MIPS-Fusion: Multi-Implicit-Submaps for Scalable and Robust Online Neural RGB-D Reconstruction

Supplemental Material

Yijie Tang*, Jiazhao Zhang*, Zhinan Yu, He Wang, Kai Xu[†]

National University of Defense Technology   Peking University

*Joint first authors
[†] Corresponding author: Kai Xu (kevin.kai.xu@gmail.com)

Outline

Metrics of reconstruction quality: Accuracy details

- Accuracy Acc is the indicator that describes the difference between two point cloud S and G, it is formulated as below:

$$Acc = \frac{1}{\sum A(s)} \sum_{s \in S} A(s) \min_{g \in G} \|s - g\|_2,$$

where $A(s)$ represents the area around vertex s.

Metrics of reconstruction quality: Completeness details

- Completeness $Compl$ is the indicator that describes the percentage of inlier vertexes between two point cloud S and G, it is formulated as below:
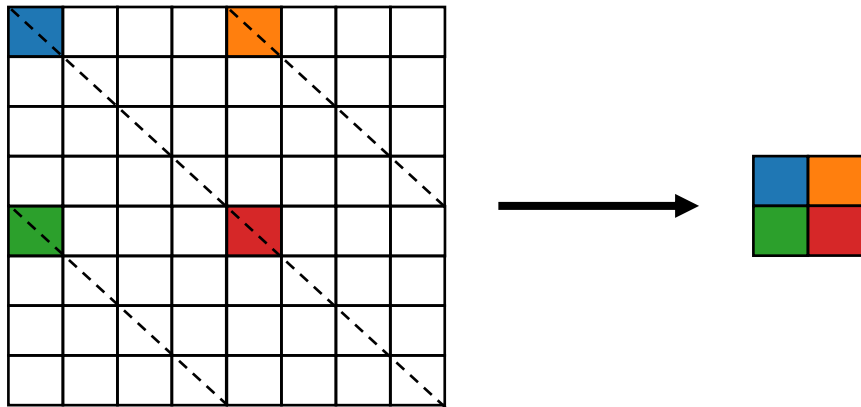
$$Compl = \frac{100}{\sum A(g)} \sum_{g \in G} A(g) \min_{s \in S} \|s - g\|_2^0,$$

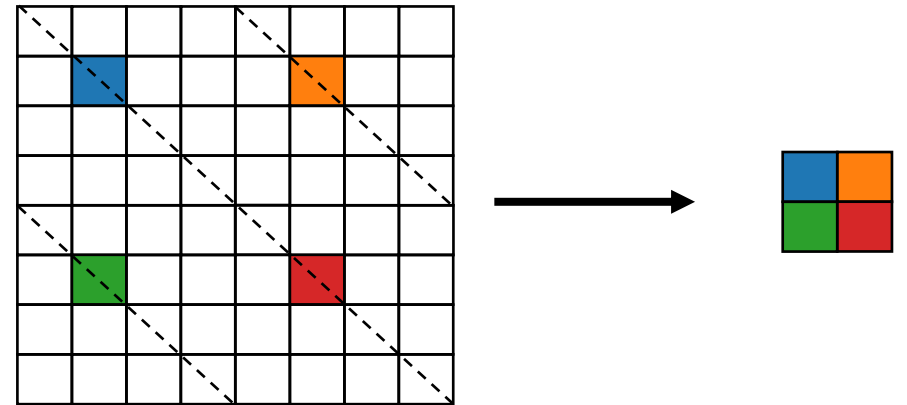where $\|s - g\|_2^0$ is defined via parameter ε as:

$$\|s - g\|_2^0 = \begin{cases} 0, & if \, \|s - g\|_2 < \varepsilon \\ 1, & otherwise \end{cases}$$

Down-sample Depth Map

We down-sample the depth map with a specific offset which vary along with the iteration.

Down-sample ×4
offset 0

Down-sample ×4
offset 1

Quantitative evaluation on FastCaMo-Synth dataset

	ElasticFusion	BundleFusion	MIPSFusion
Apartment_1	40.9cm	**4.6**cm	6.6
Apartment_2	40.7cm	**2.2**cm	3.1cm
Frl_apartment_2	43.8cm	83.6cm	**2.6**cm
Office_0	22.3cm	**2.7**cm	5.2cm
Office_1	**2.3**cm	17.3cm	7.6cm
Office_2	65.9cm	93.0cm	**17.4**cm
Office_3	94.3cm	253.5cm	**24.9**cm
Hotel_0	43.8cm	65.2cm	**6.0**cm
Room_0	-	-	**4.47**cm
Room_1	31.0cm	**0.6**cm	3.65cm

(ATE)

Results on FastCaMo-Synth (noise)

Quantitative evaluation on Replica

		office_0	office_1	office_2	office_3	office_4	room_0	room_1	room_2	avg
	accuracy(cm)⬇	2.939	**2.679**	3.476	3.359	**2.715**	2.767	2.455	**2.180**	2.821
	completion(cm)⬇	1.879	1.862	3.059	3.187	3.646	2.635	2.213	2.756	2.655
NICE-SLAM	comp ratio(%)⬆	95.02	93.97	88.49	86.99	87.05	91.87	93.47	91.36	91.03
	accuracy(cm)⬇	**1.730**	2.993	**2.231**	**2.866**	3.320	**2.407**	**1.615**	3.141	**2.374**
	completion(cm)⬇	**1.396**	**1.584**	**2.692**	**2.667**	**2.880**	2.624	2.206	**1.945**	**2.227**
Vox-Fusion	comp ratio(%)⬆	**97.14**	**95.73**	**91.12**	**90.85**	89.53	92.73	93.51	**94.18**	**92.84**
	accuracy(cm)⬇	3.211	2.827	4.138	3.278	3.316	3.084	3.083	2.559	3.187
	completion(cm)⬇	3.099	4.215	3.595	3.125	2.921	**2.015**	**2.003**	3.250	3.028
Ours	comp ratio(%)⬆	90.68	84.94	84.84	88.90	**91.31**	**96.95**	**95.49**	89.17	90.29

