A

EECS C106B / 206B Robotic Manipulation and Interaction **Spring 2022**

Lecture 6A: (Rigid Body Motion and Imaging Geometry)

Scribes: James McFarland and Jeremy Hughes

6.1 OUTLINE

- 3-D Euclidean Space and Rigid-Body Motion
  - Coordinates and coordinate frames
  - Rigid-body motion and homogeneous coordinates
- Geometric Models of Image Formation
  - Pinhole Camera Model
- Camera Intrinsic Parameters
  - From Metric to Pixel Coordinates
- Images of Basic Geometric Elements
  - Points, lines planes etc.

6.2 3-D Euclidean Space and Rigid Body motion

6.2.1 Euclidean Space

6.2.1.1 Recall Vectors in Space

1. The standard basis vectors,

$$e_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, e_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, e_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$$

   that define a coordinate in R^3 represented by $p = \begin{bmatrix}x\\y\\z\end{bmatrix}$

2. "free vector" defined by a pair of coordinates such that $v = p - q = \begin{bmatrix}X_1 - X_2\\Y_1 - Y_2\\Z_1 - Z_2\end{bmatrix}$ with magnitude and direction.

3. Vectors allow us to measure the important things in any dimensional space such as distance, length, angle

6.2.1.2 Inner Product

- Defined for vectors u and v: $\langle u, v\rangle := u^T v = u_1 v_1 + u_2 v_2 + u_3 v_3$
- L2 norm (length) of vector u: $||u|| := \sqrt{u^T u} = \sqrt{u_1^2 + u_2^2 + u_3^3}$
- Angle between vectors: $cos(\theta) = \frac{\langle u,v\rangle}{||u||||v||}$

6.2.1.3 Cross product

- Defined by $u \times v = \hat{u}v$
- Gives new vector orthogonal to both u and v
- Skew symmetric $\hat{u} := \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}$

6.2.2 Rigid-Body Motion

6.2.2.1 Rotation Matrix

Special Orthogonal group to define rotation matrices:

$$SO(3) = \left\{ R \in R^{3\times3} \mid R^T R = RR^T = I, \det R = 1 \right\}$$

Particularly this group allows for compositions of rotations that never leave the space of possible rotations and is thus invertible all the way back to an original starting point Rotating a coordinate p to coordinate q with matrix R has the relation $q = Rp$

Properties

1. Preserve inner product (vector length)

$$\|Rv\|_2^2 = v^T R^T Rv = \|v\|_2^2, \forall v \in R^3$$

2. Preserve cross product (orientation)

$$Ru \times Rv = R(u \times v)$$

3. The useful "hat" operator:

$$\widehat{Ru} = R\widehat{u}R^T \forall u \in R^3, R \in SO(3)$$

Following proof: $\widehat{Ru}Rv = R\widehat{u}v$ by property 2 implies $R^T\widehat{Ru}Rv = \widehat{u}v, \forall u, v \in R^3$ because $R^T R = I$. Note plugging in property 3 into $\widehat{Ru}$ on the LHS satisfies the equation.

6.2.2.2 Reflection Matrix

Main mathematical difference from rotation is determinant

$$R^T R = I, \det(R) = -1$$

Main intuitive/consequential difference is that reflections do not preserve orientation

$$Ru \times Rv = -R(u \times v)$$

Reflections belong to the orthogonal group:

$$O(3) = \left\{ R \in R^{3\times 3} \mid R^T R = I, \det(R) = \pm 1 \right\}$$

6.2.2.3 Homogenous Coordinates

When looking at the translation and rotation of an object, coordinates are related by: $\boldsymbol{X}_c = R\boldsymbol{X}_w + T$ and velocities are related by: $\dot{\boldsymbol{X}}_c = \widehat{\omega}\boldsymbol{X}_c + v$.

Remember for Homogenous coordinate append a 1 to position vectors and 0 to velocity vectors. Thus to express the above with homogenous coordinates (there was typo in the lecture slides for these matrices):

$$\text{X} = \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad V = \begin{bmatrix} \dot{X}_c \\ \dot{Y}_c \\ \dot{Z}_c \\ 0 \end{bmatrix} = \begin{bmatrix} \widehat{\omega} & v \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

6.3 Geometric Models of Image Formation

6.3.1 Pinhole Camera Model

We begin this section discussing the general practice of projecting a 3-D scene onto a 2-D plane (image). This leads us into a rather simple model known as the pinhole camera model. The general equaltions of a pinhole camera model are as follows.

$$X = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \longrightarrow x = \begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}$$

where X is the point within the 3-D image in camera frame coordinates and x is the 2-D point in camera frame coordinates. f is the focal length. Rewriting this equation in Homogenous Coordinates gives us

$$x = \longrightarrow \frac{1}{Z} \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} \quad where \quad X = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

