# Supplementary Material for Inferring 3D Layout of Building Facades from a Single Image

Jiyan Pan
Google Inc.†
jiyanpan@google.com

Martial Hebert
Carnegie Mellon University
hebert@ri.cmu.edu

Takeo Kanade
Carnegie Mellon University
Takeo.Kanade@cs.cmu.edu

## 1. Sampling a quadrillateral

In Section 4.1 of the main paper, we describe the quadrilateral-based sampling algorithm. Here, we further show in Figure 1 an example of how a quadrillateral is sampled from vanishing lines. First, from all the detected horizontal vanishing directions, we randomly select one of them; and among all the vanishing lines belonging to the selected horizontal vanishing direction, we randomly select a pair of them, such as the line segments in magenta in Figure 1. Also, from all the vanishing lines belonging to the unique vertical vanishing direction, we randomly select a pair of them, such as the line segments in red in Figure 1. (Note that vanishing lines located within the region occupied by already-selected quads are not sampled.) The four line segments delineate a quadrilateral as is shown by the yellow shape in Figure 1. The orientation $\mathbf{n_{si}}$ of this quad is equal to the cross-product of the horizontal vanishing direction represented by the magenta line segments, and the vertical vanishing direction represented by the red line segments. Figure 1 also shows the hybrid score of the quadrilateral.


Figure 1. Example of a randomly sampled quadrilateral.

## 2. Surface layout estimation

In Section 5.2 of the main paper, we evaluate the performance of our algorithm on estimating surface layout labels on the Geometric Context dataset [2]. Here, we include a few more success cases in Figure 2, as well as some failure cases in Figure 3.

In rows 1 and 2 of Figure 2, our algorithm recovers facade planes that the other methods fail to identify. In row 3 of Figure 2, the benefit of using plane-based reasoning by our algorithm is evident: it neither over-segments the facades as the segment-based approach usually does (column 2), nor ignores the detailed layout of the facades as the block-based approach tend to do (column 3). We can also see that the surface layout estimation directly obtained from orientation maps is usually quite noisy (e.g., row 1, column 6 in Figure 2). Our quadrilateral-based sampling algorithm helps inhibit such noise and leads to a better result (e.g., row 1, column 5 in Figure 2). While inter-planar geometric constraints in our approach effectively regularize plane depths (as we have seen in Section 5.1 of the main paper), it has a smaller effect on correcting mistakes in plane orientations, except for some cases such as the one shown in row 4 of Figure 13 in the main paper. Therefore, we only see a modest improvement on surface layout accuracy after the CRF is applied, as is shown in Figure 12 of the main paper.

Several of our failure cases are shown in Figure 3. The case in the first row fails because our approach discovers small non-Manhattan facade structures not included in the ground truth. It also failed to identify the garage in the distance. In the second row, our method again detects a fine structure not included in the ground truth, while mislabeling a left-facing region due to the deceiving vanishing lines from the rooftop. In the third row, our method also mis-labeled the left-facing part of a foreground structure due to strong noise in vanishing lines.

To get a better insight into the comparison between our

†Jiyan Pan was at the Robotics Institute, Carnegie Mellon University when the work was performed.

