Fall 2025 Principles of Safe Autonomy
ECE 484 (Sp 25)

Perception: Reconstructing 3D world from images

Sayan Mitra

Role of Perception in Autonomy

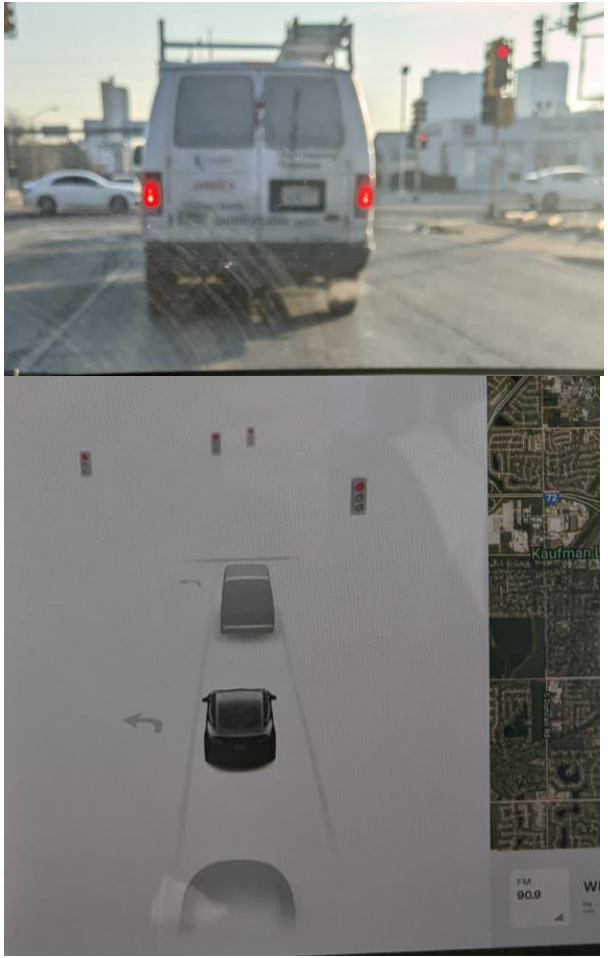
Perception module converts signals from the environment **state estimates** for the autonomous agent and its environment

Examples of state estimates:

- Type of lead vehicle, traffic sign
- Position of ego on the map, relative to the lane, distance to the leading vehicle
- Position of lead vehicle, speed, intention of the pedestrian

Types of estimates:

- Semantic: E.g., type of vehicle, sign
- Geometric: E.g., position, speed



Problem

Reconstructing the 3D structure of the scene from images

Input: image with points in pixels

Output: position of objects in millimeters in world camera frame

We will develop a method to find camera's internal and external parameters

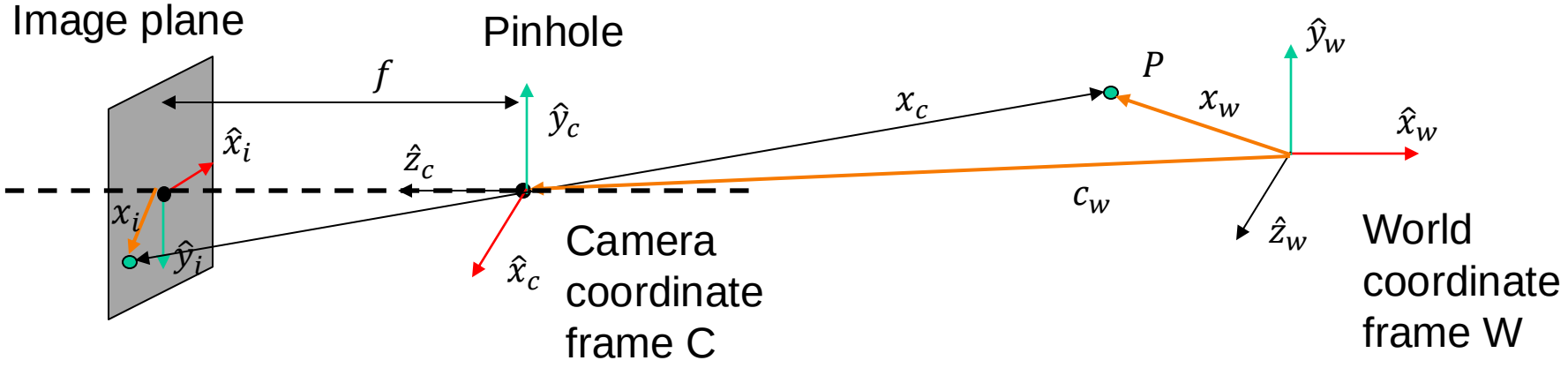
Outline:

Linear Camera Model (Projection matrix)

