3D Vision 1

Week 7

Image Formation

Camera Projection Matrix

Announcements

- Assignment 2 due in **one week** (11:59pm Friday 26 April)
  - This includes a one week extension that has already been applied
  - **Zero** marks if either report or code submitted late (unless extension)
    - Submit early; you can always resubmit an updated version later
    - Depending on your internet connection and load on the TurnItIn servers, uploading can sometimes be slow, so please factor this into your submission schedule
  - Submit your report (PDF) and code (ZIP file) **separately under the correct tab** in the submission box
  - Follow the instructions under Submission Requirements

Announcements

- Public holiday Thursday 25 April: Thursday lab rescheduled at **13:00-15:00 Tuesday Rm 109 CSIT Building**

Weekly Study Plan: Overview

Wk	Starting	Lecture	Lab	Assessment
1	19 Feb	Introduction	✗	
2	26 Feb	Low-level Vision 1	1	
3	4 Mar	Low-level Vision 2	1	
		Mid-level Vision 1		
4	11 Mar	Mid-level Vision 2	1	CLab1 report due Friday
		High-level Vision 1		
5	18 Mar	High-level Vision 2	2	
6	25 Mar	High-level Vision 3[1]	2	
	1 Apr	Teaching break	✗	
	8 Apr	Teaching break	✗	
7	15 Apr	3D Vision 1	2	CLab2 report due Friday
8	22 Apr	3D Vision 2	3	
9	29 Apr	3D Vision 3	3	
10	6 May	3D Vision 4	3	
		Mid-level Vision 3		
11	13 May	High-level Vision 4	✗	CLab3 report due Friday
12	20 May	Course Review	✗	

Weekly Study Plan: Part B

Wk	Starting	Lecture	By
7	15 Apr	3D vision: introduction, camera model, single-view geometry	Dylan
8	22 Apr	3D vision: camera calibration, two-view geometry (homography)	Dylan
9	29 Apr	3D vision: two-view geometry (epipolar geometry, triangulation, stereo)	Dylan
10	6 May	3D vision: multiple-view geometry	Weijian
		Mid-level vision: optical flow, shape-from-X	Dylan
11	13 May	High-level vision: self-supervised learning, detection, segmentation	Dylan
12	20 May	Course review	Dylan

Outline

1. Introduction to 3D vision
2. Model fitting (line fitting)
    1. Least squares
    2. M-estimation
    3. RANSAC
    4. Hough transform
3. Image formation (review): pinhole camera model
4. Camera projection matrix and single view geometry
5. Camera calibration
6. Resectioning and camera pose

Hough Transform

7

Fitting Multiple Lines Using a Hough Transform

- Given a binary edge image, find the lines (or curves) that explain the data points best in the parameter space
- This parameter space is called a Hough space

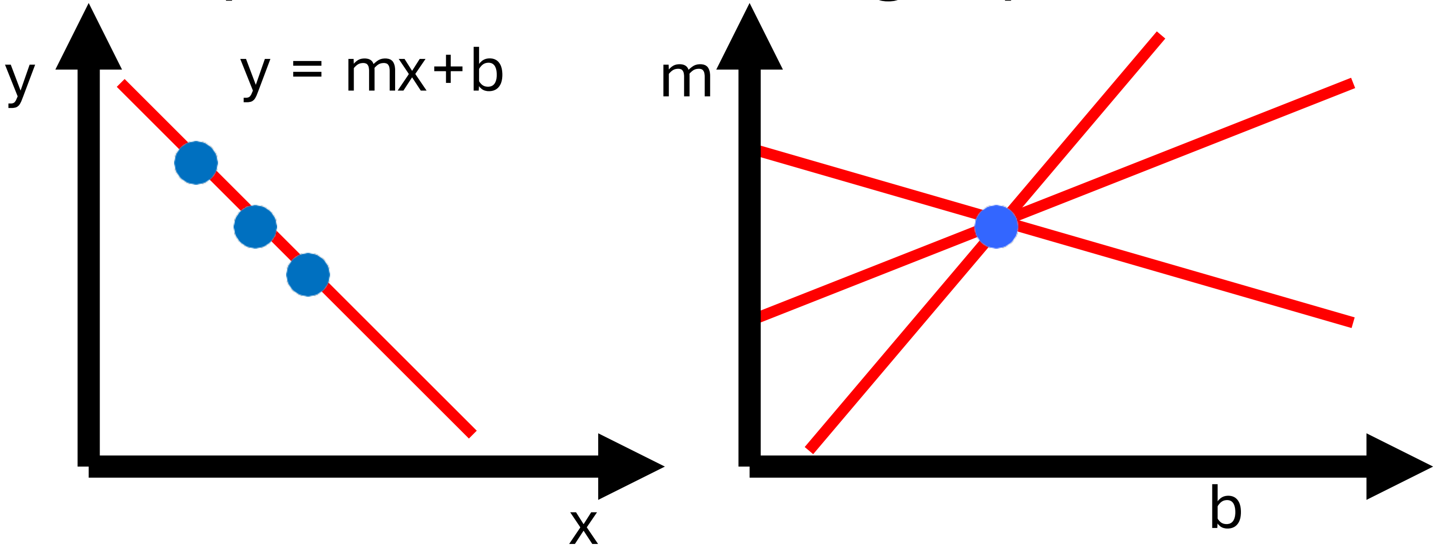


Image space Parameter (Hough) space

P.V.C. Hough, *Machine Analysis of Bubble Chamber Pictures,* Proc. Int. Conf. High Energy Accelerators and Instrumentation, 1959. (Patented)

A point (x_1, y_1) is mapped to a line in Hough space

$y_1 = m x_1 + b$

$b = -x_1 m + y_1$ (rewrite the equation in terms of m and b)

