

**Structured Prediction of Unobserved Voxels From a Single Depth Image:
Supplementary Material**

Anonymous CVPR submission

Paper ID 632

Contents

1. Tabletop test dataset: Most used voxlets	2
2. Tabletop dataset: Voxlet sizes experiment	4
3. Tabletop dataset: Number of points sampled experiment	5
4. Tabletop dataset: Additional results on the test set	6
5. Training images	16

1. Tabletop test dataset: Most used voxlets

Figure 1 shows the frequency of use of all the voxlets predicted at test time, when running the algorithm on the Tabletop test set. This distribution includes both the short, ‘floating’ voxlets together with the taller voxlets which are fixed to the ground plane (as described in Section 5.2 of the main paper).

While there is a bias towards certain voxlets, there is also a good spread of frequency across the top voxlets. This lends weight to the hypothesis that a wide range of shapes is useful for modelling scenes, rather than use of simple primitives.

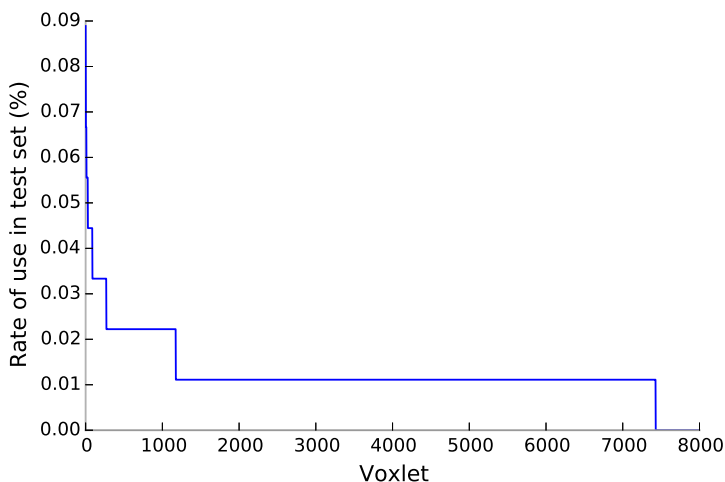


Figure 1. A distribution over the most popular voxlets in the testing dataset.

1.1. Selecting voxlets for training

When training on the Tabletop dataset, we train on 4 views from each of 60 scenes, making a total of 240 training images. From each training image we sample 1500 points, using the sampling strategy as described in Section 5 of the main paper. The voxlets extracted from each of these locations form the training set for one image. There are therefore $240 \times 1500 = 360000$ items in the training set. However, not all of these can be proposed as a voxel at a test-time location, as only the medoid of each leaf node is used for prediction.

1.2. Depiction of high and low usage voxlets

Figure 2 shows a selection of the most and least used voxlets predicted when running on the test dataset. The different types of voxel shown in each render (‘tall’ or ‘floating’) is noticeable by the varying sizes of the bounding boxes.

Some of the most used voxlets resemble primitive shapes, such as cylinders and cuboids. This is expected, as we would expect these to be used in many locations in the test scenes. However, others in the top voxlets appear to be very specialised, e.g. those at rank 1 and 50. An advantage of Voxlets over techniques which fit primitives is that we are able to learn and fit both primitive and specialist shapes, which in combination can be used to model the shape of novel structures.

