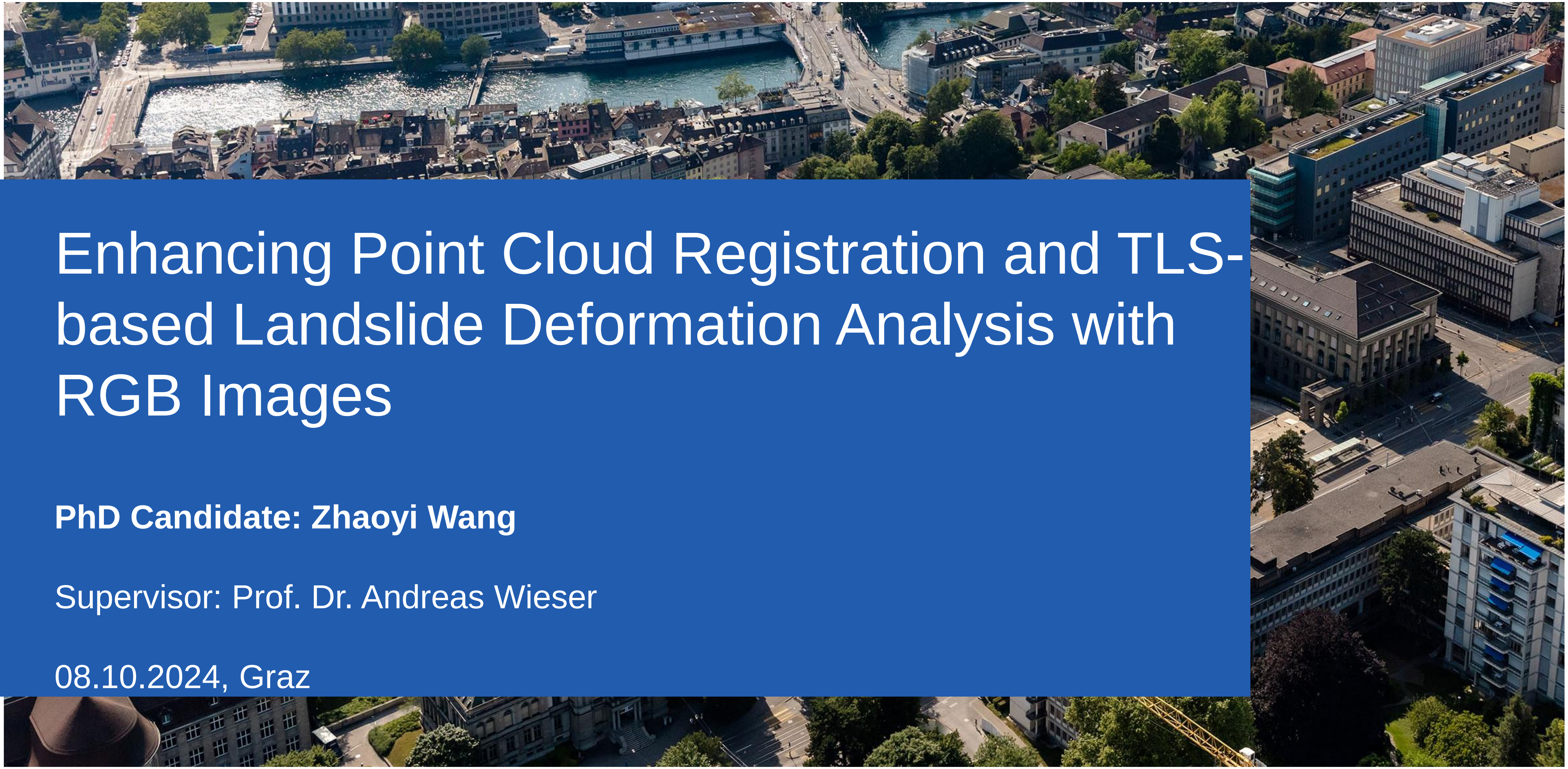
ETH zürich

Geosensors and Engineering Geodesy (GSEG)
Institute of Geodesy and Photogrammetry (IGP)



Enhancing Point Cloud Registration and TLS-based Landslide Deformation Analysis with RGB Images

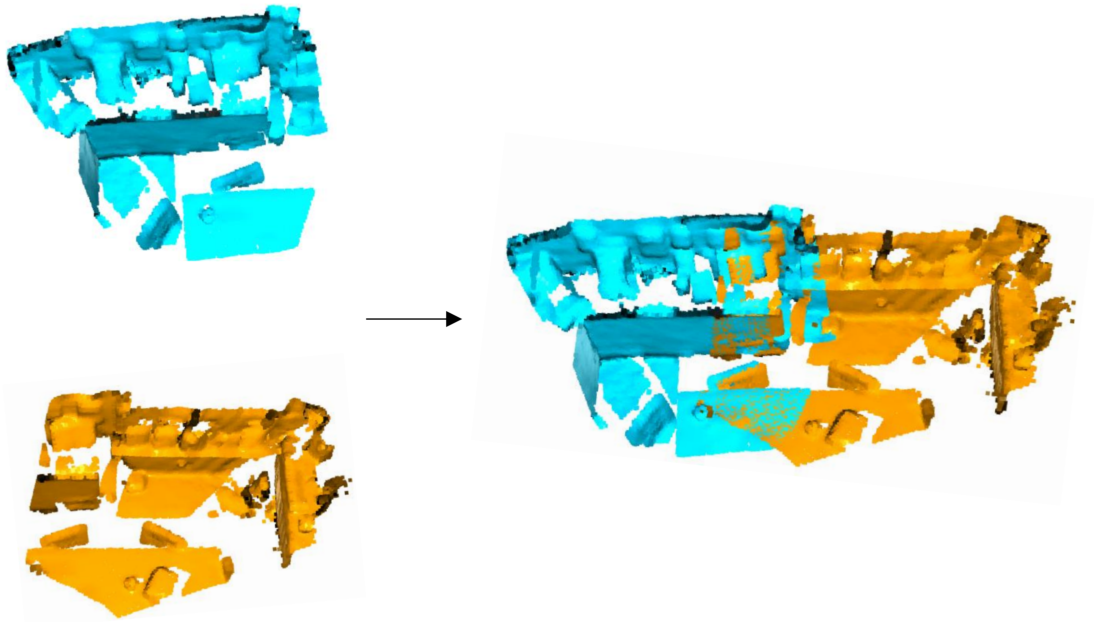
PhD Candidate: Zhaoyi Wang

Supervisor: Prof. Dr. Andreas Wieser

08.10.2024, Graz

Point Cloud Registration

- Our community
  - ICP[1] (and variants), plane-based
  - typically need good initial alignment and high overlap
- Computer vision community
  - DL-based, learn feature descriptiors
  - achieve good registration
    - without initial alignment
    - without high overlap

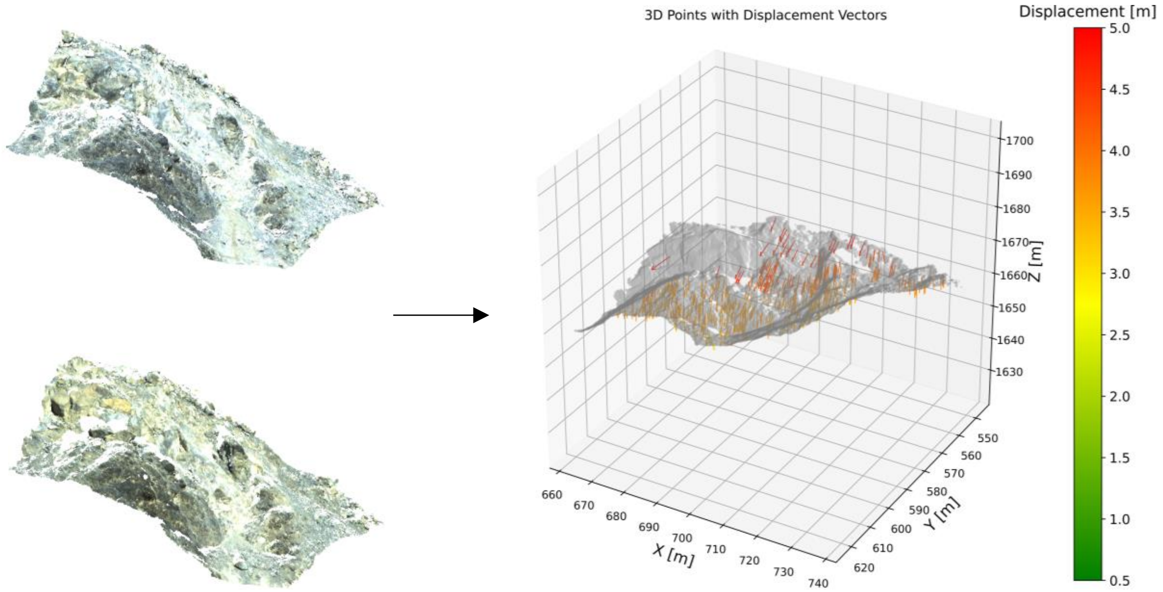


However, the SOTA method is only using geometry – GeoTransformer[2]

1. Besl et al. (1992). Method for registration of 3-D shapes. In Sensor fusion IV: control paradigms and data structures (Vol. 1611, pp. 586-606).
2. Qin et al. (2022). Geometric transformer for fast and robust point cloud registration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11143-11152).

