Mizzou INformation and Data FUsion Lab (MINDFUL)

Title: Frame Selection Strategies for Real-Time Structure-from-Motion from an Aerial Platform

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

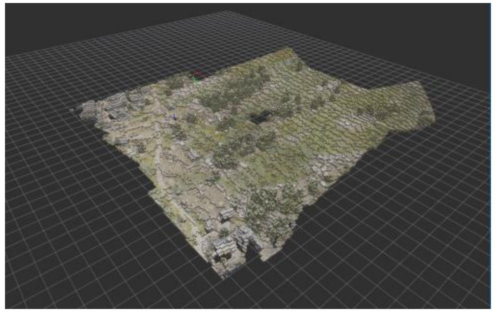
University of Missouri

Department of Electrical Engineering and Computer Science

September 27th, 2023

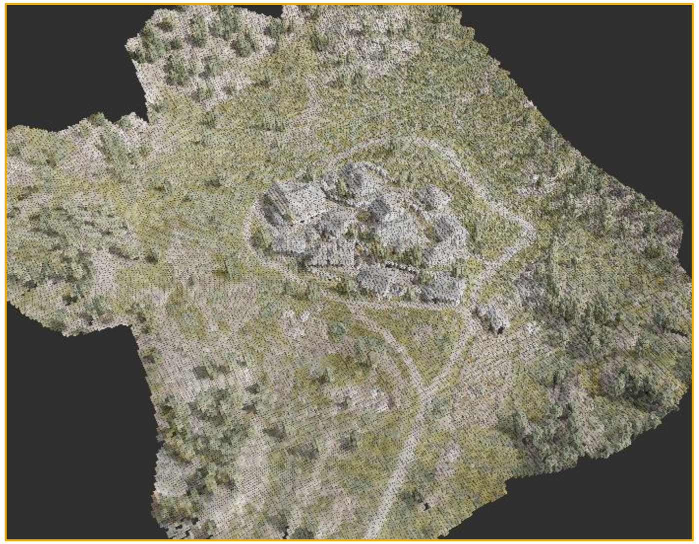
Introduction

- **What's our problem?**
  - Real-time 3D mapping from a single camera on a micro-UAV
- **Challenges?**
  - Single camera
  - Can solve with SfM on frame pairs
  - Don't control where drone goes (manual control)
  - <u>How to select frames</u> to obtain the best reconstruction?
- **Today**
  - <u>SIM environment and study</u>



Random Movement Dataset

UE native quality



Reconstructed 3D voxel space (UFOmap)

AirSim quality

Collected about 1000 frames of aligned RGB/Depth imagery

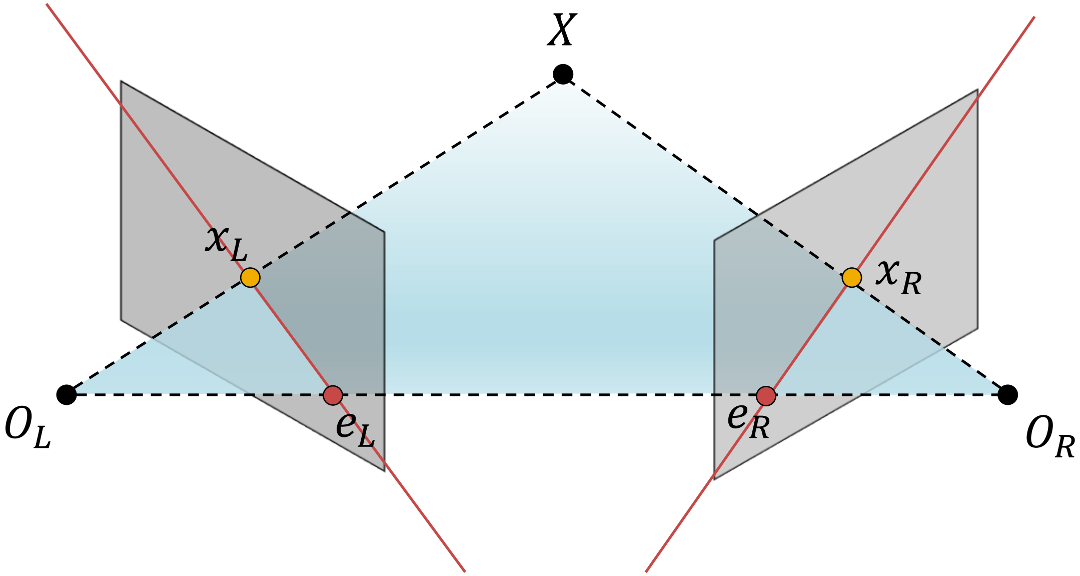
Drone moves to random poses (position and look direction)

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.

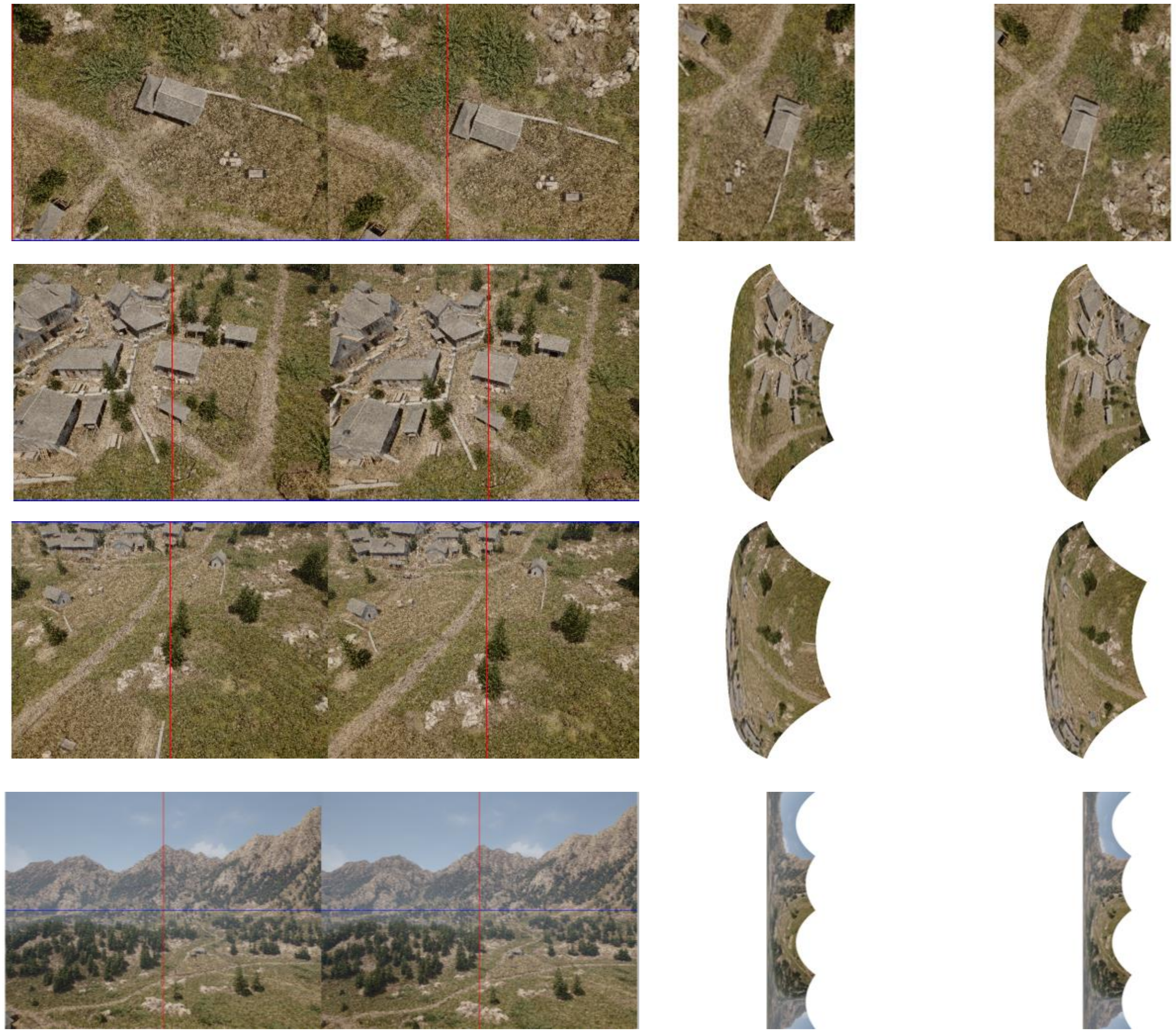
- Generally, areas around the epipole are hard to match.