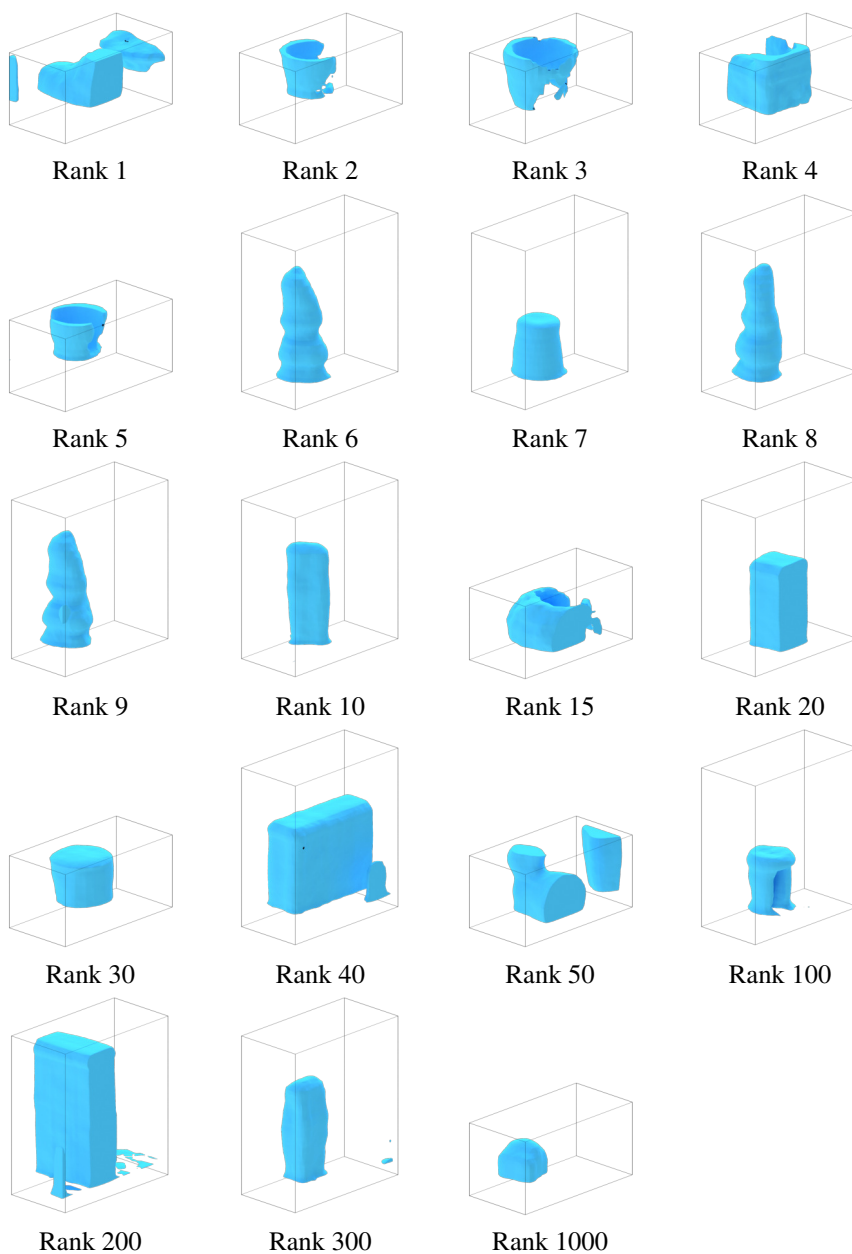


Figure 2. Some of the most and least popular voxels, when running the algorithm on the Tabletop test set.

2. Tabletop dataset: Voxlet sizes experiment

Here we present the effect of varying the voxel size parameter x , as defined in Section 5.2 of the main paper. We note that varying x effects a trade-off between precision and recall. The IoU score is maximised on the test set at $x = 0.1$. For our experiments we used $x = 0.15$. Clearly, parameter tuning on an appropriate validation set could provide better results.

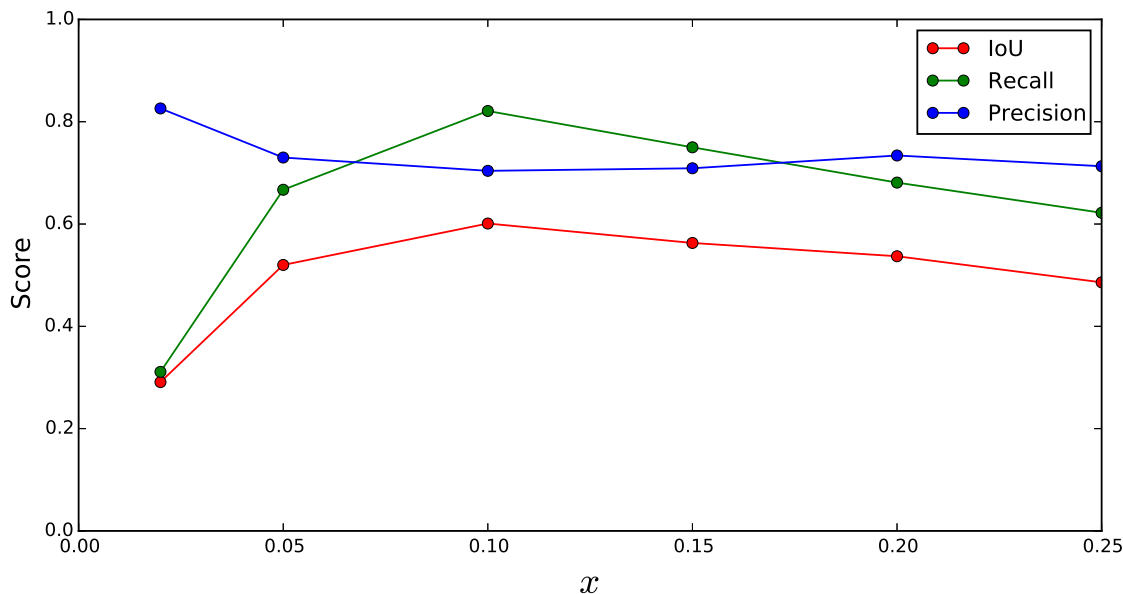


Figure 3. The effect of varying the parameter x on the tabletop dataset

3. Tabletop dataset: Number of points sampled experiment

Here we present the effect of varying the number of points sampled when making predictions into the scene. We note that precision remains reasonably consistent no matter how many points are sampled, while recall and IoU require at least 100 points sampled for equivalent results to those presented in the paper. For our experiments we sampled 300 points for each test image.

