Lecture 1, Sep 2, 2025

Introduction to Robotic Vision

- Many vision problems are *ill-posed*, i.e. a solution may not exist, may not be unique, or may not be continuous with regard to change of the initial conditions
- Vision is an *inverse problem*
  - Much information is lost, so we need to appeal to probabilistic models
  - Most of our physical models are forward models, i.e. how should a particular object look under certain conditions?
  - The inverse is non-unique, e.g. a bigger object far away appears the same as a smaller, closer object
    * This is known as *scale ambiguity*
  - By projecting 3D geometry to a 2D image we lose a lot of information – appearance only weakly depends on geometry

Definition

The *imaging function* (i.e. *perspective projection*):

$$I = \mathscr{P}(G, M, V, L, A, \epsilon)$$

where:

- G: Scene geometry (shape of the world)
- M: Materials
- V: Viewpoint (where the camera is located)
- L: Lighting (where light sources are)
- A: Atmospherics (how light interacts with the atmosphere)
- ϵ: Noise

We aim to find $\mathscr{P}^{-1}$ to recover G, M, V from I.

- Parallel lines converge at "points at infinity" under perspective projection, and shapes are distorted
- Datasets typically contain synchronized, timestamped data from GPS/INS, LiDAR, stereo, etc. and accurate intrinsic/extrinsic calibration of sensors
  - Popular datasets include KITTI, Waymo Open (100x larger than KITTI), Nuscenes (object detection and tracking)