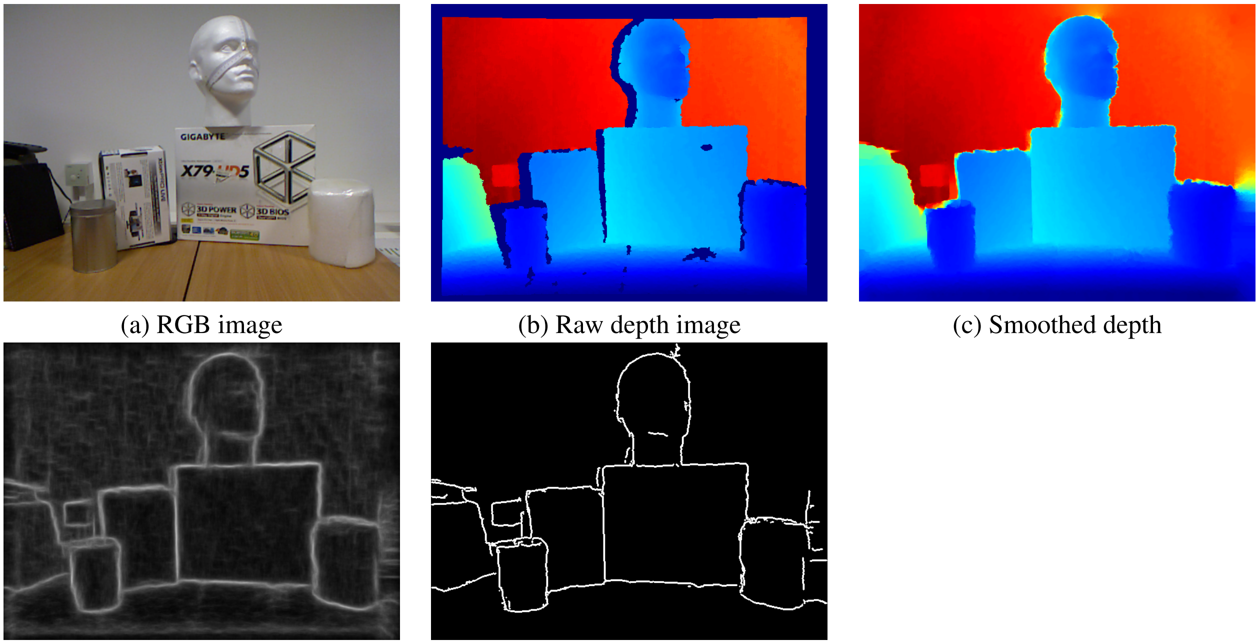
Structured Prediction of Unobserved Voxels From a Single Depth Image: Supplementary Material

Anonymous CVPR submission

Paper ID 1437

1. Full pipeline of images

In Figure 1 we show the depth image preprocessing pipeline we use before we make our predictions. The noisy depth image (b) is smoothed, and missing data is filled (c). This makes use of the RGB image in the cross-bilateral filtering, where we use the implementation provided by [3]. We then use the structured edge detection model from [2] to compute a real-valued edge map (d) for the image, which is finally binarized (e) using the Canny edge hysteresis method [1].



(a) RGB image

(b) Raw depth image

(c) Smoothed depth

(d) Structured edges

(e) Binary edge map

Figure 1. Pre-processing pipeline

2. ‘Bigbird’ turntable dataset: Additional results

Figure 2 shows additional results from the Bigbird turntable dataset. We note that we are able to gain good reconstructions even where the object suffers from heavy self-occlusions (e.g. (c) and (d)). However, where much of the data is missing through the height of the object, we can fail to recover the main bulk (e.g. (f) and (g)). We note that the baseline algorithms can also perform poorly under such conditions, under- or over- predicting the volume.