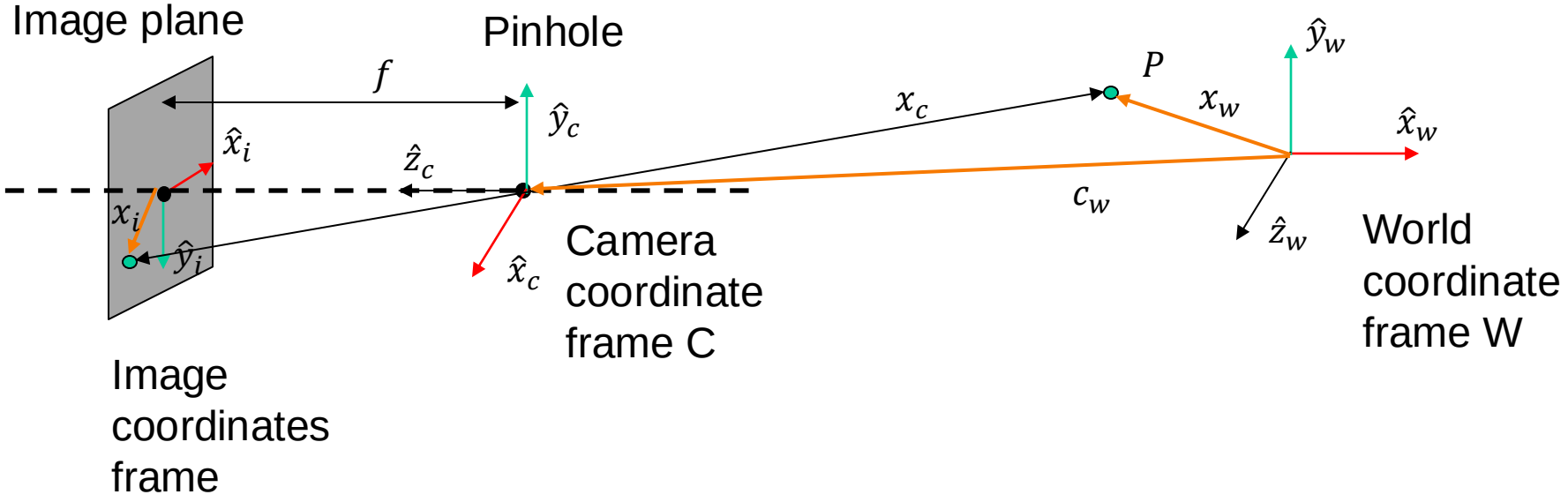
Camera calibration

Simple stereo

Forward Imaging Model: 3D to 2D

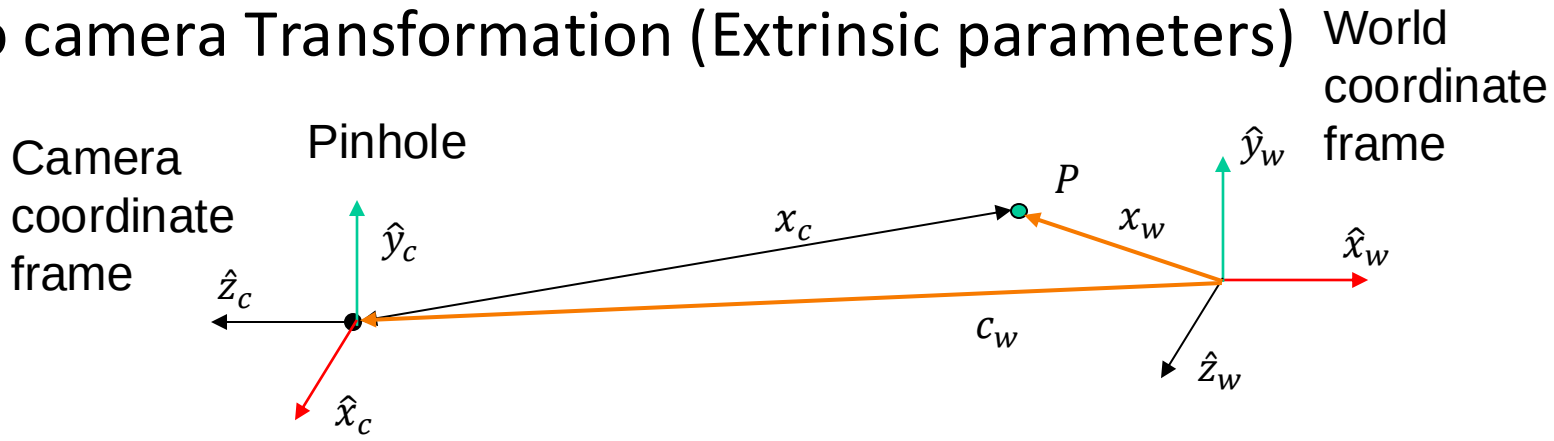


Forward Imaging Model: 3D to 2D



$$\boldsymbol{x}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad \text{3D-2D} \quad \boldsymbol{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \quad \text{3D-3D} \quad \boldsymbol{x}_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}$$

