

SE 461 Computer Vision

Nazar Khan

PUCIT

Lectures 21, 22 and 23

Announcement

- Talk on Image Restoration
- Friday 2:30 pm, Al-Khwarzmi Lecture Theater
- All of you should attend!

Motivation

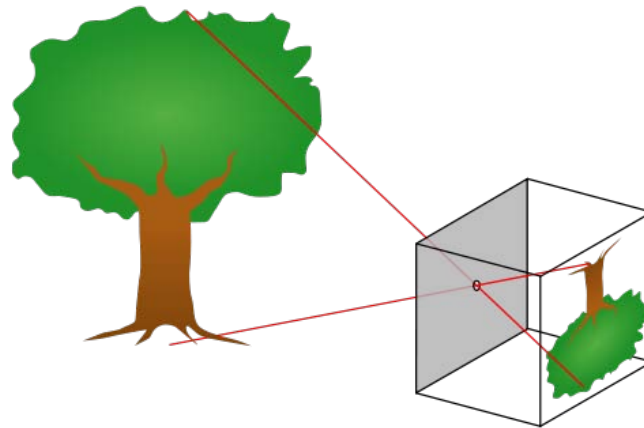
- One central goal of computer vision is to extract information about a 3-D world from 2-D images.
- To understand this mechanism, we first have to investigate how 2-D images arise from a 3-D world.
- First we consider a single camera (monocular vision).
- Usually the imaging process is modelled with the so-called pinhole camera model.
- This requires some single view projective geometry.
- More advanced situations such as two cameras (binocular vision, stereo vision) and the corresponding epipolar geometry will be treated in the next lecture.

Camera

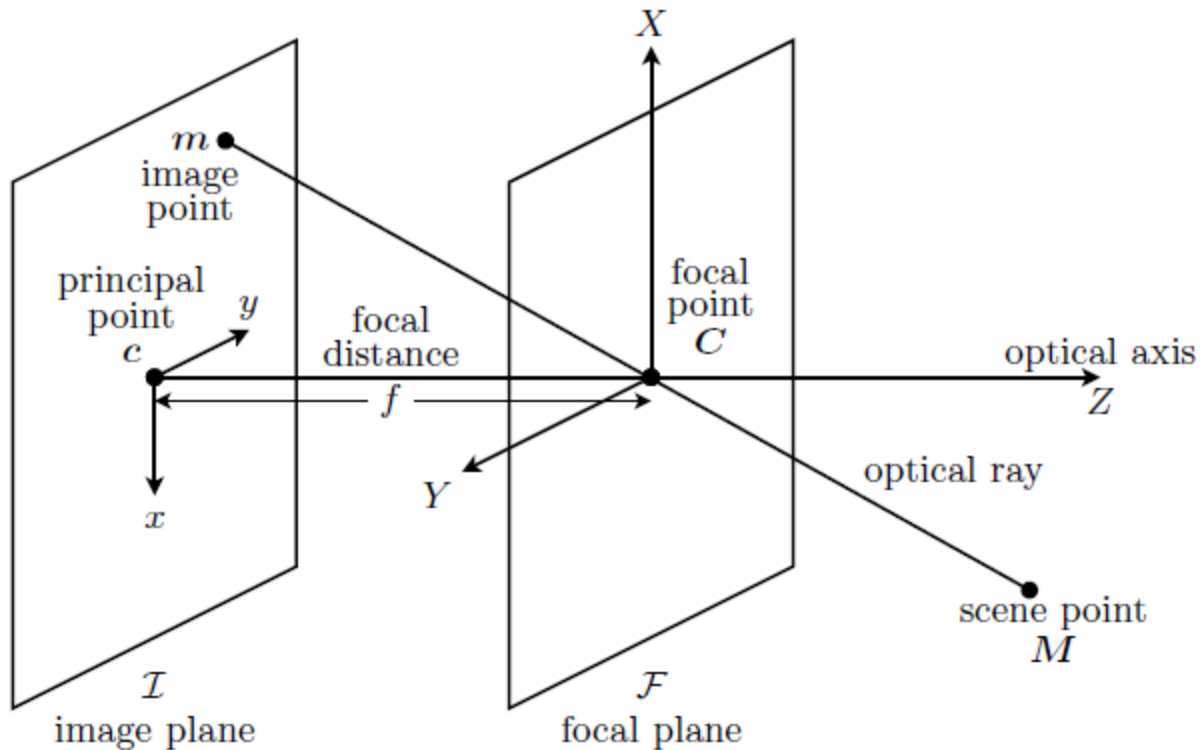
- A camera **projects** 3D world points to 2D pixel coordinates.
- In this lecture, we study the whole process of going from 3D world coordinates to 2D image pixel coordinates.
- Summary: **The whole process can be encoded in a 3x4 camera projection matrix P .**
 - $x = PX$ where x is the 2D pixel location of the 3D world point X when projected by camera P .

Pinhole Camera

- A dark box with just a small hole.
 - No lens.
- Light from the scene passes through the pinhole and projects an inverted image of the scene on the side opposite to the pinhole.



Pinhole Camera Model



Pinhole camera model. Author: M. Mainberger (2010).

