RRC Summer School 2023

Multi-View Geometry: Camera Calibration

Pranjal Paul

Computer Vision In General

- See and Perceive visual inputs (images/videos)
  - Given an image, we want :
    - Segmentation
    - Recognition
    - Reconstruction

- Are Human Vision perfect ?
  - Does CV mimic HV ?
  - Should computers process visual inputs like humans ?

What’s Wrong ?

What's Wrong ?

Thatcher Effect

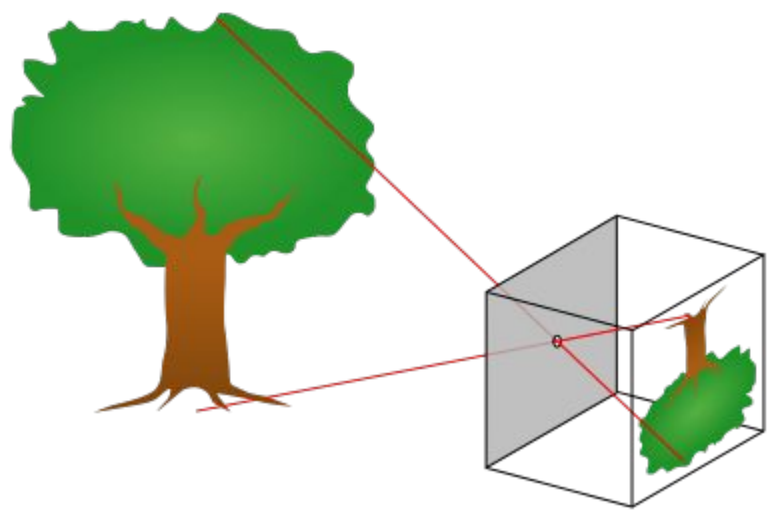
Objective

- Camera Model
  - Mathematically model what a camera does
- Camera Calibration
  - Estimation from a real world measurements
- Intrinsic and Extrinsic
- Application:
  - Stereo Setup
  - LiDAR-Camera Calibration

Camera Imaging

The Pinhole Camera

- Image formation can be approximated using PhC
- Entire world is in focus
- Affine Camera