Geomonitoring

- M3C2[1] and its variants
  - only 1D
- Piece-wise ICP[2]
  - sparse 3D
- Hillshade[3]
  - a bit less sparse, but still sparse 3D
- F2S3[4]
  - dense 3D (current SOTA)

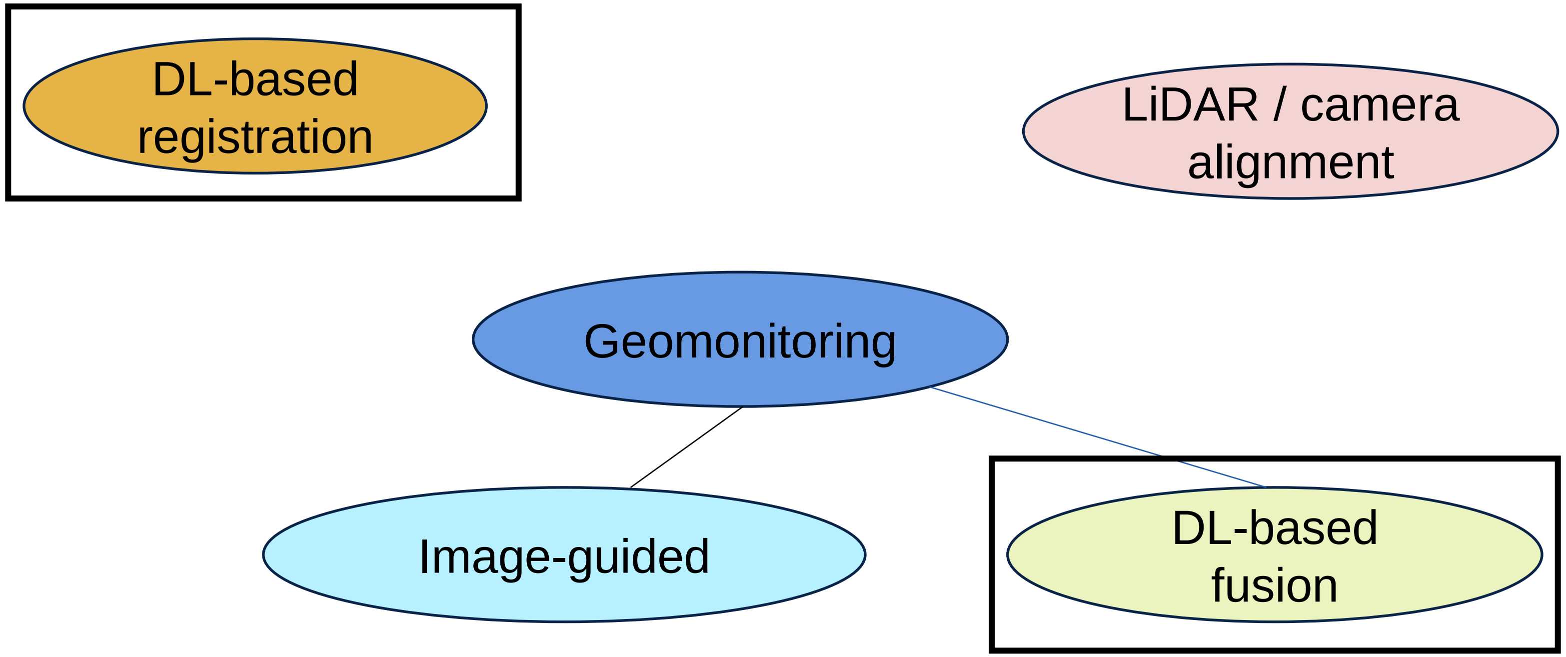


All methods rely only on geometry

1. Lague et al. (2013). Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ). ISPRS journal of photogrammetry and remote sensing, 82, 10-26.
2. Friedli et al. (2016). Identification of stable surfaces within point clouds for areal deformation monitoring. In Proc. of 3rd Joint International Symposium on Deformation Monitoring (JISDM).
3. Holst et al. (2021). Increasing spatio-temporal resolution for monitoring alpine solifluction using terrestrial laser scanners and 3d vector fields. Remote Sensing, 13(6), 1192.
4. Gojcic et al. (2021). Dense 3D displacement vector fields for point cloud-based landslide monitoring. Landslides, 18, 3821-3832.

Contributions

- In general, my PhD topic is about fusing point cloud geometry with RGB image.

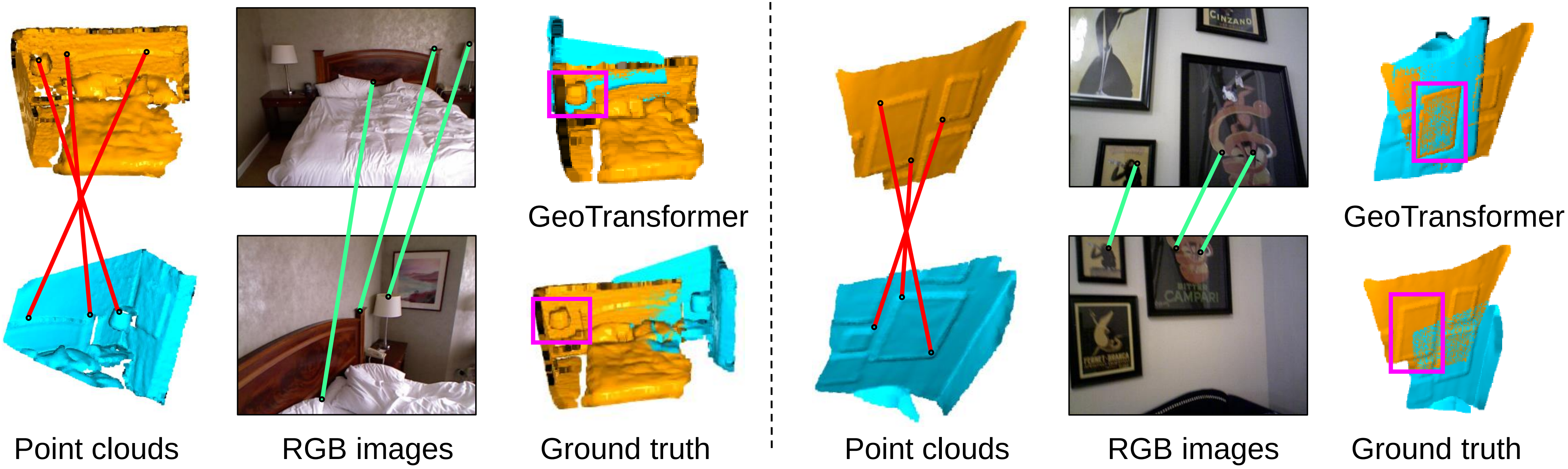


Roadmap

- Point cloud registration using both 2D RGB and 3D point cloud features
- TLS-based Landslide deformation analysis using both 2D and 3D matches

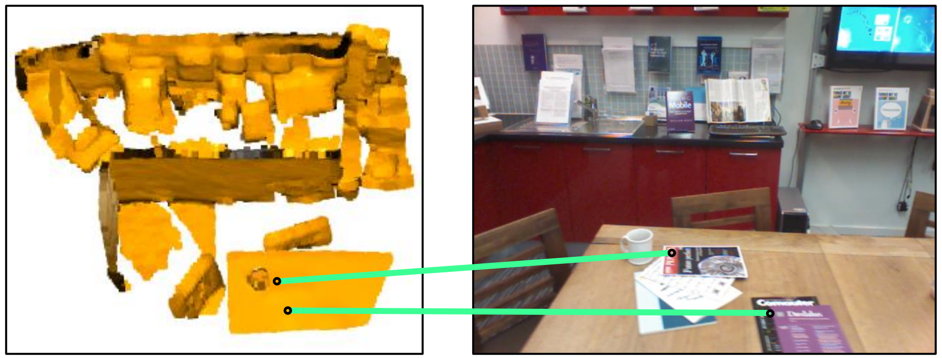
Problem

- **Geometric ambiguity**

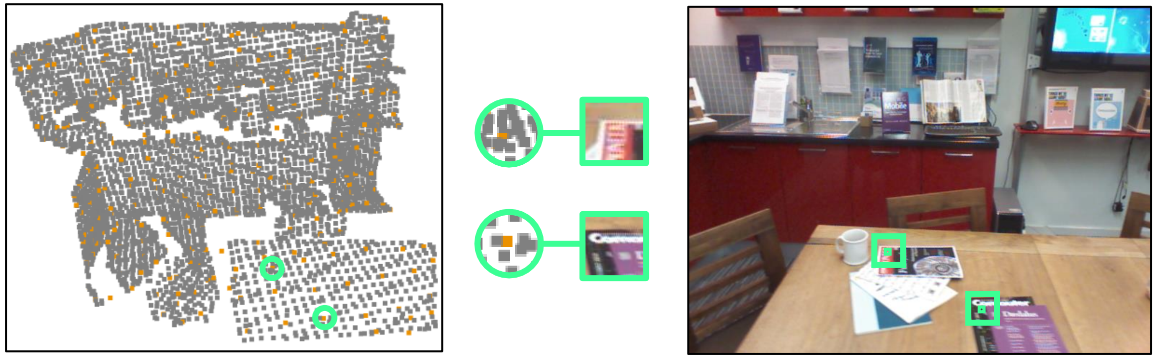


Solution

- **Propose a two-stage cross-modal feature fusion, including:**
  - assigning pixel-wise image features to input point clouds
  - incorporating patch-wise image features with subsampled point features

