Prof. Kjersti Engan

ELE510 Image processing and computer vision

Camera and calibration, (chap 13.4, 13.5 Birchfield) 2023

University of Stavanger

Camera and calibration

- Three points from the topic

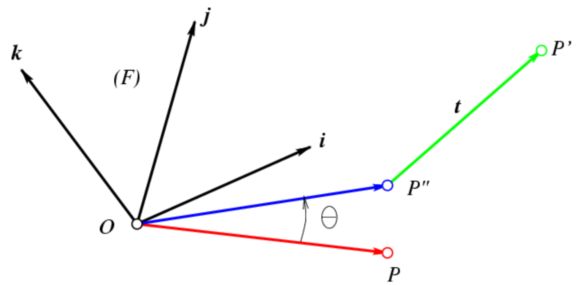
1. Which three different coordinate-systems do we need to deal with when calibrating camera and stereo vision systems?
2. What are the intrinsic and extrinsic camera parameters?
3. What is the principles of Zangs camera calibration routine?

Homogeneous representation of Rigid Transformations

Rotation and translation:

$$\begin{bmatrix} {}^{B}\mathbf{P} \\ 1 \end{bmatrix} = \begin{pmatrix} {}^{B}_{A}\mathbf{R} & {}^{B}\mathbf{O}_{A} \\ \mathbf{0}^{T} & 1 \end{pmatrix} \begin{bmatrix} {}^{A}\mathbf{P} \\ 1 \end{bmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} {}^{A}x \\ {}^{A}y \\ {}^{A}z \\ 1 \end{bmatrix}$$

Rigid Transformations as mappings:



$${}^{F}P' = \mathcal{R}\,{}^{F}P + \mathbf{t} \quad \Leftrightarrow \quad \begin{pmatrix} {}^{F}P' \\ 1 \end{pmatrix} = \begin{pmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^{T} & 1 \end{pmatrix} \begin{pmatrix} {}^{F}P \\ 1 \end{pmatrix}$$

