Supplementary Material: Aerial Path Planning for Urban Scene Reconstruction: A Continuous Optimization Method and Benchmark

NEIL SMITH, King Abdullah University of Science and Technology
NILS MOEHRLE, TU Darmstadt
MICHAEL GOESELE, TU Darmstadt
WOLFGANG HEIDRICH, King Abdullah University of Science and Technology

ACM Reference format:
Neil Smith, Nils Moehrle, Michael Goesele, and Wolfgang Heidrich. 2018. Supplementary Material: Aerial Path Planning for Urban Scene Reconstruction: A Continuous Optimization Method and Benchmark. *ACM Trans. Graph.* 37, 6, Article 183 (November 2018), 19 pages.
https://doi.org/10.1145/3272127.3275010

1 BENCHMARK DATASET

Below we present the Synthetic Benchmarks in further detail. The Benchmark consists of five large urban scenes created from 34 uniquely different buildings.

NY-1. This collection of 10 buildings share a common theme of multi-story apartment buildings sharing exterior walls for compactness. The streets are narrow and the walls and roofs are cluttered with fire escapes, water containers and signs. The natural textured variant uses both tiled material textures and rectified images from the original models. The street and ground plane use Simplex noise to avoid miscalculation of false matches from highly repeated tiling of texture. Windows in all datasets are textured with a chrome highly reflective mirroring material. This is to both make capturing windows difficult but also to avoid false reconstructions if the glass was clear.

GOTH-1. These five large buildings modeled with Gothic and Parisian architectural styles are situated across a large open channel with a columned bridge connecting the two banks. This scene is especially challenging with its many arches on the cathedral, bridge and apartment entryways. The natural colored variant is primarily textured with marble or decayed stone materials which create a good texture but in some cases high reflectivity. After initial runs we found that the ground plane and bridge should still be textured using Simplex noise since otherwise it would lead to the SfM pipelines overmatching the two symmetric sides.

CA-1. This street intersection contains 11 modern buildings with small unflyable alleys between them. Every two-story building has a complex store front with signs, balconies, low canopies and fenced porches. This scene was made especially challenging for path planning by restricting accessibility and occluding buildings with canopies, which are a common architectural feature found in the area the scene is modeled after. The natural color variant is primarily textured using differently colored stucco materials and normal maps to create diversity. CA-1 proves to be the most difficult dataset in reconstruction due to the high amount of homogenous texture. We texture the roofs with a material blend of hatched texture tiles and Perlin noise. Similarly, the street is a combination of Simplex noise and concrete both to prevent against repeating textures but maintain closer appearance to roads.

UK-1. These eight neo-classical architectural style buildings represent the largest of the modeled scenes in the benchmark. The face of the buildings are studded with brickwork, moldings, hundreds of windows, and pillars. The buildings encompass a large open courtyard from which the majority of all the buildings can be seen. The challenge of this scene is to evaluate whether the flight plan descends into the open courtyard to optimize its view space. A similar approach to colored texturing was conducted for UK-1 while the two largest building had highly reflective wood plank texture applied.

Old-City. A less detailed scene of 60 mass model buildings completely attached together with very narrow roads. The scene has been thoroughly modeled to ensure there are no faces that cannot be theoretically captured in the path planning. This scene is designed to examine how path planning scales to larger scenes and can be used as a fast metric for measuring completeness of reconstruction.

2 BENCHMARK EVALUATION MEASURES

For the research community we release a downloadable stand-alone tool in UE4 for image generation, alignment and evaluation against the original mesh datasets:

https://vccimaging.org/Publications/Smith2018UAVPathPlanning

The original mesh datasets and textures will not be released to the public. Rather they will be only accessible inside the evaluation tools or packaged inside UE4 which takes care that the original models cannot be extracted. Although we assume that the research community will strive to not abuse the evaluation tools, it is possible that people may attempt to combine the rendered images of previous runs to augment the planned image set and thus illegally increase the coverage. We believe the best approach to avoiding this activity when suspected is to request from the submitter the original bundle file and images used to create the reconstruction so that we can verify only these images were used. A more extreme measure is to change the random texture with a new seed number in each run but this would be more complicated to implement and not possible for the "naturally" textured version.

3 WORKSTATION SPECIFICATIONS FOR RUNTIME COMPUTATION

All run time experiments conducted on our approach are conducted on a 2013 Dell Workstation with Intel Xeon 2.7GHZ 32 core processor with 128GB ram and Nvidia Tesla X graphics card. Note that the computation is GPU based and thus primarily affected by clock speed and available CUDA cores of the GPU. For Roberts *et al.*'s [2017] approach and their implementation of NBV approach we use a Mac Pro 2013 Intel Xeon 2.7 GHZ 24 core processor with 64GB ram and AMD FirePro D700. Note Roberts *et al.*'s [2017] original implementation was conducted on a 2015 Macbook Pro with 2.8GHZ Core i7 with 8 cores and 16GB ram. More nodes could be visited and better optimization achieved in less time using the much more powerful Mac Pro. Their approach is CPU bound and regularly consumed the entire memory limit of the Mac Pro and all cores.

4 ADDITIONAL RESULTS OF COMPARISONS

In addition to comparisons of GOTH-1 and CA-1, we compare the NY-1 dataset for Roberts *et al.*'s [2017] submodular approach (SUB) and our approach (OURS). Note that both approaches do not place any cameras closer than 4m from the scene geometry. In Table 1 , we show the comparison between SUB and OURS for reconstruction of NY-1. Similar to the datasets compared in the paper, the SUB approach fairs well but does not achieve similar levels of accuracy and completion as our approach.

Table 1. Comparison between SUB and OURS in reconstruction quality for flight paths on the NY-1 dataset.

	NY-1				
	Error[m] 90%	Error[m] 95%	Comp[%] 0.075 m	Comp[%] 0.050 m	Comp[%] 0.020 m
SUB (473)	0.034	0.052	35.43	31.30	21.09
OURS (433)	**0.024**	**0.047**	**44.62**	**40.80**	**33.60**

5 BENCHMARK EVALUATION OF SFM PIPELINES

In the first set of graphs, we graph completeness and error for the four synthetic scenes reconstructed from the Perlin noise variant. In earlier work we also ran VSFM which produced poor results so it was not used in the final study of the natural color variant. Second, we present the visualization of the flight trajectories computed for the synthetic scenes. In the final section of figures, we provide the qualitative visual representation of accuracy and reconstruction for the four scenes and comparisons against other work.

The benchmark scores indicate that MVE and COLMAP perform very similarly. Overall, all the pipelines except VSFM+PMVS are able to fully reconstruct the entire scene. Considering that the targeted pixel size is 1.0cm we find the majority of the reconstruction pipelines achieving accuracy scores close to pixel resolution promising. This also demonstrates the efficacy of the path planning to insure sufficient overlap and consistent GSD for the entire scene. The high accuracy and completeness is especially surprising in difficult regions such as below the the short balconies and store of CA-1, or the pillars under the bridge of GOTH-1. Many of the cameras drop down to street level and take horizontal capture of these areas. Window sills and details in the stained glass windows of GOTH-1 are all captured with high accuracy.

There are also some notable areas of improvement for all pipelines. Several of the MVS approaches leave a halo of false reconstructed points along the edges of the buildings. Pix4D suffers most from this problem and the use of the Simplex textured surface reconstruction results in large bulbous mesh artifacts. The fire escapes of NY-1, balconies and fences of CA-1, and rows of short columns in UK-1 have many errors, which all appear related to vertical objects separated by narrow spaces. Inaccuracy is seen on most sharp edged features, especially along window sills, wall moldings and building edges. This is in contrast to the high accuracy reconstruction of the rounded pillars found in GOTH-1 and the dome roof in UK-1. Finally, in the UK-1 scene both Pix4D and MVE incorrectly reconstruct a ghost version of one of the buildings.

The hardest scene for all pipelines to reconstruct is CA-1 which has many overhangs and tight spaces that are not well reconstructed underneath. This highlights the limitations of aerial capture and the restraints we have set in our formulation, where the camera gimbal cannot view upward due to being under the copter. It also has to be noted that the height map and nadir capture also cannot fully represent the details of what are under these overhangs limiting how the camera path can be optimized offline.

The textured color variants produce similar error (accuracy) results to the Perlin textured datasets but noticeably fail to reconstruct textureless and reflective surfaces leading to significantly reduced completeness scores.

