3D space is represented in homogeneous coordinates as a 4-vector

- A projective transformation of the 3D space is a 4x4 non singular matrix, with 15 d.o.f.

PROJECTIVE TRANSFORMATIONS

- since homogeneous points define equivalence classes, *projective transformations are unique up to a scale factor*
- x=Hx’
- x=3Hx’
- …
- x=aHx’ for each non zero constant a

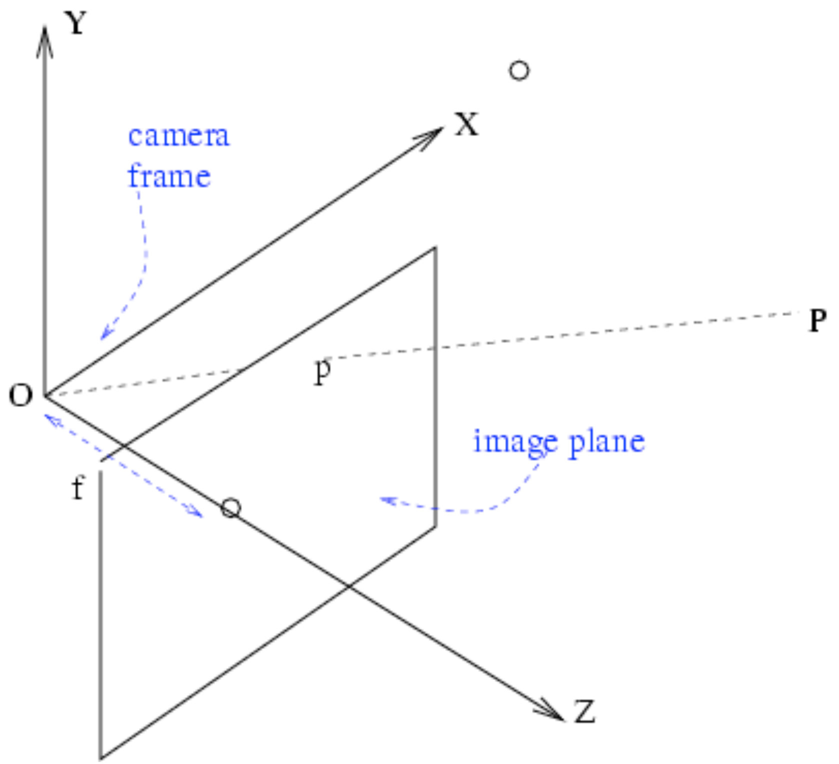
FROM SPACE TO PLANE

- a perspective projection can be represented as a projective transformation $x=PX$
- not true the vice-versa…

CAMERA MODELS
AND
CAMERA PARAMETERS

BASIC PIN HOLE CAMERA

- a camera is a mapping between the 3D world and a 2D image



$$x = f\frac{X}{Z}$$

$$y = f\frac{Y}{Z}$$

in matrix form (homogeneous coordinates)

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

more compactly: $\mathbf{x} = \mathrm{P}\ \mathbf{X}$ $\quad \mathrm{P} = \mathrm{diag(f,\ f,\ 1)}\ [\mathrm{I}\ ;\ \mathbf{0}]$

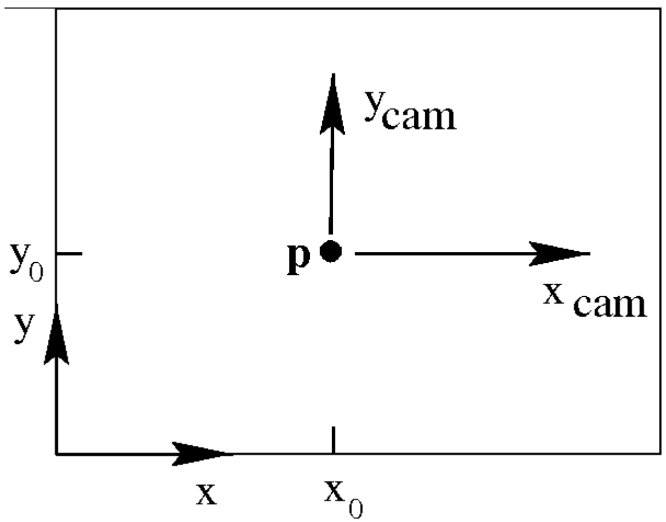
GENERAL MODEL

- a general model for a camera should take into account:
  - rigid transformation between the camera and the 3D world reference frames
  - the transformation from metric coordinates to pixels coordinates

