Workshop on Computer Vision, Graphics and Image processing

Projective Geometry and Camera Models

Dr. Uma Mudenagudi

Professor,

Department of Electronics and Communication,

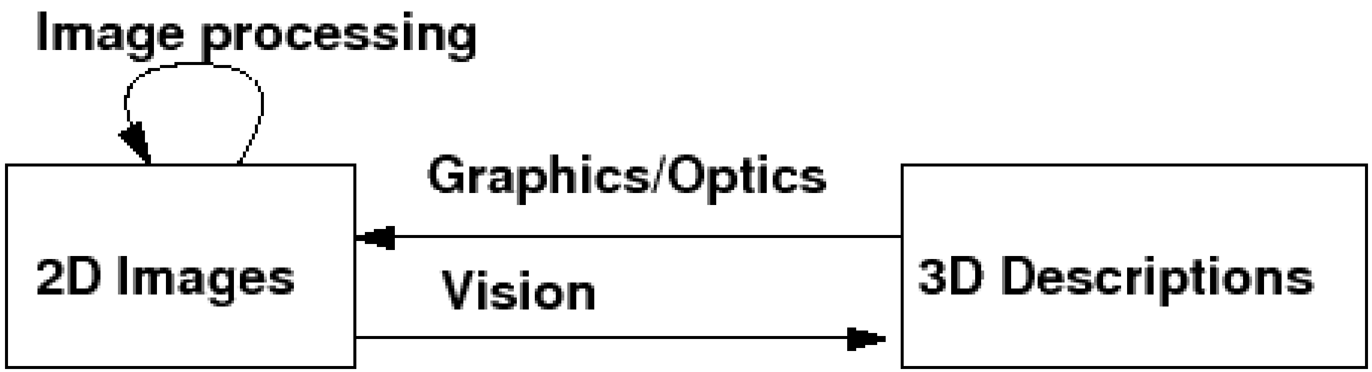
BVB College of Engineering and Technology, Hubli

Class 14,15, 16 04/05.07.2011 10.00-1.15pm

Overview

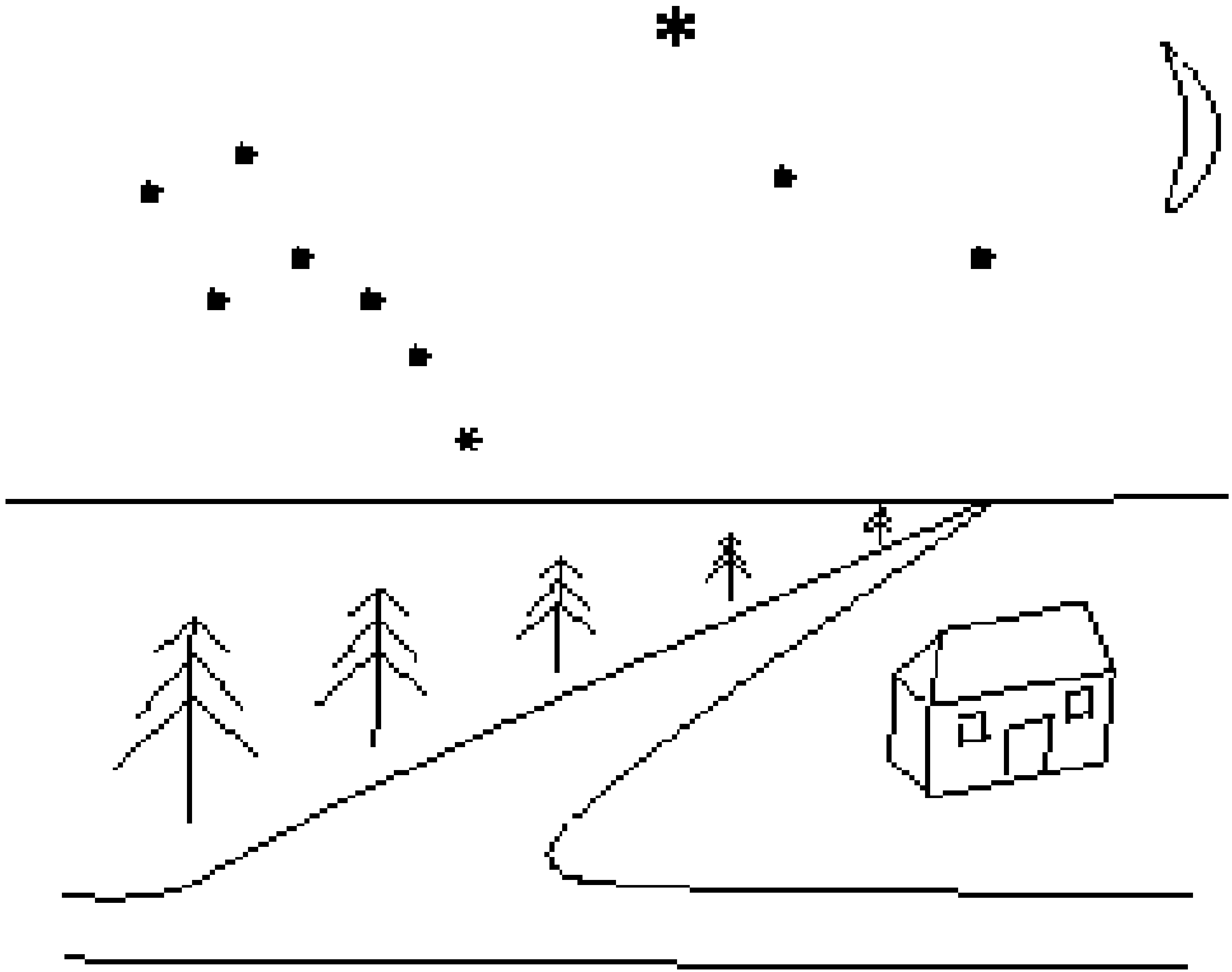
- Introduction to projective geometry
- 2D and 3D transformations
- Different camera models
- Conclusions

Projective geometry



- Correspondence problem: Match image projections of a 3D configuration.
- Reconstruction problem: Recover the structure of the 3D configuration from image projections.
- Re-projection problem: Is a novel view of a 3D configuration consistent with other views? (Novel view generation) All of these require camera calibration in some form.