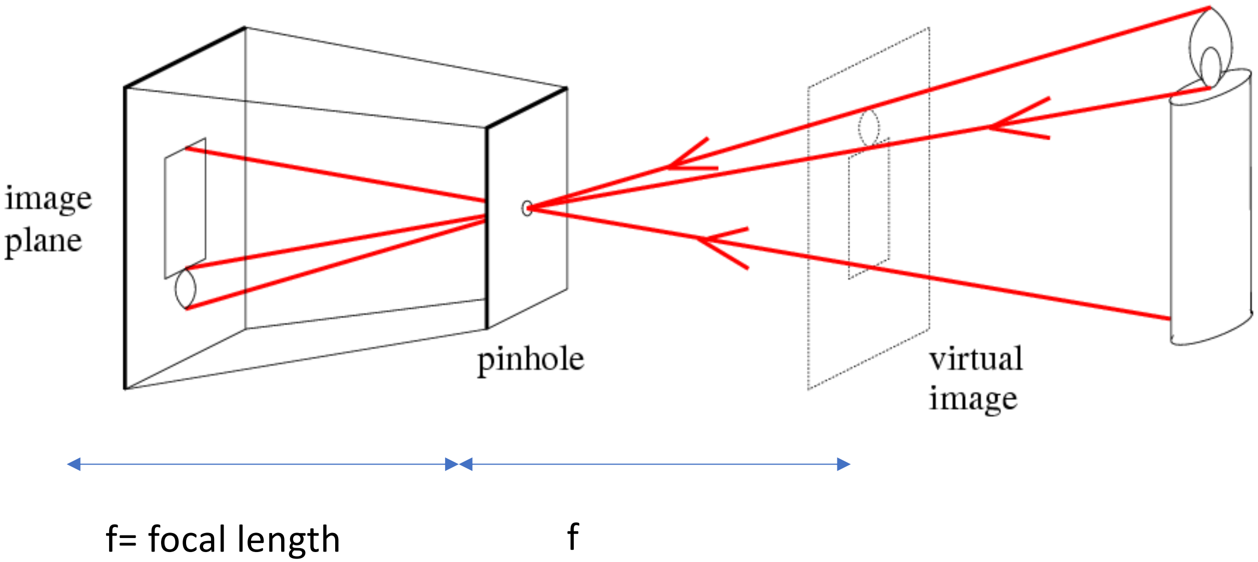
Fundamental rotation matrices

$$\mathbf{R}_{x,\theta} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}$$

$$\mathbf{R}_{y,\theta} = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

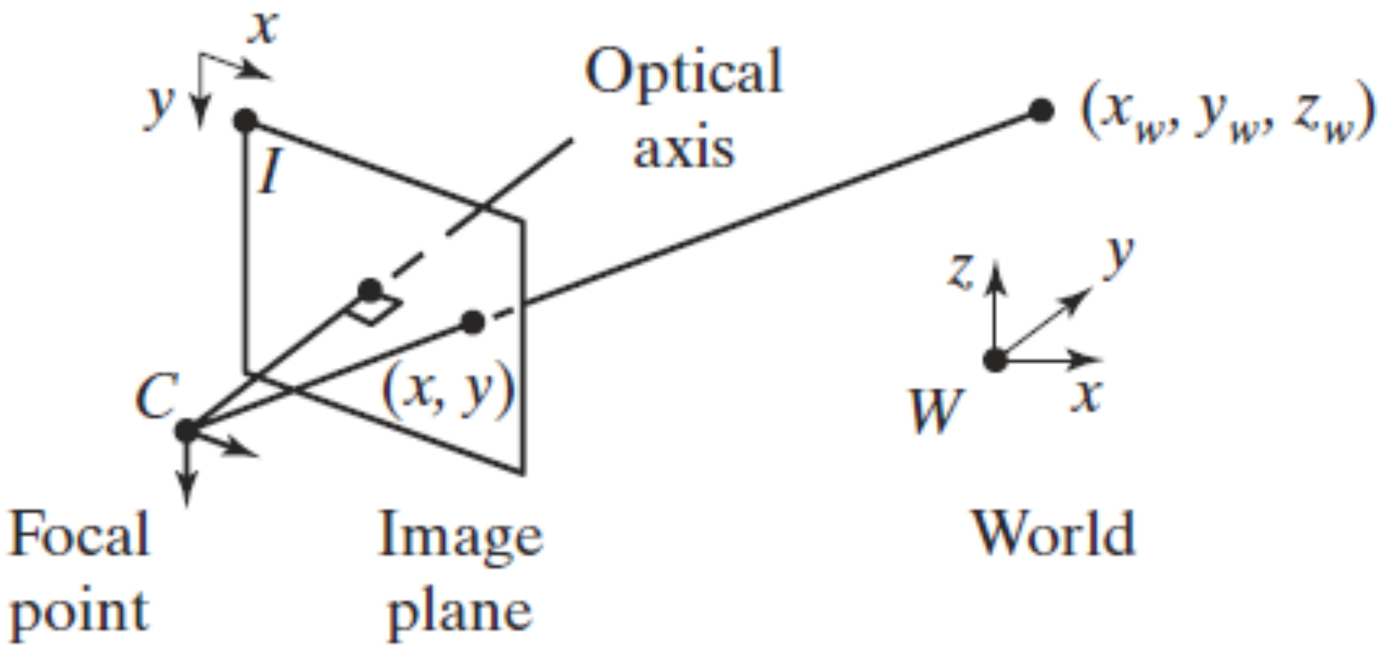
$$\mathbf{R}_{z,\theta} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Pinhole camera - revisited