World to camera Transformation (Extrinsic parameters)

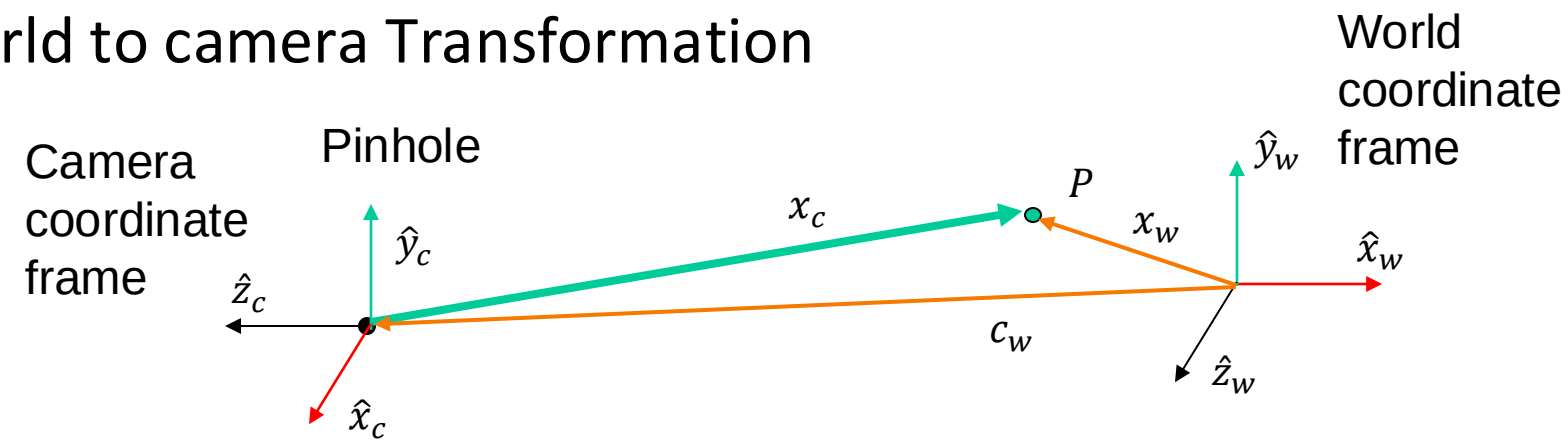


Position c_w and the orientation R of the camera in the world coordinate frame (W) are the camera's **Extrinsic Parameters**

$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$ → row 1 is the direction of $\hat{x}_c$ in world coordinates, 2 for $\hat{y}_c$, …

This is an **orthonormal matrix**, i.e., the row vectors or the colum vectors are orthonormal $R^{-1} = R^T$ i.e., $R^T R = R R^T = I$

World to camera Transformation



Position c_w and the orientation R of the camera in the world coordinate frame (W) are the camera's **Extrinsic Parameters**

Given the extrinsic parameters (R, c_w) of the camera, the camera-centric location of the point P in the world coordinate (w) is simply $(x_c)_w = x_w - c_w$