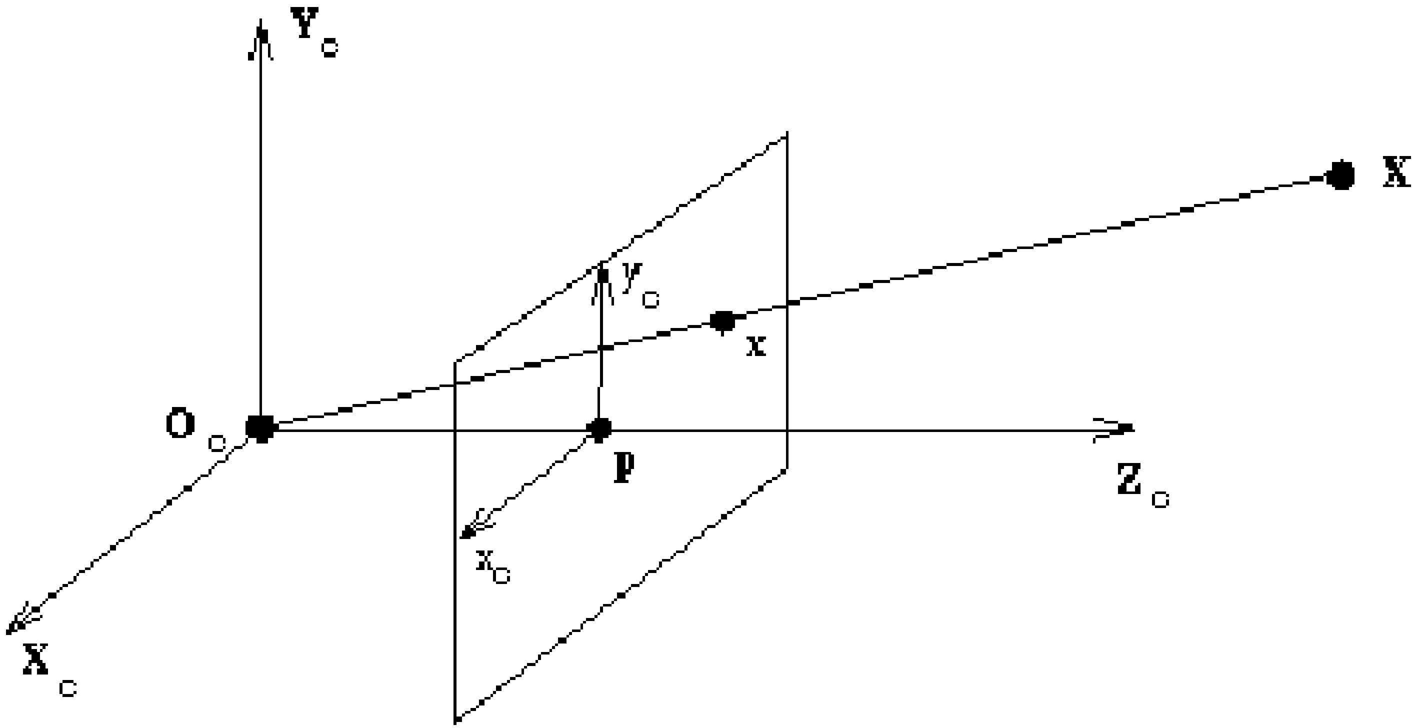
Projective geometry



Projective geometry

- Parallel lines in 3D space converge in images.
- The line of the horizon is formed by ‘infinitely’ distant points (vanishing points).
- Any pair of parallel lines meet at a point on the horizon corresponding to their common direction.
- All ‘intersections at infinity’ stay constant as the observer moves.
- The effects can be modelled mathematically using the ‘linear perspective’ or a ‘pin-hole camera’ (realized first by Leonardo?)

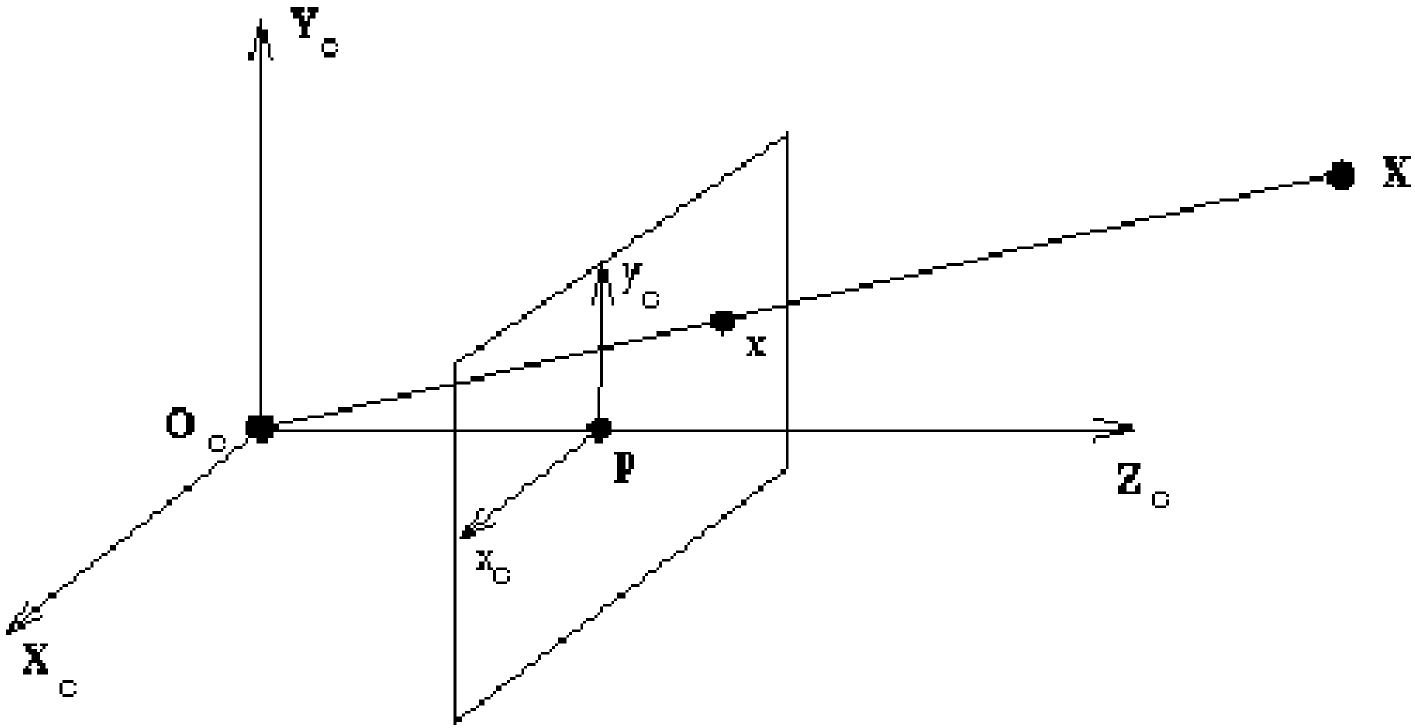
Pinhole model of the camera



- world co-ord (X,Y,X) the image coordinates are (x,y) then

$$x = \frac{fX}{Z} and y = \frac{fY}{Z}$$

Pinhole model of the camera



- world co-ord (X,Y,X) the image coordinates are (x,y) then

$$x = \frac{fX}{Z} and y = \frac{fY}{Z}$$

2D Transformations

- Euclidean: 3 degrees of freedom
- Invariant: length and area

$$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

2D Transformations contd..

- Similarity: 4 degrees of freedom
- Invariant: ratio of lengths and angle

$$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

2D Transformations contd..

- Affine: 6 degrees of freedom
- Invariant: parallelism, ratio of areas, ratio of lengths on collinear or parallel lines (midpoints), line at ∞

$$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$$

2D Transformations contd..

- Projective: 8 degrees of freedom
- Invariant: Concurrency, collinearity order of contact (intersection), parallelism, cross ratio (ratio of ratio of lengths)

$$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

3D Transformations

- Euclidean: 6 degrees of freedom
- Invariant: volume

$$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$$

3D Transformations contd..

- Similarity: 7 degrees of freedom
- Invariant: The absolute conic

$$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$$

3D Transformations contd..

- Affine: 12 degrees of freedom
- Invariant: parallelism of planes, volume ratios, centroids, the plane at ∞

$$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$$

3D Transformations contd..