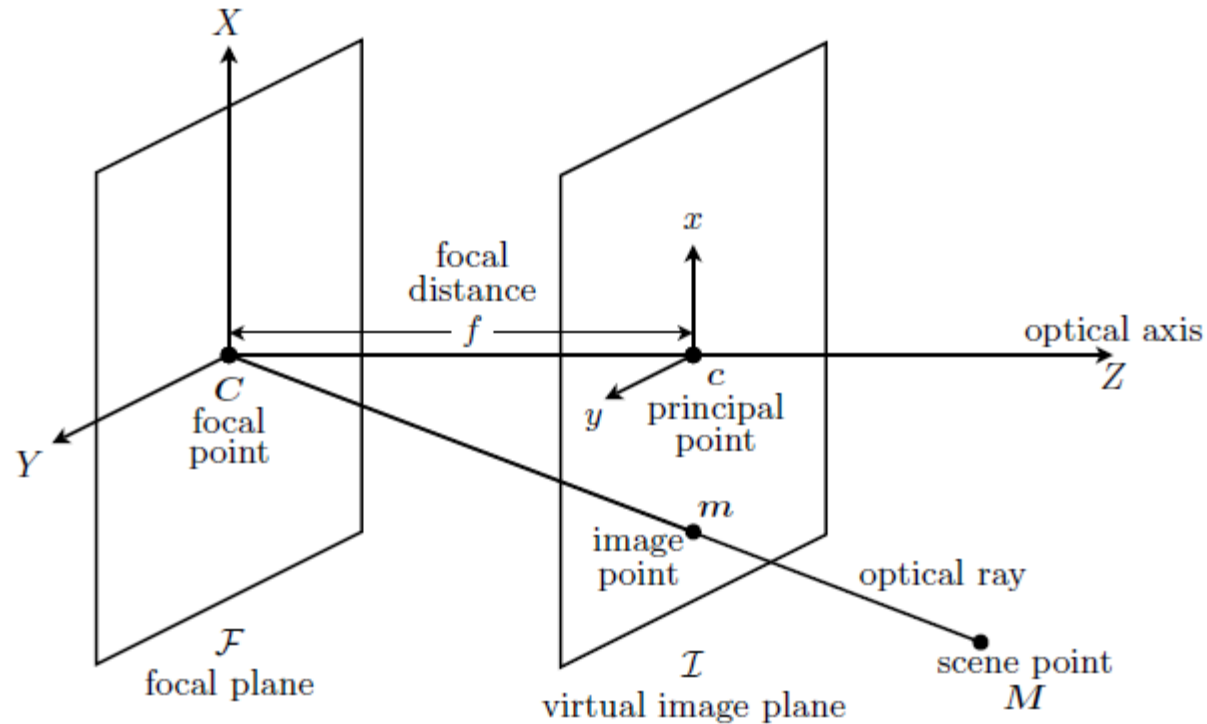
Pinhole Camera Model

- Simple but fairly realistic model of a camera system.
- Perspective projection of the 3-D space onto a 2-D image plane
- Maps 3D camera coordinates $M = (X, Y, Z)^T$ with centre C to 2D image coordinates $m = (x, y)^T$ with centre c .
- Notations:
 - M : scene point
 - C : focal point, camera centre, optical centre: location of the pinhole
 - F : focal plane: specifies camera orientation, contains focal point C
 - optical axis, principal axis : orthogonal to focal plane, passes through C
 - I : image plane: parallel to focal plane
 - f : focal distance: distance focal plane F – image plane I
 - c : principal point: intersection between image plane and optical axis
 - optical ray: passes through M and C
 - m : image point: intersection between optical ray and image plane

Pinhole Camera Model

- **Drawback** of this representation: The projective mapping inverts the orientation. Objects oriented upwards in the real world appear downwards in the image.
- **Simplification:** Instead of the image plane behind the focal plane, consider a (virtual) image plane in front of the focal plane (with distance f).
- In this way, objects that are oriented upwards in the real world appear upwards in the image.