Nadir

Climb

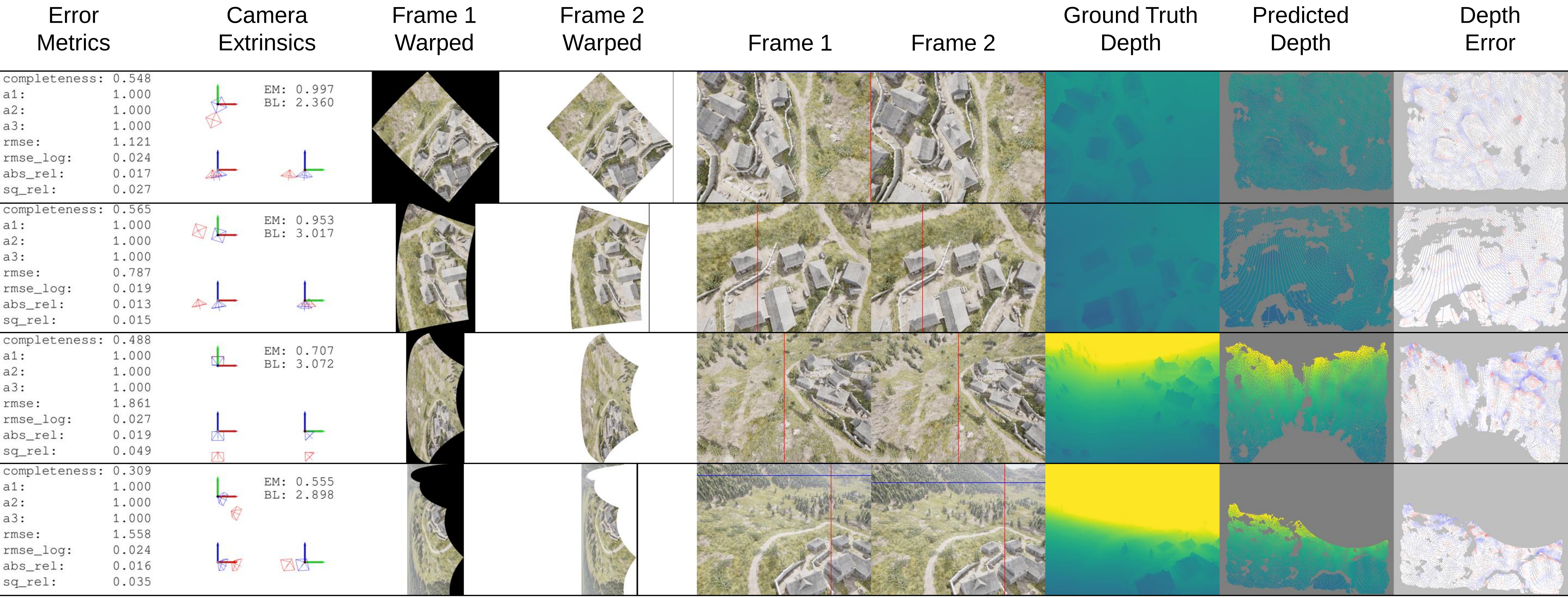
Forward 45°

Straight Ahead



EpiDepth Matching Examples

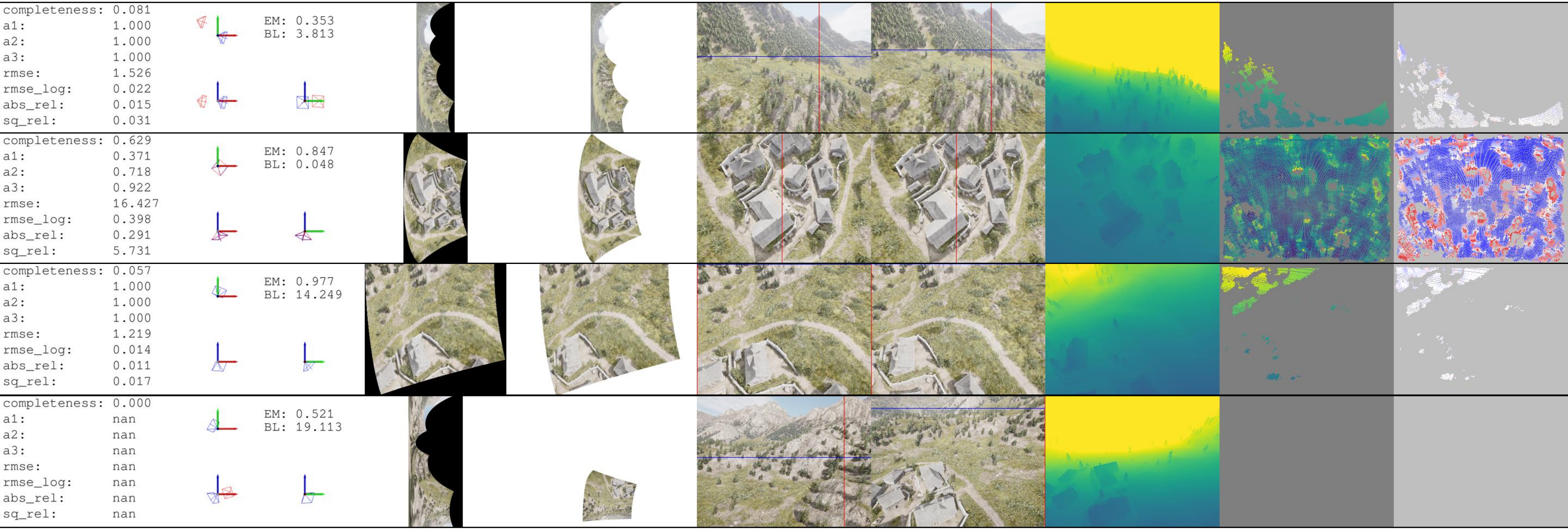
- Good cases for frame pair matching…



EpiDepth Matching Examples

- Not so good cases for frame pair matching…

Error Metrics | Camera Extrinsics | Frame 1 Warped | Frame 2 Warped | Frame 1 | Frame 2 | Ground Truth Depth | Predicted Depth | Depth Error



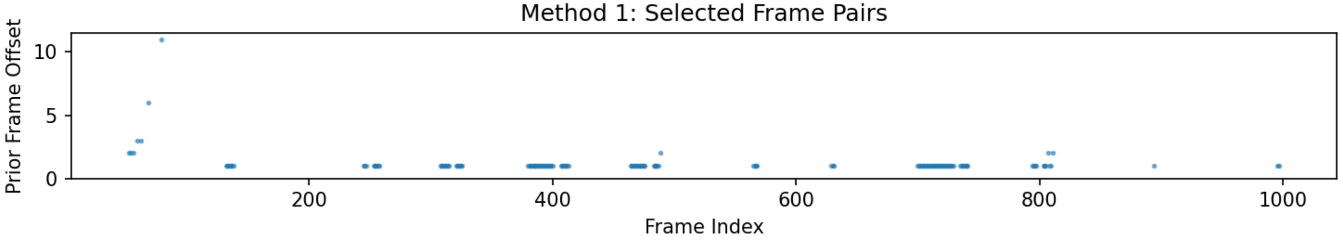
Simple Frame Picker (Version 1)

- Keeps a running candidate frame and tries to match new incoming frames to this one.
- Reject if frames are too close or if the rotation difference is too large.
- If a pair is found, yield it and keep the latest frame as the new frame to match to.
- Otherwise, if distance becomes too large, replace the candidate frame with the latest frame.

<u>Frame Picker v1</u>

