A. ADDITIONAL EXPERIMENTS

A.1. Overlap investigation

To extend our investigation beyond the overlap ratio range of 0.2 to 0.8 discussed in the main paper, we analyzed the impact of both lower and negative overlap ratios. Since our method is based on reconstructing the scene without requiring access to the original training images, it remains robust to variations in overlap, provided that some elements of the scene remain co-visible in the submap frames. This property allows us to examine cases where the two submaps are increasingly separated in time, leading to what we define as *negative overlap ratios*.

Negative overlap ratios are characterized by the percentage of frames that separate the two submaps. More formally, for a given submap containing N frames, a negative overlap of -1.0 corresponds to a separation of N frames between the starting points of the two submaps. In the case of the ScanNet and TUM RGB-D datasets, where we typically use 400-frame submaps, a -0.1 overlap means the second submap begins 40 frames after the first submap ends, ensuring no direct overlap between them. The Fig. 1 illustrates this overlap negative ratio between two captures on one continuous scene. Larger negative overlap ratios (e.g., -0.4 or beyond) further increase the temporal gap, meaning the two submaps originate from entirely distinct portions of the sequence.

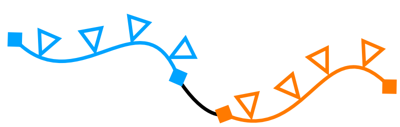

Fig. 1. Illustration of two disctinct sequences (blue sequence and orange sequence) with an negative overlap ratio (represented in black) on the figure.

This analysis is particularly important because, in real-world scenarios, mapping and localization systems may need to establish correspondences between spatially distant or temporally disjoint observations. Traditional methods relying on direct feature matching between overlapping images often struggle in such cases, whereas our approach, by leveraging scene-level reconstruction, can still establish meaningful correspondences as long as there are shared structural elements in the environment.

Table 1 presents the results obtained on sequence *rgbd_dataset_freiburg2_desk* of the TUM RGB-D dataset, evaluating our method under various negative overlap ratios with a constant rotation of 10 degrees between the two submaps and a resolution of 0.005m per pixel.

Results show that the proposed method demonstrates strong robustness across a wide range of overlap ratios, successfully aligning submaps even under challenging conditions. For high overlap ratios (0.8 to 0.4), the method performs exceptionally well, achieving low rotational (RRE) and translational (RTE) errors. Notably, at 0.6 overlap, the RRE is as low as 0.0011, highlighting the method's ability to achieve near-perfect registration when a sufficient portion of the scene is shared. This confirms that the approach effectively utilizes overlapping information to optimize submap alignment. As the overlap decreases to 0.2 and 0.1, the method continues to provide meaningful results, although with slightly increased errors. For instance, at 0.1 overlap, RRE remains below 1.0 (0.509), and RTE is still relatively small (0.0516), indicating that the method is still capable of extracting useful correspondences even with limited shared visual data. A particularly interesting result is that even at 0.0 overlap and beyond into negative overlap ratios (-0.1 to -0.4), the method still manages to produce alignments. While errors naturally increase as the overlap vanishes, the fact that meaningful transformations can still be estimated suggests that the approach effectively leverages structural consistencies in the scene. Fig. 2 illustrates the worst-case registration scenarios at overlap ratios of -0.1 and at -0.4. At -0.3 and -0.4 overlap, RRE stabilizes at 0.900 and 0.376, respectively, indicating that in certain cases, even temporally disjoint submaps contain enough co-visible elements for successful registration.

Table 1. Performance evaluation on `rgbd_dataset_freiburg2_desk` with varying overlap ratio. Lower is better for RRE and RTE.

