Team 3

Range Finder Disciplined Monocular Depth Estimation

Penn Engineering

Project Objectives

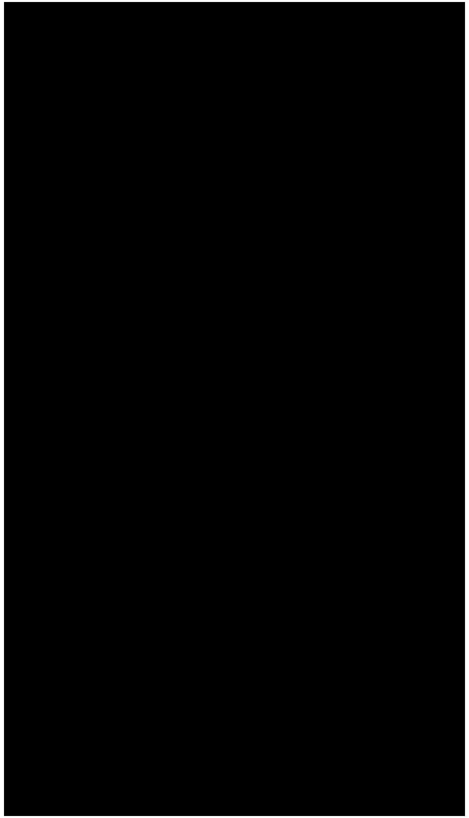
- Lidar is one of the most expensive components on the F1TENTH car
- Develop a lower-cost alternative with usable depth accuracy

- Monocular depth estimation models:
  - Excellent azimuth/elevation accuracy
  - Returns relative depths
- Low-cost laser rangefinders
  - Excellent depth accuracy
  - Azimuth/elevation precision limited to sensor FOV

- Combine strengths of Depth Estimation models and low-cost laser rangefinders through sensor fusion

Summary

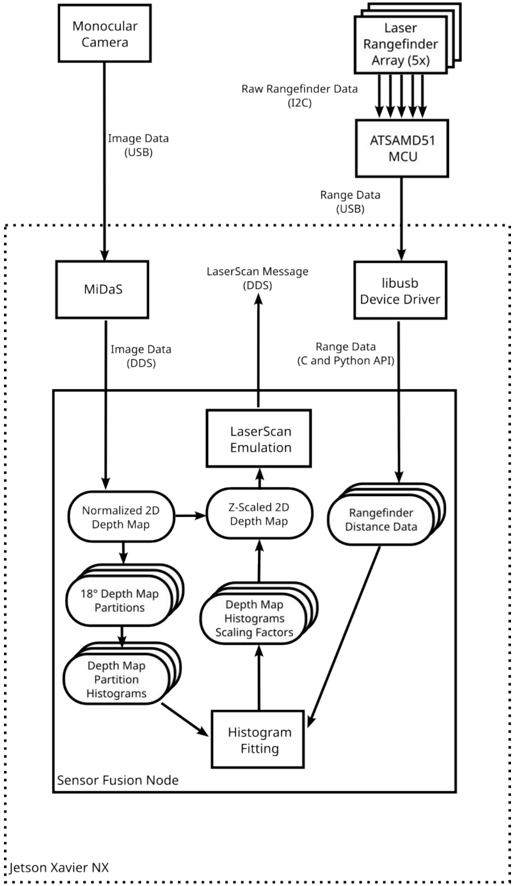
- Able to execute follow the gap using vision-based inputs
- Depth Estimation model performed reasonably well
- Sensor fusion outputs currently too "jumpy" to use reliably
  - Can show some degree of absolute depth
  - Would need more integration time to refine implementation



