SiLVR: Scalable Lidar-Visual NeRF with Uncertainty Quantification

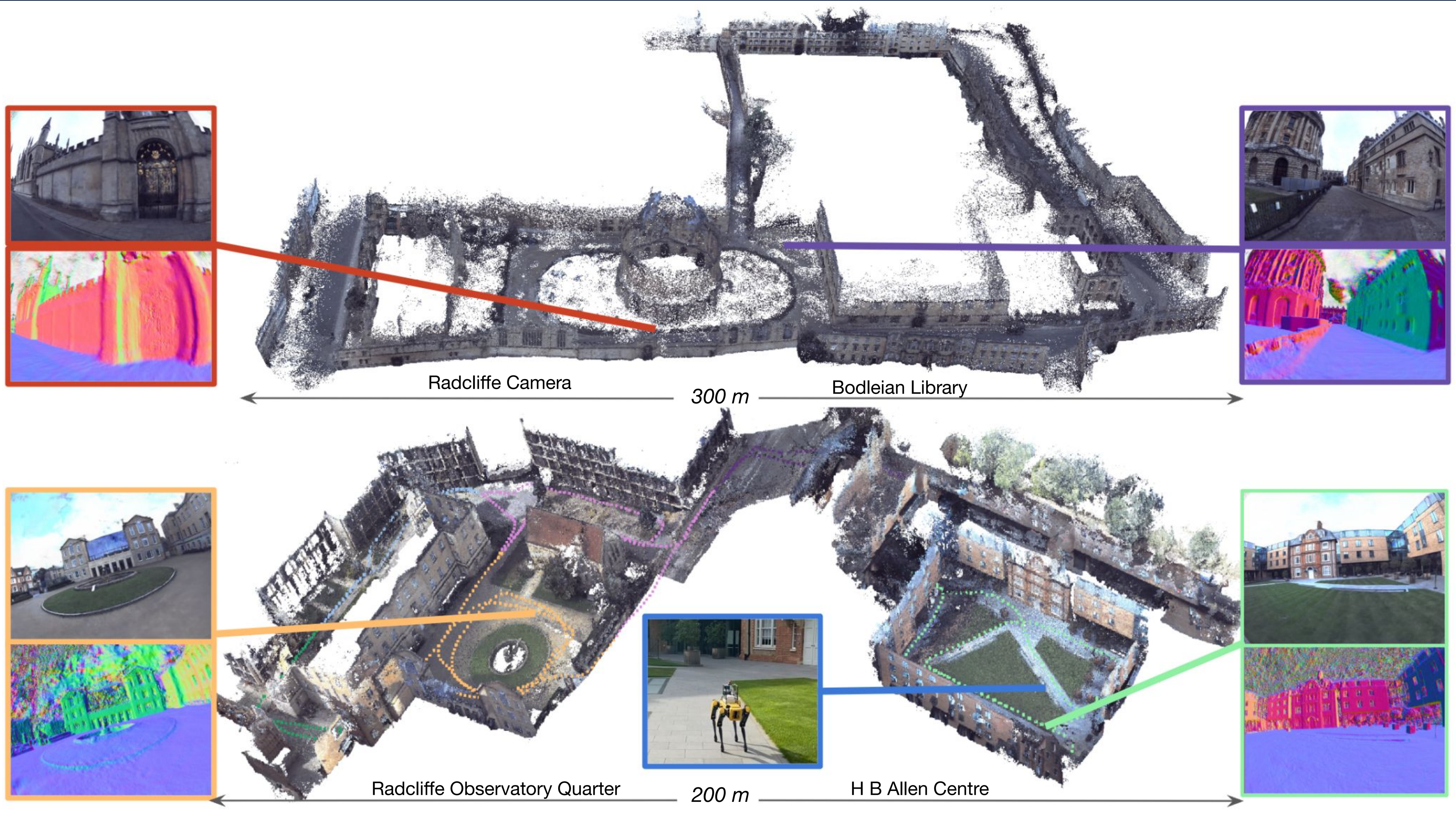
Yifu Tao, Maurice Fallon

https://dynamic.robots.ox.ac.uk/projects/silvr/
{yifu,mfallon}@robots.ox.ac.uk

UNIVERSITY OF OXFORD
OXFORD ROBOTICS INSTITUTE
DYNAMIC ROBOT SYSTEMS GROUP

Overview

- **Lidar-Visual NeRF** for 3D textured mapping
- Decoupled **epistemic uncertainty** estimation based on Fisher Information
- **Submapping** strategy based on visibility
- Pose estimation with online **SLAM** and offline Bundle Adjustment (**BA**).
- Lidar-visual 3D reconstruction dataset of sites over *100,000* m^2 with precise ground truth map for both SLAM and NeRF.

