CS231A Notes

Sean Wu

June 28, 2020

Contents

1 Camera Models

1.1 Piece of Film (Camera Obscura) - da Vinci 1500s

- Camera obscura $\longrightarrow$ means dark chamber
- Each point on the 3D object emits multiple rays of light outwards
- Add a barrier to block off most of the light rays
  - Reduces blurring
  - Establishes a one-to-one mapping between points on 3D object and the film

aperture : opening in barrier to allow some light rays to pass through and hit film

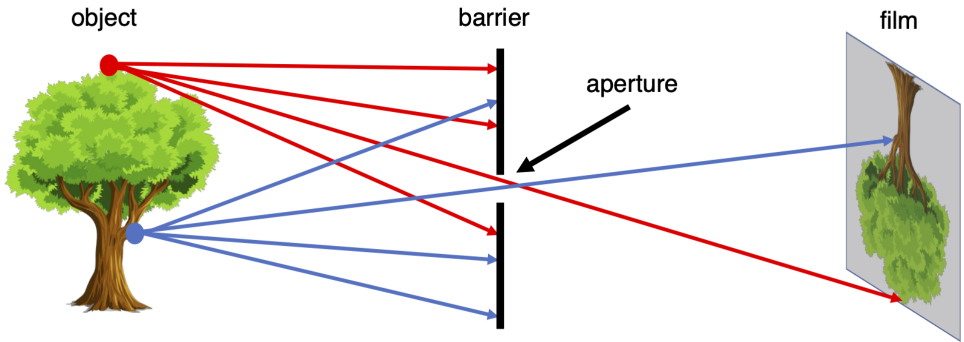

Figure 1: A simple working camera model: the pinhole camera model.

1.2 Pinhole Camera

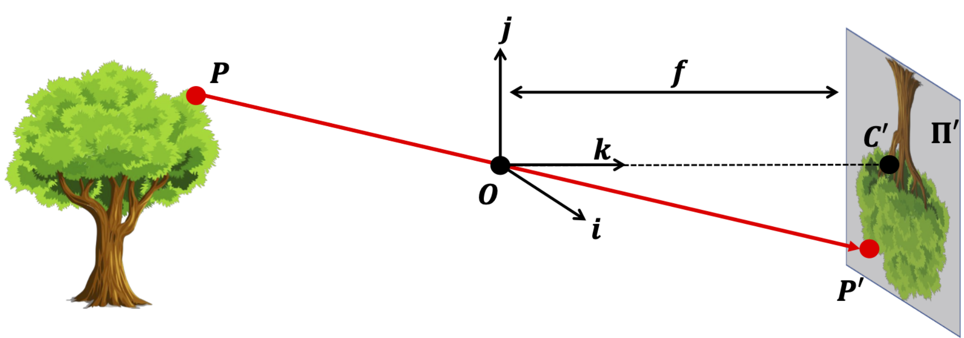

Figure 2: A formal construction of the pinhole camera model.

image or retinal plane : the film inside the camera

f : focal length

O : aperture; pinhole centre of camera

virtual image or virtual retinal plane : when image or retinal plane is placed in between the camera centre O and the 3D object at a distance f from O

optical axis : the line defined by C' and O

convention : indicate projected or complementary points using the prime superscript (ex. P' indicates point on image plane)

- **Key Assumption:** the aperture in the pinhole model is a single point
  - In most real world scenarios, we cannot assume that the aperture is infinitely small
  - As aperture size increases $\longrightarrow$ number of light rays passing through the barrier increases
  - With more light rays passing through, each point on the film can be affected by light rays from multiple points in 3D space, blurring the image
  - Smaller aperture $\longrightarrow$ crisper but darker images
- Note: projection of the object in the image plane and the image of the object in the virtual image plane are identical up to a scale (similarity) transformation
- Let $P = \begin{bmatrix} x & y & z \end{bmatrix}^T$ be a point on some 3D object visible to the pinhole camera
- P gets mapped or **projected** onto the image plane Π', resulting in point $P' = \begin{bmatrix} x' & y' \end{bmatrix}^T$
- Similarly, the pinhole itself can be projected onto th image plane, giving a new point C'
- Define a coordinate system $\begin{bmatrix} i & j & k \end{bmatrix}$ centered at the pinhole O such that the axis k is perpendicular to the image plane and points toward it
- Since the point P' is obtained by the projection of the 3D point P, on the image plane Π', we can understand the way the 3D world is captured on the 2D image by finding the relationship between 3D point P and image plane point P'
- Note: the triangle $P'C'O$ is similar to the triangle formed by P, O and $(0, 0, z)$