Evaluation on rendering quality

ScanNet

	PSNR	MSE	SSIM
NICE-SLAM	**26.6**	**0.0022**	**0.9717**
MIPSFusion	25.4	0.0029	0.9647

FastCaMo-Large

	PSNR	MSE	SSIM
NICE-SLAM	15.2	0.0298	0.6139
MIPSFusion	**23.0**	**0.0049**	**0.9118**

FastCaMo-Synth

	PSNR	MSE	SSIM
NICE-SLAM	13.3	0.0464	0.7855
MIPSFusion	**27.2**	**0.0019**	**0.9521**

Average results of NICE-SLAM and our method on FastCaMo-Synth and ScanNet

Evaluation on rendering quality

| GT color inputs | MIPSFusion | NICE-SLAM | VoxFsuion |
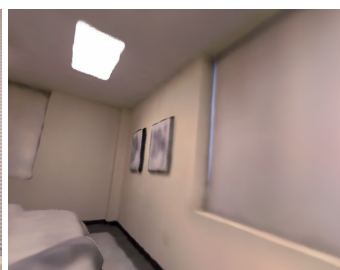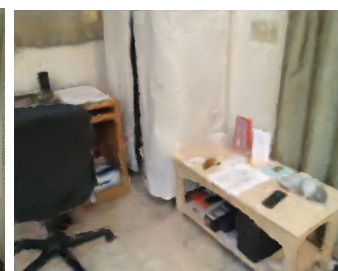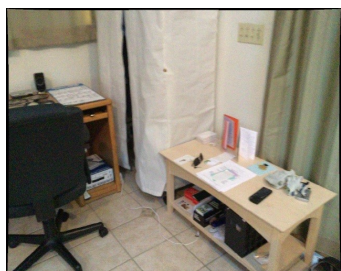
Evaluation on w./w.o. non-trivial loop
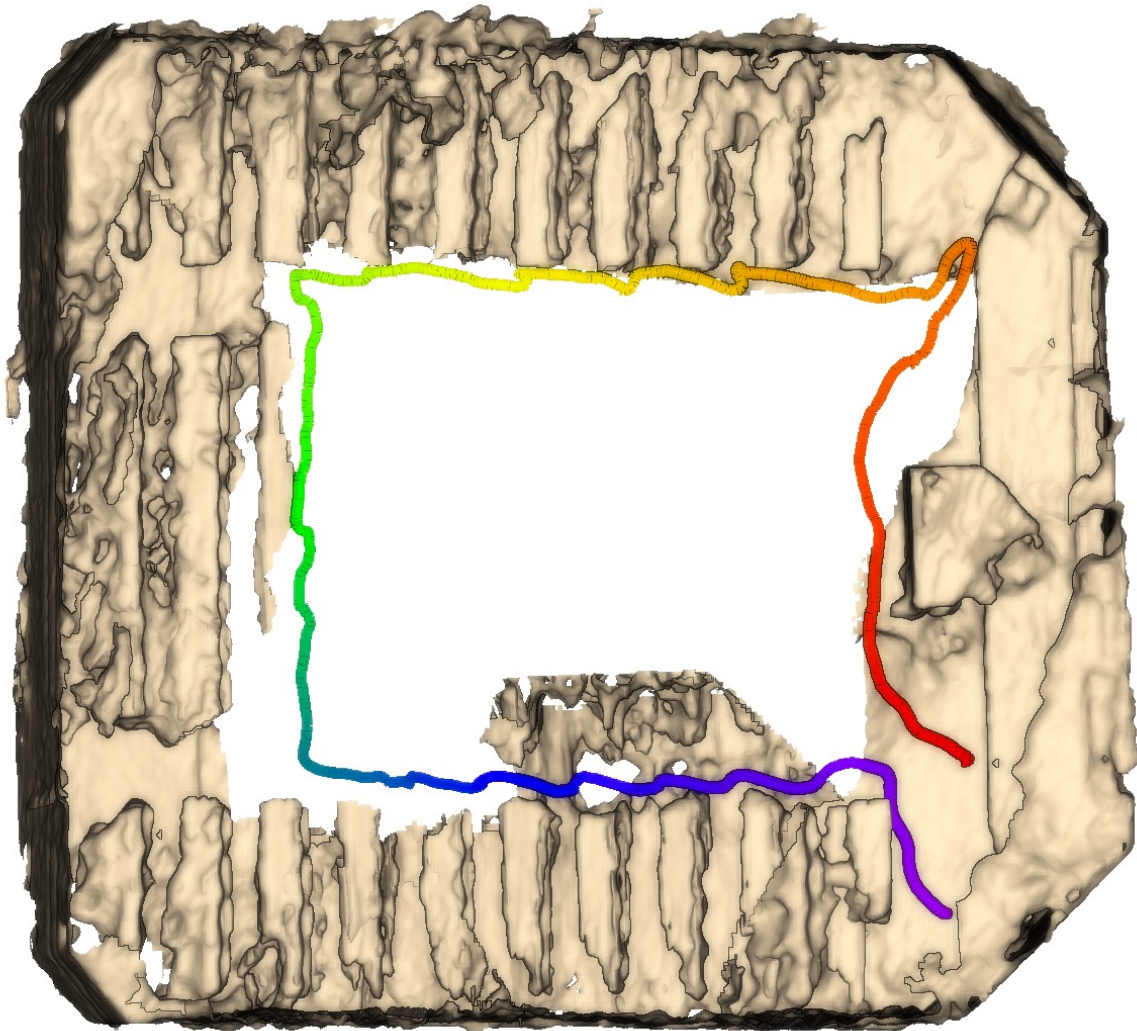
One round
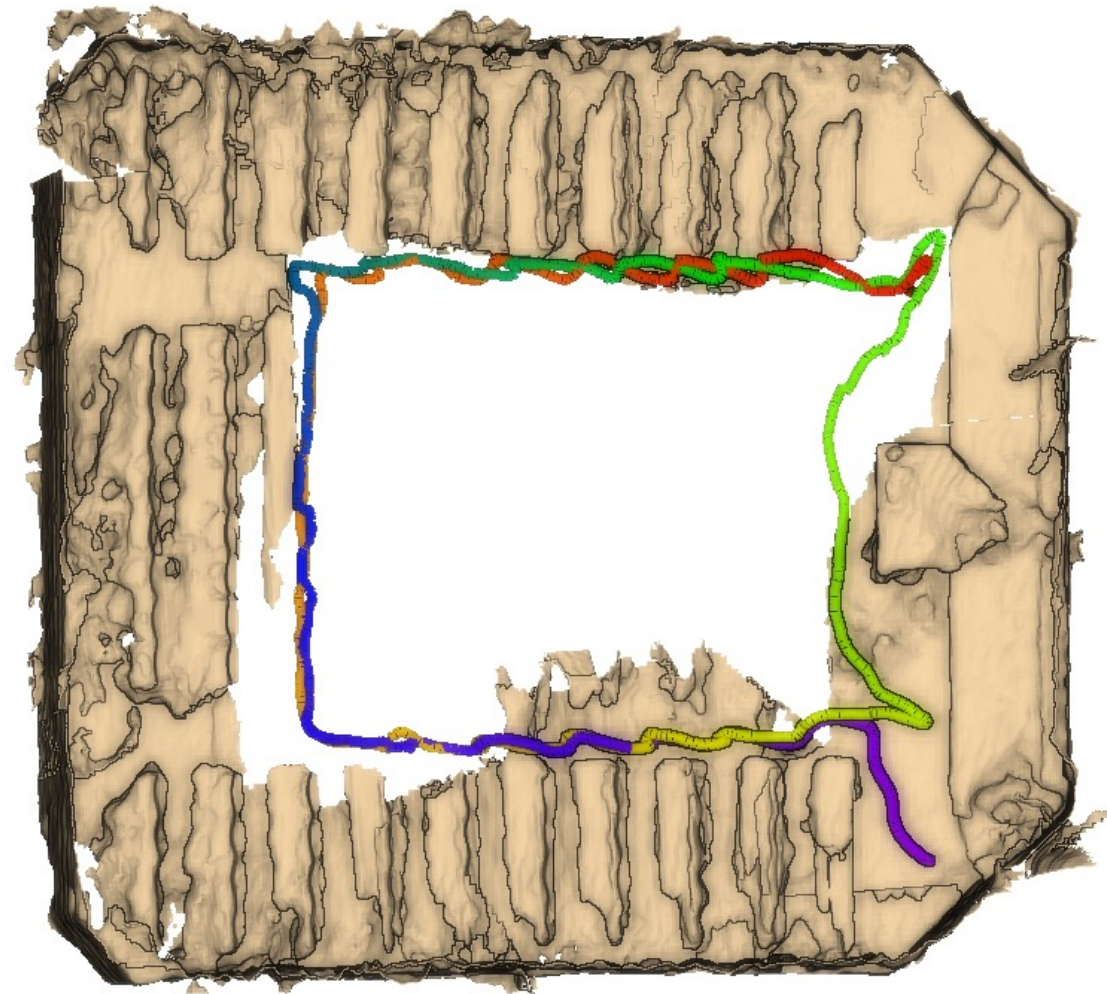
Second round