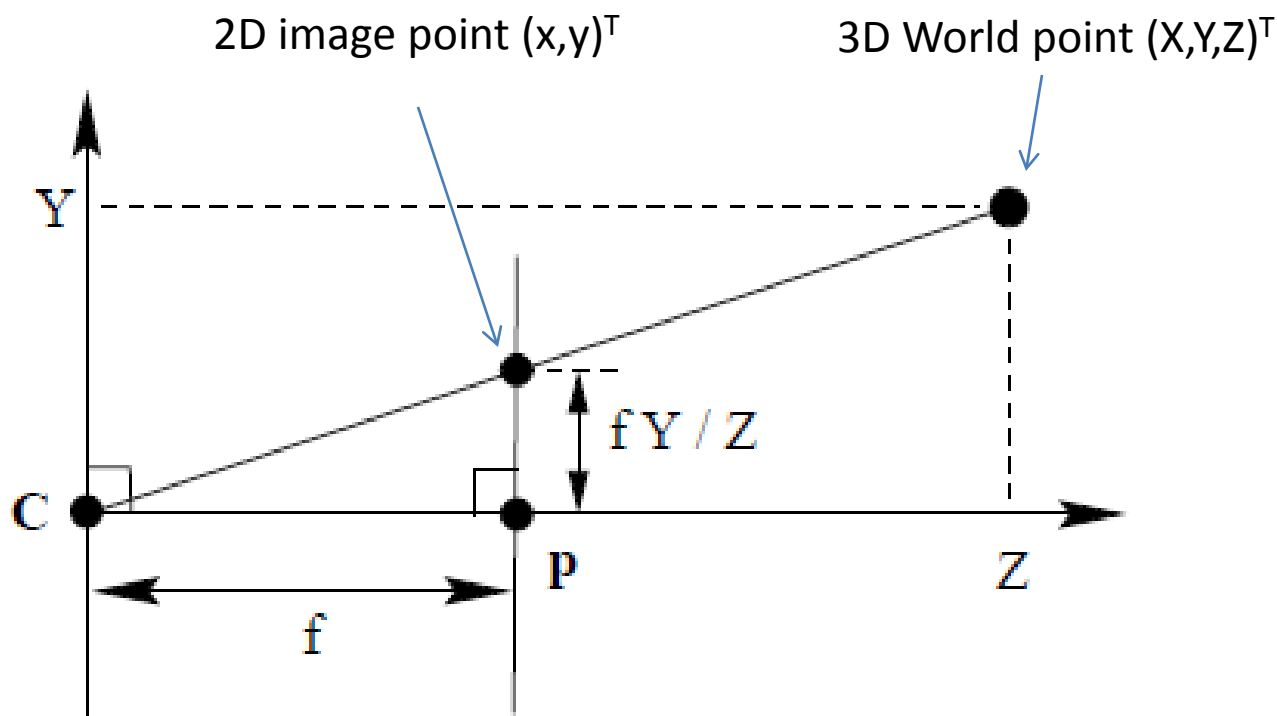
Pinhole Camera Model



Pinhole camera model with virtual image plane in front of the focal plane. Author: M. Mainberger (2010).

Pinhole Camera Model

- Using similar triangles, $x/X = y/Y = f/Z$
- So $x=fX/Z$ and $y=fY/Z$



Pinhole Camera Model

- **Non-uniqueness:** All points on the optic ray are mapped onto the same image point: Two points $M1 := (X, Y, Z)^T$ and $M2 := (wX, wY, wZ)^T$ are mapped to the same image point $m = (x, y)^T$.
 - Thus, the depth information is lost.
- The equations $x=fX/Z$ and $y=fY/Z$ describe the projective geometry, but is unpleasant to work with:
 - It is a nonlinear transformation between the 3D camera coordinates $(X, Y, Z)^T$ and the 2D image coordinates $(x, y)^T$, that involves divisions.
 - Remedy: Homogenous Coordinates

Homogenous Coordinates

- An elegant tool for describing the nonlinear projective camera geometry by means of matrices, i.e. linear mappings.
- The price one has to pay for this is one additional coordinate.

Homogenous Coordinates

- From Greek
 - homos = same
 - genos = kind
- Transformation from standard coordinates to homogeneous coordinates:
 - $(x,y)^T \rightarrow (wx,wy,w)^T$ with some arbitrary scaling factor $w \neq 0$.
- For the back-transformation, pick the first two coordinates and divide by the third.

Camera Projection Matrix P

- The non-linear equations $x=fX/Z$ and $y=fY/Z$ describe the projective geometry in R^3
- In homogenous space P^3 , we can write them as a matrix-vector multiplication (i.e., linear equations)

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$wm = PM$$

Camera Projection Matrix P

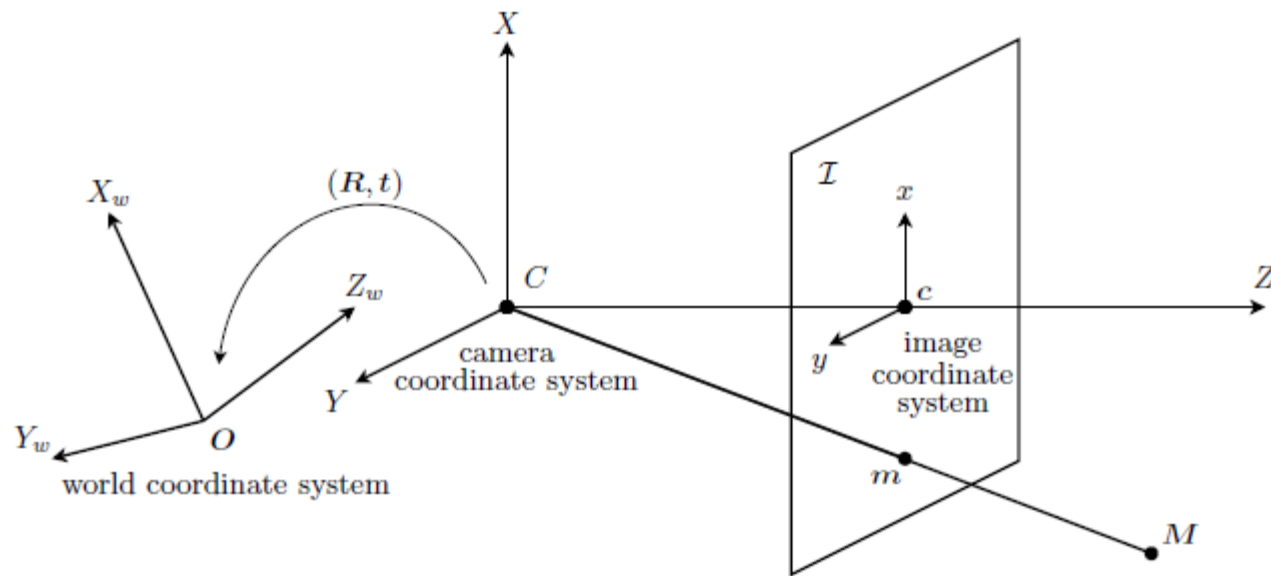
Short notation of the projection equations