$$P = \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\text{3D}} \longrightarrow P' = \underbrace{\begin{bmatrix} x' \\ y' \end{bmatrix}}_{\text{2D}} = \begin{bmatrix} f\frac{x}{z} \\ f\frac{y}{z} \end{bmatrix} \qquad (\mathbb{R}^3 \xrightarrow{E} \mathbb{R}^2)$$

- If aperture is too small, less light passes through
  - Soln: use a lens

1.3 Lenses

lens : focuses light onto the film; addresses crispness vs brightness tradeoff for smaller apertures

- If pinhole is replaced with the right lens (position and size), it satisfies:

– **Property of Correct Lens:** all light rays emitted by a point P are refracted by the lens such that they converge to a single point P' in the image plane
– Note: this only occurs for a specific point P that is in focus
– Another point Q that is closer/further from the image plane than P will be blurred or out of focus

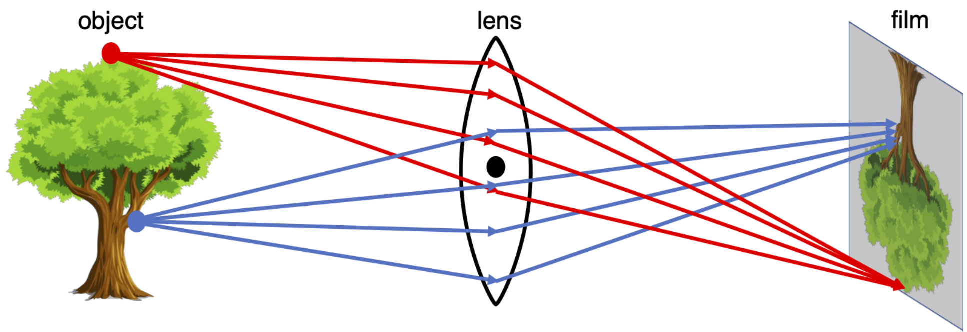

Figure 3: A setup of a simple lens model. Notice how the rays of the top point on the tree converge nicely on the film. However, a point at a different distance away from the lens results in rays not converging perfectly on the film (blurred).

- Lenses have a special distance at which objects are "in focus"
- Related to concept of depth of field
- All rays parallel to the optical (or principal) axis converge to one point (the **focal point**) on a plane located at the **focal length** f from the centre of the lens
– Caused by light refraction in lens
- Rays passing through the centre are not diverted

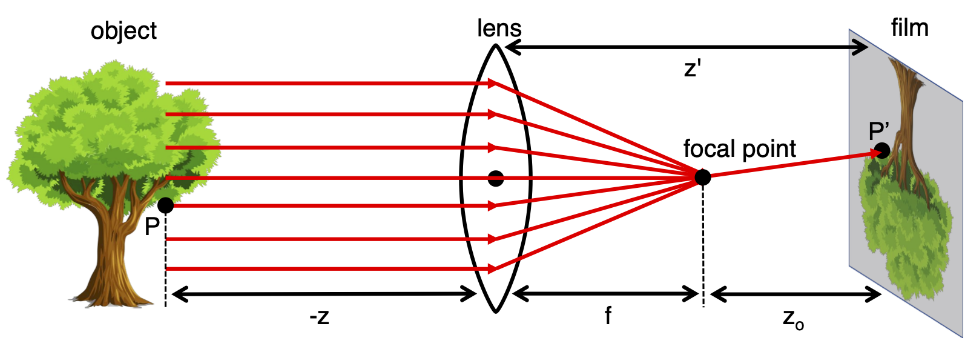

Figure 4: Lenses focus light rays parallel to the optical axis into the focal point. Furthermore, this setup illustrates the paraxial refraction model, which helps us find the relationship between points in the image plane and the 3D world in cameras with lenses.

