Supplementary material: Local detection of stereo occlusion boundaries

Jialiang Wang and Todd Zickler
Harvard University
jialiangwang@g.harvard.edu, zickler@seas.harvard.edu

1. Additional Details Regarding Middlebury Experiments

We test our detector on five scenes from the Middlebury 2014 training set [6] using stereo occlusion boundaries that we manually label (in (x, y)) as ground truth. These five scenes are held out when training the Siamese matching network that provides input cost volumes C to our detector. Figure 1 shows the results using Middlebury "half resolution" images.

Figure 1. First column: Manually-labelled ground truth stereo occlusion boundaries in (x, y). Second column: Score $max_d B(x, y, d)$ with B being our detector's output before non-maximum suppression. Third column: Detection results in (x, y, d), using $B > 0.5$, superimposed on the stereo left image. For visualization, ground truth labels are dilated to five-pixel wide.

2. Additional Results: Sintel

Figure 2 shows additional results on the Sintel [1] "clean pass" dataset, along with the labelled ground truth stereo occlusion boundaries in (x, y, d). We require a pixel to be mutually-visible (*i.e.* in a binocular region) and to occlude at least two adjacent pixels in the cyclopean view in order for it to be considered as a stereo occlusion boundary.

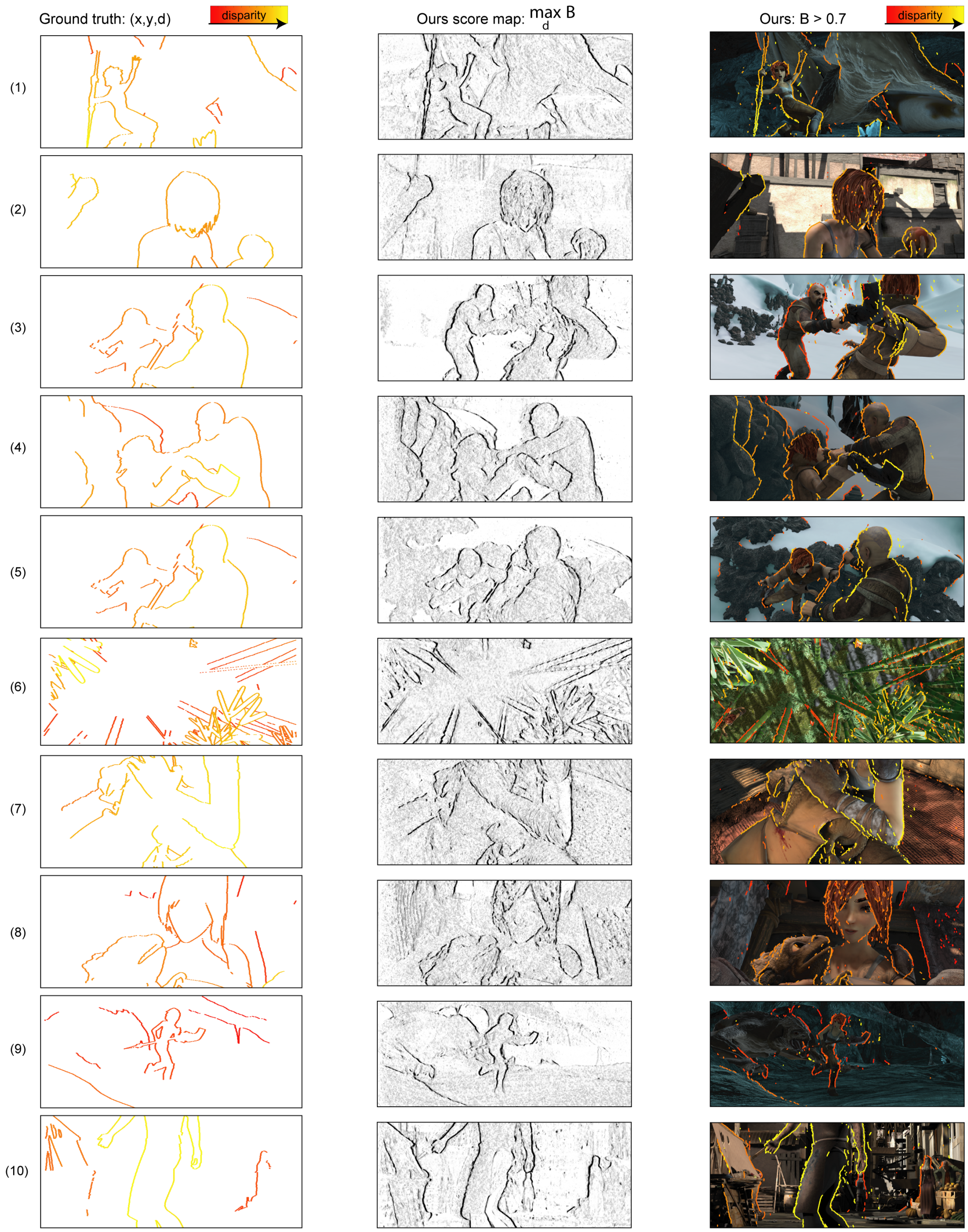

Figure 2. First column: Labelled ground truth stereo occlusion boundaries in (x, y, d), with d encoded using color. Second column: $max_d B(x, y, d)$ where B is our detector's output before non-maximum suppression. Third column: three-dimensional detection results (x, y, d) (using $B > 0.7$) superimposed on the left image. For visualization, ground truth labels are dilated to five-pixel wide.