$$w\tilde{m} = P\tilde{M}$$

where \sim denotes the additional component 1

$$\tilde{m} = \begin{bmatrix} m \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$
$$\tilde{M} = \begin{bmatrix} M \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Extrinsic and Intrinsic Camera Parameters



World coordinate system (X_w, Y_w, Z_w) , camera coordinate system (X, Y, Z) , and image coordinate system (x, y) . The transitions between these coordinate systems depend on the position of the camera and its internal characteristics. Extrinsic and intrinsic camera parameters characterise these transformations. Author: M. Mainberger (2010).

Extrinsic Camera Parameters

- Denote the position of the world coordinate system relative to the camera coordinate system.
- In homogeneous coordinates, 3D transformations such as translations and rotations can be expressed by multiplication with 4×4 matrices.
- Rotation and translation of world coordinates by R and T is described by multiplication with

$$TR = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Extrinsic Camera Parameters

- 6 degrees of freedom
 - 3 rotation angles: 1 around each axis
 - 3 translation parameters: 1 along each axis
- Since they only depend on the camera orientation, but not on internal camera specifics, they are called **extrinsic camera parameters**.

$$TR = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

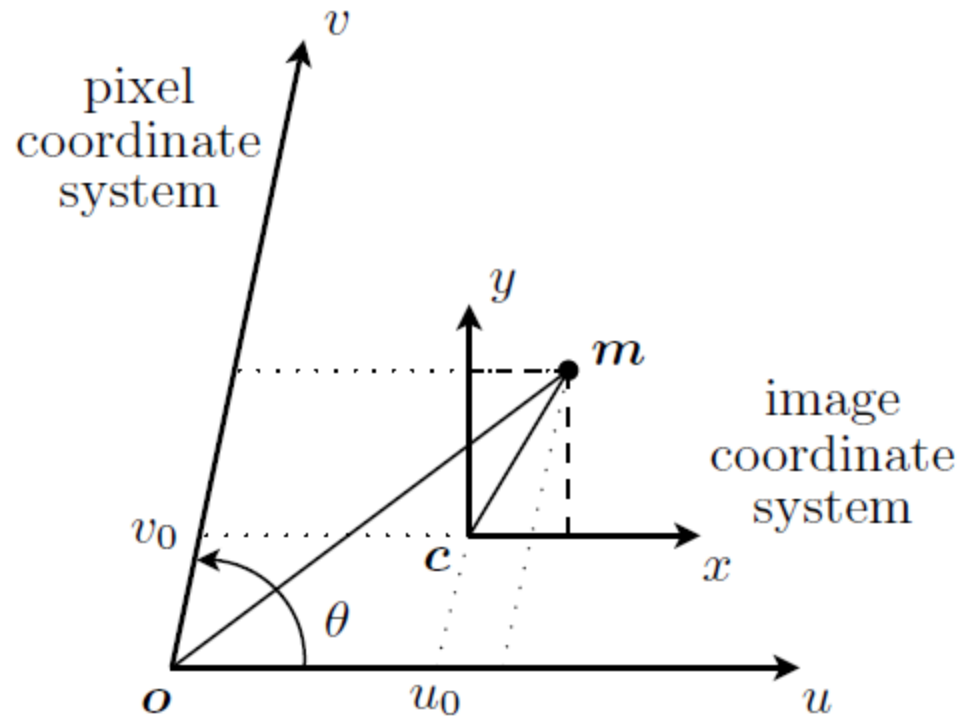
Intrinsic Camera Parameters

- Characterise the geometry of the image plane inside the camera
- Problems:
 - Origin of the image plane can be located in another point than the principal point, e.g. at the top left. Let the principal point in this coordinate system be located in $(u_0, v_0)^T$.
 - Pixels may have different dimensions h_u and h_v .
 - In the worst case, the coordinate axis may have an angle $\theta \neq 90$ degrees.
- These 5 intrinsic parameters lead to a matrix

$$H = \begin{pmatrix} h_u & -h_u \cot \theta & u_0 \\ 0 & h_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

that describes the **transition from the ideal image coordinates to the real (pixel) coordinates**.

Intrinsic Camera Parameters



The intrinsic camera parameters describe the transition from the ideal image coordinates $(x, y)^\top$ to the real (pixel) coordinates $(u, v)^\top$. Author: M. Mainberger (2010).