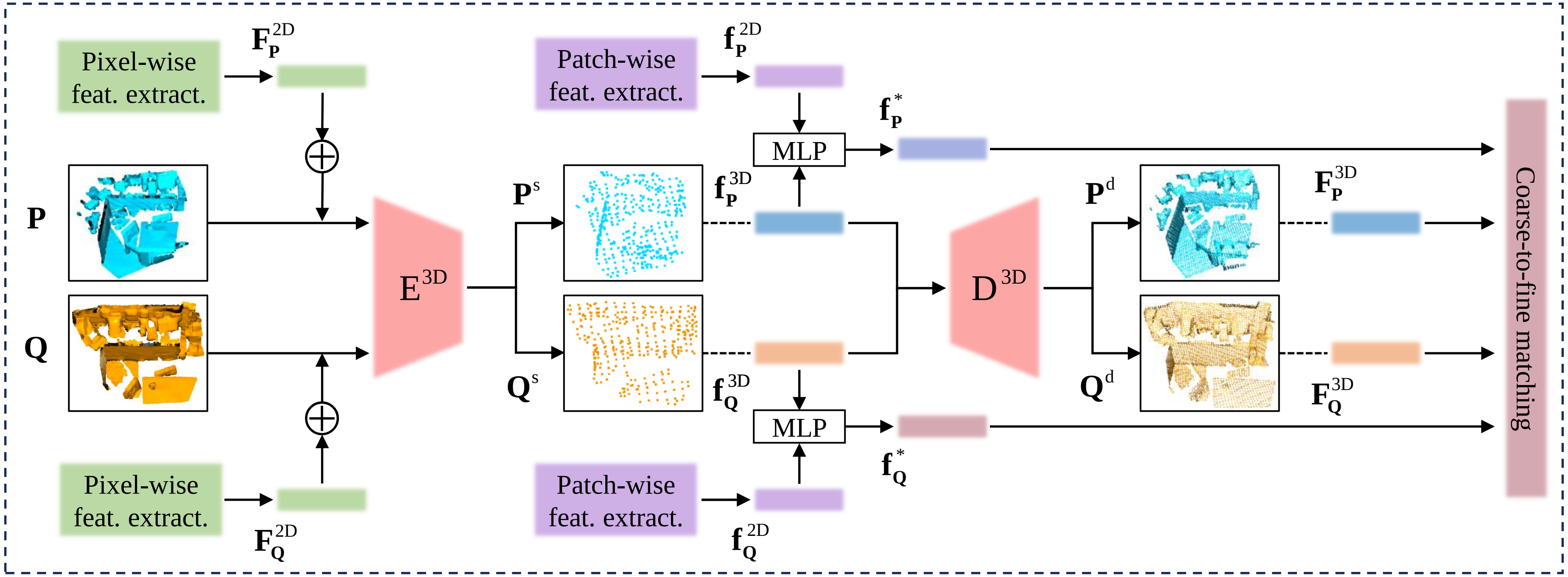


First stage



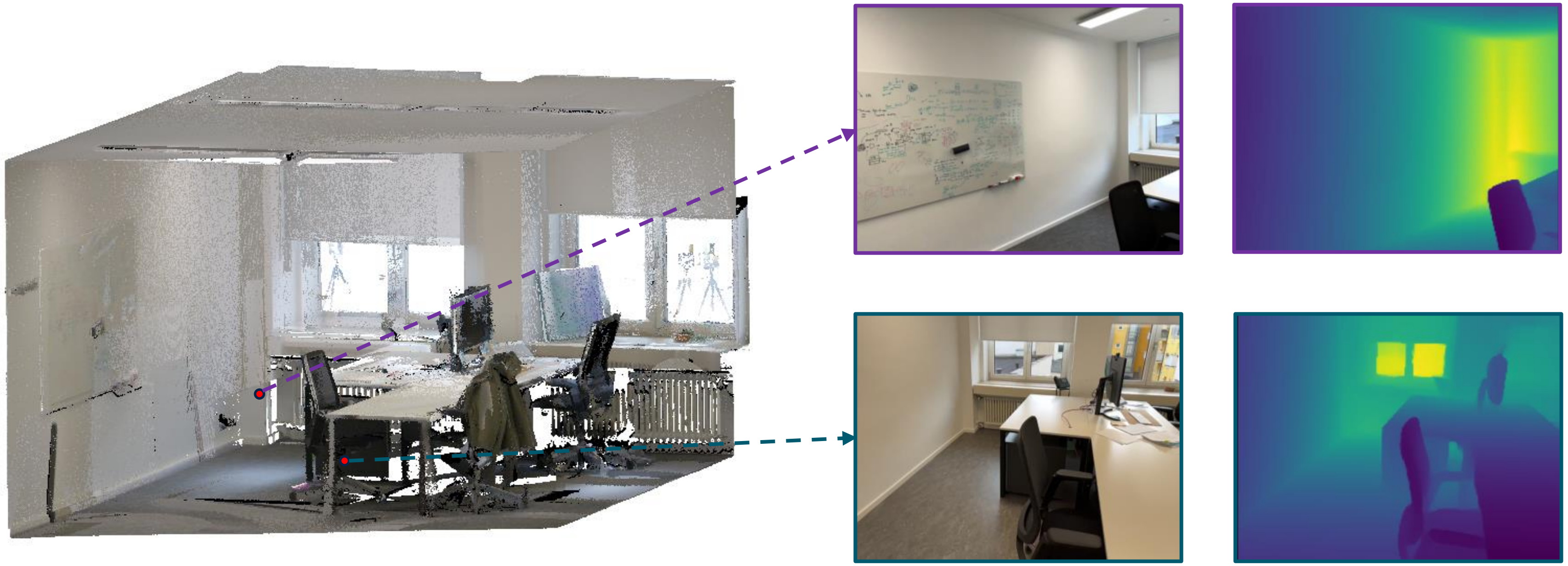
Second stage

Overview of our Method



Datasets

- 3DMatch[1]
- IndoorLRS[2]
- ScanNet++[3]

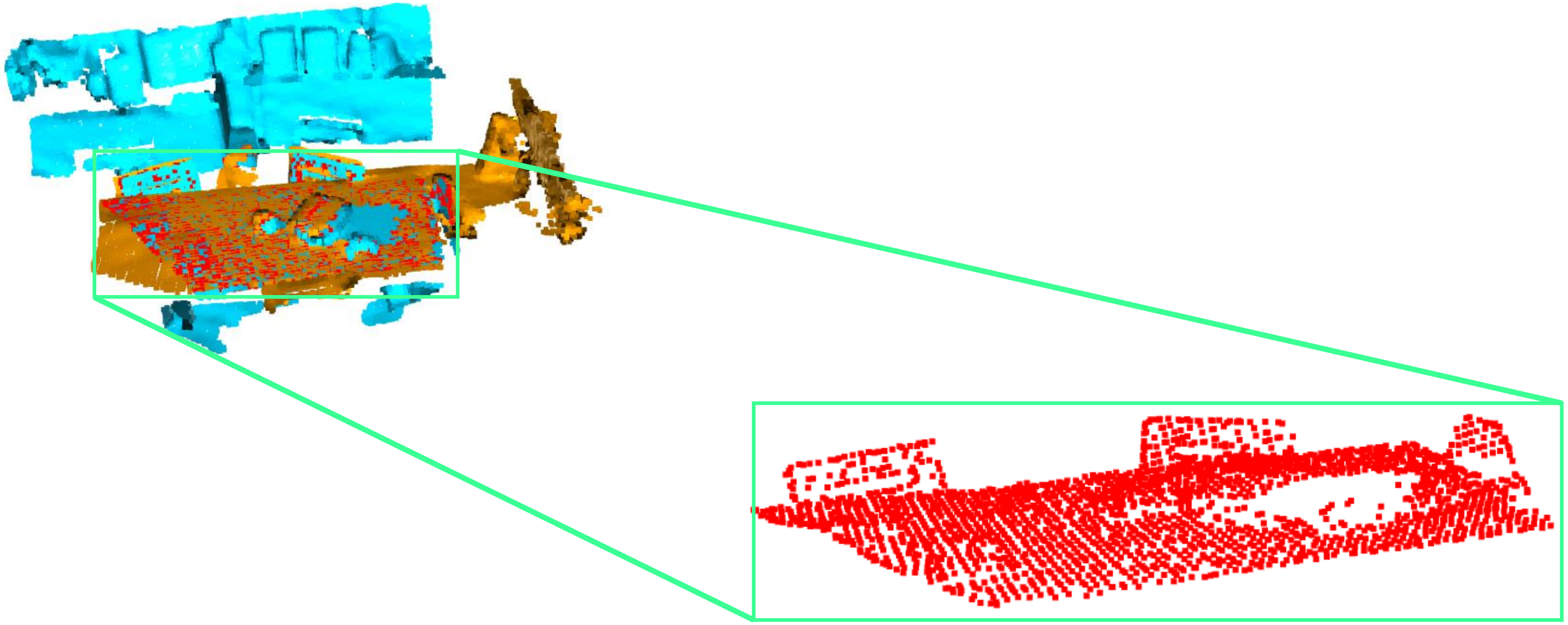


Point cloud RGB Depth

1. Zeng et al. (2017) 3DMatch: Learning local geometric descriptors from rgb-d reconstructions. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1802-1811).
2. Park et al. (2017) Colored point cloud registration revisited. In Proceedings of the IEEE international conference on computer vision (pp. 143-152).
3. Yeshwanth et al. (2023) ScanNet++: A high-fidelity dataset of 3d indoor scenes. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 12-22).