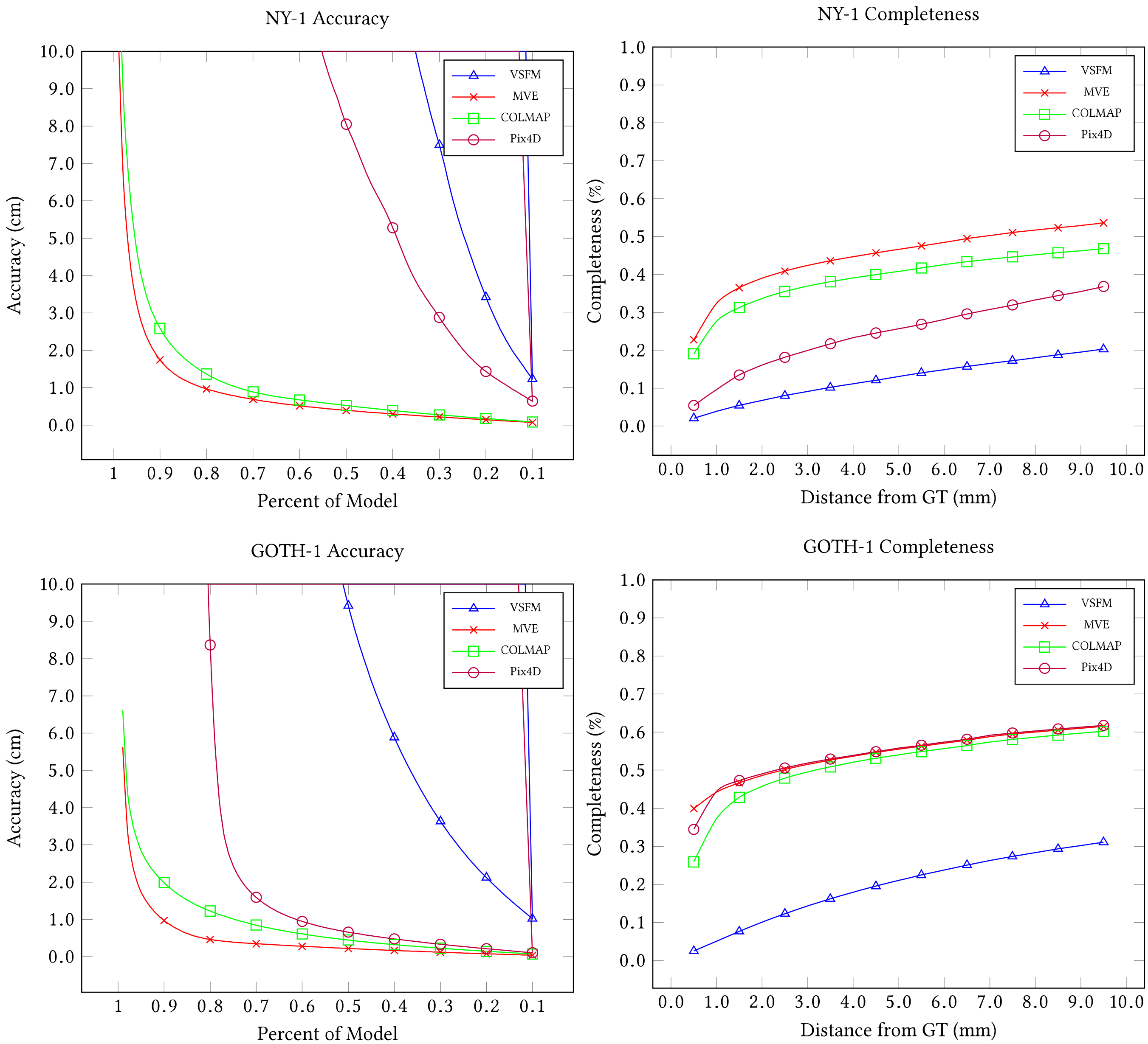
6 ADDITIONAL OBSERVATIONS

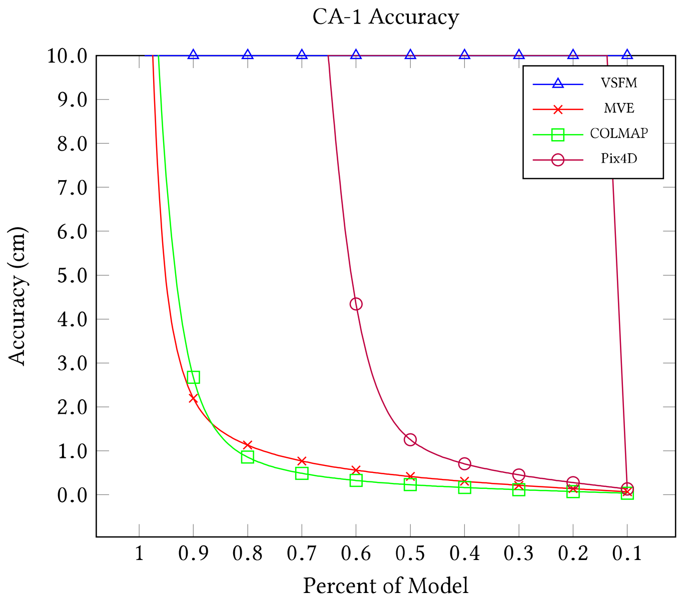
The results from the synthetic benchmark highlight the major advances in reconstruction pipelines in the last several years; however, there are still several noticeable areas for improvement. First, edges and corners are not well reconstructed but rather tend towards either being overly rounded or miscalculated. We believe this is an inherent problem with feature detection in general which focuses on blob detection and rejects edges in feature matching, making the amount of 3D measured points on the edge very sparse. For natural scenes that are more rounded this is not as noticeable, but it is very apparent in the Manhattan scenes of this benchmark. This has probably an additional cause in that surface extraction techniques have problems with sharp edges as shown in Aroudj *et al.* [2017]. Second, artifacts regularly appear at the transition between buildings and the sky or between small spaces such as fencing or lattice walls. These errors are exacerbated during meshing leading to either bulbous meshes along edges or the filling of the gaps between fencing. This is a segmentation problem that could be addressed in future work in image based reconstruction. Finally, perfectly planar surfaces both in the streets and on the buildings are not perfectly smooth but show a random distribution of small depth calculation errors. Despite having highly heterogeneous texture these errors persist in all reconstruction pipelines. A hybrid based approach of reconstruction that also seeks to detect planes and edges may be the best way forward in this domain.

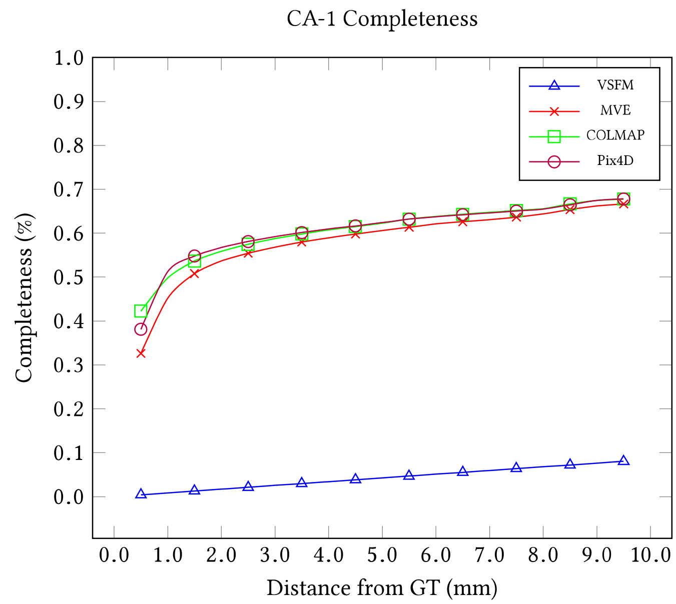
REFERENCES

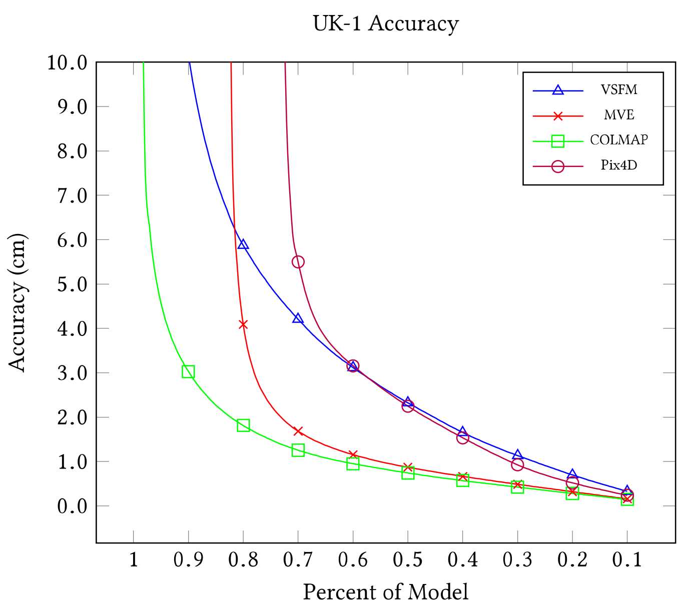
Samir Aroudj, Patrick Seemann, Fabian Langguth, Stefan Guthe, and Michael Goesele. 2017. Visibility-consistent Thin Surface Reconstruction Using Multi-scale Kernels. *ACM Trans. Graph.* 36, 6, Article 187 (Nov. 2017), 13 pages. https://doi.org/10.1145/3130800.3130851

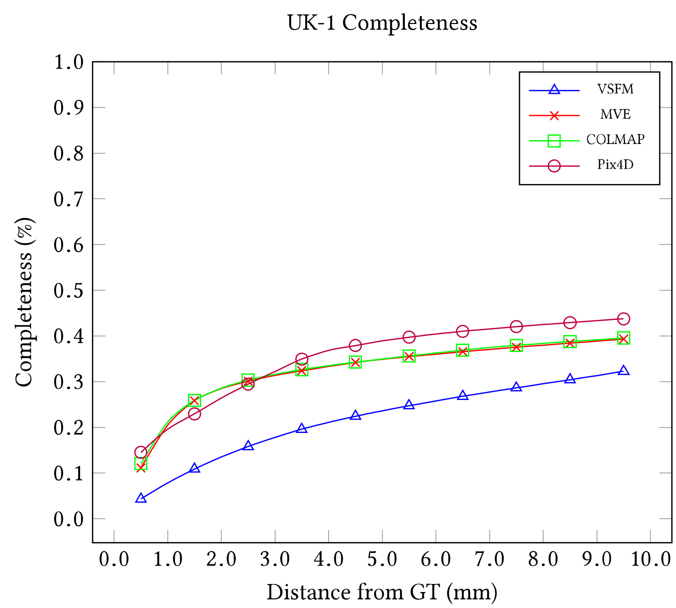
Mike Roberts, Debadeepta Dey, Anh Truong, Sudipta Sinha, Shital Shah, Ashish Kapoor, Pat Hanrahan, and Neel Joshi. 2017. Submodular Trajectory Optimization for Aerial 3D Scanning. In *Proc. ICCV*.


