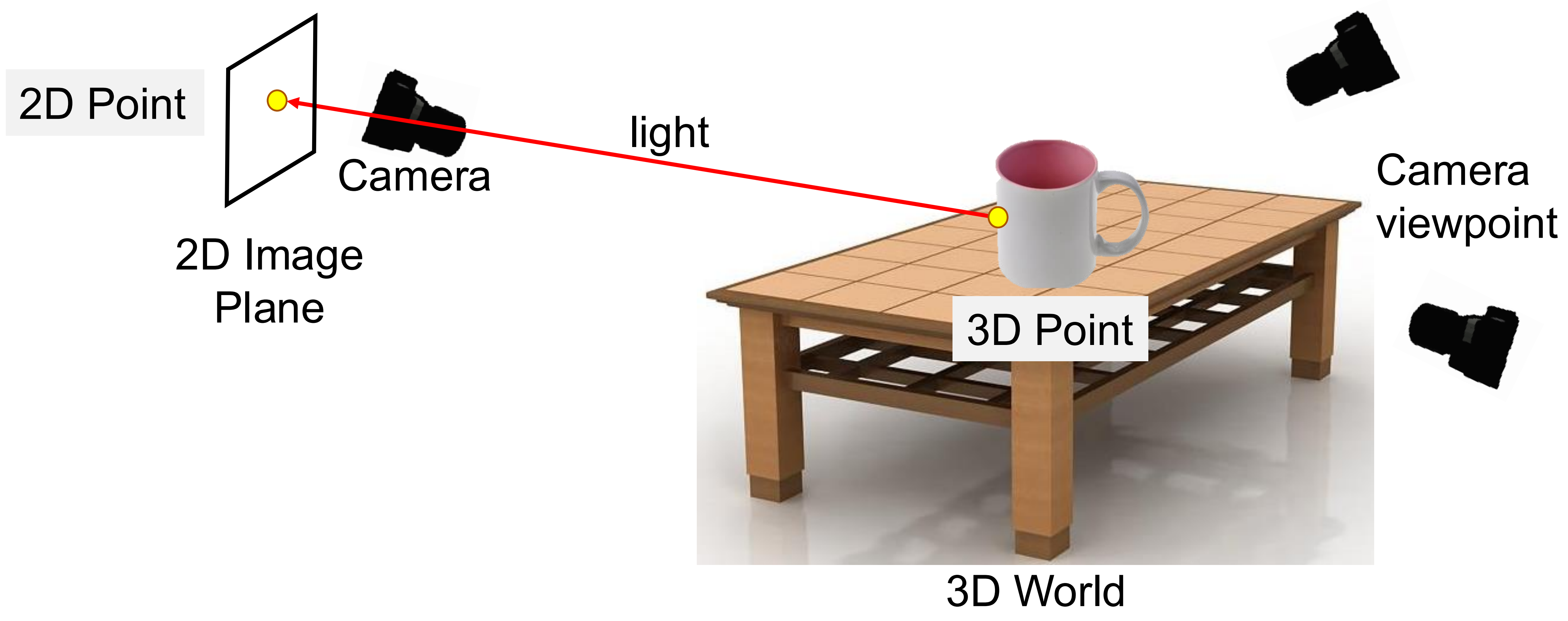
Basics on Computer Vision

Jikai Wang

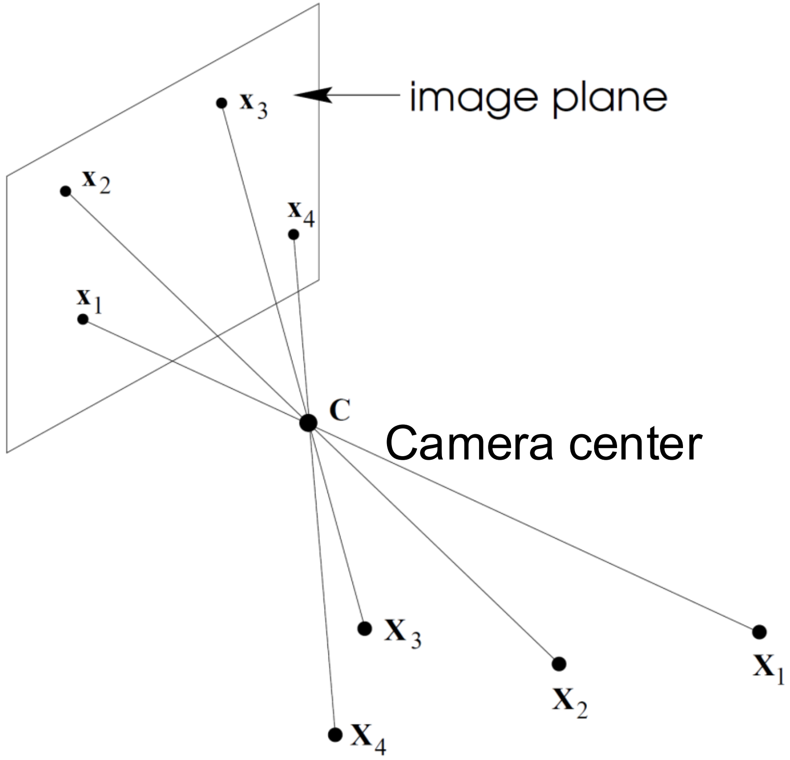
THE UNIVERSITY OF TEXAS AT DALLAS

Geometric Primitives

Geometry in Image Generation



2D Points and 3D Points



A 2D point is usually used to indicate pixel coordinates of a pixel

$$\mathbf{x} = (x, y) \in \mathcal{R}^2 \qquad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

A 3D point in the real world

$$\mathbf{x} = (x, y, z) \in \mathcal{R}^3 \qquad \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Homogeneous Coordinates

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