Datasets

- **Create extremely geometry-poor datasets**

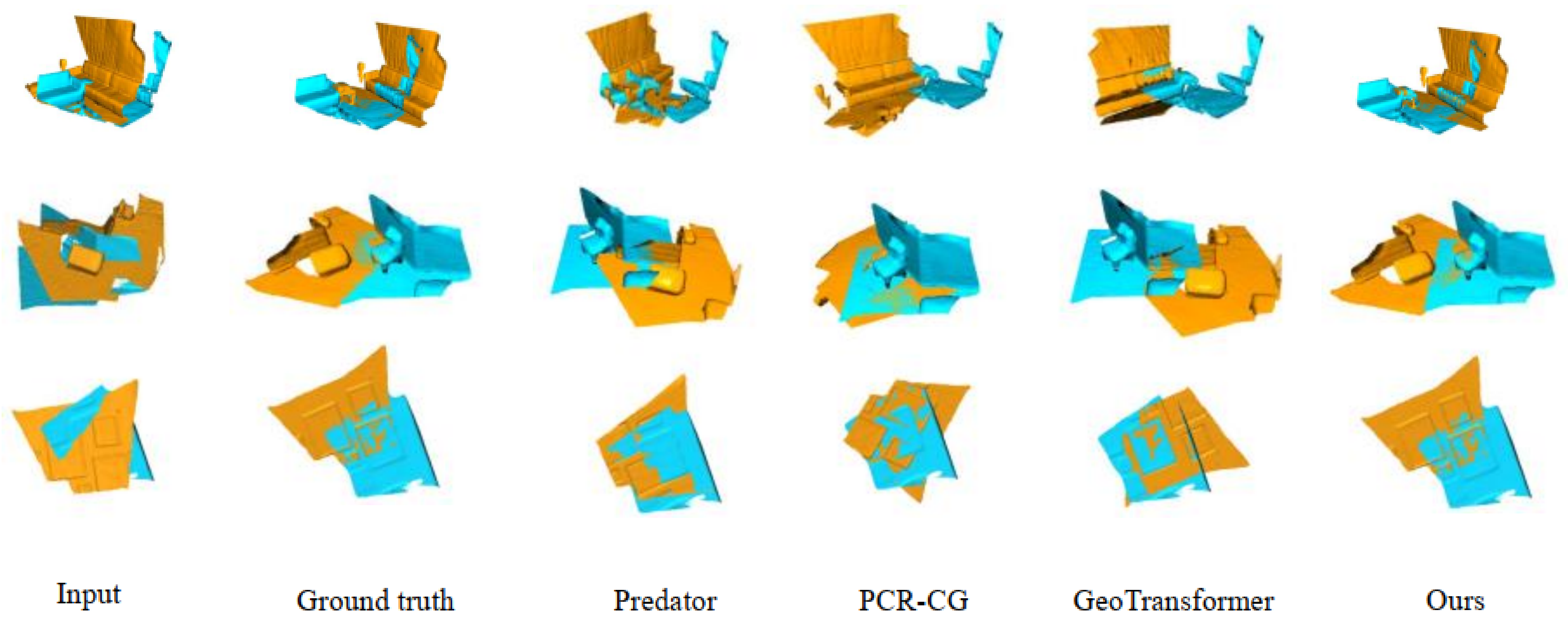


Quantitative Results

Method	3DMatch_Planar			3DLoMatch_Planar		
	FMR (%)↑	IR (%)↑	RR (%)↑	FMR (%)↑	IR (%)↑	RR (%)↑
FCGF [45]	84.5	39.1	45.5	61.6	14.6	27.4
Predator [14]	83.4	39.8	55.5	68.2	19.1	47.9
CoFiNet [15]	89.9	31.4	65.2	73.2	17.4	52.6
GeoTransformer [22]	96.5	57.1	76.4	75.8	32.7	59.5
PCR-CG(retrain) [35]	80.7	38.0	64.1	64.5	17.4	48.1
CoFF(ours)	**99.0**	**60.8**	**90.5**	**87.2**	**36.5**	**70.4**

Method	IndoorLRS_Planar			ScanNet++_Planar		
	FMR (%)↑	IR (%)↑	RR (%)↑	FMR (%)↑	IR (%)↑	RR (%)↑
FCGF [45]	98.6	52.8	60.8	35.9	9.9	15.1
Predator [14]	96.0	48.9	76.9	37.8	9.6	27.2
CoFiNet [15]	97.2	44.2	87.3	53.3	13.7	33.4
GeoTransformer [22]	98.6	64.5	91.2	49.9	14.6	37.6
PCR-CG [35]	91.8	39.6	68.3	67.9	**26.7**	31.9
CoFF(ours)	**99.2**	**78.1**	**94.2**	**70.6**	20.0	**56.0**

Qualitative Results



Take-Home Message

- We propose a cross-modal feature fusion method that
  - specializes for pairwise point cloud registration
  - works well even the geometry is poor

Limitation

- Need accurate alignment between RGB images and point clouds[1].

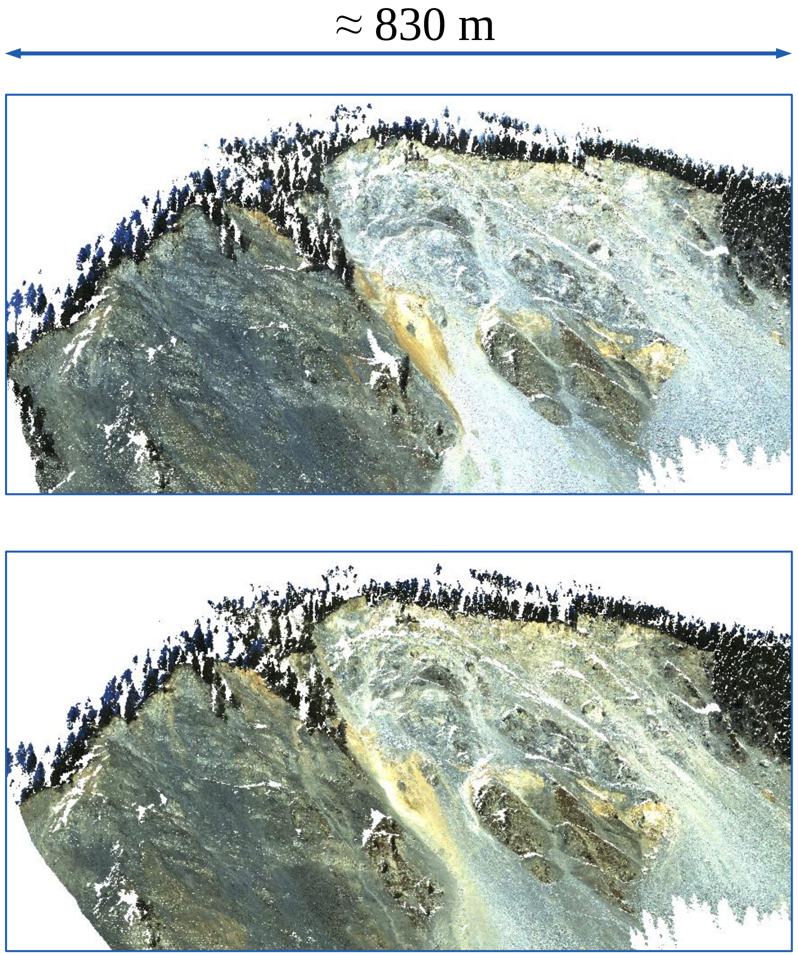
1. Wang, Z., Varga, M., Medic, T., & Wieser, A. (2023). Assessing the alignment between geometry and colors in TLS colored point clouds. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., X-1/W1-2023, 597–604.

Roadmap

- Point cloud registration using both 2D RGB and 3D point cloud features
- TLS-based Landslide deformation analysis using both 2D and 3D matches

Landslide Dataset

- Epochs: 02.2020 and 11.2020
- Point cloud data[1]: long range (1.5+ km) scans with a Riegl VZ-6000
- RGB image: built-in cameras