Visual comparison on Scannet, FastCaMo-Real

NICE-SLAM   Vox-Fusion   MIPS-Fusion   Ground Truth

Apartment-I
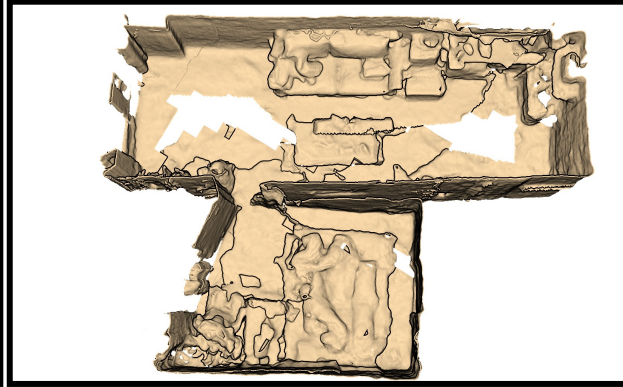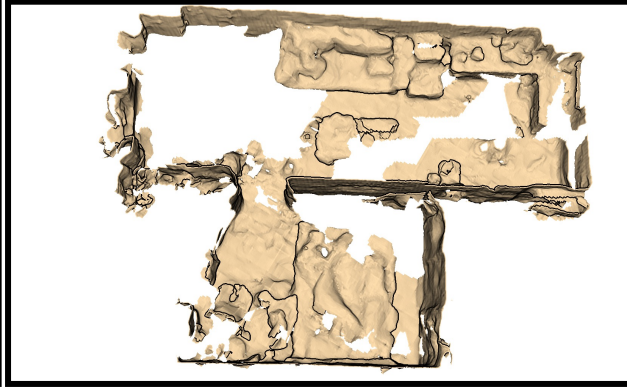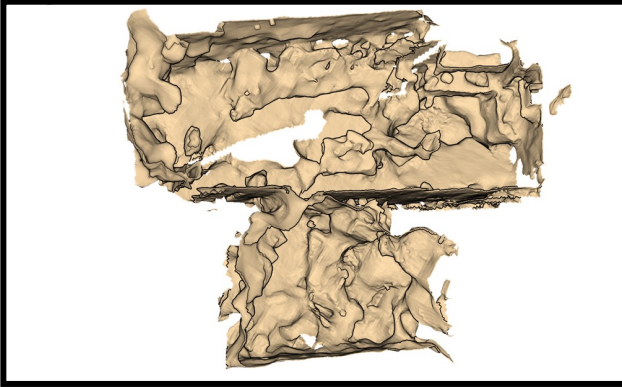
Lab
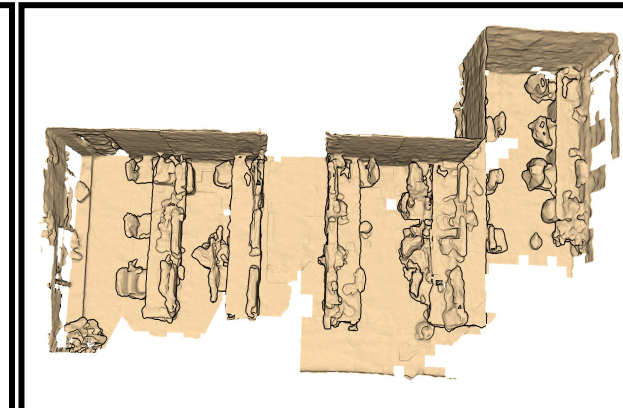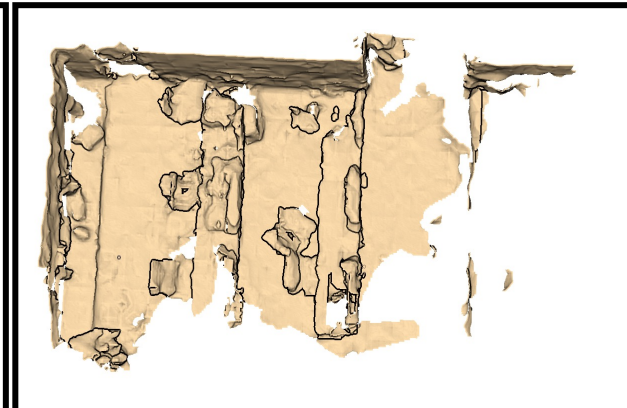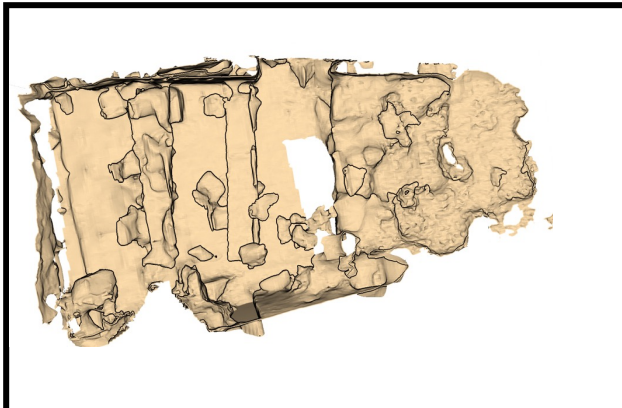
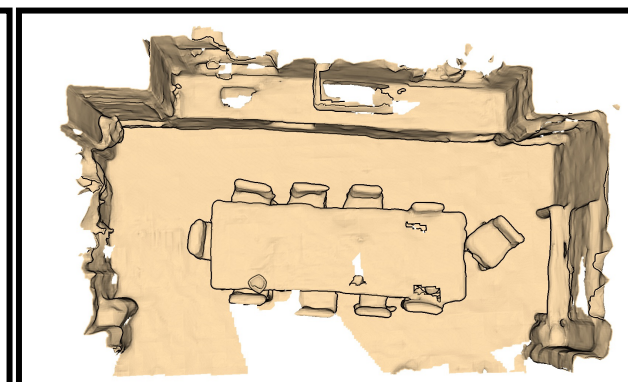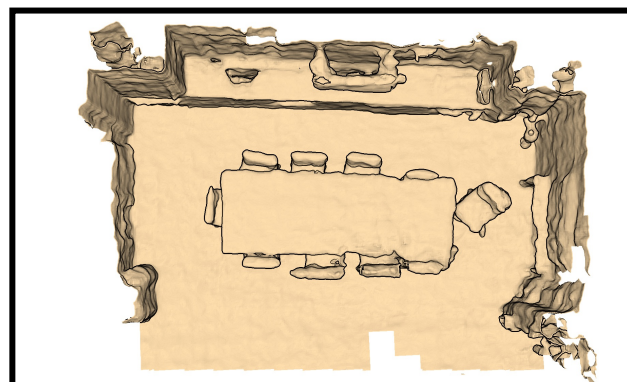# Visual comparison on Scannet, FastCaMo-Real