Overlap ratio	RRE	RTE
0.8	0.119	0.0075
0.7	0.315	0.0977
0.6	0.0011	0.0125
0.5	0.317	0.0075
0.4	0.142	0.0131
0.3	1.080	0.0198
0.2	1.245	0.0243
0.1	0.509	0.0516
0.0	0.885	0.6545
-0.1	5.645	1.0068
-0.2	0.470	0.0435
-0.3	0.900	0.2069
-0.4	0.376	0.3086

A.2. Impact of the features on MapClosure

We also investigated the integration of more recent feature detectors and matchers within the MapClosure approach to ensure a fair comparison. In particular, we tested the use of SuperPoint + LightGlue (SP+LG) in the MapClosure pipeline. Table 2 presents the results of this evaluation on the TUM RGB-D dataset. The results demonstrate that our method con-

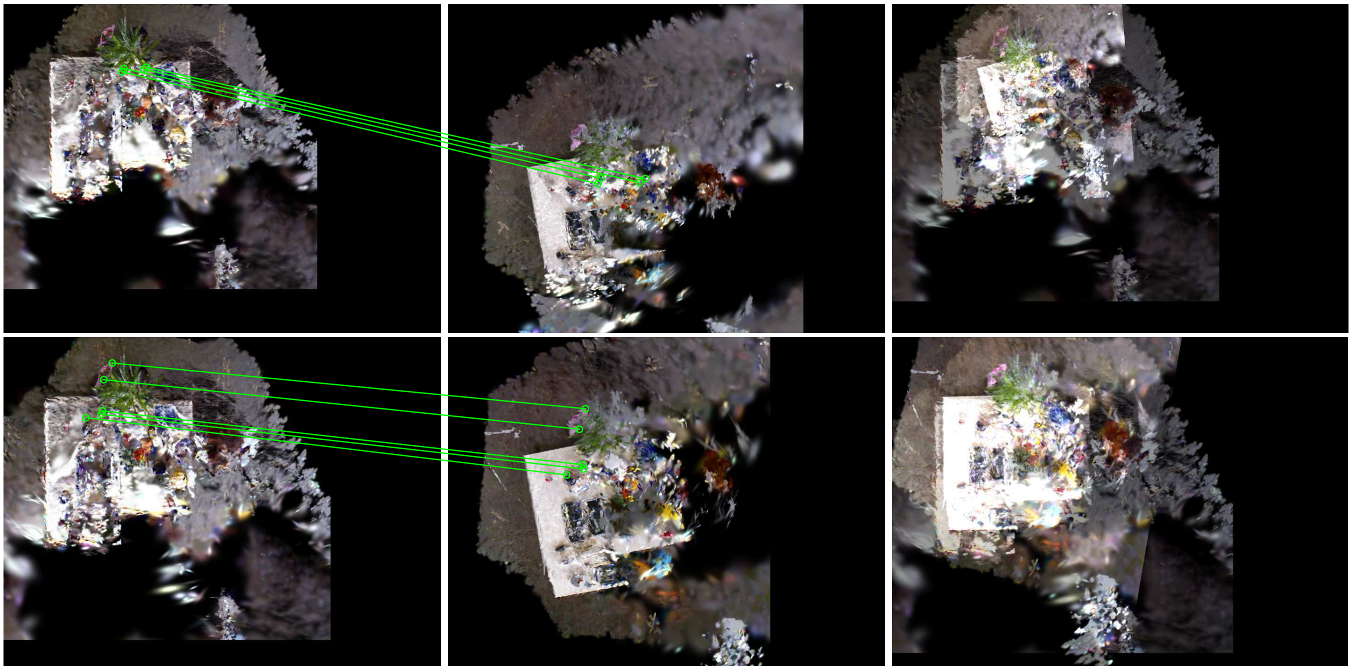

Fig. 2. Visualization of the worst-case matching scenarios. The top images represents the matching and registered scene for an overlap ration of -0.1 while the bottom images represent the matching and registration for an overlap ratio of -0.4. The inliers of the RANSAC estimation are shown in green.

sistently outperforms MapClosure across all tested configurations. Notably, while the incorporation of SP+LG proves to make MapClosure less reliable overall, as indicated by lower success rates (SR) and higher registration errors (RRE, RTE) compared to our approach.