It follows that

$$Z\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{K_f}\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Pi_0}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

6.4 Camera Intrinsic Parameters

6.4.1 From Metric to Pixel Coordinates

We are now faced with the challenge of converting our 2-D coordinates into pixel frame coordinates within our image. Letting $x^{'}$ represent our pixel coordinates:

$$\mathrm{x}^{'} = \begin{bmatrix} x^{'} \\ y^{'} \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Combing the above ideas of both the Pinhole Camera $\lambda x = K_f \Pi_0 \mathbf{X}$ and Pixel coordinates $x^{'} = K_s x$

$$\lambda x^{'} = K_s K_f \Pi_0 \mathbf{X} = \begin{bmatrix} f s_x & f s_\theta & O_x \\ 0 & f s_y & O_y \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Where $K = K_s K_f$
Our camera model then becomes:

$$\lambda x^{'} = K \Pi_0 \mathbf{X} = \mathbf{\Pi X}$$

6.5 Images of Basic Geometric Elements

6.5.1 Points, Lines, Planes etc

We now discuss Image formation with regards to a single point as well as a line From the above section, we learned that we can map a point from camera coordinates to pixel coordinates using the formula

$\lambda x^{'} = K\Pi_0\mathbf{X} = \mathbf{\Pi X}$\$

Representing a line is slightly more tricky. A homogenous representation of a line $\mathbf{L}$ follows the form

$$\mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix} + \mu \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ 0 \end{bmatrix}$$

And the homogenous representation of its 2-D image is

$$l = [\mathbf{a}, \mathbf{b}, \mathbf{c}]^{\mathbf{T}} \in \mathbf{R^3}$$

Therefore the projection of a 3-D line to an image plane is

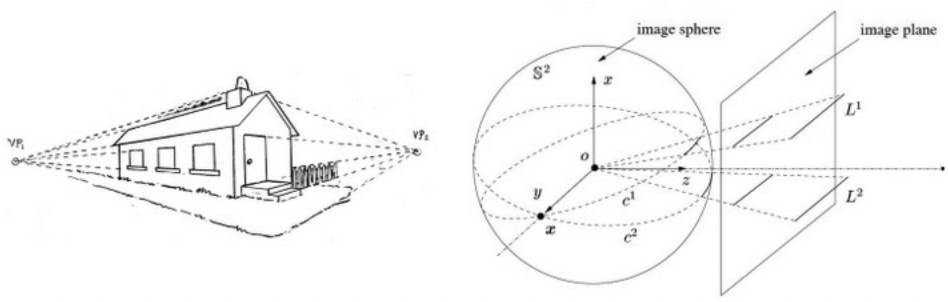
$$lx = l^T \Pi \mathbf{X} = \mathbf{0}$$

$$where \quad [KR, KT] \in R^{3\times 4}$$

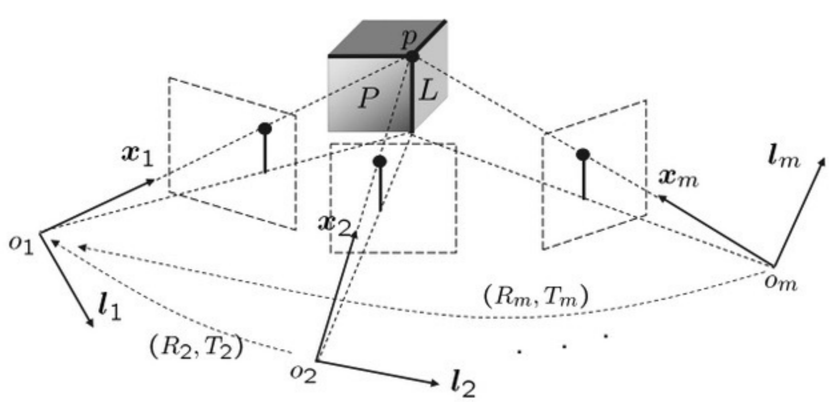
This idea of lines led into the topic of vanishing points. We discussed the idea that as the above value

$$\mu \to \infty$$

(or λ from the slides) we reach a point in which is considered a vanishing point



We briefly discussed 3-D reconstruction from multiple images. The following image was helpful in understanding this concept



1. Images are all "incident" at the corresponding features in space.
2. Features in space have many types of incidence relationships.
3. Features in space have many types of metric relationships.

The lecture concluded with an inspiring speech from Professor Ma as he discussed the many innovations yet to be made in the field of computer vision, as well as potential final project ideas.