Perspective Imaging

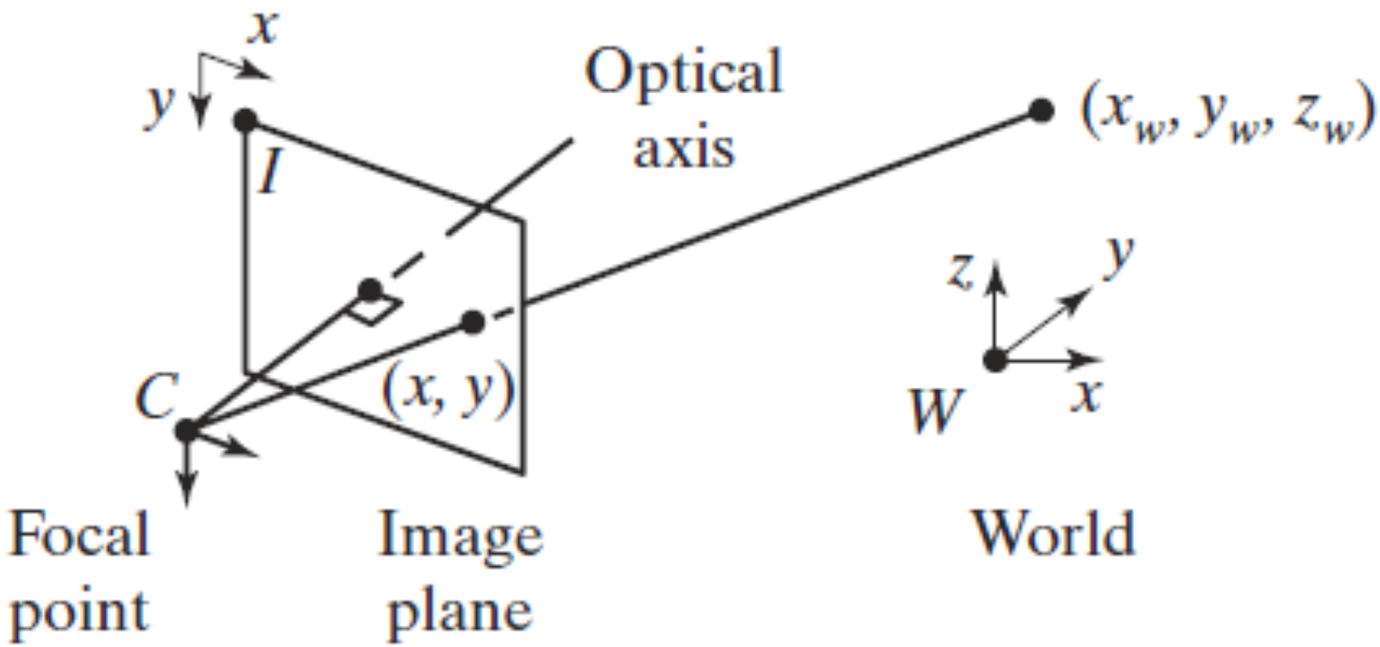
- Three coordinate systems are involved in the process of imaging:
  - W – world coordinate system
  - C - camera coordinate system
  - I – image coordinate system



Points in the world are described in the *world coordinate system*, in which lengths are measured in meters.

The camera coordinate system is centered at the focal point, *x* and *y* axes aligned with the rightward horizontal and downward vertical directions of the image plane. Positive *z* axis points along the optical axis toward the world. lengths are measured in meters.

The *image coordinate system* is centered at the top left corner of the image, with the positive *x* and *y* axes pointing along the rows and columns of the imaging sensor. In the image coordinate system, measurements are made in pixels.



Perspective Imaging

- Using homogeneous coordinates, the imaging process can be described:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}}_{\mathbf{P}_{\{3\times 4\}}} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

Image coord. system (I)

Proportional to

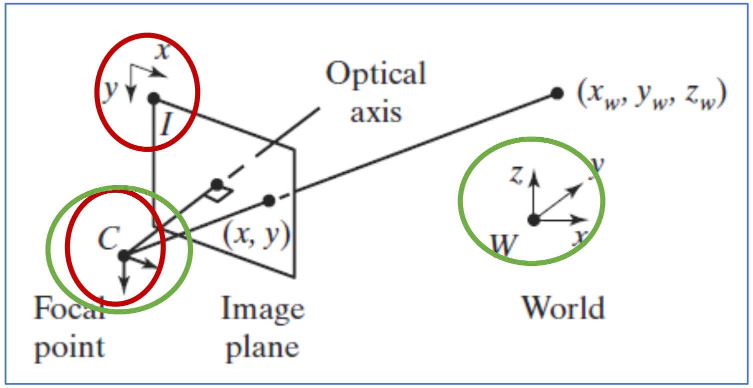
World coord. system (W)