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

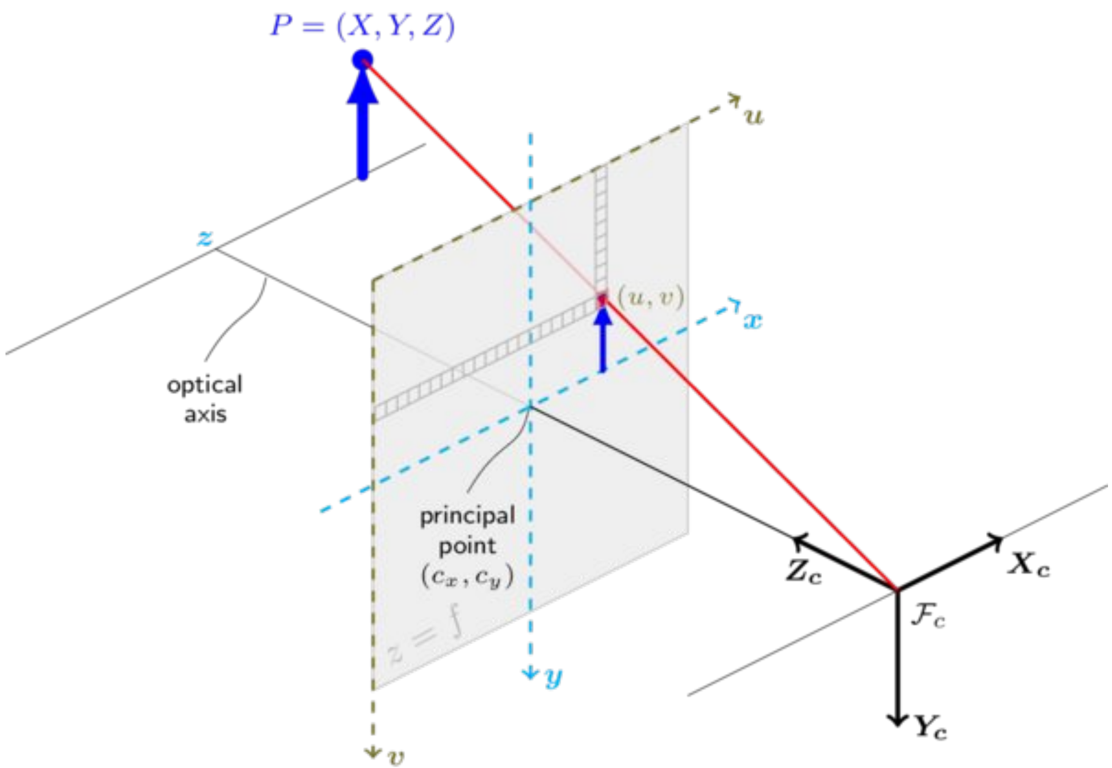
Image Courtesy: Wikipedia

The Pinhole Camera Effect

Virupaksha Temple, Hampi, Karnataka

Image Courtesy: Twitter

The Pinhole Camera Model

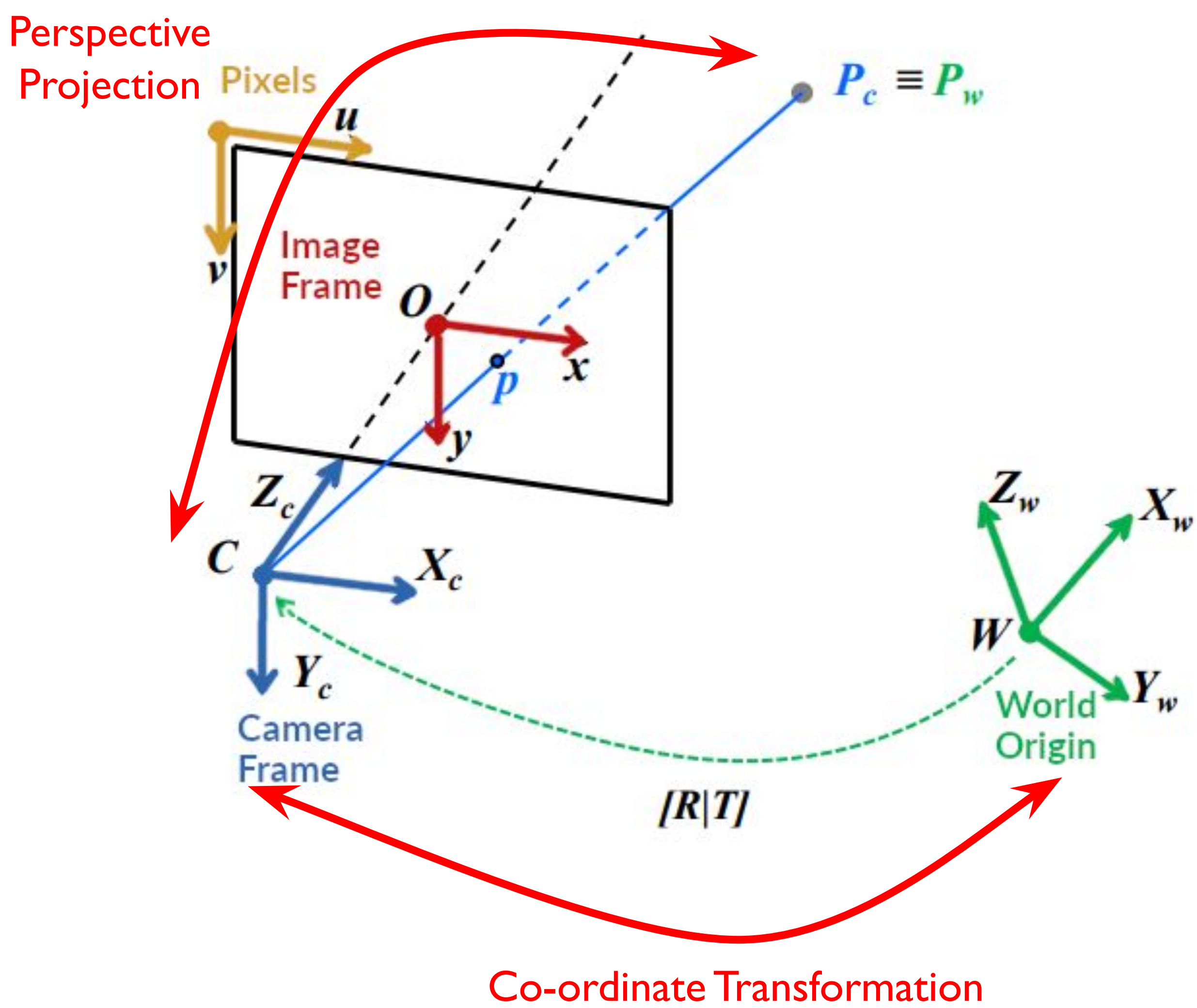


The transformation above is equivalent to the following (when $z \neq 0$):

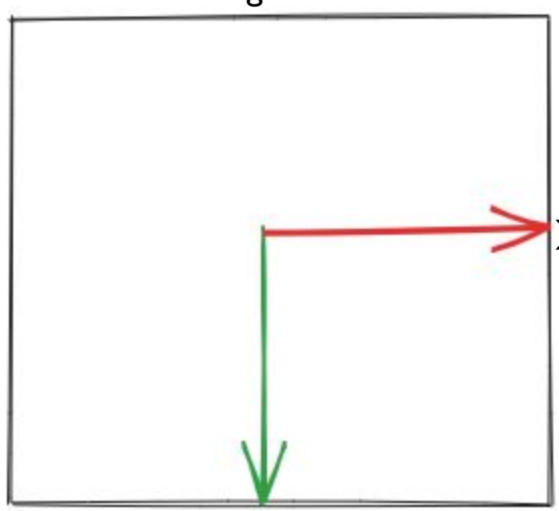
$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t$$