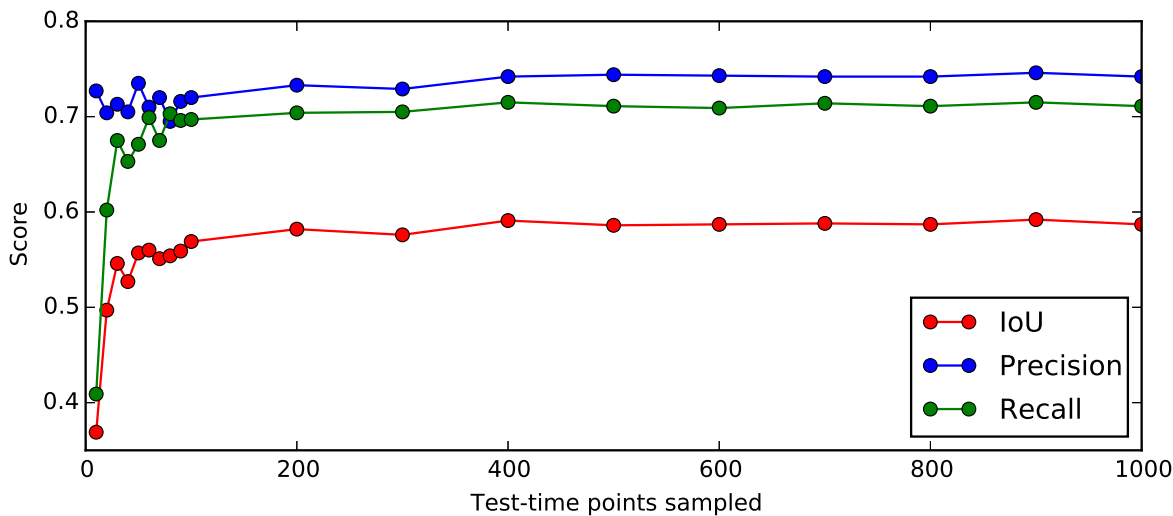


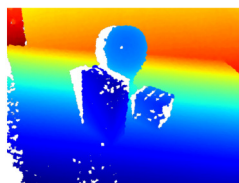
Figure 4. The effect of varying the number of test-time points sampled, on the tabletop dataset

4. Tabletop dataset: Additional results on the test set

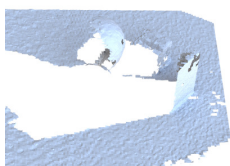
4.1. Sequence: saved_00219_[666]



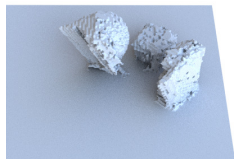
Input RGB



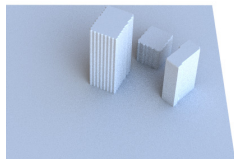
Input Depth



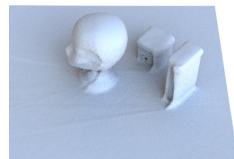
Observed surfaces



Zheng et al. (GT)



Bounding box (GT)

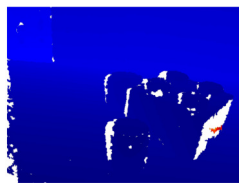
**Voxlets**

Ground truth

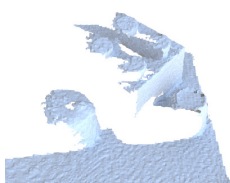
4.2. Sequence: saved_00238_[403]



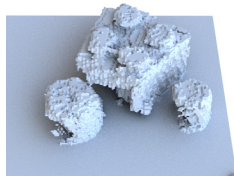
Input RGB



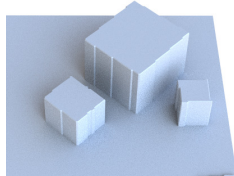
Input Depth



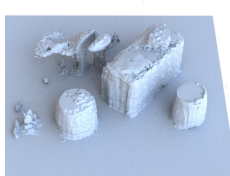
Observed surfaces



Zheng et al. (GT)



Bounding box (GT)

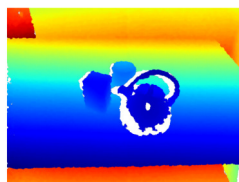
**Voxlets**

Ground truth

4.3. Sequence: saved_00237_[136]



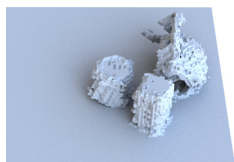
Input RGB



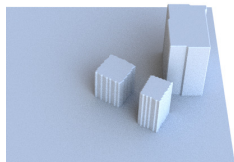
Input Depth



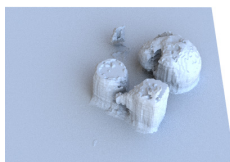
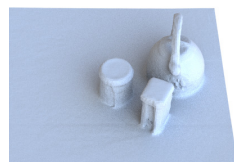
Observed surfaces



Zheng et al. (GT)



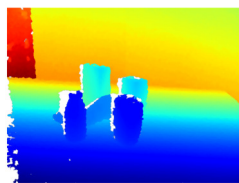
Bounding box (GT)

**Voxlets**

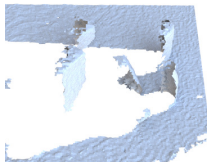
Ground truth

4.4. Sequence: saved_00231_[55]

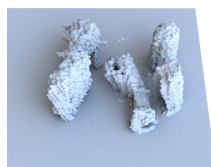
Input RGB



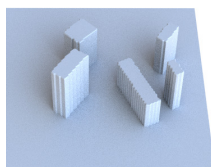
Input Depth



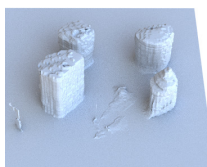
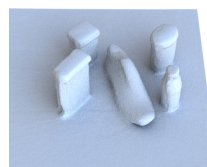
Observed surfaces



Zheng et al. (GT)



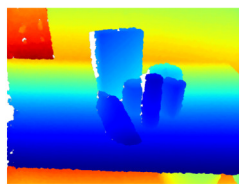
Bounding box (GT)

**Voxlets**

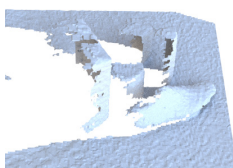
Ground truth

4.5. Sequence: saved_00236_[147]

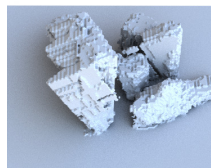
Input RGB



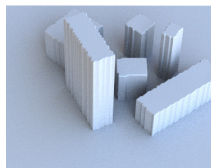
Input Depth



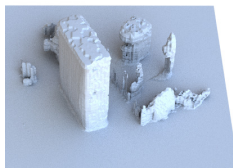
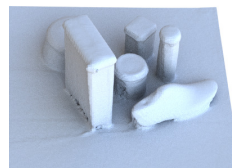
Observed surfaces



Zheng et al. (GT)