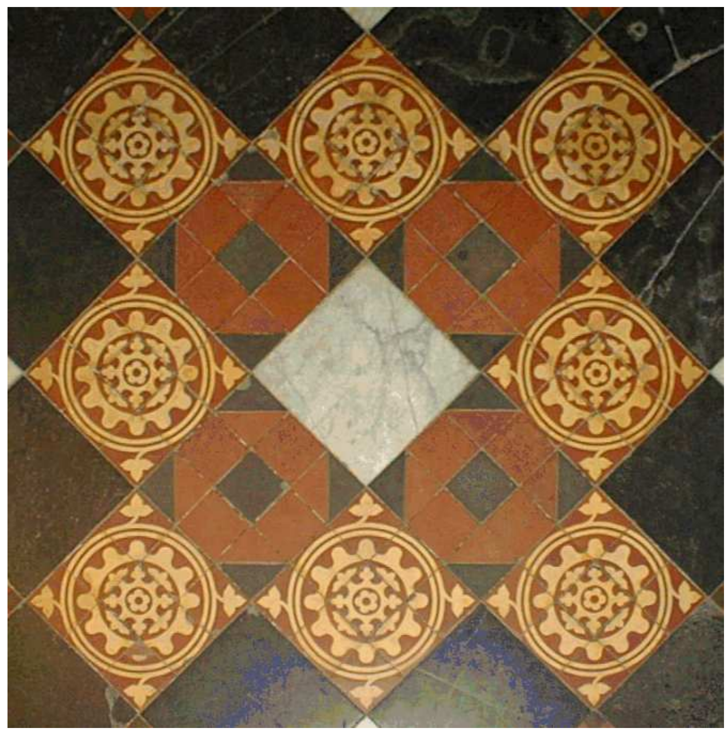
- Projective: 15 degrees of freedom
- Invariant: Intersection and tangency of surfaces in contact, sign of Gaussian curvature

$$\begin{bmatrix} A & t \\ V^T & 1 \end{bmatrix}$$

Similarity transform

Under similarity transform the circular pattern is imaged as circle

and a square tile is imaged as a square



Affine transform

Under affine transform the circle is imaged as an ellipse and

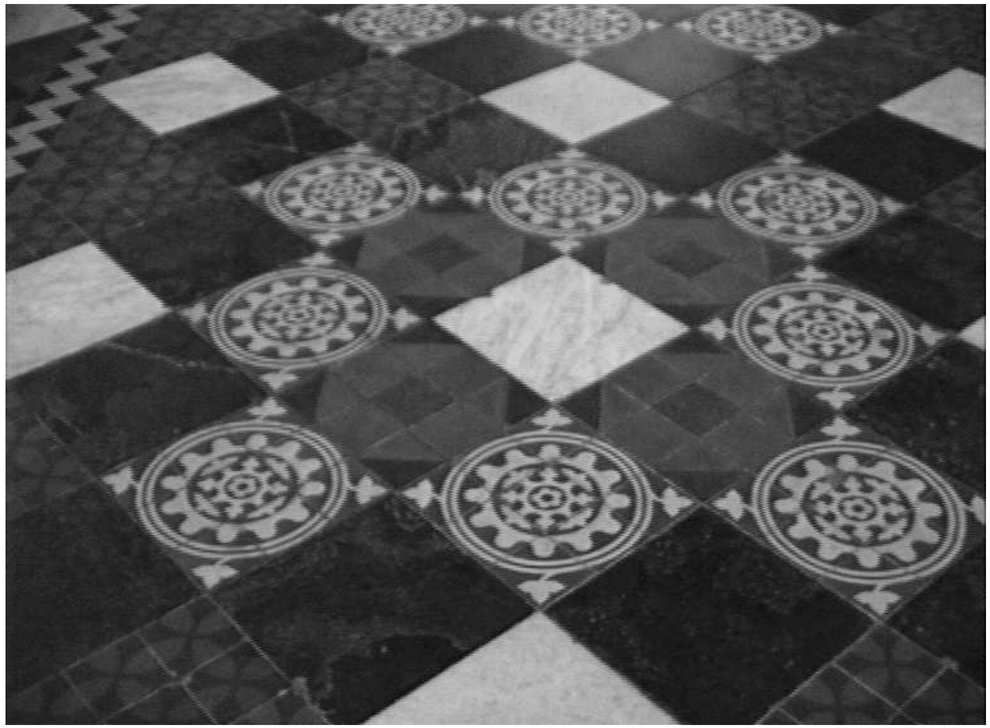
orthogonal lines are not imaged as a orthogonal lines

Projective transform

Under projective transform, parallel lines are imaged as converging lines. Tiles closer to the camera have larger image than those far away