NY-1 Accuracy
Accuracy (cm)
Percent of Model
NY-1 Completeness
Completeness (%)
Distance from GT (mm)
GOTH-1 Accuracy
Accuracy (cm)
Percent of Model
GOTH-1 Completeness
Completeness (%)
Distance from GT (mm)
VSFM
MVE
COLMAP
Pix4D

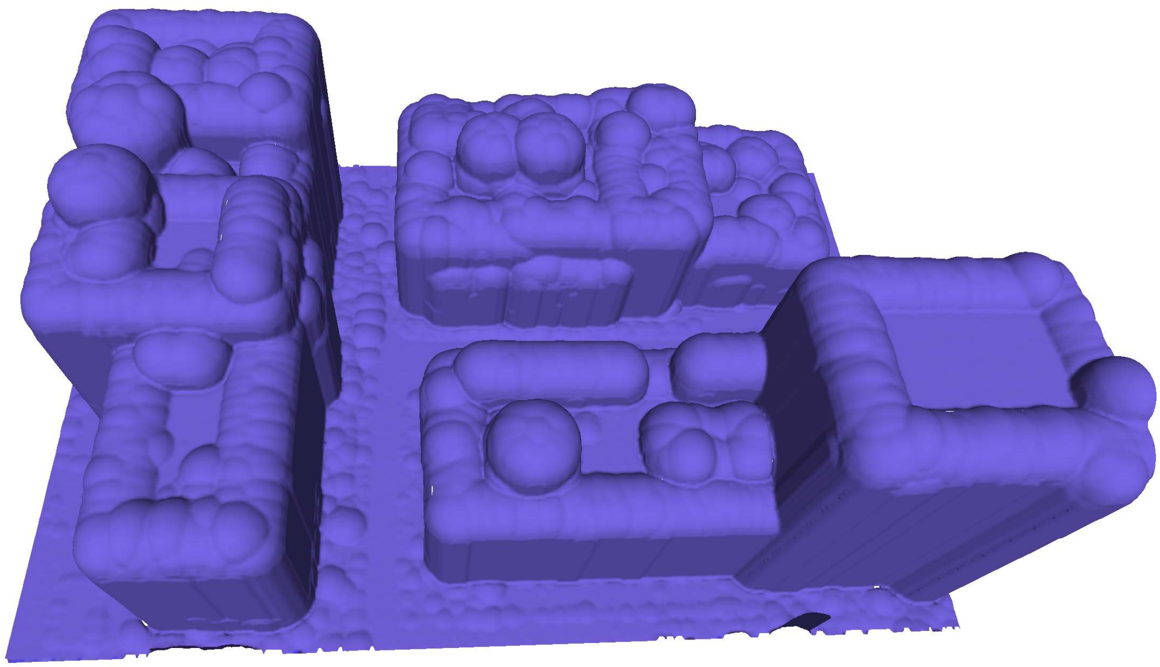

CA-1 Accuracy
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Accuracy (cm)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Percent of Model
VSFM
MVE
COLMAP
Pix4D


CA-1 Completeness
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Completeness (%)
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
Distance from GT (mm)
VSFM
MVE
COLMAP
Pix4D

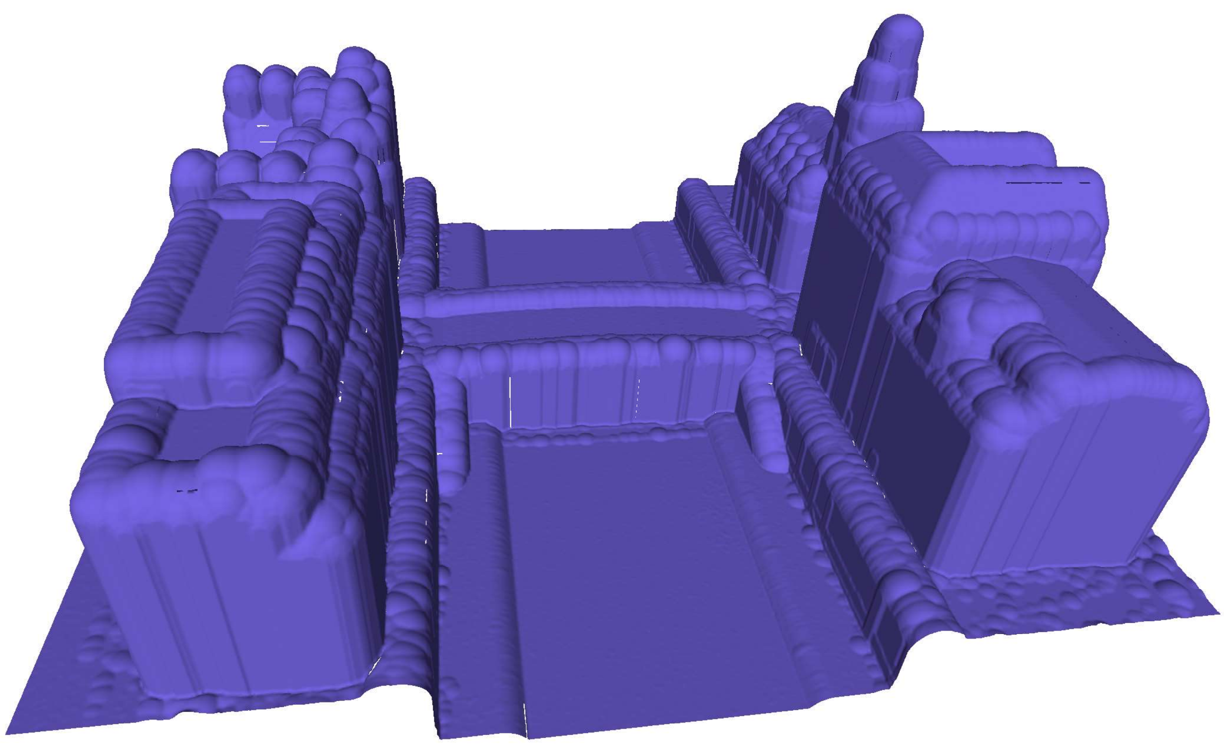

UK-1 Accuracy
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Accuracy (cm)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Percent of Model
VSFM
MVE
COLMAP
Pix4D


UK-1 Completeness
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Completeness (%)
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
Distance from GT (mm)
VSFM
MVE
COLMAP
Pix4D