Camera Projection Matrix P

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \underbrace{\begin{pmatrix} h_u & -h_u \cot \theta & u_0 \\ 0 & h_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{intrinsic matrix}} \underbrace{\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{projection matrix}} \underbrace{\begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{extrinsic matrix}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

$$=: \underbrace{\begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \end{pmatrix}}_{\text{full projection matrix}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

- Concatenating the extrinsic, projection and intrinsic matrices gives the full projective mapping.
- It maps a 3D point in homogeneous world coordinates $(X_w, Y_w, Z_w, 1)^T$ to a 2D image point with homogeneous pixel coordinates $(u, v, w)^T$
- 12 parameters in total but with 1 free scaling parameter. So 11 degrees of freedom: 6 extrinsic plus 5 intrinsic.

Anatomy of P

- The 3x4 camera matrix P encodes very rich geometric information.
- **The advantage of linear algebra is that we handle all of this geometric information through algebra (manipulation of symbols).**

Anatomy of P

Camera centre

- Let M = first 3×3 matrix of the 3×4 matrix P
- When M is non-singular, P has rank 3 and therefore a null-space of dimensionality 1.
- Therefore there exists a vector v such that $Pv=0$
- Vector v must be the camera centre C .

Anatomy of P

Camera centre

- Consider the set of points along the line joining some point A and the camera centre C

$$X(\lambda) = (1 - \lambda)C + \lambda A \quad \leftarrow \text{Join of A and C}$$

- All such points will map to the same image point PA

$$\begin{aligned} \lambda PA &= PX(\lambda) \\ &= (1 - \lambda)PC + \lambda PA \\ \Rightarrow PC &= 0 \end{aligned}$$

- In Matlab, $C = \text{null}(P)$

Anatomy of P

Camera centre

- Camera can image every point in 3-D but it's own centre! **Why?**
- If $\text{Rank}(M) = 2$, then C will be a point at infinity, i.e. the last coordinate of C will be zero!
 - This is called the camera at infinity model.

Vanishing Point

- **Vanishing Point:** Point where parallel lines meet in the image.
- In the real world, parallel lines meet at infinity.
- So a vanishing point is the image of infinity!



- **Why do parallel lines meet in a projected image?**
(Hint: use the projection equations $x=fX/Z$ and $y=fY/Z$)

Anatomy of P

Notation

- Let p_i be the i^{th} column of P .
- Let P^{iT} be the i^{th} row of P .

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}$$
$$p_i = \begin{bmatrix} p_{1i} \\ p_{2i} \\ p_{3i} \end{bmatrix} \rightarrow P = [p_1 \quad p_2 \quad p_3 \quad p_4]$$
$$P^i = \begin{bmatrix} p_{i1} \\ p_{i2} \\ p_{i3} \\ p_{i4} \end{bmatrix} \rightarrow P = \begin{bmatrix} P^{1T} \\ P^{2T} \\ P^{3T} \end{bmatrix}$$

Points at Infinity

- In homogenous coordinates we can express points at infinity.
 - In P^2 $[a,b,0]$ is a point at infinity in the direction of the 2D vector $[a,b]$. (**Why?**)
 - In P^3 $[a,b,c,0]$ is a point at infinity in the direction of the 3D vector $[a,b,c]$.
- **Setting the last coordinate to 0 in homogenous coordinates, yields a point at infinity in Euclidean space.**
- Every direction is represented as a point at infinity in that direction.
- **Write down the representation of the x-axis in P^3 .**

Anatomy of P

Columns of P

- $P [1 \ 0 \ 0 \ 0]^T = p_1$ (first column of P)
- But $[1 \ 0 \ 0 \ 0]^T$ is the direction of the x-axis.
- So p_1 is the image of the point at infinity in the direction of the x-axis.
 - Also called the **vanishing point in the x-direction**.
- p_2 is the image of ...?
- Which point at infinity maps to p_3 ?
- p_4 is the projection of ...?

Anatomy of P

- Column p_1 is the vanishing point in the x-direction.
- Column p_2 is the vanishing point in the y-direction.
- Column p_3 is the vanishing point in the z-direction.
- Column p_4 is the image of the world origin.

Equation of a Plane

- Non-homogenous: $aX+bY+cZ+d=0$
- Homogenous:

$$\underbrace{\begin{bmatrix} a & b & c & d \end{bmatrix}}_{n^T} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0$$

$$n^T \vec{X} = 0$$

Anatomy of P

Rows of P

- $P = [P^{1T}; P^{2T}; P^{3T}]$ where each P^{iT} is a plane in P^3 .
- All points in plane P^{3T} satisfy $P^{3T} X = 0$
 - In other words, their images are of the form $(x, y, 0)^T$
- Therefore, **P^{3T} is the principle plane.**