In the camera coordinate (c) $x_c = R(x_w - c_w) = Rx_w - Rc_w = Rx_w + t$

$$t = -Rc_w$$

$$x_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \qquad x_c = Rx_w + t$$

Extrinsic Matrix

$$x_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

We have an affine transformation: $x_c = Rx_w + t$

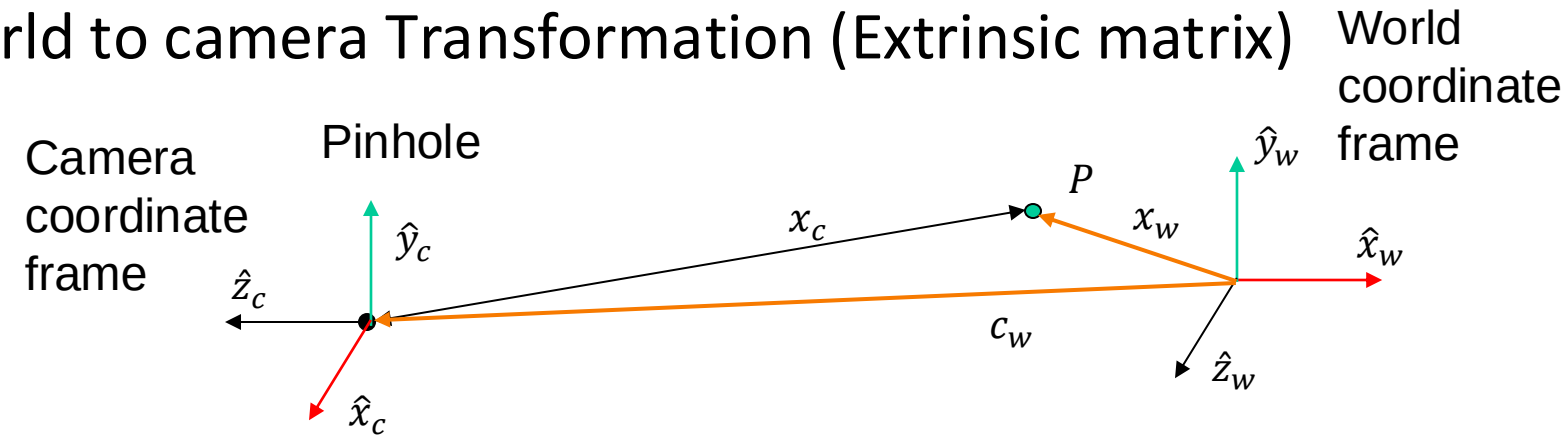
Can we represent it as $x_c = Mx_w$? No

We can introduce a new coordinate $\tilde{x}_c = [\tilde{x}, \tilde{y}, \tilde{z}, 1]^T$

Now can we represent this as a matrix multiplication $\tilde{x}_c = M\tilde{x}_w$

$$\tilde{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

World to camera Transformation (Extrinsic matrix)



Given the extrinsic parameters (R, c_w) of the camera, the camera-centric location of the point P in the world coordinate is

$$x_c = R(x_w - c_w) = Rx_w - Rc_w = Rx_w + t \qquad t = -Rc_w$$

$$x_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$ Using homogeneous coordinates

$$\tilde{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$ **Extrinsic matrix** $M_{ext}\ \tilde{x}_c = M_{eext}\ \tilde{x}_w$

Geometry of Homogeneous coordinates (for 2D)

Affine transformation: $\mathbf{x_c} = \mathbf{R}\mathbf{x_w} + \mathbf{t}$

$$x_c = \begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

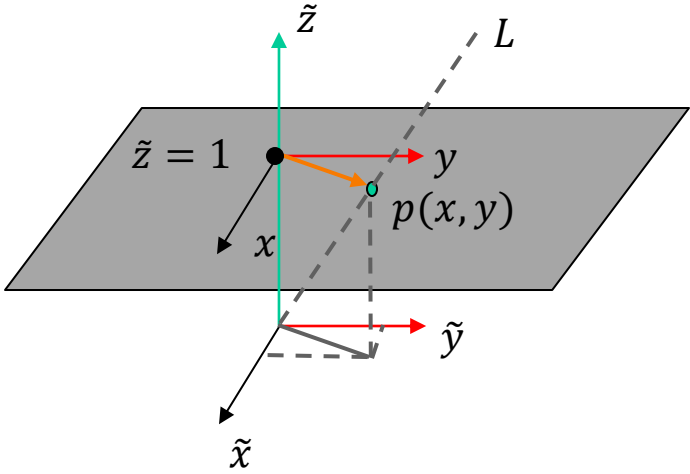
How to represent this as $\tilde{\mathbf{x}}_\mathbf{c} = \mathbf{M}\tilde{\mathbf{x}}_\mathbf{w}$

The homogeneous representation of a 2D point $p = (x, y)$ is a 3D point $\tilde{p} = (\tilde{x}, \tilde{y}, \tilde{z})$.

The third coordinate $\tilde{\mathbf{z}} \neq \mathbf{0}$ is fictitious such that:

$$p = (x, y) \quad x = \frac{\tilde{x}}{\tilde{z}} \; y = \frac{\tilde{y}}{\tilde{z}}$$