1.3.1 Paraxial Refraction Model - Thin Lens Assumption

- With the lens, we can relate a point P in 3D space with its corresponding image plane point P' using Snell's law (w/ small angle approximation $sin\theta \approx \theta$ as $\theta \to 0$)

$$\begin{cases} x' = z'\frac{x}{z} & z' = f + z_0 \\ y' = z'\frac{y}{z} & f = \frac{R}{2(n-1)} \end{cases} \tag{1}$$

1.3.2 Issues with Lenses

- Because paraxial refraction model approximates using the thin lens assumption, many aberrations can occur

1.3.3 Radial Distortion

radial distortion : image magnification decreases/increases as a function of the distance to optical axis

pincushion distortion : magnification increases further from the optical axis

barrel distortion : magnification decreases further from the optical axis; usually occurs with fish-eye lenses

- Deviations are most noticeable for rays that pass through the edge of the lens

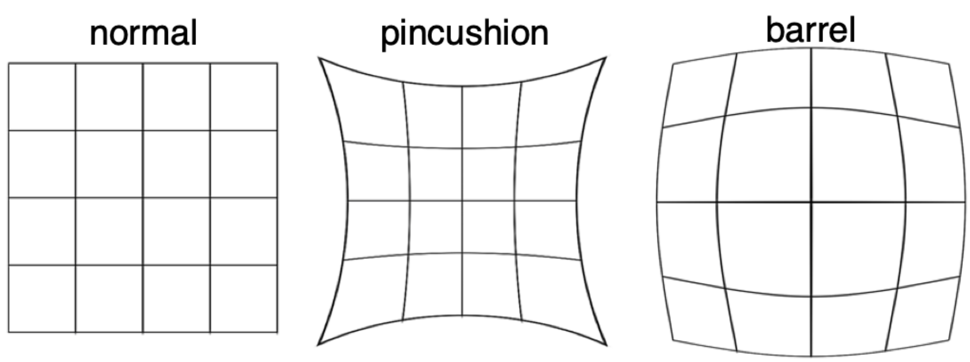

Figure 5: Demonstrating how pincushion and barrel distortions affect images.

1.4 Going to Digital Image Space

- Results derived using the pinhole model, but they hold as well for the paraxial refraction model
- A point P in 3D space can be mapped (or projected) into a 2D point P' in the image plane Π' using a projective transformation

projective transformation : a $\mathbb{R}^3 \to \mathbb{R}^2$ mapping of points in 3D space to 2D points in the image plane

- Projective transformations do not directly correspond to what we see in actual digital images

1. Points in digital are, in general, in a different reference system than those in the image plane
2. Digital images are divided into discrete pixels, whereas points in the image plane are continuous
3. Physical sensors can introduce non-linearity (ex. distortion) to the mapping

- Thus have to introduce additional transformations to be able to map any point from the 3D world to pixel coordinates

1.5 Coordinate systems

- Image coordinates have their origin C' at the image centre where the k axis intersects the image plane
- However, digital images typically have their origin at the lower-left corner of the image
- Thus, 2D points in the image plane and 2D points in the image are offset by a translation vector $\begin{bmatrix} c_x & c_y \end{bmatrix}^T$

$$(x, y, z) \longrightarrow \left(f\frac{x}{z} + c_x, f\frac{y}{z} + c_y \right) \tag{2}$$

- Points in digital images are expressed in pixels, while points in image plane are represented in physical measurements (ex. metres)
- Convert from metric to pixels with new parameters k and ℓ

$$(x, y, z) \longrightarrow \left(\underbrace{fk}_{\alpha} \frac{x}{z} + c_x, \underbrace{f\ell}_{\beta} \frac{y}{z} + c_y \right) \tag{3}$$

- $[k] = [\ell] = \frac{\text{pixel}}{m}$ measure pixels
- Note: k and ℓ may be different because the aspect ratio of the unit element is not guaranteed to be one
- If $k = \ell$, we say that the camera has **square pixels** (aspect ratio of 1)
- $[f] = m$ and $[\alpha] = [\beta] = \text{pixels}$

1.5.1 Is this Projective Transformation Linear?

$$P = (x, y, z) \longrightarrow P' = (\alpha\frac{x}{z} + c_x, \beta\frac{y}{z} + c_y) \tag{4}$$