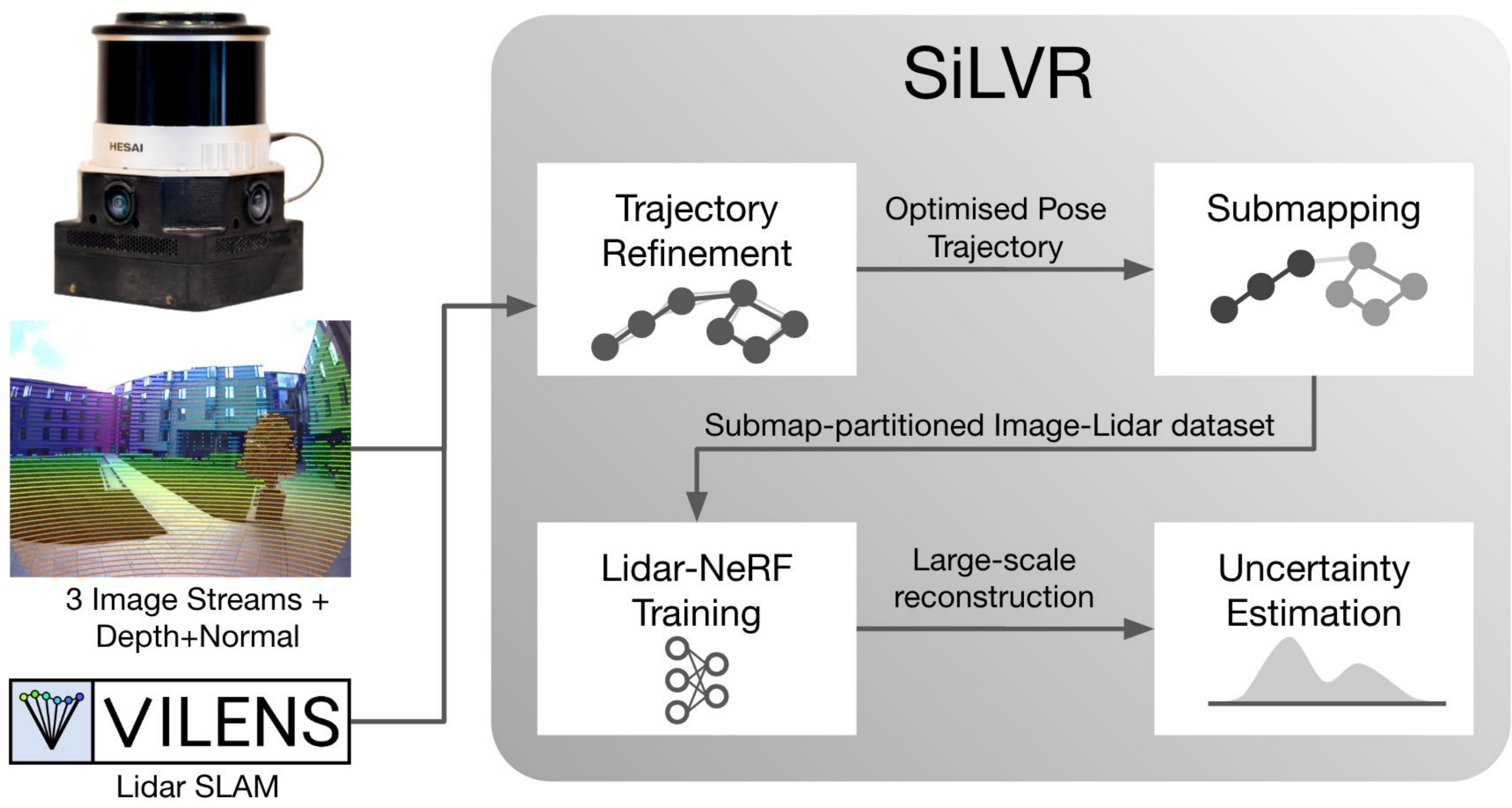


Hardware and System

Sensor payload: Three Sevensense Alphasense fisheye cameras, IMU and Hesai Pandar QT64 Lidar

Pipeline: Camera images, Lidar depth and normal images are used to train the NeRF, with poses from lidar SLAM refined by COLMAP and partitioned into submaps. Uncertainty is estimated after training.