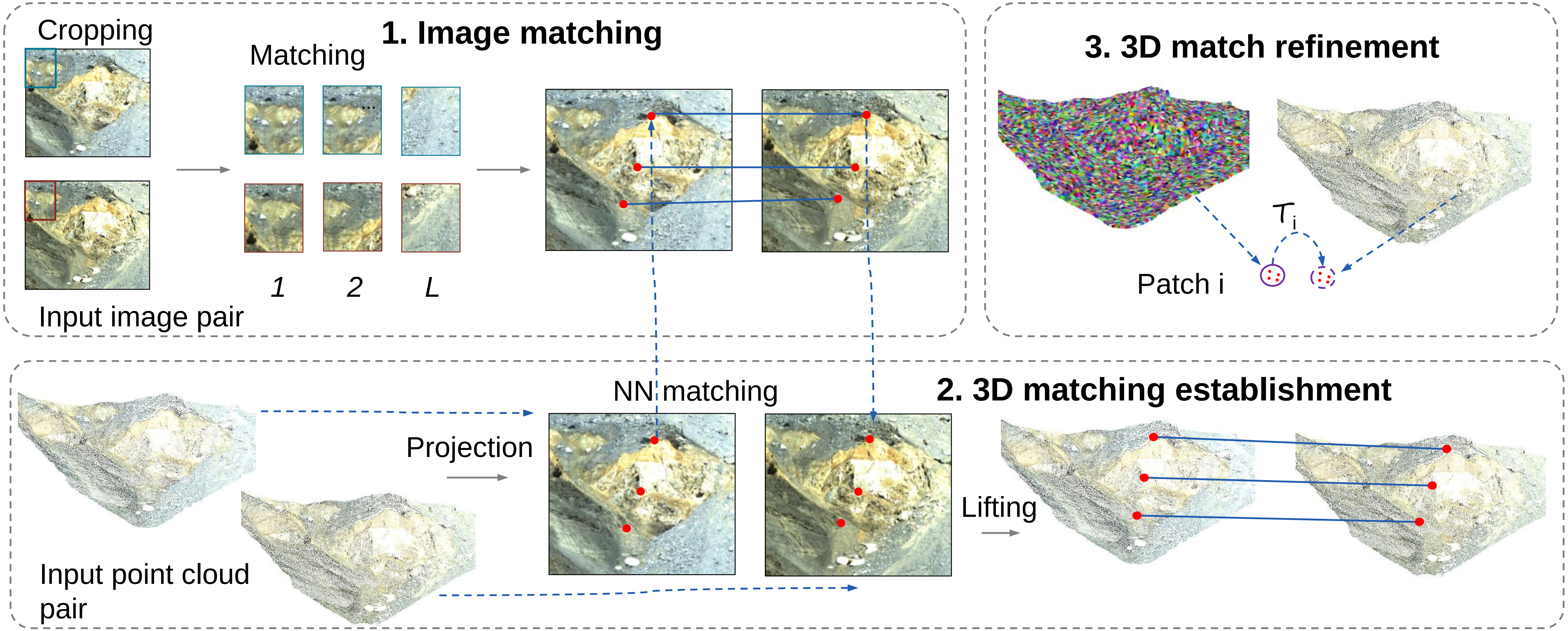


Goal

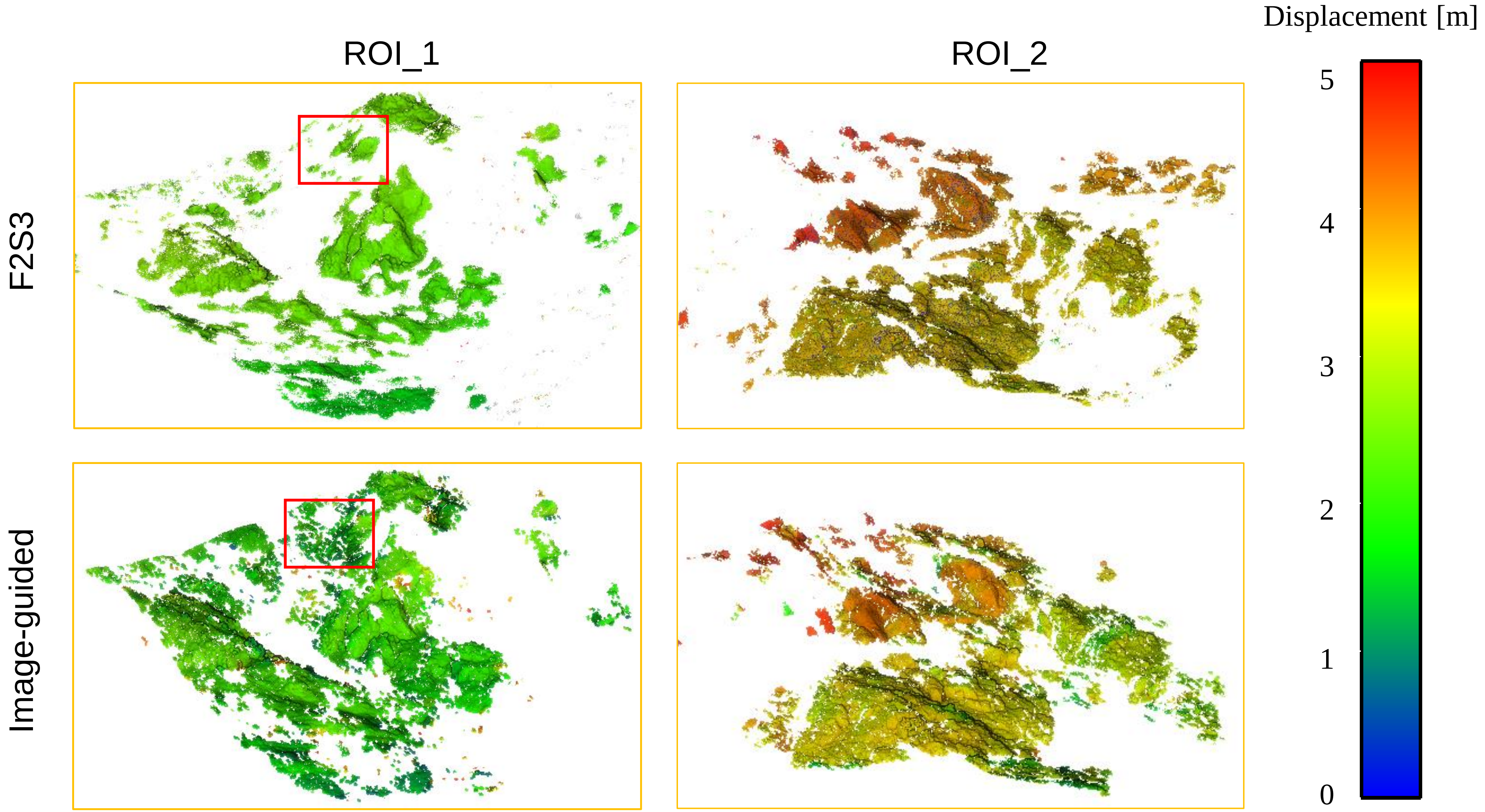
- Estimate 3D displacement vector fields from two epoch point clouds

1. Kenner et al. (2022). The potential of point clouds for the analysis of rock kinematics in large slope instabilities: examples from the Swiss Alps: Brinzauls, Pizzo Cengalo and Spitze Stei. Landslides, 19(6), 1357-1377.