1. Is the projection $P \to P'$ a linear transformation?

- No because the division of the input parameter z is not linear
- Can we express it with a matrix?
- Yes, use homogenous coordinates

1.6 Homogenous Coordinates

- Consider transformations between Euclidean $\mathbb{E}$ and Homogenous $\mathbb{H}$ coordinates

(a) $\mathbb{E} \longrightarrow \mathbb{H}$

$$\underbrace{(x, y) \longrightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}_{\text{homogenous image coordinates}} \qquad \underbrace{(x, y, z) \longrightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}}_{\text{homogenous scene coordinates}} \tag{5}$$

(b) $\mathbb{H} \longrightarrow \mathbb{E}$

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \longrightarrow \left(\frac{x}{w}, \frac{y}{w}\right) \qquad \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \longrightarrow \left(\frac{x}{w}, \frac{y}{w}, \frac{z}{w}\right) \tag{6}$$

- Note: equality between a vector and its homogenous coordinates only occurs when the final coordinate is 1

1.6.1 Projective Transformation in Homogenous Coordinate System

- Represent the projective transformation from eqn 3 in Homogenous Coordinates as follows

$$P'_h = \underbrace{\begin{bmatrix} \alpha x + c_x z \\ \beta y + c_y z \\ z \end{bmatrix}}_{\text{Homogenous Image Coords}} = \underbrace{\begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}}_{\text{Homogenous Scene Coords}} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} P_h \tag{7}$$

- Can then convert Homogenous image coordinates to Euclidean image coordinates by dividing by z

$$\underbrace{P'_h}_{Homogenous} \longrightarrow \underbrace{P' = (\alpha x + c_x z, \beta y + c_y z)}_{\text{Euclidean}} \tag{8}$$

1.7 Camera Matrix

- Note: from now on, assume we are using homogenous coordinates (unless stated otherwise)
- Drop the h index, so any point P or P' can be assumed to be in homogenous coordinates

- From eqn 7, can represent the relationship between a point in 3D space and its image coordinates by a vector relationship

$$P' = \begin{bmatrix} x' \\ y' \\ z \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} P = \mathbf{M}P \tag{9}$$

- Can then decompose this transformation into

$$P' = \mathbf{M}P = \begin{bmatrix} \alpha & 0 & c_x & 0 \\ 0 & \beta & c_y & 0 \\ \underbrace{0 \quad 0 \quad 1}_{\mathbf{K}} & & & 0 \end{bmatrix} P = \begin{bmatrix} \alpha & 0 & c_x \\ 0 & \beta & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3\times 3} & \mathbf{0} \end{bmatrix} P = K \begin{bmatrix} \mathbf{I}_{3\times 3} & \mathbf{0} \end{bmatrix} P \tag{10}$$

- In summary,

$$P = \mathbf{M}P \tag{11}$$

$$= K \begin{bmatrix} \mathbf{I}_{3\times 3} & \mathbf{0} \end{bmatrix} P \tag{12}$$

K camera matrix; contains some of the critical parameters that are useful to characterize a camera model

- The rows of the camera matrix can also be directly obtained from the projective transformation eqn expressing the image plane coordinates x', y' in terms of the scene coordinates

$$\mathbf{K} = \begin{bmatrix} x' \\ y' \\ z \end{bmatrix} = \begin{bmatrix} \alpha & 0 & c_x \\ 0 & \beta & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

- Note: currently missing 2 parameters from our formulation: **skewness** and **distortion**

1.8 Camera Skewness

- An image is called skewed when the camera coordinate system is skewed
  - Here the angle between the two axes are slightly larger or smaller than 90 degrees
- Most cameras have zero-skew, but some degree of skewness may occur because of sensor manufacturing errors
- New camera matrix accounting for skewness is

$$P' = \begin{bmatrix} \alpha & -\alpha\cot\theta & c_x & 0 \\ 0 & \frac{\beta}{\sin\theta} & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{14}$$

$$\mathbf{K} = \begin{bmatrix} x' \\ y' \\ z \end{bmatrix} = \begin{bmatrix} \alpha & -\alpha\cot\theta & c_x \\ 0 & \frac{\beta}{\sin\theta} & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{15}$$

