Mizzou INformation and Data FUsion Lab (MINDFUL)

Title: Capturing Uncertainty in Monocular Depth Estimation: Towards Fuzzy Voxel Maps

Authors: Andrew R. Buck, Derek T. Anderson, Raub Camaioni, James M. Keller, Robert H. Luke III, and Jack Akers

University of Missouri

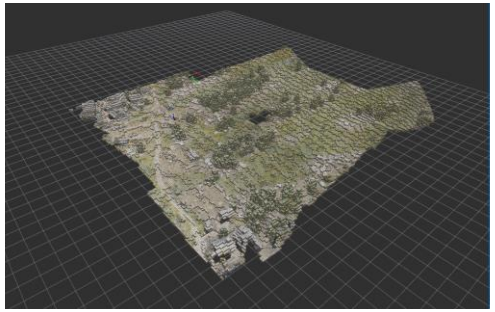
Department of Electrical Engineering and Computer Science

August 15th, 2023

Introduction

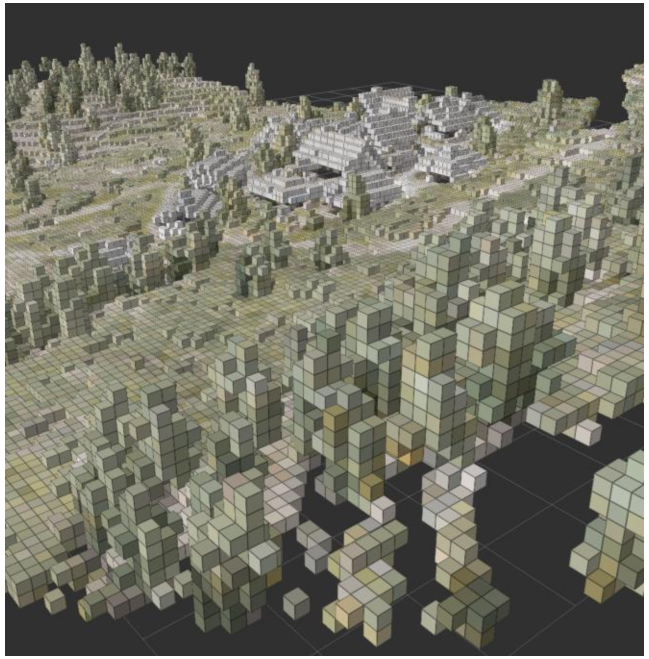
- **Consider this outdoor scene:**
  - We want to fly a UAV here and need a 3D model of the environment.
  - Our only sensors are a monocular camera and a GPS/IMU.
  - This model will be used for navigation and collision avoidance.
  - Using simulation here (with ground truth)

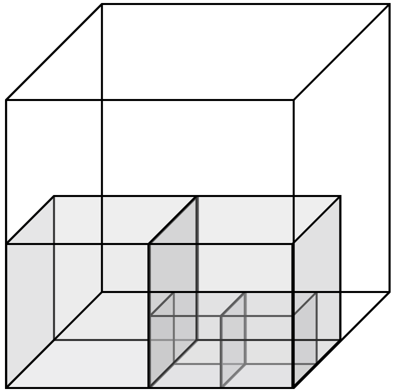
- **Questions:**
  - How to represent the map?
  - How to construct it?
  - How to handle uncertainty?



OctoMap and UFOMap

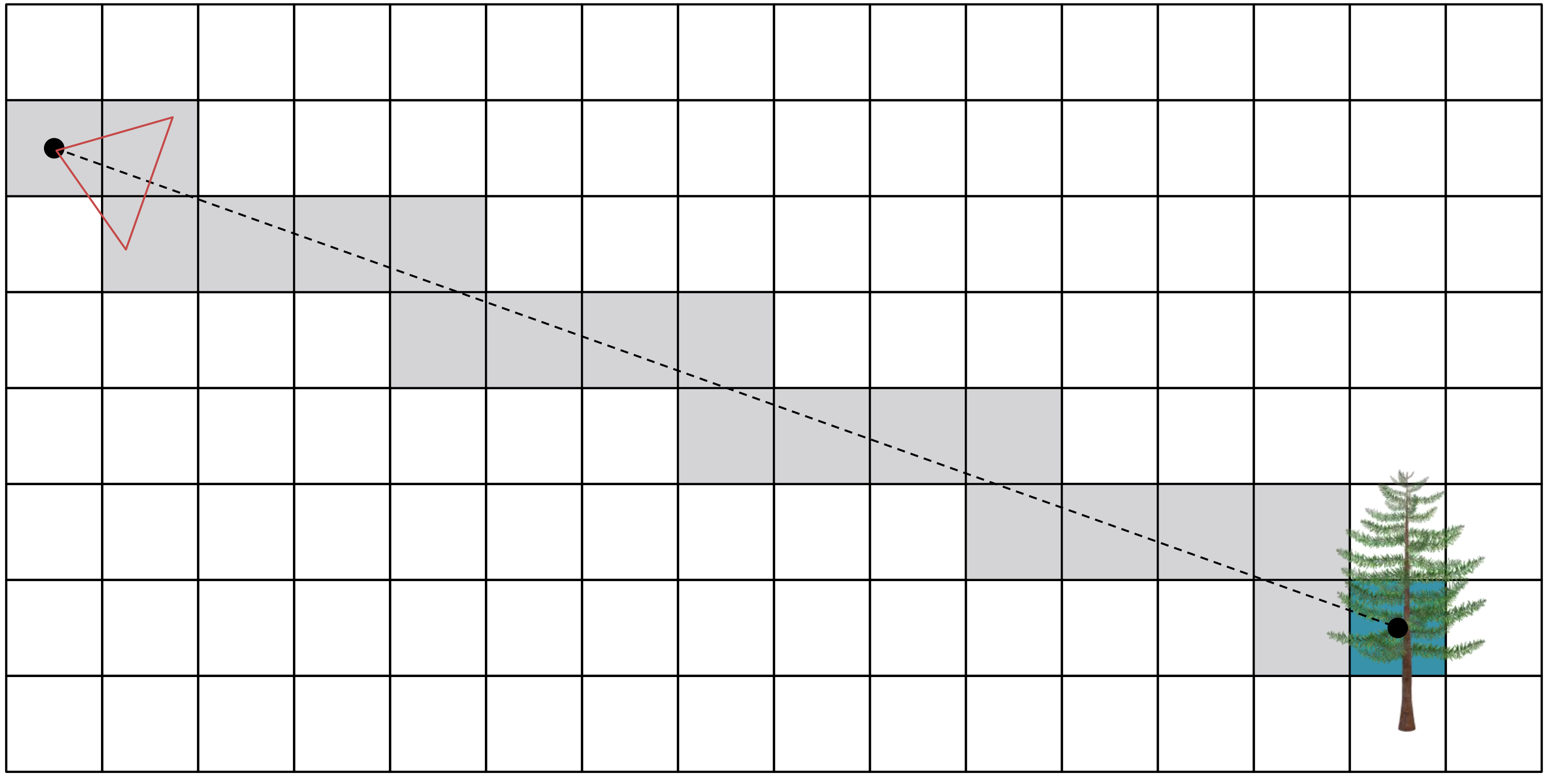
- **A common approach is to use a probabilistic occupancy grid.**
  - Examples: OctoMap and UFOMap
- **In these models, space is hierarchically partitioned into fixed-size voxels.**
  - Each voxel stores a value [0, 1] that represents the probability that the cell is occupied.
  - Observations are incrementally added as depth images or point clouds.
  - Can model free and unknown space
- **This captures one type of uncertainty, but not all!**