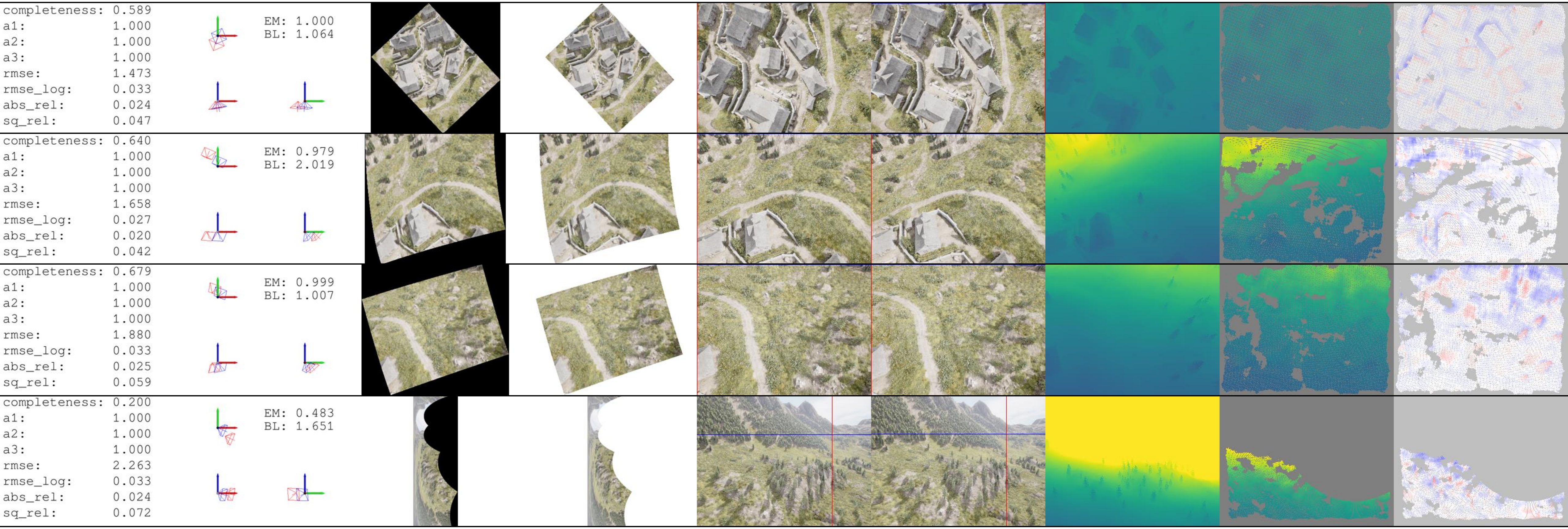
```
Define d_min, d_max, and r_max
Initialize frame f_0
For each new frame f_t:

    // Get baseline distance
    d ← DISTANCE(f_0, f_t)

    // Check if extrinsics are acceptable
    If d_min < d < d_max:
        r ← ROTATION(f_0, f_t)
        If r < r_max:
            yield (f_0, f_t)
            f_0 ← f_t

    // Drop old frame
    If d > d_max:
        f_0 ← f_t
```

Method 1 (Simple) Results



Total frame pairs selected: **142**
Completeness: 0.432 ± 0.226
RMSE-Log: 0.035 ± 0.010

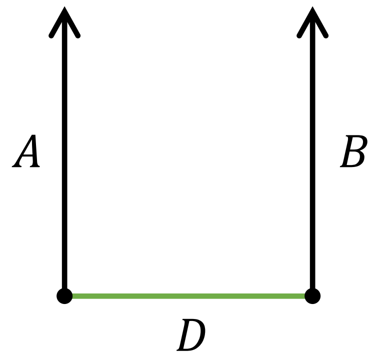
Examples:



Extrinsic Quality Metric

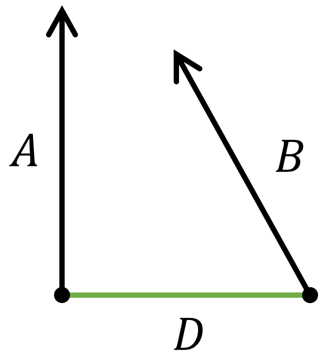
- **We can define some heuristics to judge the quality of the two frame poses.**
  - Let A and B be the look vectors of the two image frames
  - Let D be the displacement between the focal points of the two image frames