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

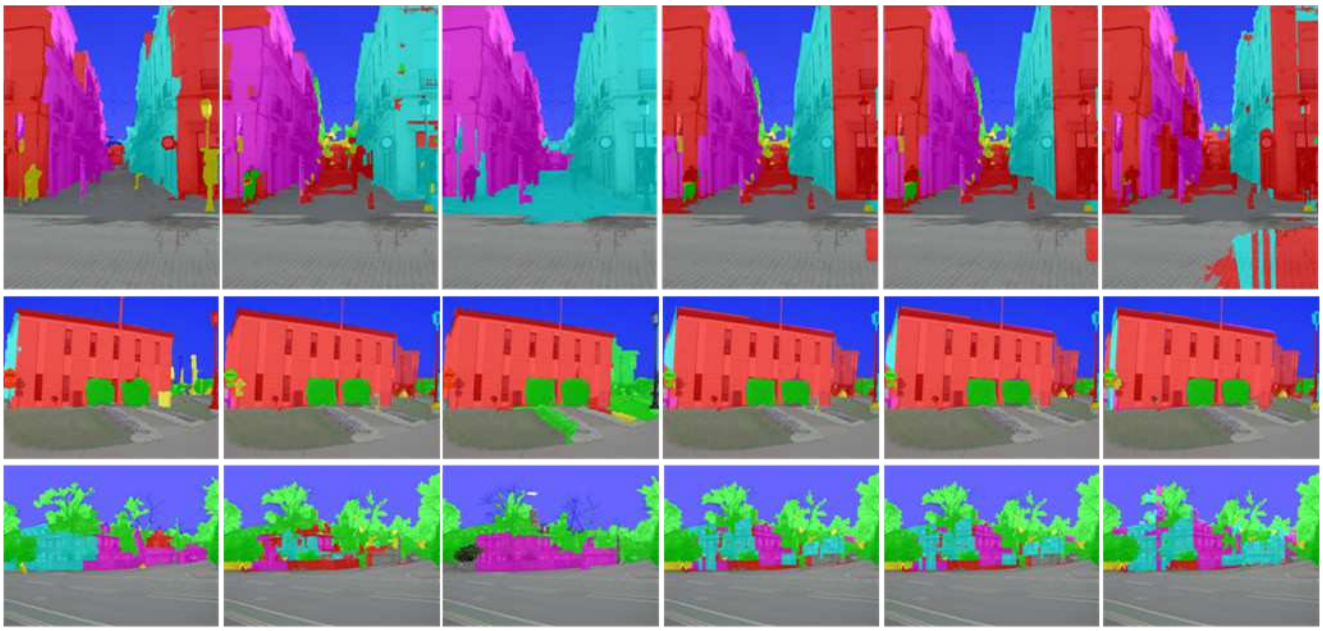

Figure 2. Additional success cases of surface layout estimation. From left to right: Ground truth; Hoiem *et al.* [3]; Gupta *et al.* [1]; Ours; Ours w/o CRF; Orientation map from the line-sweeping algorithm [4]. Surface layout color code: Magenta – planar right; Cyan – planar left; red – planar center; green – non-planar porous; yellow – non-planar solid; blue – sky; grey – support.

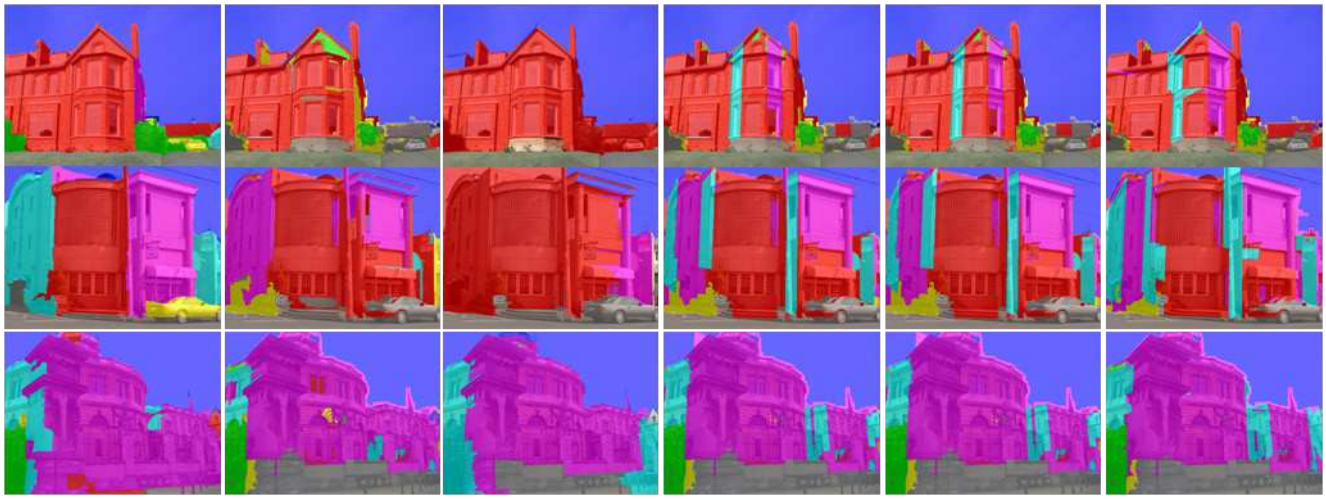

Figure 3. Failure cases of surface layout estimation. From left to right: Ground truth; Hoiem *et al.* [3]; Gupta *et al.* [1]; Ours; Ours w/o CRF; Orientation map from the line-sweeping algorithm [4]. Surface layout color code: Magenta – planar right; Cyan – planar left; red – planar center; green – non-planar porous; yellow – non-planar solid; blue – sky; grey – support.