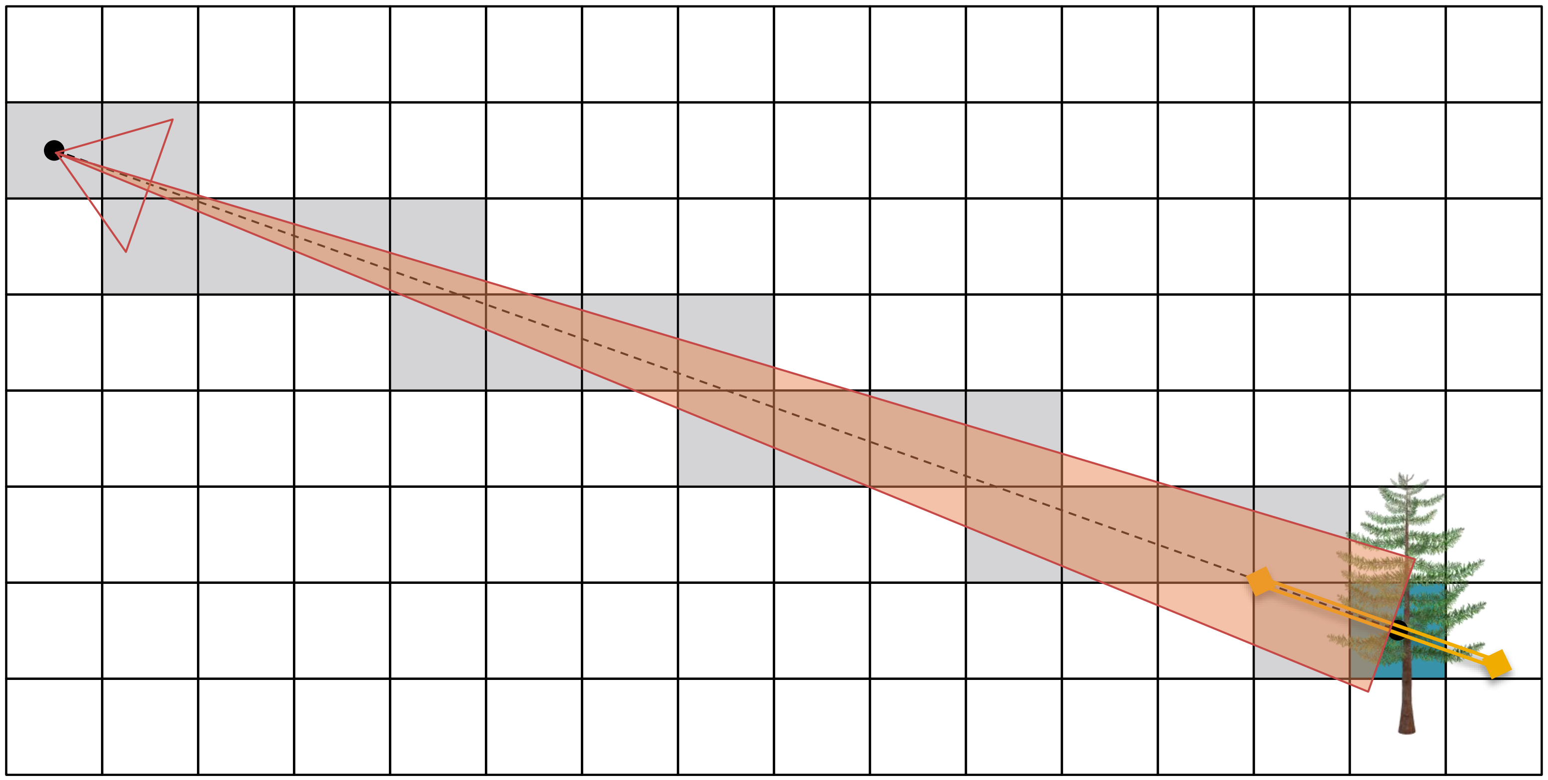
Probabilistic Occupancy Maps

- **Probabilistic methods typically work this way:**
  - **Suppose there is an object in the scene.**
  - **A sensor measures the distance to the object.**
  - **The final grid cell is marked as occupied:** $p(occ)\uparrow$
  - **The cells along the ray are marked as free:** $p(occ)\downarrow$



Towards Fuzzy Voxel Maps

- **Some issues:**
  - **All measurements are considered equal.**
    - We may have ways to assign confidence.
  - **Distances are crisp.**
    - Farther measurements should have more uncertainty (interval?)
  - **Only a single ray is considered for each measurement point.**
    - Area of influence should expand at long ranges.