Implementation

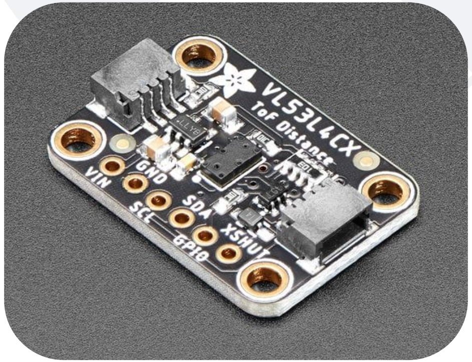
Sensor Design

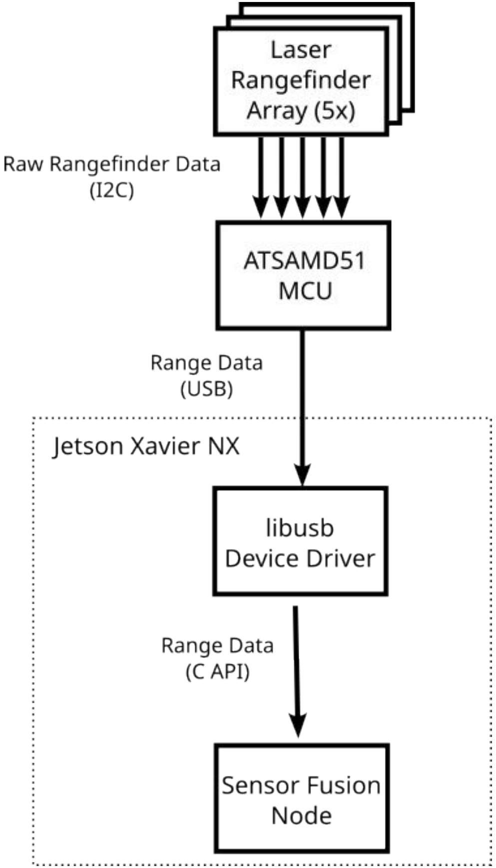
- Receive data from TOF sensor array and camera
- Process camera data with MiDaS
- Provide range measurements via USB
- Fuse MiDaS depth map with direct range measurements
- Emulate a ROS LaserScan message by taking a slice through the fused depth map
- Execute follow-the-gap based on emulated LaserScan



VL53L4CX Time-of-Flight Sensor Array

- VL53L4CX: low cost and excellent depth accuracy, but fixed 18-degree FOV
- Array of rangefinders to cover full FOV of the camera
- Sensor –I2C–> SAMD51 –USB–> libusb –> …
- 3D-printed mount at fixed 18-degree increments



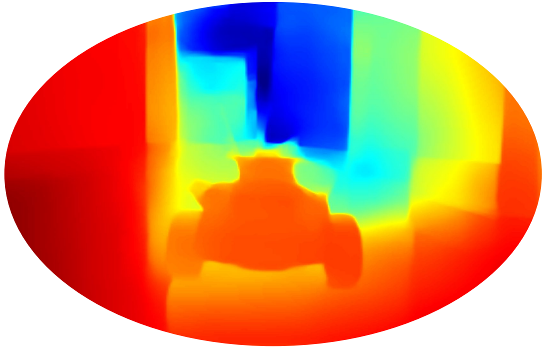


Monocular Depth Estimation

Given a Single RGB image, Predict depths of each pixel. Each pixel has a class (0-1) after normalization

Why use Monocular Depths ?

- More Convenient and Cost-effective.
- Stereo depth estimation relies on triangulating points between the two cameras, which can lead to lower resolution in areas with fewer features or where the two views are not well-aligned.
- Monocular depth models can be trained on a wider range of images making it more robust to changes in lighting conditions and other factors.



Adopted Approaches

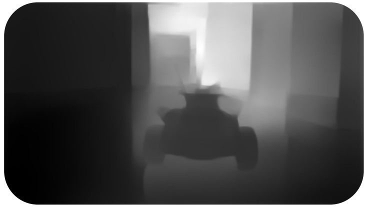
Custom U-Net Model

Input Size: (256,256)

Bad Accuracy !
High Temporal inconsistency

Inference:
Xavier NX: Not Tested

UNet - DensNet-201 Backbone



Input Size: (640,480)

Better Accuracy,
Computationally Expensive !

Inference:
Xavier NX: 4 FPS

MiDaS v2.1 Small

Input Size: (256,256)

Moderate Accuracy,
Light weight & Efficient!

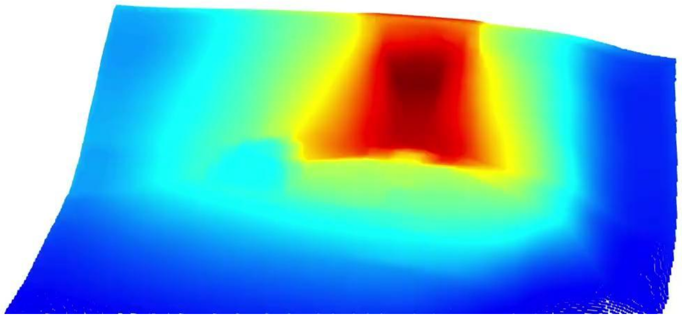
Inference:
Xavier NX: 40 FPS

Transfer Learning -
MiDas v21 small

Built the Levine Dataset with Ground Truth Depth Maps from Real-sense camera for Transfer Learning