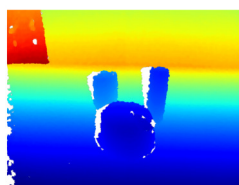
Bounding box (GT)

**Voxlets**

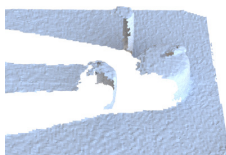
Ground truth

4.6. Sequence: saved_00243_[532]

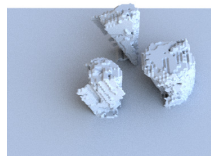
Input RGB



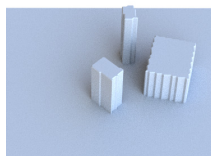
Input Depth



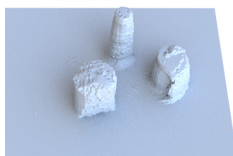
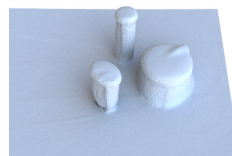
Observed surfaces



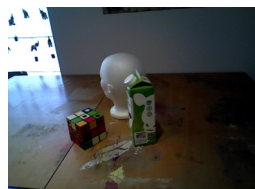
Zheng et al. (GT)



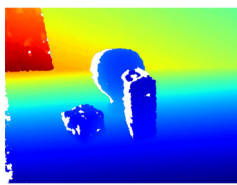
Bounding box (GT)

**Voxlets**

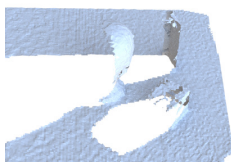
Ground truth

4.7. Sequence: saved_00216_[548]

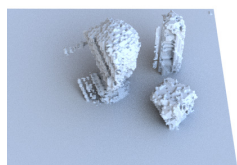
Input RGB



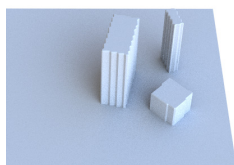
Input Depth



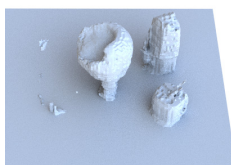
Observed surfaces



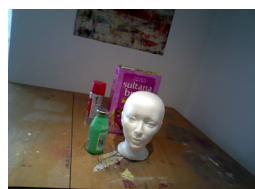
Zheng et al. (GT)



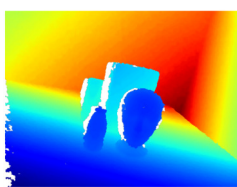
Bounding box (GT)

**Voxlets**

Ground truth

4.8. Sequence: saved_00233_[134]

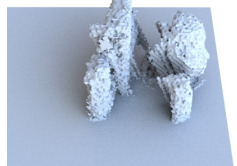
Input RGB



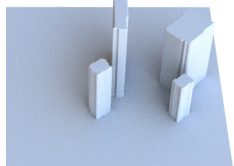
Input Depth



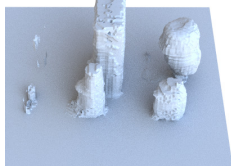
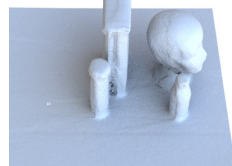
Observed surfaces



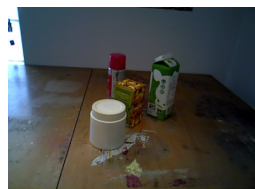
Zheng et al. (GT)



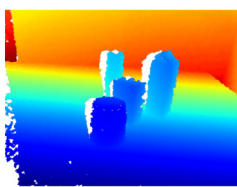
Bounding box (GT)

**Voxlets**

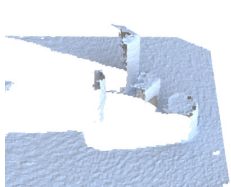
Ground truth

4.9. Sequence: saved_00234_[419]

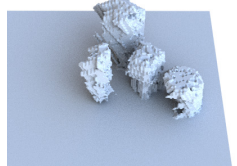
Input RGB



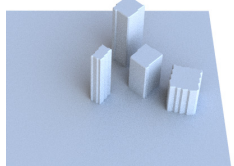
Input Depth



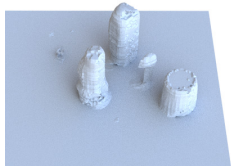
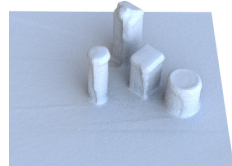
Observed surfaces



Zheng et al. (GT)



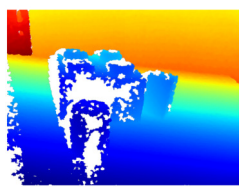
Bounding box (GT)

**Voxlets**

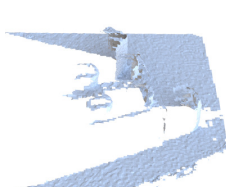
Ground truth

4.10. Sequence: saved_00215_[364]

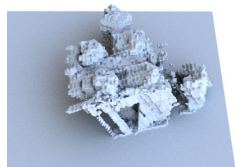
Input RGB



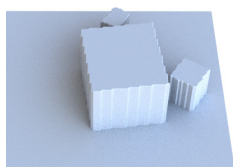
Input Depth



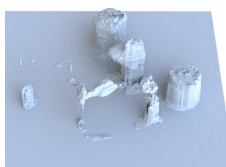
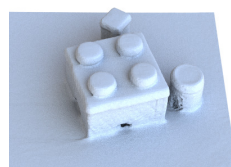
Observed surfaces



