Multiple View Object Cosegmentation using Appearance and Stereo Cues

Adarsh Kowdle[1], Sudipta N. Sinha[2], and Richard Szeliski[2]

[1]Cornell University, Ithaca, NY, USA
apk64@cornell.edu
[2]Microsoft Research, Redmond, WA, USA
{sudipsin,szeliski}@microsoft.com

1 Additional Results

In the supplementary material, we present detailed experimental results on the eight datasets reported in the paper. These are presented in the following order.

1. The piecewise planar labeling, the associated depth maps and foreground segmentations recovered by our method for a few images from each of the eight datasets are visualized (see Fig 1-8).
2. A comparison of our approach against [3] and [1] is shown in Fig 9. These are the images used to report our quantitative accuracy in Fig. 6 of the main paper.

References

1. Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *CVPR*, 2010.
2. C. Rother, V. Kolmogorov, and A. Blake. Grabcut: interactive foreground extraction using iterated graph cuts. In *SIGGRAPH*, 2004.
3. S. Vicente, C. Rother, and V. Kolmogorov. Object cosegmentation. In *CVPR*, 2011.

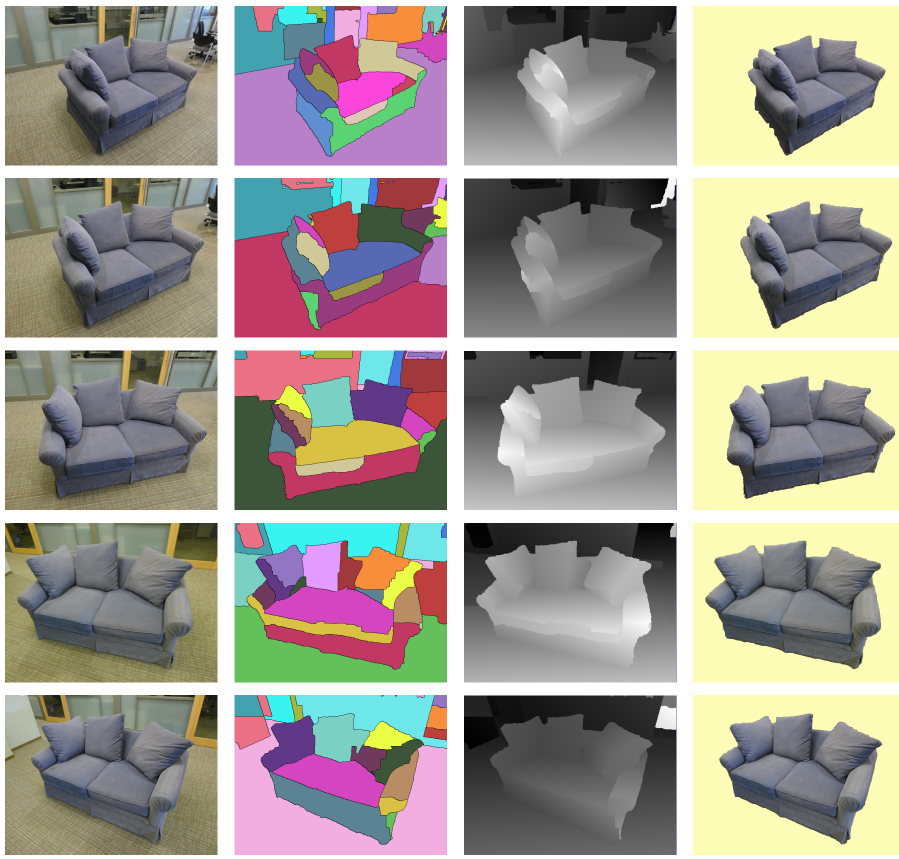

Fig. 1: **COUCH DATASET** (Five out of 9 images): For each of the images shown in COLUMN 1, the corresponding piecewise planar labeling and the induced depth maps recovered by our method are shown in COLUMN 2 and COLUMN 3 respectively. The final segmentation of the foreground object is shown in COLUMN 4.