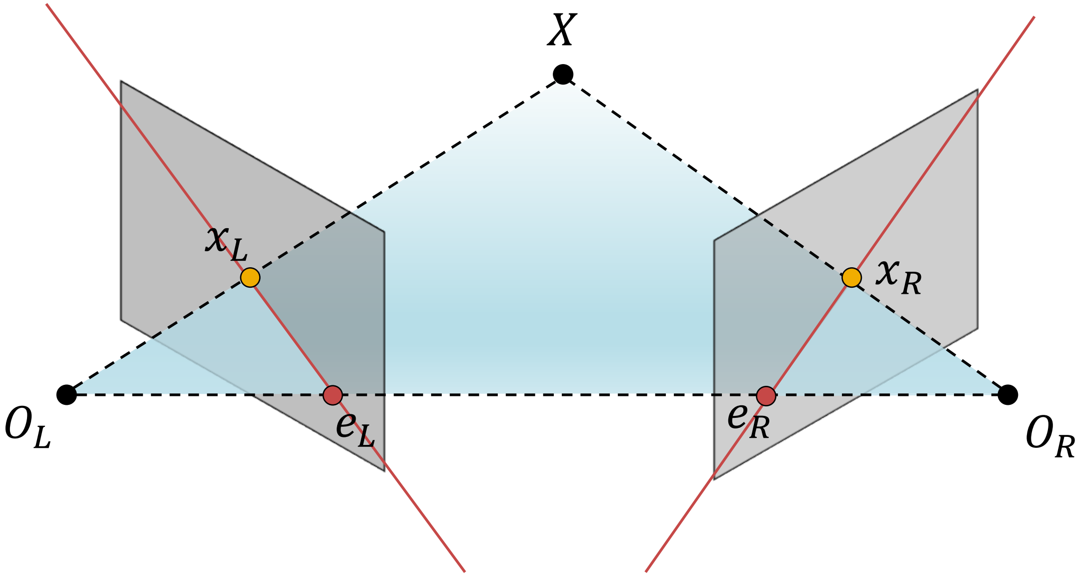
EpiDepth

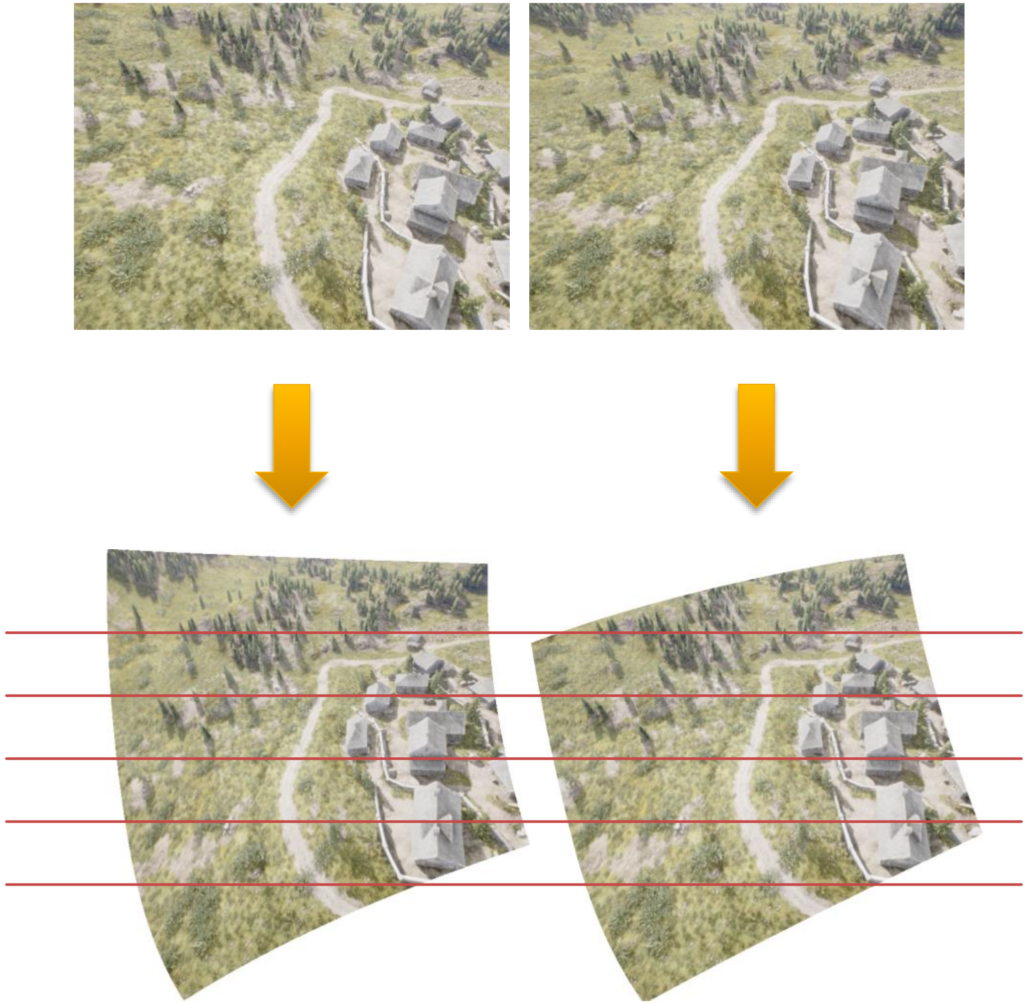
- **A moving camera on a UAV provides a stream of images with known poses (thanks to onboard GPS/IMU).**

- **For a given frame pair, we can align the images and perform stereo matching to estimate depth.**

Epipolar Warping

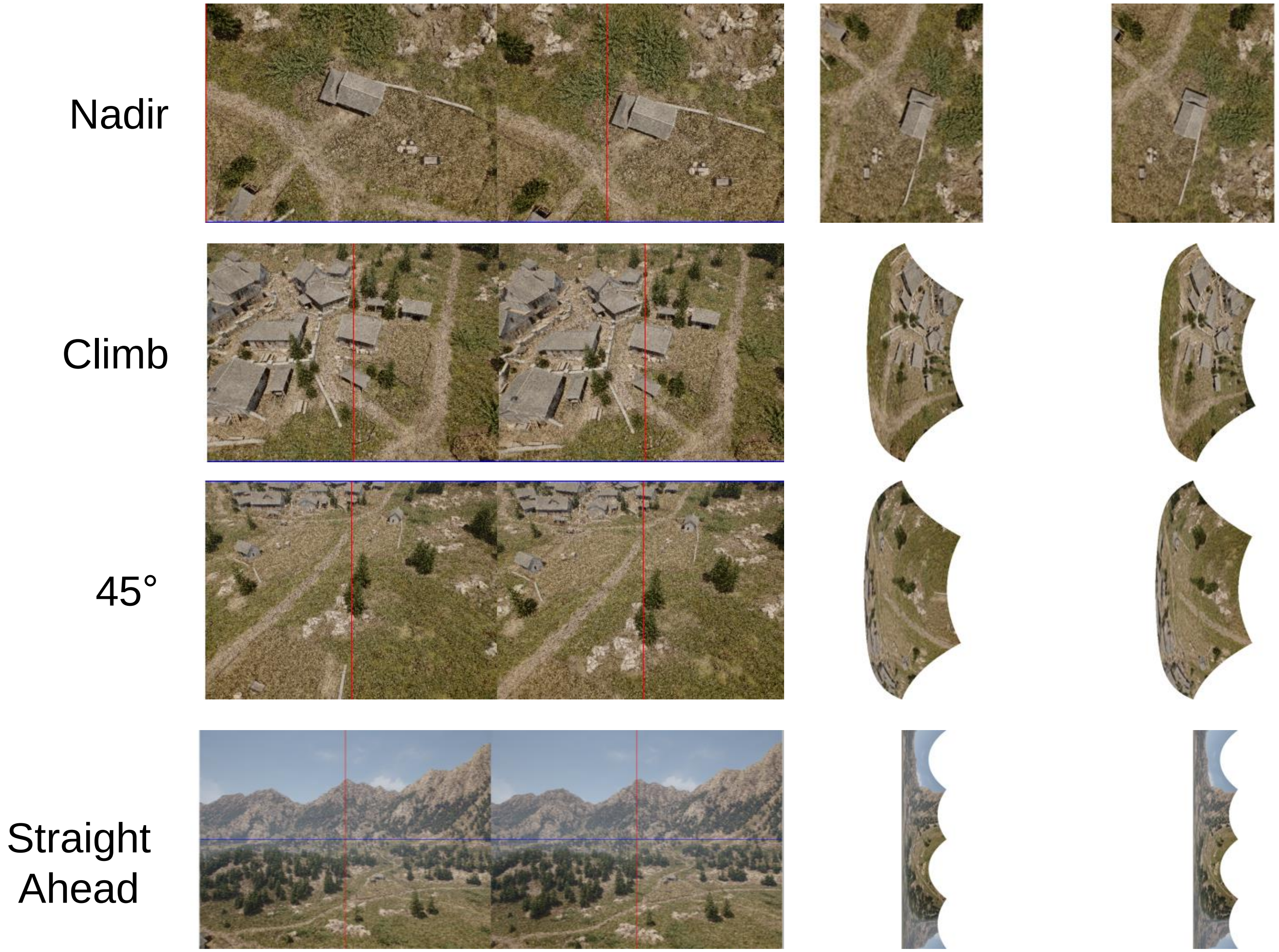
- **The epipolar geometry of two camera views defines how to warp the images.**
- **Feature pairs are aligned on the same row and the pixel disparity is used to estimate depth.**