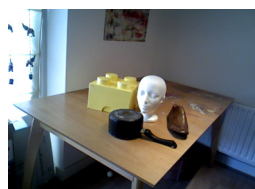
Zheng et al. (GT)



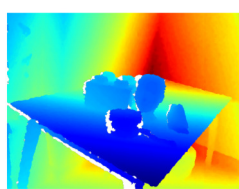
Bounding box (GT)

**Voxlets**

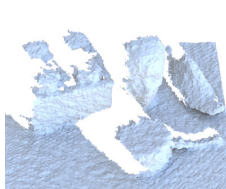
Ground truth

4.11. Sequence: saved_00196_[663]

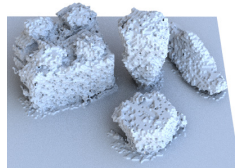
Input RGB



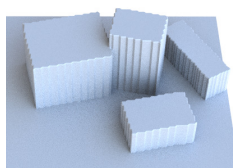
Input Depth



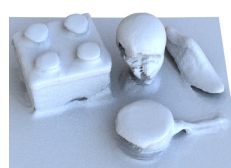
Observed surfaces



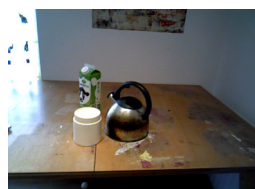
Zheng et al. (GT)



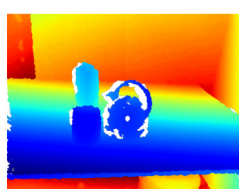
Bounding box (GT)

**Voxlets**

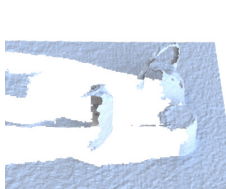
Ground truth

4.12. Sequence: saved_00221_[106]

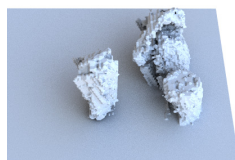
Input RGB



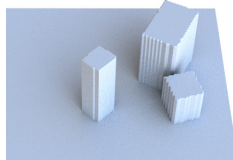
Input Depth



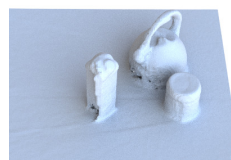
Observed surfaces



Zheng et al. (GT)

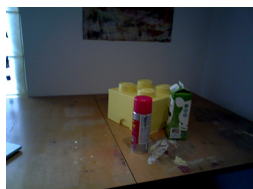


Bounding box (GT)

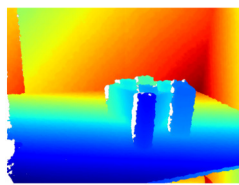
**Voxlets**

Ground truth

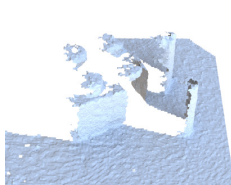
4.13. Sequence: saved_00239_[100]



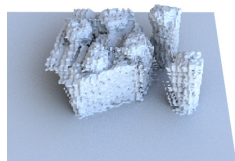
Input RGB



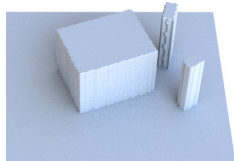
Input Depth



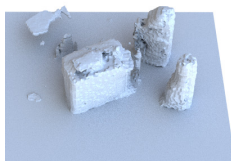
Observed surfaces



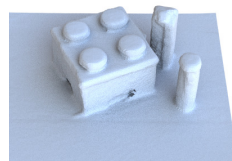
Zheng et al. (GT)



Bounding box (GT)



Voxlets

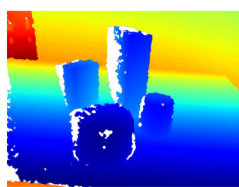


Ground truth

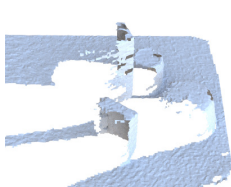
4.14. Sequence: saved_00211_[469]



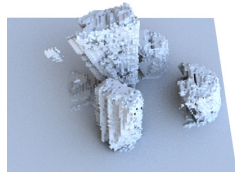
Input RGB



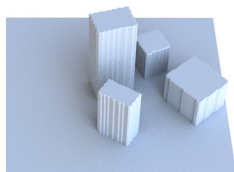
Input Depth



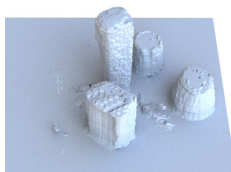
Observed surfaces



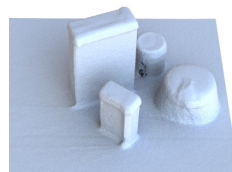
Zheng et al. (GT)



Bounding box (GT)

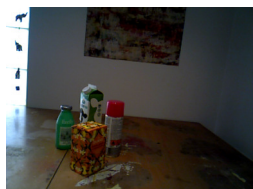


Voxlets

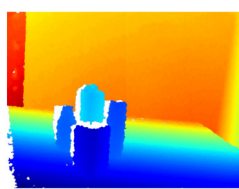


Ground truth

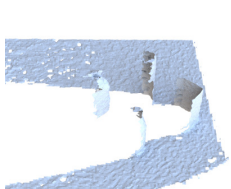
4.15. Sequence: saved_00229_[63]