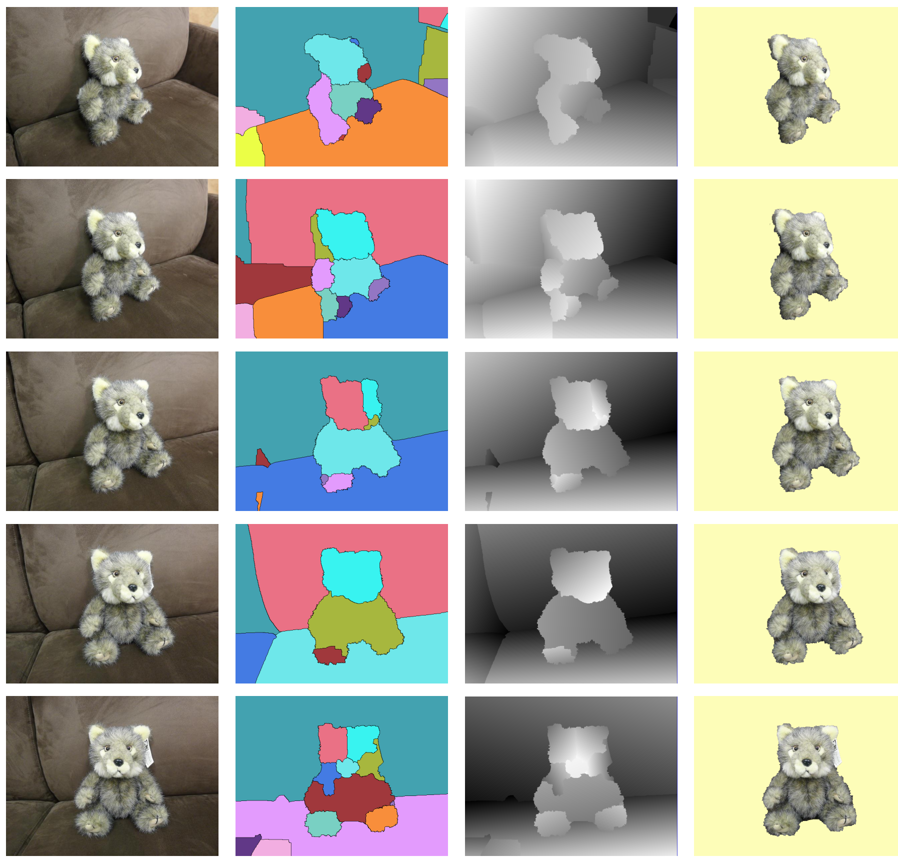

Fig. 2: **TEDDY DATASET** (Five out of 15 images): This example demonstrates the use of multi-view reasoning to infer which surfaces constitutes the object of interest. Even though the surfaces of the sofa are visible in multiple images, they are correctly labeled as background.