On the other hand, our method demonstrates strong robustness to changes in resolution. Our approach maintains consistently high success rates (SR) and low registration errors (RRE, RTE) across different resolution levels, highlighting the reliability of our method in varying conditions.

A.3. Impact of the initial capture

In this section, we present additional quantitative results by testing different amounts of input initial points for GS training, ranging from 100k to 10k initial points, on the first ScanNet scene with a fixed overlap of 0.6 and a 15° rotation. These experiments help evaluate the performance of our method under varying levels of input data. The results in Table 3 suggest that the initial number of points used for the GS optimization does not drastically impact the overall performance of our method. While minor variations are observed in both Relative Rotation Error (RRE) and Relative Translation Error (RTE), the differences remain relatively small across different input sizes. Notably, reducing the number of initial points from 100k to 50k results in a sharp drop in RRE, but beyond this point, further reductions to 30k and 10k do not lead to significant degradation. Similarly, the RTE remains stable across

Table 2. Evaluation of registration performance across different overlap ratios and with a random rotation amount up to 30 degrees on TUM RGB-D. Lower is better for RRE and RTE while higher is better for SR.

Overlap Ratio	Method	SR↑	RRE↓	RTE↓
0.2	MapClosure-ORB-0.01	0.500	1.208	0.030
	MapClosure-ORB-0.005	0.333	1.482	0.056
	MapClosure-SP+LG-0.01	0.166	0.768	0.022
	MapClosure-SP+LG-0.005	0.000	-	-
	Ours-0.01	1.000	0.351	0.031
	Ours-0.005	1.000	0.802	0.035
0.4	MapClosure-ORB-0.01	0.666	2.611	0.067
	MapClosure-ORB-0.005	0.666	1.039	0.041
	MapClosure-SP+LG-0.01	0.166	0.600	0.017
	MapClosure-SP+LG-0.005	0.000	-	-
	Ours-0.01	1.000	0.485	0.013
	Ours-0.005	1.000	0.262	0.015
0.6	MapClosure-ORB-0.01	1.000	1.038	0.055
	MapClosure-ORB-0.005	1.000	1.123	0.066
	MapClosure-SP+LG-0.01	0.166	0.078	0.010
	MapClosure-SP+LG-0.005	0.000	-	-
	Ours-0.01	1.000	0.352	0.034
	Ours-0.005	1.000	0.119	0.026
0.8	MapClosure-ORB-0.01	0.833	1.598	0.064
	MapClosure-ORB-0.005	1.000	1.024	0.058
	MapClosure-SP+LG-0.01	0.166	0.636	0.017
	MapClosure-SP+LG-0.005	0.000	-	-
	Ours-0.01	1.000	0.181	0.023
	Ours-0.005	1.000	0.109	0.034

different input sizes, with only a slight increase when using 10k initial points. These results indicate that our approach remains robust even with a reduced number of points, suggesting that the method does not heavily depend on the initial points density to maintain accurate estimations.

Input Data (initial points)	**RRE**	**RTE**
100k	0.4702	**0.004**
50k	**0.001**	0.005
30k	0.057	0.009
10k	0.199	0.020

Table 3. Performance of our method with varying amounts of input data (initial points) for the first ScanNet scene.

B. ADDITIONAL QUALITATIVE RESULTS

In this section, we introduce more quantitative results using our method on the ScanNet and TUM RGB-D datasets. Specifically, we present additional keypoint matching results, showcasing MapClosure's ability to capture key spatial structures and achieve robust matches on the ScanNet dataset illustrated Fig. 3. Additionally Fig. 4 and 5 provide a deeper look at the performance of our approach, further illustrating its effectiveness with different overlap ratios and scene complexities.