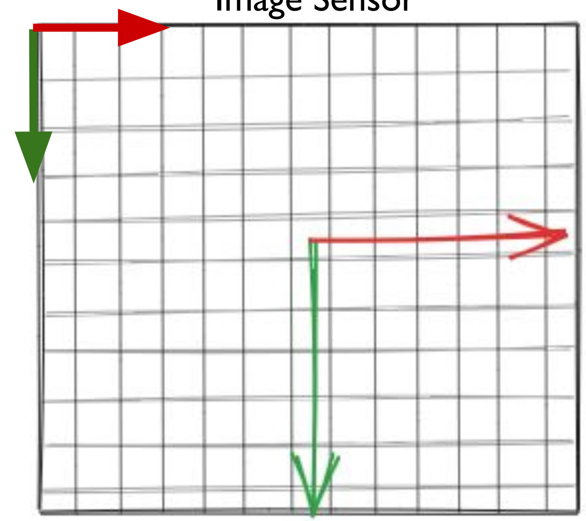
Image Courtesy: OpenCV

Forward Imaging model



Mapping from img plane to img sensor





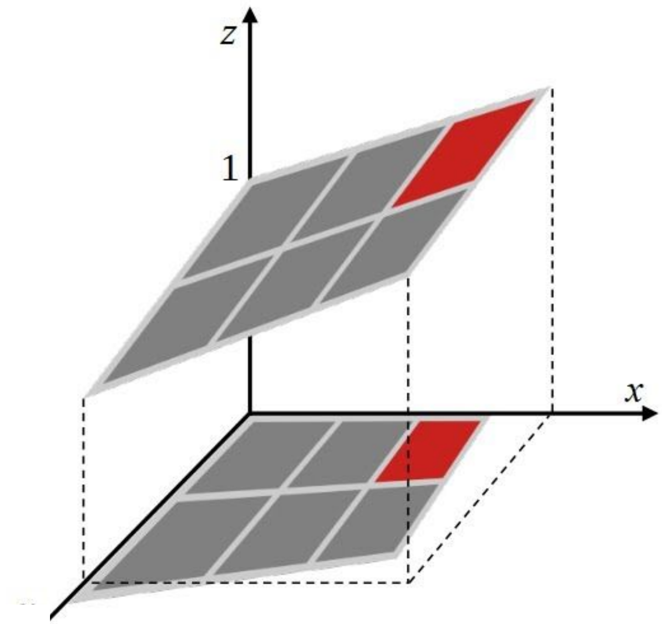
$$u = m_x x_i = m_x\ f\frac{x_c}{z_c} + o_x$$

$$v = m_y y_i = m_y\ f\frac{y_c}{z_c} + o_y$$

m_x and m_y are pixel density (pixels/mm)

Projective Transformation

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



x’ = Hx

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} xh_{11} + yh_{12} + h_{13} \\ xh_{21} + yh_{22} + h_{23} \\ xh_{31} + yh_{32} + h_{33} \end{bmatrix}$$

$$x' = \frac{x'_1}{x'_3} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}$$

$$y' = \frac{x'_2}{x'_3} = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}$$

Projective Geometry

Perspective Projection

- Does not preserve relative proportion
- In perspective projection, size varies inversely with distance - looks realistic
- But, can’t judge distances as we can with parallel projection
- Parallel lines seems to meet at a single point, called Vanishing Point

Perspective Projection

One-point Perspective

Perspective Projection

Two-point Perspective

Camera Calibration: Intrinsic

Why Calibrate ?

- Calibration provides the direction in space for each pixel
- Goal:
  - Metric reconstruction of the 3D scene from the given images
- What do we get from calibration ?
  - After calibration, we know the precise direction of the projection ray for each pixel
- To achieve this, we need:
  - Extrinsic parameters
  - Intrinsic parameters

Popular approach

- Assumption: Pinhole Camera Model
- Two approaches
  a. Direct Linear Transform
     - Requires at-least 7 known 3D points
  b. Zhang's Method
     - Estimate 5 linear parameters using a checkerboard

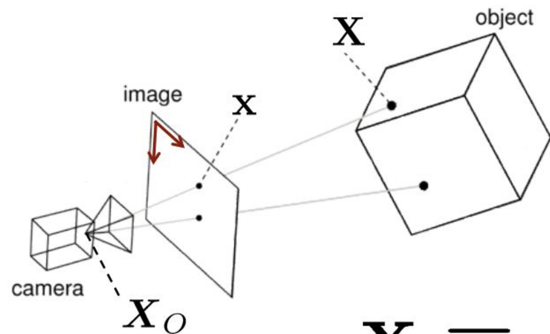
What are we determining ?