algorithm and the state-of-the-art approaches, we compute, for each image, the accuracy gain of our algorithm over the segment-based approach [3] and the block-based approach [1], respectively, and plot the histogram of accuracy gain over the test images.

The histograms for comparisons with the two state-of-the-art approaches are shown in Figures 4 and 5, respectively. In each figure, typical examples are shown for positive, neutral, and negative accuracy gains, respectively. In both figures, we can see that our algorithm achieves good surface classification accuracy where vanishing lines are correctly detected and grouped. However, in image regions where missing or distracting vanishing lines dominate, the accuracy of our algorithm deteriorates, because vanishing lines play an important role in proposing and evaluating quadrilateral samples. A poor semantic segmentation result

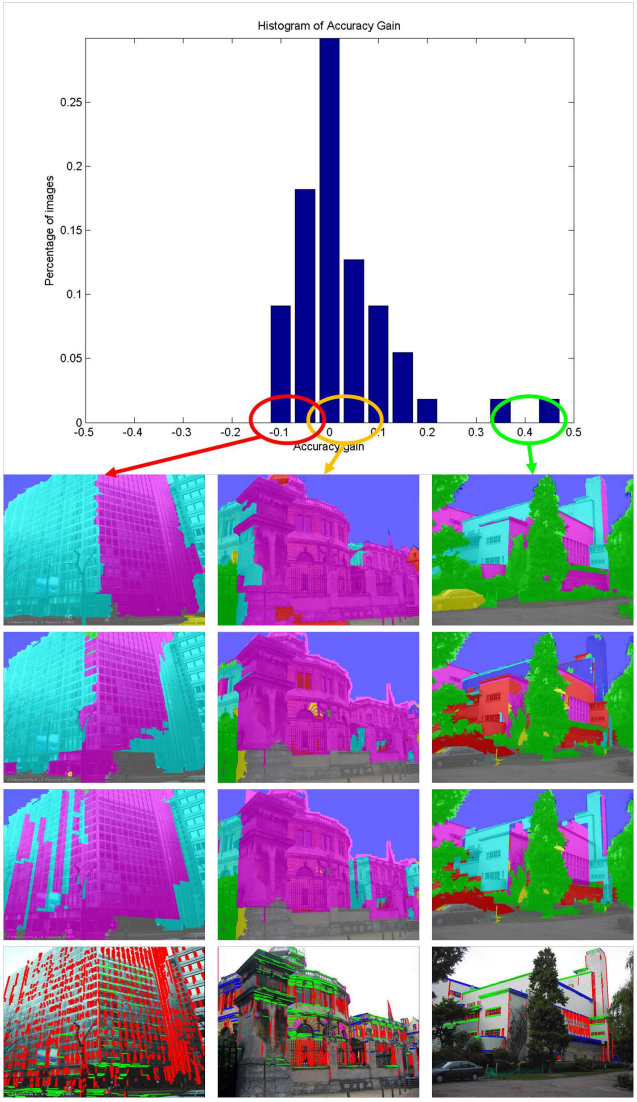


Figure 4. Histogram of accuracy gain of our algorithm over the segment-based approach in [3]. Typical examples corresponding to three different ranges of accuracy gain are also displayed, where the first row is the ground truth, the second row is the result of [3], the third row is our result, and the fourth row shows the vanishing lines. The color code is the same as in Figure 2.