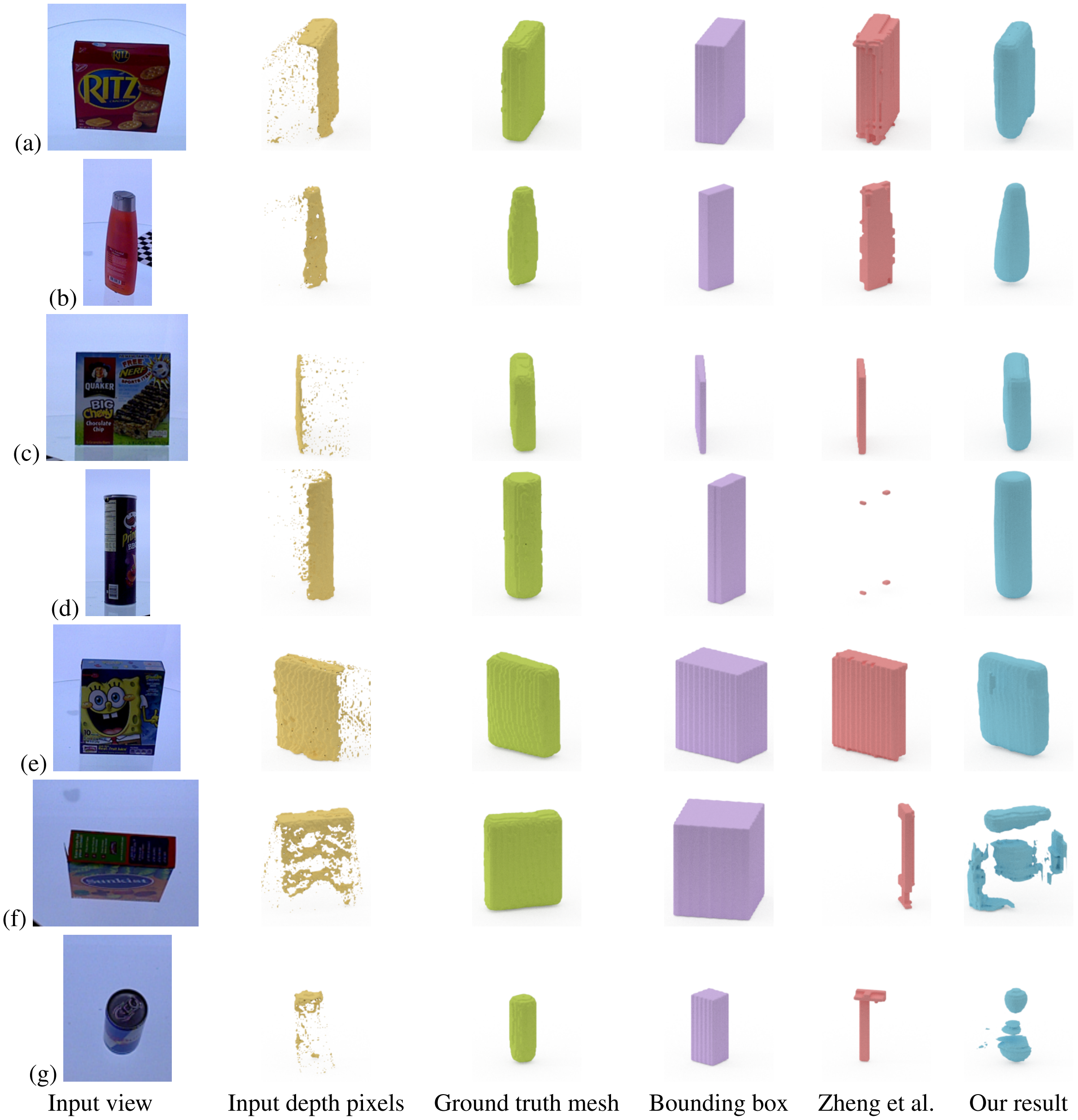


Figure 2. Additional results from the Bigbird dataset

3. Slices through the TSDF

For most of our images, we have displayed the zero level-set from the truncated signed distance function (TSDF). However, the underlying TSDF could be useful for some application. It additionally provides some introspection to reveal the inner workings of the algorithm. We also note that using marching cubes to find the level set is a fairly naive version of regularization.

The images in Figures 3 and 4 show slices through the TSDF volume at different heights in a prediction. These test scenes fall well outside the scope of our training data, which was much smaller objects on a turntable. However, these images show the ability of our algorithm to degrade gracefully in the presence of novel types of data. It can be seen that some regions are predicted to be occupied (red regions) while actually landing in a region known to be empty, given the camera image. Sometimes this occurs where predictions are made from points above or below the slice being viewed. We have not in these images made any attempt to remove these regions.