$$\overbrace{s\begin{bmatrix}u\\v\\1\end{bmatrix} = \underbrace{\boxed{\begin{bmatrix}f_x & 0 & c_x\\0 & f_y & c_y\\0 & 0 & 1\end{bmatrix}}}_{\text{K (Camera Intrinsics)}} \begin{bmatrix}r_{11} & r_{12} & r_{13} & t_1\\r_{21} & r_{22} & r_{23} & t_2\\r_{31} & r_{32} & r_{33} & t_3\end{bmatrix}}^{\text{Given this}} \overbrace{\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix}}^{\text{Find This}}$$

where:

- (X, Y, Z) are the coordinates of a 3D point in the world coordinate space
- (u, v) are the coordinates of the projection point in pixels
- A is a camera matrix, or a matrix of intrinsic parameters
- (cx, cy) is a principal point that is usually at the image center
- fx, fy are the focal lengths expressed in pixel units.

Setting up…



$$\mathbf{x} = \mathsf{P}\mathbf{X}$$

$$\mathbf{x} = \mathsf{K}R[I_3| - \boldsymbol{X}_O]\mathbf{X}$$

observed image point

c, s, m, x_H, y_H

3 rotations

3 translations

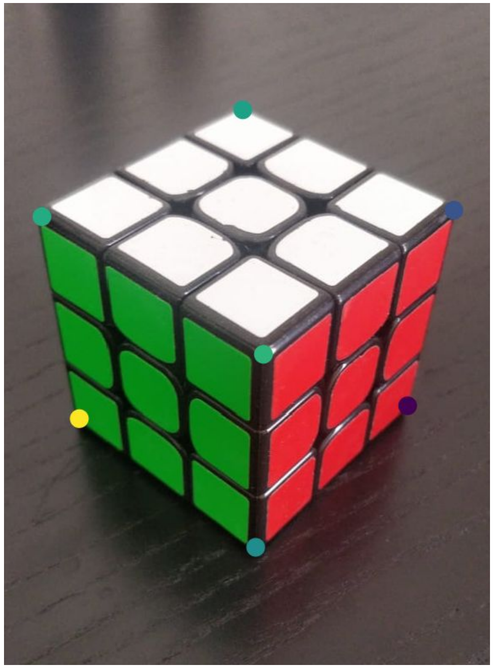
control point

Direct Linear Transform

DLT

- Gives both Intrinsic and Extrinsic parameters
- Procedure:
  a. Capture an image of an object with known geometry
  b. Identify correspondences between 3D and 2D points

```
image_points = ([128.88842013, 581.73122702],
                [337.61167244, 734.08396593],
                [517.38790436, 566.49595313],
                [84.70612585, 342.53742693],
                [346.75283678, 505.55485757],
                [572.23489036, 334.91978999],
                [322.37639855, 216.08465364])
world_points = ([0, 0, 0], [i, 0, 0], [i, i, 0],
                [0, 0, i], [i, 0, i], [i, i, i],
                [0, i, i])
```



DLT

c. Each 3D point gives two observation equations, one for each image coordinate

$$x = \frac{p_{11}X + p_{12}Y + p_{13}Z + p_{14}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}}$$
$$y = \frac{p_{21}X + p_{22}Y + p_{23}Z + p_{24}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}}$$

$$\mathbf{x}_i = \underset{3\times 4}{\mathsf{P}}\,\mathbf{X}_i = \begin{bmatrix} \boxed{p_{11} \quad p_{12} \quad p_{13} \quad p_{14}} \\ \boxed{p_{21} \quad p_{22} \quad p_{23} \quad p_{24}} \\ \boxed{p_{31} \quad p_{32} \quad p_{33} \quad p_{34}} \end{bmatrix} \mathbf{X}_i$$

So what we can rewrite the equation as

$$\begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} = \mathbf{x}_i = \mathsf{P}\mathbf{X}_i = \begin{bmatrix} \boxed{\mathbf{A}^\top} \\ \boxed{\mathbf{B}^\top} \\ \boxed{\mathbf{C}^\top} \end{bmatrix} \mathbf{X}_i = \begin{bmatrix} \mathbf{A}^\top\mathbf{X}_i \\ \mathbf{B}^\top\mathbf{X}_i \\ \mathbf{C}^\top\mathbf{X}_i \end{bmatrix}$$

DLT

$$x_i = \frac{\mathbf{A}^\top \mathbf{X}_i}{\mathbf{C}^\top \mathbf{X}_i} \quad \Rightarrow \quad x_i \, \mathbf{C}^\top \mathbf{X}_i - \mathbf{A}^\top \mathbf{X}_i = 0$$

$$y_i = \frac{\mathbf{B}^\top \mathbf{X}_i}{\mathbf{C}^\top \mathbf{X}_i} \quad \Rightarrow \quad y_i \, \mathbf{C}^\top \mathbf{X}_i - \mathbf{B}^\top \mathbf{X}_i = 0$$