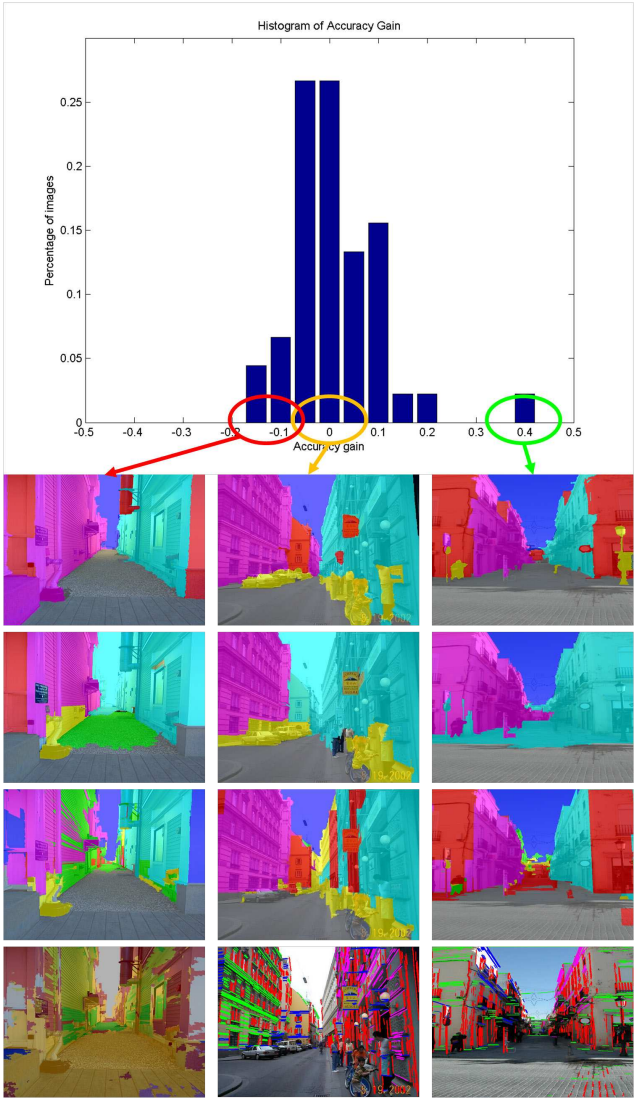


Figure 5. Histogram of accuracy gain of our algorithm over the block-based approach in [1]. Typical examples corresponding to three different ranges of accuracy gain are also displayed, where the first row is the ground truth, the second row is the result of [1], the third row is our result, and the fourth row shows the vanishing lines – except for the leftmost image which shows semantic segmentation, where the "building"region is indicated by the red shade. The color code of surface layout is the same as in Figure 2.

would also significantly undermine the surface classification accuracy of our algorithm, as is shown in the leftmost example in Figure 5.

While our approach does not perform significantly better than the state-of-the-art algorithms in terms of coarse surface layout classification, our approach is able to generate a quantitative, continuous 3D layout of the facade scene, which is beyond the capability of competing methods. Several examples are shown in Figure 6.

## 3. Depth map estimation

In Section 5.3 of the main paper, we present the quantitative results of depth map estimation on the Make3D dataset [6, 5]. Here, we further show some qualitative results of depth map estimation in Figure 7. We could see that our method makes better depth estimation in facade regions than Liu's super-pixel based approach [5]. Also, as our method does not require training using ground-truth