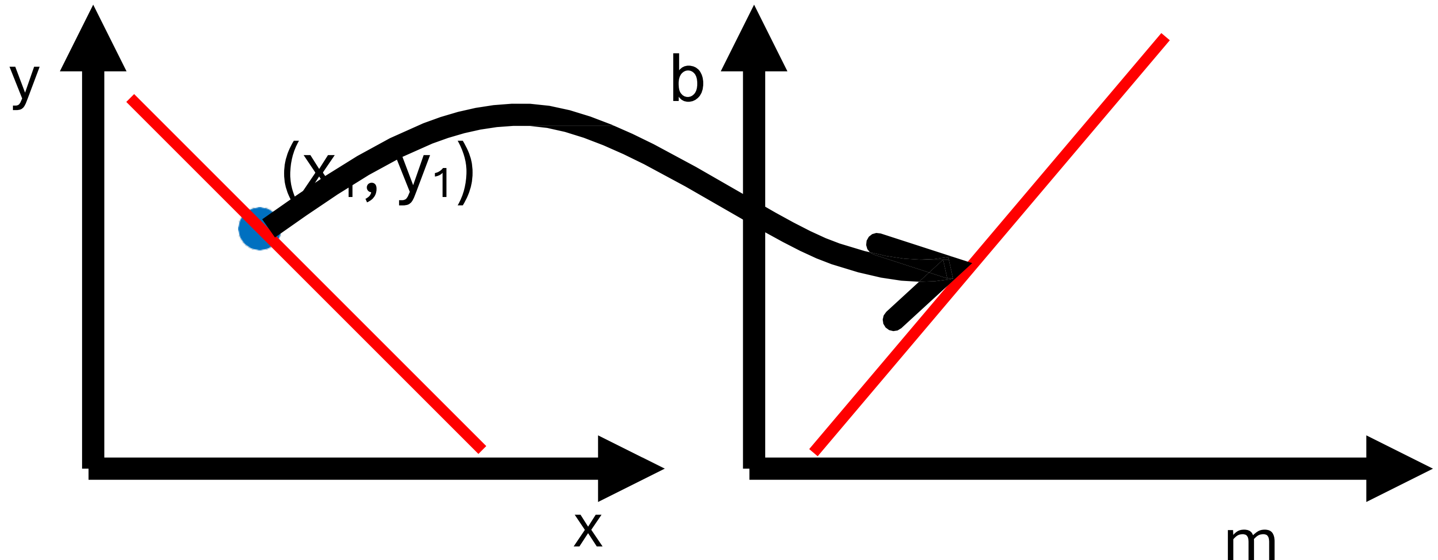


Image space

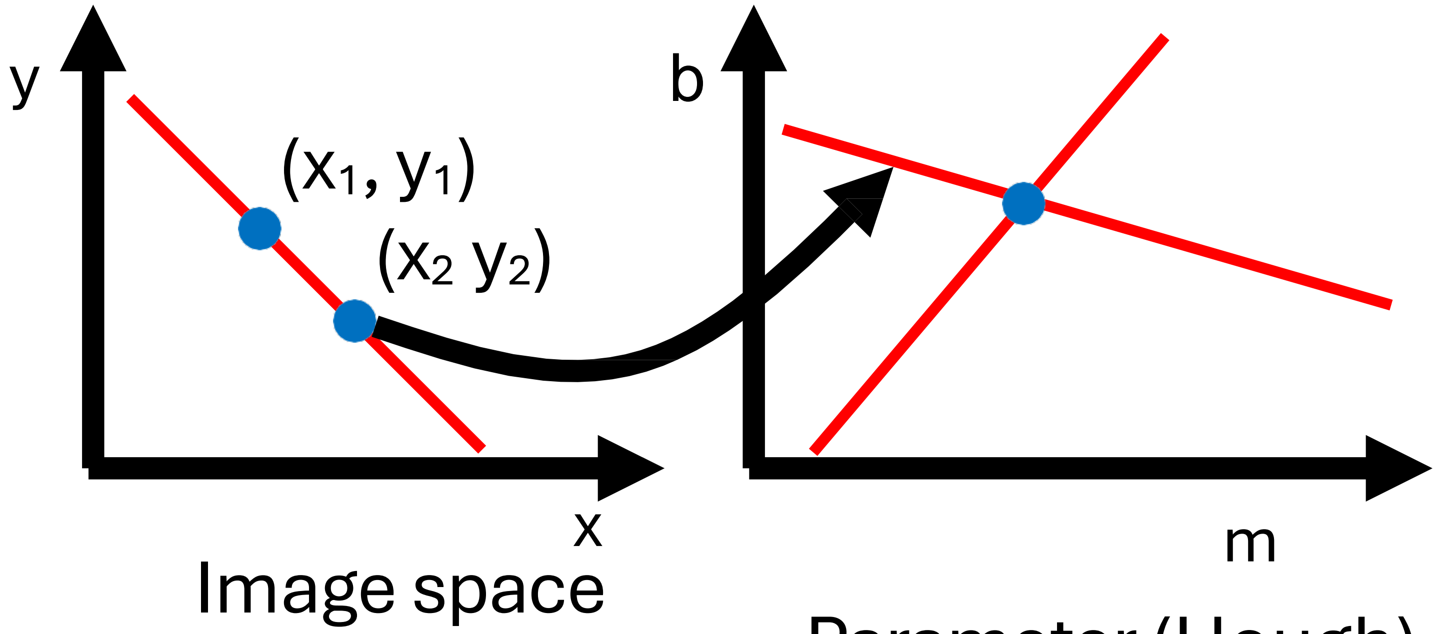
Parameter (Hough) space

A point (x_2, y_2) is mapped to a line in Hough space

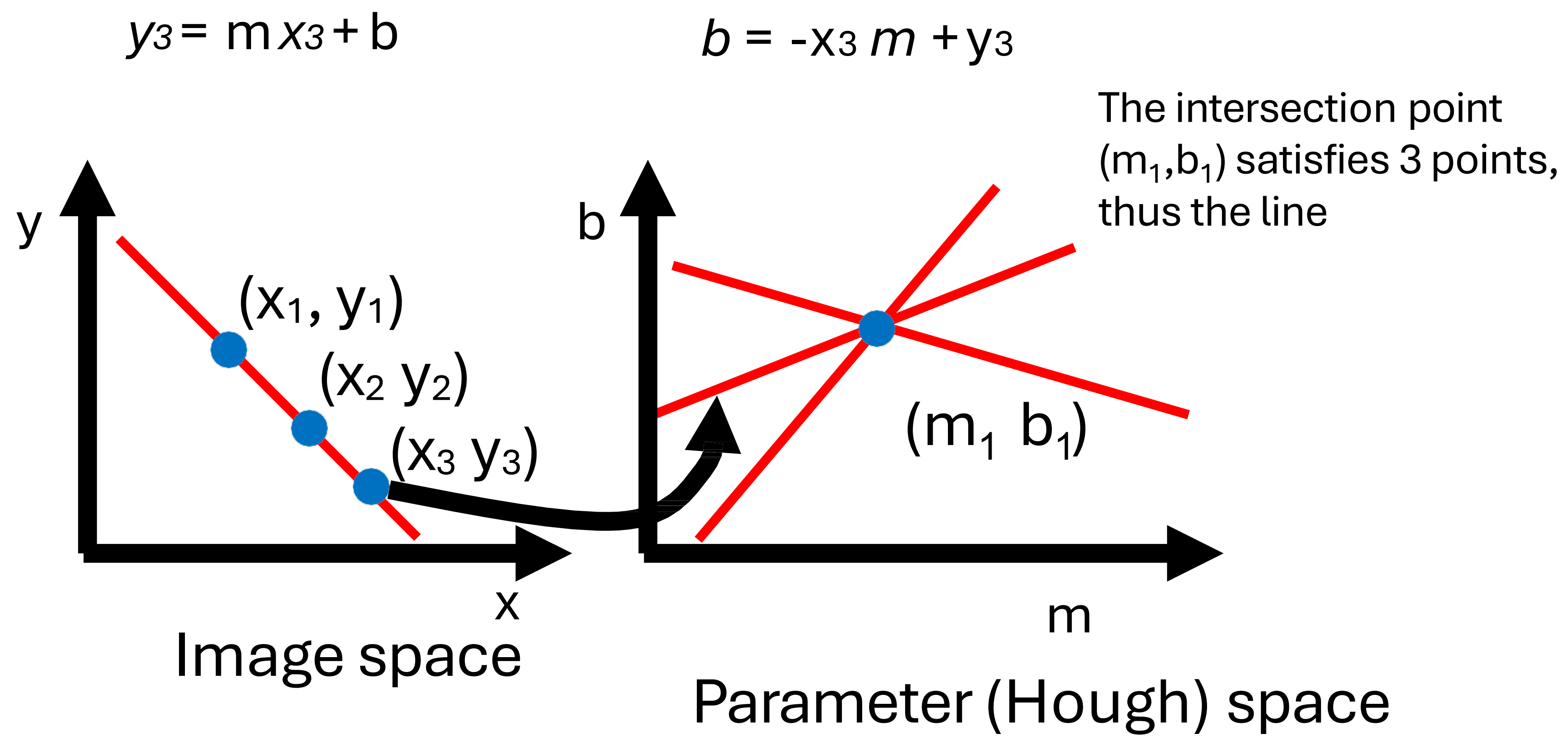
The intersection point satisfies both (x_1, y_1) and (x_2, y_2), thus the line

$y_2 = m x_2 + b$

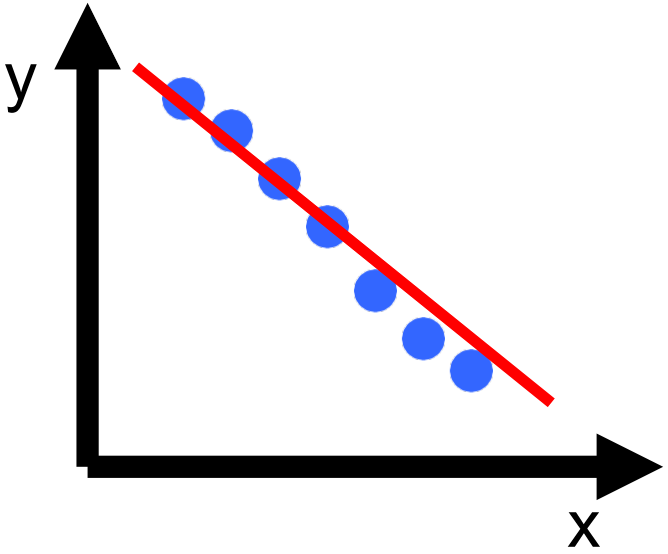
$b = -x_2 m + y_2$

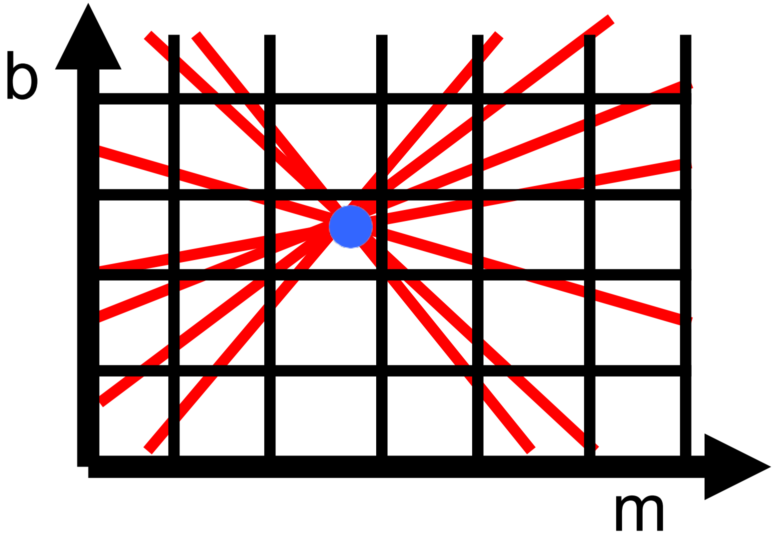


A point (x_3, y_3) is mapped to a line in Hough space

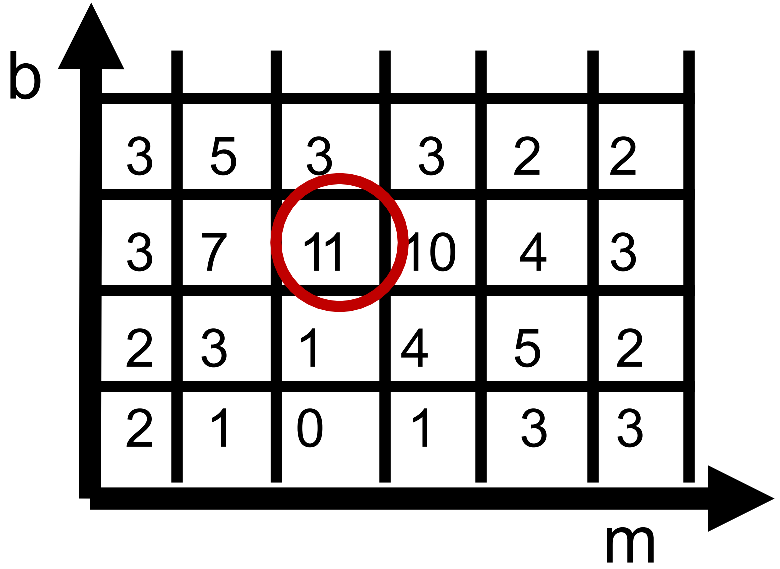


Hough Transform: Accumulator





- For each pixel, draw a line in the discretised Hough space, assigning a value of one to all the discretised positions it passes through
- Process all image pixels

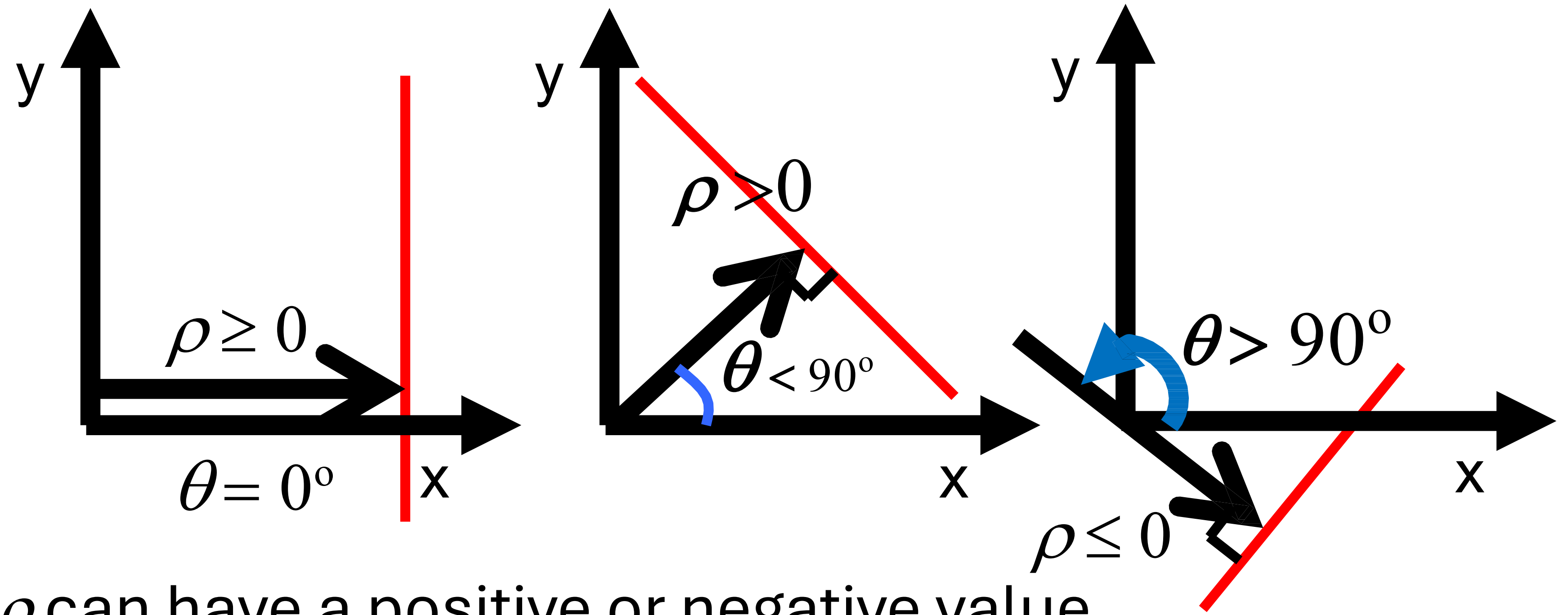


Hough Transform

- Can detect multiple lines in an image
  - from multiple local maxima
- Can easily be extended for circles and ellipses
- Computationally efficient
- Problem: *(m, b)* are unbounded
  - E.g., the slope parameter *m* can have an infinite value

Hough Transform: Polar Form

- Use a polar representation for the parameter space
- $x\cos\theta + y\sin\theta = \rho \rightarrow y = -\dfrac{\cos\theta}{\sin\theta}x + \dfrac{\rho}{\sin\theta}$ $\qquad (0 \le \theta \le 180^o)$