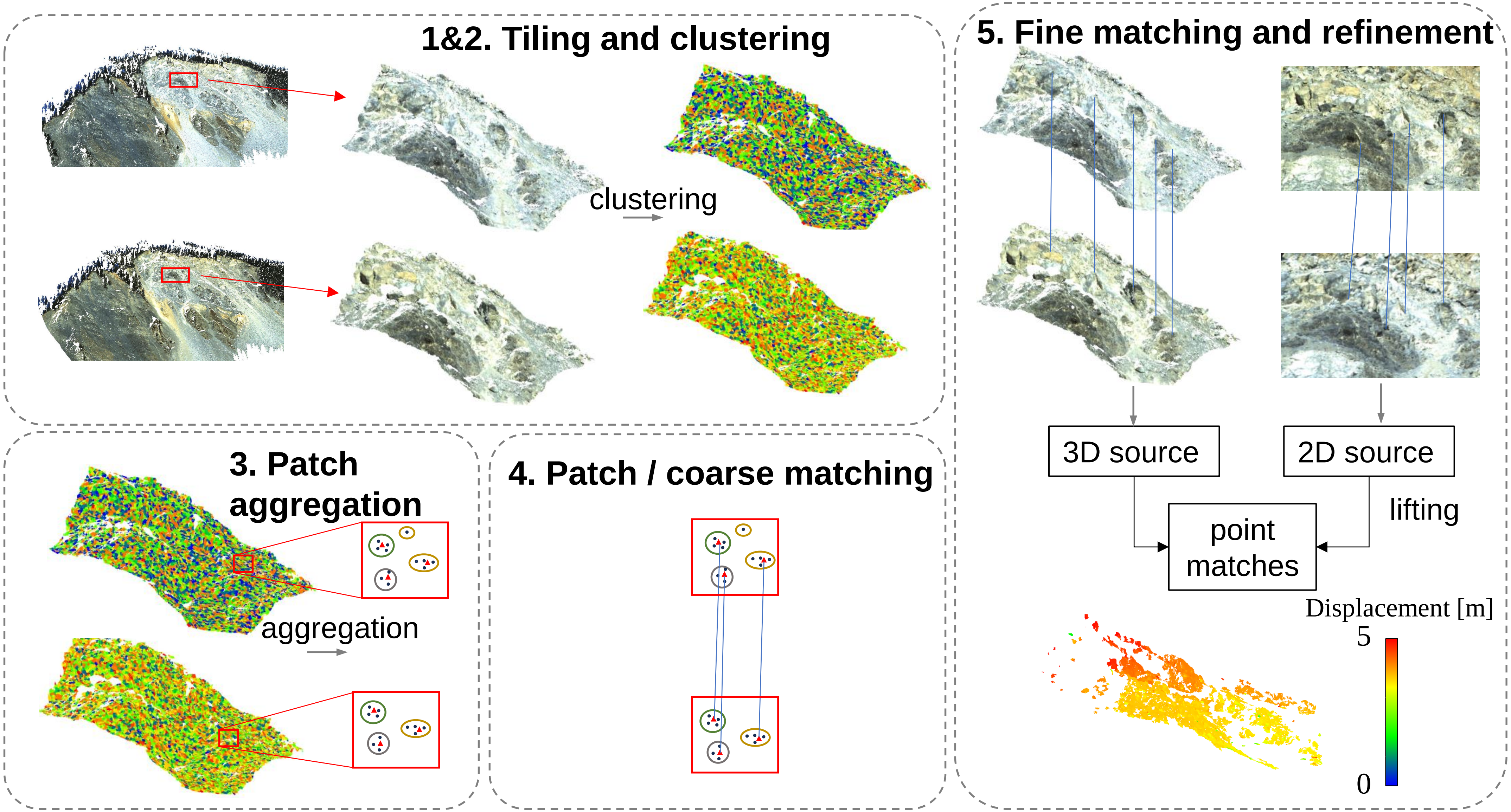
Overview of our Image-Guided Method



Results of F2S3 and our Image-Guided Method



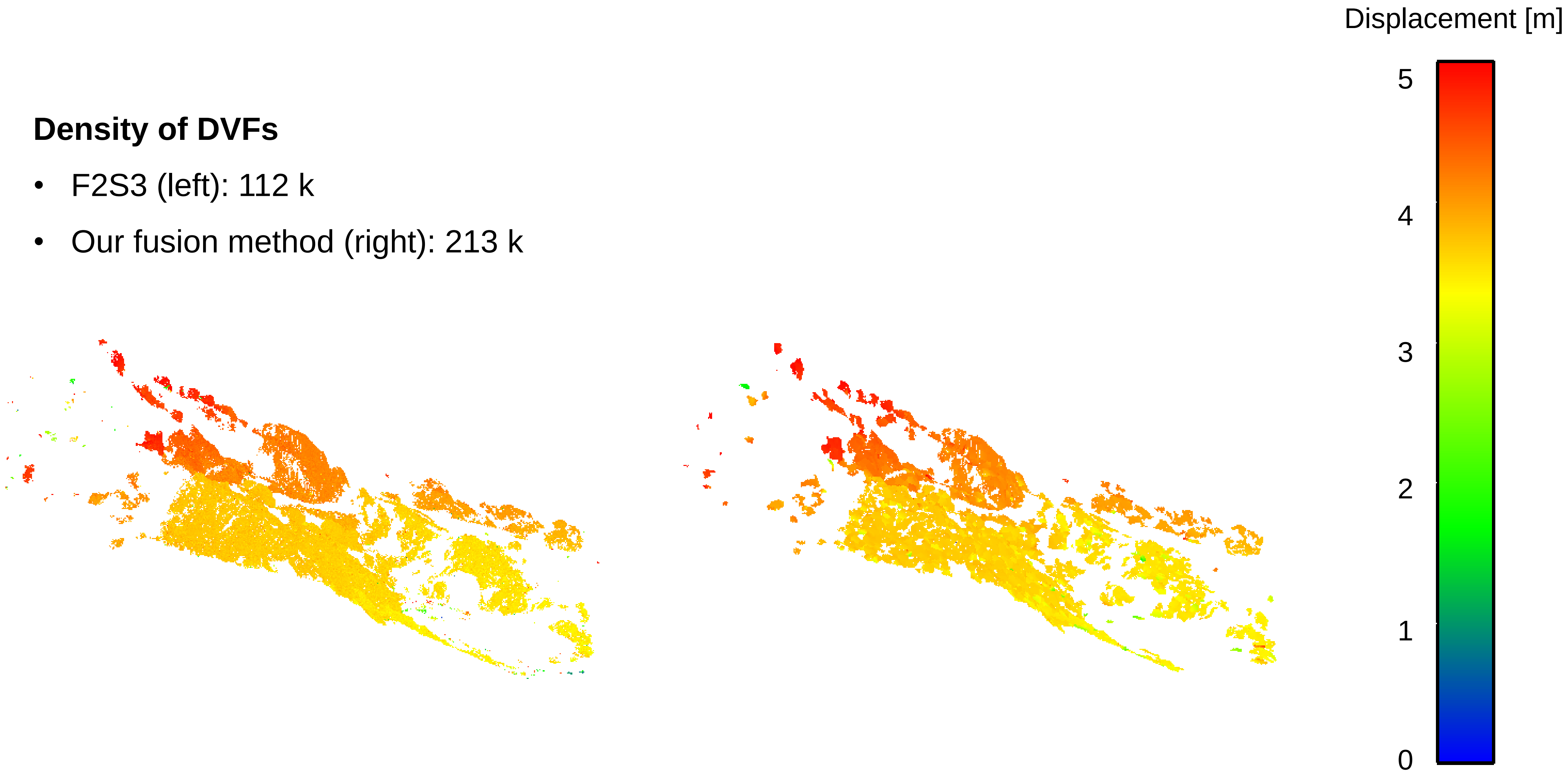
Overview of our DL-based Fusion Method



Results

Density of DVFs

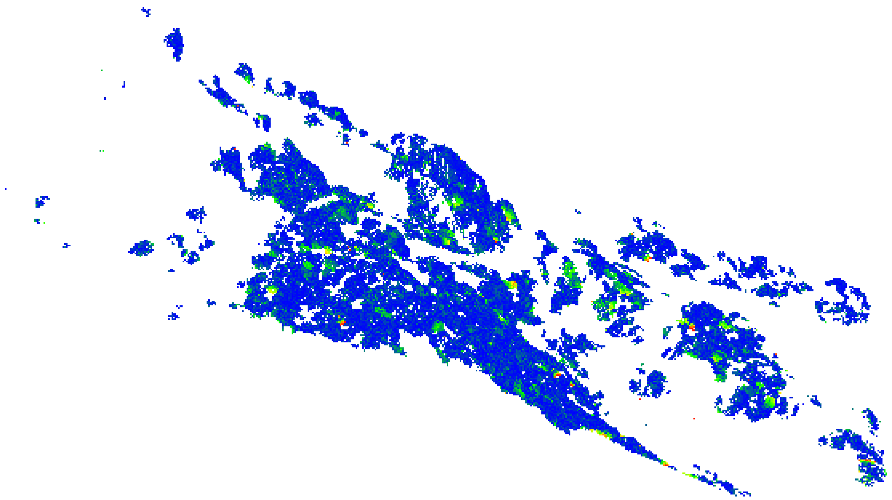
- F2S3 (left): 112 k
- Our fusion method (right): 213 k

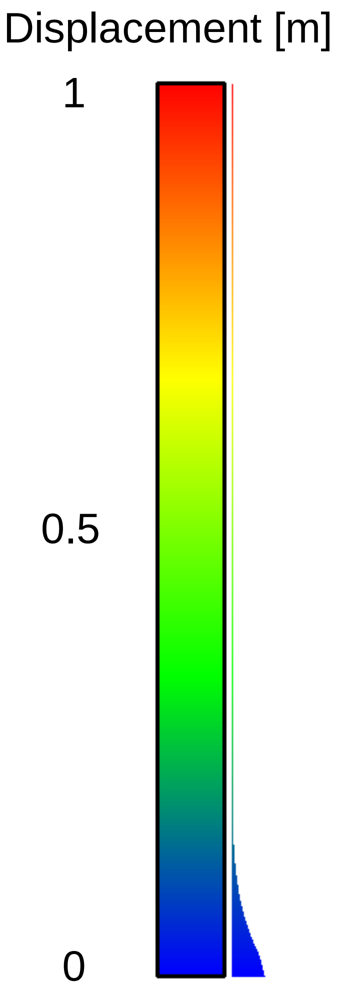


Results

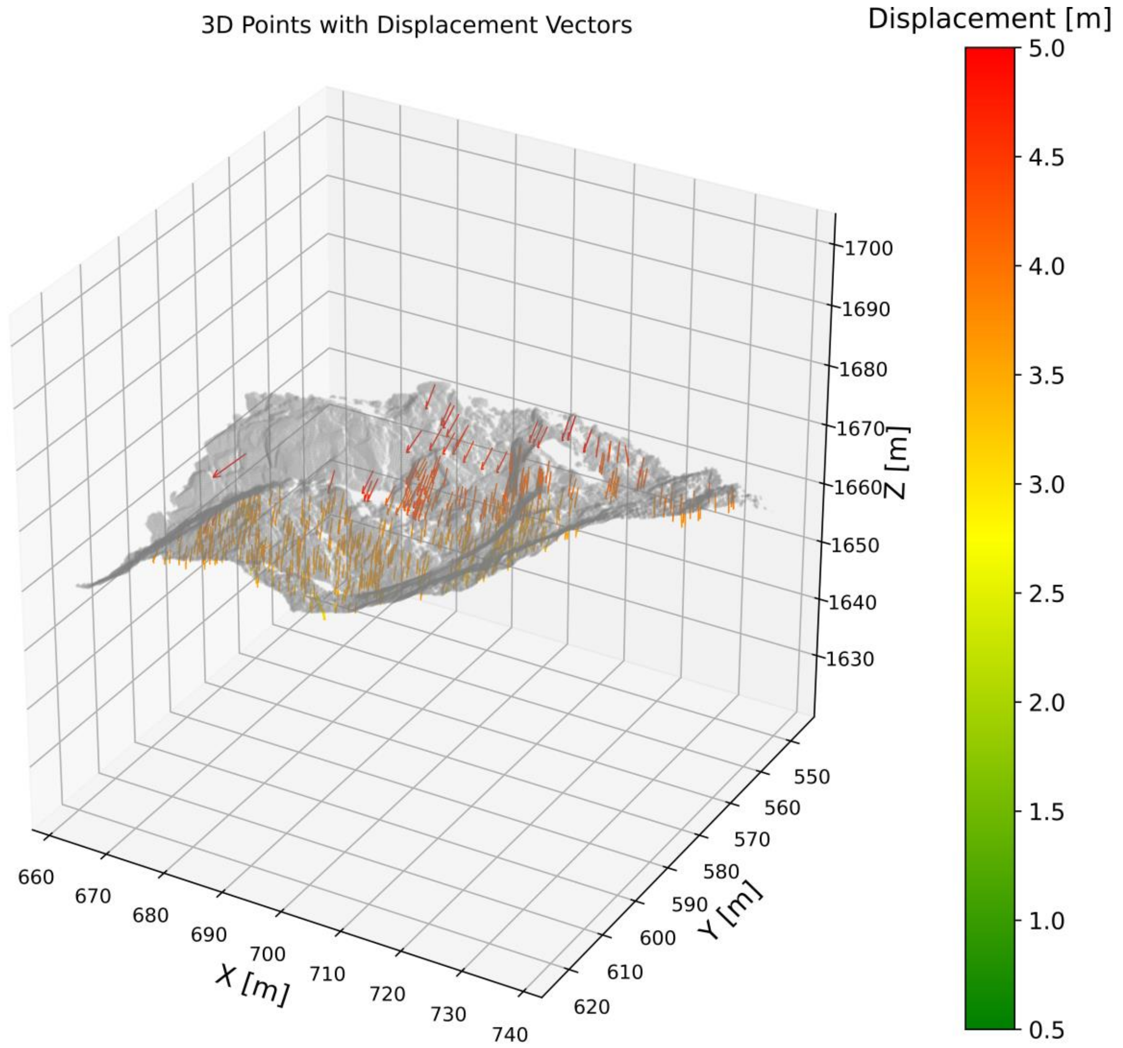
Average deviation

- 0.081 m





Estimated 3D DVFs



Only 0.3% of vectors are visualized for better readability

Take-Home Message

- We are focusing on proposing a DL-based fusion method for estimating 3D DVFs.
  - We hope the results can be more dense and more accurate.