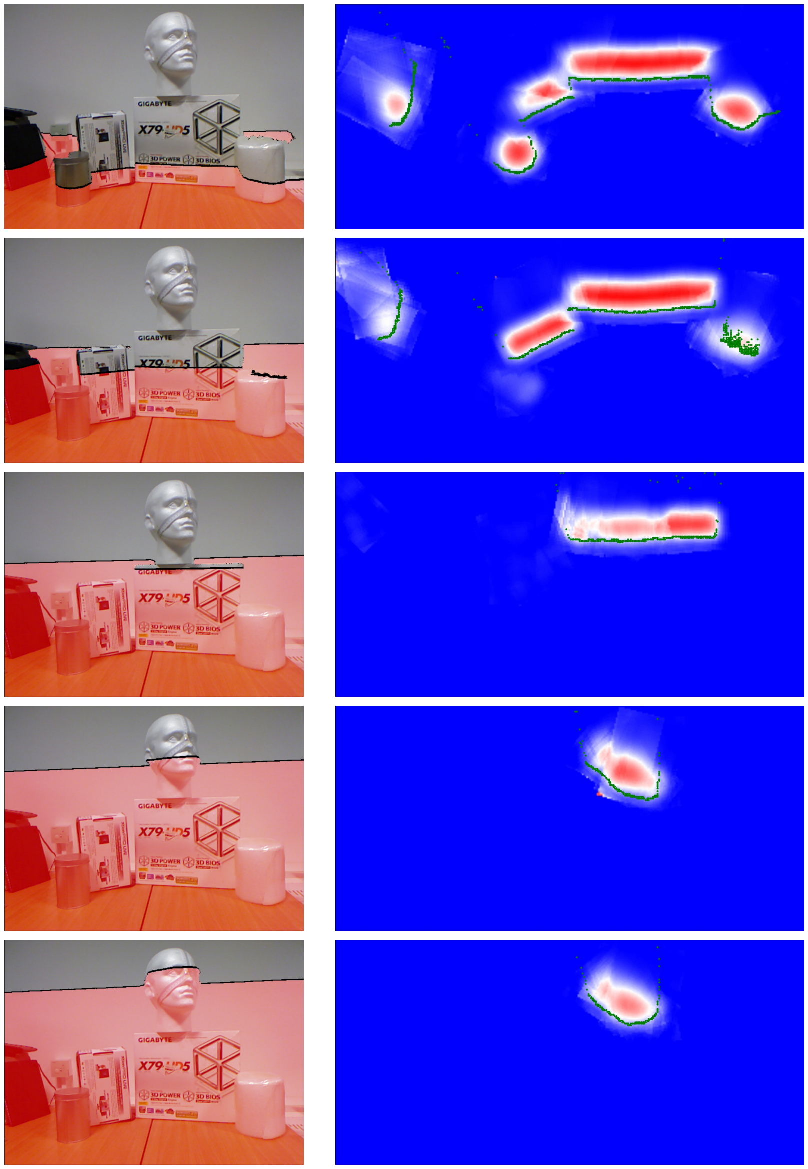


Figure 3. Slices through a TSDF prediction. The black line on the raw image on the left of each image pair indicates the height of the slice, while the image on the right shows the values of the TSDF. The TSDF values range from -3cm (red) to +3cm (blue). The zero level-set falls in the middle of the white region. The green dots show the position of the raw Kinect input in the indicated region.

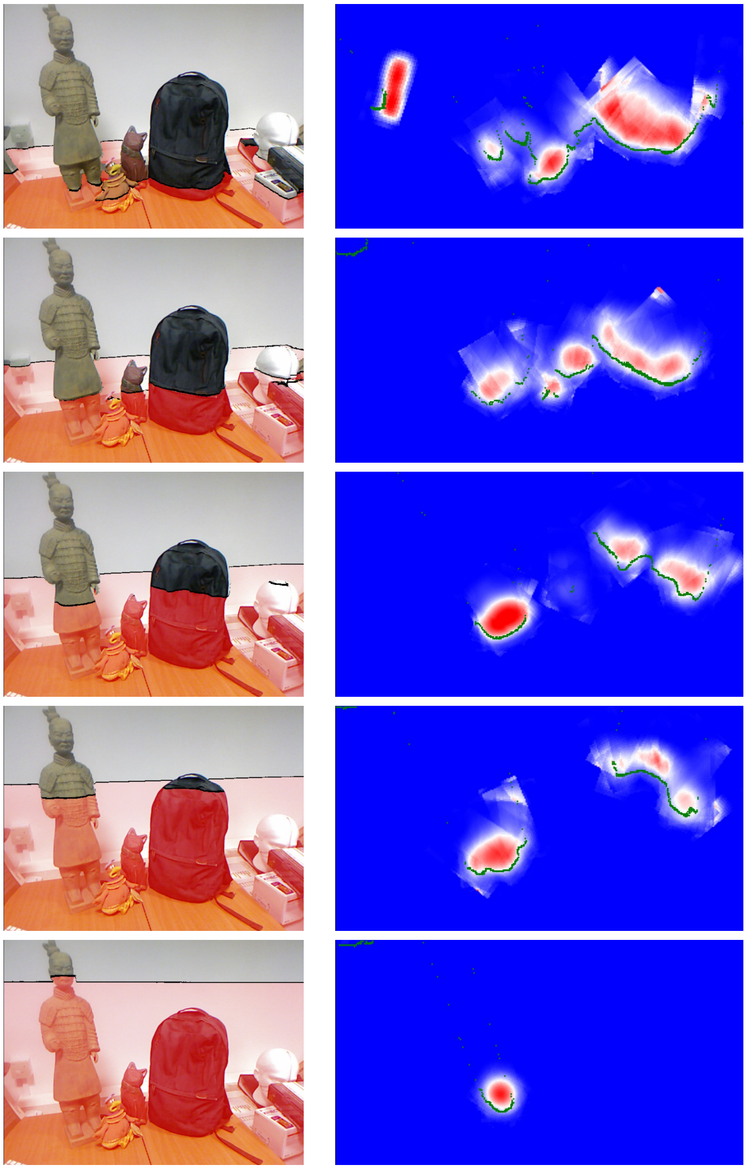


Figure 4. Slices through a TSDF prediction. Colors and images are as described in the caption of Figure 3.

References

[1] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence (PAMI)*, 1986.

[2] P. Dollár and C. L. Zitnick. Structured forests for fast edge detection. In *International Conference on Computer Vision (ICCV)*, 2013.

[3] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from RGBD images. In *European Conference on Computer Vision (ECCV)*, 2012.