- Note: ρ can have a positive or negative value

A point (x_1, y_1) is mapped to a sinusoid in the polar parameter space

$$x_1 \cos\theta + y_1 \sin\theta = \rho$$

$$\rho = x_1 \cos\theta + y_1 \sin\theta$$
$$= \alpha_1 \sin(\theta + \beta_1)$$

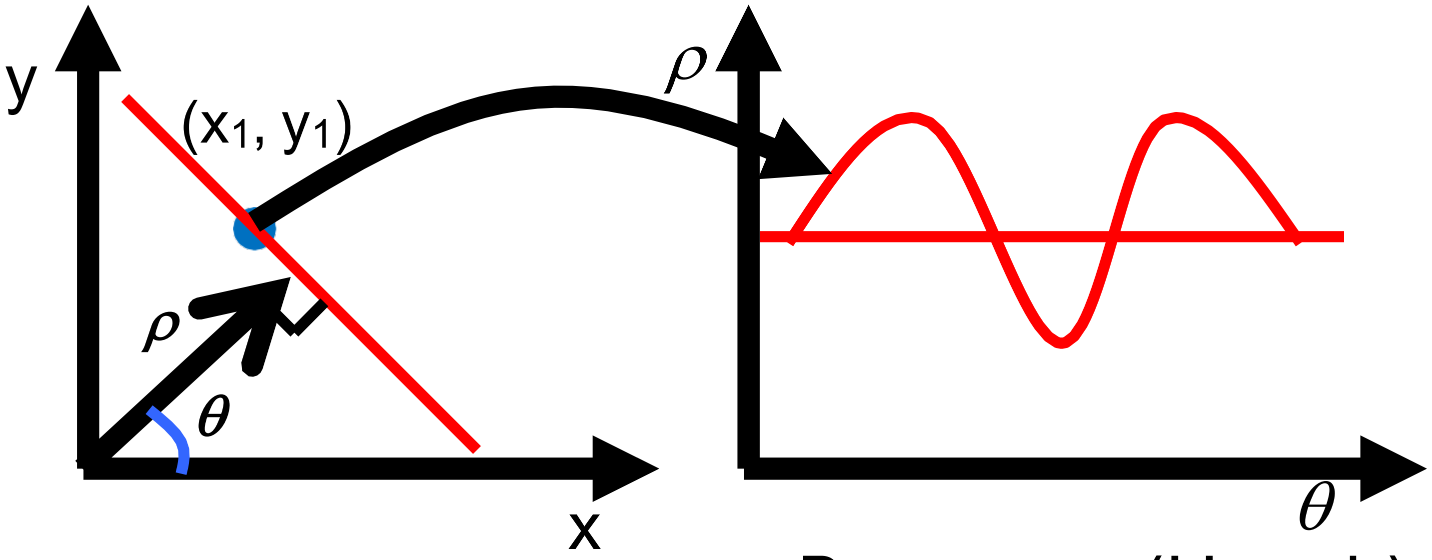


Image space

Parameter (Hough) space

$$0 \le \theta \le 180^o$$

How to map: just divide the right-hand side by $\sqrt{x_1^2 + y_1^2}$ and multiply by the same

A point (x_2, y_2) is mapped to a sinusoid in the polar parameter space

$$x_2 \cos\theta + y_2 \sin\theta = \rho$$

$$\rho = x_2 \cos\theta + y_2 \sin\theta$$
$$= \alpha_2 \sin(\theta + \beta_2)$$

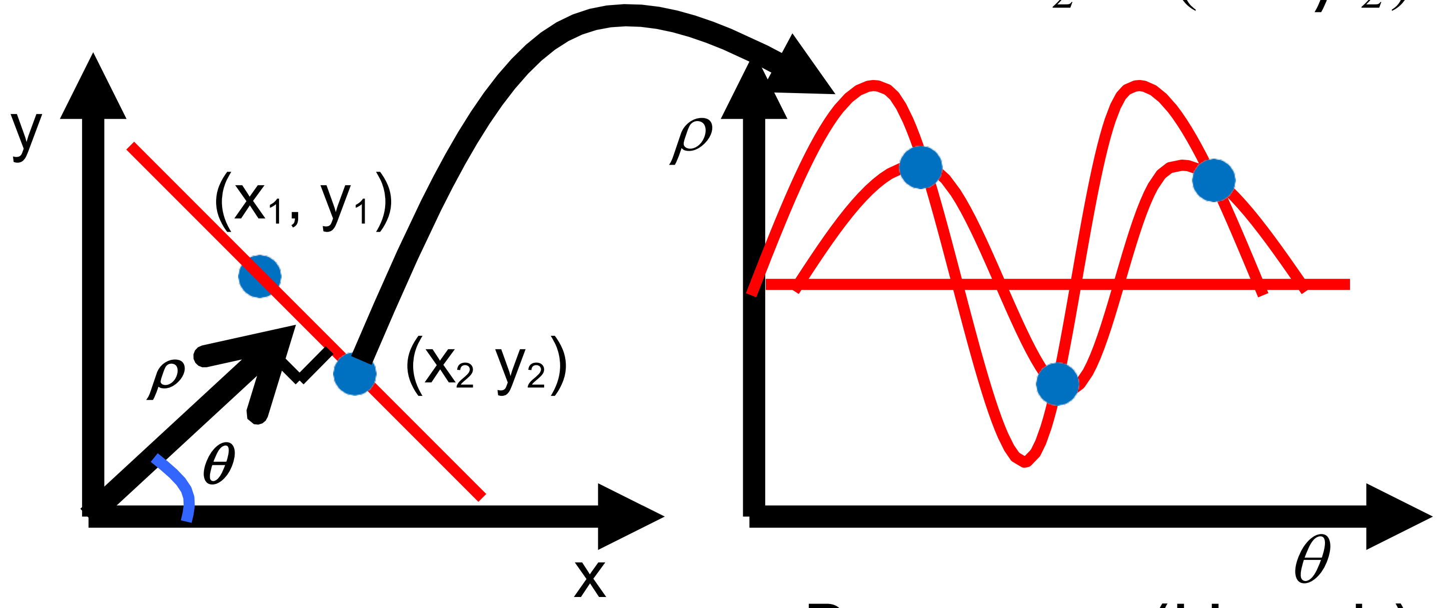


Image space

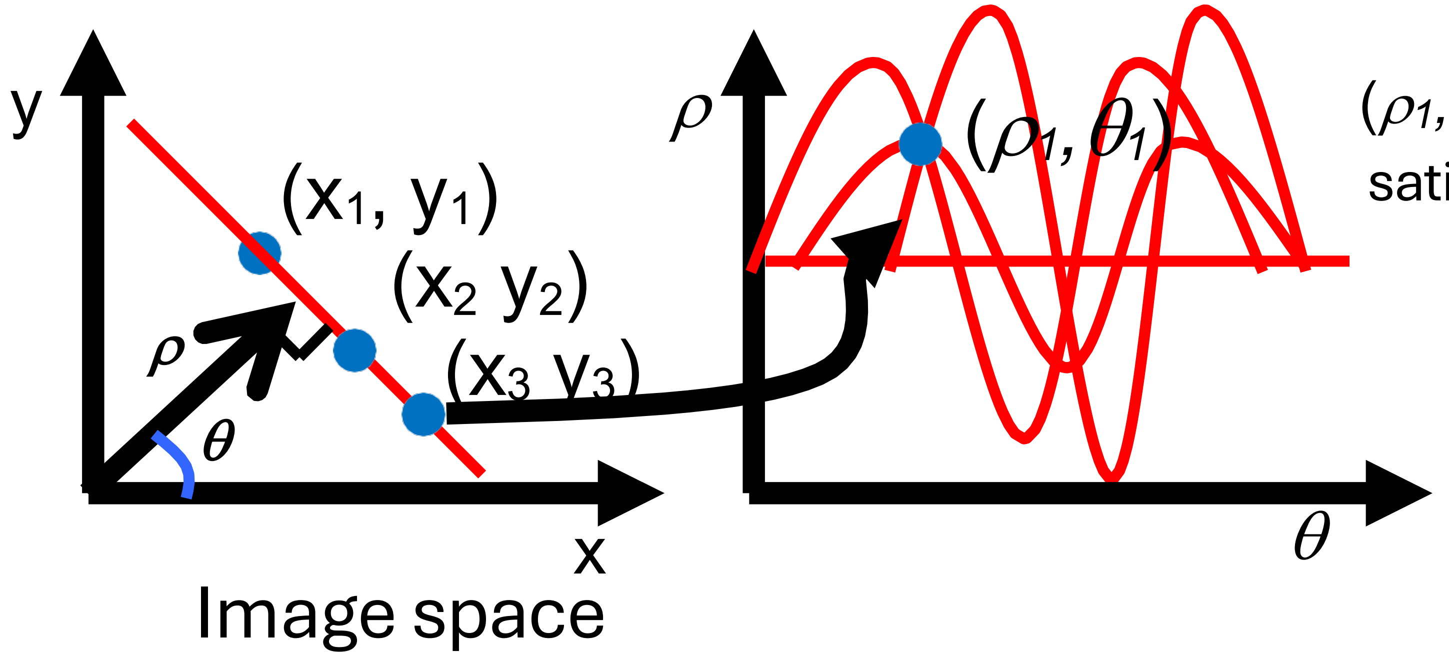
Parameter (Hough) space

$$0 \leq \theta \leq 180^o$$

A point (x_3, y_3) is mapped to a sinusoid in the polar parameter space

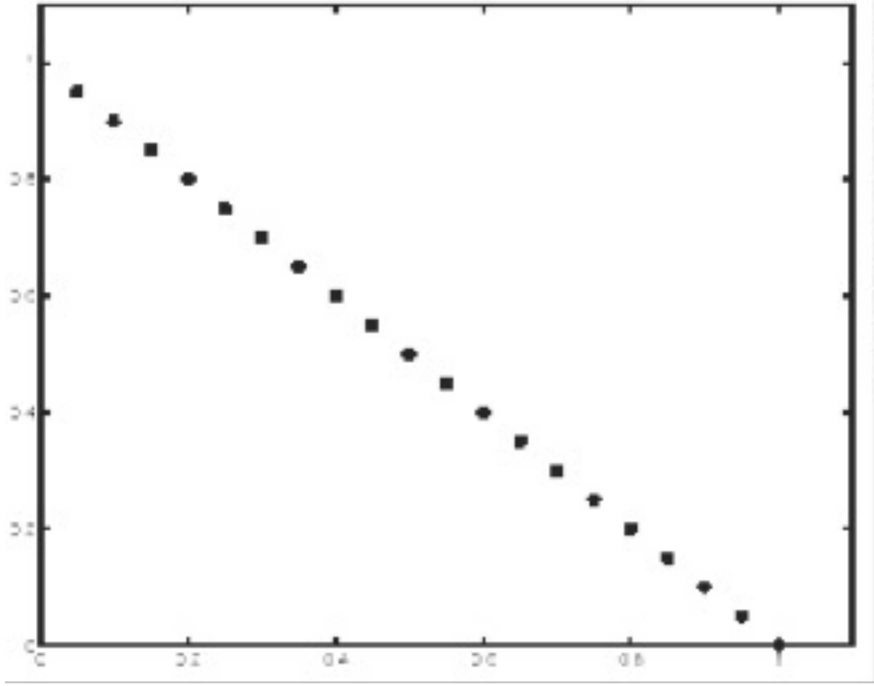
$$x_3 \cos\theta + y_3 \sin\theta = \rho$$