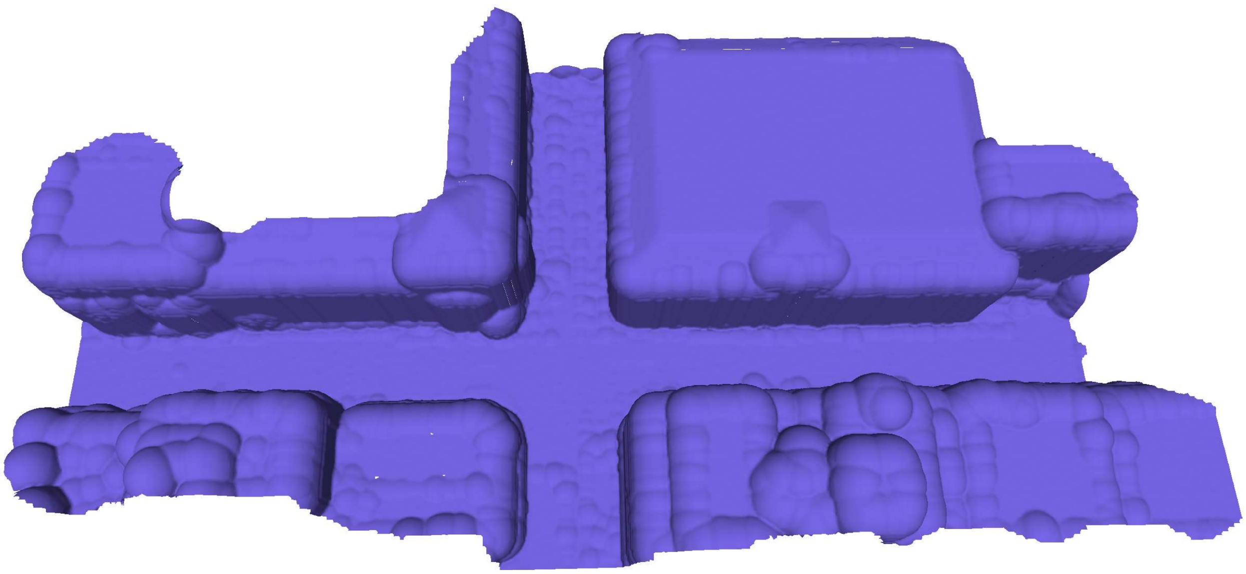
Airspace NY-1



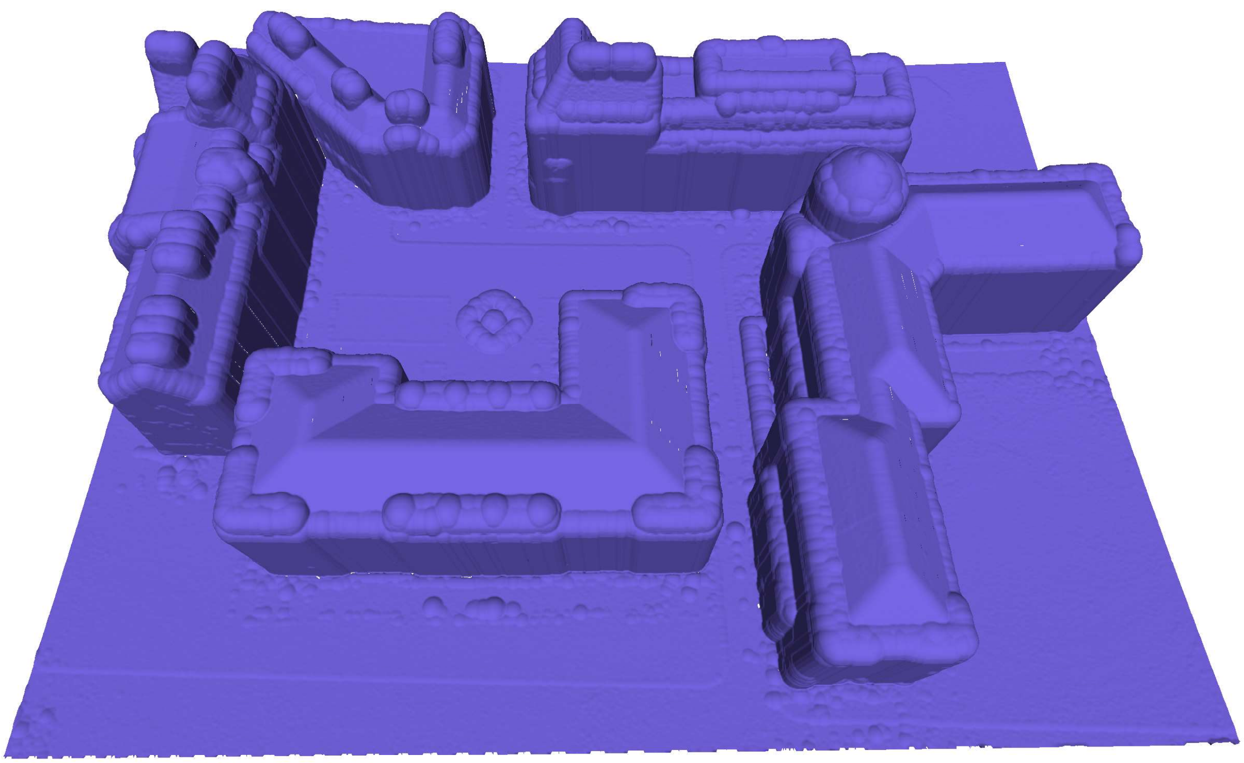
Airspace GOTH-1



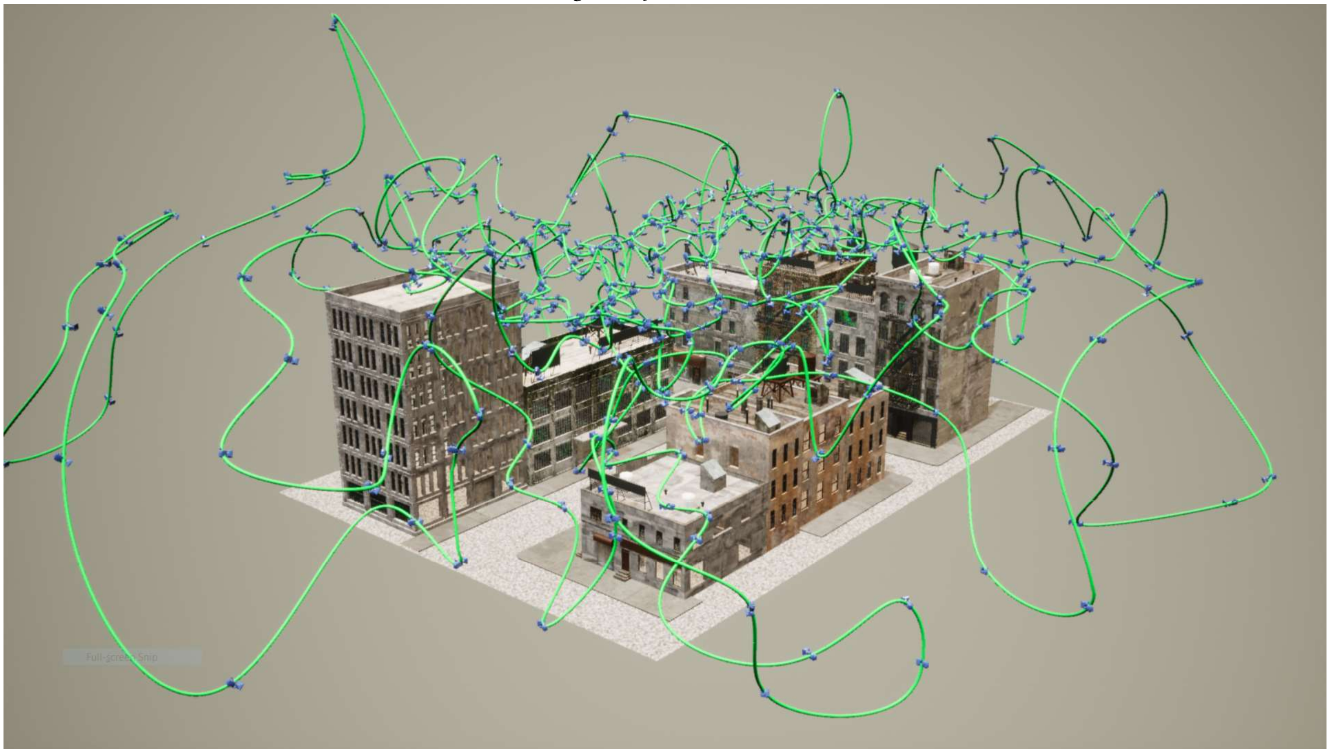
Airspace CA-1



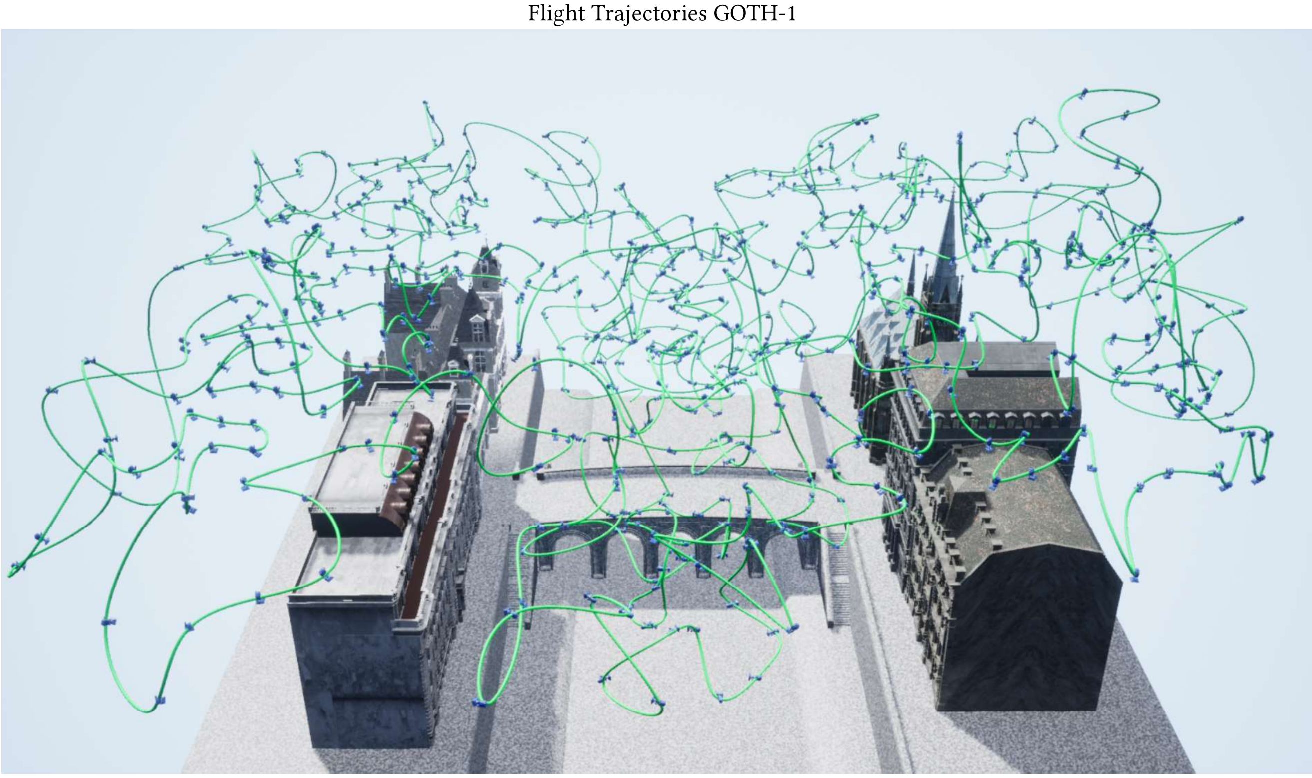
Airspace UK-1



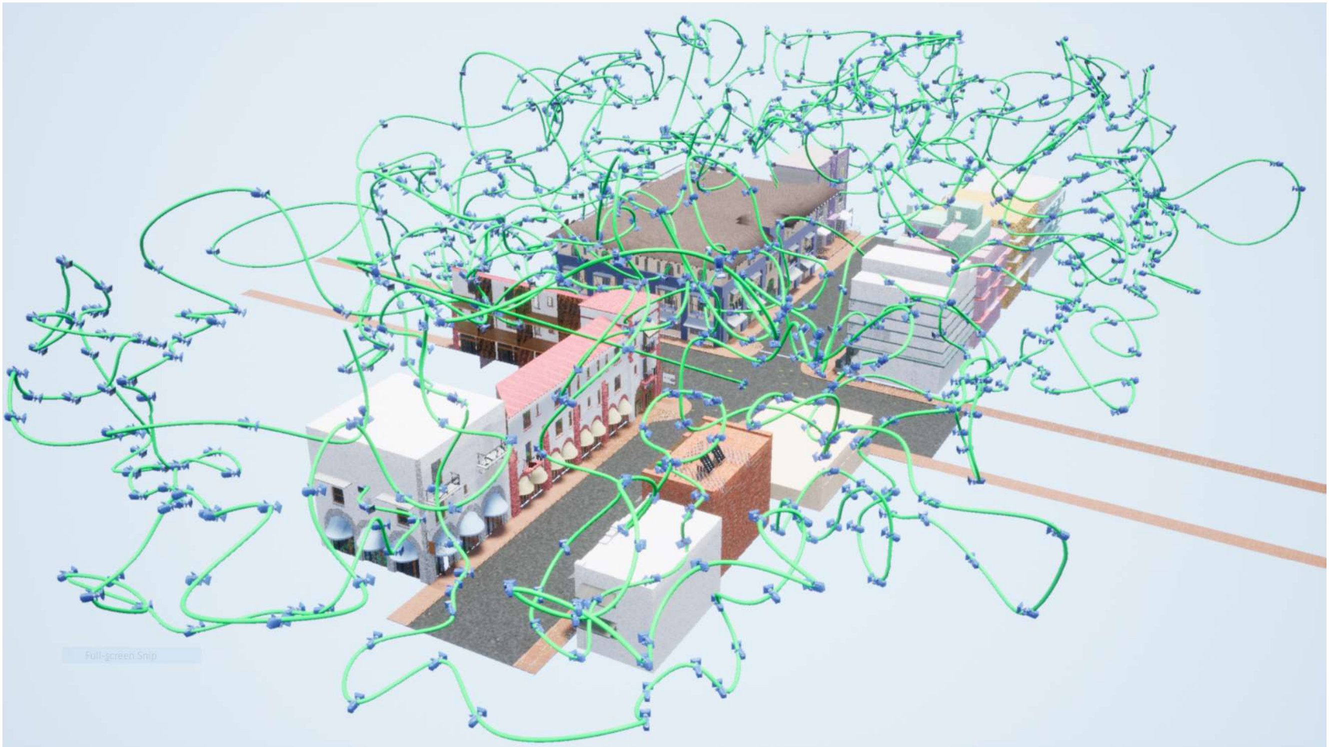
Flight Trajectories NY-1



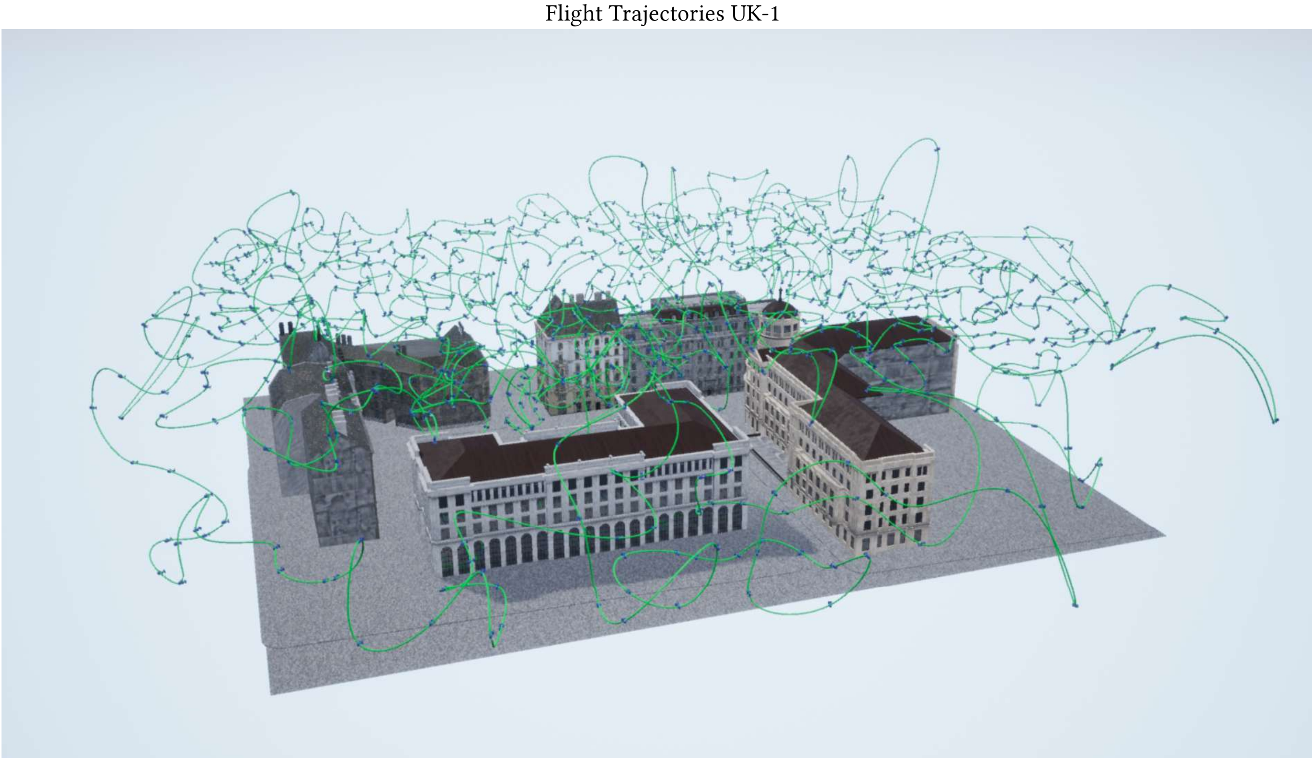