- Ignoring distortion effects, the camera matrix **K** has 5 DOF: α and β for focal length, c_x and c_y for image plane offset, and θ for skewness

1.9 Canonical Projective Transformation

- $P' = \mathbf{M}P$
- i.e. $\mathbb{R}^4 \xrightarrow{H} \mathbb{R}^3$

$$P' = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{M}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{16}$$

$$P'_i = \begin{bmatrix} \frac{x}{z} \\ \frac{y}{z} \end{bmatrix} \tag{17}$$

1.10 World Reference System

- Mapping so far defined within Cartesian reference system
- Need to introduce an additional mapping from world reference system to camera reference system

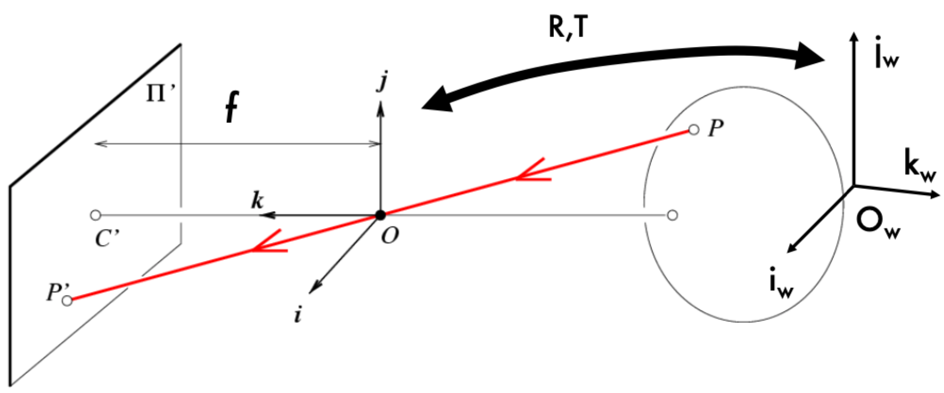

Figure 6: Mapping from world reference system (i_w, j_w, k_w) to camera reference system (i, j, k)

1.11 2D Transformations

1.11.1 2D Translation

- $P = (x, y)$
- $\mathbf{t} = (t_x, t_y)$
- $P' = P + \mathbf{t} = (x + t_x, y + t_y)$
- Translation requires using homogenous coordinates

$$P' \longrightarrow \begin{bmatrix} x+t_x \\ y+t_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{18}$$

$$= \begin{bmatrix} \mathbf{I} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{T} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{19}$$

1.11.2 2D Scaling

- $P = (x, y) \longrightarrow P' = (s_x x, s_y y)$
- $P = (x, y) \longrightarrow (x, y, 1)$

$$P' \longrightarrow \begin{bmatrix} s_x x \\ s_y y \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{S}} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{S}' & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{S} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{20}$$

1.11.3 2D Rotation

- CCW rotation by angle θ
- $x' = x\cos\theta - y\sin\theta$
- $y' = x\sin\theta\cos\theta$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{21}$$

- $P' = \mathbf{R}P$
- 1 DOF

$$P' = \longrightarrow \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{22}$$

1.11.4 2D Scale + Rotation + Translation

- $P' = \mathbf{TRS}P$

$$P' \longrightarrow \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{23}$$

$$= \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{24}$$

$$= \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{S} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{RS} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{25}$$

- Note: if $s_x = s_y$, this is a similarity transformation

1.12 3D Transformations

1.12.1 3D Rotation of Points

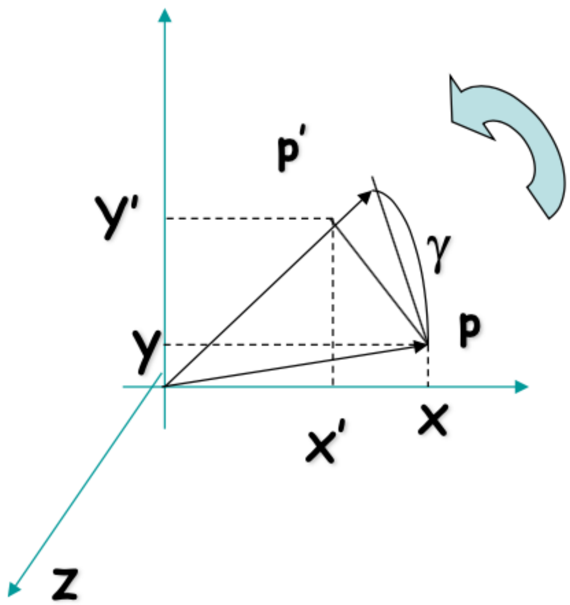

Figure 7: CCW rotation of γ around the z-axis using $R_z(\gamma)$.