Limitations

- Further comparison with GNSS observations is required for validation.
- Only a limited number of TLS datasets include RGB images, which restricts broader application.

Summary of my Work

List of current works

1. Assessing the alignment between geometry and colors in TLS colored point clouds (published)
2. Cross-modal feature fusion for robust point cloud registration with ambiguous geometry (under review)
3. An approach for RGB-guided dense 3D displacement estimation in TLS-based geomonitoring (in-process)
4. TLS-based Landslide deformation analysis using both 2D and 3D matches (in-process)

Future work

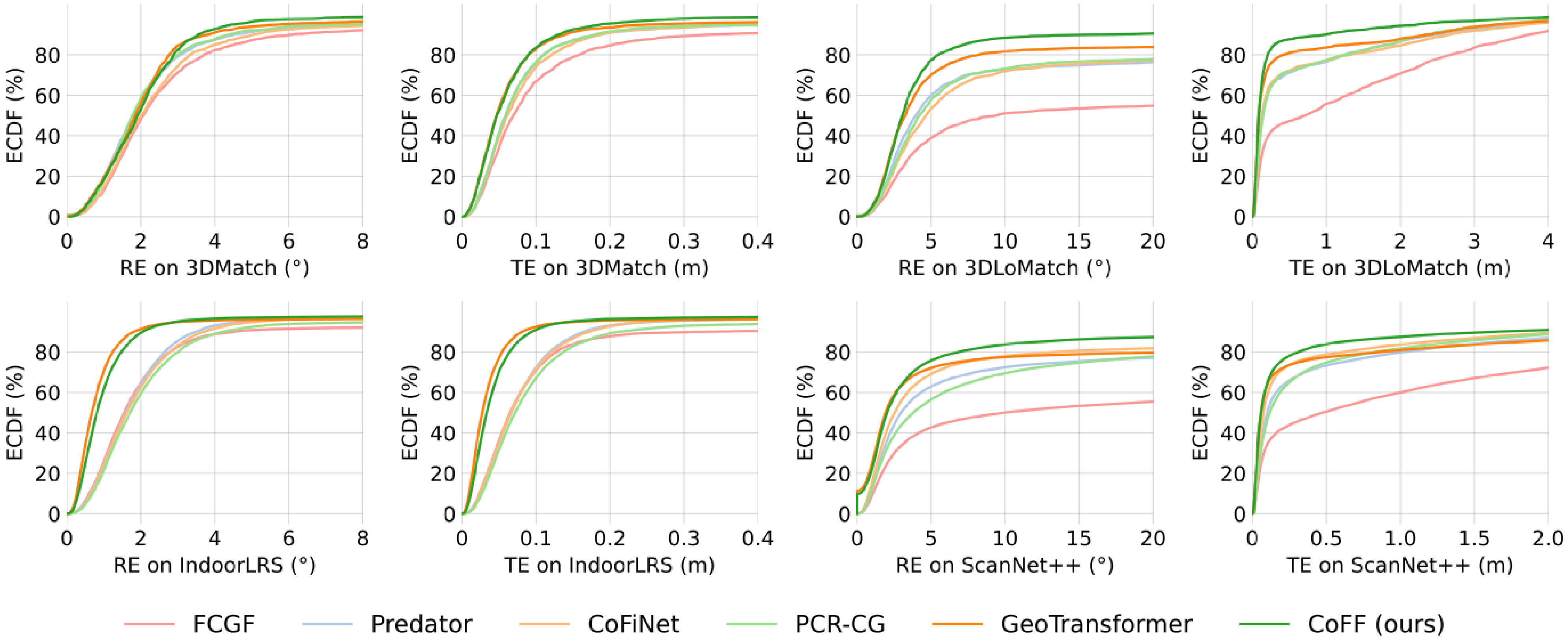
1. Further improve the performance of DL-based fusion method
2. Densify point clouds using RGB images

Thank you!

More information at https://gseg.igp.ethz.ch/people/people-group.html

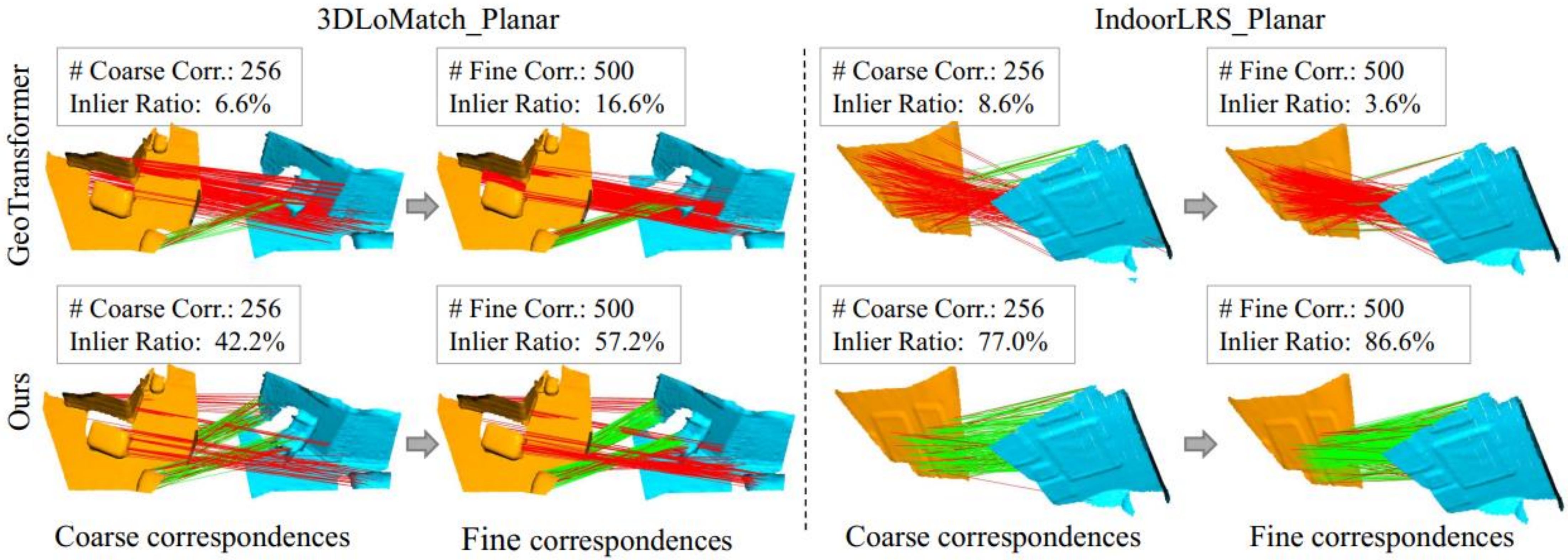
Backup Slides

RE and TE



ECDF (%)
RE on 3DMatch (°)
TE on 3DMatch (m)
RE on 3DLoMatch (°)
TE on 3DLoMatch (m)
RE on IndoorLRS (°)
TE on IndoorLRS (m)
RE on ScanNet++ (°)
TE on ScanNet++ (m)
FCGF
Predator
CoFiNet
PCR-CG
GeoTransformer
CoFF (ours)

Quality of Correspondences



Further Discussion

- Effectiveness of different image features

$\mathbf{F}^{3D}+\mathbf{f}^{3D}$	$\mathbf{F}^{2D}$	$\mathbf{f}^{2D}$	3DMatch FMR(%)↑	3DMatch IR(%)↑	3DMatch RR(%)↑	3DLoMatch FMR(%)↑	3DLoMatch IR(%)↑	3DLoMatch RR(%)↑
✓			97.9	71.9	92.0	88.3	43.5	74.0
✓	✓		<u>98.8</u>	**76.2**	<u>94.2</u>	90.1	**48.2**	<u>78.0</u>
✓		✓	98.3	68.7	91.5	92.7	40.5	77.8
✓	✓	✓	**99.5**	<u>74.3</u>	**95.9**	**93.6**	<u>47.4</u>	**81.6**

Evaluation Metrics

- Inlier Ratio (IR): denotes the fraction of inlier correspondences among all estimated dense point correspondences.

$$\mathrm{IR} = \frac{1}{|\boldsymbol{\Omega}^{\mathrm{f}}|} \sum\nolimits_{(\mathbf{p}^{\mathrm{d}}_{m_j}, \mathbf{q}^{\mathrm{d}}_{n_j}) \in \boldsymbol{\Omega}^{\mathrm{f}}} \left(\left\| \mathbf{R}^*_i \cdot \mathbf{p}^{\mathrm{d}}_{m_j} + \mathbf{t}^*_i - \mathbf{q}^{\mathrm{d}}_{n_j} \right\|_2^2 < \Delta_r \right)$$

- Feature Match Recall (FMR): denotes the fraction of point cloud pairs whose IR is above 5%.