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



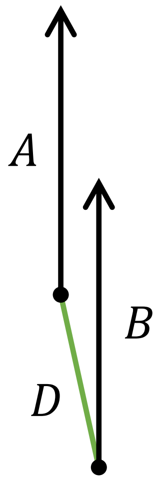
Excellent

Frames are separated perpendicular to the look direction and aligned



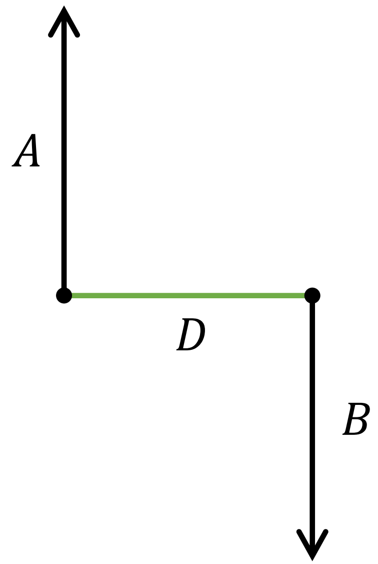
Good

Frames are separated and mostly aligned



Poor

Frames are aligned, but separated in the look direction



Bad

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° (●)

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

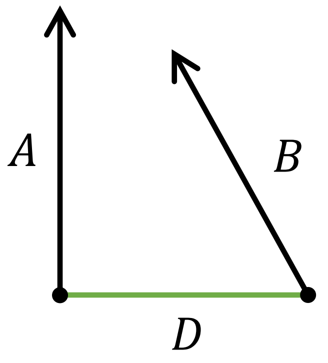
$$S_{BD} = \cos(\angle BD)$$

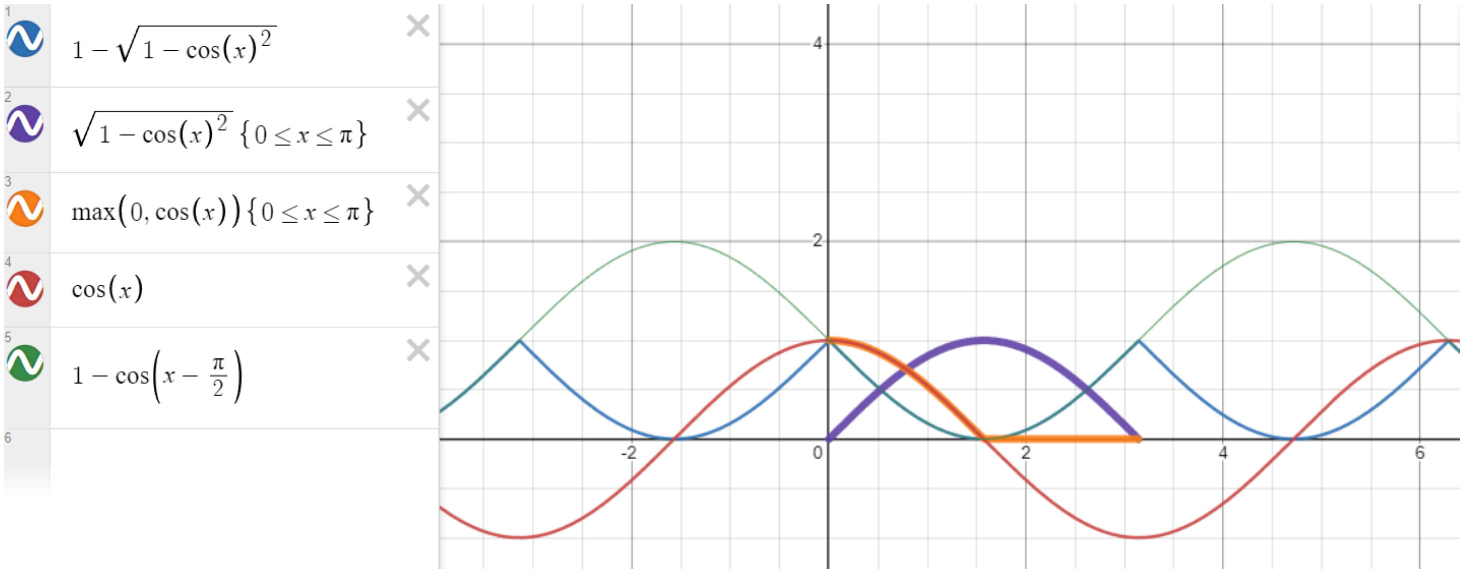
$$R_{AD} = \sqrt{1 - S_{AD}^2}$$

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

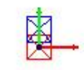
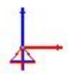
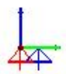
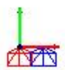
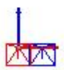
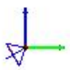
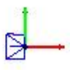
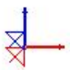
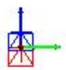
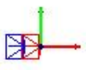
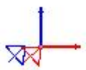
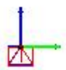
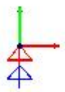
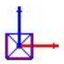
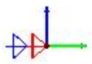
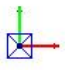
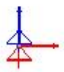
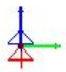
- Overall metric is the minimum of these,