Leads to an system of equation, which is **linear in the parameters $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$**

$$\begin{aligned} -\mathbf{X}_i^\top \mathbf{A} \qquad\qquad\qquad & + x_i \, \mathbf{X}_i^\top \mathbf{C} = 0 \\ -\mathbf{X}_i^\top \mathbf{B} \quad & + y_i \, \mathbf{X}_i^\top \mathbf{C} = 0 \end{aligned}$$

DLT

- Rewrite $$\begin{aligned} -\mathbf{X}_i^\top \mathbf{A} \qquad\qquad\qquad & +x_i\,\mathbf{X}_i^\top \mathbf{C} = 0 \\ -\mathbf{X}_i^\top \mathbf{B} \quad & +y_i\,\mathbf{X}_i^\top \mathbf{C} = 0 \end{aligned}$$
- as $$\begin{aligned} \boldsymbol{a}_{x_i}^\top \boldsymbol{p} = 0 \\ \boldsymbol{a}_{y_i}^\top \boldsymbol{p} = 0 \end{aligned}$$
- with

$$\begin{aligned}
\boldsymbol{p} &= (p_k) = \mathrm{vec}(\mathsf{P}^\top) \\
\boldsymbol{a}_{x_i}^\top &= (-\mathbf{X}_i^\top, \mathbf{0}^\top, x_i\mathbf{X}_i^\top) \\
&= (-X_i,\ -Y_i,\ -Z_i,\ -1, 0,\ 0,\ 0,\ 0, x_iX_i,\ x_iY_i,\ x_iZ_i,\ x_i) \\
\boldsymbol{a}_{y_i}^\top &= (\mathbf{0}^\top, -\mathbf{X}_i^\top,\ y_i\mathbf{X}_i^\top) \\
&= (0,\ 0,\ 0,\ 0, -X_i,\ -Y_i,\ -Z_i,\ -1, y_iX_i,\ y_iY_i,\ y_iZ_i,\ y_i)
\end{aligned}$$

DLT

- For each point, we have

$$\boldsymbol{a}_{x_i}^\mathsf{T} \boldsymbol{p} = 0$$

$$\boldsymbol{a}_{y_i}^\mathsf{T} \boldsymbol{p} = 0$$

- Stacking everything together

$$\begin{bmatrix} \boldsymbol{a}_{x_1}^\mathsf{T} \\ \boldsymbol{a}_{y_1}^\mathsf{T} \\ \cdots \\ \boldsymbol{a}_{x_i}^\mathsf{T} \\ \boldsymbol{a}_{y_i}^\mathsf{T} \\ \cdots \\ \boldsymbol{a}_{x_I}^\mathsf{T} \\ \boldsymbol{a}_{y_I}^\mathsf{T} \end{bmatrix} \boldsymbol{p} = \underset{2I \times 12}{\mathsf{M}} \; \underset{12 \times 1}{\boldsymbol{p}} \stackrel{!}{=} 0$$

DLT

- Solving a system of linear equations of the form Ax = 0 is equivalent to finding the null space of A
- Thus, we can apply the SVD on M
- Choose *p* as the singular vector belonging to the singular value of 0
- i.e., choose the last column of V^T

$$\underset{2I\times 12}{\mathsf{M}} = \underset{2I\times 12}{U}\ \underset{12\times 12}{\mathsf{S}}\ \underset{12\times 12}{V^{\mathsf{T}}} = \sum_{i=1}^{12} s_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathsf{T}}$$

DLT

- Reshape 12x1 to 3x4

$$\boldsymbol{p} = \begin{bmatrix} p_{11} \\ \vdots \\ p_{34} \end{bmatrix} \Rightarrow \mathsf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}$$

- Now, how do we determine K, X_0, R ?

Decomposition of P

- Structure of P_{3x4}

$$P = \left[KR \mid -KRX_0\right] = \left[H \mid h\right]$$

- Projection Center

$$X_0 = -H^{-1}h$$

- Now, what do we know about H = KR
  - K is a triangle matrix
  - R is a rotation matrix
- Is there a matrix decomposition into a rotation matrix and a triangular on?