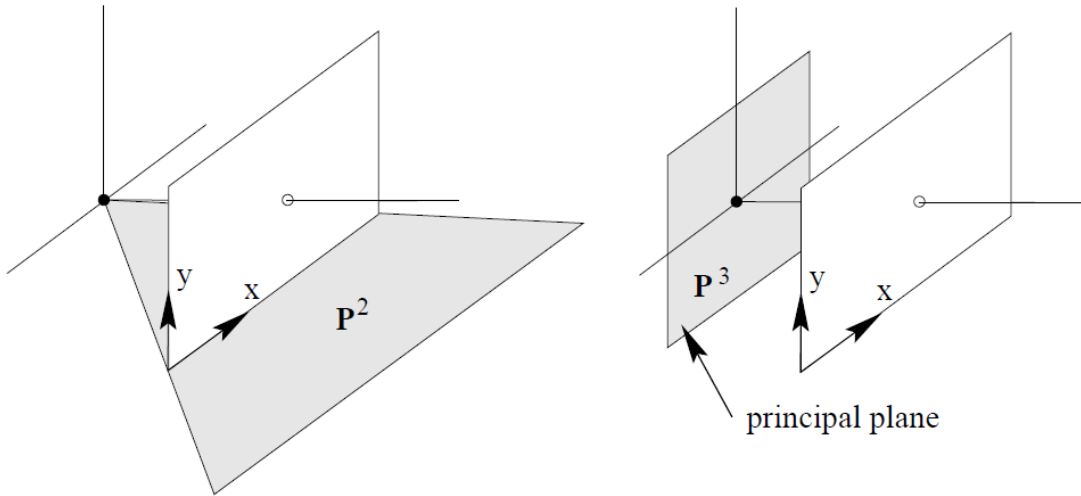
Anatomy of P

Rows of P

- All points in the plane P^1 satisfy $P^{1T}X=0$
- In other words, their images are of the form $(0,y,w)^T$ which are points on the image y -axis.
- Since $PC=0$, $P^{1T}C=0$ also. So, C also lies on the plane P^{1T} .
- Therefore, **P^{1T} is the plane defined by the camera centre C and the line $x=0$ in the image.**

Anatomy of P

- Row P^{1T} is the plane defined by the camera centre C and the image y-axis.
- Row P^{2T} is the plane defined by the camera centre C and the image x-axis.
- Row P^{3T} is the principle plane.



**H.W. Prove that $PC=0$
using these 3 points.**

Anatomy of P

Principal Axis Vector

- Normal to plane $[a;b;c;d]$ is the vector $[a;b;c]$.
- Principal axis vector is the normal vector of the principal plane P^3 .
 - Therefore, it is given by $\mathbf{m}^3 = [p_{31} \ p_{32} \ p_{33}]^T$ which is the 3rd row of M where M is the left 3x3 matrix of P.
- But since P is defined only upto scale \mathbf{m}^3 can point in the –ve Z direction too. (Why?)
- The principal axis vector pointing to the front of the camera is given by $\det(M)\mathbf{m}^3$.

Anatomy of P

Principal Point

- Since a vector is a direction, it can be represented as $[a;b;c;0]$ which is a point at infinity in direction $[a;b;c]$.
- Principal axis vector $m^3 = [p_{31} \ p_{32} \ p_{33}]^T$ can be represented as a point at infinity $P^3_\infty = [p_{31} \ p_{32} \ p_{33} \ 0]^T$.
- Principal point x_0 is the projection of P^3_∞ .
 - $x_0 = PP^3_\infty = Mm^3$.

Camera Geometry

(MVG Ch - 6)

All Images Courtesy: Hartley & Zisserman

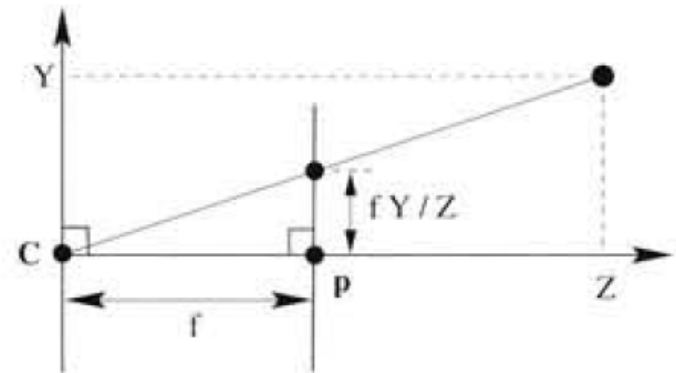
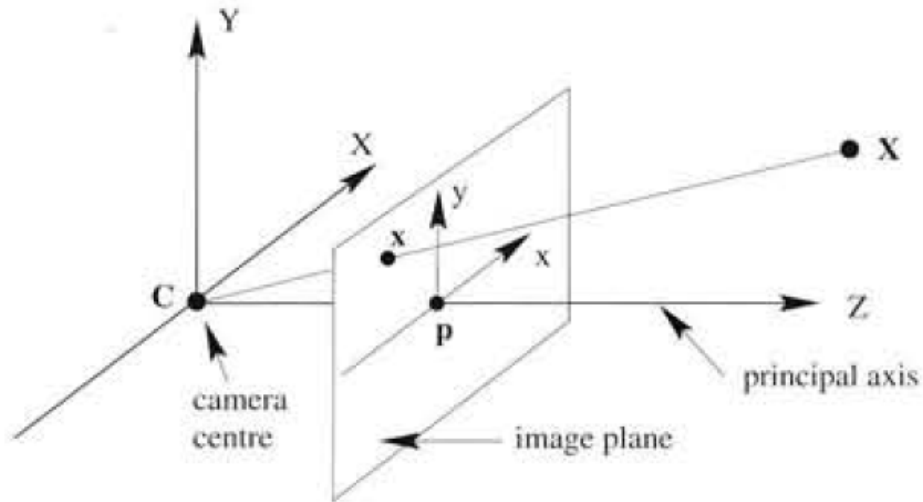