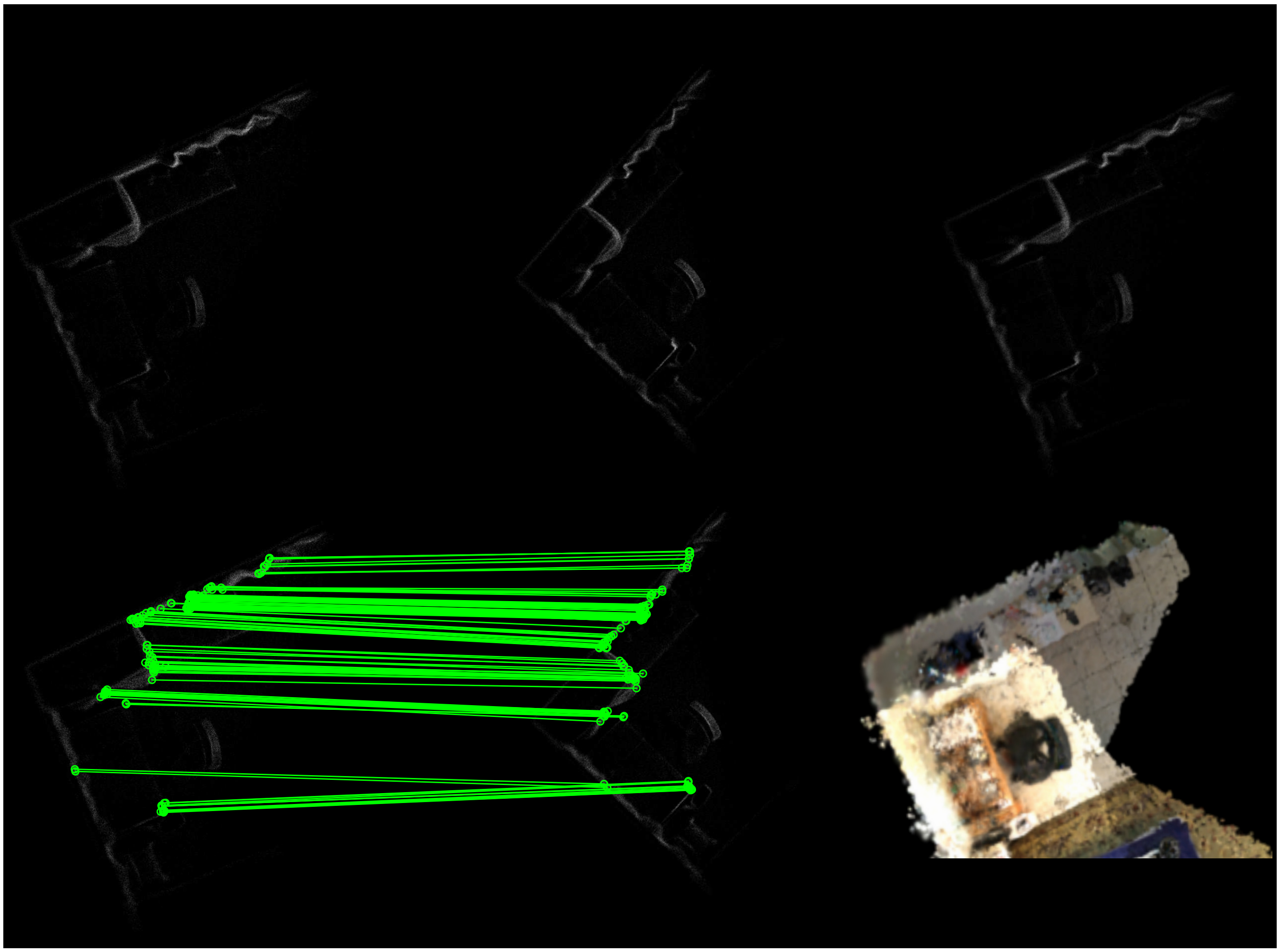

Fig. 3. Matching and registered density maps on ScanNet using the MapClosure method at a 0.005m per pixel resolution and a 0.8 overlap ratio. The top-left and top-center images represent the input density maps, while the top-right image shows the registered result. The bottom-left image illustrates the matching process between the input density maps, and the bottom-right image presents our BEV rendering for improved visualization and understanding. The inliers after the RANSAC-based rigid transformation estimation are shown in green.