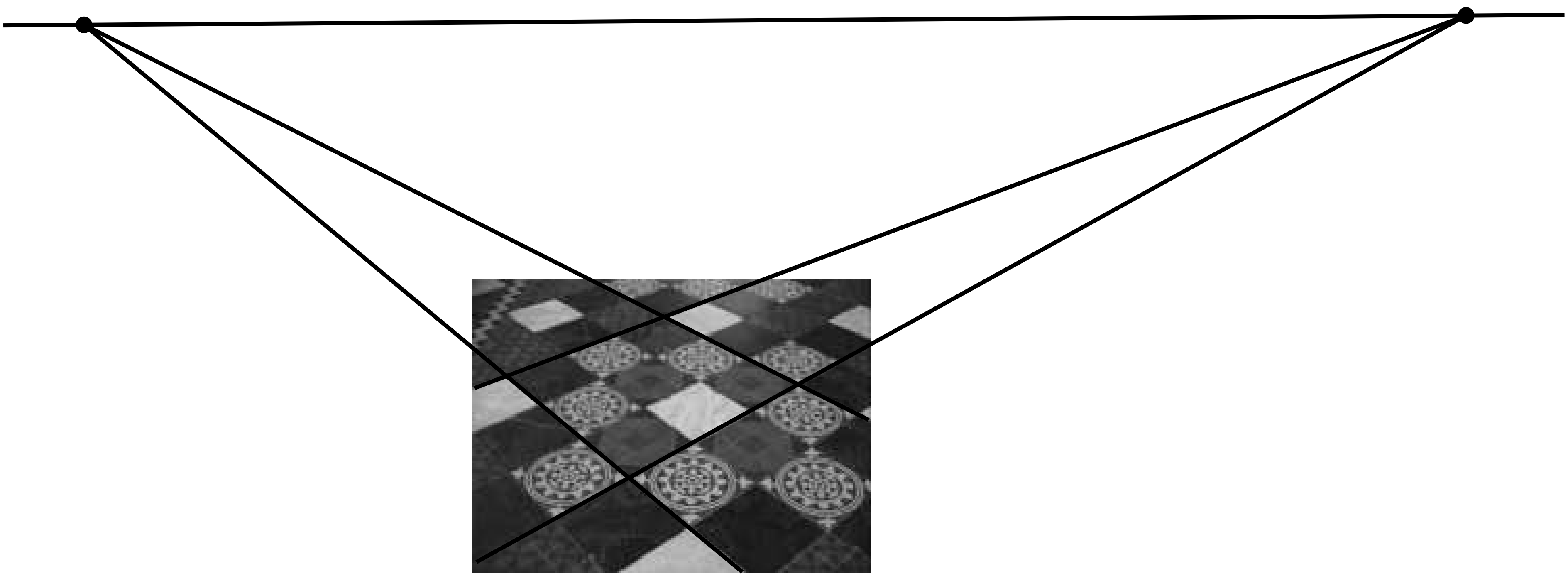
Affine rectification via vanishing line

The vanishing line of the plane imaged is computed



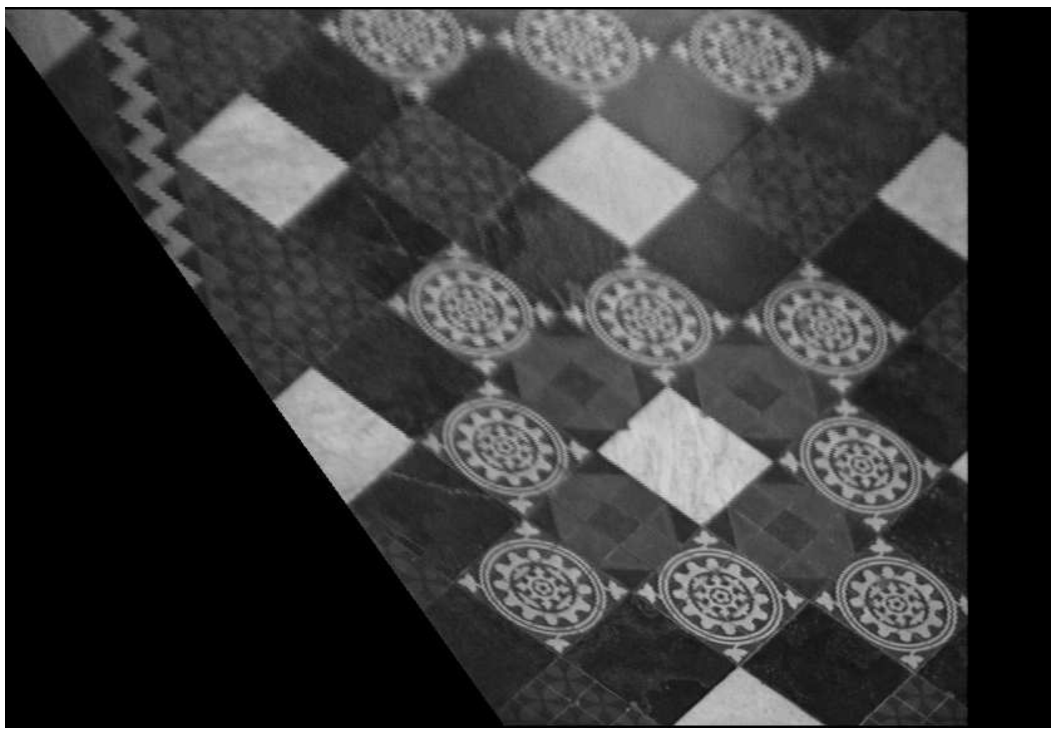
Affine rectification via vanishing line

From the intersection of two sets of imaged parallel lines



Affine rectification via vanishing line

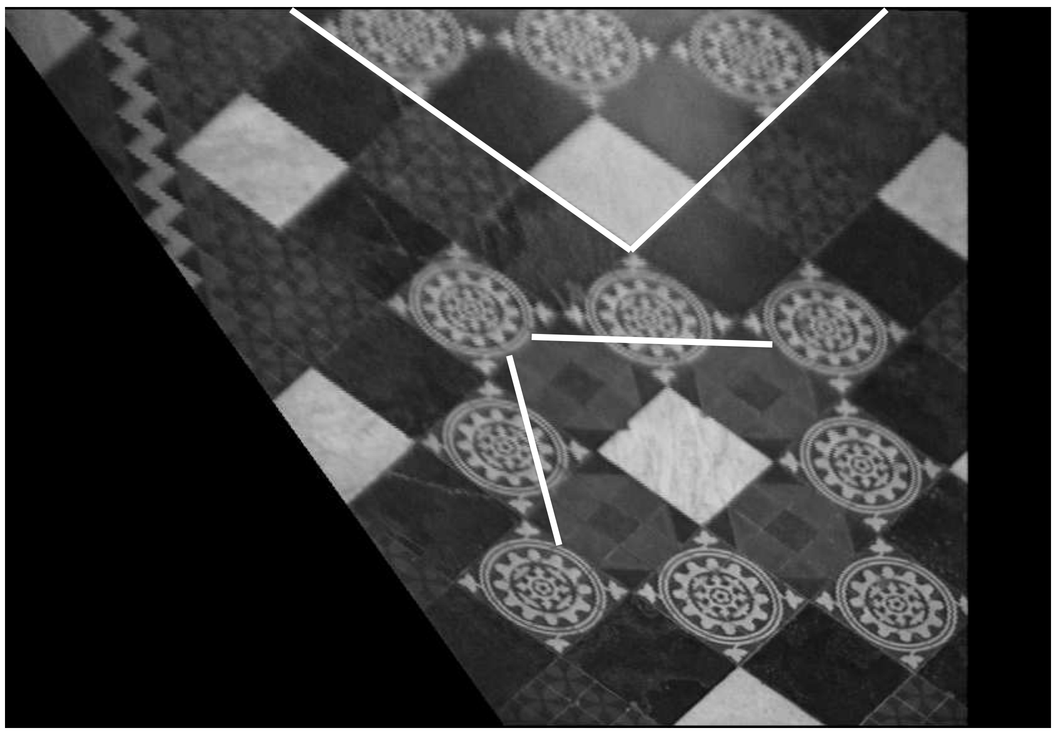
The image is then projectively warped to produce the affine rectified image



Metric rectification

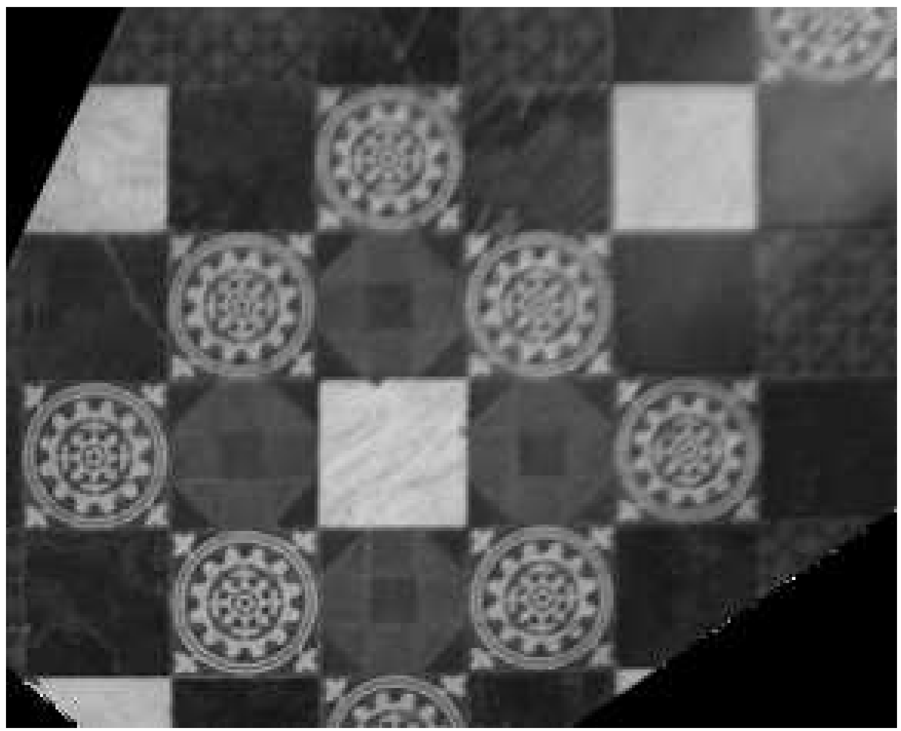
Two non parallel lines indentified on the affine rectified image

correspond to orthogonal lines on the world plane



Metric rectification contd..