Image point: Intersection of 3-D line(joining C and word point) and the image plane

Let a point in space be :

$$P_i = [X_i \quad Y_i \quad Z_i]^T$$

The Line joining the point and Camera Center C is :

$$l(\lambda) = (1 - \lambda)C + \lambda P_i$$

$$C = [0 \quad 0 \quad 0]^T$$

Parametric
eq. of line

$$\begin{aligned} X &= 0 + \lambda X_i \\ Y &= 0 + \lambda Y_i \\ Z &= 0 + \lambda Z_i \end{aligned}$$

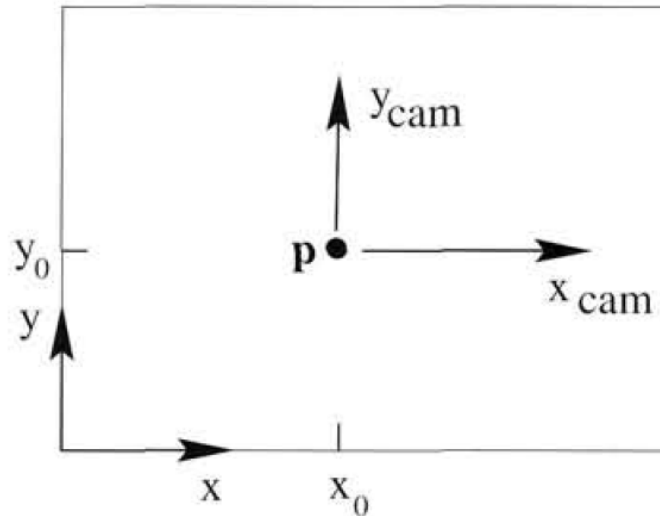
Image Plane $Z = f$

$$\Rightarrow \lambda = \frac{f}{Z_i}$$

$$\begin{aligned} P_i &\mapsto \begin{bmatrix} \frac{fX_i}{Z_i} & \frac{fY_i}{Z_i} & f \end{bmatrix}^T \\ x_i &= \begin{bmatrix} \frac{fX_i}{Z_i} & \frac{fY_i}{Z_i} \end{bmatrix}^T \end{aligned}$$

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

Principal Point Offset



What if the origin of the image plane is not principal point? The model assumes it! The generalized formulation for an arbitrary choice of image origin is given below:

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

$$\mathbf{x} = \mathbf{K}[\mathbf{I} \mid \mathbf{0}]\mathbf{X}_{cam} \quad \mathbf{K} = \begin{bmatrix} f & p_x \\ f & p_y \\ 1 \end{bmatrix}$$

Pixel Coordinates

Let the number of pixels per unit distance in image coordinates are

m_x and m_y in x and y directions respectively,

then the world coordinates to pixel coordinates mapping is given below

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} m_x \left(\frac{fX}{Z} + p_x \right) \\ m_y \left(\frac{fY}{Z} + p_y \right) \end{bmatrix}$$

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

where $\alpha_x = fm_x$ and $\alpha_y = fm_y$

$$x_0 = m_x p_x \quad y_0 = m_y p_y$$

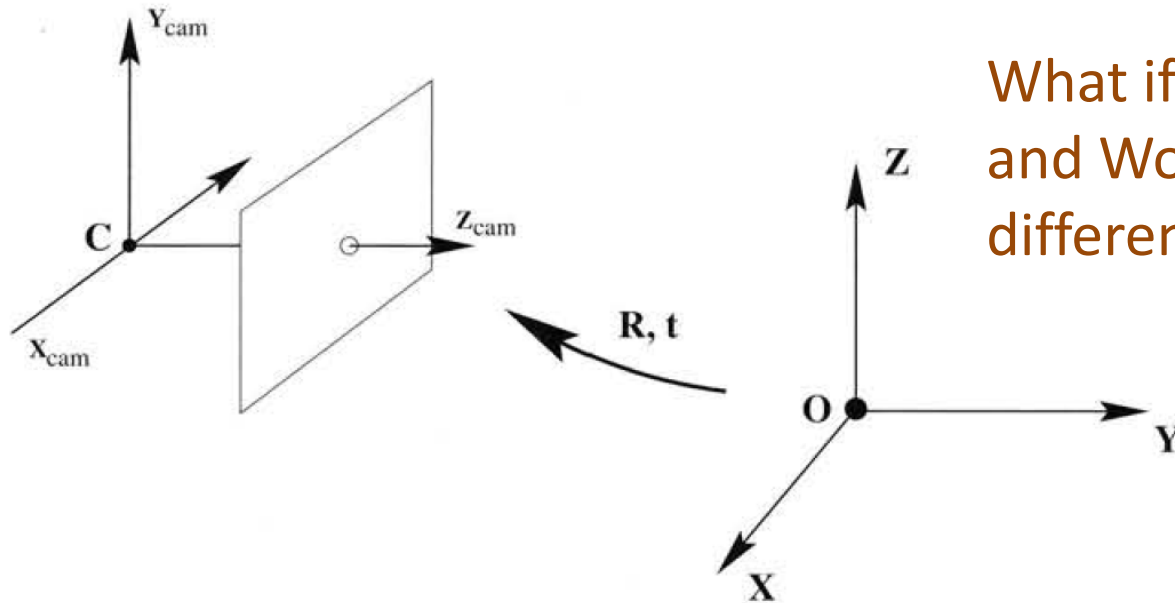
$$K = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}$$

