Problem 1. Estimate Camera Parameters

Straight roads and cubic buildings are characteristic of the Cornell campus. While walking around, you may notice that the parallel lines formed by these structures carry useful geometric information about the scene. Using your mobile phone, take a photo in a location where such cues are visible. From this photo, determine the focal length and the height of the camera. Provide a clear derivation of your results.

Problem 2. 360° Images

Background. A *360° image* (a.k.a. panorama) captures all viewing directions from a single point on the viewing sphere. A common raster representation is the *equirectangular map (ERM)*, which linearly samples longitude and latitude:

$$(u, v) \in [0, W) \times [0, H), \quad \lambda(u) = 2\pi \frac{u}{W} - \pi, \quad \phi(v) = \frac{\pi}{2} - \pi \frac{v}{H},$$

where λ is longitude (yaw) and ϕ is latitude (elevation).

Figure 1: Equirectangular panorama of Goldwin Smith Hall (link to the image).

Rotation convention. Intuitively, yaw is rotating your head left/right, pitch is up/down, and roll is tilt; in our convention, positive yaw is right, positive pitch is up, and positive roll is tilting left (counterclockwise). We use rotation in the order *roll–pitch–yaw* (RPY).

(1) Perspective view extraction. Extract a perspective view with $(\mathrm{roll}, \mathrm{pitch}, \mathrm{yaw}) = (25°, 10°, 30°)$ and square field of view fov $= 90°$ from the given ERM image.

(2) Cube map orientations. Besides ERM, a 360° image can be represented as a *cube map* of six square faces (each a 90° FOV pinhole view) pointing along the $\pm x$, $\pm y$, $\pm z$ axes. Suppose the front direction is the image you derived from (1), give one valid set of $(\mathrm{roll}, \mathrm{pitch}, \mathrm{yaw})$ for the five faces other the front image.

Note. Depending on your target convention or library, you may add $\pm 90°$ roll to the Up/Down faces to keep "up" visually consistent across faces; any self-consistent set earns full credit.

(3) Derive the sampling Jacobian. Show that the solid-angle element on the sphere maps to ERM area by $\mathrm{d}\Omega = \cos\phi \, \mathrm{d}\lambda \, \mathrm{d}\phi$ and discuss one implication for importance sampling or antialiasing near the poles.

Problem 3. Real-world aligned 3D Reconstruction

Reconstructing static 3D scenes from multiview images is a long-established challenge in 3D vision. In this task, you will recover a 3D mesh from a video that captures the scene from multiple viewpoints. Unlike ordinary reconstructions, however, we require the resulting mesh to align with real-world scale and coordinates. To achieve this, we rely on chessboard calibration, illustrated in Figure 2. Each square on the chessboard has a scale of 0.01 meters, and the target coordinate system is indicated below. The input images are provided for you.

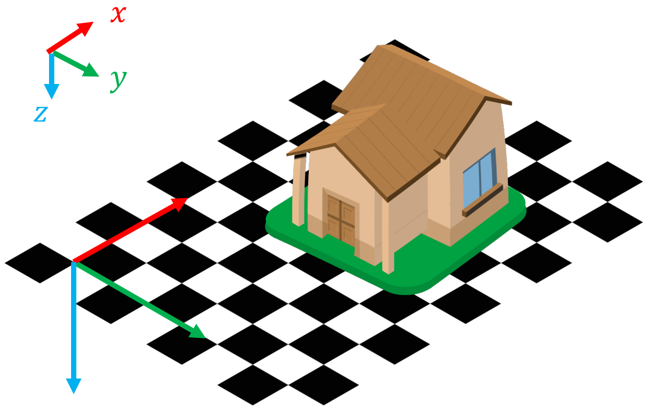

Figure 2: Coordinate system defined by the chessboard.

The suggested pipeline:

(1) Structure from motion (SfM). Use structure from motion methods (e.g., COLMAP) to estimate the relative poses of the cameras.

(2) Chessboard calibration. Use the chessboard to calibrate the absolute camera poses of some frames where the chessboard is visible.

(3) Align the coordinate system. Transform the camera poses estimated from SfM to the real-world coordinate system calibrated by the chessboard.

(4) Run reconstruction method. Reconstruct the 3D scene with the input images and camera poses using reconstruction methods (e.g., Gaussian Splatting or NeRF).

(5) Convert the representation into a mesh. Using the existing methods to convert the optimized representation into a mesh.