Perspective Imaging

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}}_{\mathbf{P}_{\{3\times 4\}}} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

P is a 3 X 4 projection matrix that itself is composed of two parts:

$$\mathbf{P}_{\{3\times 4\}} = \underbrace{\begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\{3\times 3\}}} [\mathbf{R}_{\{3\times 3\}} \;\; \mathbf{t}_{\{3\times 1\}}]$$



- **Intrinsic** (camera dependent) parameters are found in **K.** Converts meters to pixels, moves the origo in the image plane etc., From C (camera coordinate system) to I (image coordinate system)
- **Extrinsic** (how do we describe the world relative to the camera) parameters in **R** and **t.** From W (world coordinate system) to C

Intrinsic Camera parameters

$$\underbrace{\begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\{3\times3\}}}$$

- u0 and v0 specifies the principal point, i.e. The intersection of the optical axis and the image plane.
- α, β, γ are related to the focal length in x direction (fx), in y direction (fy) and the skew θ between x and y axis in the image plane.

α = fx
β = fy/sin(θ)
γ = -fx/tan(θ)

fx= f/Δx, fy = f/Δy   Often: fx ≈ fy (tolerance of 5%)

Often we can assume θ≈(π/2) :

$$\boxed{\mathbf{K} = \begin{pmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}}$$

you would need to account for axis skew when calibrating unusual cameras or cameras taking photographs of photographs, else you can usually ignore the skew parameter (we will ignore it).

Extrinsic parameters

- Rotation and translation of the World coordinates with respect to the Camera coordinates:

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{TR} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \qquad {}^C\mathbf{P} = \mathbf{TR}\,{}^W\mathbf{P}$$

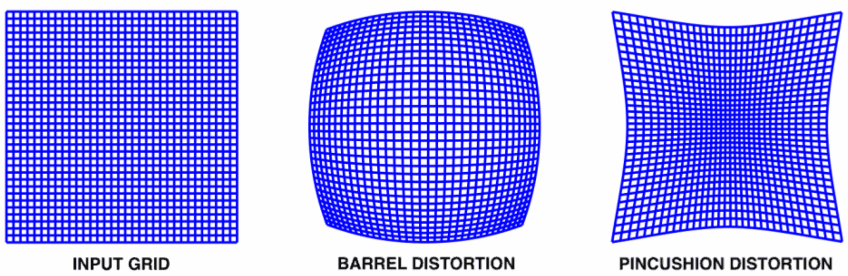
Lens Distortion

- Real cameras have lenses. Light bends due to curvature in the lens.
- The dominant distortion of a typical lens is **radial distortion**
- If a more accurate model is desired, **tangential distortion** (or *decentering distortion*) can be included
- Distortion parameters/coefficients have to be

estimated, thereafter the image can be unwarped

- We are not looking at that - but in a real situation

You might need to take that into account



(13.5) Camera calibration

- Camera calibration is the operation to *measure* and *compute* an estimate of the **Intrinsic camera parameters**, and **Extrinsic camera parameters**

- Camera calibration is *necessary in machine-* and *robot vision* because we want to
- **Reconstruct** a world model of the scene or objects in the scene, and or
- **Measure** physical distances in the scene, and or
- **Interact** with objects in the scene (robot, hand-eye coordination).
- **Stereo vision**

- The most popular technique for calibrating the intrinsic camera parameters is **Zhang's algorithm**.

Camera calibration

- Estimate intrinsic and extrinsic parameters from a set of images. Intrinsic parameters most interestng – >about the camera
- Calibrating a pinhole camera involves estimating 11 parameters. The matrix $\mathbf{P_M}$ contains 12 elements, But 3D-> 2D means we loose depth, therefore one free parameter.
- It is common to set Pm34 = 1 as a normalization of $\mathbf{P_M}$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}}_{\mathbf{P}_{\{3\times 4\}}} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