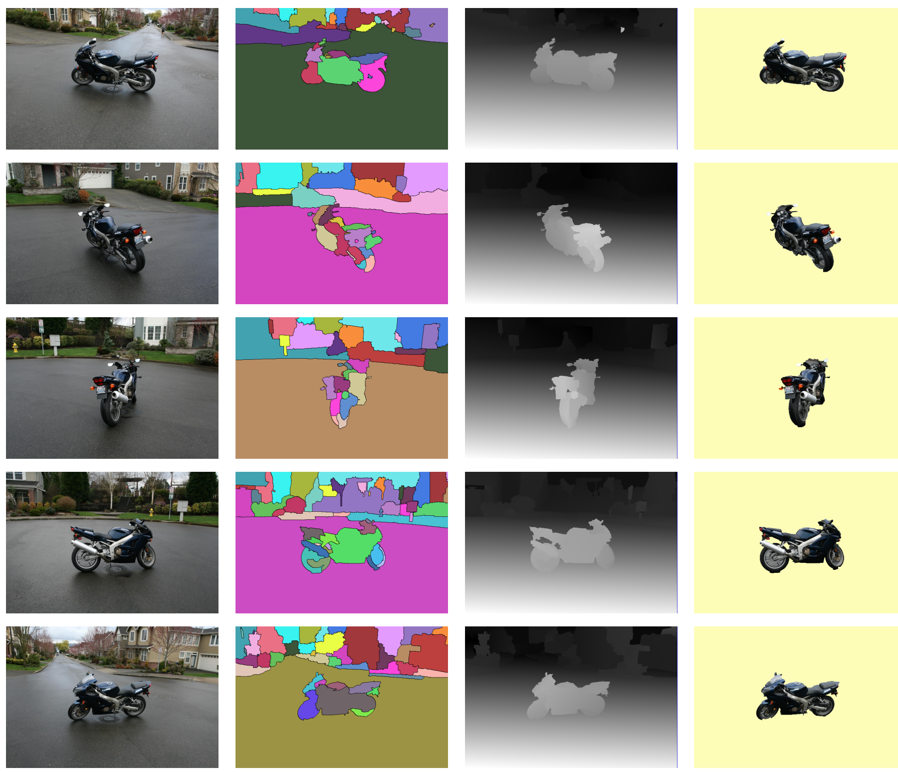

Fig. 3: **BIKE DATASET** (Five out of 34 images): Our method is well suited for situations where the foreground and background color distributions overlap significantly. In this case, the bike seat and tires share the same color as the road in the background.

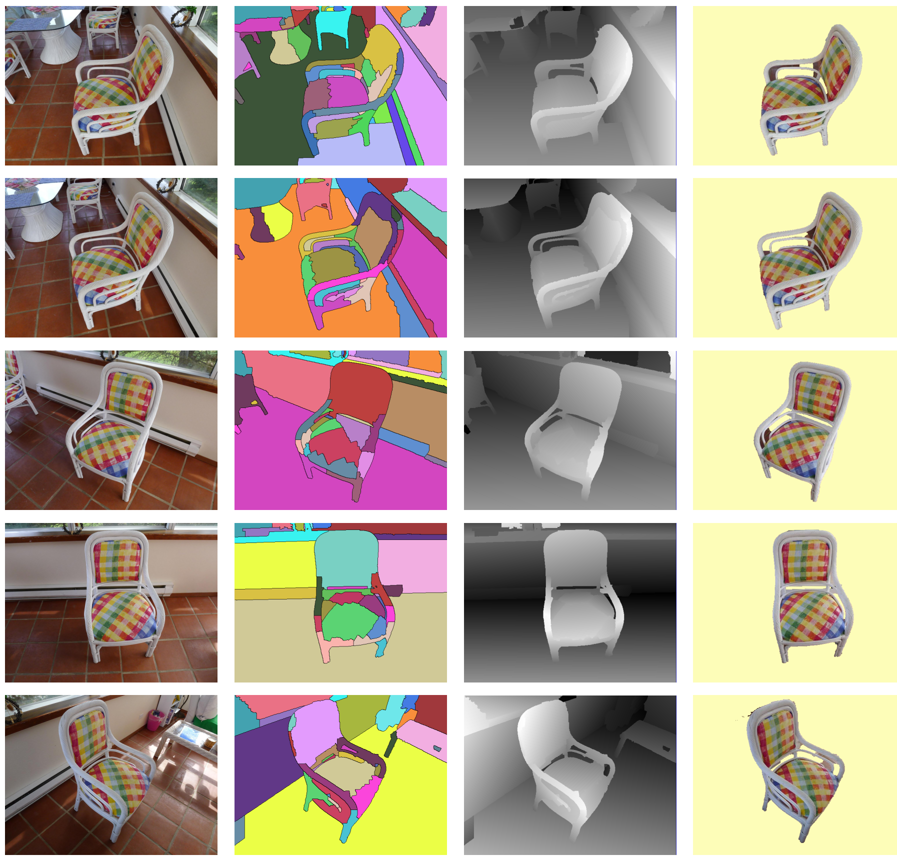

Fig. 4: **CHAIR1 DATASET** (Five out of 17 images): Using a per-surface appearance model is advantageous when the object's appearance has a diverse color distribution as in this case. The object is accurately segmented even though the contrast along its occlusion boundaries is quite low in many cases.