Flight Trajectories GOTH-1

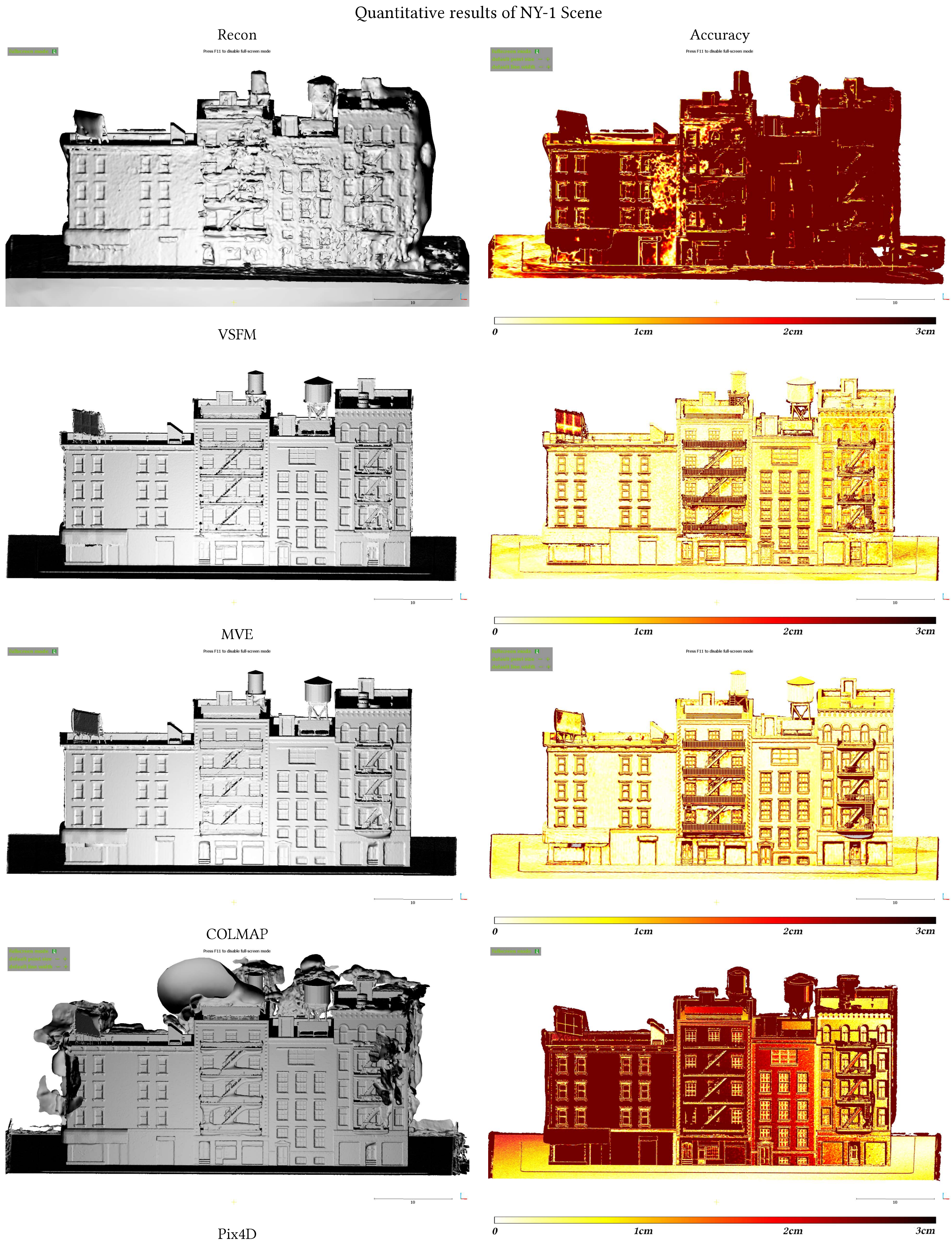


Flight Trajectories CA-1



Flight Trajectories UK-1

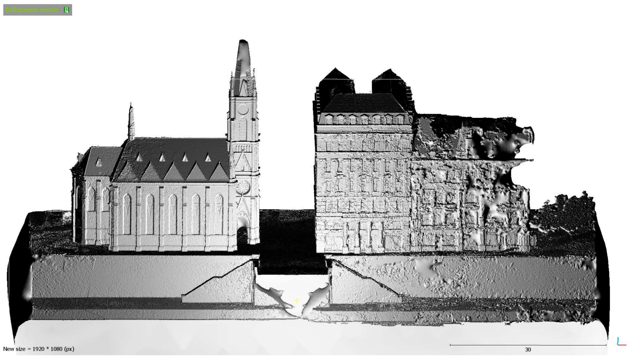


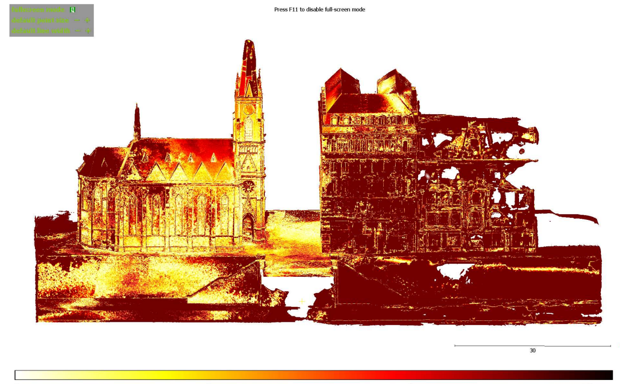

Quantitative results of NY-1 Scene
Recon
Accuracy
0
1cm
2cm
3cm
VSFM
0
1cm
2cm
3cm
MVE
0
1cm
2cm
3cm
COLMAP
0
1cm
2cm
3cm
Pix4D