Warping Effects

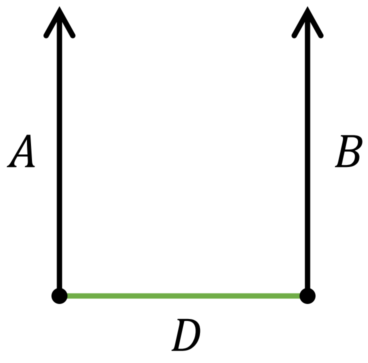
- **The relative pose between images has a big impact on how much warping is required.**
- **Generally, areas around the epipole are hard to match.**



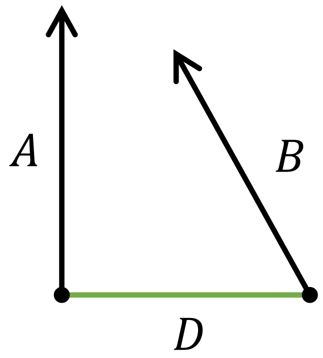
Extrinsic Quality Metric

- **We can define some heuristics to judge the quality of the two frame poses.**
  - Let A and B be the look vectors of the two image frames
  - Let D be the displacement between the focal points of the two image frames

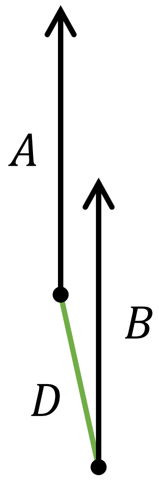
Heuristics:
- $\angle AB$ should be small
- $\angle AD$ and $\angle BD$ should both be close to $90°$



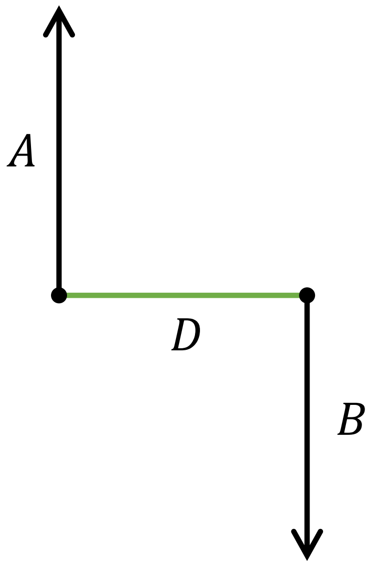
<u>Excellent</u>
Frames are separated perpendicular to the look direction and aligned



<u>Good</u>
Frames are separated and mostly aligned



<u>Poor</u>
Frames are aligned, but separated in the look direction



<u>Bad</u>
Frames are not aligned

EQ Metric Function Crafting

- $\angle AB$ should be small (●)

$$S_{AB} = \cos(\angle AB)$$

$$H_{AB} = \begin{cases} 0, & S_{AB} < 0 \\ S_{AB}, & S_{AB} \geq 0 \end{cases}$$

- $\angle AD$ and $\angle BD$ should both be close to 90° (●)