Camera calibration

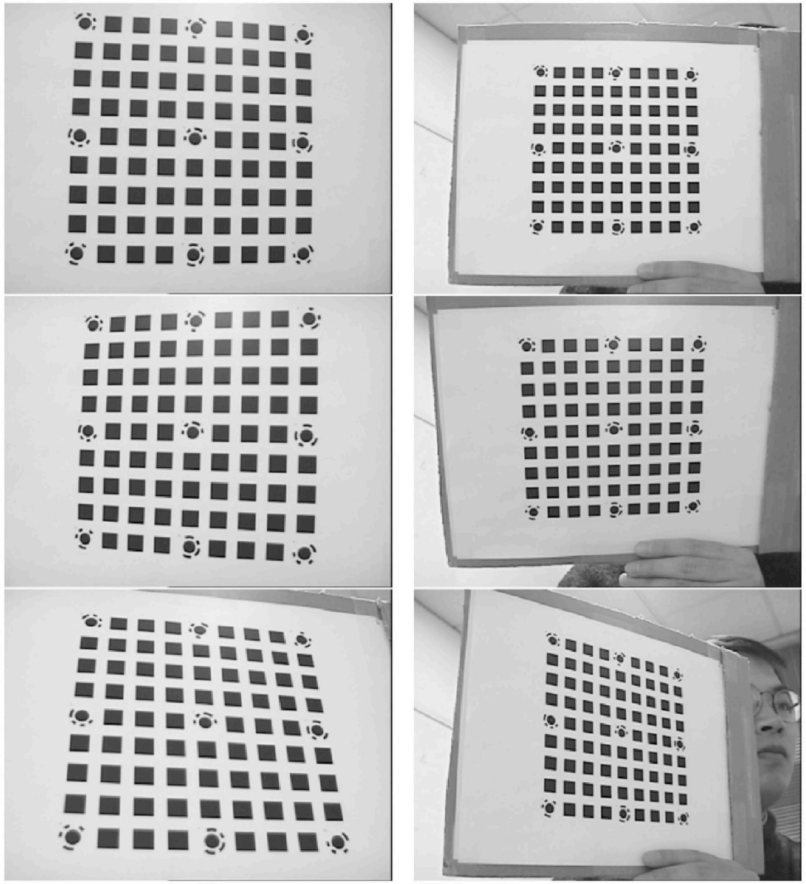
$$\mathbf{P}_{\{3\times 4\}} = \underbrace{\begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\{3\times 3\}}} [\mathbf{R}_{\{3\times 3\}} \; \mathbf{t}_{\{3\times 1\}}]$$

- 6 values for the Euclidean rotation and translation (3 values each, extrinsic parameters), and 5 values for the internal calibration matrix **K** (intrinsic parameters).

Zhangs algorithm - overview

- Require a planar calibration target with known coordinates /dimensions in meter. Typically used checkerboard
- Require 6-10 (more than 3) different images captured of the target at different positions.
- Find feature points (corners in the chekerboard, use Harris/SIFT) in the images (x,y)coordinates in the image, along with knowing the physical distance of the corner points in meters.
- From this we can estimate the intrinsic parameters K. (will look at some more details)
- Can also find extrinsic parameters. This is not always intersting

Checkerboard calibration pattern



(a) (b)

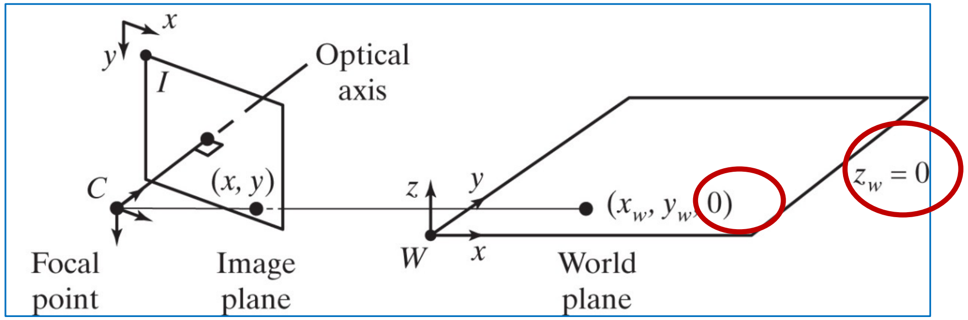
“Checkerboard” pattern suggested by Zhengyou Zhang (Microsoft Research) 1999

A plane viewed from different unknown orientations.