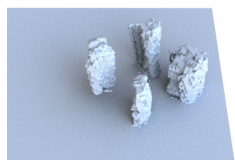
Input RGB



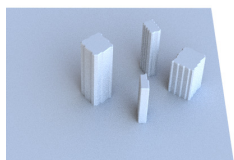
Input Depth



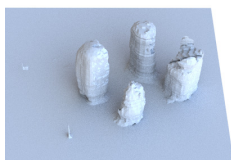
Observed surfaces



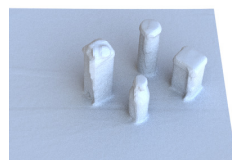
Zheng et al. (GT)



Bounding box (GT)



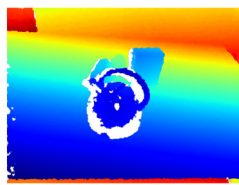
Voxlets



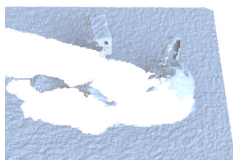
Ground truth

4.16. Sequence: saved_00224_[616]

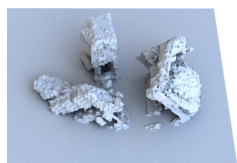
Input RGB



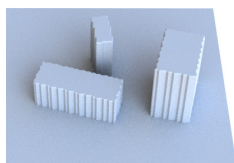
Input Depth



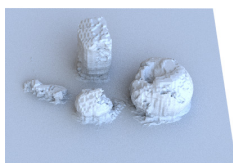
Observed surfaces



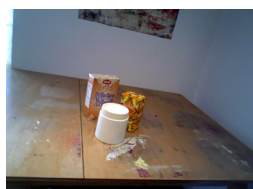
Zheng et al. (GT)



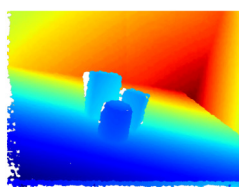
Bounding box (GT)

**Voxlets**

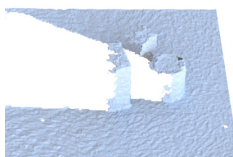
Ground truth

4.17. Sequence: saved_00227_[120]

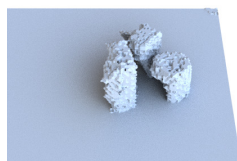
Input RGB



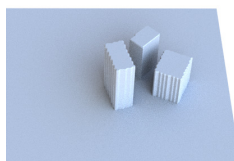
Input Depth



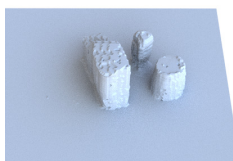
Observed surfaces



Zheng et al. (GT)



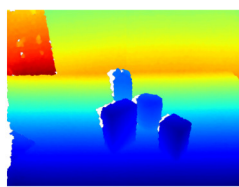
Bounding box (GT)

**Voxlets**

Ground truth

4.18. Sequence: saved_00241_[103]

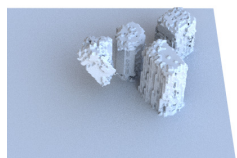
Input RGB



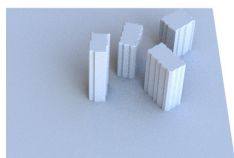
Input Depth



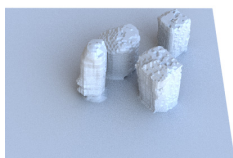
Observed surfaces



Zheng et al. (GT)

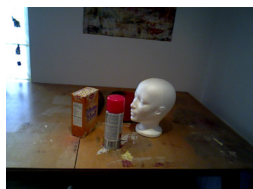


Bounding box (GT)

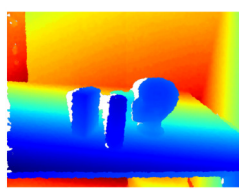
**Voxlets**

Ground truth

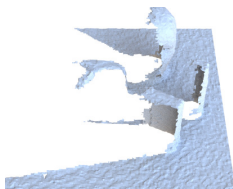
4.19. Sequence: saved_00218_[121]



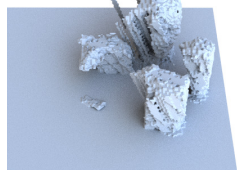
Input RGB



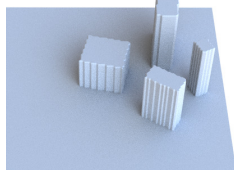
Input Depth



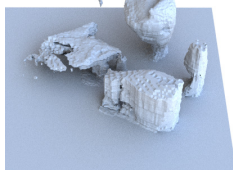
Observed surfaces



Zheng et al. (GT)



Bounding box (GT)



Voxlets

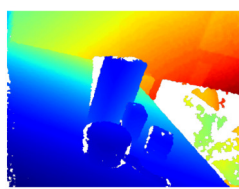


Ground truth

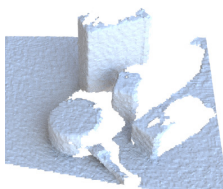
4.20. Sequence: saved_00197_[633]



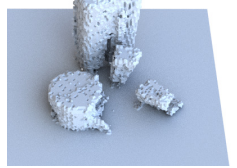
Input RGB



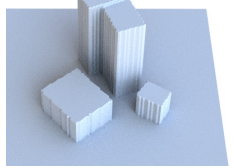
Input Depth



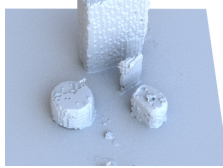
Observed surfaces



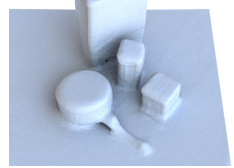
Zheng et al. (GT)



Bounding box (GT)