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





Extrinsic Quality Examples

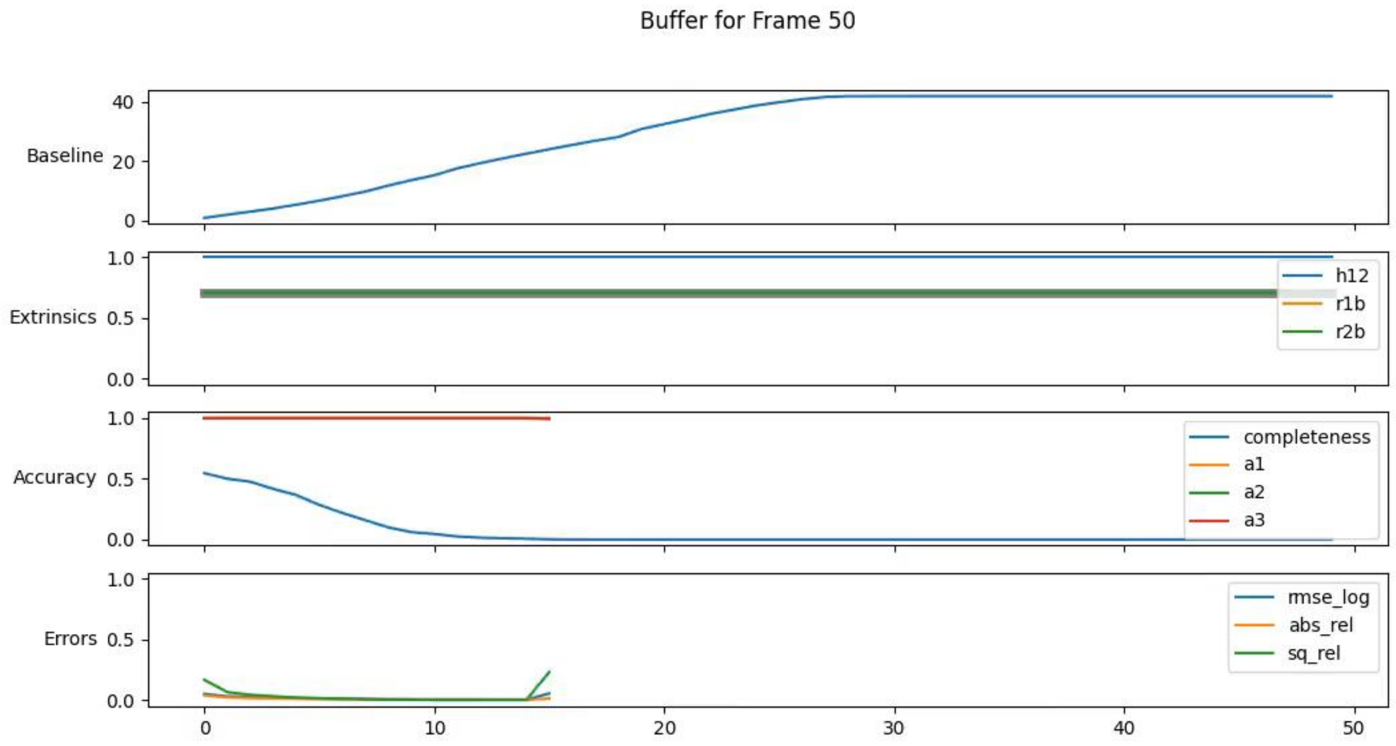
Extrinsic Quality	Top	Front	Side	Frame 1	Frame 2	
1.000						Nadir
1.000						Strafe
0.707						45° Climb
0.680						45° Forward
0.001						Straight Ahead
0.000						Nadir Climb

Rolling Frame Buffer

As new frames arrive, add them to the buffer.

Compute baseline distance and extrinsic metric with each frame in the buffer.

Use these features to select the best frame to match with.



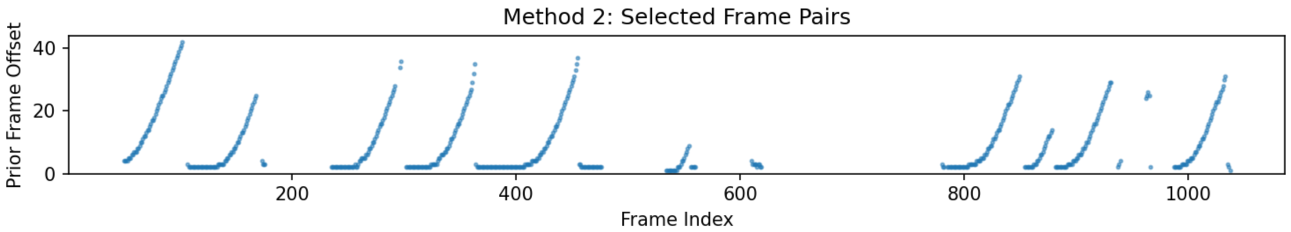
Heuristic Frame Picker (Version 2)

- Keep rolling buffer, past N frames.
- For each new frame, look back in the buffer for the most recent frame that has a minimum baseline distance.
- If this frame has an acceptable extrinsic metric when compared with the current frame, yield the frame pair.
- Otherwise, move on to the next incoming frame.

<u>Frame Picker v2</u>

Define N, d_{min}, q_{min}
Initialize rolling frame buffer B
For each new frame f_t:
    $B.insert(f_t)$
    For $i = 1 \ldots N$:
        $d \leftarrow \textsc{Distance}(f_{t-i}, f_t)$
        If $d < d_{\text{min}}$:
            continue
        $q \leftarrow \textsc{Extrinsic\_Quality}(f_{t-i}, f_t)$
        If $q \geq q_{\text{min}}$:
            yield (f_{t-i}, f_t)
        break

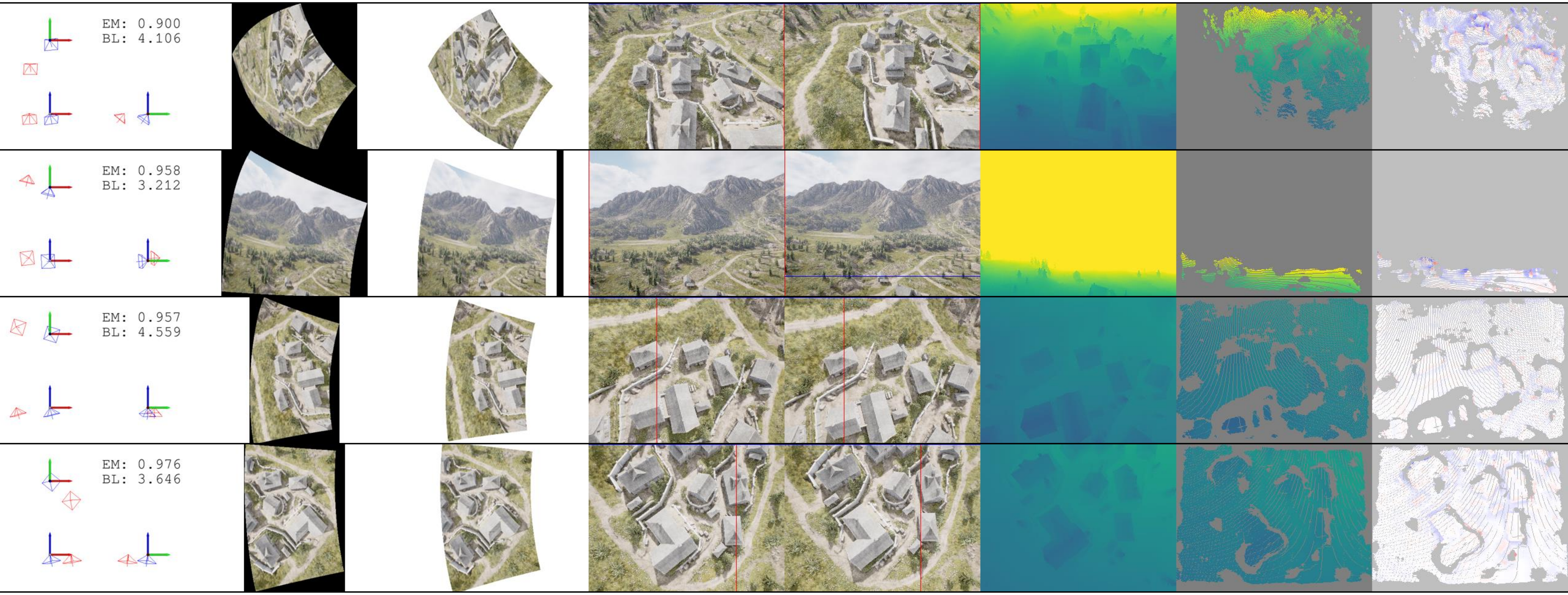
Method 2 (Heuristic) Results



Total frame pairs selected: **592**
Completeness: 0.364 ± 0.190
RMSE-Log: 0.030 ± 0.138