Zhangs Calibration Algorithm

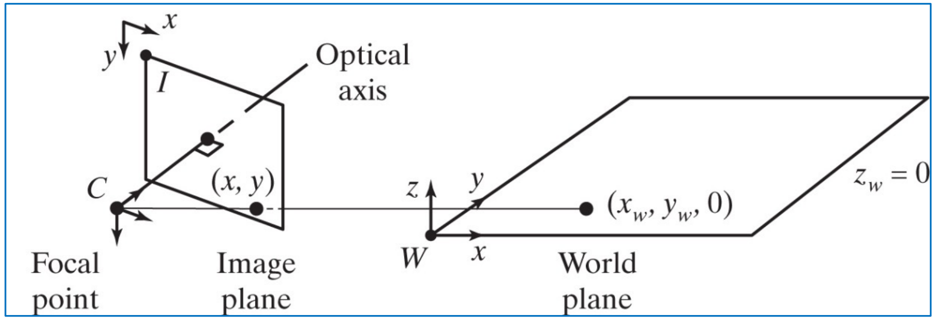
- The projection of a target point onto the image plane is given by:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\{3\times 3\}}} \underbrace{\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix}}_{[\mathbf{R}_{\{3\times 3\}} \quad \mathbf{t}_{\{3\times 1\}}]} \begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix} = \underbrace{\mathbf{K}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}}_{\mathbf{H}_{\{3\times 3\}}} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$$



How to estimate the homography matrix H?

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \propto \mathbf{H} \begin{pmatrix} x_w \\ y_w \\ 1 \end{pmatrix} = [\mathbf{h}_1 \, \mathbf{h}_2 \, \mathbf{h}_3] \begin{pmatrix} x_w \\ y_w \\ 1 \end{pmatrix}$$



Using SIFT / Harris on calibration images we find corresponding pairs of points:

$$(x_i\,, y_i) \leftrightarrow (x_{w_i}\,, y_{w_i}\,)$$

From these we can estimate **H**.

Normalized Direct Linear Transfomration (DLT) algorithm can be used.

We will not look at those details, but with a number of points this can be done.

Have as many **H** estimates as we have images.

Zhangs Calibration Algorithm

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\{3\times 3\}}} \underbrace{[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}]}_{[\mathbf{R}_{\{3\times 3\}} \quad \mathbf{t}_{\{3\times 1\}}]} \begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix} = \underbrace{\mathbf{K}\,[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}]}_{\mathbf{H}_{\{3\times 3\}}} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$$

Can estimate H from images, from that – find K

1. $\boxed{\mathbf{K}\mathbf{r}_i \propto \mathbf{h}_i \implies \mathbf{r}_i \propto \mathbf{K}^{-1}\mathbf{h}_i}$
2. $\boxed{\mathbf{r}_1^T \mathbf{r}_2 = 0}$
3. $\boxed{\mathbf{r}_1^T \mathbf{r}_1 = \mathbf{r}_2^T \mathbf{r}_2}$

H is the 3 X 3 projective transformation matrix known as a **homography.**

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Define image of the absolute conic IAC: $\mathbf{w}_\infty = \mathbf{K}^{-T}\mathbf{K}^{-1} = (\mathbf{K}\mathbf{K}^T)^{-1}$

Combine with equations from prev. slide:

1. $\boxed{\mathbf{K}\mathbf{r}_i \propto \mathbf{h}_i \implies \mathbf{r}_i \propto \mathbf{K}^{-1}\mathbf{h}_i}$
2. $\boxed{\mathbf{r}_1^T\mathbf{r}_2 = 0}$
3. $\boxed{\mathbf{r}_1^T\mathbf{r}_1 = \mathbf{r}_2^T\mathbf{r}_2}$

1+2 $\quad \mathbf{h}_1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 = 0 \qquad \boxed{\implies \mathbf{h}_1^T\mathbf{w}_\infty\mathbf{h}_2 = 0}$