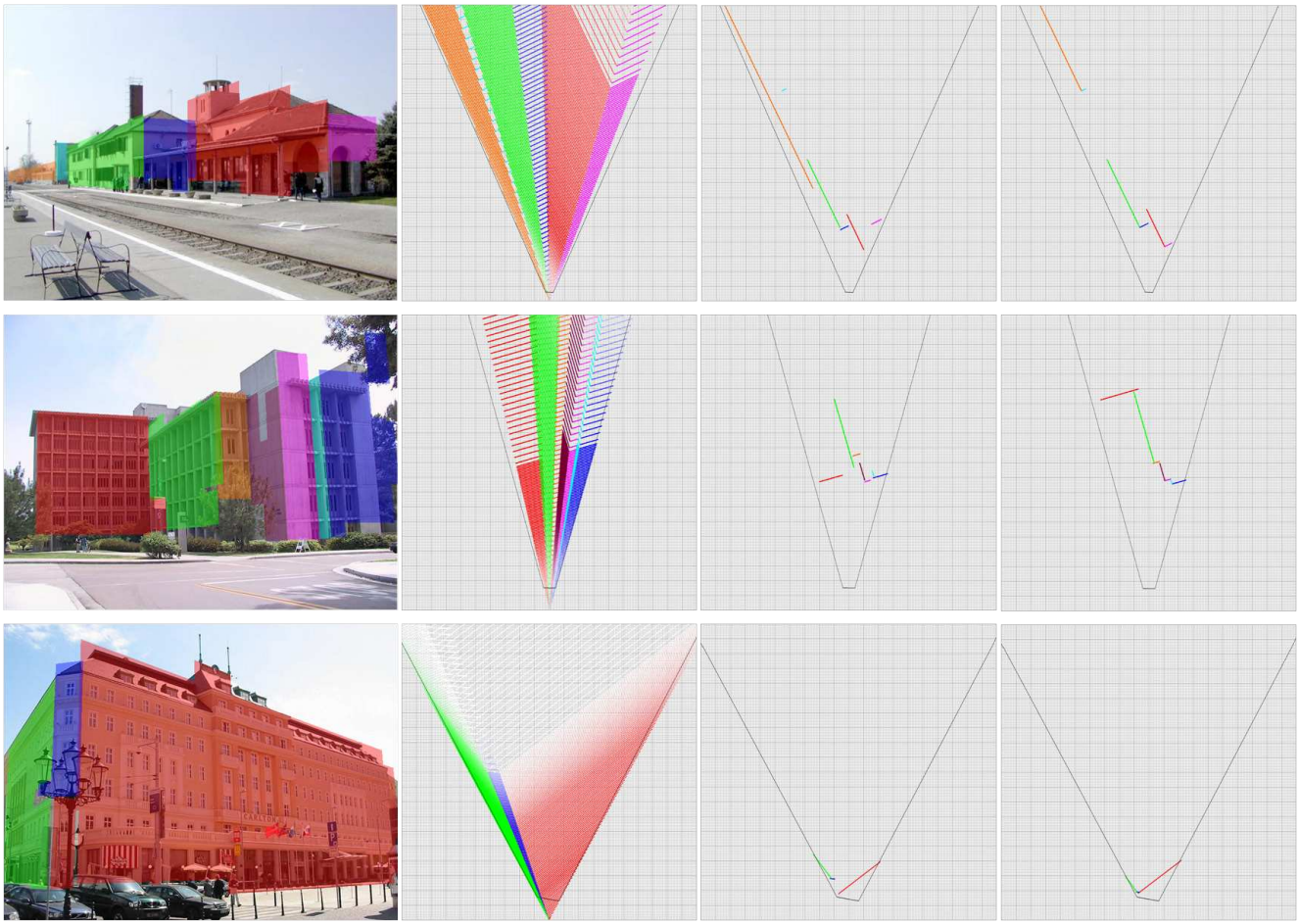

Figure 6. 3D facade layout estimation by our method on images from the Geometric Context dataset [2]. Left to right: 1) Original image overlaid with color shades representing quads from distinctive facade planes. Quads from the same distinctive facade plane share the same color. 2) Depth distribution of candidate distinctive facade planes before running the CRF inference. 3) Best locations of candidate distinctive facade planes before running the CRF inference. 4) Optimal locations of valid distinctive facade planes after running the CRF inference. The viewing boundary is marked with black lines, and the coarser grid spacing is 10m.

depth maps, it does not try to fit blindly to the defects of the ground truth such as the "infinity holes"at windows in row 2 of Figure 7 (although it might negatively affect the final quantitative number). Depth estimation of ground region by our method is also better due to the 3D stage we have established using the facades. The benefit of imposing inter-planar geometric constraints could be seen in rows 5 and 6 of Figure 7. In row 5, the depth of the second-floor camera-facing facade is correctly estimated after CRF inference, while in row 6, two misaligned left-facing facades in column 5 are correctly aligned after CRF inference in column 4.

In row 6, however, our method fails to identify the structure closest to the camera as it did not regard it as a facade region. Also, our method is not designed to estimate the depth of trees or foreground objects, and the make-shift algorithm described in the main paper only provides a very rough estimation. Liu's algorithm [5], nevertheless, is effective at estimating the depth of those regions. Please see row 4 of Figure 7 for an example.