In the metric rectified image all lines orthogonal in the world are orthogonal, squares have unit aspect ration and circles are circular



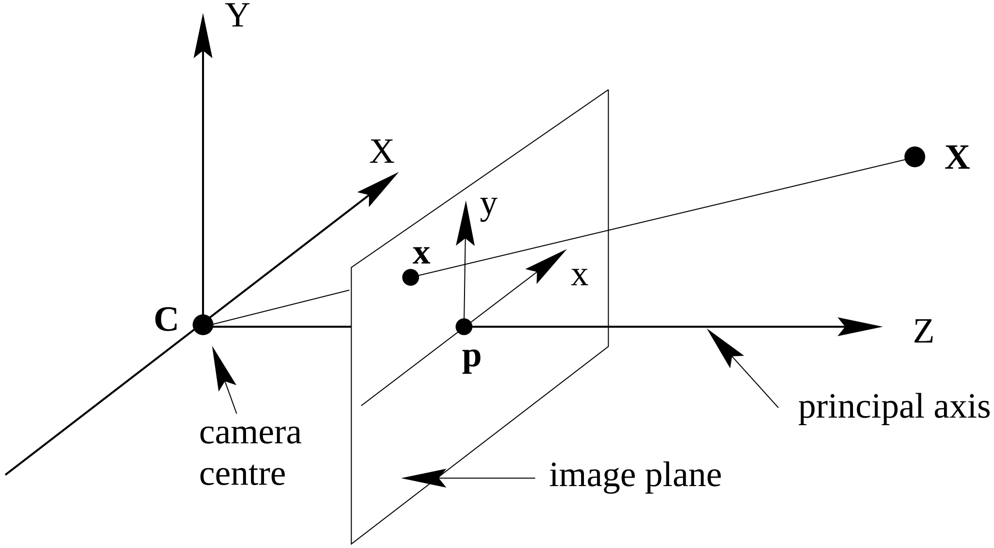
Different camera models

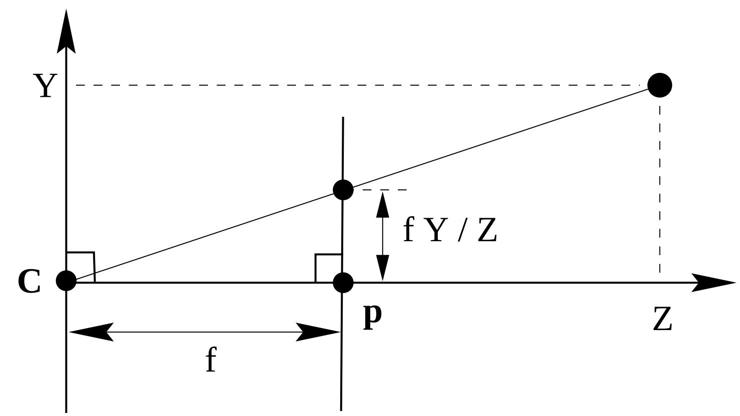
- Projective camera: mapping from the 3D world to 2D image
- Two major class:
  - with finite centre: finite cameras
    - pinhole model
    - finite projective camera
  - with centre at infinity
    - affine camera: natural generalization of parallel projection

Pinhole model of camera

- Let centre of projection is the origin of a Euclidean coordinate system
- $z = f$, image plane or focal plane
- A point in 3D space $\mathbf{X}$ is mapped to a point on the image plane where a line joining the point $\mathbf{X}$ to the centre of projection meets the image plane

Pinhole model of camera





- By similar triangles

$$\mathbf{X} = (X, Y, Z)^T \mapsto \mathbf{x} = (fX/Z, fY/Z)^T$$

Pinhole model of camera contd..

- Camera centre: centre of projection (optical centre)
- Principal axis (ray): the line from the camera centre perpendicular to the image plane
- Principal point: the point where the Principal axis meets the image plane

Central (perspective) projection

- Using homogeneous coordinates

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

Central projection contd..

- The matrix is $diag(f, f, 1)[I|0]$, where $diag(f, f, 1)$ is a diagonal matrix and $[I|0]$ represents a matrix divided up to $3\times$ block (identity) plus a column vector
- Let $\mathbf{X} = (X, Y, Z, 1)^T$ be the world point and $\mathbf{x}$ be the image point (3-vector) in homogeneous coordinate systems

Central projection contd..

- Let P be a 3×4 homogeneous camera projection matrix, then

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

- Which defines camera matrix for pinhole model as

$$P = diag(f, f, 1)[I|0]$$

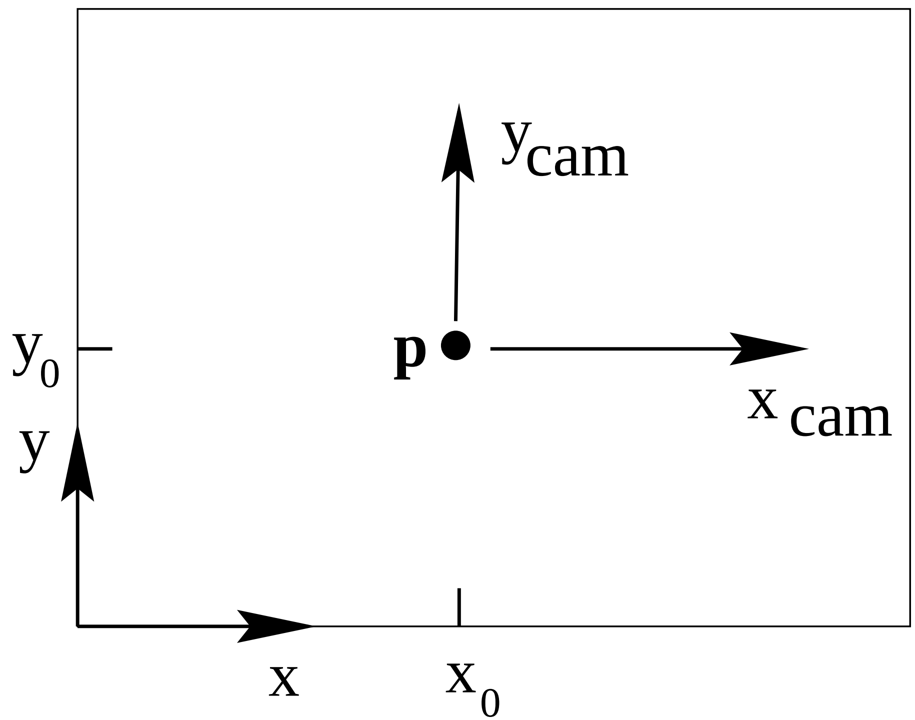
Principal point offset

- The mapping assumed that the origin of coordinates in the image plane is at the principal point
- In practice, it may not be, so the mapping is

$$(X, Y, Z)^T \mapsto (fX/Z + p_x, fY/Z + p_y)^T$$

Principal point offset contd..

where $(p_x, p_y)^T$ are the coordinates of the principal point



Principal point offset contd..

- In homogeneous coordinates

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

Principal point offset contd..