Decomposition of P

- QR decomposition of H^{-1} yields *R* & *K*

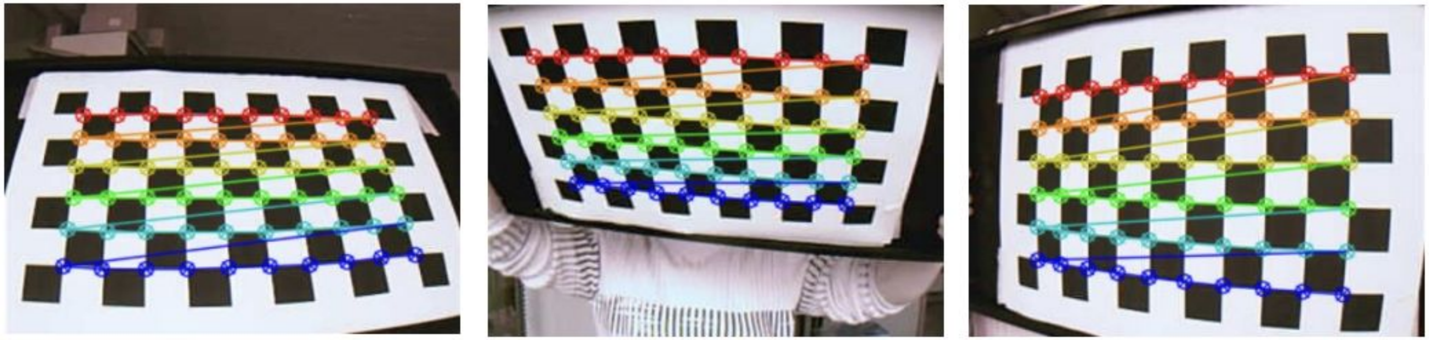
$$H^{-1} = (K\,R)^{-1} = R^{-1}\,K^{-1} = \underbrace{R^{T}}_{Q}\,\underbrace{K^{-1}}_{R}$$

- H = KR is homogeneous
- Thus is Calibration matrix K
- Due to homogeneity, normalize K: K/K_{33}

Checkerboard Method

Zhang's Method

- Observed 2D pattern (checkerboard)
- Known size and structure
- Set the world coordinate system to the corner of the checkerboard for each image
- All points on the checkerboard lie in the X/Y plane, i.e., Z = 0



Zhang's Method

- The Z coordinate of each point on the checkerboard is equal to zero

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} c & cs & x_H \\ 0 & c(1+m) & y_H \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ \boxed{Z} \\ 1 \end{bmatrix}$$

- Each point observed on the checkerboard generates such an equation

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathsf{K}[\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t}] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

Zhang's Method

- For multiple observed points on the checkerboard (in the same image), we obtain

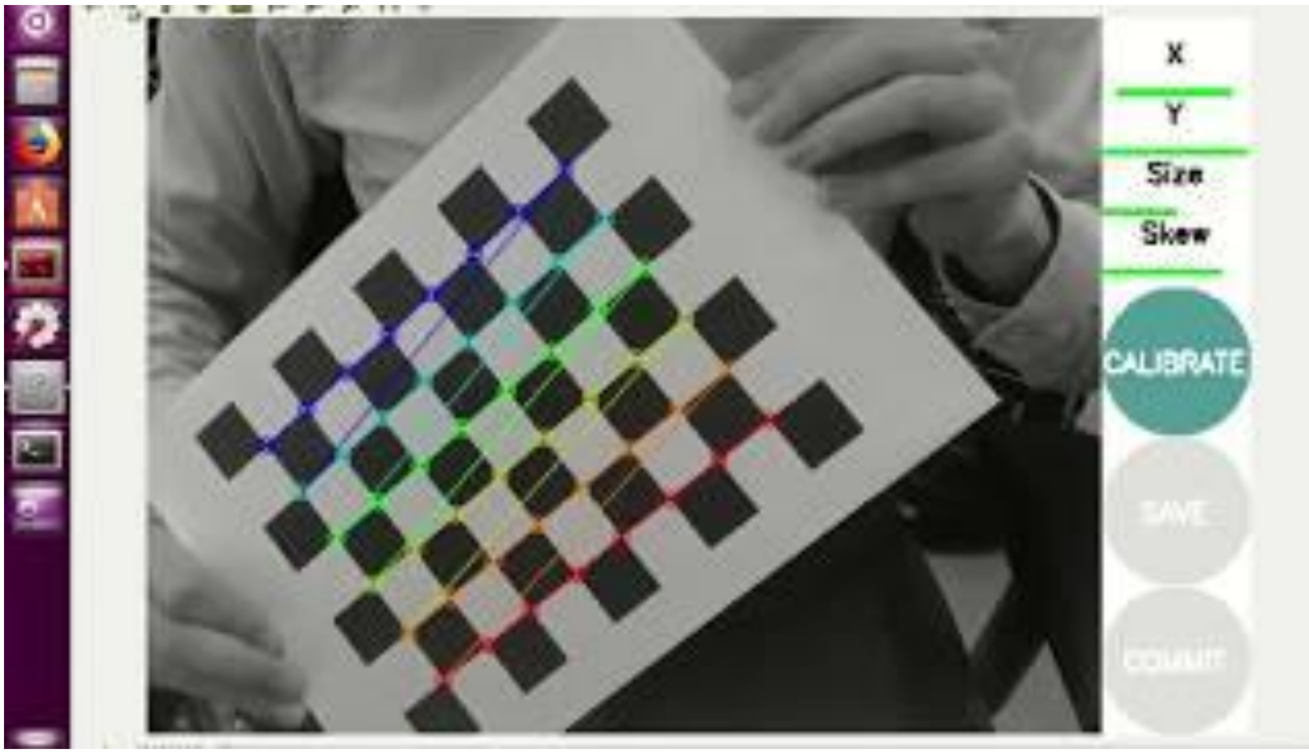
$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \underset{3\times 3}{\mathsf{H}} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} \qquad i = 1, \ldots, I$$