- Rotation Error (RE): represents the rotation deviation between estimated and ground-truth rotation matrices:

$$\mathrm{RE} = \arccos\left(\frac{\mathrm{trace}((\mathbf{R}^\dagger)^T \cdot \mathbf{R}^* - 1)}{2} \right)$$

- Translation Error (TE): represents the Euclidean distance between estimated and ground-truth translation vectors:

$$\mathrm{TE} = \|\mathbf{t}^\dagger - \mathbf{t}^*\|_2$$

- Registration Recall (RR): for three indoor datasets is defined as the fraction of point cloud pairs whose RMSE is below an acceptance radius 0.2 m.

$$\mathrm{RMSE} = \sqrt{\frac{1}{|\overline{\boldsymbol{\Omega}}|} \sum\nolimits_{(\mathbf{p}^{\mathrm{d}}_{m_j}, \mathbf{q}^{\mathrm{d}}_{n_j}) \in \overline{\boldsymbol{\Omega}}} \left(\left\| \mathbf{R}^\dagger_i \cdot \mathbf{p}^{\mathrm{d}}_{m_j} + \mathbf{t}^\dagger_i - \mathbf{q}^{\mathrm{d}}_{n_j} \right\|_2^2 \right)}$$

Failure Cases

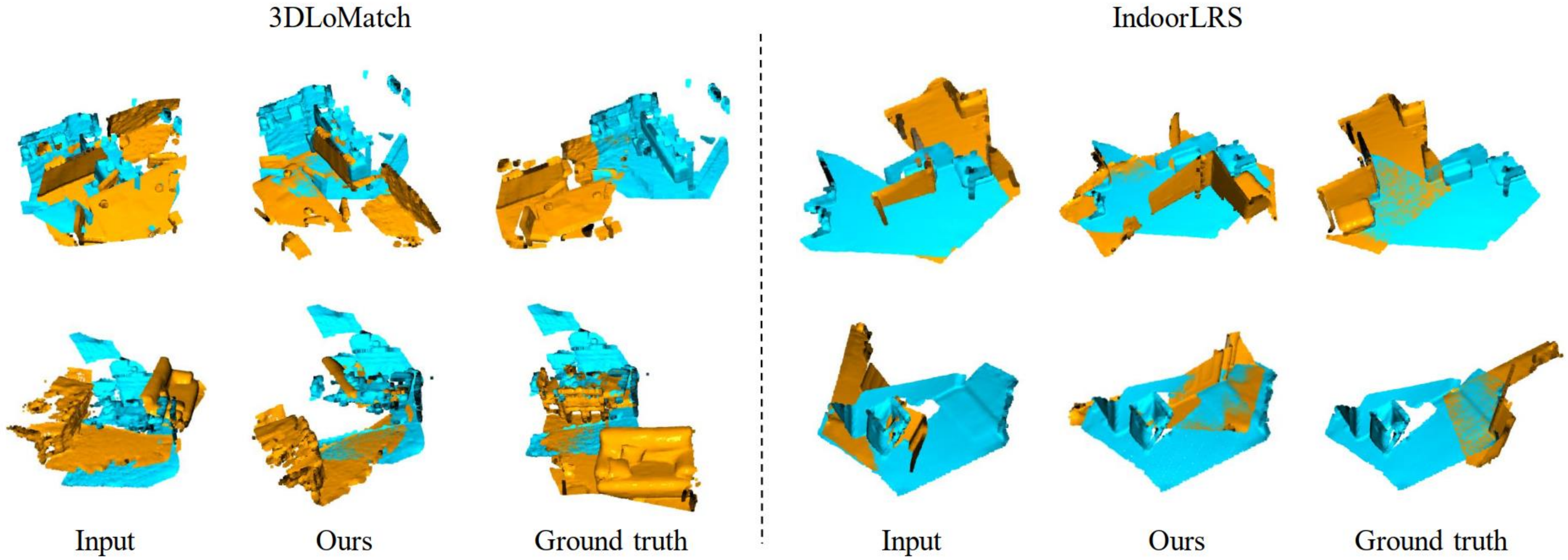
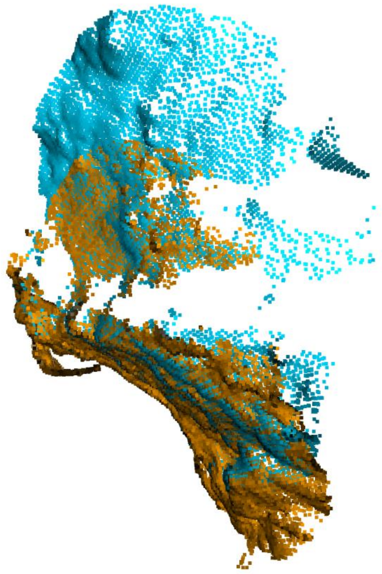
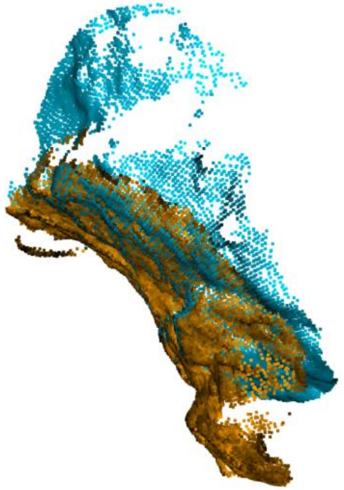


Figure D.11: Failure cases on 3DLoMatch and IndoorLRS datasets, where both geometry and color information are ambiguous.

Ground-truth DVFs

- RMSE: 0.05 m
- visually check:
- blue: source; yellow: target



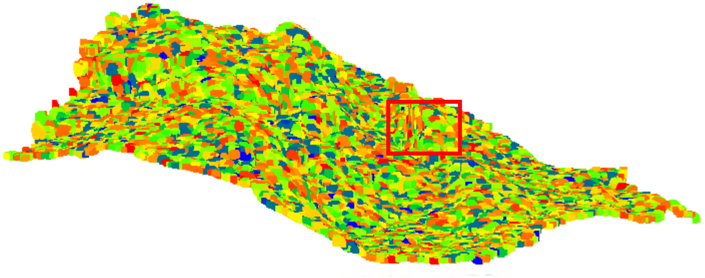


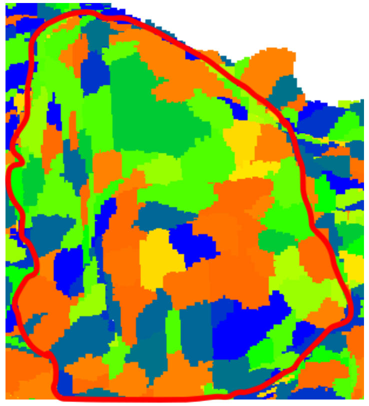
Clustering using a segmentation algorithm

Used features include:

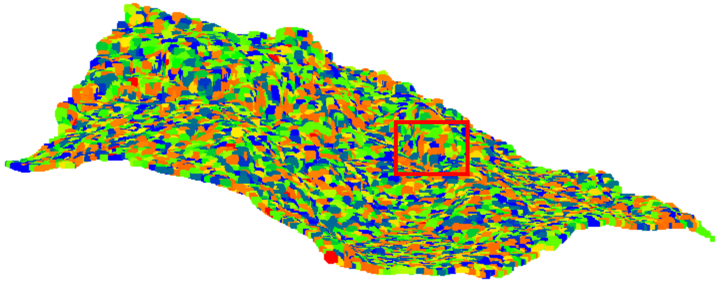
- Geometric: linearity, planarity, scattering
- Radiometric: intensity, or grayscaled RGB

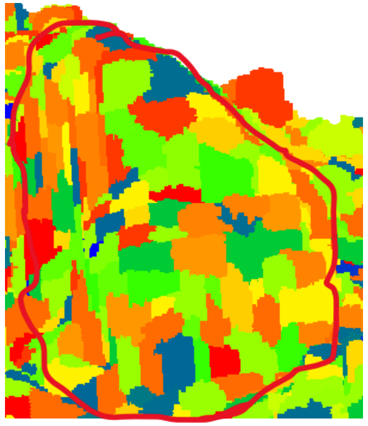
source





target

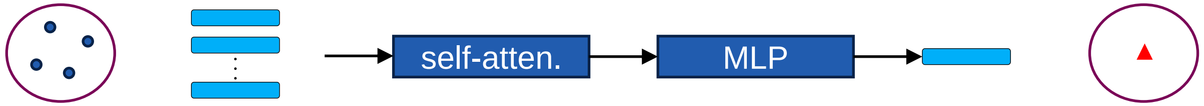


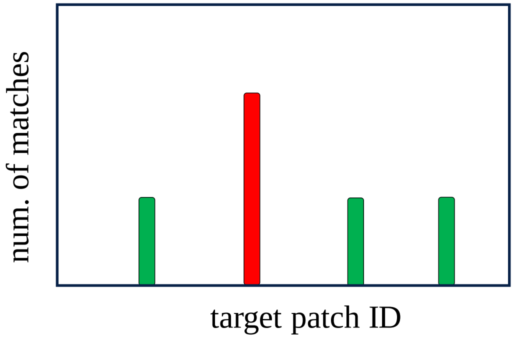


clustering results

Coarse-to-Fine Matching

- Goal of coarse matching: establish cluster matches between two epochs
- Steps:
- use shape-invariant information
- self-attention + MLP, and random noise to make the feature are transformation-invariant
- additionally, count the point matches from 2D source
- find the most matched cluster





ECDF of Point Errors