Tested with a handheld device, legged robot and a drone



Multi-Modal Epistemic Uncertainty

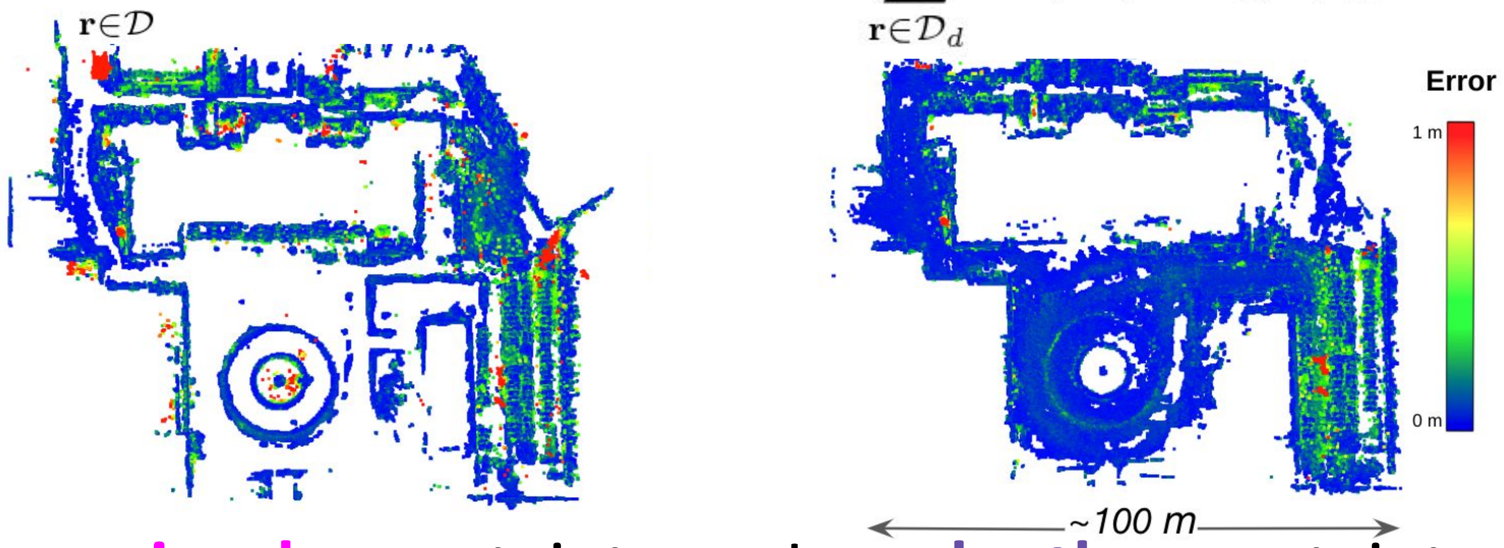
We estimate the model uncertainty by calculating the **spatial covariance** of each point location.

We use the **Laplace approximation** and **Fisher Information matrix** to approximate the covariance.