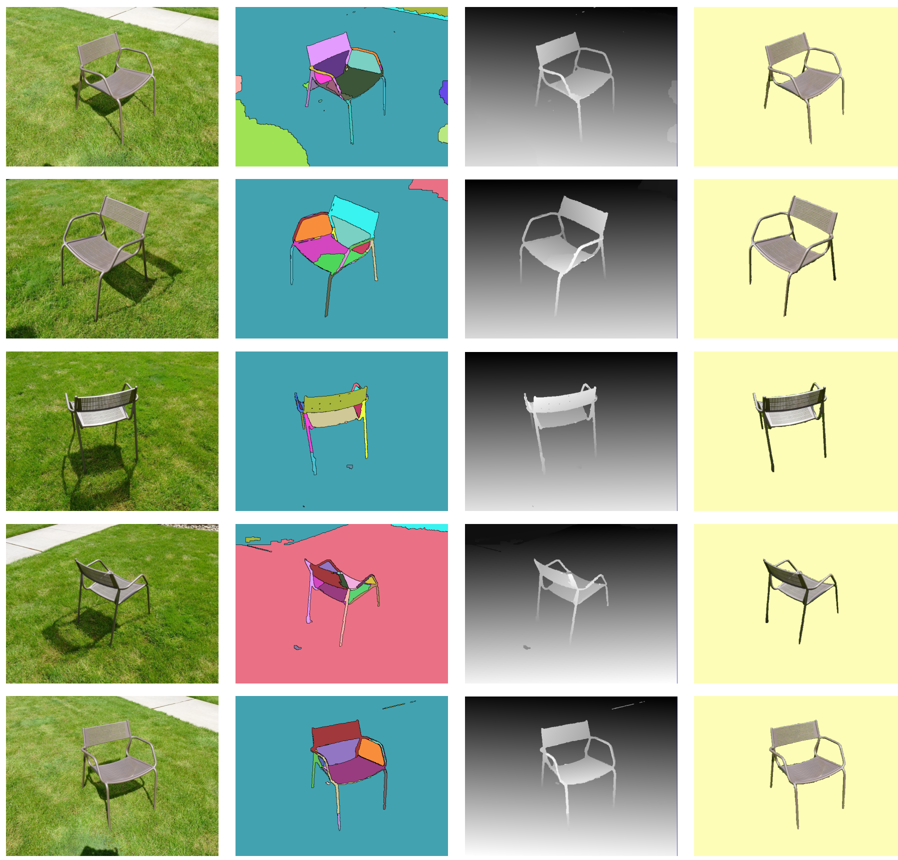

Fig. 5: **CHAIR2 DATASET** (Five out of 45 images): For this dataset, we draw attention to the narrow legs of the chair that our method accurately segments in spite of the shadow, which decreases the contrast between the foreground and background surfaces. (see ROWS 2–4).