$$S_{AD} = \cos(\angle AD)$$

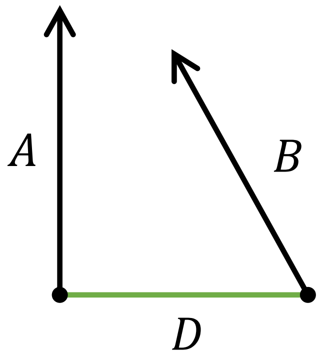
$$S_{BD} = \cos(\angle BD)$$

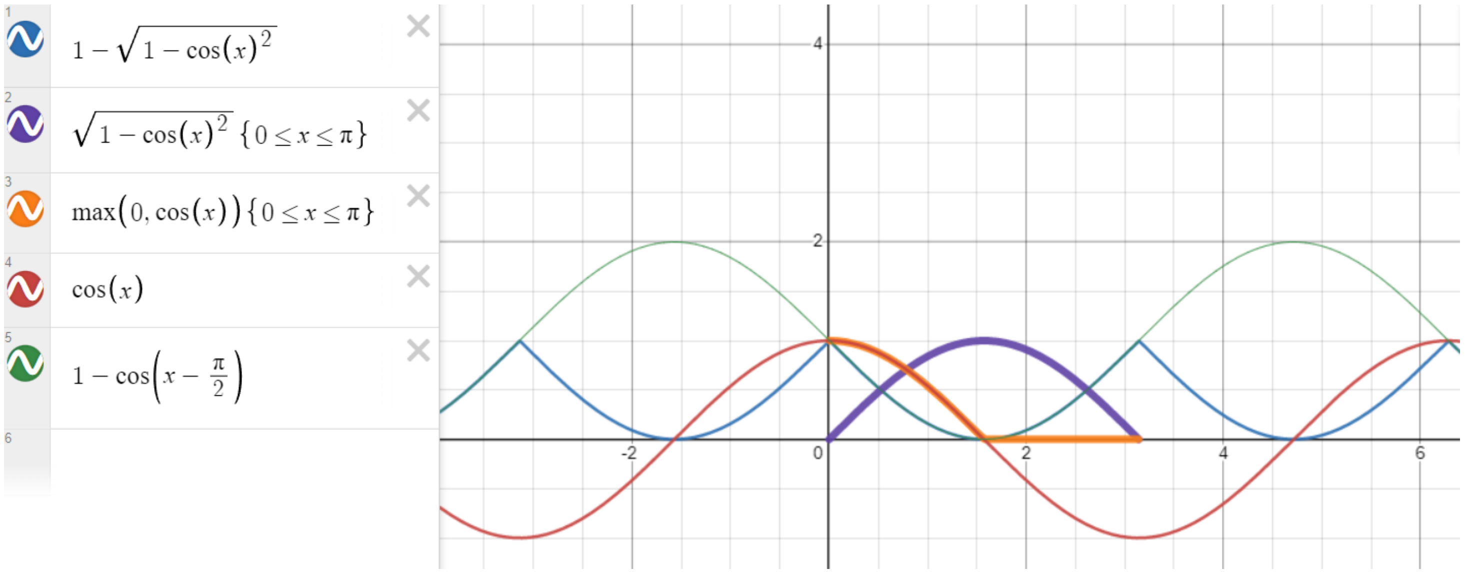
$$R_{AD} = \sqrt{1 - S_{AD}^2}$$

$$R_{BD} = \sqrt{1 - S_{BD}^2}$$

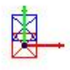
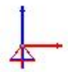
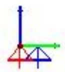
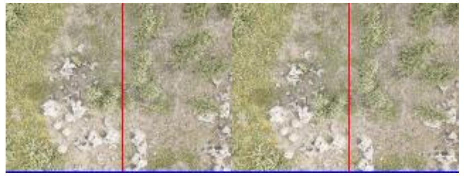
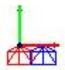
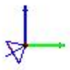
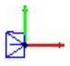
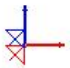
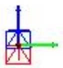
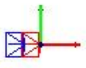
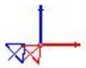
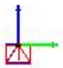
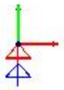
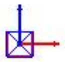
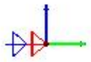
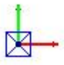
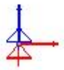
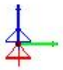
- Overall metric is the minimum of these,

$$Q_{ABD} = \min(H_{AB}, R_{AD}, R_{BD})$$



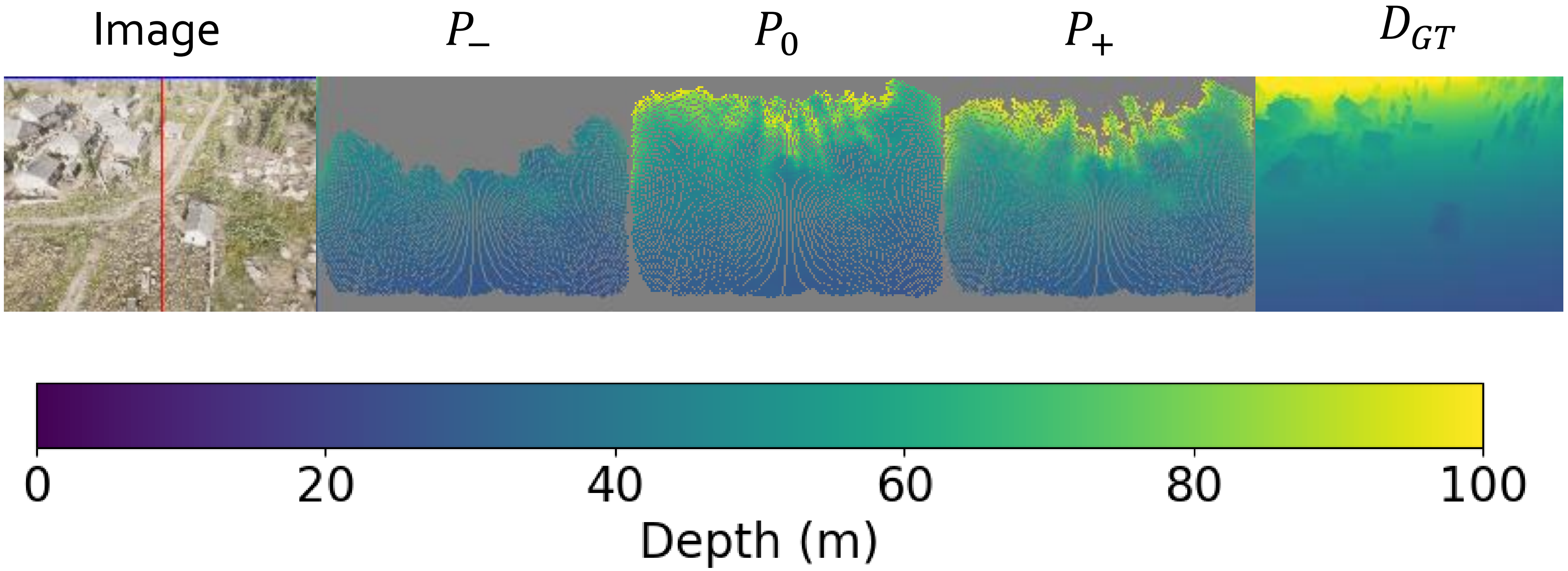


Extrinsic Quality Examples

Extrinsic Quality	Top	Front	Side	Frame 1 / Frame 2	
1.000					Nadir
1.000					Strafe
0.707					45° Climb
0.680					45° Forward
0.001					Straight Ahead
0.000					Nadir Climb

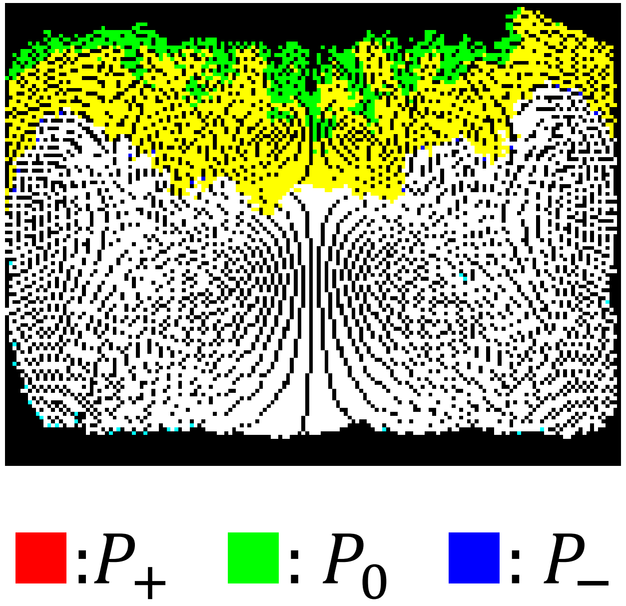
EpiDepth Prediction Confidence

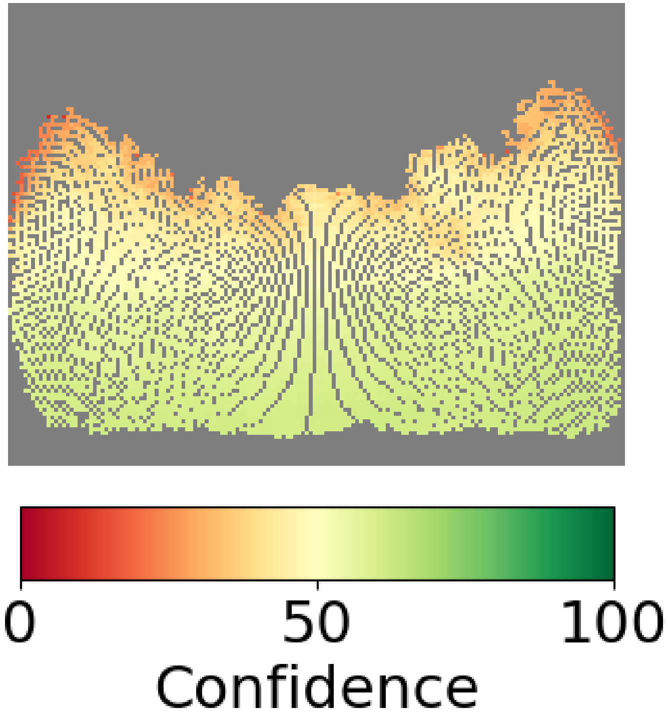
- EpiDepth generates a depth prediction P_0.
- Changing the parameters gives two additional depth predictions, P_- and P_+.
- The difference $|P_+ - P_-|$ gives a measure of sensitivity.
- In simulation, we also have a ground truth depth D_{GT}.