- We can write

$$x = K[I \mid 0]X_{cam}$$

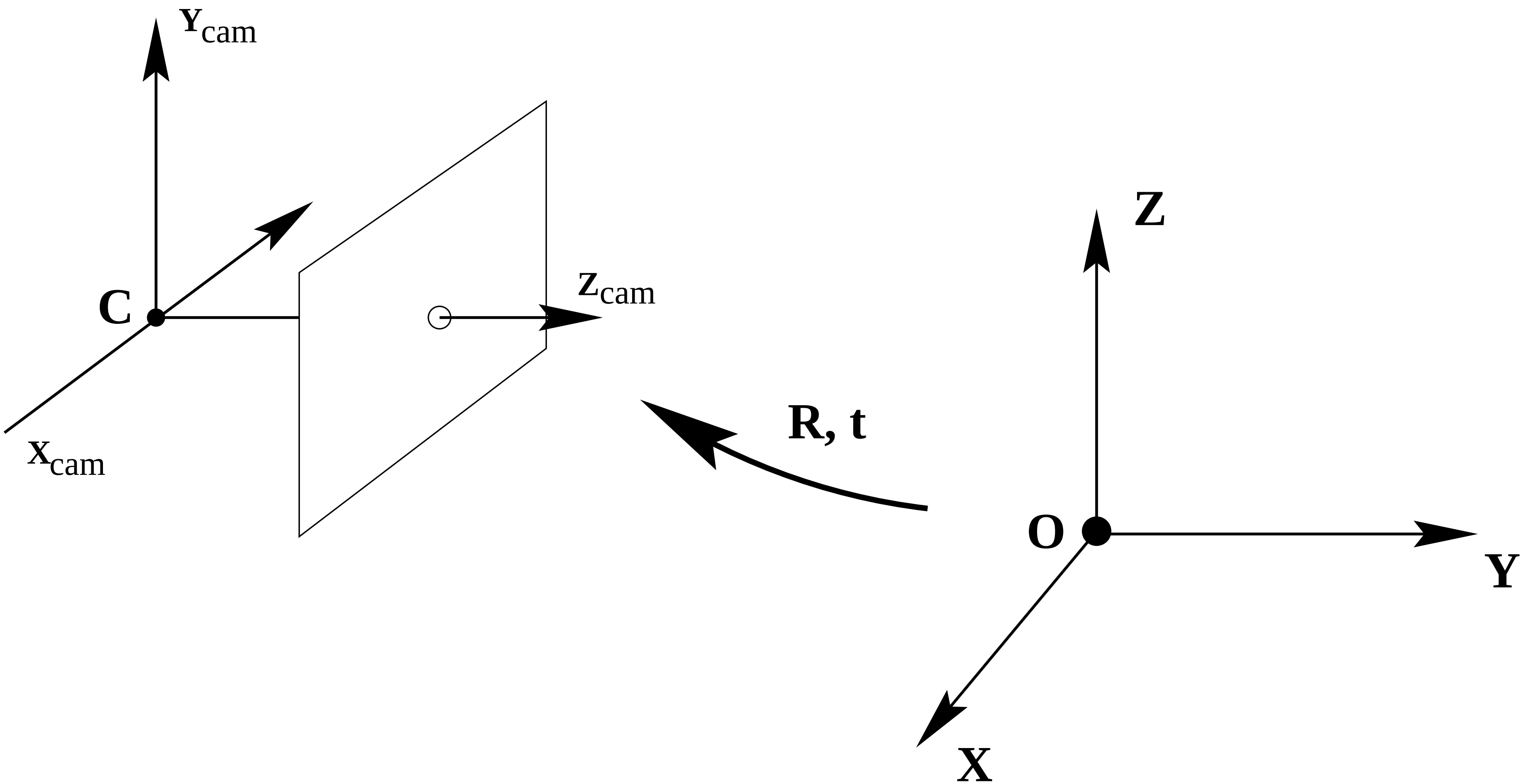
where k is camera calibration matrix

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

Principal point offset contd..

- $X_{cam} = (X, Y, Z, 1)^T$, ie. the camera is assumed to be located at the origin of a Euclidean coordinate system with the principal axis of the camera pointing down the z axis, which is called camera coordinate system

Camera rotation and translation



Camera rotation and translation contd..

- $\tilde{X}$: co-ord of point in world co-ord frame.
- $\tilde{X}_{cam}$: represents the same point in the camera frame, then

$$\tilde{X}_{cam} = R(\tilde{X} - \tilde{C})$$

where $\tilde{C}$ represents the co-ord of the camera centre in the world co-ord frame and R is the 3×3 rotation matrix

Camera rotation and translation contd..

- In homogeneous coordinates

$$\mathbf{X}_{cam} = \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix} \mathbf{X}$$

- We get $\mathbf{x} = KR[I \mid -\tilde{C}]\mathbf{X}$

Camera rotation and translation contd..

- Now $\mathbf{X}$ is in the world co-ord frame
- General pinhole camera $P = KR[I \mid -\tilde{C}]$
  nine degrees of freedom
- Three for k: f, p_x, p_y
- Three for R
- Three for $\tilde{C}$

Camera rotation and translation contd..

- Internal camera parameters: the parameters of k
- External camera parameters: the parameters of R and $\tilde{C}$, which relate the camera orientation and position
- If $\tilde{X}_{cam} = R\tilde{X} + t$ then $P = K[R|t]$

Finite projective camera

- For generality consider calibration matrix K of the form

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

- The added parameter s is referred as skew parameter

Finite projective camera contd..

- The skew parameter is zero for most of the cameras
- A camera $P = KR[I \mid -\tilde{C}]$ for which K takes this form is finite projective camera
- A finite projective camera has 11 degrees of freedom, same as 3×4 matrix up to scale

General projective camera

- A general projective camera may be decomposed into blocks $\tilde{P} = [M|p_4]$, where M is 3×3 matrix and if M is non-singular then this is finite camera otherwise it is not.
- The projective camera $\tilde{P}$ has a rank 3 whereas it has 4 columns. Clearly, it has a one dimensional null (right) space.

The optical center: (camera center)

- Suppose the null space is generated by the 4-vector C, that is

$$\tilde{P}C = 0,$$

then $\tilde{C}$ is the optical center of the camera $\tilde{P}$.

- Consider line containing C and any other point A in 3-space, points on this line can be expressed as

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

- Under the mapping $x = \tilde{P}\mathbf{X}$, points on this line are projected to

$$\mathbf{x} = \tilde{P}\mathbf{X}(\lambda) = \lambda\tilde{P}A + (1-\lambda)\tilde{P}C = \lambda\tilde{P}A, \qquad (\because \tilde{P}C = 0)$$

General projective camera contd..

- Since every point on the line are mapped on to the same image point, the line must be a ray through the camera center. It follows that C is the camera center because for all choices of A the line passes through the optical center.
- Writing