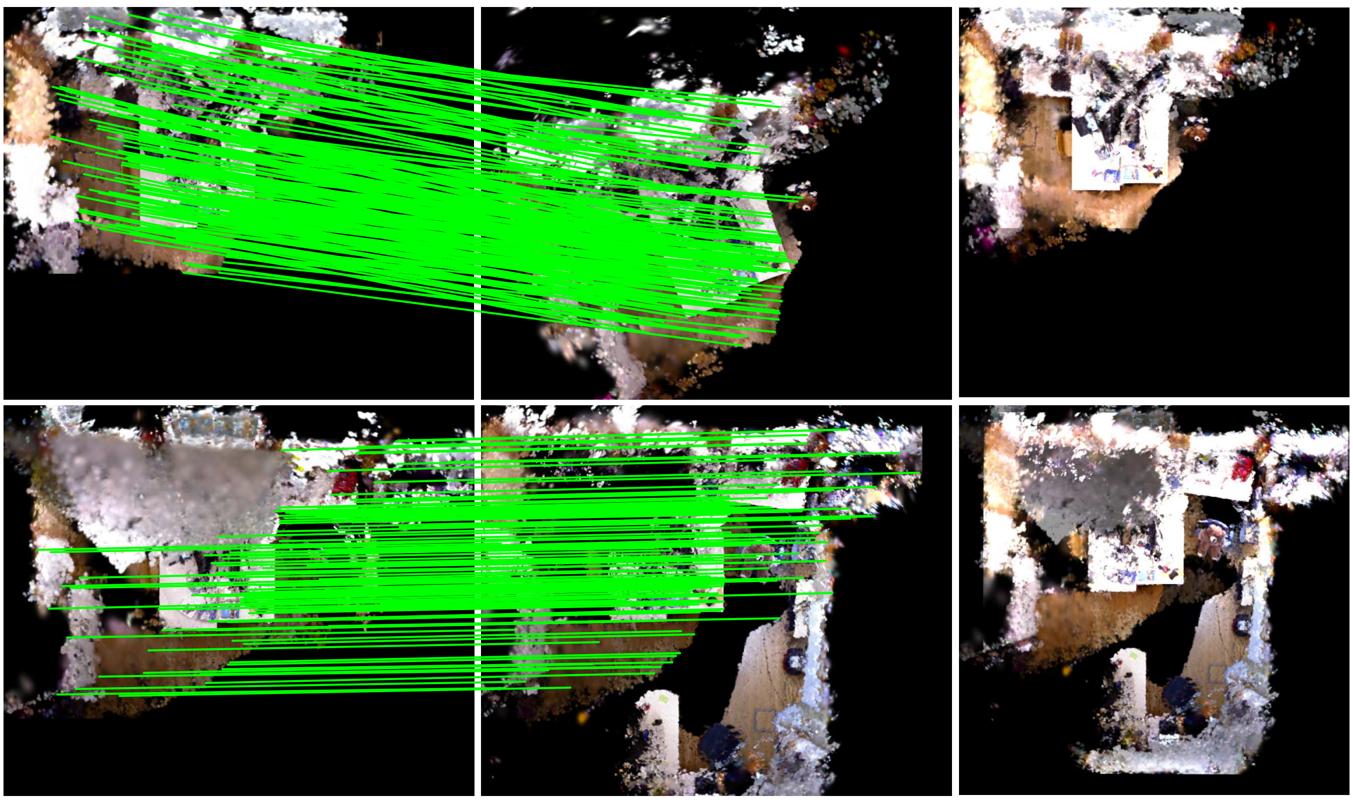

Fig. 4. Matching and matched submaps on TUM RGB-D using our method at 0.005m per pixel resolution and at an overlap ratio of 0.2. The inliers after the RANSAC-based rigid transformation estimation are presented in green.