homogeneous image coordinates

$$(x, y, z) \Rightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = w \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

homogeneous scene coordinates

Up to scale

Conversion

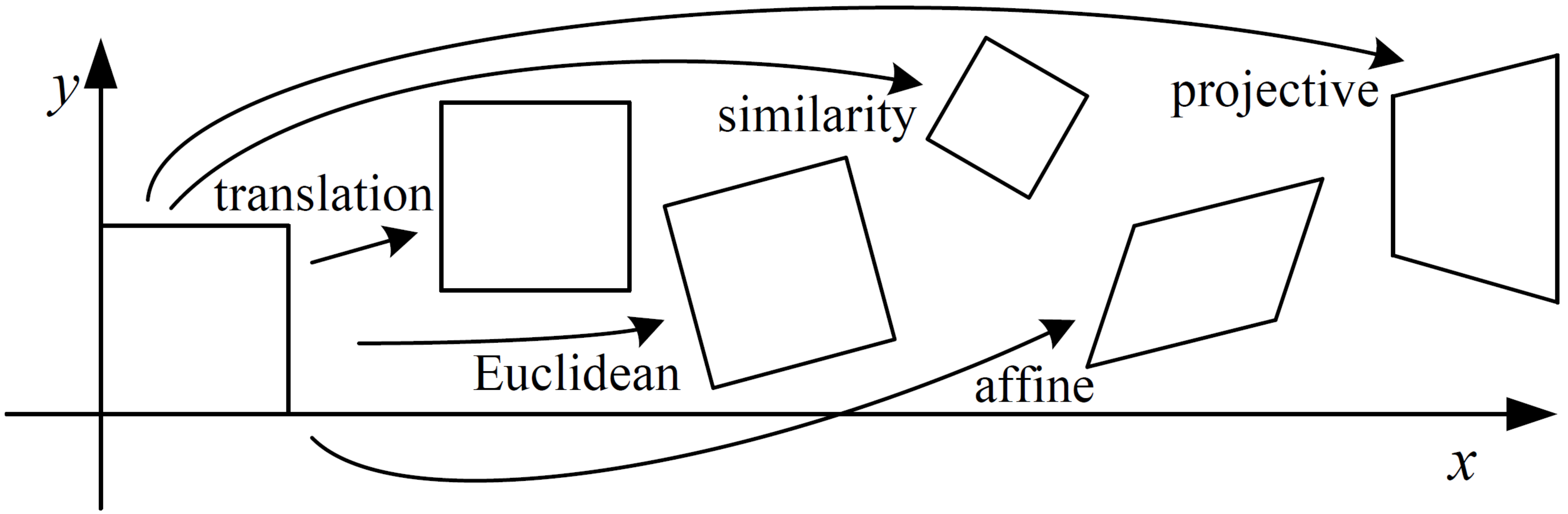
$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$$

THE UNIVERSITY OF TEXAS AT DALLAS

2D Transformation

2D Transformations



2D Translation

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

$$\mathbf{x}' = \mathbf{x} + \mathbf{t}$$

$$\mathbf{x}' = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{t} \end{bmatrix}}_{2 \times 3} \bar{\mathbf{x}}$$