1+3 $\quad \mathbf{h}_1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_1 = \mathbf{h}_2^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 \quad \boxed{\implies \mathbf{h}_1^T\mathbf{w}_\infty\mathbf{h}_1 = \mathbf{h}_2^T\mathbf{w}_\infty\mathbf{h}_2}$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\{3\times3\}}} \underbrace{[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}]}_{[\mathbf{R}_{\{3\times3\}} \quad \mathbf{t}_{\{3\times1\}}]} \begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix} = \underbrace{\mathbf{K}\,[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}]}_{\mathbf{H}_{\{3\times3\}}} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$$

Define image of the absolute conic IAC: $\mathbf{w}_\infty = \mathbf{K}^{-T}\mathbf{K}^{-1} = (\mathbf{K}\mathbf{K}^T)^{-1}$

$$\mathbf{h}_1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 = 0 \qquad \boxed{\Longrightarrow \ \mathbf{h}_1^T\mathbf{w}_\infty\mathbf{h}_2 = 0}$$

$$\mathbf{h}_1^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_1 = \mathbf{h}_2^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 \quad \boxed{\Longrightarrow \ \mathbf{h}_1^T\mathbf{w}_\infty\mathbf{h}_1 = \mathbf{h}_2^T\mathbf{w}_\infty\mathbf{h}_2}$$

Rewrite! aij are known from **H** matrix, **w**i contains unknown from $\mathbf{w}_\infty$

$$\left.\begin{aligned} &\mathbf{a}_{12}^T\mathbf{w} = 0 \\ &(\mathbf{a}_{11} - \mathbf{a}_{22})^T\mathbf{w} = 0 \end{aligned}\right\}$$

From these we can estimate $\mathbf{w}_\infty$ when we have a number of different H´s. We have as many H´s as we have images.

From $\mathbf{w}_\infty$ we can find **K**

When we have K and Hj, we can find Rj and tj for each image j

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}_{\{3\times3\}}} \underbrace{[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}]}_{[\mathbf{R}_{\{3\times3\}} \quad \mathbf{t}_{\{3\times1\}}]} \begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix} = \underbrace{\mathbf{K}\,[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}]}_{\mathbf{H}_{\{3\times3\}}} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$$

Thereafter we can do a reprojection , i.e. With these parameters where would the physical point be on the image plane? Is that far from the true image point? -> Reprojection error:

$$RepErr = \sum_i \sum_j ||\mathbf{x}_{ij} - g(\mathbf{K}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{x}_w^j, \phi)||^2$$

Can update iteratively the intrinsic and extrinsic parameters minimizing the reprojection error -> bundle adjustment

ALGORITHM 13.12 Zhang's camera calibration routine

CALIBRATECAMERAZHANG

Input: $n \geq 3$ images of a known planar target at different orientations and directions
Output: intrinsic parameters $\mathbf{K}$, lens distortion parameters ϕ, and extrinsic parameters

for each image **do**
    Detect image features corresponding to known points on target.
    Compute the homography using the normalized DLT algorithm.
Stack the entries from the homographies into matrix $\mathbf{A}$.
Solve $\mathbf{Aw} = \mathbf{0}$ for $\mathbf{w}$, then reshape into ω_∞.
Compute the five intrinsic parameters of $\mathbf{K}$ from ω_∞ using either
    Cholesky decomposition, or
    Equation (13.143) for v_0, then Equation (13.142) for λ, then Equations (13.144)–(13.147).
Compute extrinsic parameters $\mathbf{R}_i$ and $\mathbf{t}_i$ for each image $i = 1, \ldots, n$ using Equations (13.148)–(13.151).
Using these results as a starting point, minimize Equation (13.152) to perform bundle adjustment.

Camera and calibration

- Three points from the topic

1. Which three different coordinate-systems do we need to deal with when calibrating camera and stereo vision systems?
   - ✓ World, camera and image coordinate systems
2. What are the intrinsic and extrinsic camera parameters?
   - ✓ Instrinsic- camera dependent. Extrinsic- how do we describe the world relative to the camera system?
3. What is the principles of Zangs camera calibration routine?
   - ✓ Planar (reduce dimentions with one) chckerboard-like (easy to find corners with SIFT etc) object with known physical measures. Take minimum 3(6-10) images, estimate homography images based on corner point matching and physical measures. Thereafter find intrinsic (and possibly extrinsic) parameters.