ECE276A: Sensing & Estimation in Robotics
Lecture 8: Projective Geometry

Lecturer:

Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:

Siwei Guo: s9guo@eng.ucsd.edu

Anwesan Pal: a2pal@eng.ucsd.edu

UC San Diego

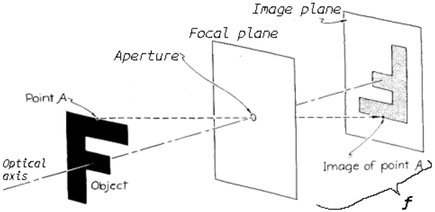
JACOBS SCHOOL OF ENGINEERING

Electrical and Computer Engineering

Image Formation

- **Image formation model**: must trade-off physical constraints and mathematical simplicity
- The values of an image depend on the shape and reflectance of the scene as well as the distribution of light
- **Image intensity/brightness/irradiance** $I(u, v)$ describes the energy falling onto a small patch of the imaging sensor (integrated both over the shutter interval and over a region of space) and is measured in power per unit area (W/m^2)
- A camera uses a set of lenses to control the direction of light propagation by means of *diffraction*, *refraction*, and *reflection*
- **Thin lens model**: a simple geometric model of image formation that considers only refraction
- **Pinhole model**: a thin lens model in which the lens aperture is decreased to zero and all rays are forced to go through the optical center and remain undeflected (diffraction becomes dominant).

Pinhole Camera Model



- **Focal plane**: perpendicular to the **optical axis** with a circular aperture at the **optical center**
- **Image plane**: parallel to the focal plane and a distance f (**focal length**) in **meters** from the optical center
- The pinhole camera model is described in an **optical frame** centered at the optical center with the optical axis as the z-axis:
  - optical frame: x = right, y = down, z = forward
  - world frame: x = forward, y = left, z = up
- **Ideal perspective projection**: relates the coordinates (X, Y, Z) of point A to its image coordinates (x, y) using similar triangles:

$$x = -f\frac{X}{Z} \qquad y = -f\frac{Y}{Z}$$

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \frac{1}{Z}\begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Pinhole Camera Model

- **Image flip**: the object appears upside down on the image plane. To eliminate this effect, we can simply flip the image $(x, y) \to (-x, -y)$, which corresponds to placing the image plane $\{z = -f\}$ in front of the optical center instead of behind $\{z = f\}$.
- **Field of view**: the angle subtended by the spatial extend of the image plane as seen from the optical center. If m is the side of the image plane in **meters**, then the field of view is $\boxed{\theta = 2\arctan\left(\frac{m}{2f}\right)}$.
  - For a flat image plane: $\theta < 180^\circ$.
  - For a spherical or ellipsoidal imaging surface, common in omnidirectional cameras, θ can exceed 180°.
- **Ray tracing**: under assumptions of the pinhole model and Lambertian surfaces, image formation can be reduced to tracing rays from points on objects to pixels. A mathematical model associating 3-D points in the world frame to 2-D points in the image frame must account for:
  1. **Extrinsics**: world-to-camera frame transformation
  2. **Projection**: 3D-to-2D coordinate projection
  3. **Intrinsics**: scaling and translation of the image coordinate frame

Extrinsics

- Let $p_{wc} \in \mathbb{R}^3$ and $R_{wc} \in SO(3)$ be the camera position and orientation in the world frame

- Rotation from a regular to an optical frame: $R_{oc} := \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}$

- Let (X_w, Y_w, Z_w) be the coordinates of point A in the world frame. The coordinates of A in the optical frame are then:

$$\begin{pmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{pmatrix} = \begin{bmatrix} R_{oc} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_{cw} & p_{cw} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \begin{bmatrix} R_{oc} R_{wc}^T & -R_{oc} R_{wc}^T p_{wc} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

Projection

- The 3D-to-2D projection in homogeneous coordinates from the optical frame to the image frame for a frontal pinhole camera model is:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \frac{1}{Z_o} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{pmatrix}$$

- The above can be decomposed into:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \underbrace{\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{image flip: } R_f} \underbrace{\begin{bmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{focal scaling: } K_f} \underbrace{\frac{1}{Z_o} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{canonical projection: } \Pi_0} \begin{pmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{pmatrix}$$

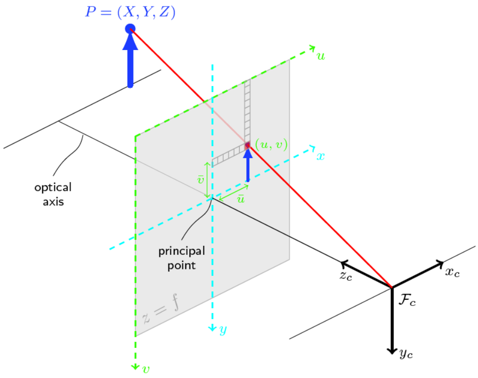
- The focal scaling K_f and image flip R_f are intrinsic parameters.