Homogeneous coordinate

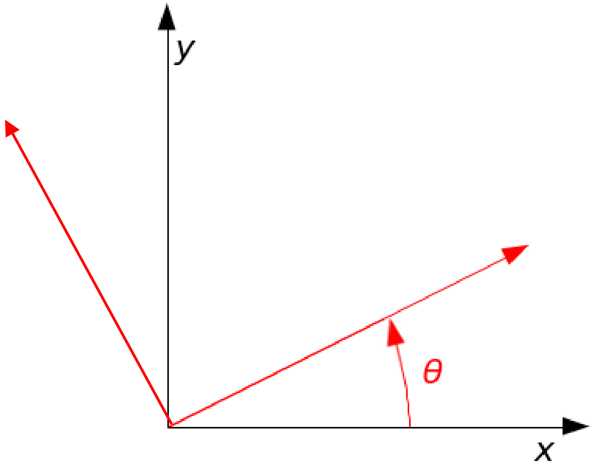
$$\bar{\mathbf{x}}' = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}}_{3 \times 3} \bar{\mathbf{x}}$$

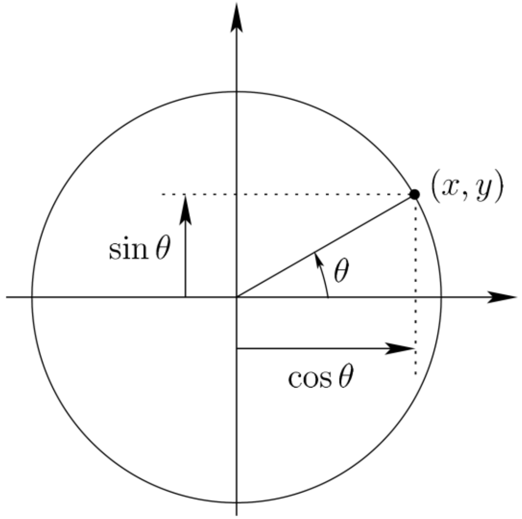
augmented vector $\bar{\mathbf{x}} = (x, y, 1)$

2D Rotation

$$\mathbf{x}' = \mathbf{R}\mathbf{x}$$

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$





$$[\hat{\mathrm{x}}_{\mathrm{b}} \quad \hat{\mathrm{y}}_{\mathrm{b}}] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

orthonormal rotation matrix

$$\mathbf{R}\mathbf{R}^T = \mathbf{I} \text{ and } |\mathbf{R}| = 1$$

2D Euclidean Transformation

- 2D Rotation + 2D translation

$$\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t} \qquad \mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \end{aligned}$$

2D Euclidean Transformation

- 2D Rotation + 2D translation

$$\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t} \qquad \mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$\mathbf{x}' = \underbrace{\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}}_{2 \times 3} \bar{\mathbf{x}}$$

$$\bar{\mathbf{x}} = (x, y, 1)$$

- Degree of freedom (DOF)
  - The maximum number of logically independent values
  - 2D Rotation?
  - 2D Euclidean transformation?

2D Similarity Transformation

- Scaled 2D rotation + 2D translation

$$\mathbf{x}' = s\mathbf{R}\mathbf{x} + \mathbf{t} \qquad \mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$\mathbf{x}' = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \end{bmatrix} \bar{\mathbf{x}} = \begin{bmatrix} a & -b & t_x \\ b & a & t_y \end{bmatrix} \bar{\mathbf{x}} \qquad \bar{\mathbf{x}} = (x, y, 1)$$

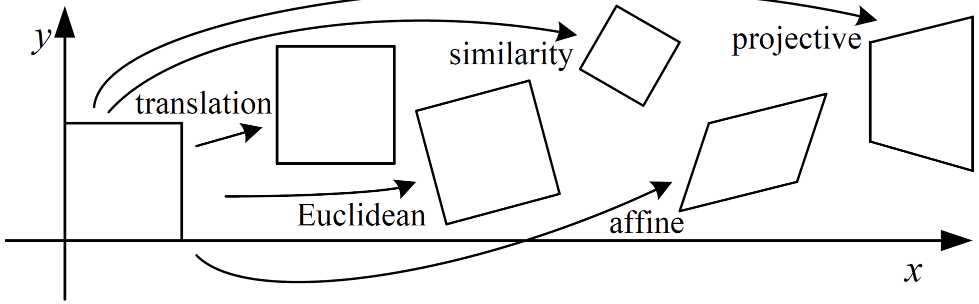
The similarity transform preserves angles between lines.

2D Affine Transformation

- Arbitrary 2x3 matrix

$$\mathbf{x}' = \mathbf{A}\bar{\mathbf{x}} \qquad \bar{\mathbf{x}} = (x, y, 1)$$

$$\mathbf{x}' = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \bar{\mathbf{x}}$$



Parallel lines remain parallel under affine transformations.