Camera Calibration

Matrix

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

Camera Orientation and Position



What if Camera frame and World frame are different?

$$\mathbf{x}_{\text{cam}} = \begin{bmatrix} \mathbf{R} & -\mathbf{R}\tilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{R} & -\mathbf{R}\tilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \mathbf{x} \quad \mathbf{x} = \mathbf{K}\mathbf{R}[\mathbf{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}$$

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

$$\mathbf{t} = -\mathbf{R}\tilde{\mathbf{C}}$$

$$x = PX$$

Projective Camera

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \quad P = [M \mid \mathbf{p}_4]$$

- P is Projective Camera if $\text{Rank}(P) = 3$
- If $\text{Rank}(M) = 3$ then camera is finite, i.e. the camera center is not at infinity
- If $\text{Rank}(M) = 2$ then the camera centre is at infinity, (Camera at infinity)

Camera Centre

Consider the set of points along the Line joining some point A and the Camera Center C :

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

All such points will map to the same image point PA

$$\lambda PA = PX(\lambda)$$

$$PX(\lambda) = (1 - \lambda)PC + \lambda PA$$

$$\Rightarrow (1 - \lambda)PC + \lambda PA = \lambda PA$$

$$\Rightarrow PC = 0$$

$$P = [M \mid p_4]$$

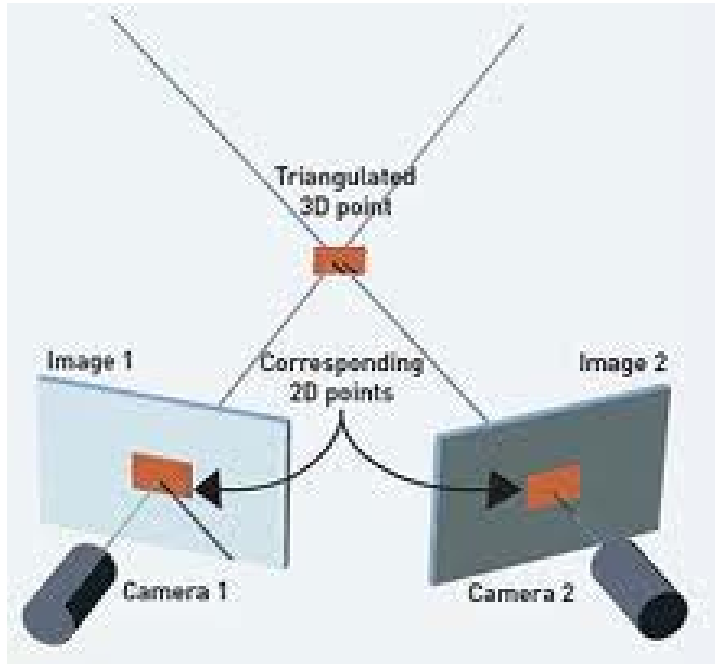
$$C = \text{null}(P)$$

Camera can image every point in 3-D but it's own centre! Why?

So with respect to imaging, the camera centre is a unique point in space

If $\text{Rank}(M) = 2$, then C will be a point at infinity, i.e. the last coordinate of C will be zero! (Camera at infinity!)

3-D Reconstruction: Triangulation



The intersection of both the 3-D lines is the 3-D point X

What if camera matrix is not known?

Assumptions

1. Camera Matrices are known
2. Image correspondences are known

$$P^+ = P^T (PP^T)^{-1}$$

$$PP^+ = I$$

$$x_1 = P_1 X \quad A_1 = P_1^+ x_1$$

$$x_2 = P_2 X \quad A_2 = P_2^+ x_2$$

$$X_1(\lambda) = (1 - \lambda)A_1 + \lambda C_1$$

$$X_2(\lambda) = (1 - \lambda)A_2 + \lambda C_2$$

Estimating Camera Matrix and Calibration Matrix

DLT

$$x_i = PX_i$$

$$x_i \times PX_i = 0$$

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

$$Ax = 0$$

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

$$P = [M \mid \mathbf{p}_4]$$

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

$$M = KR$$

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

$$[K \ R] = RQ(M)$$

Cameras at Infinity

Orthographic Camera

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

Scaled Orthographic Camera

$$P_s = \begin{bmatrix} k & & \\ & k & \\ & & 1 \end{bmatrix} P_o$$

Weak Perspective Camera

$$P_w = \begin{bmatrix} \alpha_x & & \\ & \alpha_y & \\ & & 1 \end{bmatrix} P_o$$

Affine Camera

$$P_a = \begin{bmatrix} \alpha_x & s & \\ & \alpha_y & \\ & & 1 \end{bmatrix} P_o$$

$$P_a = \begin{bmatrix} m_{11} & m_{12} & m_{13} & t_1 \\ m_{21} & m_{22} & m_{23} & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

What is the form of Camera at infinity which is not Affine?