Note that, unlike Liu's algorithm [5], our approach does not require being trained on the ground-truth depth maps of any specific dataset, and therefore would generalize better. Even on the Make3D dataset [6, 5] that Liu's algorithm [5] is specifically trained upon, our approach still achieves comparable results. Moreover, the output of our approach is beyond a depth map – it is a higher-level interpretation of building facades with mutually constrained planes, as is shown in Figure 8. The reason why inter-planar interactions do not play a significant role here is that facade planes in this dataset are relatively simple and have clearly visible ground contact lines in the image. As a result, cues from an individual facade plane are already sufficiently informative in locating it in 3D (please see the peaky

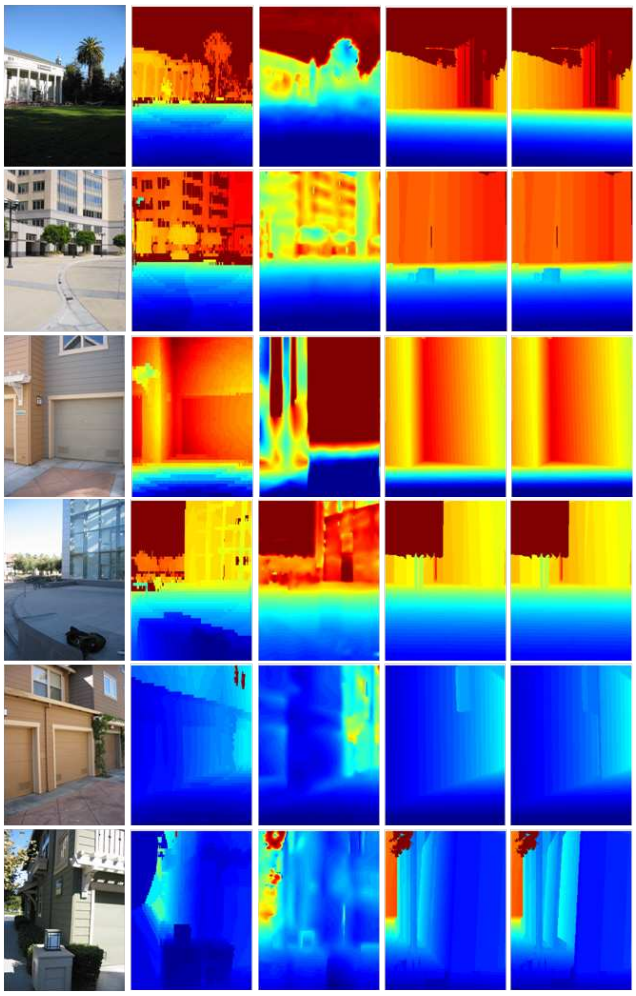

Figure 7. Qualitative comparisons of depth map estimation. Left to right: Original image; Ground truth; Liu [5]; Ours; Ours w/o CRF. Depth value increases from blue to red. The color code is scaled according to the ground-truth depth map.

distributions in the second column of Figure 8). However, when ambiguity is high (*e.g.*, the green facade plane in the first row of Figure 8), inter-planar interactions are critical in forcing facade planes into more geometrically plausible locations.

## References


[1] A. Gupta, A. A. Efros, and M. Hebert. Blocks world revisited: image understanding using qualitative geometry and mechanics. *ECCV*, 2010.

[2] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *IJCV*, 2007.

[3] D. Hoiem, A. A. Efros, and M. Hebert. Closing the loop on scene interpretation. *CVPR*, 2008.

[4] D. C. Lee, M. Hebert, and T. Kanade. Geometric reasoning for single image structure recovery. *CVPR*, 2009.

[5] B. Liu, S. Gould, and D. Koller. Single image depth estimation from predicted semantic labels. *CVPR*, 2010.

[6] A. Saxena, M. Sun, and A. Y. Ng. Make3d: learning 3-d scene structure from a single still image. *PAMI*, 2009.