Quantitative results of GOTH-1 Scene

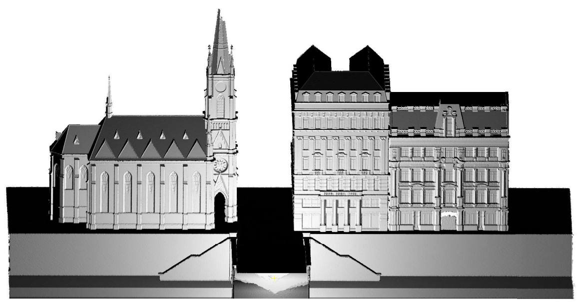
Recon

Accuracy

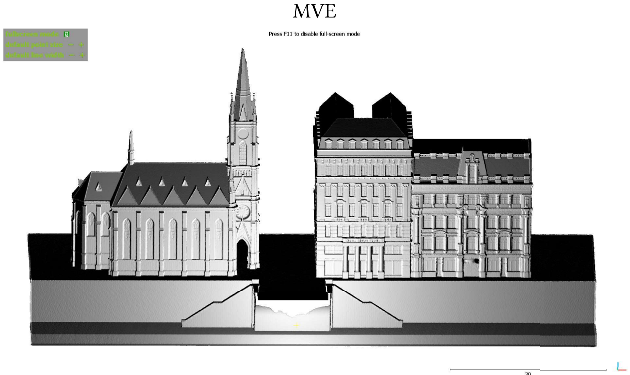


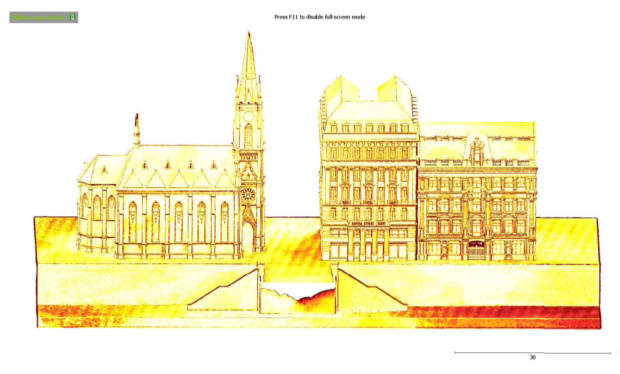


VSFM

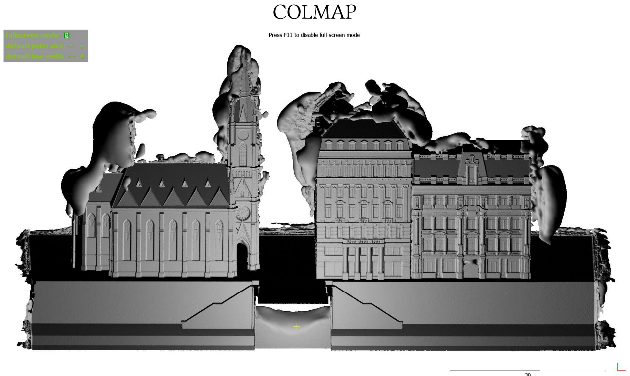


MVE



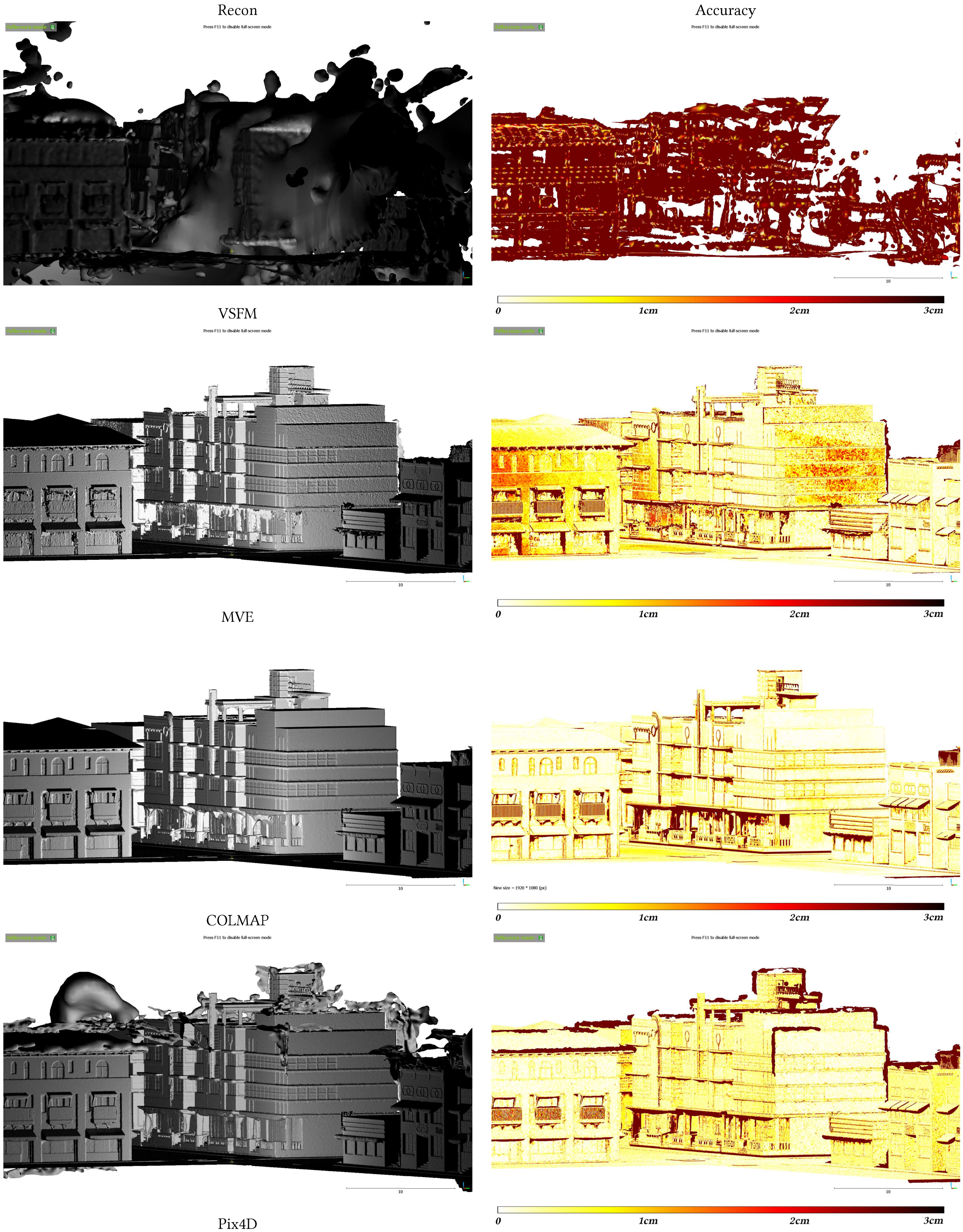


COLMAP

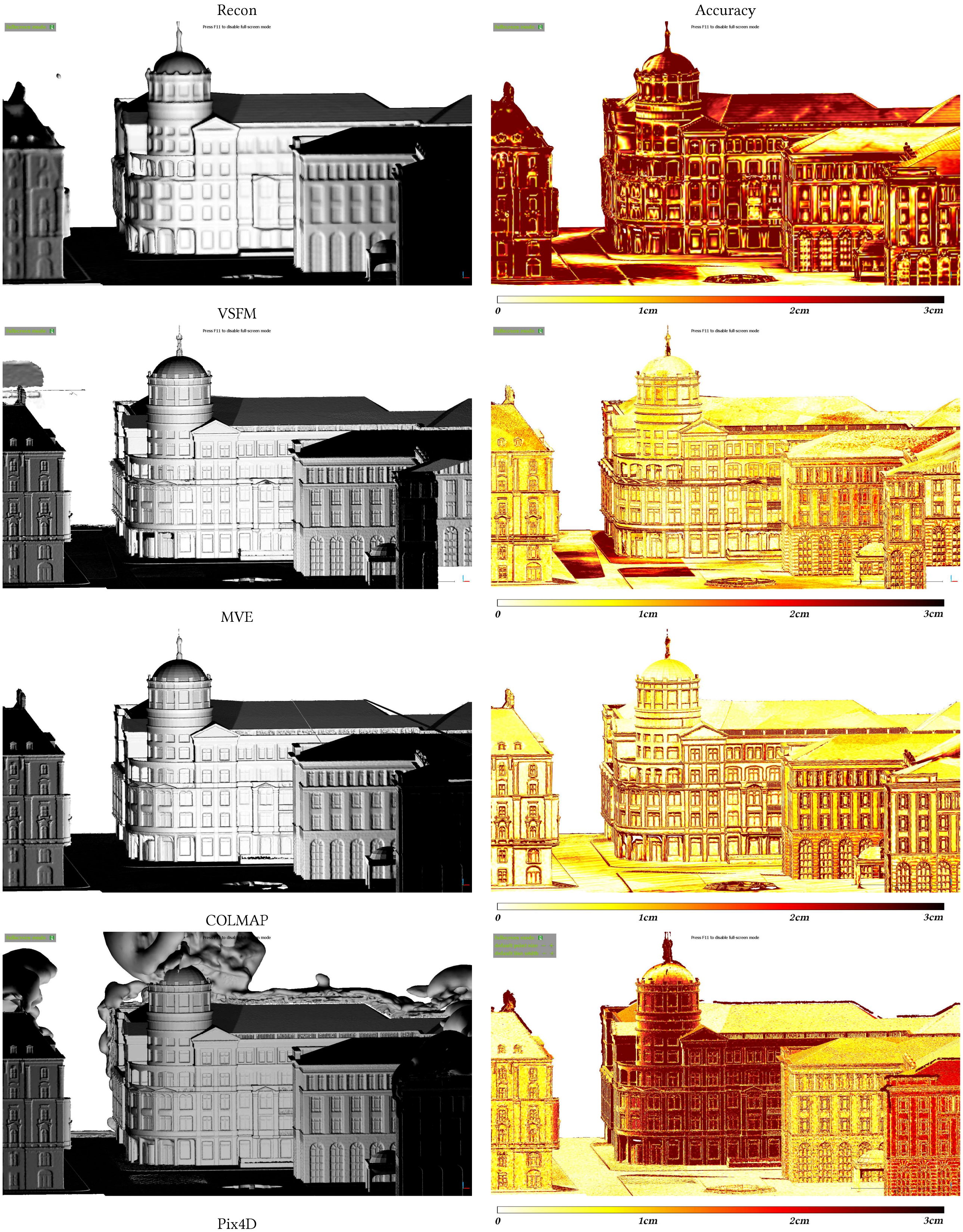


Pix4D

Quantitative results of CA-1 Scene

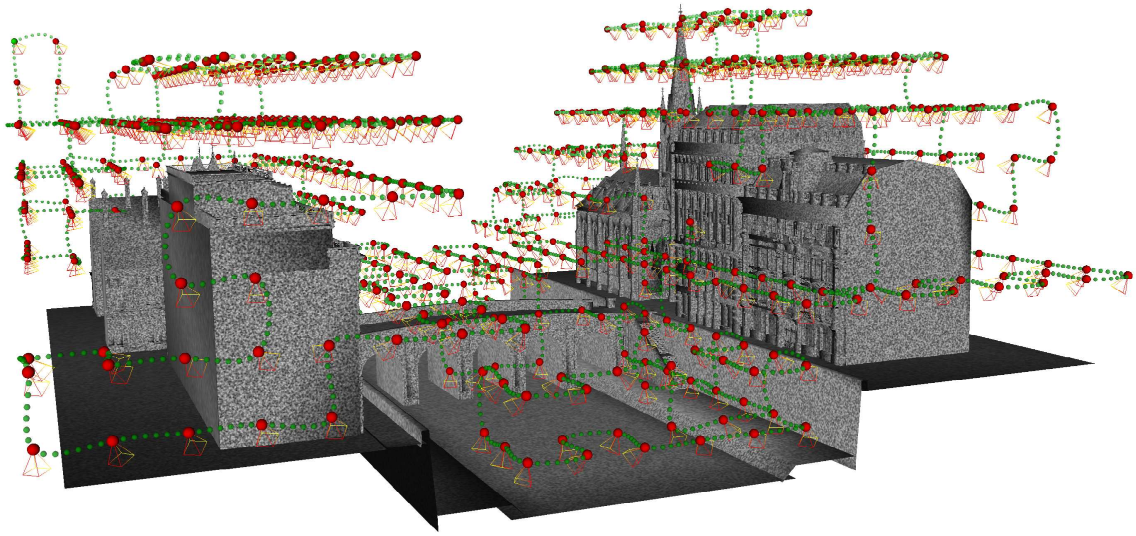


Quantitative results of UK-1 Scene