INTRINSIC PARAMETERS

- in the basic model we assumed the origin of the image plane to be the principal point

- in practice this will not be
  - we may add a translation of the points in the image plane



$$\mathbf{x} = \mathrm{P}\ \mathbf{X} \qquad \mathrm{P} = \mathrm{K}\ [\mathrm{I}\ ;\ \mathbf{0}]$$

$$K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

INTRINSIC PARAMETERS

- in images units are expressed in pixels
- then the matrix K of intrinsic parameters becomes

$$K = \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

all entries in pixel coordinates

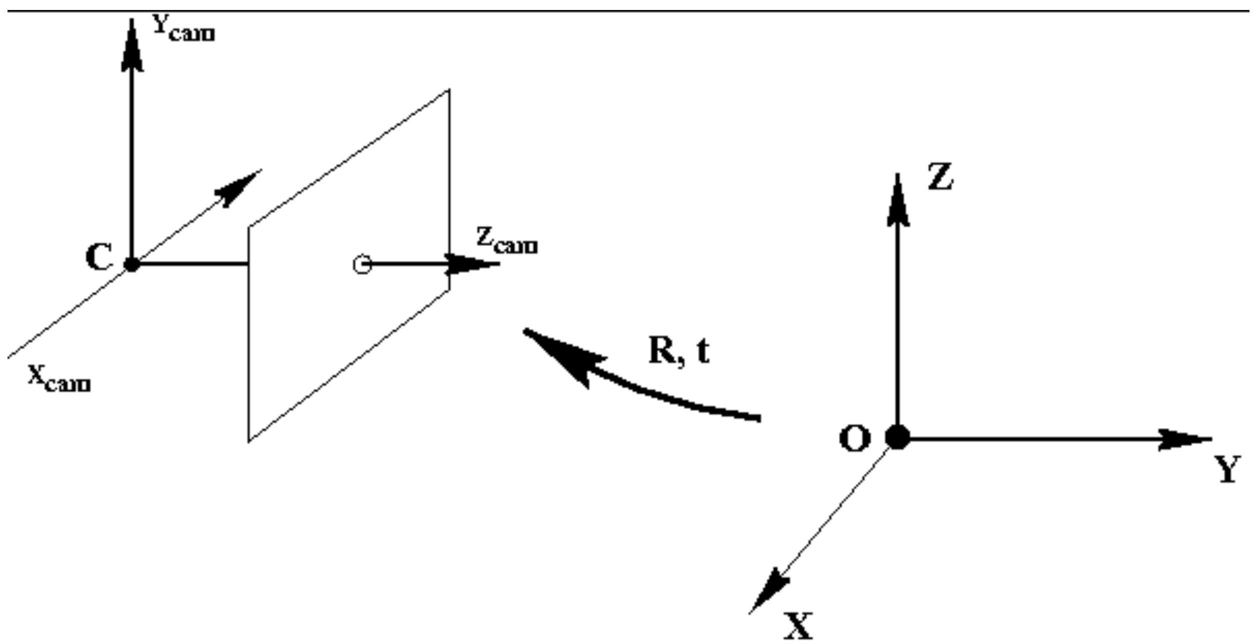
- with

$$\alpha_x = f m_x$$
$$\alpha_y = f m_y$$
$$x_0 = m_x p_x$$
$$y_0 = m_y p_y$$

number of pixels per unit distance

EXTRINSIC PARAMETERS

- world reference system in general does not coincide with the camera reference frame