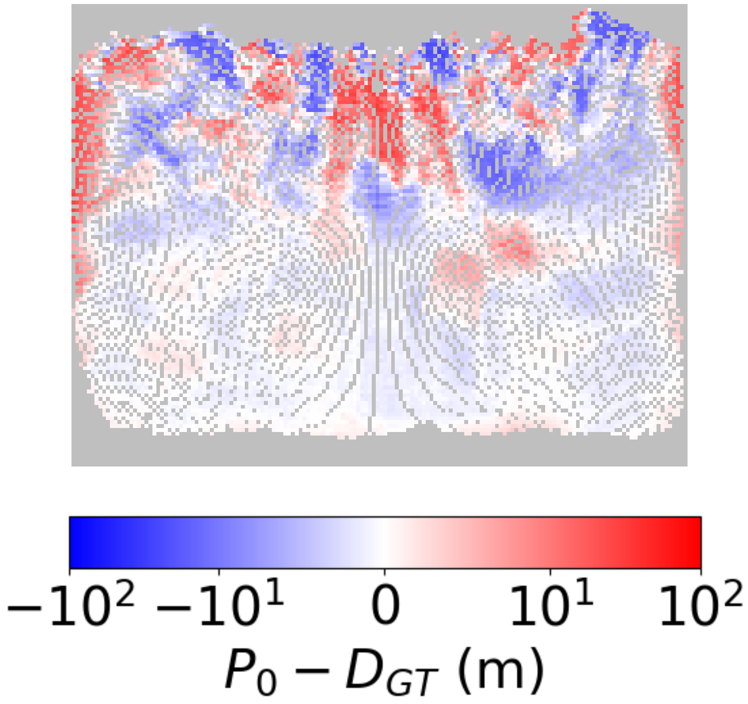


Combined Confidence

- Plotting P_-, P_0, and P_+ as different color channels shows where they all overlap.
- The absolute difference $|P_+ - P_-|$ is scaled to the range [0, 100] as a measure of confidence.
- Comparing P_0 with the ground truth D_{GT} validates our confidence.
- The scores are multiplied by the extrinsic quality metric to give the combined confidence.

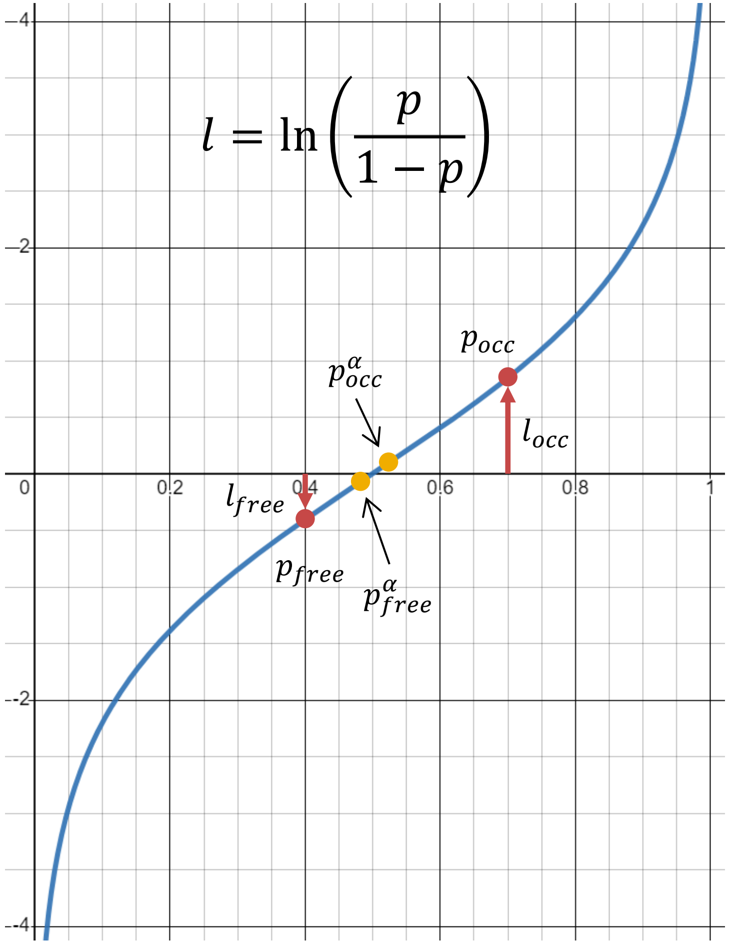






Voxel Map Updates

- **We use UFOMap to store the voxel map.**
  - By default, log-odds probabilities are updated by constant values l_{occ} and l_{free}.
- **To work around this limitation, we insert points multiple times based on the computed confidence.**
  - Confidence values are interpreted as α-cuts.
- **We use $N_t = 10$ thresholds, so the new update values are $1/N_t$ of the original values.**
  - The default values become:
    - $p_{occ} = 0.7\ (l_{occ} = 0.847) \quad \rightarrow p_{occ}^{\alpha} = 0.521$
    - $p_{free} = 0.4\ (l_{free} = -0.405) \quad \rightarrow p_{free}^{\alpha} = 0.490$