$$\tilde{P} = [P_1 | -P_1 t]$$

where P_1 is a 3×3 non singular and t is the optical center

$$[P_1 | -P_1 t][t \;\; 1]' = 0$$

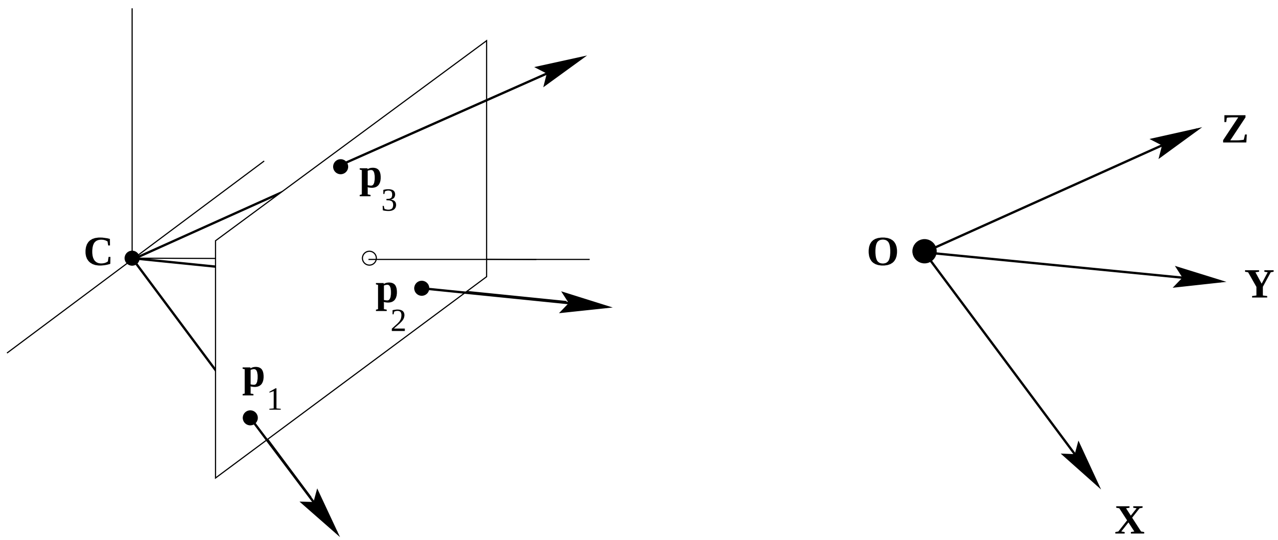
General projective camera contd..

- An image point $\mathbf{x}$ defines a line $\lambda P^{-1}\mathbf{x} + t$ in 3-space

$$\mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \lambda \begin{bmatrix} P^{-1} & \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \\ 1 \end{bmatrix}$$

- Clearly all points $\mathbf{X}$ project to $\mathbf{x}$ under $\mathbf{x} = \tilde{P}\mathbf{X}$.
- This result is not unexpected
- $(0,0,0)^T = PC$ is not defined
- The camera centre is the unique point in space for which the image undefined

The column vectors of the camera matrix



- Let the columns of a projective camera $\tilde{P}$ be p_i for $i = 1, \ldots, 4$. p_i's have geometric interpretations as special image points. p_1, p_2, p_3 are the vanishing points of the world coordinates axes X, Y and Z respectively. For example, the X-axis has direction $D = (1, 0, 0, 0)^T$, which is imaged as $p_1 = \tilde{P}D$.
- The column p_4 is the image of the world origin $(0, 0, 0, 1)^T$

The row vectors of the camera matrix

- The row vectors of the projective camera are 4-vectors which have the geometric interpretations as particular world planes.
- Let us write the rows of the projective camera as π_1^T, π_2^T and ${\pi_3}^T$. That is

$$\tilde{\mathbf{P}} = \begin{bmatrix} \pi_1^T \\ \pi_2^T \\ {\pi_3}^T \end{bmatrix}$$

- **The focal plane (principal plane)** is the plane parallel to the image plane containing the optical center. It is the plane of equation

$$\pi_3^T \mathbf{X} = 0$$

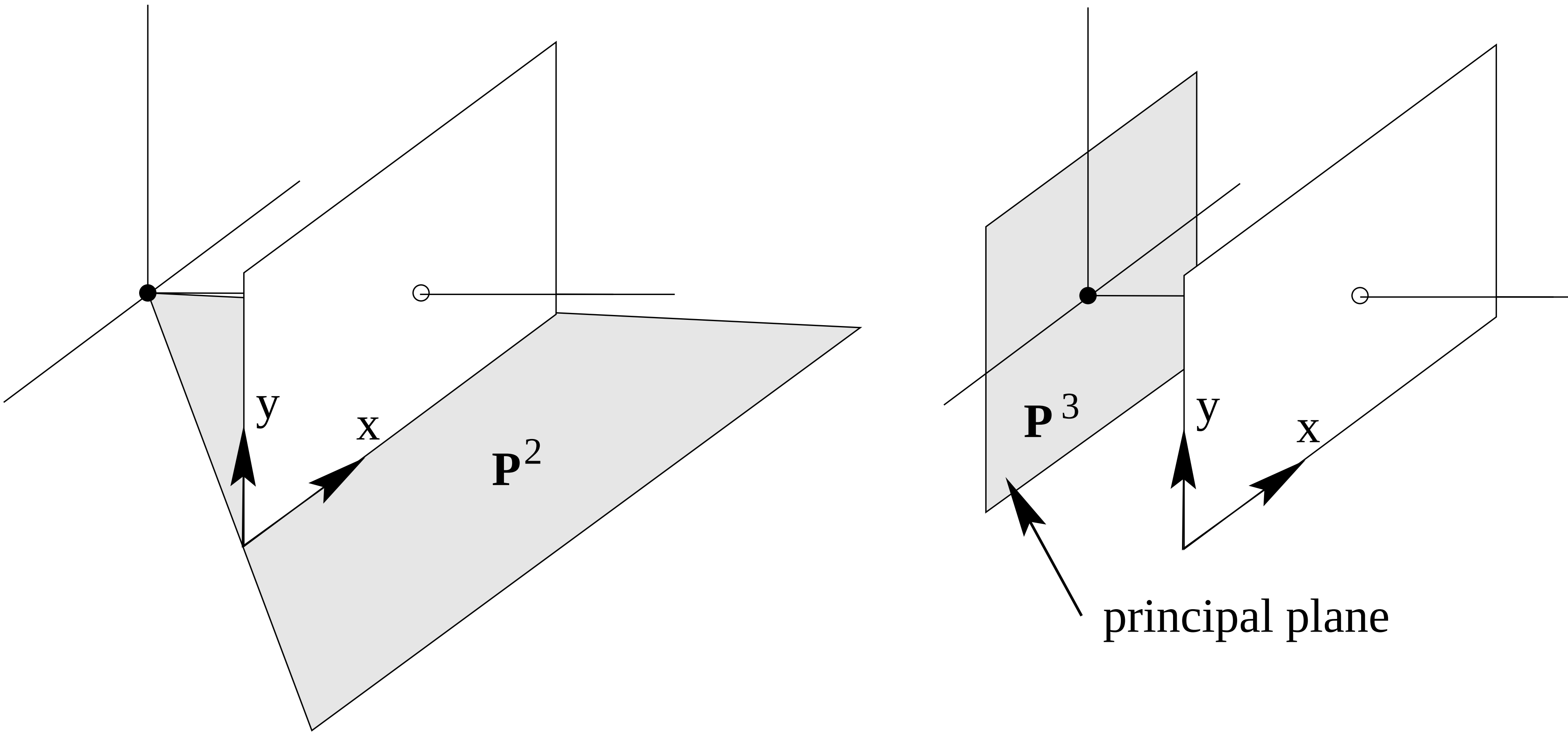
Points on this plane project on to image points $(x, y, 0)^T$, i.e., points at infinity on the image plane.