$$\rho = x_3 \cos\theta + y_3 \sin\theta = \alpha_3 \sin(\theta + \beta_3)$$

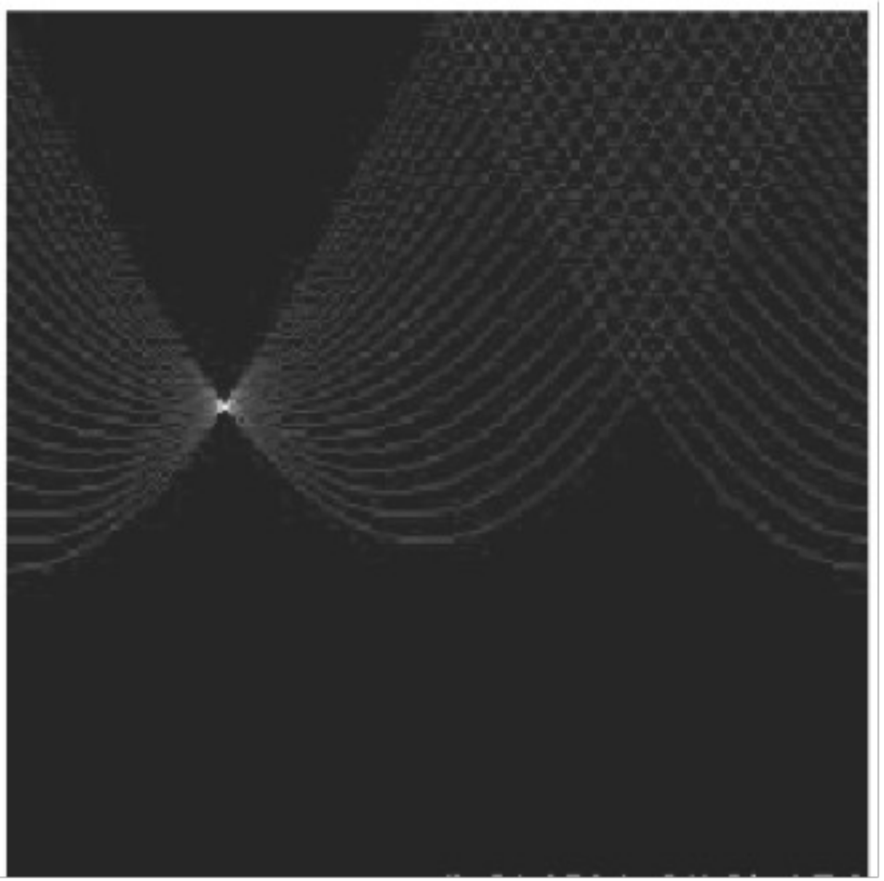


(ρ_1, θ_1) satisfies all 3 pixel points

Hough Transform: Example



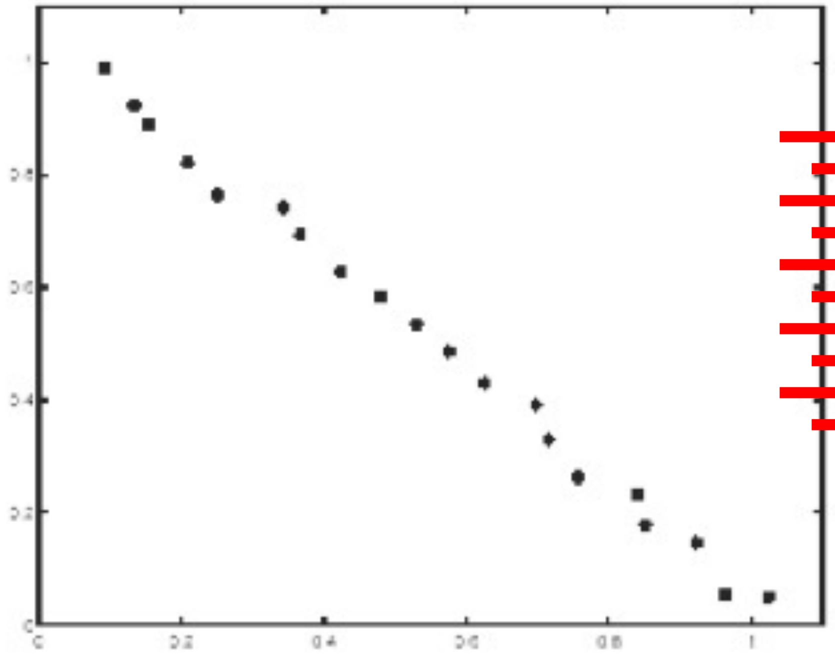
features



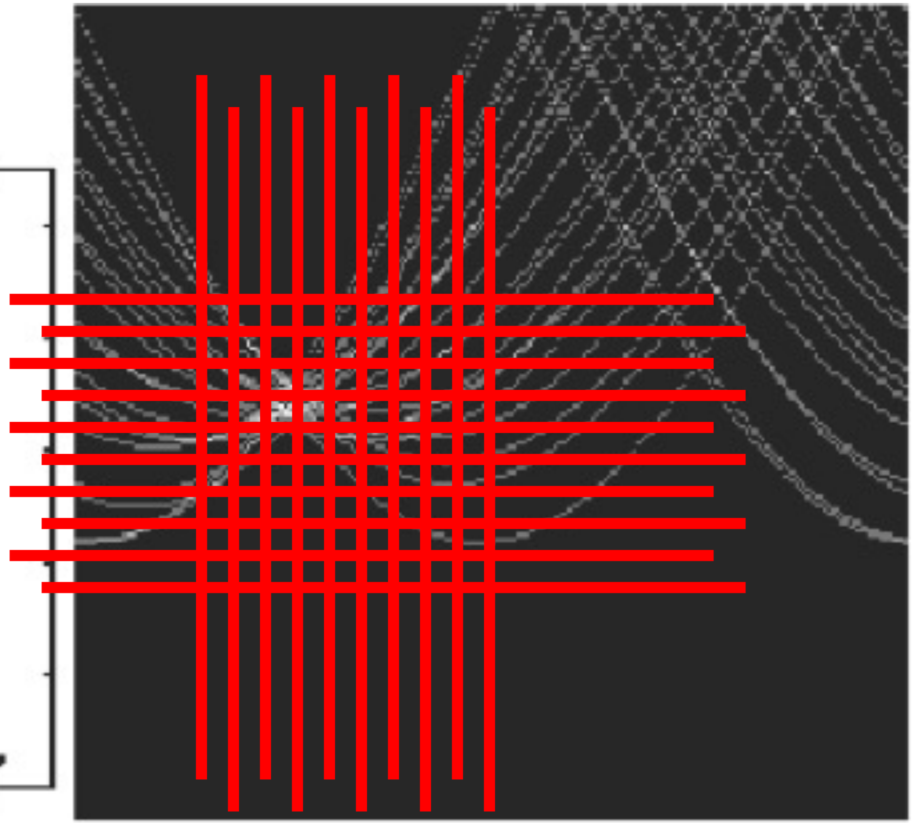
votes

Hough Transform: Example

Noisy data



features



votes

Issue: the grid size needs to be adjusted…

Hough Transform: Example

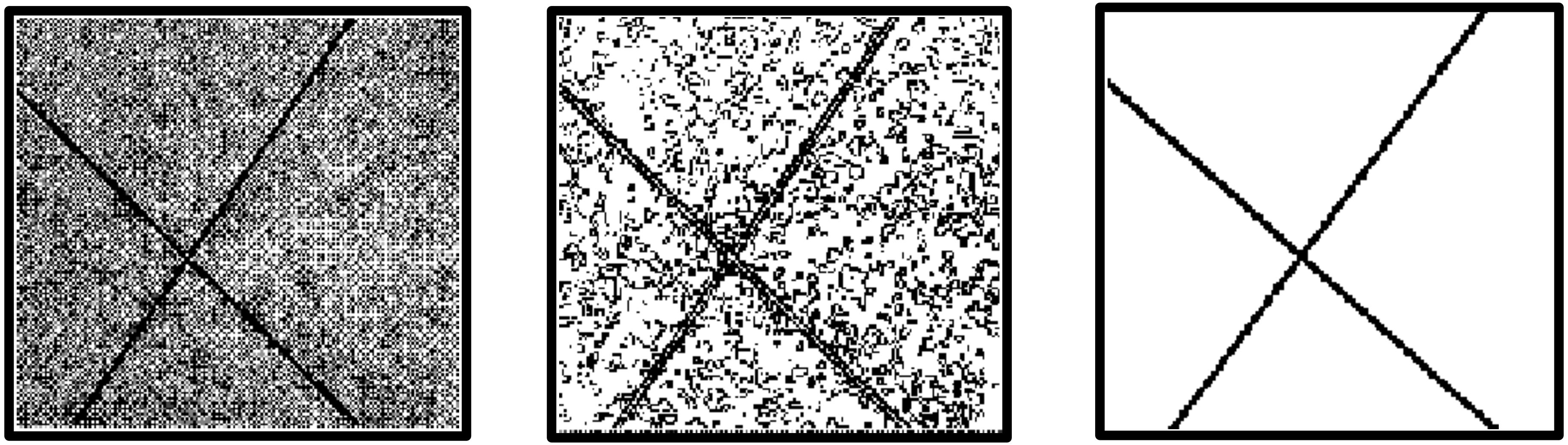
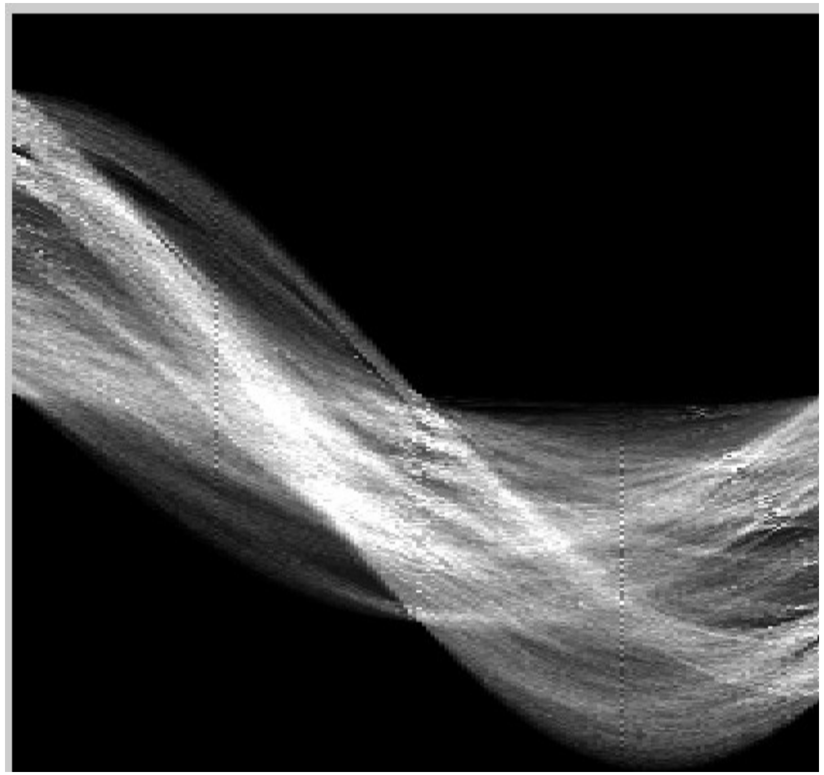