- We estimate a 3x3 homography

$$\begin{aligned} \boldsymbol{h} &= (h_k) = \text{vec}(\mathsf{H}^{\mathsf{T}}) \\ \boldsymbol{a}_{x_i}^{\mathsf{T}} &= (-X_i,\ -Y_i,\ -\cancel{Z_i},\ -1, 0,\ 0,\ \cancel{0},\ 0, x_iX_i,\ x_iY_i,\ x_i\cancel{Z_i},\ x_i) \\ \boldsymbol{a}_{y_i}^{\mathsf{T}} &= (0,\ 0,\ \cancel{0},\ 0, -X_i,\ -Y_i,\ -\cancel{Z_i},\ -1, y_iX_i,\ y_iY_i,\ y_i\cancel{Z_i},\ y_i) \end{aligned}$$

$$\begin{aligned} \boldsymbol{h} &= (h_k) = \text{vec}(\mathsf{H}^{\mathsf{T}}) \\ \boldsymbol{a}_{x_i}^{\mathsf{T}} &= (-X_i,\ -Y_i,\ -1, 0,\ 0,\ 0, x_iX_i,\ x_iY_i,\ x_i) \\ \boldsymbol{a}_{y_i}^{\mathsf{T}} &= (0,\ 0,\ 0, -X_i,\ -Y_i,\ -1, y_iX_i,\ y_iY_i,\ y_i) \end{aligned}$$

