1 Projecting Aerial LiDAR Points onto Color Images for Depth Map Generation

This document outlines the process of aligning aerial LiDAR data with a SLAM-based point cloud, transforming the points into the camera coordinate system, and projecting them onto the image plane to generate a depth map.

1.1 Understanding the Coordinate Frames

We deal with multiple coordinate frames:

- **Aerial LiDAR frame** ($\mathcal{F}_{\text{aerial}}$): The coordinate system in which aerial LiDAR points are.
- **SLAM frame** ($\mathcal{F}_{\text{SLAM}}$): The coordinate system of the global SLAM point cloud.
- **LiDAR frame** ($\mathcal{F}_{\text{LiDAR}}$): The local frame of the LiDAR sensor.
- **Camera frame** ($\mathcal{F}_{\text{cam}}$): The local coordinate frame of the color camera.
- **Image plane** ($\mathcal{F}_{\text{image}}$): The 2D pixel space of the color image.

The transformation chain follows this sequence:

$$\mathcal{F}_{\text{aerial}} \rightarrow \mathcal{F}_{\text{SLAM}} \rightarrow \mathcal{F}_{\text{LiDAR}} \rightarrow \mathcal{F}_{\text{cam}} \rightarrow \mathcal{F}_{\text{image}}$$

1.2 Aligning Aerial LiDAR Points with the SLAM Point Cloud

First, the aerial point cloud is aligned with the SLAM point cloud.

$$\mathbf{P}_{\text{SLAM}} = \mathbf{R}_{\text{align}}\mathbf{P}_{\text{aerial}} + \mathbf{t}_{\text{align}}$$

where $\mathbf{P}_{\text{aerial}}$ is a 3D point from the aerial scan, $\mathbf{P}_{\text{SLAM}}$ is the corresponding point in the SLAM coordinate system, and $\mathbf{R}_{\text{align}}, \mathbf{t}_{\text{align}}$ are the rotation and translation aligning the aerial points with the SLAM map. Rotation and translation can be obtained via manual or GPS-based alignment. If manual aligned is used, the ICP algorithm can be applied to refine the alignment.

1.3 Transforming Points to the Camera Coordinate System

Once the aerial points are in the SLAM coordinate system, we need to transform them into the camera coordinate system for projection. This is done in two steps:

- **Step 1: Transforming from SLAM Frame to LiDAR Frame:**

  $$\mathbf{P}_{\mathrm{LiDAR}} = \mathbf{R}_i \mathbf{P}_{\mathrm{SLAM}} + \mathbf{t}_i$$

  where $\mathbf{R}_i$ and $\mathbf{t}_i$ correspond to the rotation and translation estimated by the SLAM system for the frame i.

- **Step 2: Transforming from LiDAR Frame to Camera Frame**.

  If the LiDAR and color camera have different coordinate frames, we need to apply an additional transformation:

  $$\mathbf{P}_{\mathrm{cam}} = \mathbf{R}_{\mathrm{LiDAR}\to\mathrm{cam}} \mathbf{P}_{\mathrm{LiDAR}} + \mathbf{t}_{\mathrm{LiDAR}\to\mathrm{cam}}$$

  where $\mathbf{R}_{\mathrm{LiDAR}\to\mathrm{cam}}, \mathbf{t}_{\mathrm{LiDAR}\to\mathrm{cam}}$ are the extrinsic parameters that transform points from the LiDAR frame to the camera frame. If the poses provided by the SLAM system are already in the camera frame, then $\mathbf{R}_{\mathrm{LiDAR}\to\mathrm{cam}} = \mathbf{I}$, and $\mathbf{t}_{\mathrm{LiDAR}\to\mathrm{cam}} = \mathbf{0}$.

1.4 Final Homogeneous Representation

Combining these transformations into a single 4×4 homogeneous transformation matrix:

$$\mathbf{P}^h_{\mathrm{cam}} = \mathbf{T}_{\mathrm{LiDAR}\to\mathrm{cam}} \mathbf{T}_i \mathbf{T}_{\mathrm{align}} \mathbf{P}^h_{\mathrm{aerial}}$$

where each transformation matrix is:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$$

and the homogeneous coordinates of a point are:

$$\mathbf{P}^h = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}.$$

1.5 Projecting 3D Points onto the Image Plane

Once the points are transformed into the camera coordinate system, we project them onto the 2D image plane using the intrinsic camera matrix $\mathbf{K}$:

$$\mathbf{p} = \mathbf{K}\mathbf{P}_{\mathrm{cam}}$$

where:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

Expanding the projection:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_{\text{cam}} \\ Y_{\text{cam}} \\ Z_{\text{cam}} \end{bmatrix}$$

The final pixel coordinates are:

$$u = \frac{f_x X_{\text{cam}}}{Z_{\text{cam}}} + c_x, \quad v = \frac{f_y Y_{\text{cam}}}{Z_{\text{cam}}} + c_y$$

where (u, v) are the image pixel coordinates and Z_{cam} is the depth value.

1.6 Constructing the Depth Map

To construct a depth map:

- Initialize a blank image where each pixel stores the depth Z_{cam}.
- For each projected point, assign Z_{cam} to (u, v).
- Handle occlusions: If multiple points project to the same pixel, keep the closest one:

$$D(u, v) = \min(Z_{\text{cam}})$$

1.7 Alternative Method for Depth Construction: Raycasting with a Mesh

Instead of projecting 3D points, an alternative is to:

1. Convert the aerial point cloud into a 3D mesh using, for example, Poisson surface reconstruction.
2. Cast rays from the camera through each pixel given the camera pose and intrinsics.
3. Determine the first intersection with the mesh to obtain depth.

This method handles occlusions better but obtaining a good mesh is not trivial.