Image **Edge Detection** **Hough Transform**

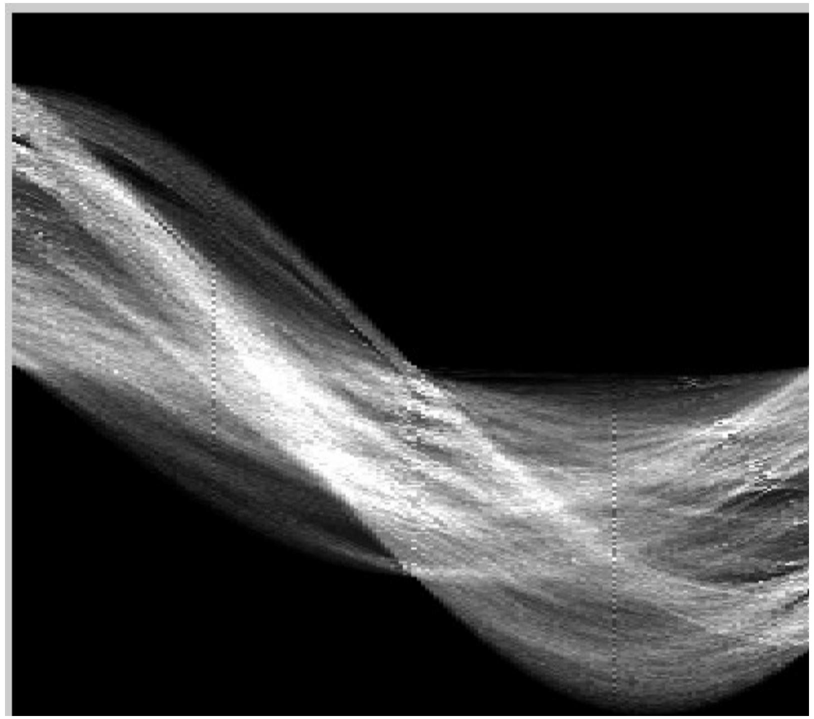
Hough Transform: 1) Image → Canny Edges

Hough Transform: 2) Canny Edges → Votes

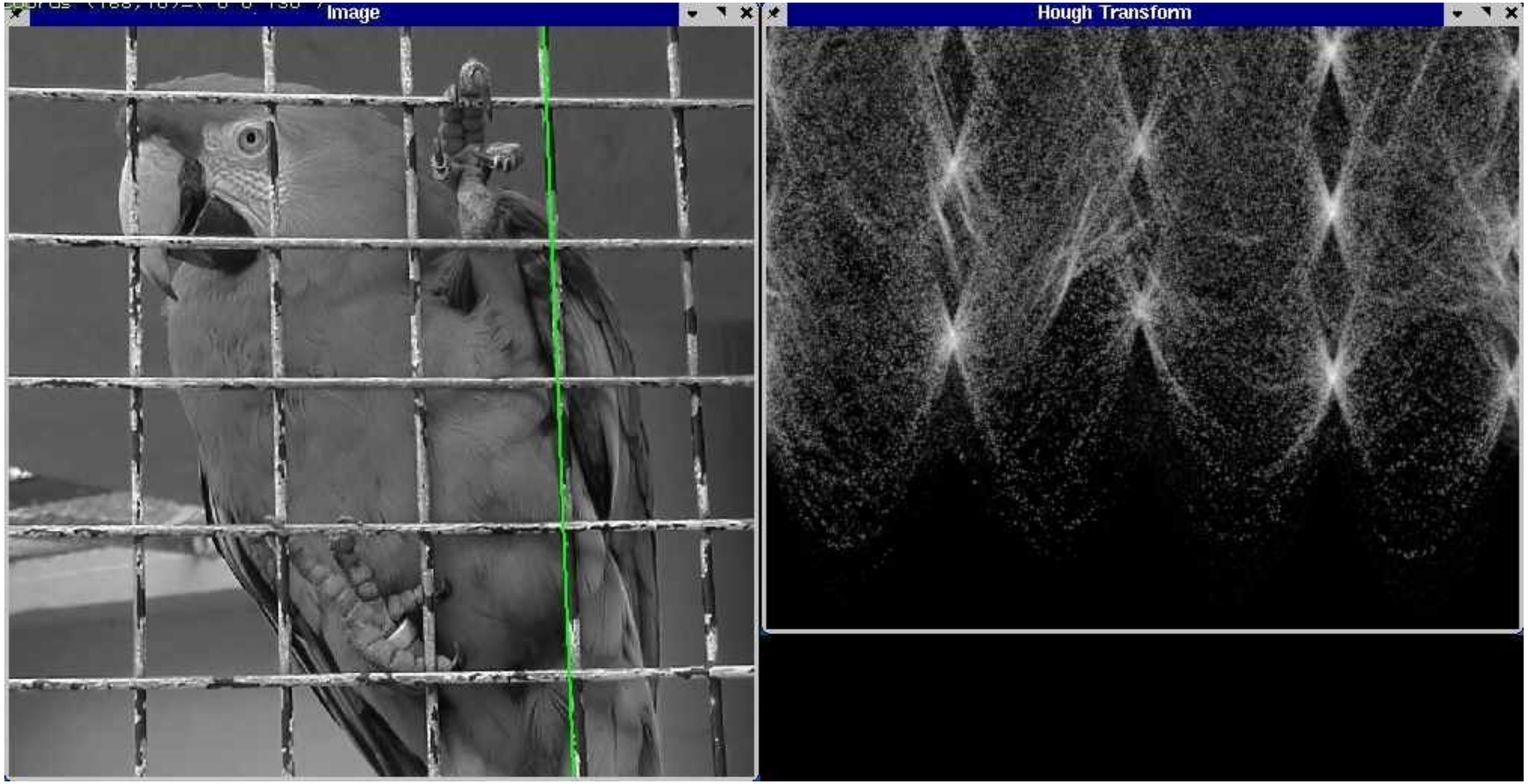


Hough Transform: 3) Votes → Lines

- Find peaks and post-process

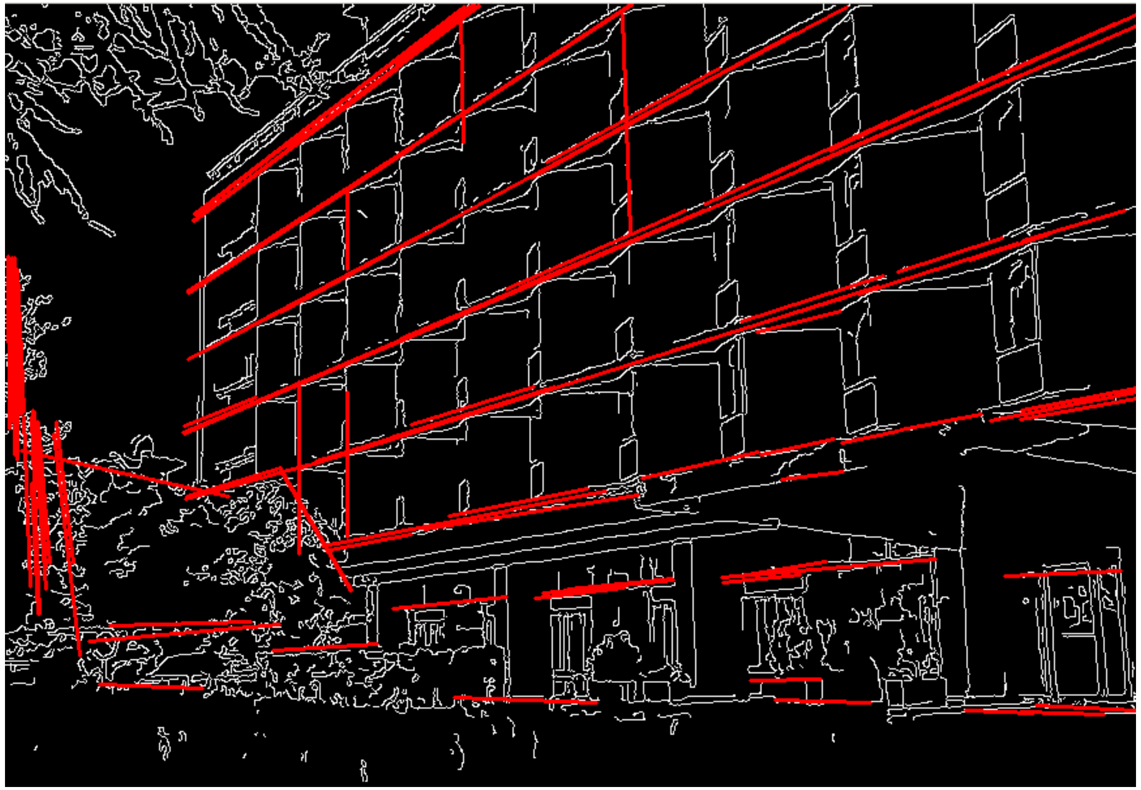


Hough Transform: Example



http://ostatic.com/files/images/ss_hough.jpg

Hough Transform: Example



Hough Transform: Pros and Cons

Pros:

- All points processed independently, so can cope with occlusion
- Some robustness to noise: noisy points are unlikely to contribute consistently to any single bin
- Can detect multiple instances of line/circle/ellipse in a single pass

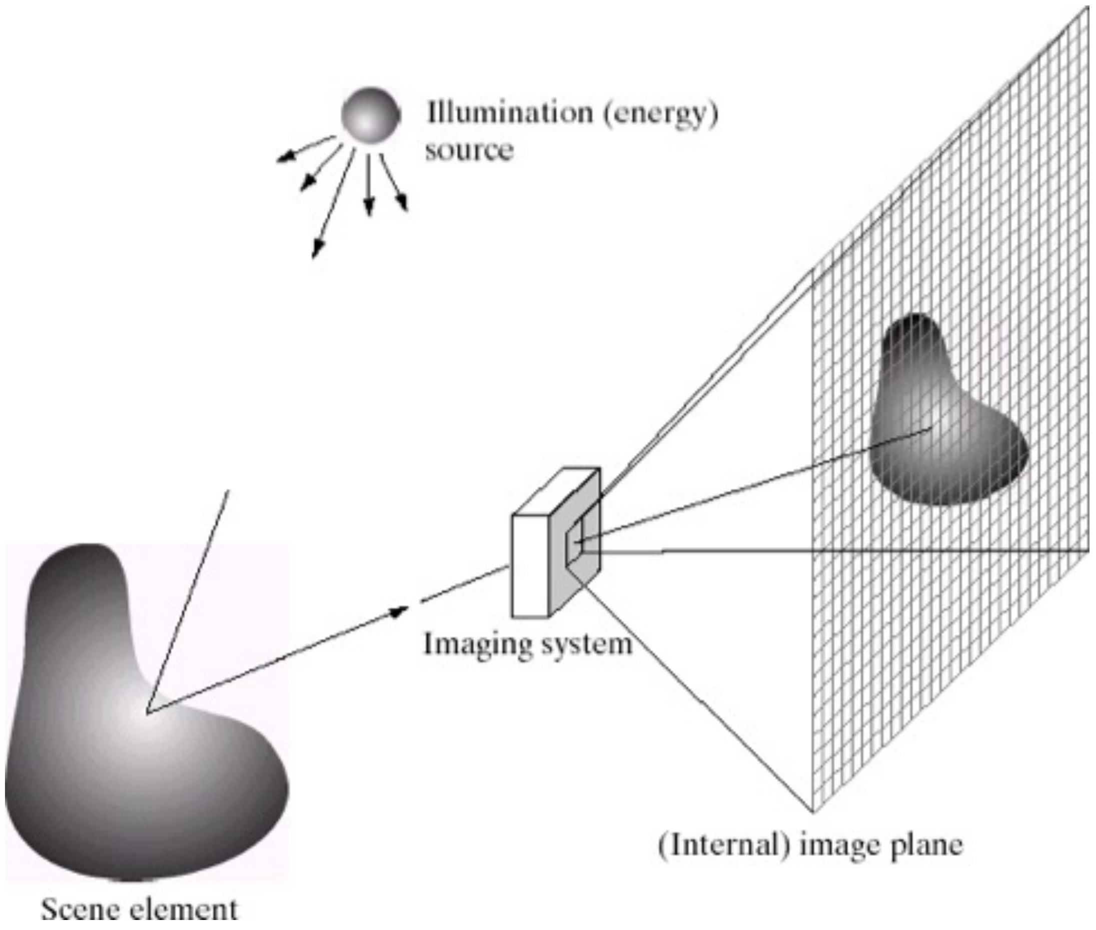
Cons:

- Complexity of search time increases exponentially with the number of model parameters
- Non-target shapes can produce spurious peaks in parameter space
- Quantization: hard to pick a grid size

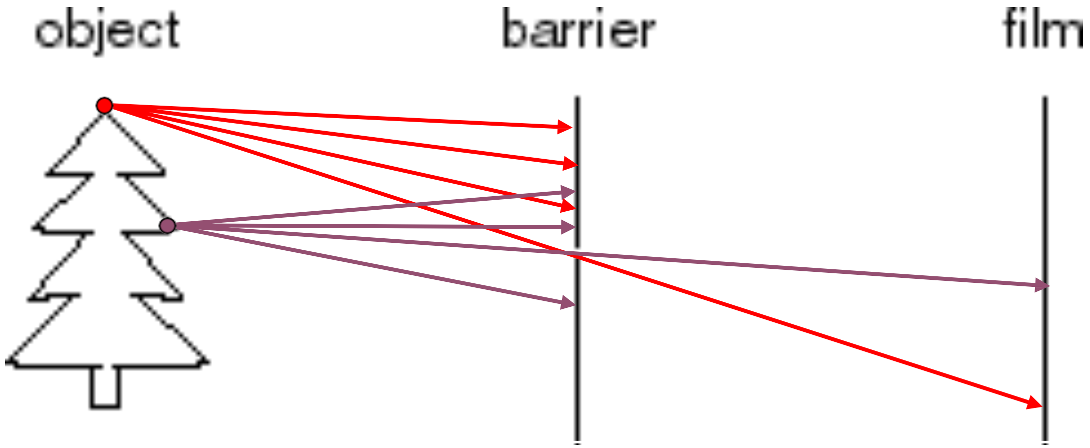
Image Formation (Review)

Pinhole Camera Model

Photometric Image Formation



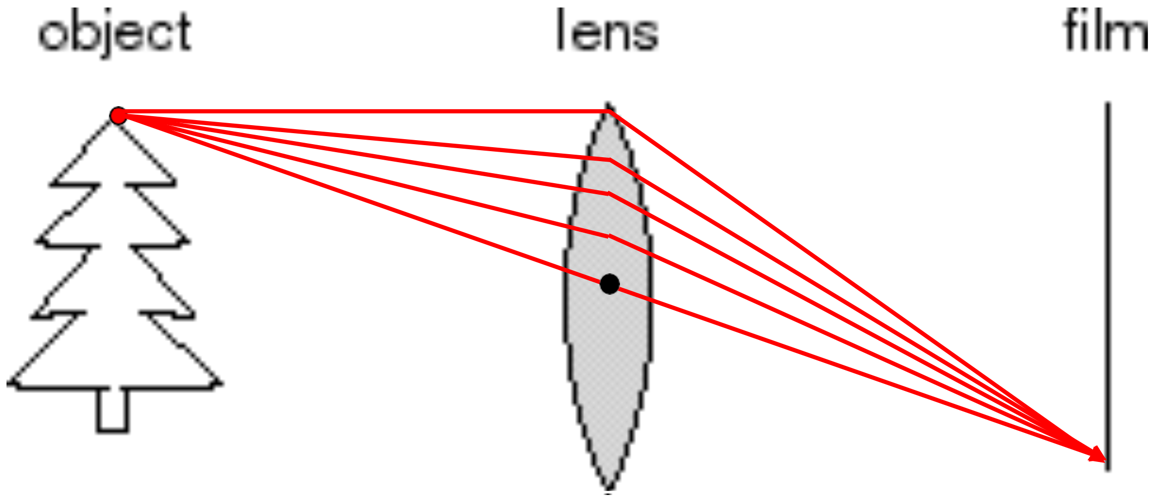
Pinhole Camera



- Add a barrier to block off most of the rays
- This reduces blurring
- The opening is known as the aperture

Image Credit: Steve Seitz

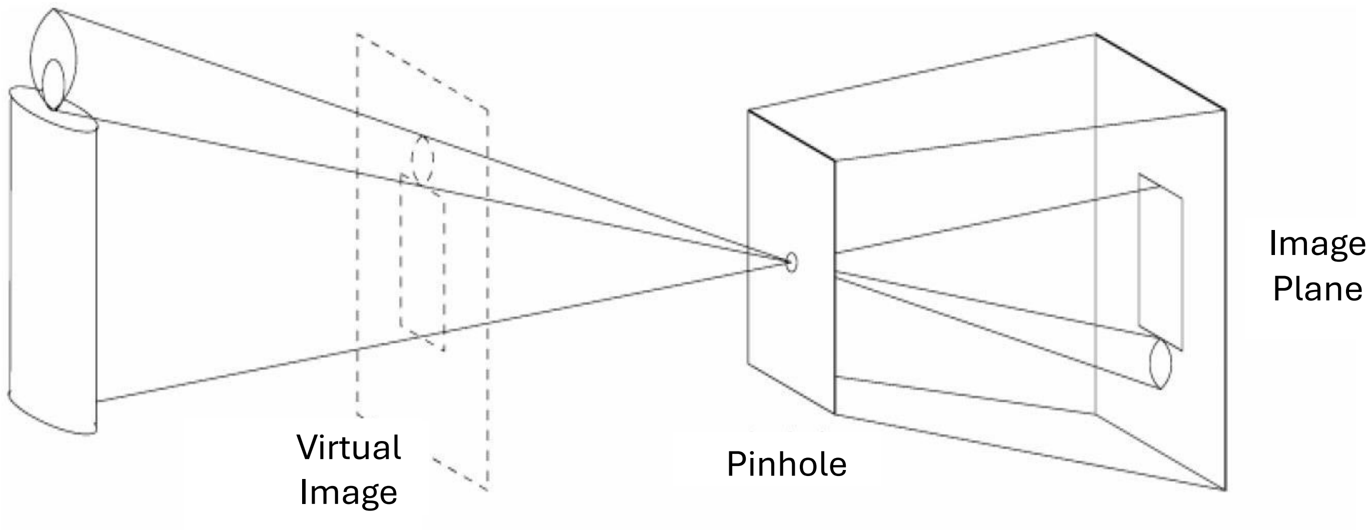
Lensed Camera



- A lens focuses light onto the film: captures more light
- Rays passing through the centre do not deviate

Image Credit: Steve Seitz

Pinhole Camera Model



- Pinhole camera is an abstract model to approximate the imaging process: **perspective projection**
- If we treat the pinhole as a point, only one ray from any given point can enter the camera

Image Credit: Forsyth & Ponce

Camera Projection Matrix and Single View Geometry

Homogeneous Coordinates

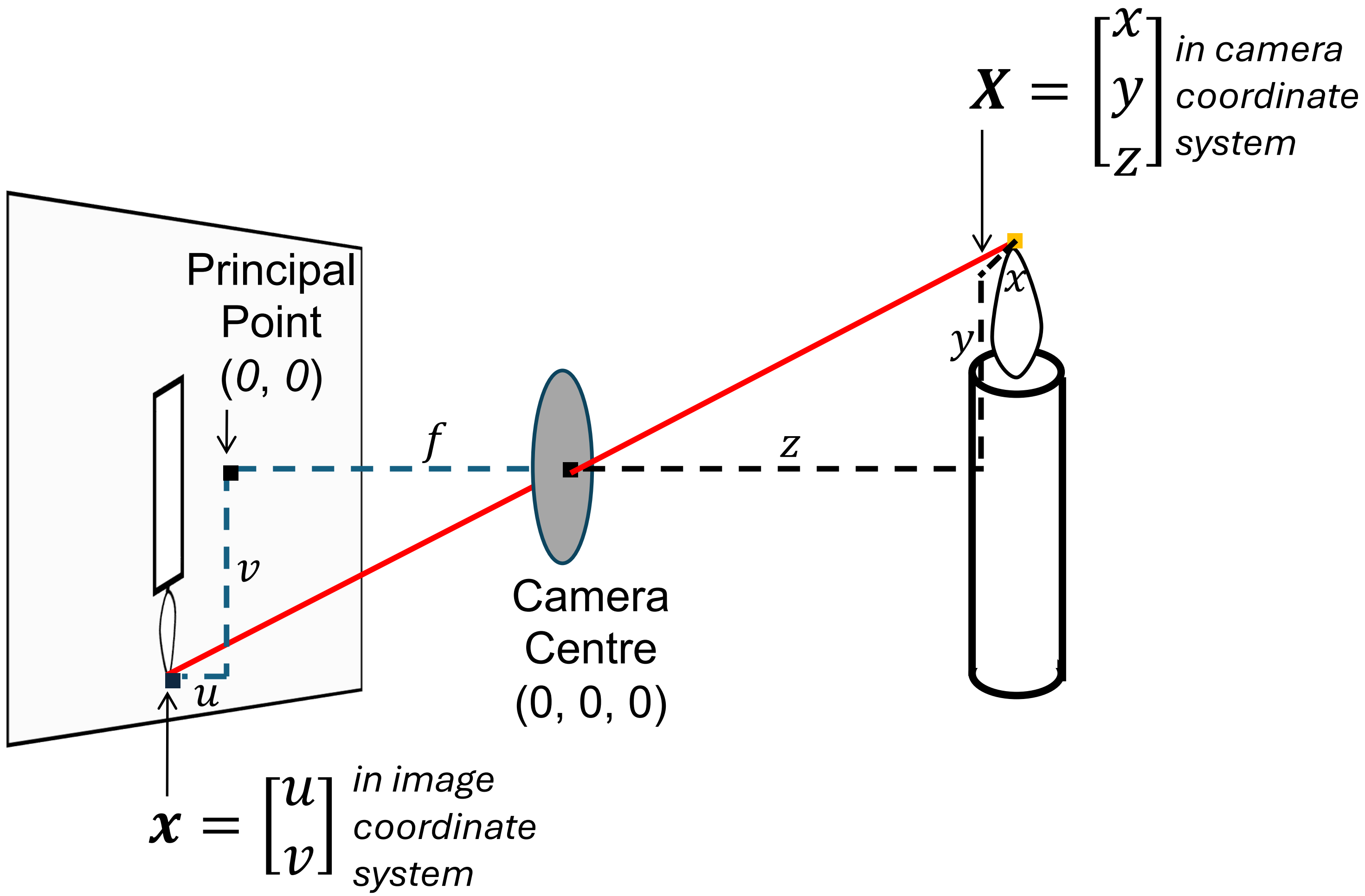
- To homogeneous: $(x, y) \rightarrow (x, y, 1)$ $\qquad$ $(x, y, z) \rightarrow (x, y, z, 1)$
- From homogeneous: $(x, y, w) \rightarrow (x/w, y/w)$ $\qquad$ $(x, y, z, w) \rightarrow (x/w, y/w, z/w)$

- Invariant to scaling: $$k \begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} kx \\ ky \\ kw \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{kx}{kw} \\ \frac{ky}{kw} \end{bmatrix} = \begin{bmatrix} \frac{x}{w} \\ \frac{y}{w} \end{bmatrix}$$