Calibration using Checkerboard



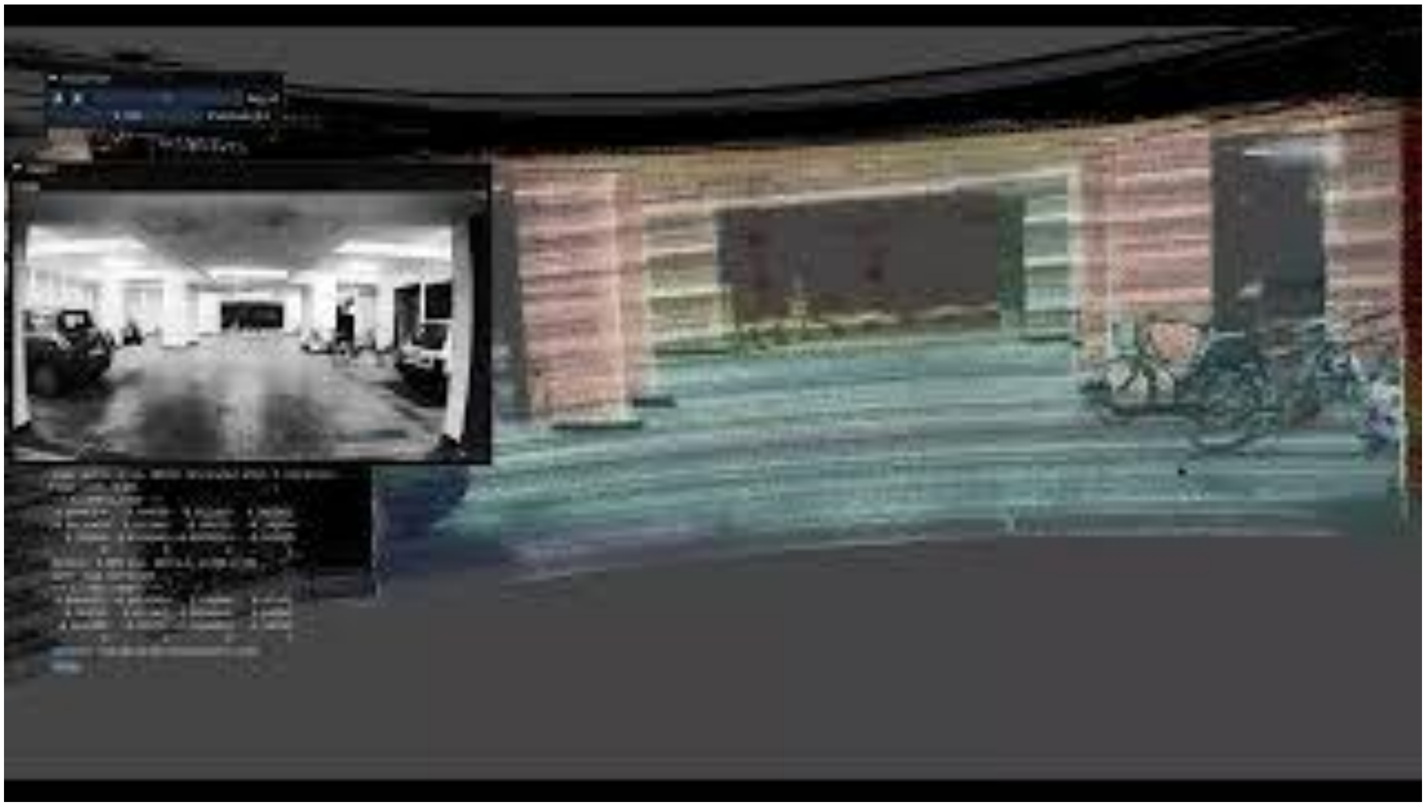
Camera Calibration: Extrinsic only

Extrinsic Parameters

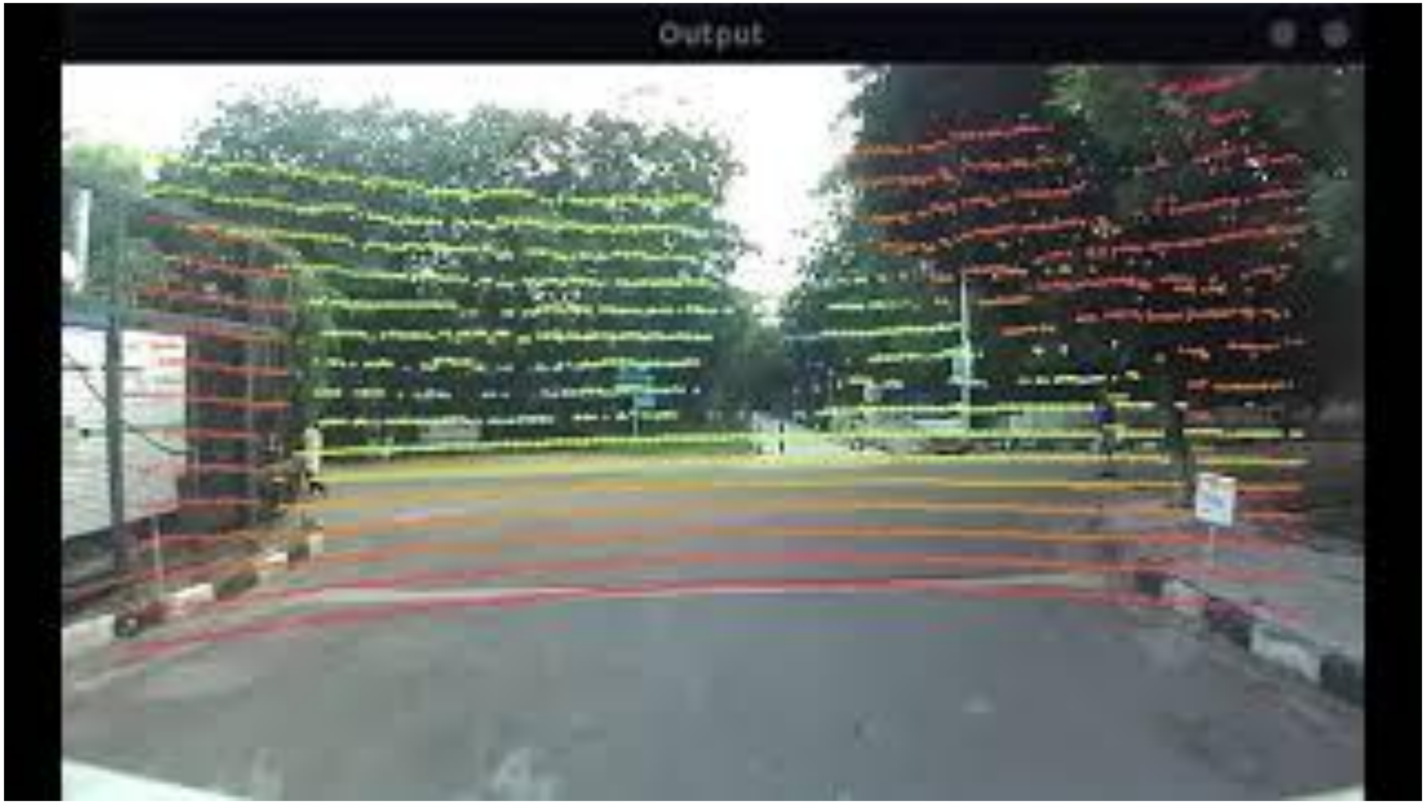
- Determines "where is the camera in 3D world?"
- 6 DoF vector
- Projection center defines the location of the camera
- T = [R | t]
- 3D Points could be obtained from other sensors
  - Stereo Setup
  - LiDAR
  - RGB-D

LiDAR-Camera Calibration

Calibration Result



Calibration Result



Resources

- Fei Fei Li (Slides)
  - Camera Models
  - Camera Calibration
- Stanford Course Notes
  - Camera Model
  - Single View Geometry
- YouTube
  - Prof. SK Nayar
    - Pinhole Camera and Image formation
    - Calibration Playlist

Resources

- YouTube
  - Prof. Cyrill Stachniss
    - Direct Linear Transform
    - Zhang's Method
  - George Lecakes
    - Video 14, 16-18
  - Lecture on Image formation CVFX