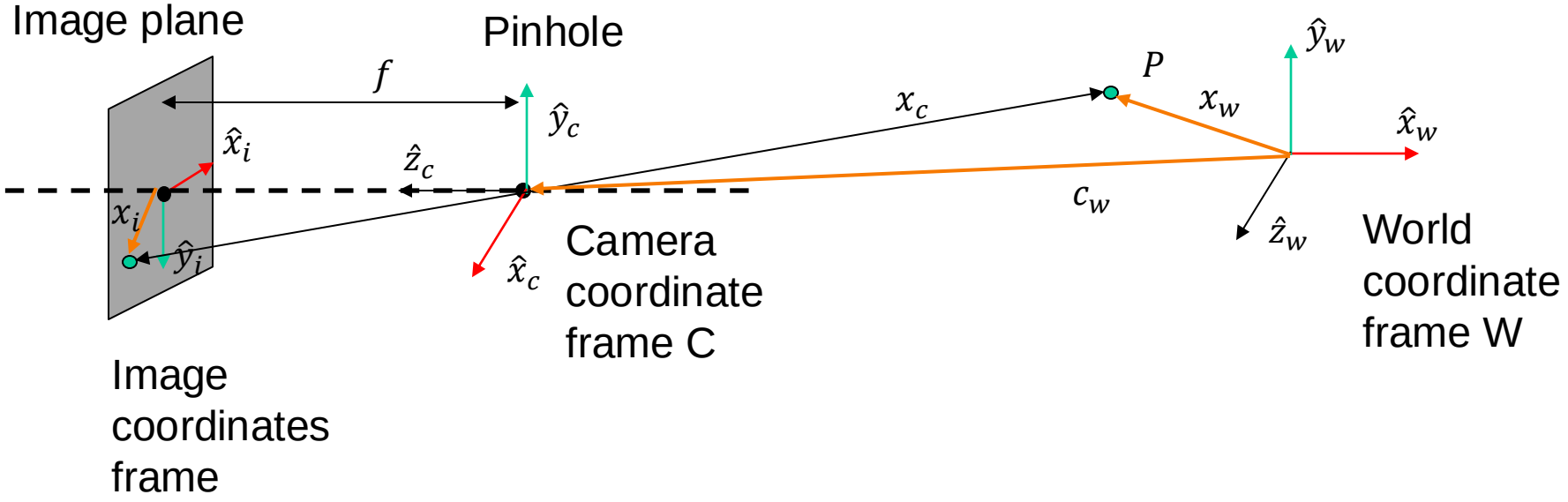
$$p \equiv \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{z}x \\ \tilde{z}y \\ \tilde{z} \end{bmatrix} \equiv \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = \tilde{p}$$



Geometric interpretation: all points on the line L (except origin) represent homogeneous coordinate $p(x, y)$

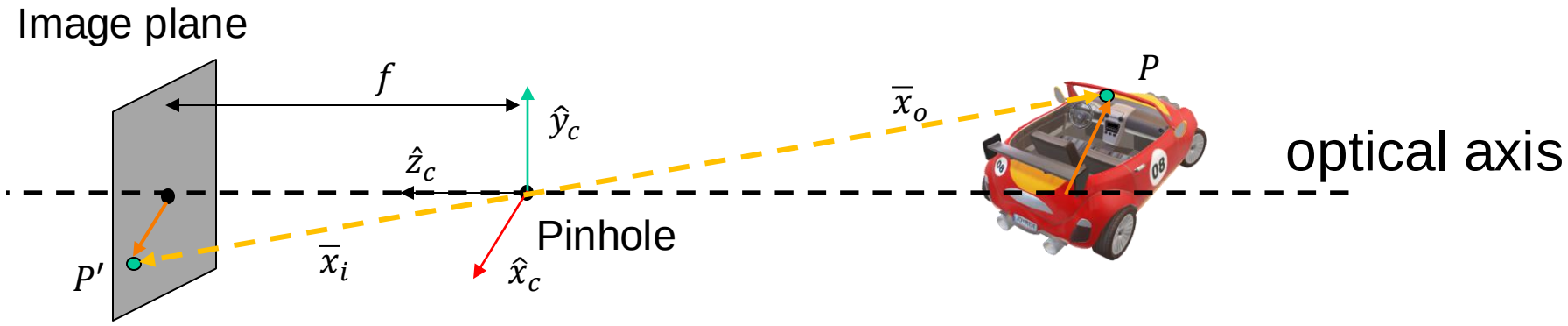
$$p \equiv \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \equiv \begin{bmatrix} wx \\ \tilde{w}y \\ \tilde{w}z \\ \tilde{w} \end{bmatrix} \equiv \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ \tilde{w} \end{bmatrix} = \tilde{p}$$

Forward Imaging Model: 3D to 2D



$$\boldsymbol{x}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad \xleftarrow{\text{3D-2D}} \quad \boldsymbol{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \quad \xleftarrow{\text{3D-3D}} \quad \boldsymbol{x}_w = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix}$$