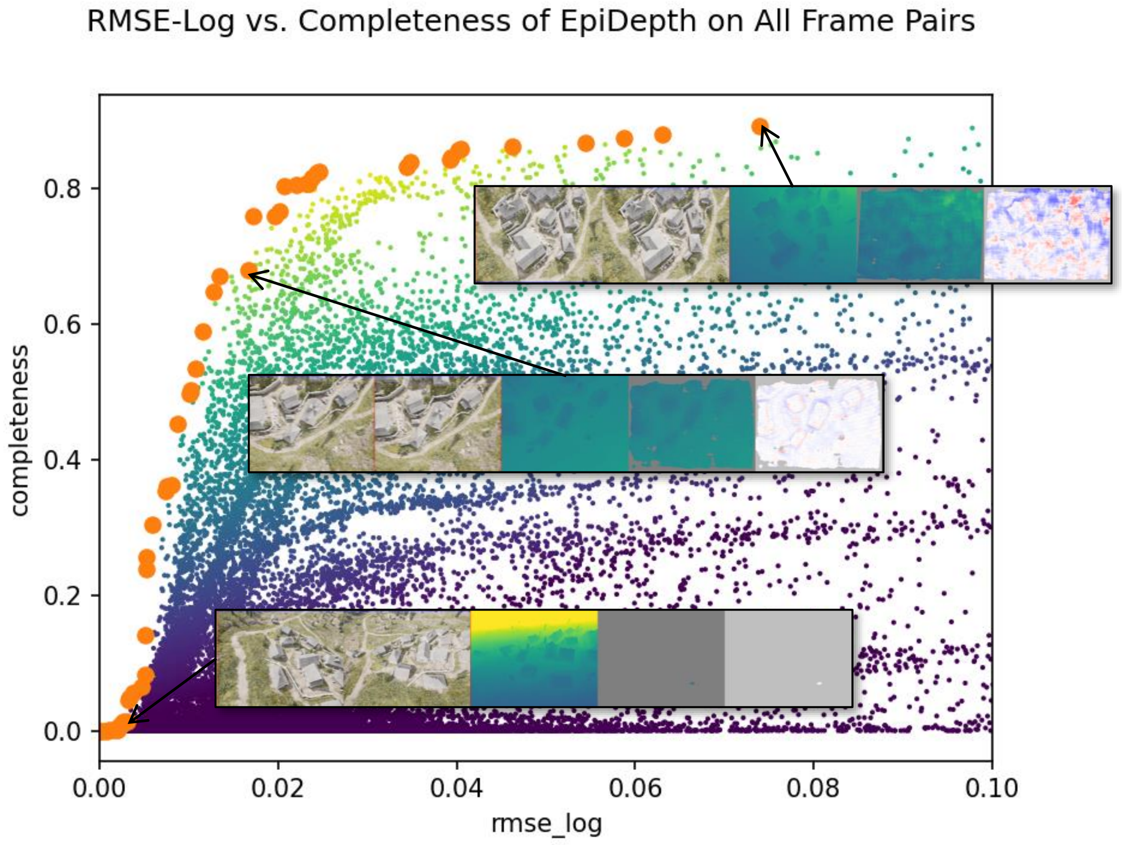
Examples:

completeness:	0.281	0.074	0.543	0.678
a1:	1.000	1.000	1.000	1.000
a2:	1.000	1.000	1.000	1.000
a3:	1.000	1.000	1.000	1.000
rmse:	2.411	2.702	0.639	0.860
rmse_log:	0.036	0.031	0.015	0.019
abs_rel:	0.026	0.022	0.010	0.012
sq_rel:	0.082	0.082	0.009	0.016



Evaluating All Frame Pairs

- Since we have computed the EpiDepth prediction for all possible frame pairs, can we make use of this data?
- We want results to have low RMSE-log values and high completeness.
- Looking at examples from the Pareto front, we select appropriate scalarization weights.
- A neural net is trained on this dataset and can be used to predict if a pose pair will perform well.



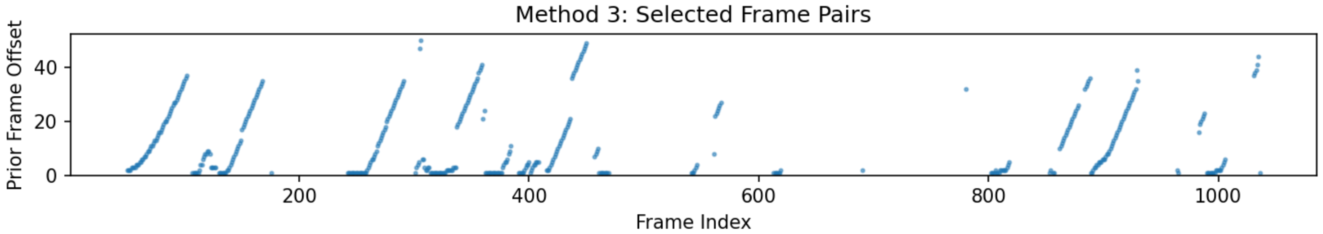
Data-Driven Frame Picker (Version 3)

- Keep a rolling buffer of the past N frames.
- For each new frame, evaluate the predicted quality with all frames in the buffer.
- If the best scoring frame satisfies the minimum baseline and extrinsic quality thresholds, yield it.
- Otherwise, move on to the next incoming frame.