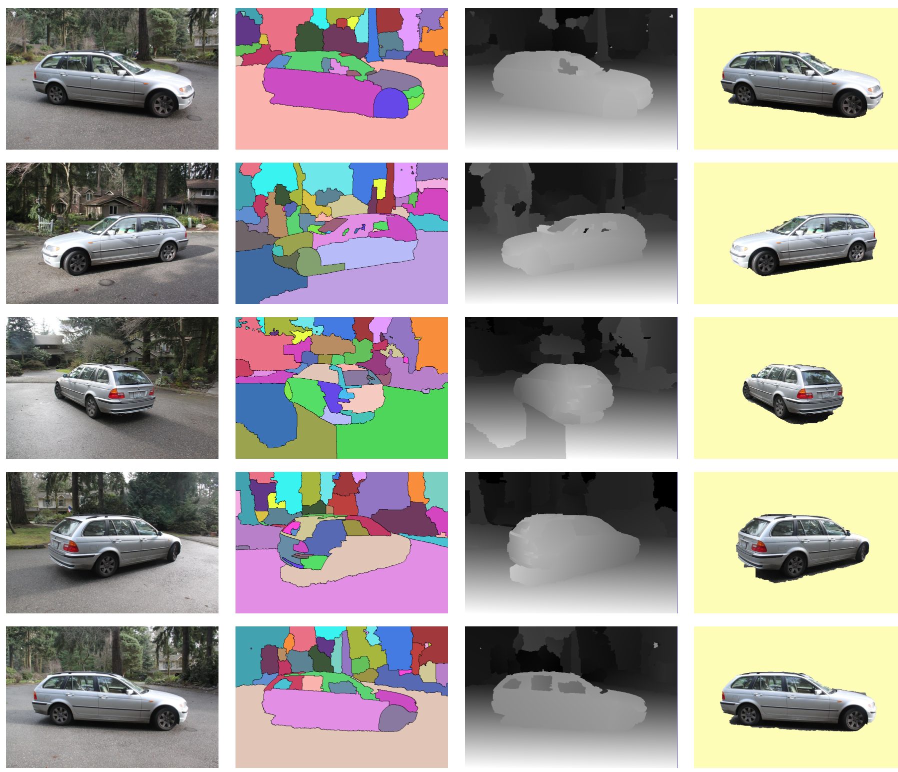

Fig. 6: **Car dataset** (Five out of 45 images): Our method handles difficulties arising from reflections in the glass windows and the specular surfaces of the car and accurately segments the car in many of the input images.