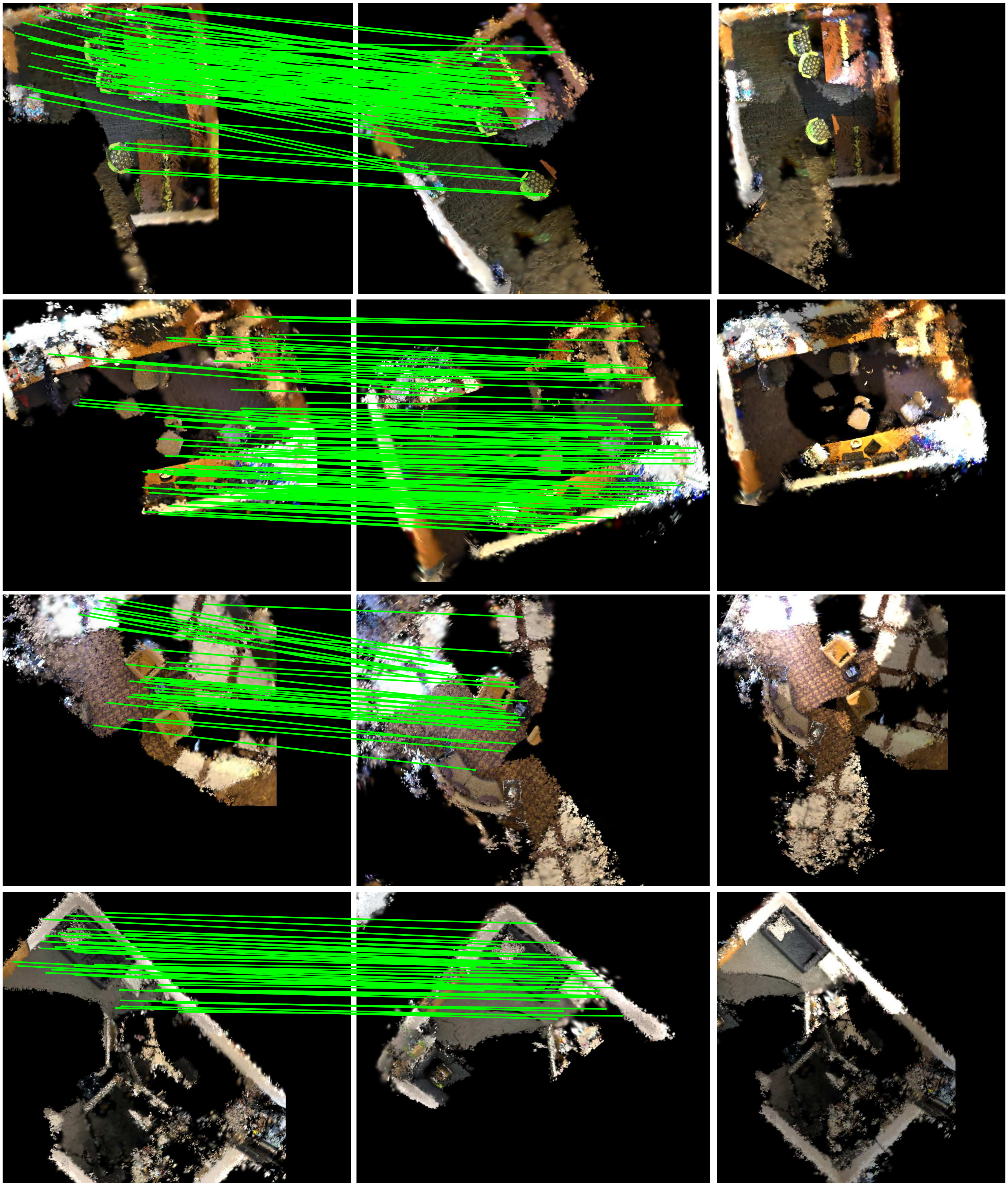

Fig. 5. Matching and matched submaps on ScanNet using our method at 0.005m per pixel resolution and at an overlap ratio of 0.4. The inliers after the RANSAC-based rigid transformation estimation are presented in green.