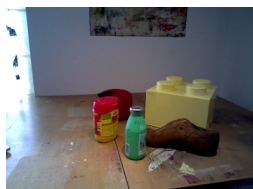


Voxlets

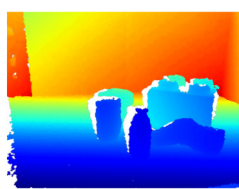


Ground truth

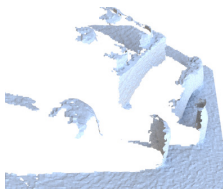
4.21. Sequence: saved_00232_[122]



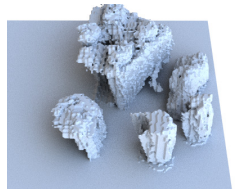
Input RGB



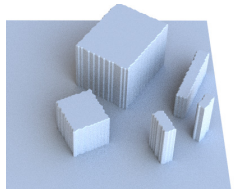
Input Depth



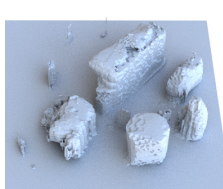
Observed surfaces



Zheng et al. (GT)



Bounding box (GT)

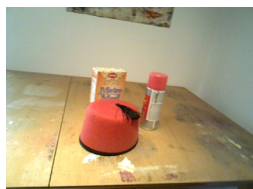


Voxlets

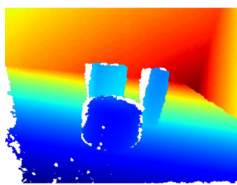


Ground truth

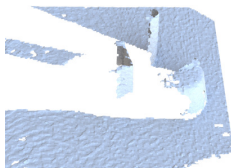
4.22. Sequence: saved_00205_[331]



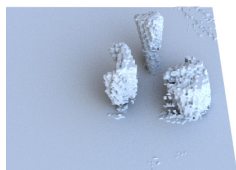
Input RGB



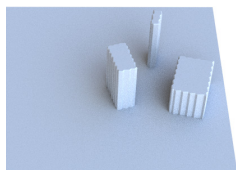
Input Depth



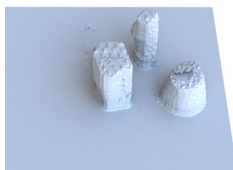
Observed surfaces



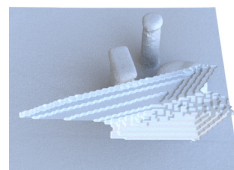
Zheng et al. (GT)



Bounding box (GT)



Voxlets

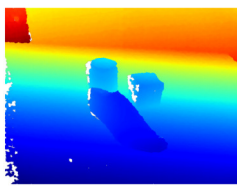


Ground truth

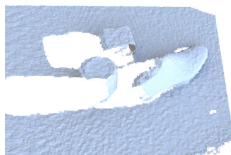
4.23. Sequence: saved_00242_[464]



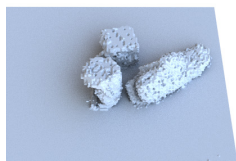
Input RGB



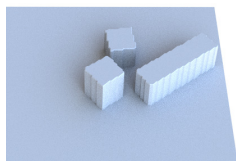
Input Depth



Observed surfaces



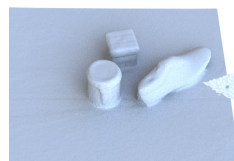
Zheng et al. (GT)



Bounding box (GT)



Voxlets

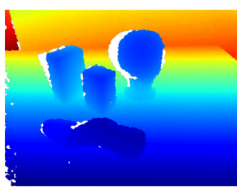


Ground truth

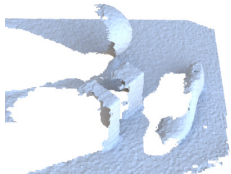
4.24. Sequence: saved_00225_[673]



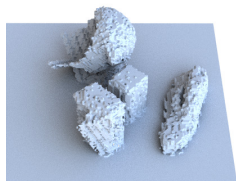
Input RGB



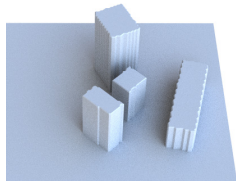
Input Depth



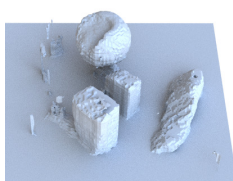
Observed surfaces



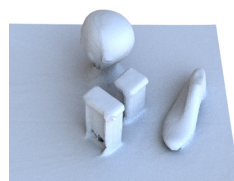
Zheng et al. (GT)



Bounding box (GT)



Voxlets

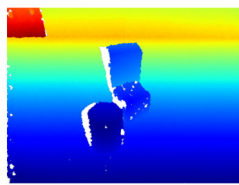


Ground truth

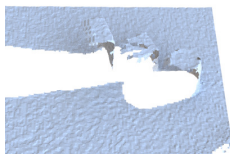
4.25. Sequence: saved_00223_[524]



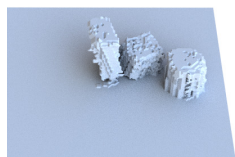
Input RGB



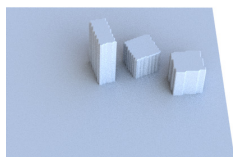
Input Depth



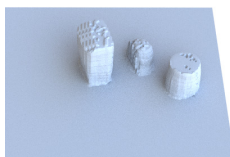
Observed surfaces



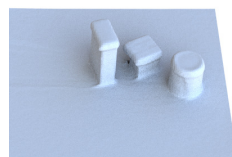
Zheng et al. (GT)



Bounding box (GT)