NICE-SLAM　　　**Vox-Fusion**　　　**MIPS-Fusion**　　　**Ground Truth**
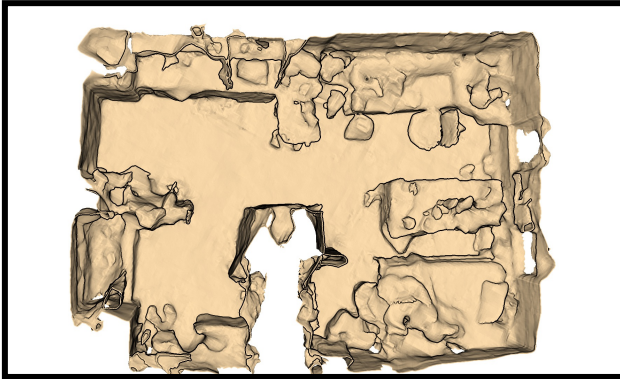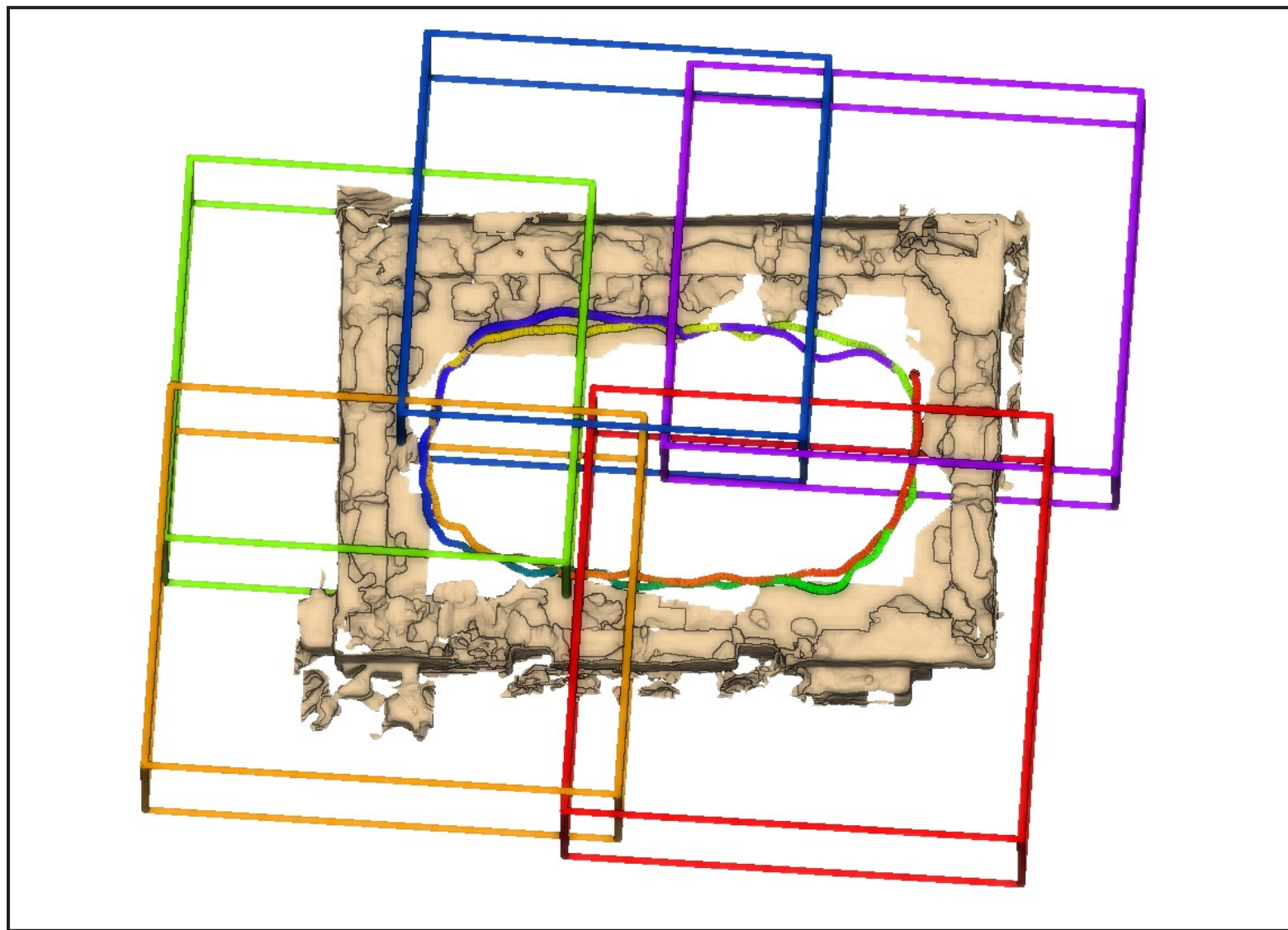
Lounge-i