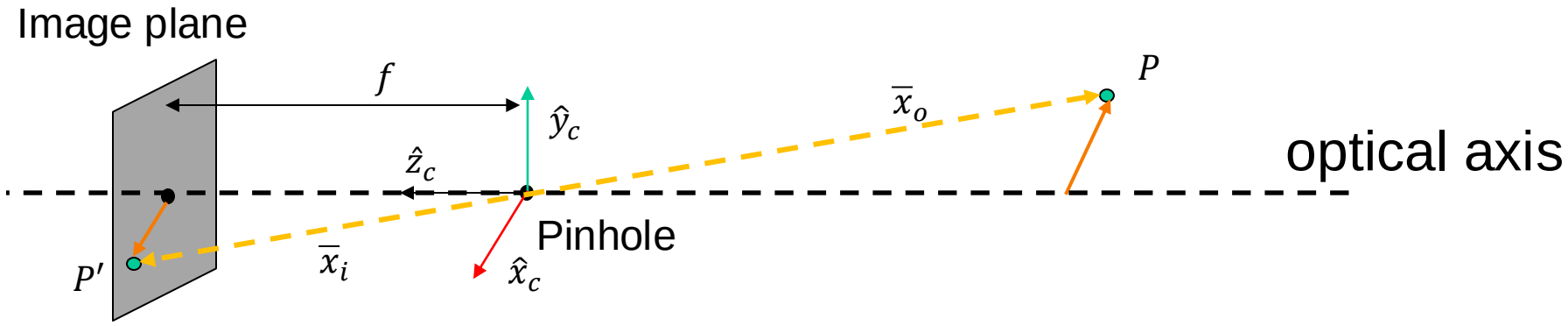
Perspective imaging with pinhole



f: Effective focal length

$$\boldsymbol{x}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} \quad \boldsymbol{x}_i = \begin{bmatrix} x_i \\ y_i \\ f \end{bmatrix}$$

Perspective imaging with pinhole

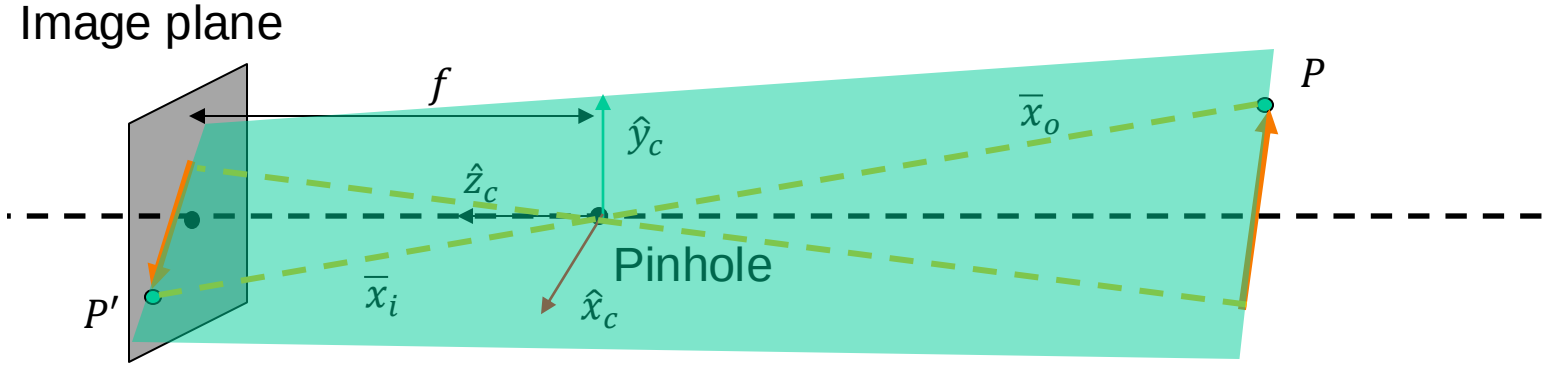


f: Effective focal length

$$\boldsymbol{x}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} \quad \boldsymbol{x}_i = \begin{bmatrix} x_i \\ y_i \\ f \end{bmatrix}$$

$$\frac{\overline{\boldsymbol{x}}_i}{f} = \frac{\overline{\boldsymbol{x}}_o}{z_0} \Rightarrow \frac{x_i}{f} = \frac{x_0}{z_0}, \frac{y_i}{f} = \frac{y_0}{z_0}$$

Perspective projection of a line



f: Effective focal length

$$\boldsymbol{x}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} \quad \boldsymbol{x}_i = \begin{bmatrix} x_i \\ y_i \\ f \end{bmatrix} \qquad \frac{\bar{x}_i}{f} = \frac{\bar{x}_o}{z_0} \quad \Rightarrow \frac{x_i}{f} = \frac{x_0}{z_0}, \frac{y_i}{f} = \frac{y_0}{z_0}$$