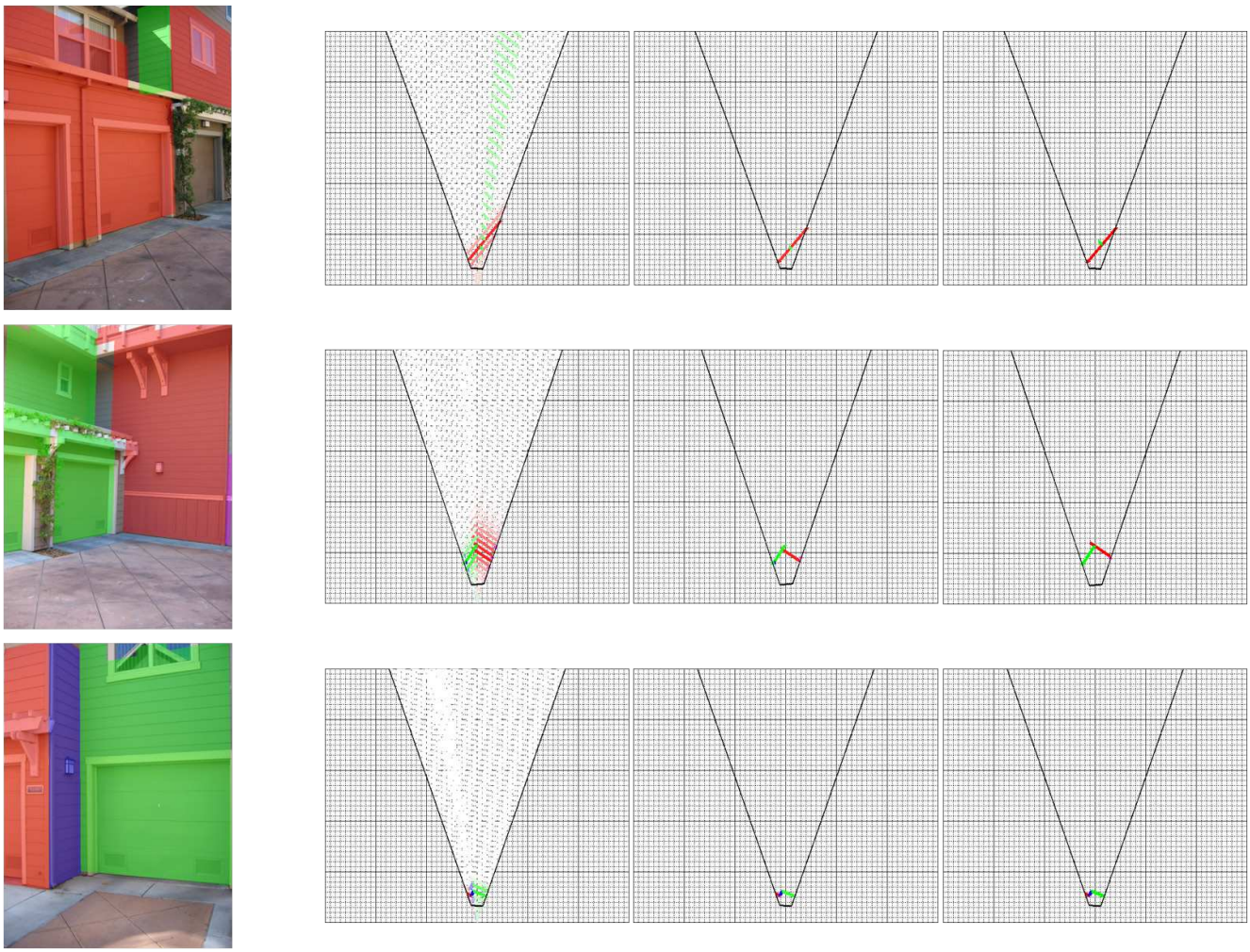

Figure 8. 3D facade layout estimation by our method on images from the Make3D dataset [6, 5]. Left to right: 1) Original image overlaid with color shades representing quads from distinctive facade planes. Quads from the same distinctive facade plane share the same color. 2) Depth distribution of candidate distinctive facade planes before running the CRF inference. 3) Best locations of candidate distinctive facade planes before running the CRF inference. 4) Optimal locations of valid distinctive facade planes after running the CRF inference. The viewing boundary is marked with black lines, and the coarser grid spacing is 10m.