Poor Ground truth depths for re-training :(

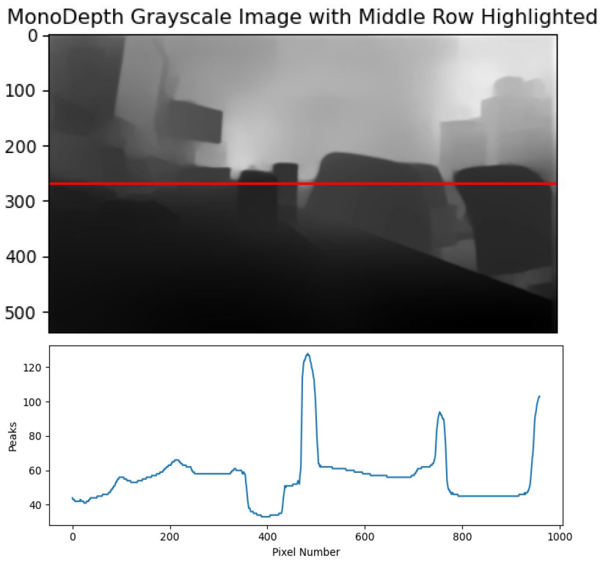
Results



Penn Engineering

Sensor Fusion

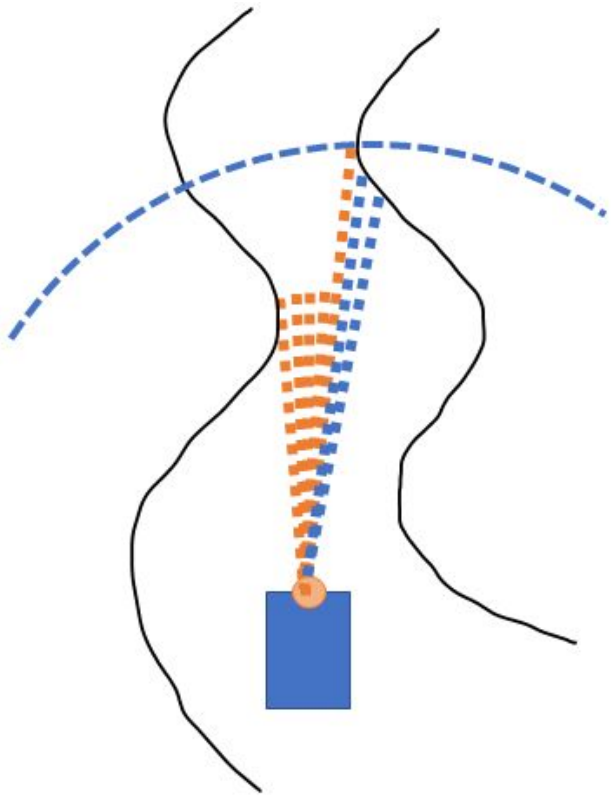
- Step 1 - Correlating Intensity Peaks and Linear Regression
  - Depth Map Overlap
  - Histogram Analysis
  - Obtaining scaling factors and biases for aligning depth map predictions with absolute inverse depth
- Step 2 - Adjusting Depth Values and Sensor Fusion:
  - Adjusting Depth Values
  - Sensor Fusion Algorithm