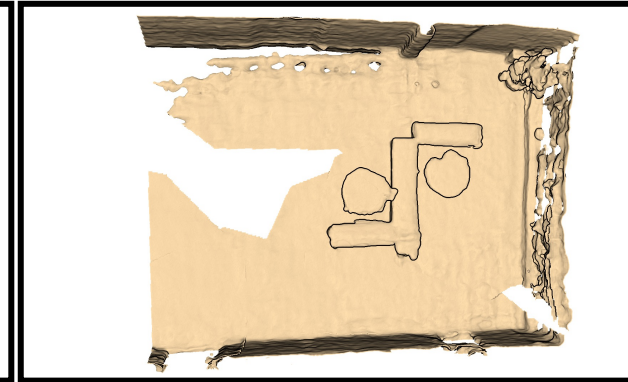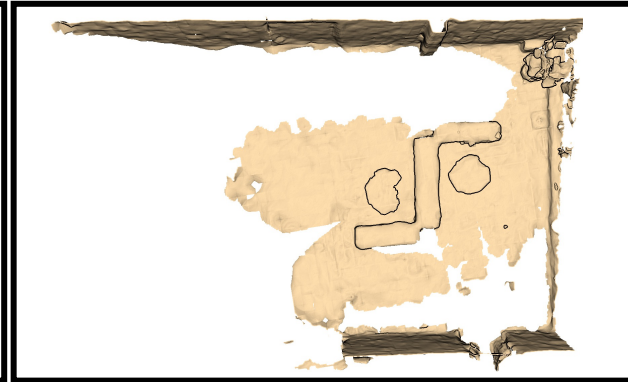
Scene0011
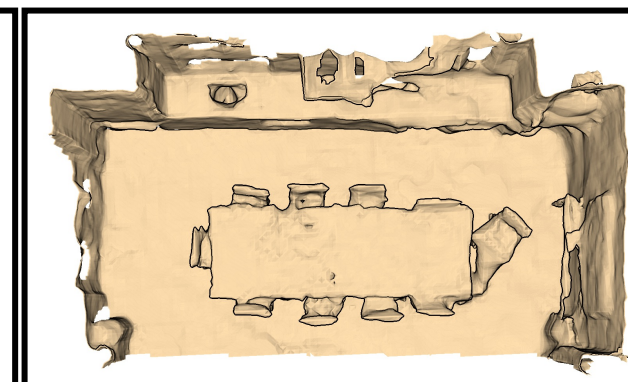
Visual comparison on Scannet, FastCaMo-Real
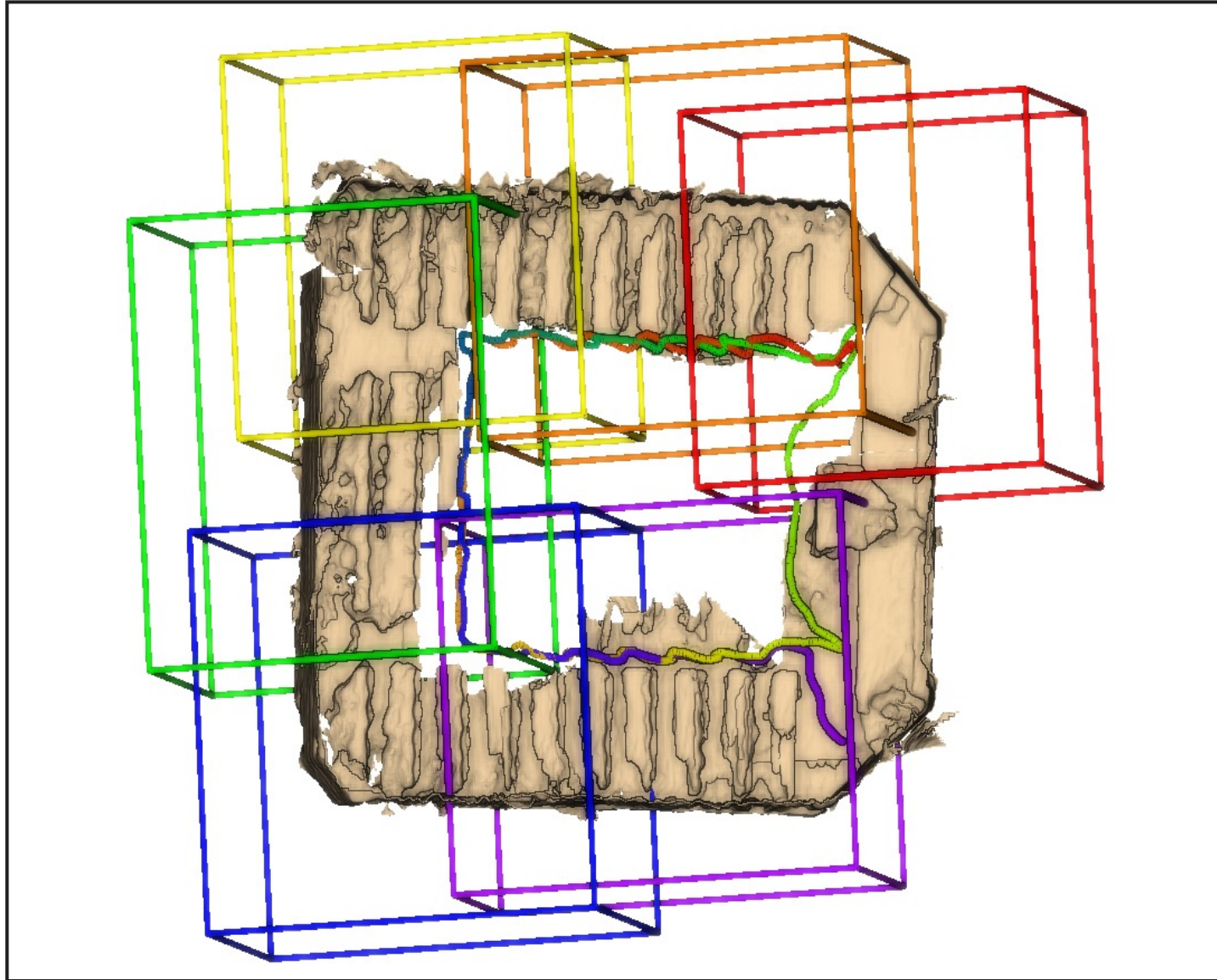