Frame Picker v3

Define N, $d_{\min}$, $q_{\min}$, $p_{\min}$
Initialize rolling frame buffer B
For each new frame f_t:
    $B.insert(f_t)$
    $f_{\text{best}} \leftarrow \emptyset$
    $p_{\text{best}} \leftarrow -\infty$
    For $i = 1 \dots N$:
        $d \leftarrow \textsc{Distance}(f_{t-i}, f_t)$
        $q \leftarrow \textsc{Extrinsic\_Quality}(f_{t-i}, f_t)$
        If $d < d_{\min}$ or $q < q_{\min}$:
            continue
        $p \leftarrow \textsc{Predict}(f_{t-i}, f_t)$
        If $p > p_{\text{best}}$ and $p > p_{\min}$:
            $p_{\text{best}} \leftarrow p$
            $f_{\text{best}} \leftarrow f_{t-i}$
    If $f_{\text{best}} \neq \emptyset$:
        yield (f_{best}, f_t)

Method 3 (Data-Driven) Results



Total frame pairs selected: **453**
Completeness: 0.322 ± 0.249
RMSE-Log: 0.024 ± 0.027

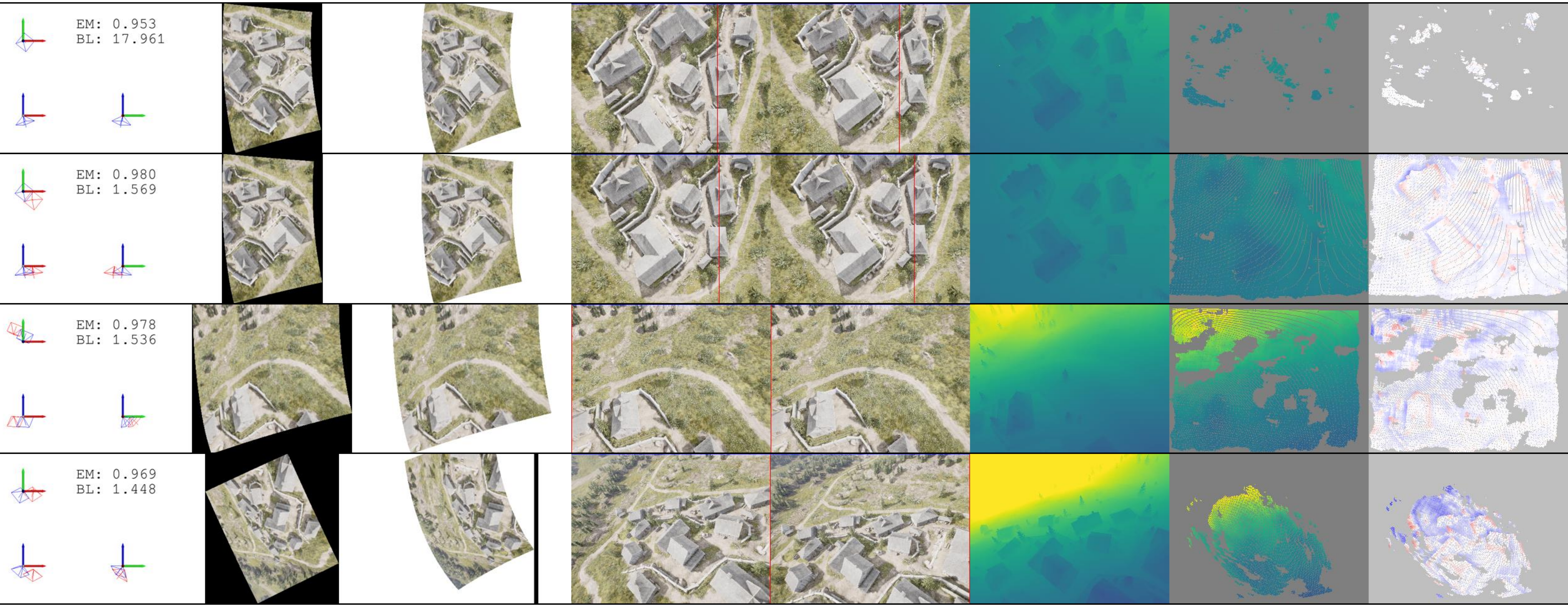
Examples:

```
completeness: 0.050
a1:           1.000
a2:           1.000
a3:           1.000
rmse:         1055.396
rmse_log:     0.118
abs_rel:      0.007
sq_rel:       17.009
```

```
completeness: 0.799
a1:           1.000
a2:           1.000
a3:           1.000
rmse:         1.235
rmse_log:     0.027
abs_rel:      0.019
sq_rel:       0.033
```

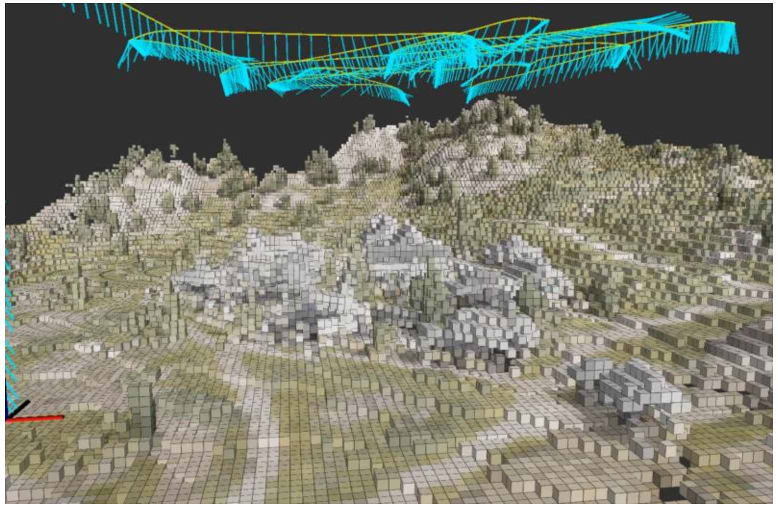
```
completeness: 0.645
a1:           1.000
a2:           1.000
a3:           1.000
rmse:         1.818
rmse_log:     0.029
abs_rel:      0.022
sq_rel:       0.050
```

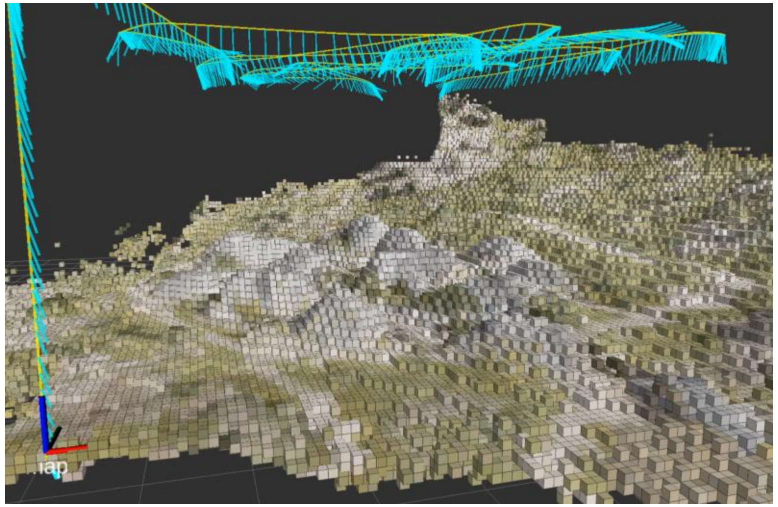
```
completeness: 0.213
a1:           0.997
a2:           1.000
a3:           1.000
rmse:         3.913
rmse_log:     0.058
abs_rel:      0.044
sq_rel:       0.209
```