2D Projective Transformation

- Also called perspective transform or homography

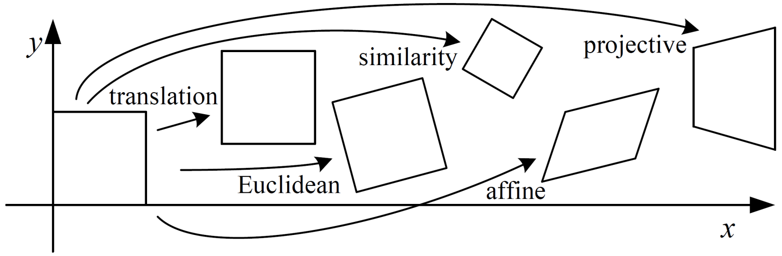
$$\tilde{\mathbf{x}}' = \tilde{\mathbf{H}}\tilde{\mathbf{x}}$$

homogeneous coordinates

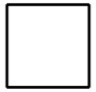
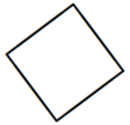
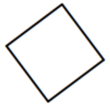
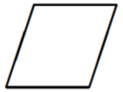
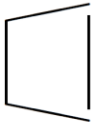
3×3 $\tilde{\mathbf{H}}$ is only defined up to a scale

$$x' = \frac{h_{00}x + h_{01}y + h_{02}}{h_{20}x + h_{21}y + h_{22}} \quad \text{and} \quad y' = \frac{h_{10}x + h_{11}y + h_{12}}{h_{20}x + h_{21}y + h_{22}}$$

Perspective transformations preserve straight lines.



Hierarchy of 2D Transformations

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix} \mathbf{I} & \mathbf{t} \end{bmatrix}_{2\times 3}$	2	orientation	
rigid (Euclidean)	$\begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix}_{2\times 3}$	3	lengths	
similarity	$\begin{bmatrix} s\mathbf{R} & \mathbf{t} \end{bmatrix}_{2\times 3}$	4	angles	
affine	$\begin{bmatrix} \mathbf{A} \end{bmatrix}_{2\times 3}$	6	parallelism	
projective	$\begin{bmatrix} \tilde{\mathbf{H}} \end{bmatrix}_{3\times 3}$	8	straight lines	

THE UNIVERSITY OF TEXAS AT DALLAS

3D Transformation

3D Translation

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

$$\mathbf{x}' = \mathbf{x} + \mathbf{t}$$

$$\mathbf{x}' = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{t} \end{bmatrix}}_{3 \times 4} \bar{\mathbf{x}}$$

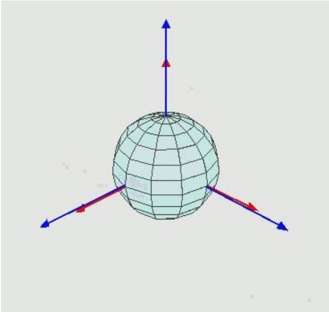
augmented vector $\bar{\mathbf{x}} = (x, y, z, 1)$

3D Rotation Representations

- Rotation matrix

$$R_{3\times 3} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

- Euler angles



- Axis-angle

$$\omega = \theta \hat{n}$$

- Unit quaternion

$$q = w + xi + yj + zk$$

3D Euclidean Transformation SE(3)

- 3D Rotation + 3D translation

$$x' = Rx + t$$

$$x' = \underset{3 \times 4}{[R|t]}\bar{x}$$

$$\bar{x} = (x, y, z, 1)$$

3D Similarity Transformation

- Scaled 3D rotation + 3D translation

$$\mathbf{x}' = s\mathbf{R}\mathbf{x} + \mathbf{t}$$

$$\mathbf{x}' = \underbrace{\begin{bmatrix} s\mathbf{R} & \mathbf{t} \end{bmatrix}}_{3 \times 4} \bar{\mathbf{x}} \qquad \bar{\mathbf{x}} = (x, y, z, 1)$$

This transformation preserves angles between lines and planes.

3D Affine Transformation

$$\mathbf{x}' = \mathbf{A}\bar{\mathbf{x}} \qquad \bar{\mathbf{x}} = (x, y, z, 1)$$

$$\mathbf{x}' = \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \end{bmatrix} \bar{\mathbf{x}}$$

$$3 \times 4$$

Parallel lines and planes remain parallel under affine transformations.