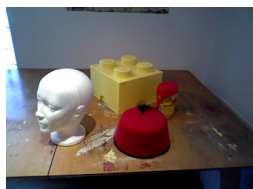


Voxlets

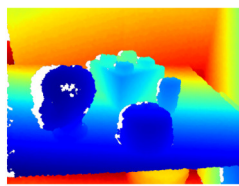


Ground truth

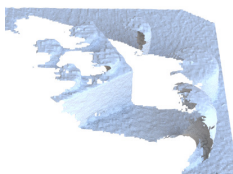
4.26. Sequence: saved_00208_[531]



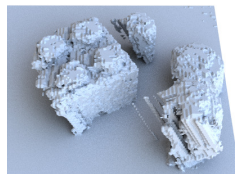
Input RGB



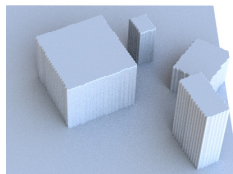
Input Depth



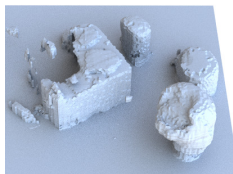
Observed surfaces



Zheng et al. (GT)



Bounding box (GT)



Voxlets

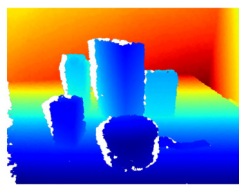


Ground truth

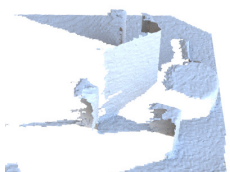
4.27. Sequence: saved_00207_[536]



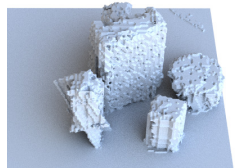
Input RGB



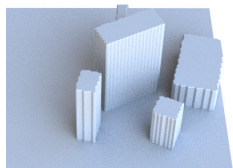
Input Depth



Observed surfaces



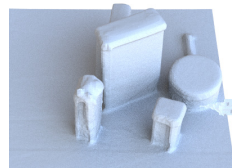
Zheng et al. (GT)



Bounding box (GT)



Voxlets

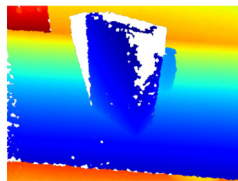


Ground truth

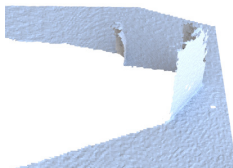
4.28. Sequence: saved_00220_[552]



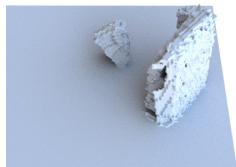
Input RGB



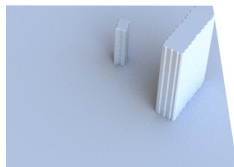
Input Depth



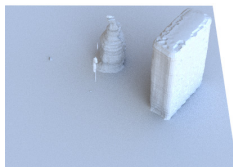
Observed surfaces



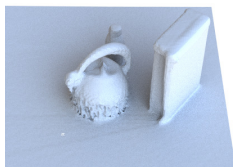
Zheng et al. (GT)



Bounding box (GT)



Voxlets

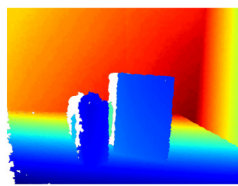


Ground truth

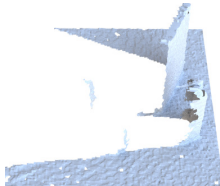
4.29. Sequence: saved_00228_[58]



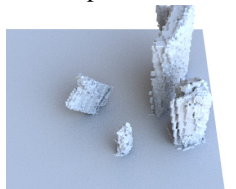
Input RGB



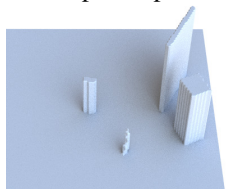
Input Depth



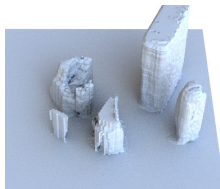
Observed surfaces



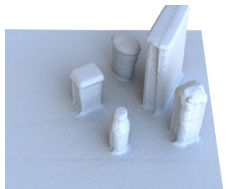
Zheng et al. (GT)



Bounding box (GT)

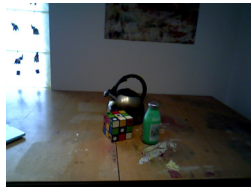


Voxlets

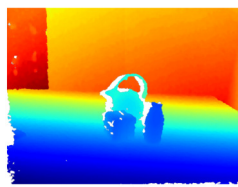


Ground truth

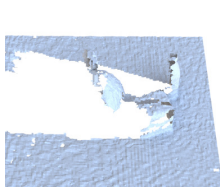
4.30. Sequence: saved_00230_[45]



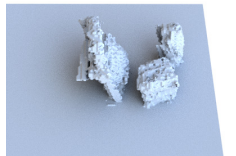
Input RGB



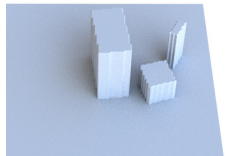
Input Depth



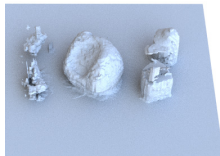
Observed surfaces



Zheng et al. (GT)



Bounding box (GT)



Voxlets



Ground truth

5. Training images

Here we present a single RGB image from of each training sequence. Each training sequence consists of several hundred images of the scene, which we fuse to get the ground truth volume. For our algorithm we selected four images from each sequence to be used as images in the training set.

Note that the images in the first half of the training set were captured under low light conditions, hence are darker in color than those in the second half.