$$G = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}$$

$$P = K[I; \mathbf{0}]G$$

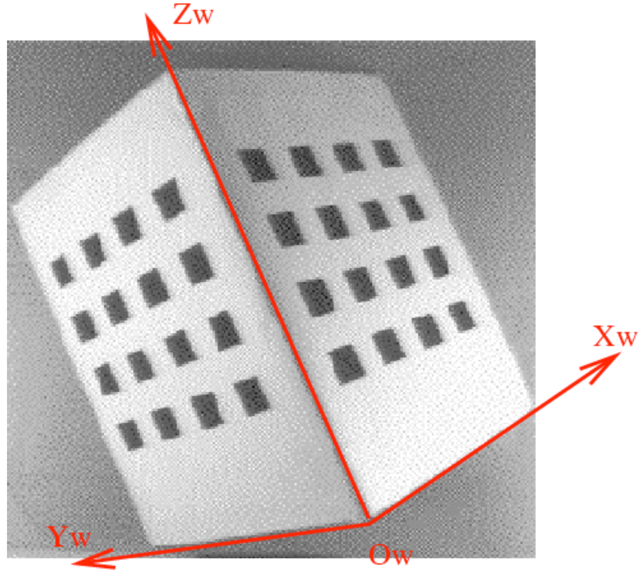
equivalent formulation

$$P = K[R; \mathbf{t}]$$

CAMERA CALIBRATION (WORLD & IMAGE PLANE)

- we start off from 2D - 3D correspondences:

$$\{(\mathbf{x}_i, \mathbf{X}_i)\}_{i=1}^{N}$$



P has 11 d.o.f. then we need N>=6

1. We may first estimate the entries of P

2. and then derive the internal and external parameters

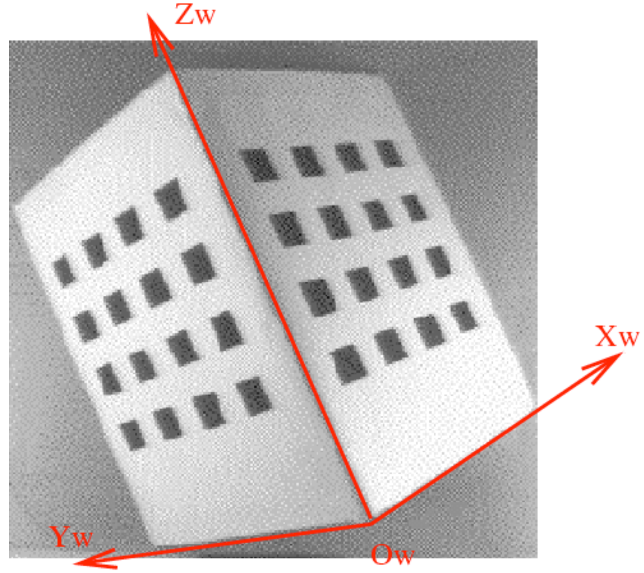
CAMERA CALIBRATION (WORLD & IMAGE PLANE)

- we start off from 2D - 3D correspondences: $\{(\mathbf{x}_i, \mathbf{X}_i)\}_{i=1}^{N}$

- ***P* has 11 d.o.f. then *N*>=6**

- the position of 3D points is known in advance (measured on the pattern)

- the position of 2D points may be estimated for instance by:
  1. canny edge detection
  2. straight line fitting
  3. intersecting the lines to obtain the imaged corners



with a careful implementation we may reach very good sub-pixel precisions

CAMERA CALIBRATION (WORLD & IMAGE PLANE)

- The main steps of a standard procedure (details in the additional material - next slides)