- A rotation matrix in 3D has 3 DOF
- Rotation matrices for **CCW** rotation about the coordinate axes are:

$$R_x(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{bmatrix} \tag{26}$$

$$R_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \tag{27}$$

$$R_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{28}$$

$$P' \longrightarrow \begin{bmatrix} \mathbf{R} & 0 \\ 0 & 1 \end{bmatrix}_{4\times 4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{29}$$

1.12.2 3D Translation of Points

$$\mathbf{T} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{30}$$

$$P' \longrightarrow \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ 0 & 1 \end{bmatrix}_{4\times 4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{31}$$

- A translation vector in 3D has 3 DOF

1.12.3 3D Translation and Rotation

$$\mathbf{T} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{32}$$

$$\mathbf{R} = R_x(\alpha)R_y(\beta)R_z(\gamma) \tag{33}$$

$$P' = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix}_{4\times 4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{34}$$

1.13 World Reference System

- In 4D homogenous coordinates

$$P = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix}_{4\times 4} P_w \tag{35}$$

$$P_w = \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{36}$$

$$P' = \mathbf{K} \begin{bmatrix} \mathbf{I} & 0 \end{bmatrix} P = \mathbf{K} \begin{bmatrix} \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix}_{4\times 4} \tag{37}$$

K camera matrix (internal parameters)
M perspective projection matrix
R, T rotation and translation matrix (external parameters)

$$P_w = \underbrace{\mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{T}\end{bmatrix}}_{\mathbf{M}} P_w \tag{38}$$

1.14 More about Projective Transformations

$$P'_{3\times4} = \mathbf{M}_{3\times4} P_w = \mathbf{K}_{3\times3}\begin{bmatrix}\mathbf{R} & \mathbf{T}\end{bmatrix}_{3\times4} \tag{39}$$

- Total DOF: $5 + 3 + 3 = 11$

$$\mathbf{M} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \tag{40}$$

$$P'_{3\times1} = \mathbf{M} P_w \tag{41}$$

$$= \begin{bmatrix} m_1 P_w \\ m_2 P_w \\ m_3 P_w \end{bmatrix} \xrightarrow{E} \left(\frac{m_1 P_w}{m_3 P_w}, \frac{m_2 P_w}{m_3 P_w}\right) \tag{42}$$

1.14.1 Faugeras Theorem

$$\mathbf{M} = \mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{T}\end{bmatrix} = \begin{bmatrix}\mathbf{KR} & \mathbf{KT}\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{T}\end{bmatrix} \tag{43}$$

$$\mathbf{A} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \tag{44}$$

- A necessary and sufficient condition for a matrix **M** to be a **perspective projection matrix** is that $\det(\mathbf{A}) \neq 0$
- A necessary and sufficient condition for a matrix **M** to be a **zero-skew perspective projection matrix** is that $\det(\mathbf{A}) \neq 0$ **AND**

$$(a_1 \times a_3) \cdot (a_2 \times a_3) = 0 \tag{Mutually Orthogonal}$$

- A necessary and sufficient condition for a matrix **M** to be a **perspective projection matrix** with **zero-skew** and **unit aspect ratio** is that $\det(\mathbf{A}) \neq 0$ **AND**

$$\begin{cases} (a_1 \times a_3) \cdot (a_2 \times a_3) = 0 \\ (a_1 \times a_3) \cdot (a_1 \times a_3) = (a_2 \times a_3) \cdot (a_2 \times a_3) \end{cases} \tag{45}$$

1.14.2 Properties of Projective Transformations

- Points project to points
- Lines project to lines
- Distant objects look further
- Angles are **NOT** preserved
- Parallel lines **meet**

Vanishing point : point at which parallel lines in the world intersect in an image