The row vectors of the camera matrix

The axes planes:

- Consider points on the plane π_1^T . This set satisfies $\pi_1^T \mathbf{X} = \mathbf{0}$, and, hence, points on this plane project on to image points, which are points on the image y-axis. It also follows from $\tilde{P}C = 0$ that C also lies on π_1^T . Hence π_1^T is the plane defined by the optical center and the y -axis in the image plane.
- Similarly, π_2^T is the plane defined by the optical center and the x-axis in the image plane.
- Thus unlike π_3^T , π_1^T and π_2^T are dependent on the choice of the coordinate system on the image plane. In particular, the intersection of these two planes is the line joining the optical center with the coordinate origin in the image plane. This line will not, in general, coincide with the principal axis

The row vectors of the camera matrix



Decomposition of the camera matrix

- Let P be a general projective camera matrix, aim is to find the camera centre C, orientation of the camera and the internal parameters of the camera from P
- Camera centre is the point for which $PC = 0$ (SVD of P)
- The matrix R gives the orientation of the camera and K is the camera calibration matrix
- The ambiguity in the decomposition is removed by putting condition that K have positive diagonal entries and is given by

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

- α_x and α_y are scale factors in x and y directions, s is the skew factor and $(x_0, y_0)^T$ are the co-ord of principal point

Affine camera

- With centre lying on the plane at infinity
- Camera center can be found using $PC = 0$
- In affine camera the last row is of the form $(0, 0, 0, 1)$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- Points at infinity are mapped to points at infinity
- Affine camera has 8 degrees of freedom (11 reduces to 8)

Affine projection contd..

- In terms of image and scene coordinates, the mapping takes the form

$$x = MX + t$$

where M is a general 2×3 matrix with elements $M_{ij} = \frac{T_{ij}}{T_{34}}$ while t is a general 2-vector representing the image center.
- Parallel world lines are projected to parallel image lines: the affine camera preserves parallelism.
- This mapping takes the point $(X, Y, Z, 1)^T$ to the image point $(X, Y, 1)^T$, dropping the z-coordinate
- Any projective camera matrix for which the principal plane is plane at infinity then it is affine camera

Affine projection contd..

- In terms of image and scene coordinates, the mapping takes the form
$$x = MX + t$$
where M is a general 2×3 matrix with elements $M_{ij} = \frac{P_{ij}}{P_{34}}$ while t is a general 2-vector representing the image center.
- Parallel world lines are projected to parallel image lines: the affine camera preserves parallelism.
- This mapping takes the point $(X, Y, Z, 1)^T$ to the image point $(X, Y, 1)^T$, dropping the z-coordinate
- Any projective camera matrix for which the principal plane is plane at infinity then it is affine camera

The Weak-Perspective Camera

- The affine camera becomes a weak-perspective camera when the rows of M form a uniformly scaled rotation matrix. The simplest form is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & Z_{av}/f \end{bmatrix}$$

which gives

$$M = \frac{f}{z_{av}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{z_{av}} \begin{bmatrix} X \\ Y \end{bmatrix}$$

- This is simply the perspective equation with individual point depths Z_i replaced by an average constant depth Z_{av}

The Weak-Perspective Camera contd..

- The weak-perspective model is valid when the average variation of the depth of the object (δz) along the line of sight is small compared to the Z_{av} and the field of view is small. We see this as follows.
- Expanding the perspective projection equation using a Taylor series, we obtain

$$x = \frac{f}{z_{av} + \delta z} \begin{bmatrix} X \\ Y \end{bmatrix} = \frac{f}{z_{av}} \left(1 - \frac{\delta z}{Z_{av}} + \left(\frac{\delta z}{Z_{av}}\right)^2 - \ldots \right) \begin{bmatrix} X \\ Y \end{bmatrix}$$

when $|\delta z| << Z_{av}$, only the zero order term remains and gives weak perspective projection.

The Weak-Perspective Camera contd..

- The error in image position then is $x_{err} = x_p - x_{wp}$.

$$x_{err} = -\frac{f}{z_{av}}\left(\frac{\delta z}{Z_{av} + \delta z}\right)\begin{bmatrix} X \\ Y \end{bmatrix}$$

The Orthographic Camera

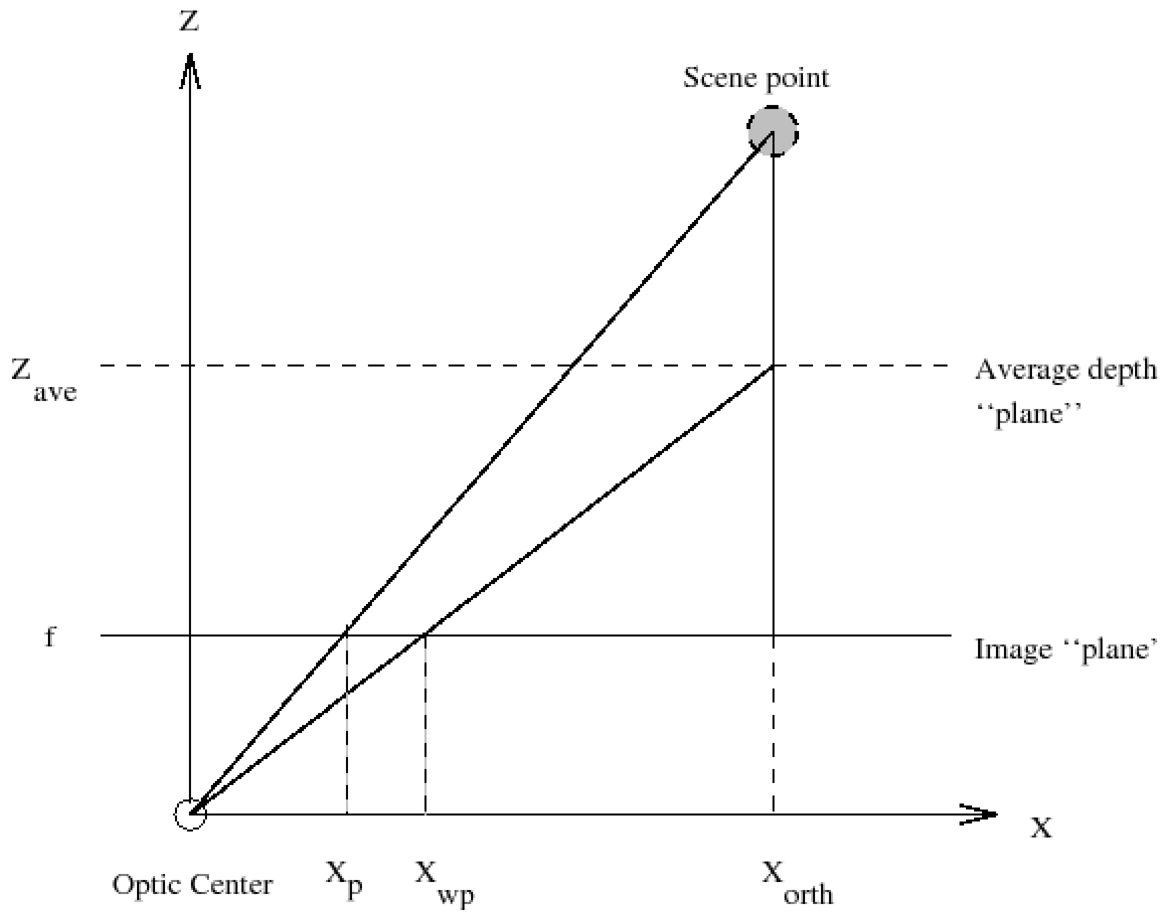
- The affine camera reduces to the case of orthographic (parallel) projection when M represents the first two rows of a rotation matrix. The simplest form is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which gives

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix}$$

The Orthographic Camera contd..



- 1D image formation with image plane at $Z = f$.
  X_p , X_{wp} and X_{orth} are the perspective, weak-perspective and orthographic projections respectively.

Conclusions

- 2D and 3D transformations
- Different camera models