3D Projective Tranformation

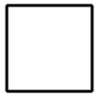
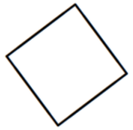
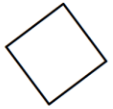
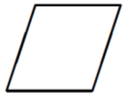
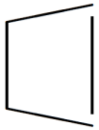
- Also called 3D perspective transform or homography

$$\tilde{\mathbf{x}}' = \tilde{\mathbf{H}}\tilde{\mathbf{x}}$$ homogeneous coordinates

4×4 $\tilde{\mathbf{H}}$ is only defined up to a scale

Perspective transformations preserve straight lines.

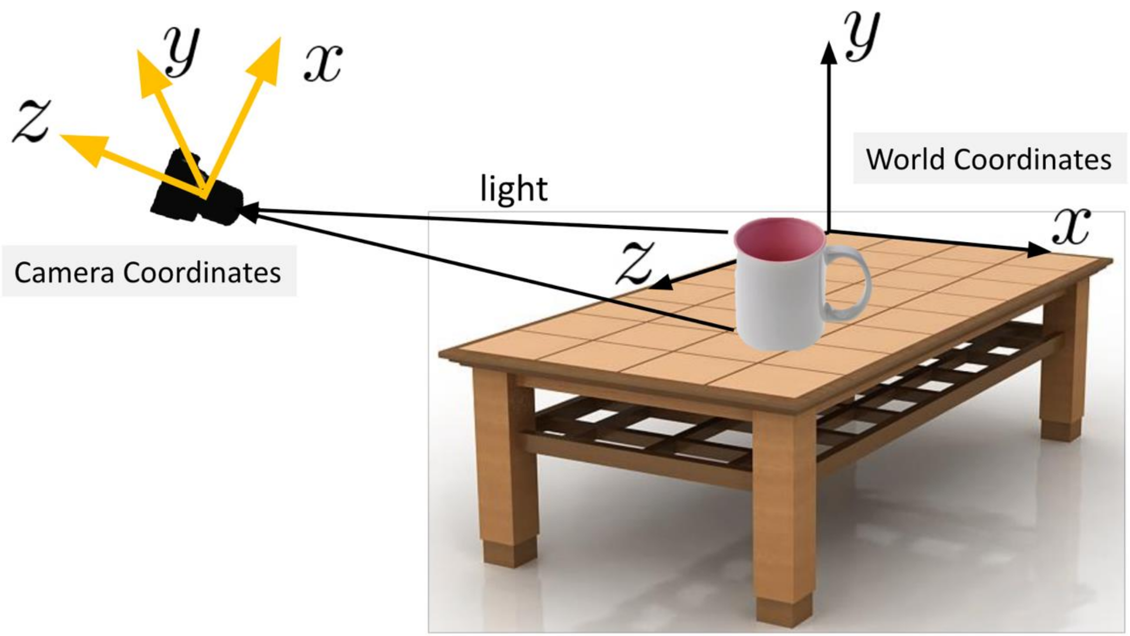
3D Transformations

Transformation	Matrix	# DoF	Preserves	Icon
translation	$\begin{bmatrix}\mathbf{I} & \mathbf{t}\end{bmatrix}_{3\times 4}$	3	orientation	
rigid (Euclidean)	$\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}_{3\times 4}$	6	lengths	
similarity	$\begin{bmatrix}s\mathbf{R} & \mathbf{t}\end{bmatrix}_{3\times 4}$	7	angles	
affine	$\begin{bmatrix}\mathbf{A}\end{bmatrix}_{3\times 4}$	12	parallelism	
projective	$\begin{bmatrix}\tilde{\mathbf{H}}\end{bmatrix}_{4\times 4}$	15	straight lines	

THE UNIVERSITY OF TEXAS AT DALLAS

Perspective Camera Model

Perspective Camera Model



- Camera intrinsics

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

- Camera extrinsics

$$\underbrace{\mathbf{X}_{\text{cam}}}_{\text{camera coordinates}} = R\underbrace{\mathbf{X}}_{\text{world coordinates}} + \mathbf{t}$$

- Camera Projection Matrix

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

- World space point to image plane pixel

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

- Camera space point to image plane pixel

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

Back-projection to a 3D Point in Camera Coordinates

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = K^{-1} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \cdot z$$

$$K^{-1} = \begin{bmatrix} 1/f_x & 0 & -c_x/f_x \\ 0 & 1/f_y & -c_y/f_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} (u - c_x) \cdot \frac{z}{f_x} \\ (v - c_y) \cdot \frac{z}{f_y} \\ z \end{bmatrix}$$

THE UNIVERSITY OF TEXAS AT DALLAS

Thank You!