Scene0207
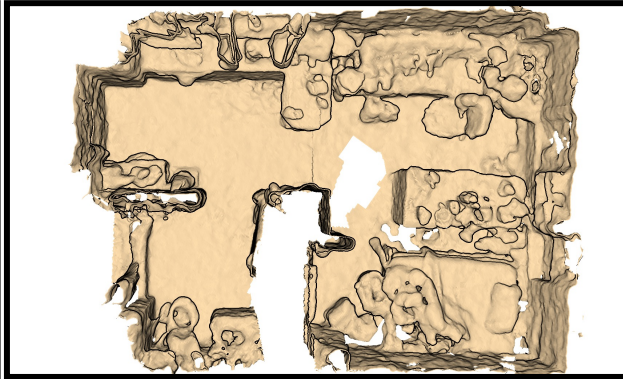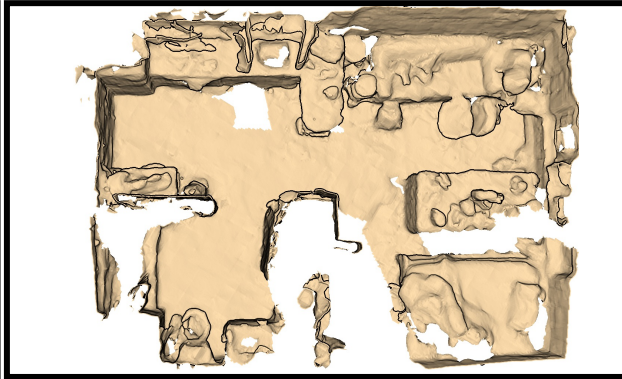
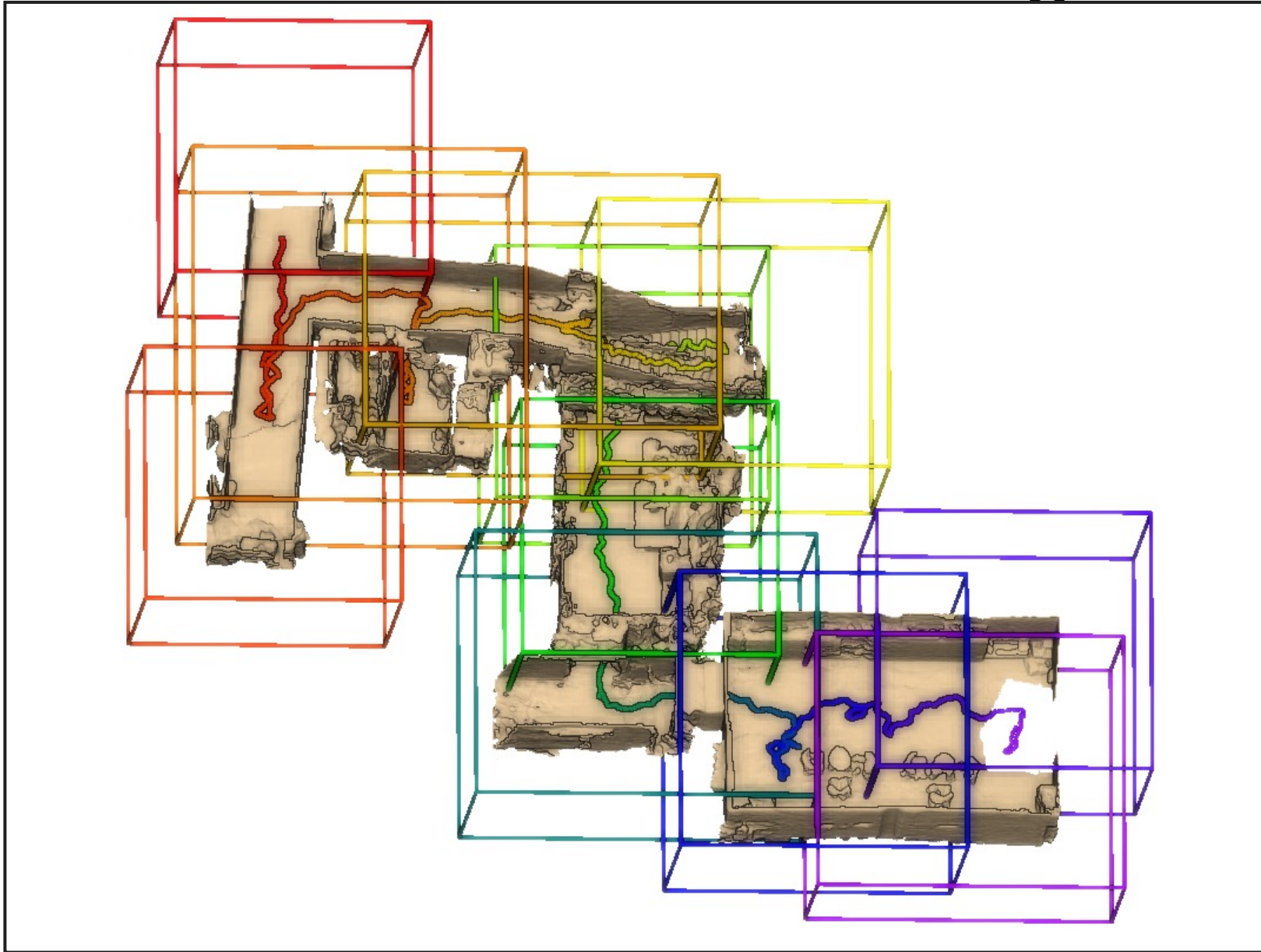# Visual results of FastCaMo-Large