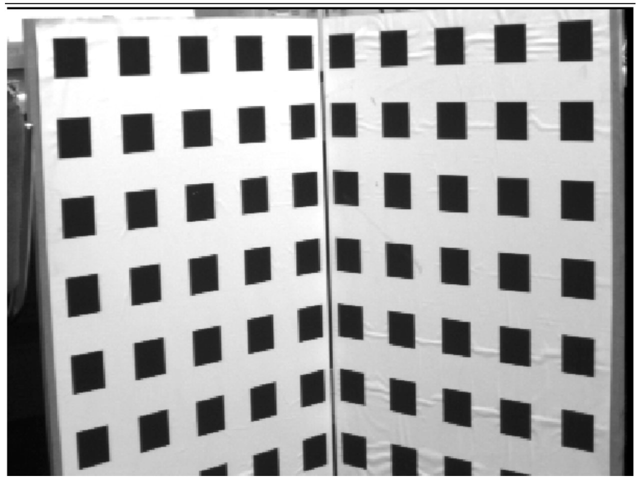
  1. Estimate the projection matrix P from world-image points correspondences

  2. Extract parameters from the estimated P

[advanced material]

1. ESTIMATION OF THE PROJECTION MATRIX

- we start off from 2D - 3D correspondences: $\{(\mathbf{x}_i, \mathbf{X}_i)\}_{i=1}^{N}$
- our goal is to estimate the entries of matrix P (3 x 4): $\mathbf{x}_i = P\mathbf{X}_i$



$$\mathbf{x}_i \times P\mathbf{X}_i = \mathbf{0}$$

$$\mathbf{x}_i = (x_i, y_i, w_i)^\top$$

$$P = \begin{bmatrix} \mathrm{P}^1 \\ \mathrm{P}^2 \\ \mathrm{P}^3 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{0}^\top & -w_i\mathbf{X}_i^\top & y_i\mathbf{X}_i^\top \\ w_i\mathbf{X}_i^\top & \mathbf{0}^\top & -x_i\mathbf{X}_i^\top \\ -y_i\mathbf{X}_i^\top & x_i\mathbf{X}_i^\top & \mathbf{0}^\top \end{bmatrix} \begin{pmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \mathbf{P}^3 \end{pmatrix} = \mathbf{0}$$

12th dim vector of unknowns

[advanced material]

1. ESTIMATION OF THE PROJECTION MATRIX DIRECT LINEAR TRANSFORM (DLT) ALGORITHM

we build a homogeneous system of equations $\mathbf{A}\mathbf{p} = \mathbf{0}$

$$A = \begin{bmatrix} \mathbf{0}^\top & -w_1\mathbf{X}_1^\top & y_1\mathbf{X}_1^\top \\ w_1\mathbf{X}_1^\top & \mathbf{0}^\top & -x_1\mathbf{X}_1^\top \\ \dots & \dots & \dots \\ \mathbf{0}^\top & -w_N\mathbf{X}_N^\top & y_N\mathbf{X}_N^\top \\ w_N\mathbf{X}_N^\top & \mathbf{0}^\top & -x_N\mathbf{X}_N^\top \end{bmatrix}$$

$$\mathbf{p} = \begin{bmatrix} P_{11} \\ P_{12} \\ \vdots \\ P_{33} \\ P_{34} \end{bmatrix} = \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \mathbf{P}^3 \end{bmatrix}$$

$$P = \begin{bmatrix} P^1 \\ P^2 \\ P^3 \end{bmatrix}$$

[advanced material]

1. ESTIMATION OF THE PROJECTION MATRIX DIRECT LINEAR TRANSFORM (DLT) ALGORITHM

- P has 11 degrees of freedom (homogeneous transformation)
- with at least 6 correspondences we may obtain a solution
  - with 11 equations or in noise free cases rank(A)= 11
  - the non trivial solution may be found in the null space of A
- if rank(A) is full we can look for a least squares solution

$$\text{minimize} \quad ||A\mathbf{x}||$$
$$\text{subj to} \quad ||\mathbf{x}|| = 1$$

it avoids the trivial solution

DLT SOLUTION OF A GENERAL PROBLEM [additional material]

$$\text{minimize} \quad ||A\mathbf{x}||$$
$$\text{subj to} \quad ||\mathbf{x}|| = 1$$

- Let $A = UDV^\top$
- Equivalent problem $||DV^\top\mathbf{x}||$ subj to $||V^\top\mathbf{x}|| = 1$