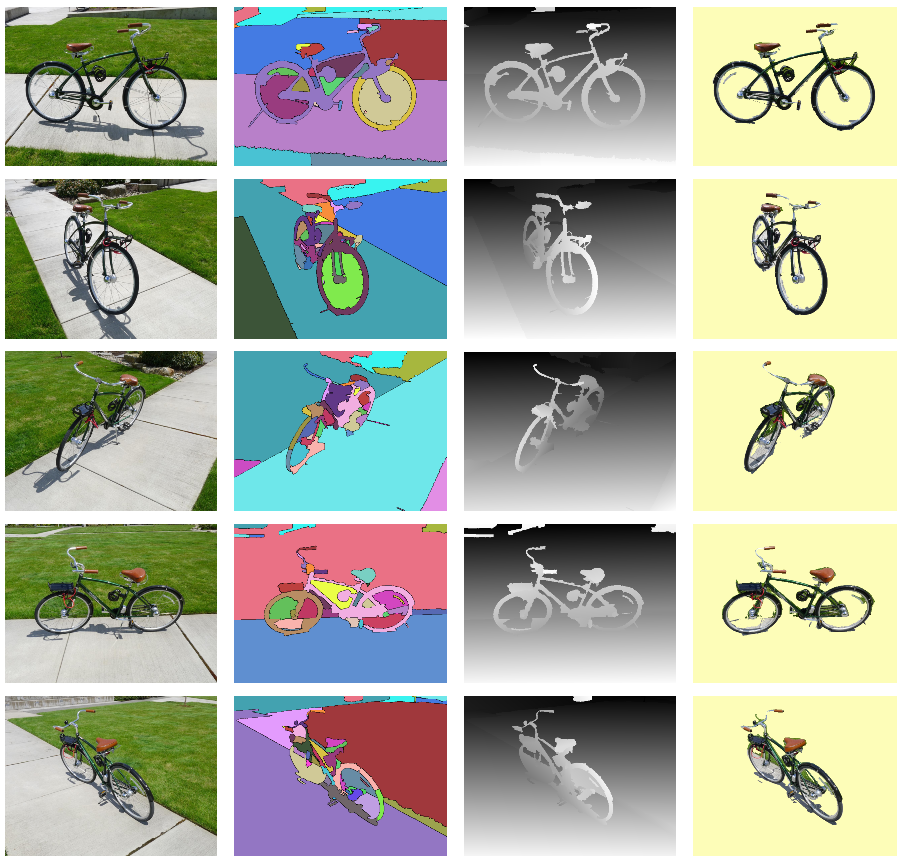

Fig. 7: **BICYCLE DATASET** (Five out of 61 images): The thin structures and holes present in the bicycle as well as some specular surfaces such as the handle bars makes this dataset challenging. Our method generates an accurate segmentation in most cases (COLUMN 4) and also recovers precise occlusion boundaries (COLUMN 3). Considerable manual effort was needed in this case when performing the segmentation using a state of the art interactive method [2].

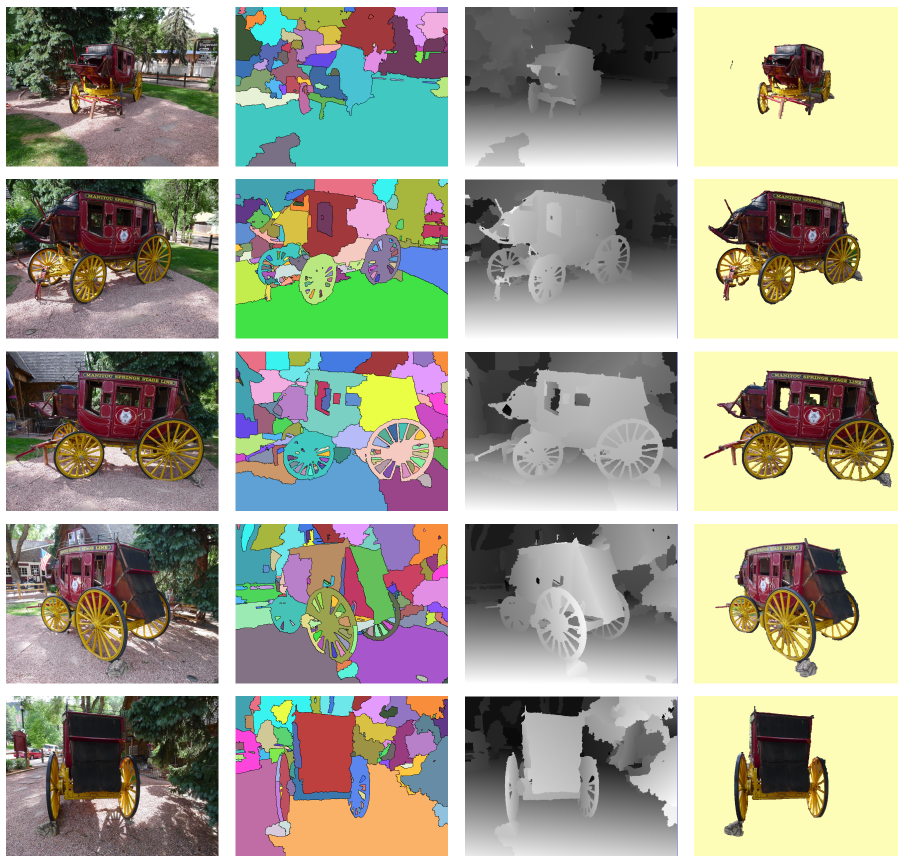

Fig. 8: **CARRIAGE DATASET:** (Five out of 33 images). Our method accurately segments out the carriage in most of the images, while accurately handling thin structures and complex topologies. In particular, the ground seen through the carriage wheels is accurately labeled as background in most cases (ROW 2 - ROW 5).

Fig. 9: COMPARISONS: Five out of 61 images from the BICYCLE sequence are shown. **[Column 1]** shows results from Vicente et. al. [3] in the red overlay. **[Column 2]** shows results from Grabcut [2] with exhaustive user input. **[Column 3]** shows our results. **[Column 4]** shows results generated from a 3D reconstuction (PMVS [1]) with manual segmentation. Our segmentations are accurate on thin structures such as the handle-bars and wheel rims and visually comparable to Grabcut [2] on this example.