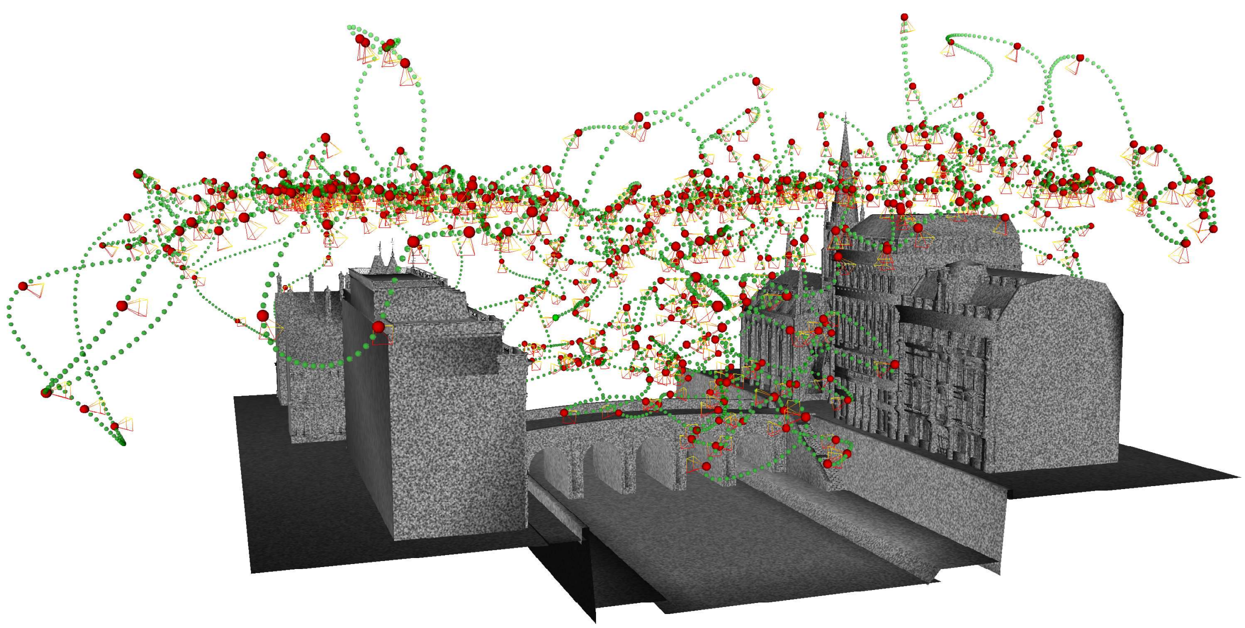


Comparison of GOTH-1 Scene Path Planning: Roberts et al. 2017 (SUB) vs. Ours

SUB

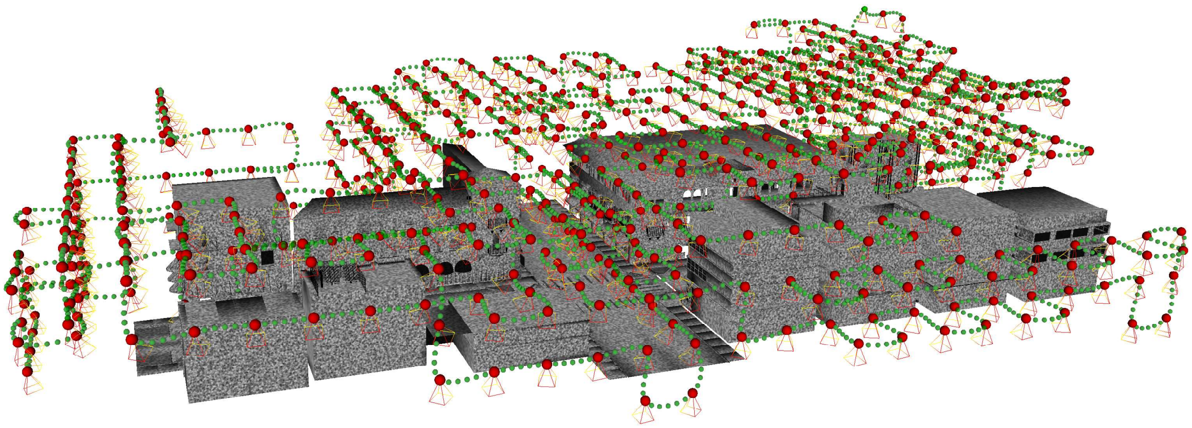


OURS

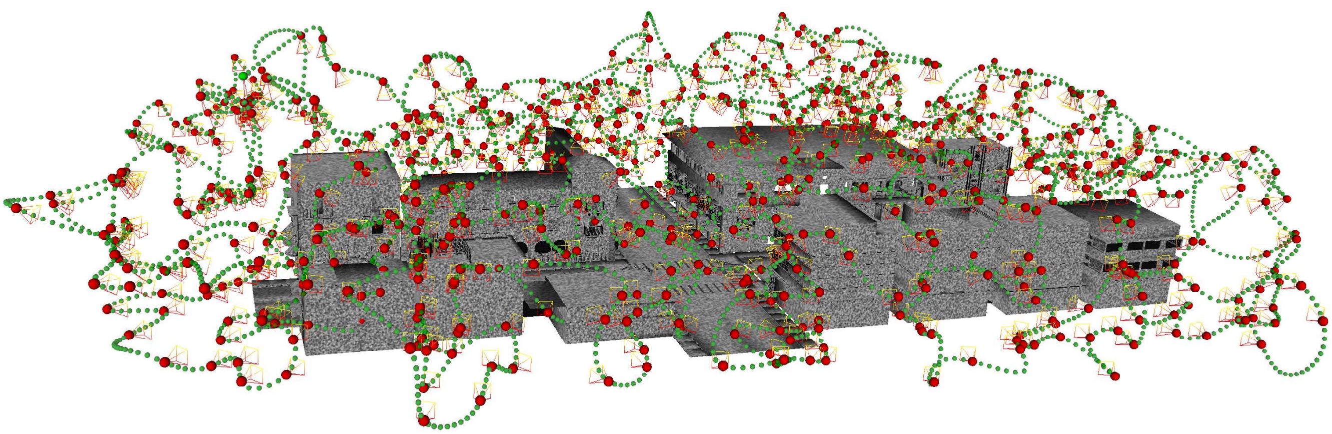


Comparison of CA-1 Scene Path Planning: Roberts et al. 2017 (SUB) vs. Ours

SUB

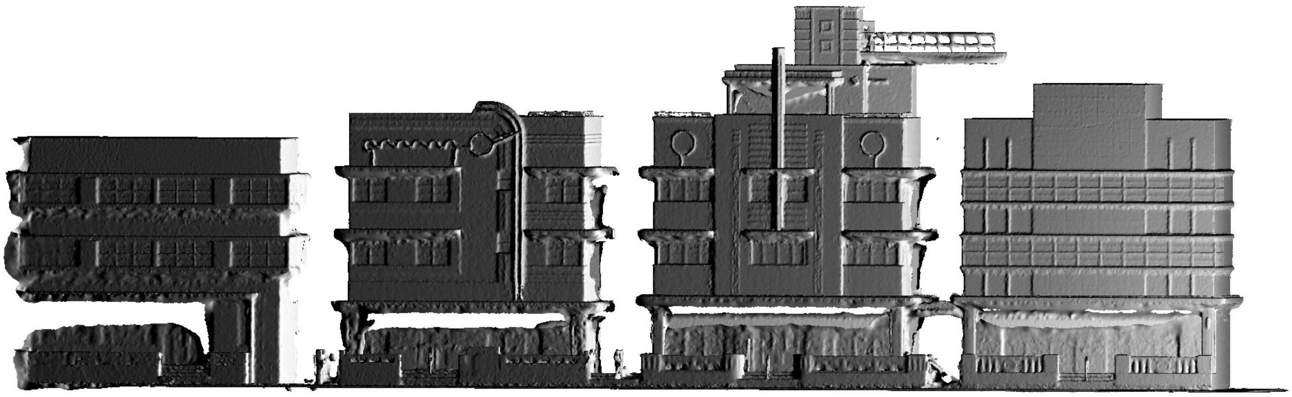


OURS

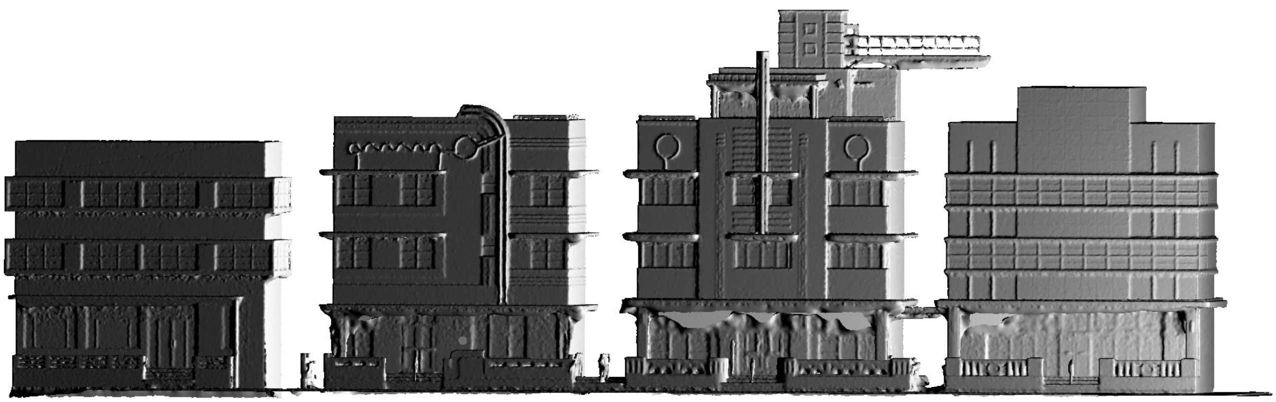


Comparison of CA-1 Scene using COLMAP: Roberts et al. 2017 (SUB), OURS, OURS with color textures