Lab (~60m²)

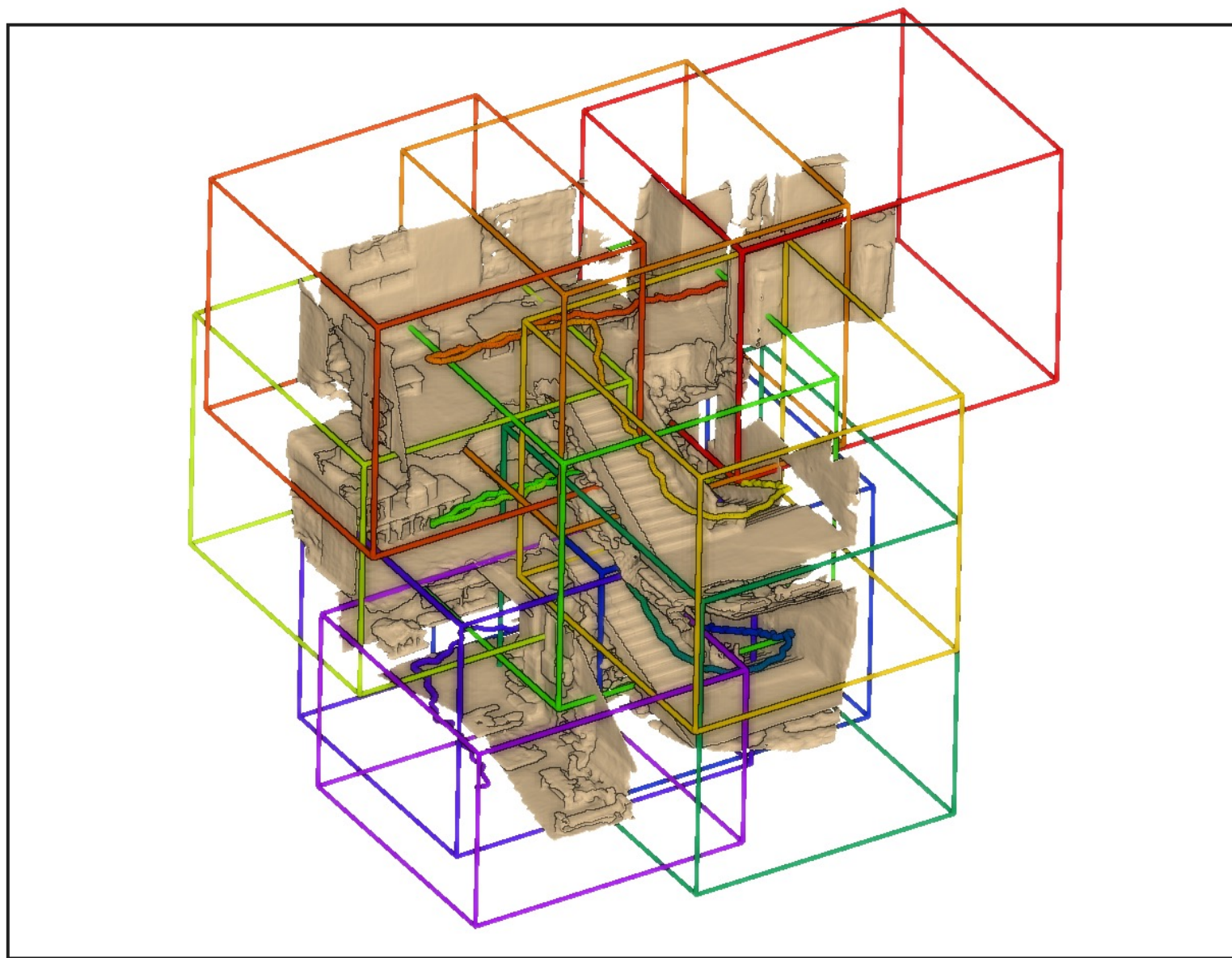# Visual results of FastCaMo-Large

Classroom (~80m^2)