The Jacobian matrix can be decomposed. This enables us to identify uncertainty due to **vision** (e.g. textureless area) and **lidar depth** (e.g. outside lidar FoV or range).

$$\mathbf{J} = -\nabla_\theta \ell(x_n, y_n; \theta)$$

$$\sum_{\mathbf{r}\in\mathcal{D}} ||\hat{\mathbf{C}}(\mathbf{r}) - \mathbf{C}(\mathbf{r})||^2 \qquad \sum_{\mathbf{r}\in\mathcal{D}_d} \mathrm{KL}[\mathcal{N}(\mathbf{D}, \hat{\sigma})||h(t)]$$



Low **visual** uncertainty   Low **depth** uncertainty

Lidar Losses for Geometric Constraints

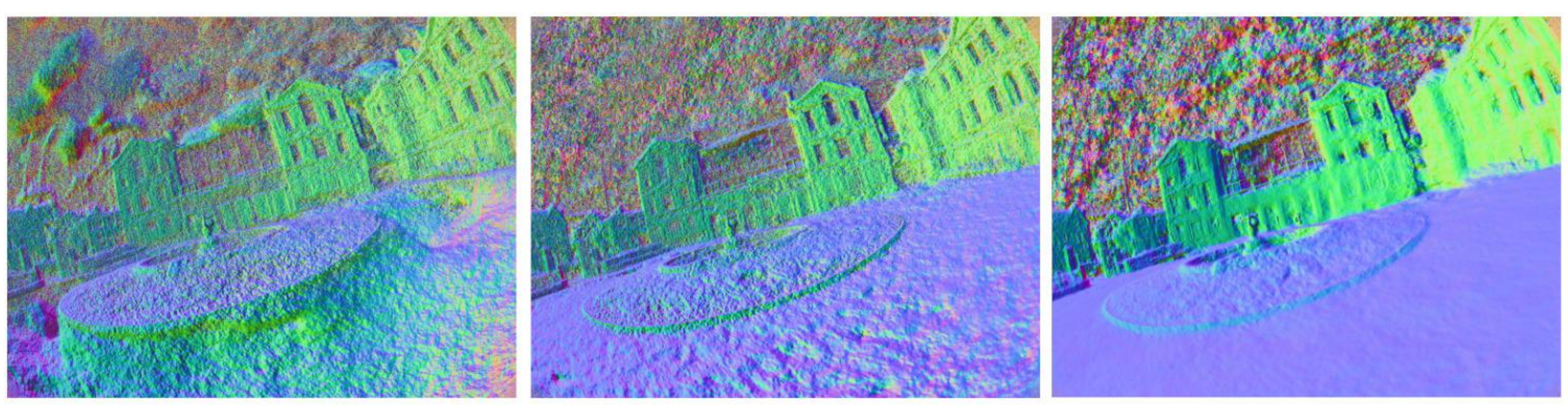
Motivation: Vision-only reconstruction methods (including NeRF and 3D Gaussian Splatting) struggle with textureless areas and limited multi-view constraints.

Solution 1: Depth Supervision
The weight distribution along the ray is optimised to follow a narrow Gaussian which encourages a surface at lidar depth.

$$\mathcal{L}_{Depth} = \sum KL[N(d, \sigma)||w(t)]$$

Solution 2: Surface Normal Supervision
The NeRF surface normal is regularised by the surface normals estimated from the lidar cloud.

$$\mathcal{L}_{Normal} = \sum ||\hat{N(r)} - N(r)||_1 + ||1 - \hat{N(r)}N(r)||_1$$



Nerfacto | **Ours - Depth Loss** | **Ours - Depth + Normal**

Oxford Spires Dataset: Radiance Field + SLAM

We curate a large-scale dataset of Oxford landmarks with multiple cameras, lidar and inertial data as well as **TLS ground truth map and trajectory**.

This enables us to evaluate visual and lidar **SLAM**, **SfM**, **MVS**, and radiance field methods (**NeRF** and **3D Gaussian Splatting**), all in one dataset.

Accepted for IJRR 2025
https://dynamic.robots.ox.ac.uk/datasets/oxford-spires/