Penn Engineering

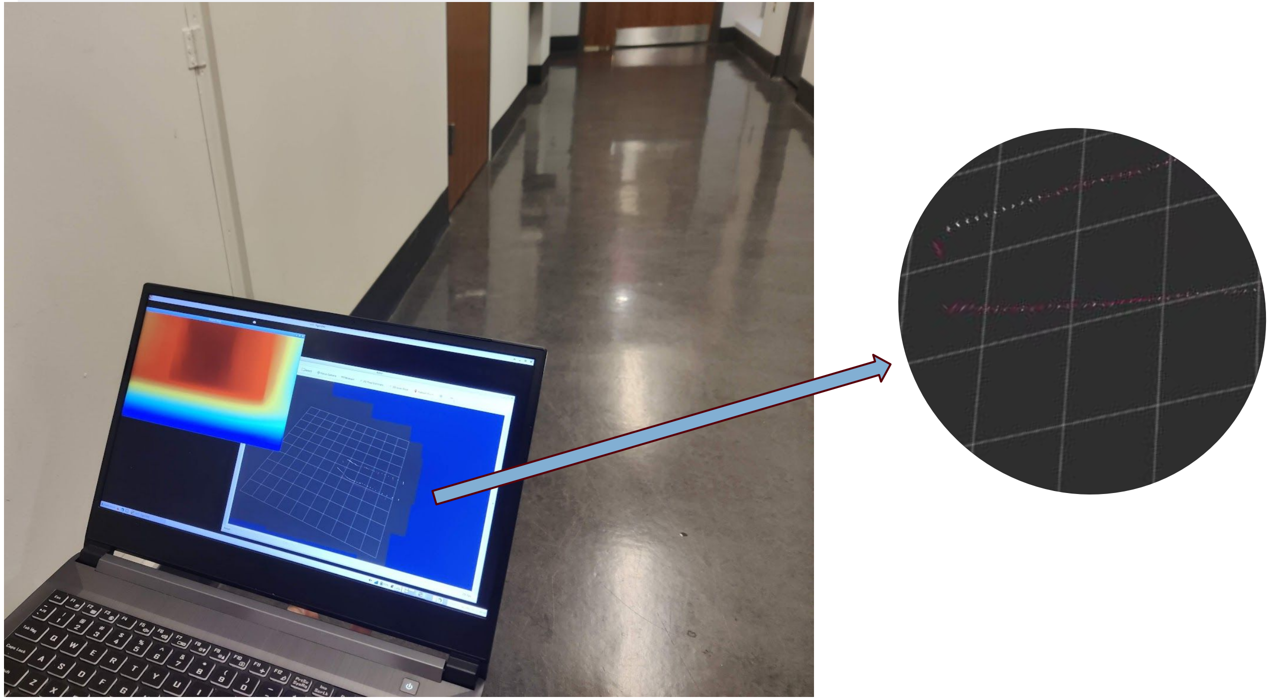
Gap Following

- Replicate 2D lidar data by taking horizontal slice of sensor fusion data
- Input pseudo-lidar data into disparity extension and gap follow algorithm from Lab 4
- Camera FOV of 87° pairs well with gap following in corridors
- Were unsure if disparity extension would work with sensor fusion data, but initial results were promising



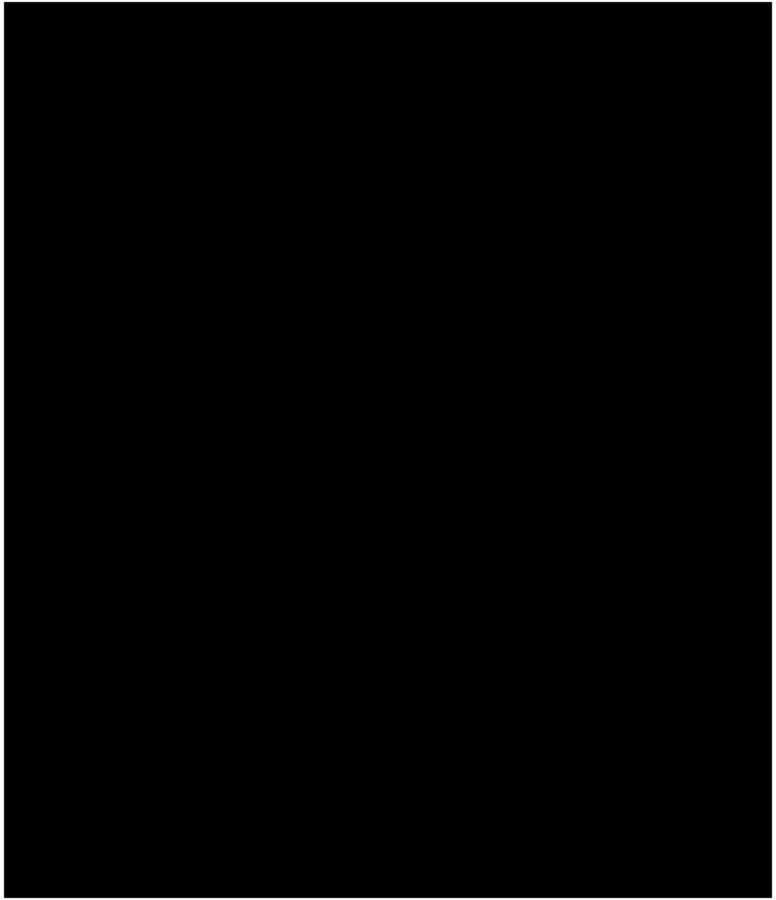
Results !