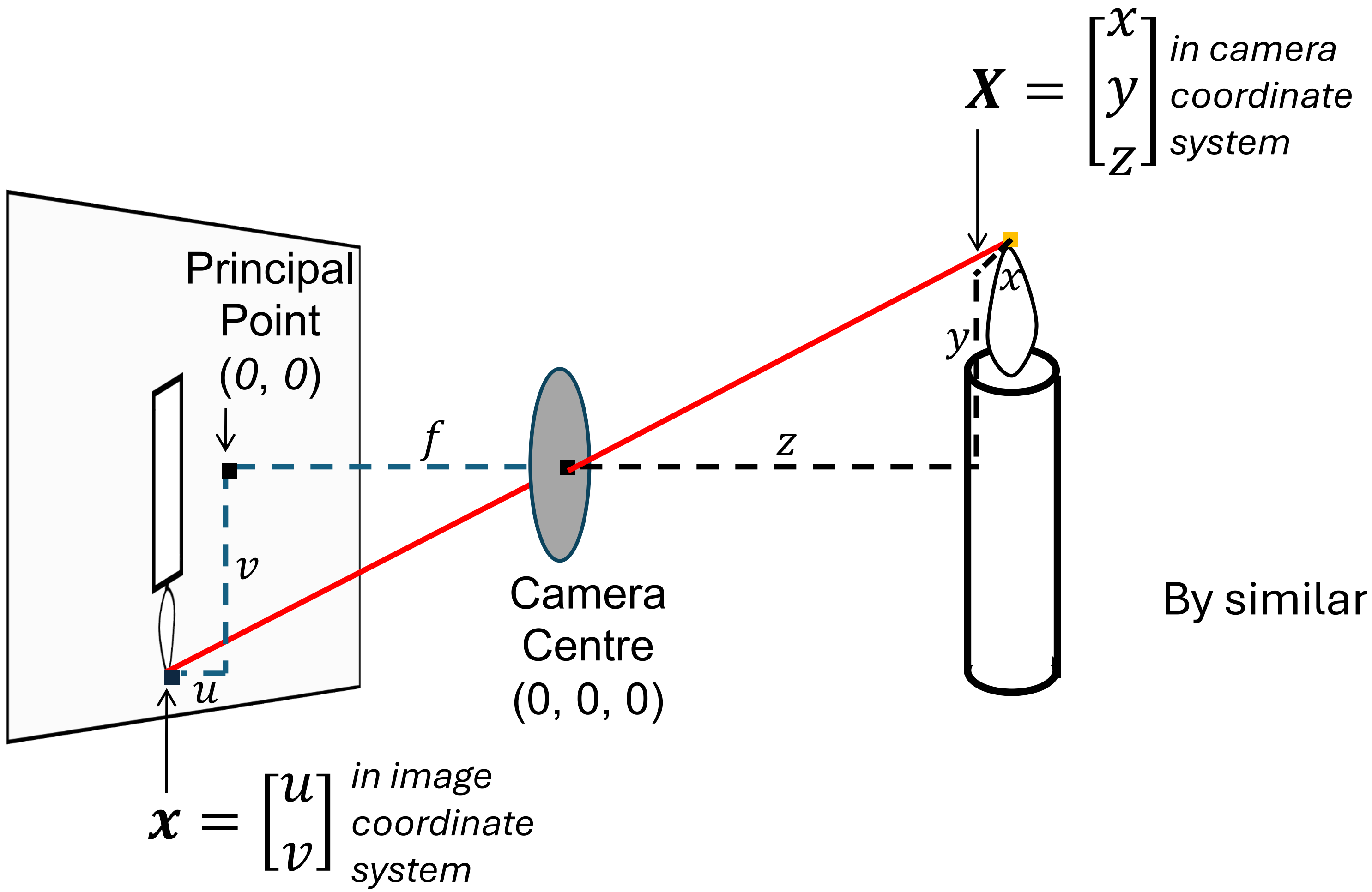
Homogeneous $\qquad$ Cartesian

- A **point** in Cartesian coordinates is a **ray** in homogeneous coords

Perspective Projection



Perspective Projection



By similar triangles:

$$u = f\,\frac{x}{z}$$

$$v = f\,\frac{y}{z}$$

Perspective Projection

- $(x, y, z) \mapsto (\frac{fx}{z}, \frac{fy}{z})$
- Using homogeneous coordinates:
  2D: $(\frac{f\mathrm{x}}{z}, \frac{fy}{z}) \mapsto (\frac{fx}{z}, \frac{fy}{z}, 1) = (fx, fy, z)$
  3D: $(x, y, z) \mapsto (x, y, z, 1)$
- Linear projection in homogeneous coordinates!

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fx \\ fy \\ z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Perspective Projection

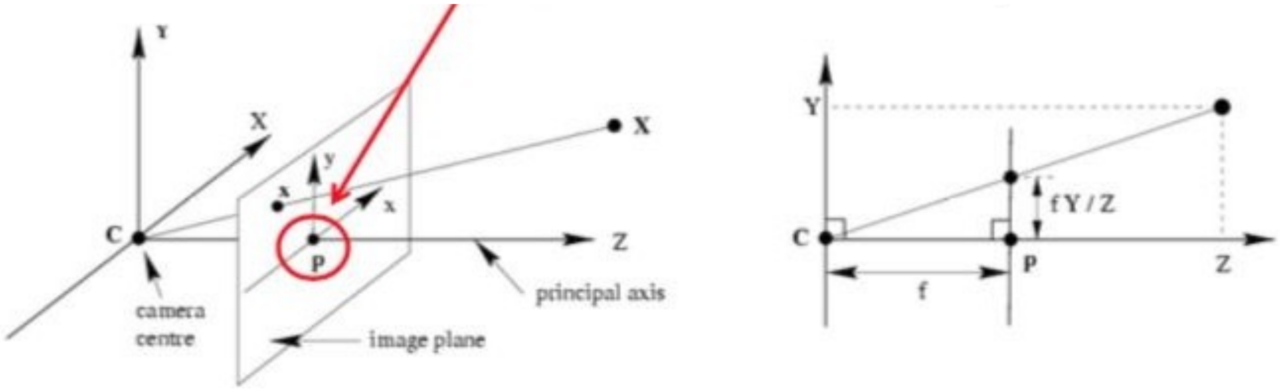
$$\begin{pmatrix} fx \\ fy \\ z \end{pmatrix} = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ & 1 & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$\Leftrightarrow \mathbf{x} = \mathbf{PX} \ \text{ with } \mathbf{P} = \mathrm{diag}(f, f, 1)\,[\,\mathbf{I} \mid 0\,]$$

$= 3\times4$ homogeneous camera projection matrix

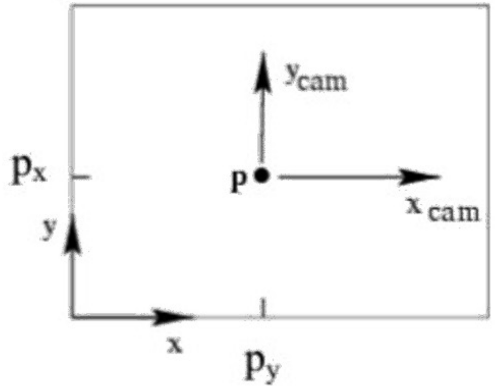
Principal Point

- The point where the principal axis intersects with the image plane



Principal Point Offset

- So far, we have assumed that the origin of points in the image plane is at principal point
- However, origin is often elsewhere (e.g., at image corner)
- Inhomogeneous: $(X, Y, Z) \mapsto \left(fX/Z + p_x, fY/Z + p_y\right)$
- Homogeneous: $\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX + Zp_x \\ fY + Zp_x \\ Z \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$



Camera (Intrinsic) Calibration Matrix

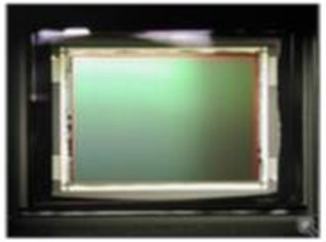
$$\begin{pmatrix} fX + Zp_x \\ fY + Zp_x \\ Z \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$= \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ & 1 & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$= \mathbf{K}[I \mid 0]\mathbf{X}$$

$$\mathbf{K} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} =$$ camera calibration matrix

Rectangular Pixels: CCD Cameras




- Charge-Coupled Device (CCD)
- From image plane to pixel coordinates

$$\mathbf{K} = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} = \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}$$

$$\begin{cases} x_0 = m_x p_x \\ y_0 = m_y p_y \\ \alpha_x = m_x f \\ \alpha_y = m_y f \end{cases}, \text{with} \begin{cases} m_x = \text{\# pixels / unit distance along x} \\ m_y = \text{\# pixels / unit distance along y} \end{cases}$$

Camera Parameters

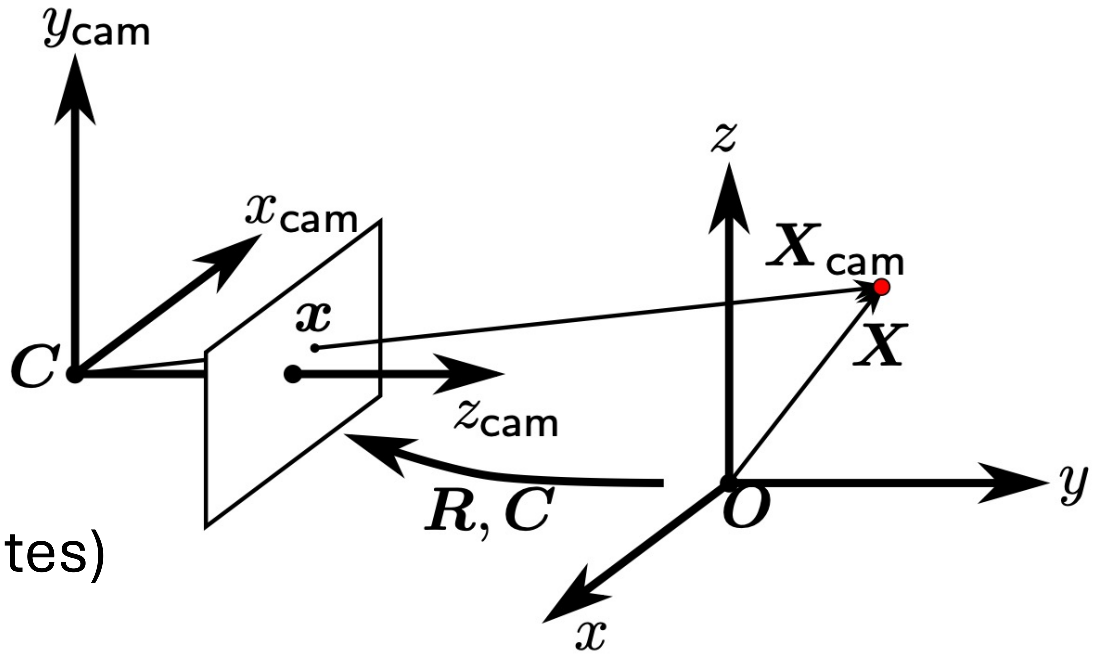
- So far: **intrinsic** camera parameters
  - How to map spatial directions to pixel coordinates
  - Focal length, principal point, pixel width/height

$$\mathbf{K} = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} = \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}$$

- What else? **Extrinsic** camera parameters
  - How to transform a 3D point into the camera frame
  - Depends on the camera rotation and translation

Extrinsic Camera Parameters

- Camera rotation and translation
- **R**: Rotation matrix
  - Orthogonal + unit determinant
  - SO(3)
- **C:** Camera centre (vector)
  - Location of the camera in the world coordinate system
- $\boldsymbol{X}_{cam} = \mathbf{R}(\boldsymbol{X} - \mathbf{C})$ (in camera coordinates)
- $\boldsymbol{X}_{cam} = \begin{bmatrix} \mathbf{R} & -\mathbf{RC} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{R} & -\mathbf{RC} \\ 0 & 1 \end{bmatrix} \boldsymbol{X}$ (in homogeneous coordinates)