$\mathbf{y} = V^\top\mathbf{x}$

- Change of variable $||D\mathbf{y}||$ subj to $||\mathbf{y}|| = 1$
- D is diagonal with elements in descending order, thus a solution that minimizes the norm is $\mathbf{y} = (0, 0, \ldots, 1)^\top$

$\mathbf{x} = V\mathbf{y}$

- thus the solution **x** is the last column of V

[advanced material]

2. MATRIX DECOMPOSITION

- the *P* we estimate is a 3x4 which we may write as $P = [M|\mathbf{p}_4]$
- we now want to extract the intrinsic and extrinsic parameters
- we know that $P = K[R| - R\tilde{C}]$
- R, K can be derived from a QR decomposition of M
- the center can be computed as the null space of *P*

THE CAMERA CENTRE

[additional material]

- *P* has a **1**-dim right null space [rank(*P*)=3 but *P* has 4 columns]
- therefore there exists a (non trivial) vector C, so that PC=0
- consider a line containing C and any other point A in 3-space
  In parametric coordinates:

$$\mathbf{X}(\lambda) = \lambda A + (1 - \lambda) C$$

$$\mathbf{x} = P\mathbf{X}(\lambda) = \lambda PA + (1 - \lambda) PC = \lambda PA$$ line projection

- all points are mapped to the same image point *PA* which means that the line is a ray through the camera centre
- since this holds for every *A* it follows that C must be the camera centre in homogeneous coordinates

[additional material]

RESTRICTED CAMERA ESTIMATION

- the decomposition method we describe does not take into account some "reasonable" assumptions on the internal parameters which may be
  - skew is 0
  - pixels are squared
  - principal point is known
  - *K is all known*

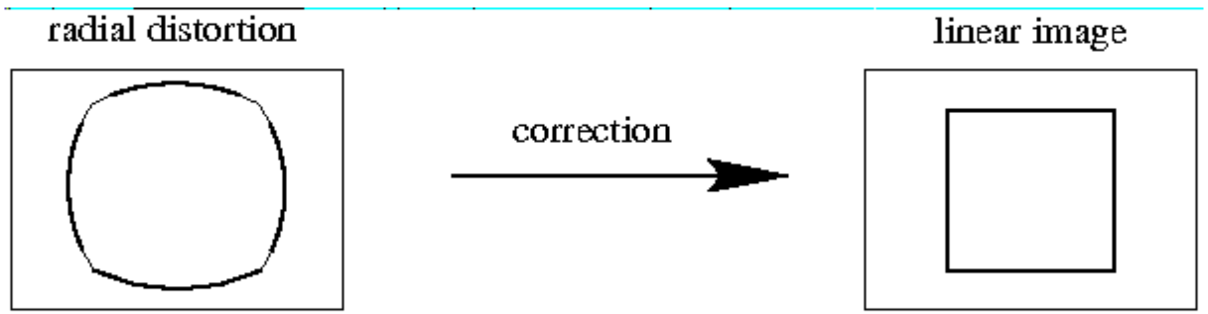
- *there exist algorithms that allow us to impose contraints*

RADIAL DISTORTION **[additional material]**

- **the linear camera model discussed so far is not perfect for the real world**
- short focal length and cheap cameras show the non linearity quite clearly

long focal length

short focal length



radial distortion correction allows us to obtain a correded image "as if" generated by a linear model

RADIAL DISTORTION

[additional material]

distortion centre

measured

$$\hat{x} = x_c + L(r)(x - x_c)$$
$$\hat{y} = y_c + L(r)(y - y_c)$$

$$r^2 = (x - x_c)^2 + (y - y_c)^2$$

corrected

distortion function

- *L may be written as a polynomial approximation*

$$L(r) = 1 + k_1 r + k_2 r^2 + k_3 r^3 + \ldots$$

- *the coefficients may be included in the calibration process (including the distortion centre which is not necessarily the principal point)*

WRAPPING UP

- Camera calibration is the process of estimating camera parameters from "ad hoc" images

- It allows us to (geometrically) relate the images acquired from the camera with the real 3D world

- Many available algorithms…

  - today we discussed a classical procedure (estimate P and then extract the parameters)