Integrated Follow the Gap



Monocular Depths from MiDaS v2.1

- Somewhat noisy
- “Likes” hands and faces
- Vulnerability to reflections, illusions
- Temporal Inconsistency due to being trained on non consecutive images



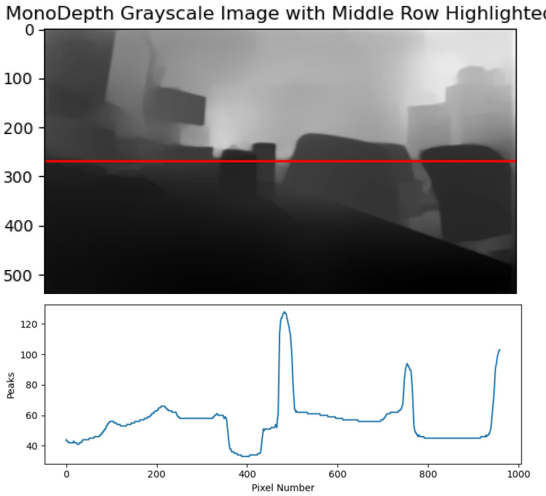
Time-of-Flight Sensor Performance

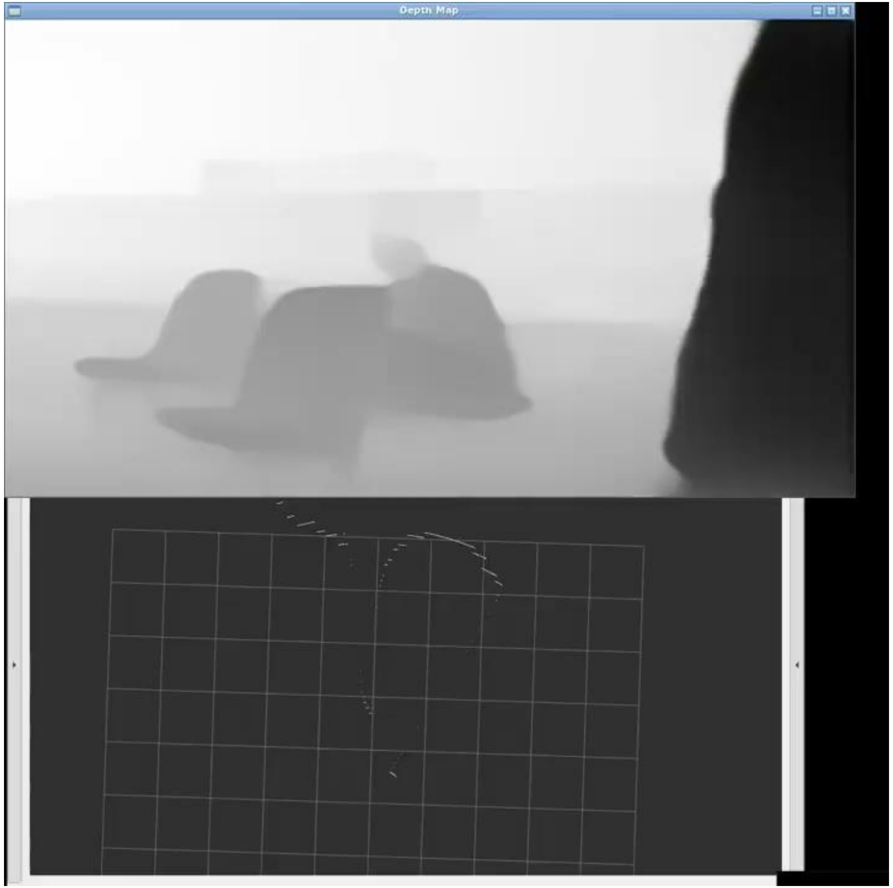
- Outputs provided by VL53L4CX were somewhat less "rich" than expected
  - Maximum range limited to about 3.5 meters instead of 6 meters
    - Tunable parameters have tradeoff between max range and latency
  - Each sensor typically returns only a single depth, occasionally two or three
    - Fewer sensor fusion inputs

- F1TENTH car consistently induces failures of I2C interfaces
  - I2C bus locks up when motor activates
  - Unshielded cables of several cm in length between SAMD51 and VL53L4CX breakout board
  - Motors are noisy; likely an EMI problem
  - Limited ability to test fused sensor data on functional platform

Sensor Fusion

- Histogram with smallest point
- Regression approach





Future Work

- Refine association logic in sensor fusion; Possible approaches:
  - Find association producing smallest change in scale
  - Retrain model with rangefinder inputs

- Build Better Dataset for Maps like Levine with a better Depth Camera
- Transfer Learning on Custom Dataset of sequential images
  for higher accuracy, reducing temporal inconsistency.
- Try better models like DPT-Hybrid, Swin v2.
- Point Cloud Segmentation for Localization based Approaches.

- Eliminate cables for rangefinder array I2C bus; a flex PCB design could potentially:
  - Reduce EMI problems
  - Be more compact mechanically
  - Reduce overall sensor cost ($5 IC vs $15 breakout board)

Bye-Bye to Lidar
Hopefully :)