Forward Imaging Model: Camera to pixel

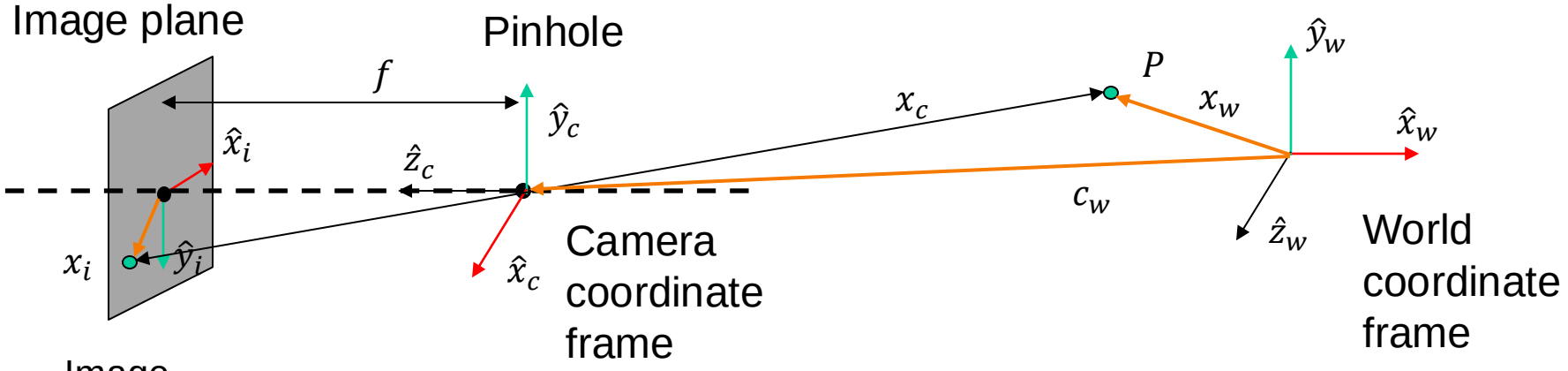


Image coordinates frame

Perspective projection

$$\frac{x_i}{f} = \frac{x_c}{z_c} \text{ and } \frac{y_i}{f} = \frac{y_c}{y_c} \quad x_i = f\frac{x_c}{z_c} \text{ and } y_i = f\frac{y_c}{y_c}$$

Pixel coordinates with pixel densities m_x and m_y

(o_x, o_y) Principle point

$$u = m_x f\frac{x_c}{z_c} + o_x \text{ and } v = m_y f\frac{y_c}{z_c} + o_y$$

$$u = f_x\frac{x_c}{z_c} + o_x \text{ and } v = f_y\frac{y_c}{z_c} + o_y$$

Perspective Projection

$$u = f_x \frac{x_c}{z_c} + o_x \text{ and } v = f_y \frac{y_c}{z_c} + o_y$$

Intrinsic parameters: f_x, f_y, o_x, o_y

Nonlinear model

We use homogeneous representation of 2D point $u = (u, v)$ as a 3D point $\tilde{u} = (\tilde{u}, \tilde{v}, \tilde{w})$ the third coordinate $\tilde{w} \neq 0$ is fictitious such that $u = \tilde{u}/\tilde{w}$ $v = \tilde{v}/\tilde{w}$.

$$u \equiv \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{w}u \\ \tilde{w}v \\ \tilde{w} \end{bmatrix} \equiv \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} \equiv \tilde{u}$$

Homogeneous representation of a 3D point in 4D

$$x \equiv \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{w}x \\ \tilde{w}y \\ \tilde{w}z \\ \tilde{w} \end{bmatrix} \equiv \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ \tilde{w} \end{bmatrix} \equiv \tilde{x}$$

Perspective Projection: From camera coordinates to pixel coordinates

$$u = f_x \frac{x_c}{z_c} + o_x \text{ and } v = f_y \frac{y_c}{z_c} + o_y$$

Homogeneous representation (u, v):

$$u \equiv \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} \equiv \begin{bmatrix} z_c u \\ z_c v \\ z_c \end{bmatrix} = \begin{bmatrix} f_x x_c + z_c o_x \\ f_y y_c + z_c o_y \\ z_c \end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x & 0 \\ 0 & f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}$$

Linear model of perspective projection $\tilde{u} = [K|0]\tilde{x}_c = M_{int}\tilde{x}_c$

Or, $\tilde{u} = K\boldsymbol{x}_c$

Intrinsic matrix (M_{int})

Calibration matrix K (upper right triangular)

Forward Camera Model

Camera to pixel

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x & 0 \\ 0 & f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}$$

$$\tilde{u} = M_{int}\ \tilde{x}_w$$

$$\tilde{u} = M_{int}\, M_{ext}\, \tilde{x}_w = P\, \tilde{x}_w$$

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

P: Projection matrix

World to camera

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

$$\tilde{x}_c = M_{ext}\ \tilde{x}_w$$

Camera Calibration Procedure

Try to cover this.