SUB



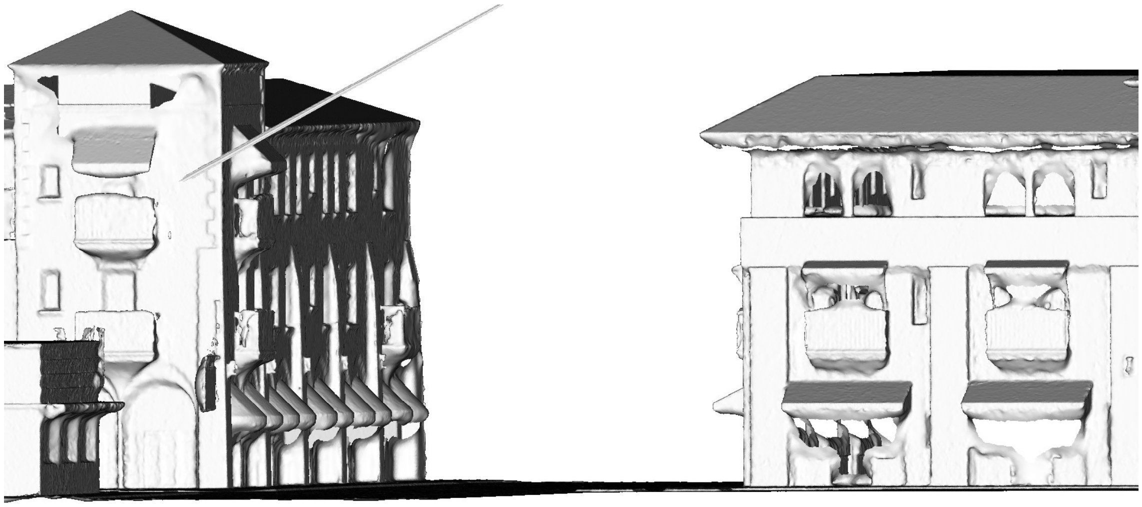
OURS



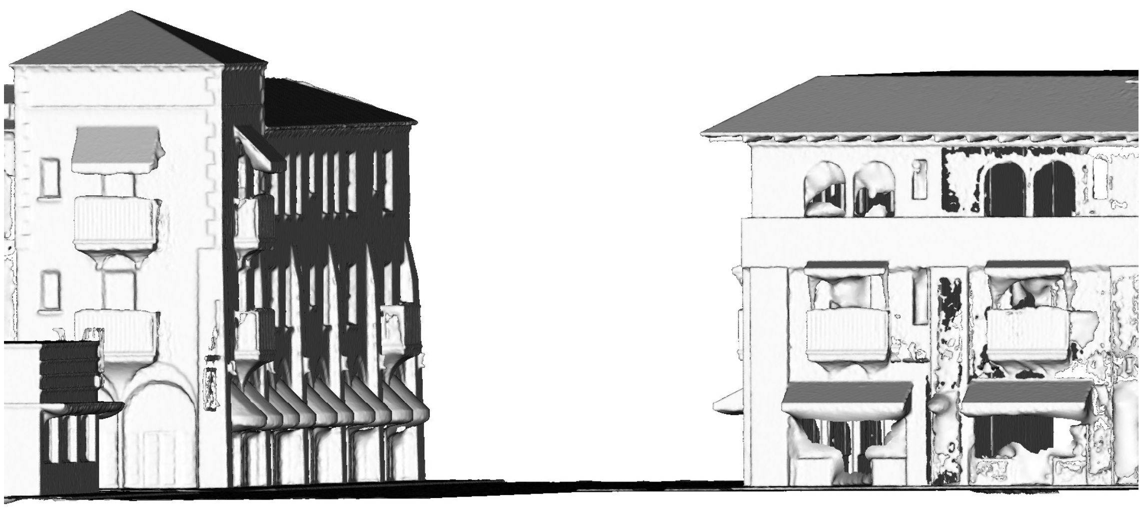
OURS (Color)

Comparison of CA-1 Scene using COLMAP: Roberts et al. 2017 (SUB), OURS, OURS with color textures

SUB



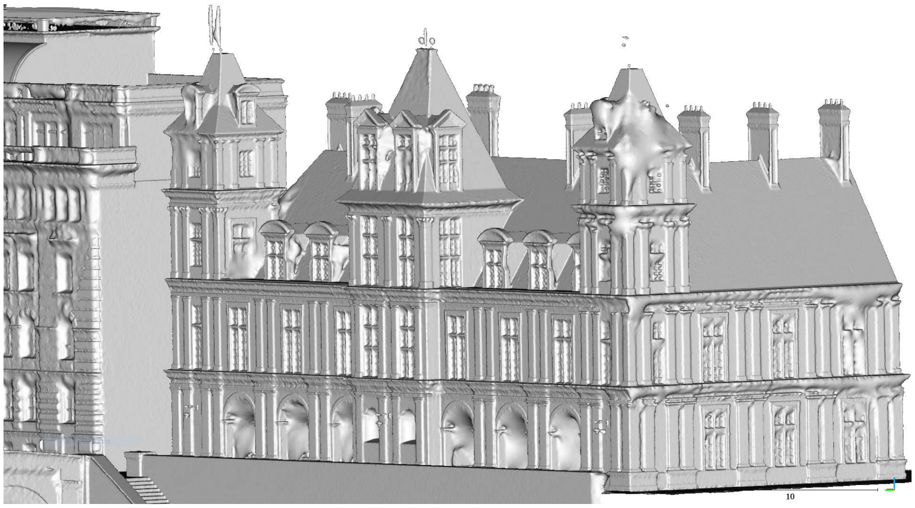
OURS



OURS (Color)

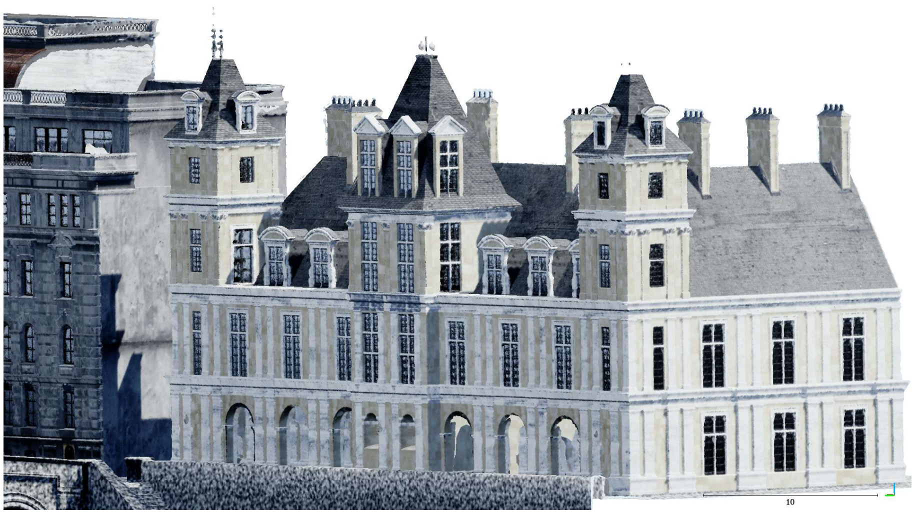
Comparison of GOTH-1 Scene using COLMAP: Roberts et al. 2017 (SUB), OURS, OURS with color textures

SUB



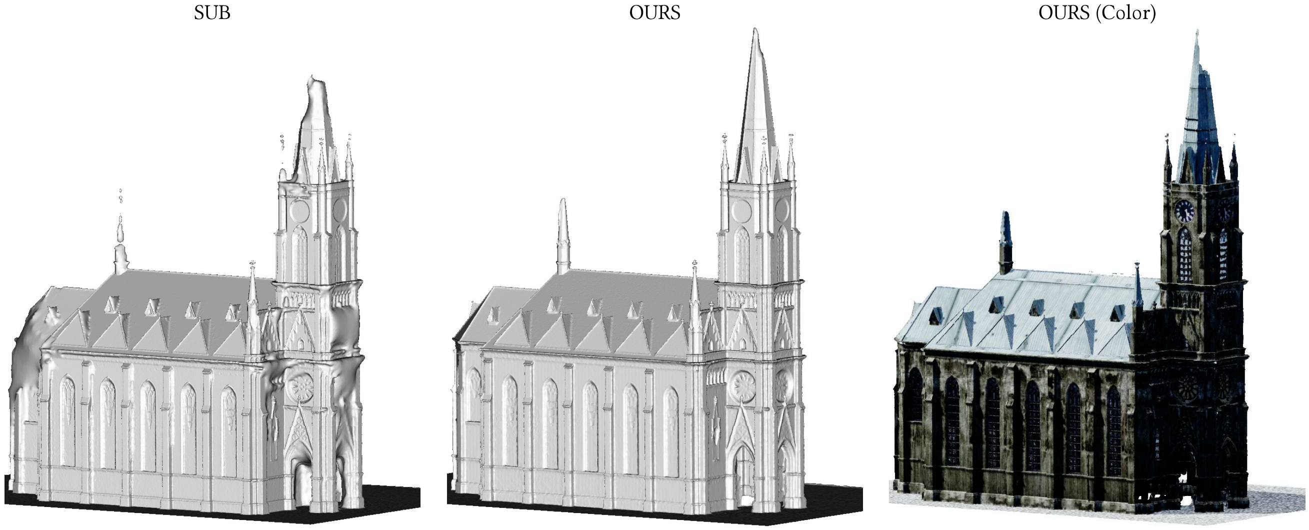
OURS

OURS (Color)



Comparison of GOTH-1 Scene Cathedral using COLMAP: Roberts et al. 2017 (SUB), OURS, OURS with color textures

SUB OURS OURS (Color)



Comparison of NY-1 Scene using MVE+FSSR: Perlin vs Color Textured regions

Perlin Color

Comparison of UK-1 Scene using COLMAP: Perlin vs Color Textured regions

Perlin Color