Rotation About Coordinate Axes in 3D

- Express 3D rotation as series of rotations around coordinate coordinate axes by angles α, β, γ
- The overall rotation is the product of these elementary rotations:
- $\mathbf{R} = \mathbf{R}_x \mathbf{R}_y \mathbf{R}_z$
- They describe clockwise rotations

$$\mathbf{R}_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix}$$

$$\mathbf{R}_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$\mathbf{R}_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Complete Camera Matrix

- $\mathrm{x} = \mathbf{K}[\,\mathbf{I} \mid 0\,]\mathbf{X}_{\mathrm{cam}}$

$$= \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} [\,\mathbf{I} \mid 0\,] \begin{bmatrix} \mathbf{R} & -\mathbf{RC} \\ 0 & 1 \end{bmatrix} \mathbf{X}$$

$$= \underbrace{\mathbf{K}}_{\text{Camera intrinsics}} \underbrace{\mathbf{R}[\,\mathbf{I} \mid -\mathbf{C}\,]}_{\text{Camera extrinsics}} \mathbf{X}$$

$$= \mathbf{PX}$$

3x4 projective camera matrix

Camera (Intrinsic) Calibration Matrix

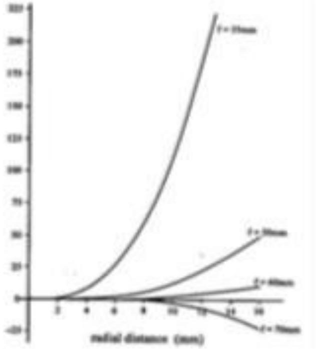
- $\mathbf{K}$ is a 3×3 upper triangular matrix, the "camera calibration matrix"

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

- Four parameters:
  - The scaling in the image x and y directions, α_x and α_y
  - The principal point (x_0, y_0), which is the point where the optical axis intersects the image plane
- The aspect ratio is α_y / α_x

Radial Lens Distortion

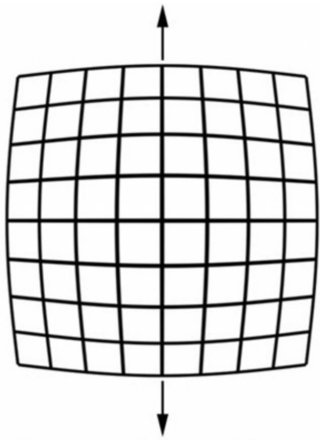
- There is no such thing as a perfect lens
- Straight lines are no longer straight!

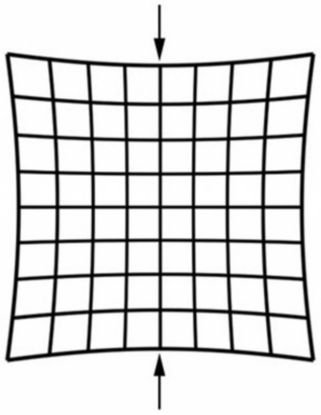


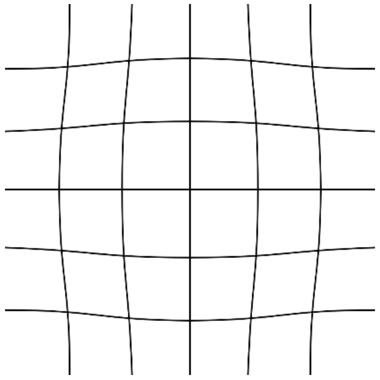
http://foto.hut.fi/opetus/260/luennot/11/atkinson_6-11_radial_distortion_zoom_lenses.jpg

Radial Lens Distortion

- Due to spherical lenses (cheaper)
  - Barrel distortion
    - Image magnification decreases with distance from optical axis
  - Pincushion distortion
    - Image magnification increases with distance from optical axis
  - Mustache distortion
    - A mixture of both types







Radial Lens Distortion

- Model for radial distortion:
  - Change based on distance of point on image plane from principal point
  - If $\mathbf{x} = \mathbf{P}\mathbf{X}_{\text{world}} = \mathbf{K}\mathbf{R}[\,\mathbf{I} \mid -\mathbf{C}\,]\mathbf{X}_{\text{world}} = \mathbf{K}[\,\mathbf{I} \mid 0\,]\mathbf{X}_{\text{cam}}$

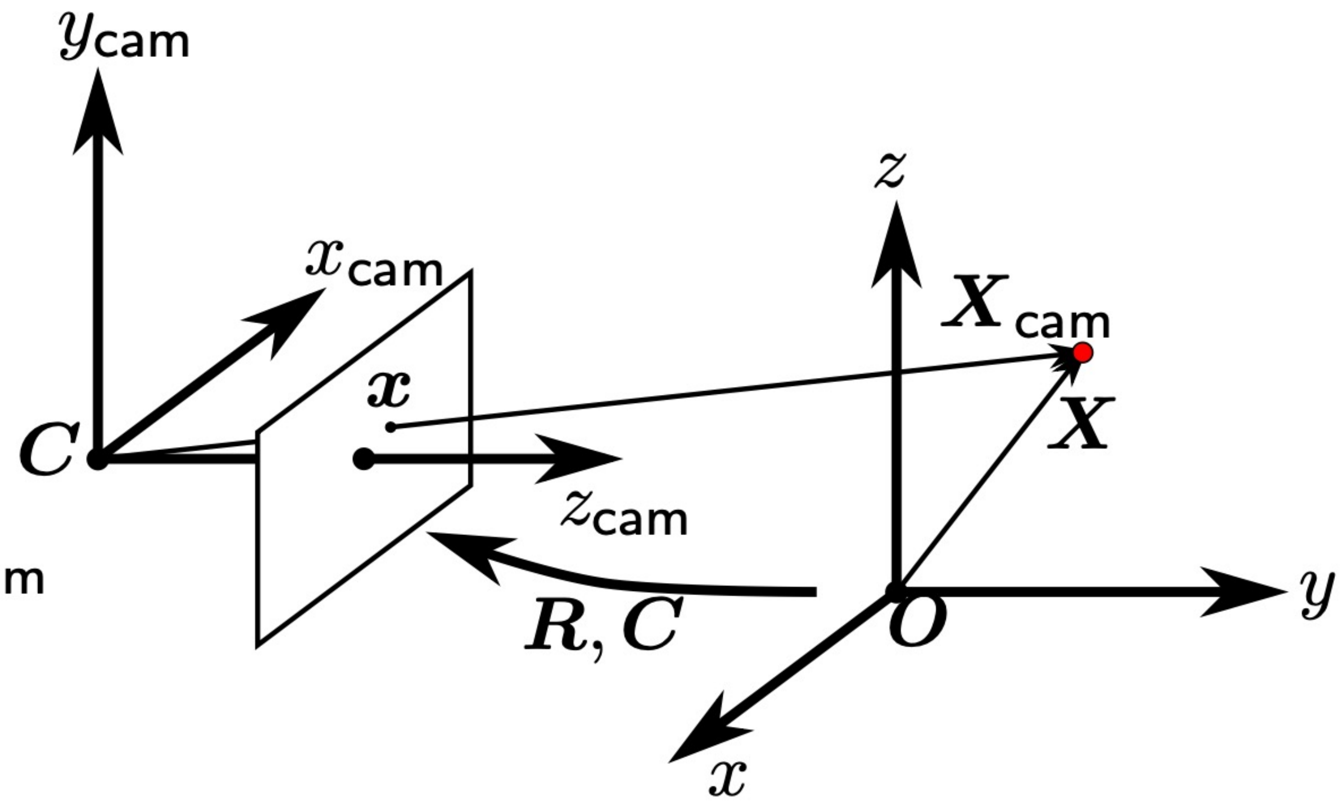
$$= \mathbf{K}\begin{bmatrix} X/Z \\ Y/Z \\ 1 \end{bmatrix} = \mathbf{K}\begin{bmatrix} x_{cam} \\ y_{cam} \\ 1 \end{bmatrix}$$

we change to $\mathbf{x} = \mathbf{K}\begin{bmatrix} r & & \\ & r & \\ & & 1 \end{bmatrix}[\mathbf{I} \mid 0]\mathbf{X}_{\text{cam}}$

with $r = 1 + k_1(x_{cam}^2 + y_{cam}^2) + k_2(x_{cam}^2 + y_{cam}^2)^2$

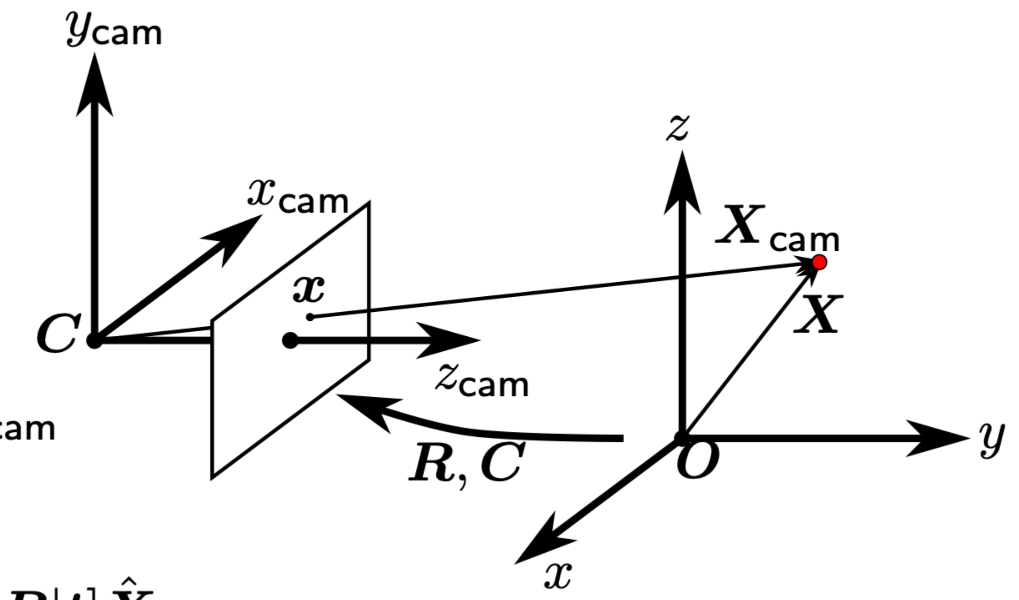
Summary: Camera Projection Matrix

- image plane
- camera centre $\boldsymbol{C}$
- principal axis z_{cam}
- image coordinates $\boldsymbol{x}$
- world coordinates $\boldsymbol{X}$
- camera coordinates $\boldsymbol{X}_{\text{cam}}$



Summary: Camera Projection Matrix

- image plane
- camera centre $\boldsymbol{C}$
- principal axis z_{cam}
- image coordinates $\boldsymbol{x}$
- world coordinates $\boldsymbol{X}$
- camera coordinates $\boldsymbol{X}_{\text{cam}}$



$$
\begin{aligned}
\hat{\boldsymbol{x}} &= \boldsymbol{P}\hat{\boldsymbol{X}} \\
&= \boldsymbol{K}\boldsymbol{R}[\boldsymbol{I}| - \boldsymbol{C}]\hat{\boldsymbol{X}} = \boldsymbol{K}[\boldsymbol{R}|\boldsymbol{t}]\hat{\boldsymbol{X}} \\
&= \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \hat{\boldsymbol{X}} \\
&= \begin{bmatrix} m_x f_x & \gamma & m_x p_x \\ 0 & m_y f_y & m_y p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \hat{\boldsymbol{X}}
\end{aligned}
$$

Next Week

- How to **calibrate** a perspective camera
- How to find the **P** matrix
- How to estimate camera focal length, etc.
- The DLT algorithm

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \propto \underbrace{\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}}_{P_{\{3\times 4\}}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$