Intrinsics

- In practice, images are obtained in terms of pixels (u, v) with the origin of the pixel array typically in the upper-left corner of the image.
- The relationship between the image frame and the pixel array is specified via the following parameters:
  - (s_u, s_v) [pixels/meter]: define the **scaling** from meters to pixels and the **aspect ration** $\sigma = s_u/s_v$
  - (c_u, c_v) [pixels]: coordinates of the *principal point* used to translate the image frame origin, e.g., $(c_u, c_v) = (320.5, 240.5)$ for a 640×480 image
  - s_θ [pixels/meter]: **skew factor** that scales non-rectangular pixels and is proportional to $\cot(\alpha)$ where α is the angle between the coordinate axes of the pixel array.
- Normalized coordinates in the image frame are converted to pixel coordinates in the pixel array using the **intrinsic parameter matrix**:

$$\underbrace{\begin{bmatrix} s_u & s_\theta & c_u \\ 0 & s_v & c_v \\ 0 & 0 & 1 \end{bmatrix}}_{\text{pixel scaling: } K_s} \underbrace{\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{image flip: } R_f} \underbrace{\begin{bmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{focal scaling: } K_f} = \underbrace{\begin{bmatrix} fs_u & fs_\theta & c_u \\ 0 & fs_v & c_v \\ 0 & 0 & 1 \end{bmatrix}}_{\text{calibration matrix: } K} \in \mathbb{R}^{3\times 3}$$

Pinhole Camera Model



- **Extrinsics**:

$$\begin{pmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{pmatrix} = \begin{bmatrix} R_{oc} R_{wc}^T & -R_{oc} R_{wc}^T p_{wc} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

- **Projection and Intrinsics**:

$$\underbrace{\begin{pmatrix} u \\ v \\ 1 \end{pmatrix}}_{\text{pixels}} = \underbrace{\begin{bmatrix} fs_u & fs_\theta & c_u \\ 0 & fs_v & c_v \\ 0 & 0 & 1 \end{bmatrix}}_{\text{calibration: } K} \underbrace{\frac{1}{Z_o} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{canonical projection: } \Pi_0} \begin{pmatrix} X_o \\ Y_o \\ Z_o \\ 1 \end{pmatrix}$$

Projection Functions

- **Canonical projection function**: for a vector $a \in \mathbb{R}^3$, define $\pi(a) := \frac{1}{a_3}a$. Then, the pixel coordinates $y \in \mathbb{R}^2$ of a point $x \in \mathbb{R}^3$ in the world frame observed by a camera at position $p \in \mathbb{R}^3$ and orientation $R \in SO(3)$ with intrinsic parameters $K \in \mathbb{R}^{3\times 3}$ are:

$$y = K\pi(R_{oc}R^T(x - p))$$

- **Spherical perspective projection**: if the imaging surface is a sphere $\mathbb{S}^2 := \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$ (motivated by retina shapes in biological systems), we can define a spherical projection $\pi_s(x) = \frac{x}{\|x\|_2}$. Similar to the planar perspective projection, the relationship between pixel coordinates y of a point and their 3-D metric counterpart x is:

$$y = K\pi_s(R_{oc}R^T(x - p))$$

- **Catadioptric model**: uses an ellipsoidal imaging surface

Radial distortion

- **Wide field of view camera**: in addition to linear distortions described by the intrinsic parameters K, one can observe distortion along radial directions.
- The simplest effective **model for radial distortion**:

$$x = x_d(1 + a_1 r^2 + a_2 r^4)$$
$$y = y_d(1 + a_1 r^2 + a_2 r^4)$$

where (x_d, y_d) are the coordinates of distorted points and $r^2 = x_d^2 + y_d^2$ and a_1, a_2 are additional parameters modeling the amount of distortion.

Panorama

- **Input**: image I and camera-to-world orientation R
- Suppose the image lies on a sphere and get the world coordinates of each pixel:
  1. Find longitude (λ) and latitude (ϕ) of each pixel using the number of rows and columns and the horizontal (60°) and vertical (45°) fields of view
  2. Convert Spherical $(\lambda, \phi, 1)$ to Cartesian assuming depth 1
  3. Rotate the Cartesian coordinates to the world frame using R
- Project world pixel coordinates to a cylinder and unwrap:
  1. Convert Cartesian to Spherical
  2. Inscribe the sphere in a cylinder so that a point $(\lambda, \phi, 1)$ on the sphere has height ϕ on the cylinder and longitude λ along the cylinder circumference
  3. Unwrap the cylinder surface to a rectangular image with width 2π radians and height π radians
  4. Different options for sphere to plane projection: equidistant, equal area, Miller, etc.
     (see `https://en.wikipedia.org/wiki/List_of_map_projections`)

Panorama