Qualitative 3D Evaluation

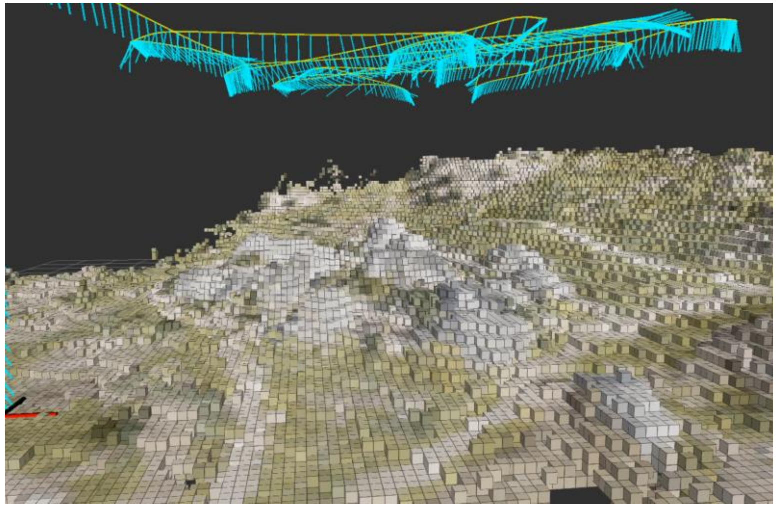
Ground Truth

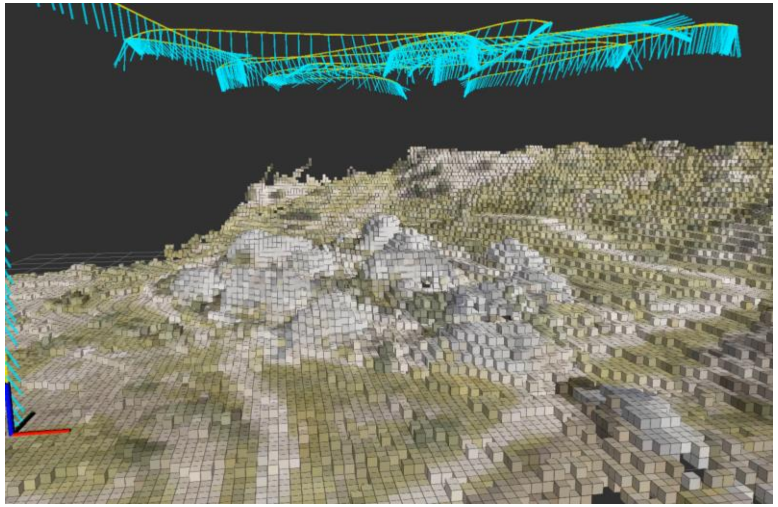




Method 1 (Simple)

Method 2 (Heuristic)



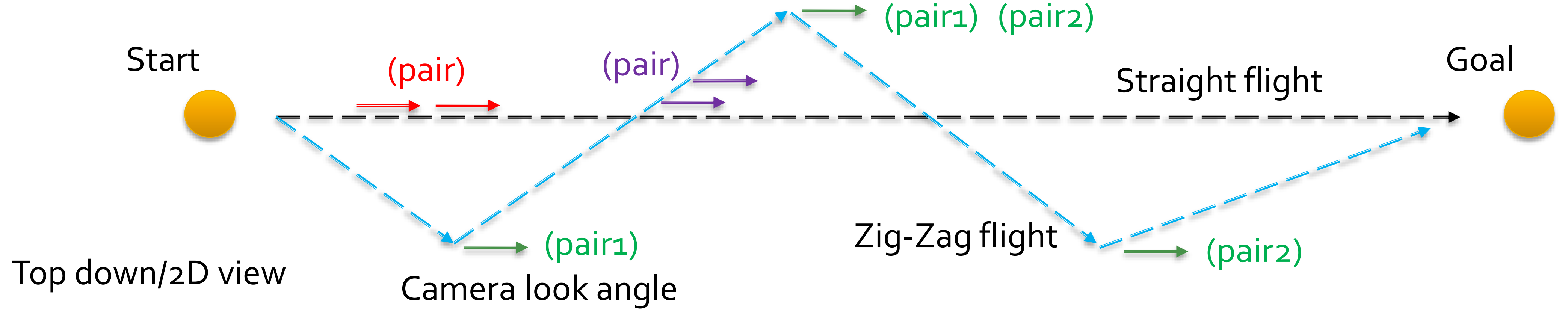


Method 3 (Data-Driven)

Trust These SIM Results?

- **Well, as real as SIM (and our setup at that) is real.**
  - What are the important variables to SIM?

1. **Extrinsic error**
   - BIG real-world problem for sure! (position, pose, …)
   - Can drastically impact UFOmap (our 3D aggregate structure), even when/if EpiDepth's good.
   - Does this mean "fewer projected points are more dangerous…"?
2. **Specific behaviors**
   - We ran a random movement sequence, wanted to remove any bias.
   - But did that wash out behavior specific benefits?
     - *e.g., fly a zig-zag vs straight flight pattern, larger baselines, see further out*



Summary

- **SIM framework for SfM frame selection**
  - Real-time micro-UAV
  - Code base is nearly one-2-one with real platform
- **Explored: (1) naïve, (2) heuristic, (3) data-driven**
  - Best-2-worse ➔ data-driven (3) then (2) then (1)
  - Quant: image space error, # of frames picked, % completeness
  - Qual: 3D voxel space reconstruction and 2D visualizations
- **Induced too easy/clean of a SIM setup?**
  - Expected bigger differences (like we have been seeing on real)
  - Perfect extrinsic data and random flight pattern
  - Goal was to generate a dataset that covers the distribution of all possible pose configurations and study in detail

Next Steps

- **SIM environment**
  - Suspected we might need to include more factors
    - Add extrinsic error and specific flight patterns to SIM
- **SIM experiments**
  - Quantitative 3D voxel space metrics (from our SPIE paper)
    - How good at free space, occupied, time-varying, etc.
  - Decomposition by “variables” (from our MSS paper)
    - How good w.r.t. object ID type, breakdown by range, etc.
- **Frame selection**
  - Preliminary work
  - More in-depth analysis of these three real-time solutions
  - Improved data-driven solution
- **SIM vs real**
  - Real world confirmational experiments

Questions?