Example: Wall Reconstruction

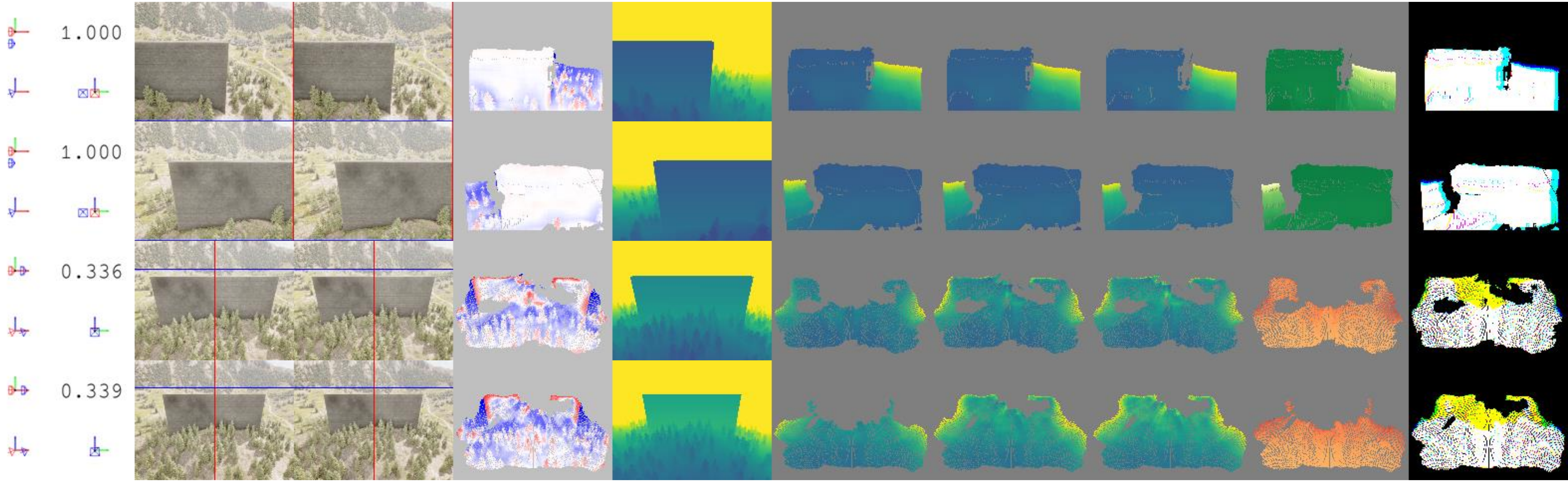
- First, we strafe along a large known reference object and generate a highly confident voxel map.
- Then, we move away and generate less confident measurements.
- We want to show that we don’t erase the wall with poor depth estimates.