Useful for GEM

Homography

Image transformations

2x2 transformations

3x3 transformations

Computing homography

Dealing with outliers RANSAC

Image manipulation

Image filtering: Change range (e.g., brightness)

$g(x,y) = Tr(f(x,y))$

Tr

Image warping: Change domain (e.g., rotation)

$g(x,y) = f(Td(x,y))$

2x2 Linear Transformations

$p_1 = (x_1, y_1)$

$p_2 = (x_2, y_2)$

$p_2 = Tp_1$

T can be represented by a matrix

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Scaling (stretching or squishing)

Forward

$$x_2 = ax_1 \quad y_2 = by_1$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = S \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

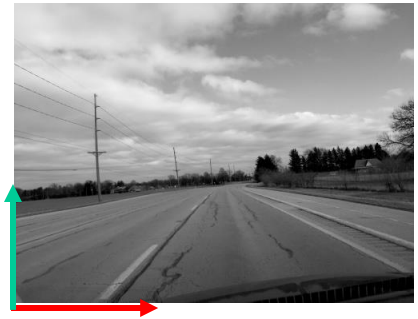
Inverse

$$x_2 = \frac{1}{a} x_1 \quad y_2 = \frac{1}{b} y_1$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = S^{-1} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{b} \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

2D Rotation

$$x_1 = r\,cos(\psi)$$
$$y_1 = r\,sin(\psi)$$



Forward

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = R \begin{bmatrix} x \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Inverse

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = R^{-1} \begin{bmatrix} x \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

2x2 Matrix Transformations

Any transformation of the form:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Examples: Scaling, rotation, skew, mirror

Properties

- Origin maps to origin
- Lines map to lines
- Parallel lines remain parallel
- Closed under composition

If $p_2 = T_{21}p_1$ $p_3 = T_{32}p_2$ then $p_3 = T_{31}p_1$ where $T_{31} = T_{32}T_{21}$

Translation

Forward

$x_2 = x_1 + t_x$

$y_2 = y_1 + t_y$

Can a 2x2 matrix express this transformation? No

Homogeneous coordinates

The homogeneous representation of a 2D point $p = (x, y)$ is a 3D point $\tilde{p} = (\tilde{x}, \tilde{y}, \tilde{z})$. The third coordinate $\tilde{z} \neq 0$ is fictitious such that:

$$x = \frac{\tilde{x}}{\tilde{z}} \; y = \frac{\tilde{y}}{\tilde{z}}$$

$$p \equiv \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{z}x \\ \tilde{z}y \\ \tilde{z} \end{bmatrix} \equiv \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = \tilde{p}$$

Translation

Forward

$$x_2 = x_1 + t_x$$

$$y_2 = y_1 + t_y$$

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{x}_2 \\ \tilde{y}_2 \\ \tilde{z}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix}$$

3x3 Affine Transformations

Scaling

$$\begin{bmatrix}\tilde{x}_2\\ \tilde{y}_2\\ \tilde{z}_2\end{bmatrix} = \begin{bmatrix}s_x & 0 & 0\\ 0 & s_y & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ y_1\\ 1\end{bmatrix}$$

Skew

$$\begin{bmatrix}\tilde{x}_2\\ \tilde{y}_2\\ \tilde{z}_2\end{bmatrix} = \begin{bmatrix}1 & m_x & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ y_1\\ 1\end{bmatrix}$$

Translation

$$\begin{bmatrix}\tilde{x}_2\\ \tilde{y}_2\\ \tilde{z}_2\end{bmatrix} = \begin{bmatrix}1 & 0 & t_x\\ 0 & 1 & t_y\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ y_1\\ 1\end{bmatrix}$$

Rotation

$$\begin{bmatrix}\tilde{x}_2\\ \tilde{y}_2\\ \tilde{z}_2\end{bmatrix} = \begin{bmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ y_1\\ 1\end{bmatrix}$$

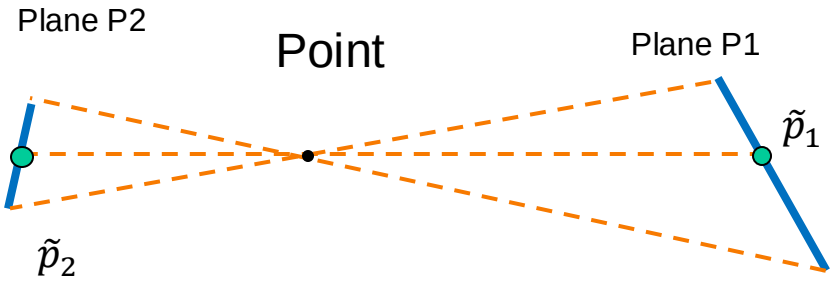
Composition of Transformations

General form $\begin{bmatrix}\tilde{x}_2\\ \tilde{y}_2\\ \tilde{z}_2\end{bmatrix} = \begin{bmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_1\\ y_1\\ 1\end{bmatrix}$

Projective Transformations

General form $\begin{bmatrix} \tilde{x}_2 \\ \tilde{y}_2 \\ \tilde{z}_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}$ $\tilde{p}_2 = H\tilde{p}_1$

H is called a homography



Origin does not necessarily map to origin

Lines map to lines

Parallel lines do not necessarily remain parallel

Closed under composition

Homographies are defined up to scaling