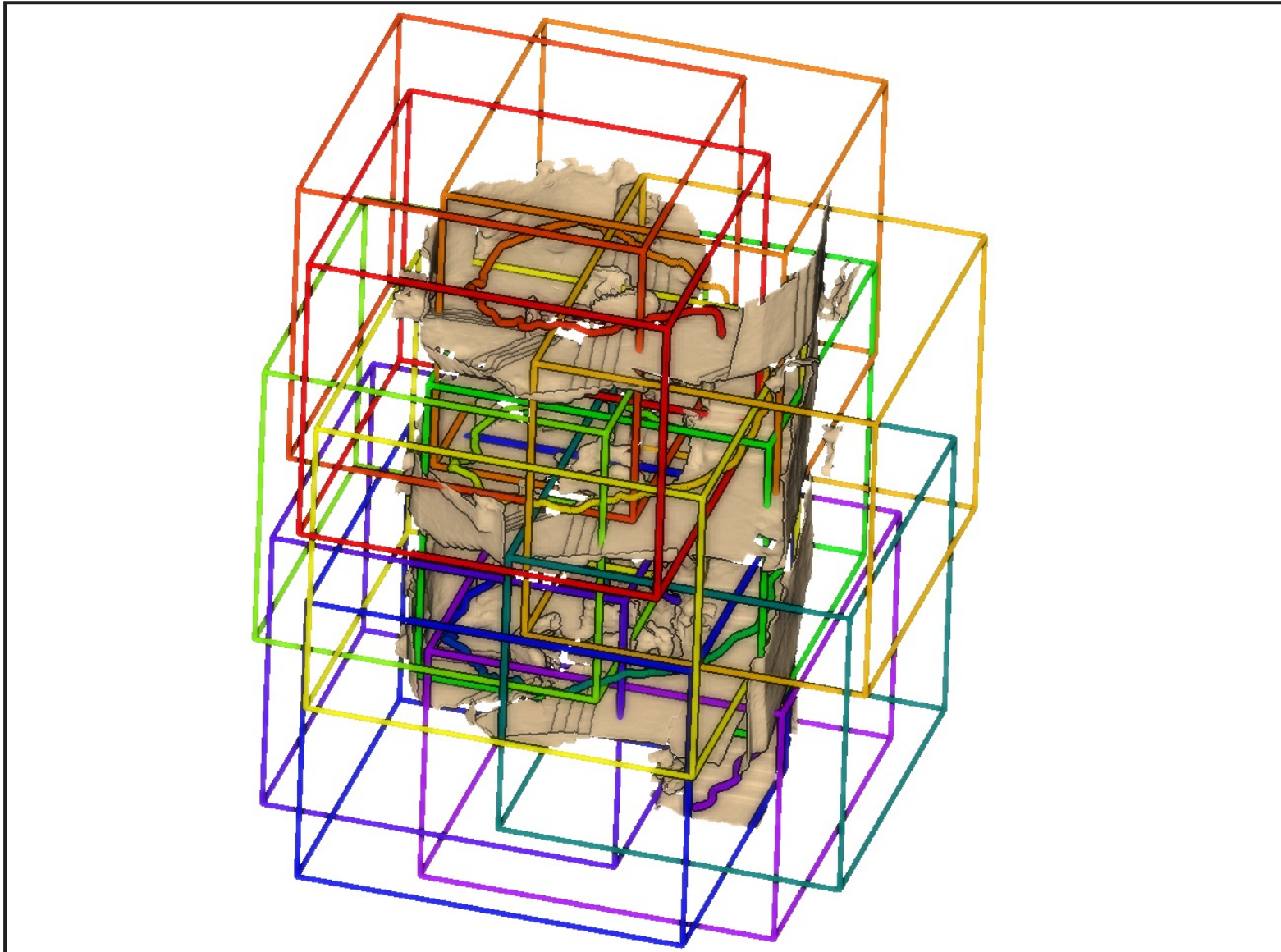# Visual results of FastCaMo-Large

Floors-I (~230 m²)
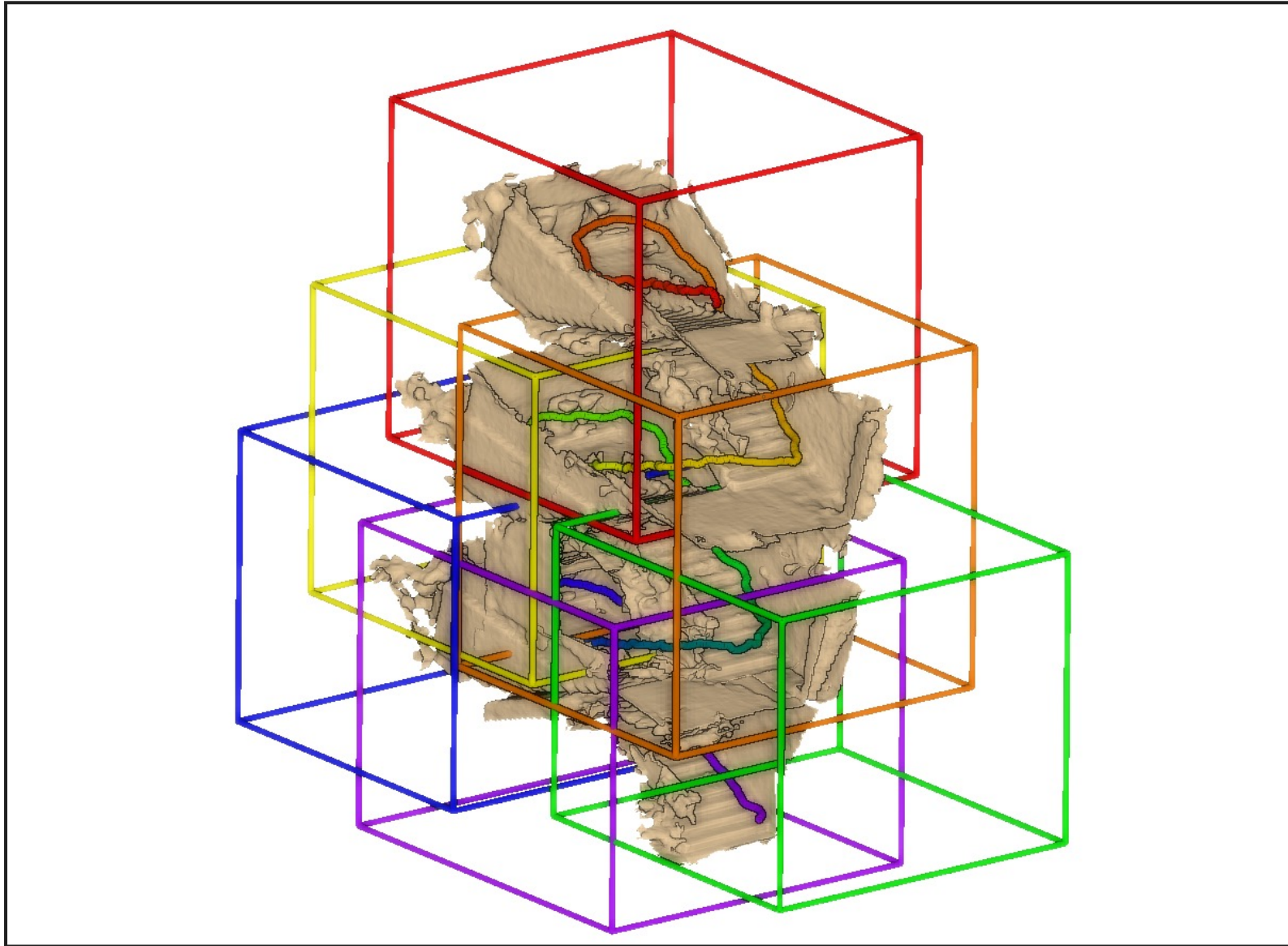
Visual results of FastCaMo-Large

Floors-II (~250 m²)

Visual results of FastCaMo-Large

Stairs-I (~100 m²)

Visual results of FastCaMo-Large

Stairs-II (~90 m²)