Initial Strafe

Moving Backwards

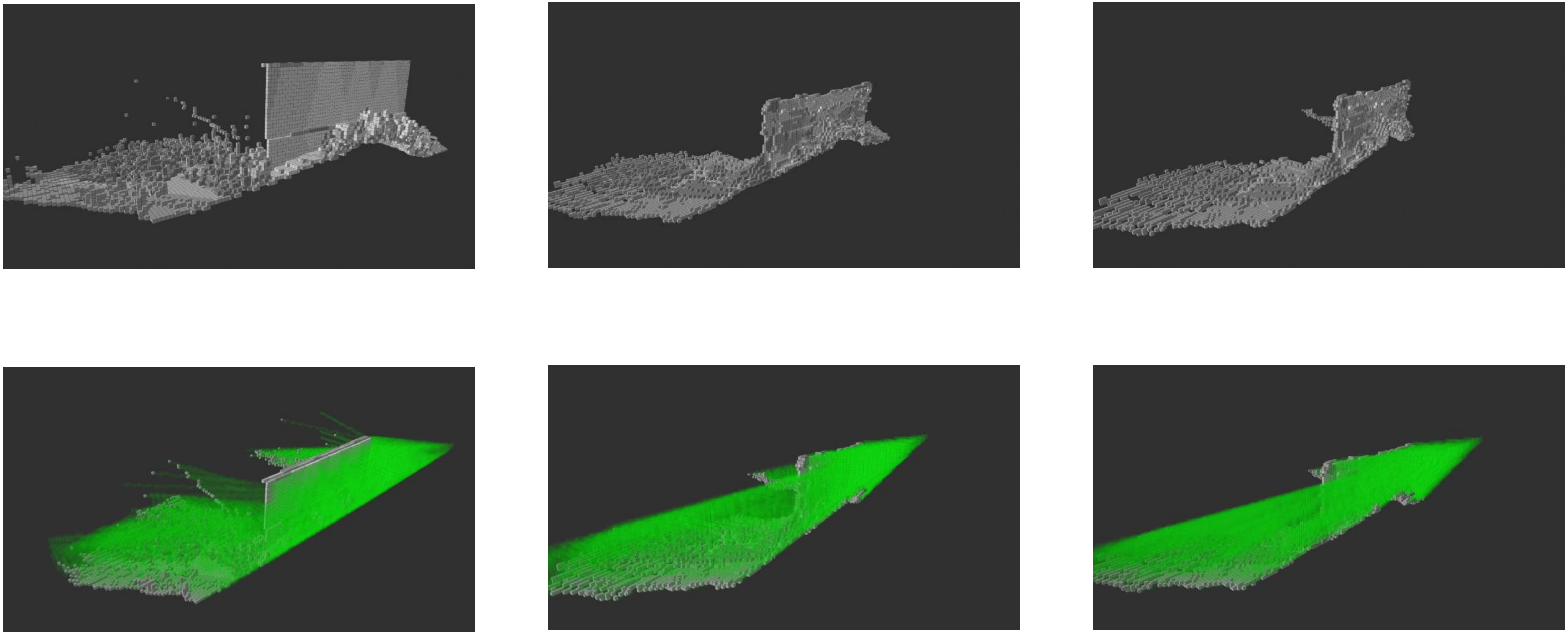
Wall Evaluation Images

Extrinsics	Frame 1	Frame 2	$P_0 - D_{GT}$	D_{GT}	P_-	P_0	P_+	C	Overlap



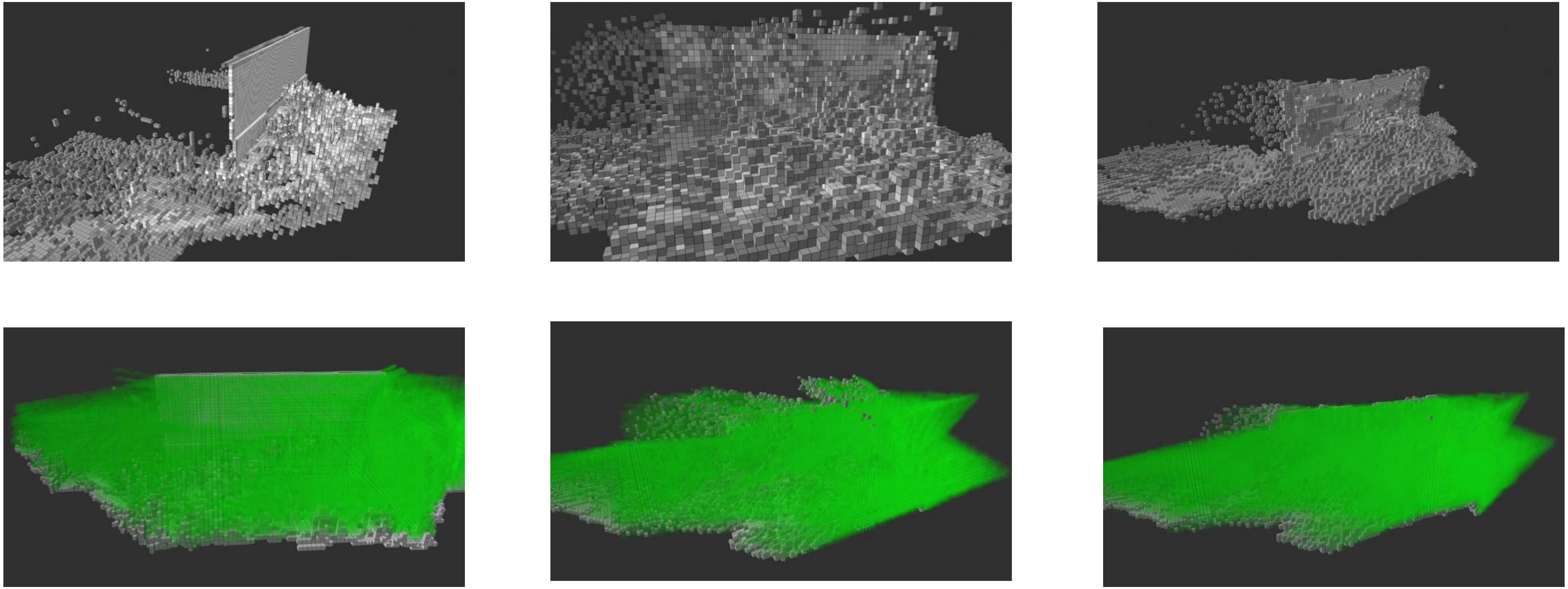
Initial Wall Scan

Ground Truth | Standard | Fuzzy



Backing Away from Wall

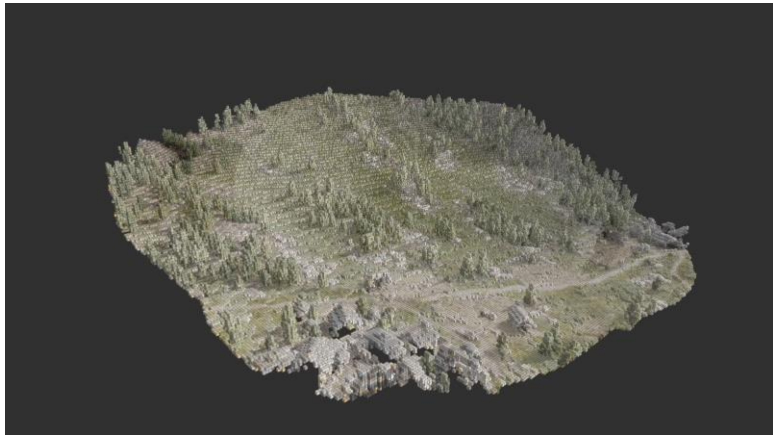
Ground Truth | Standard | Fuzzy



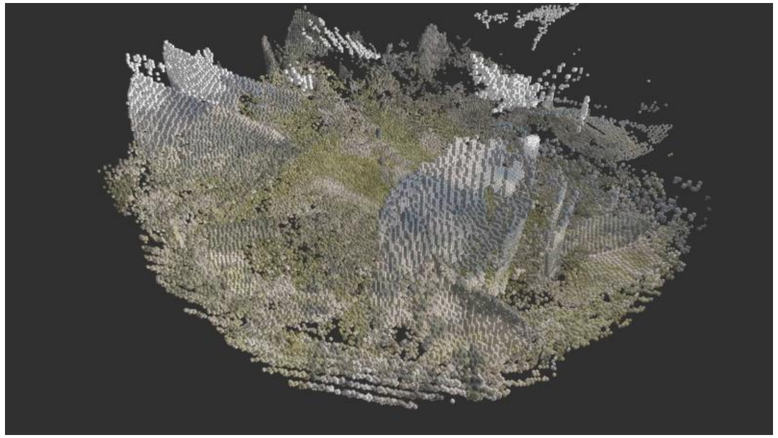
Example: Random Movement

- In this scenario, the UAV moves to random locations with random poses within a fixed area.
- We compare the standard and fuzzy approaches with the ground truth.

Ground Truth

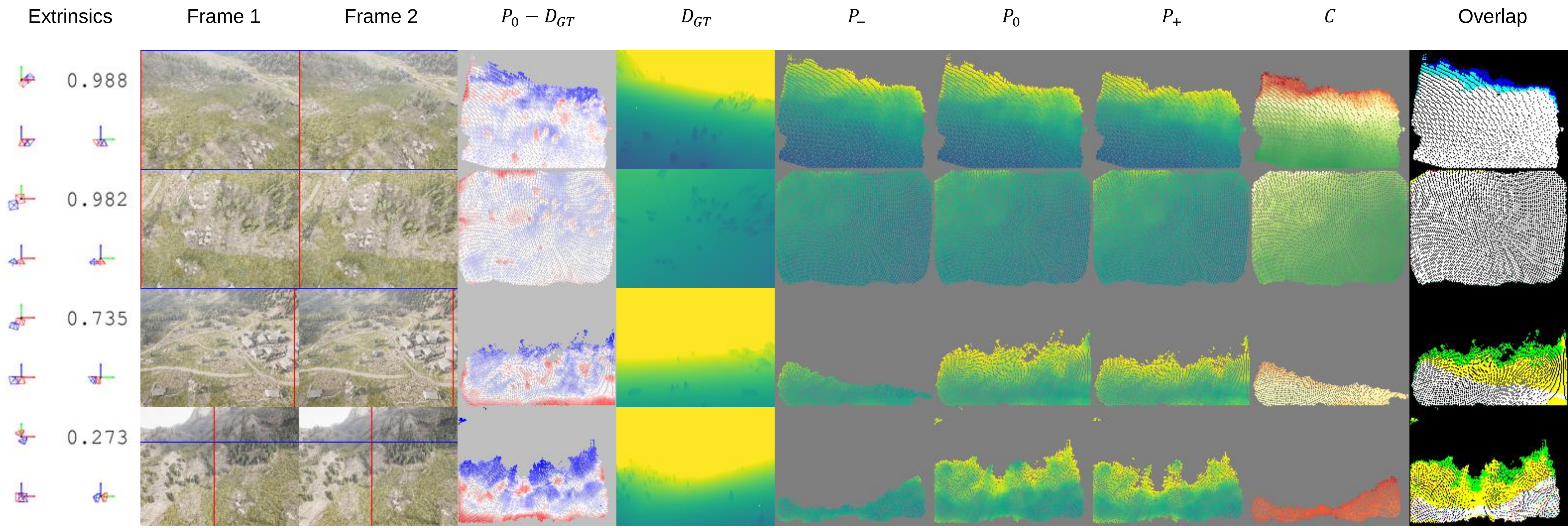


Standard



Fuzzy

Random Movement Evaluation



Conclusions

- **Probabilistic occupancy grids like OctoMap and UFOMap don’t capture all the uncertainty.**
  - It can be hard to tell if a cell had conflicting observations or was rarely observed.
- **SfM techniques like EpiDepth are fundamentally different than range-based approaches like LiDAR.**
  - We can utilize known confidence values to improve 3D map quality.
- **There are many more ways to extend this!**
  - Distance intervals and spatial uncertainty can be included.
  - Currently using a single map to represent belief that a voxel is occupied.
  - Could use multiple maps as membership sets (free, occupied), as with the work of Oriolo et al.

Questions?