3. Additional Results: Perceptual Stimuli

There are twelve stimuli in the Perceptual Stimuli Dataset [7]. Here we show the remaining eight stimuli that are not included in Section 5 of our paper, along with comparisons to three other stereo methods [5, 3, 2]. All methods work well for stimuli with sufficient matching cues (*e.g.* (1), (7) and (8)). Some other methods break down when ambiguities arise (*e.g.* (2) and (3)). All other methods fail when there is no matching information. Our method detects all stereo occlusion boundaries, as well as some E, F-type points, as discussed in Section 3 of the paper. Since our detector is a local one, we have some noise in some examples, especially (4), where the repeated textures are ambiguous and nearly E or F-type points in our taxonomy.

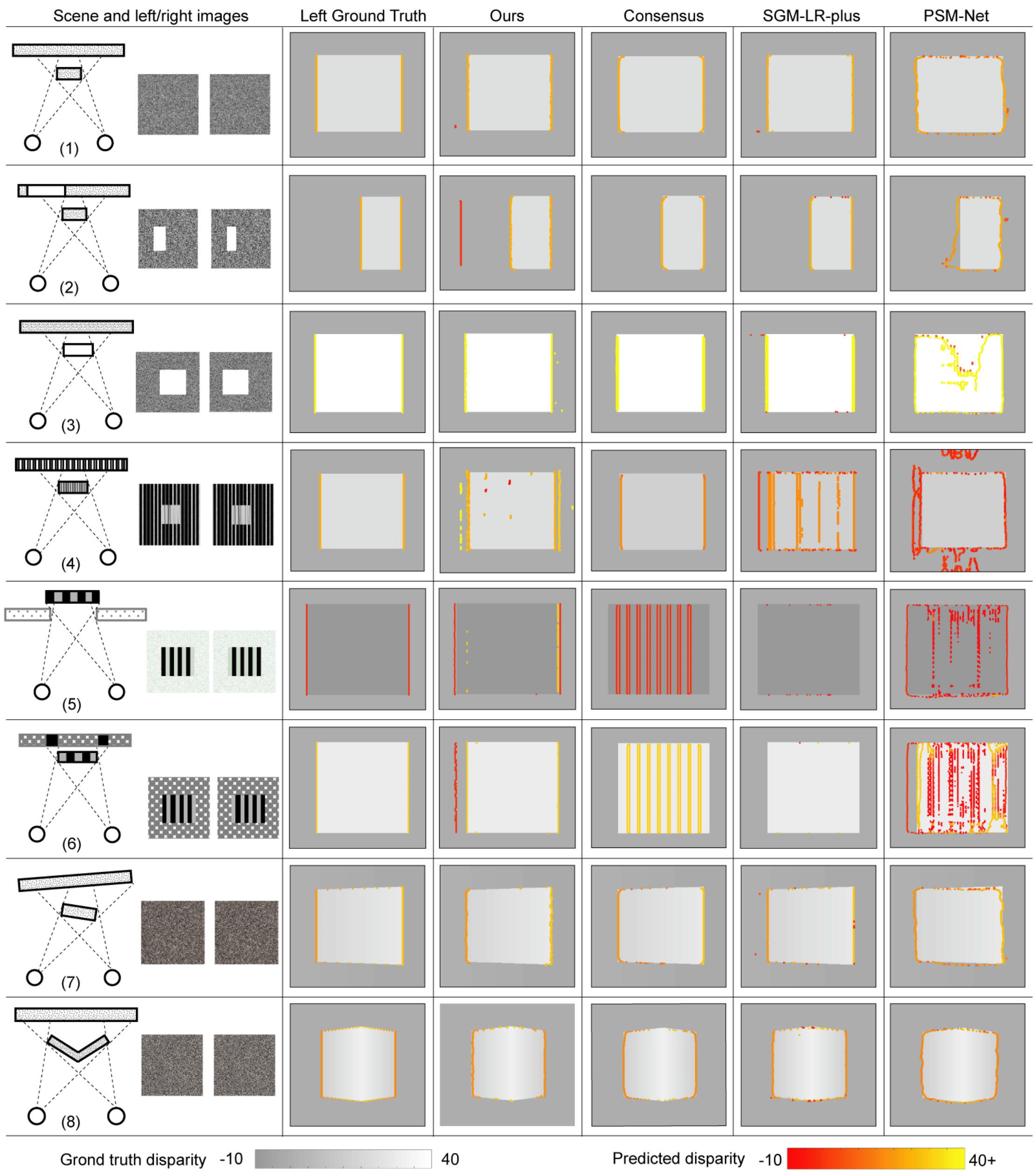

Figure 3. Perceptual Stimuli: comparison of our method's stereo occlusion boundaries (using $B > 0.7$) and those generated by four other stereo algorithms. Some results are cropped to show the region of interest.

4. Examples of Synthetic Training Images

Figure 4 shows some examples of the synthetic training images we rendered to train our stereo occlusion boundary detector. Notice that we randomly sample textures for each scene. In general, we found it helpful to include as many textures as possible. Our detector is robust and works well on all three datasets in test time despite only trained on these simple and abstract images.

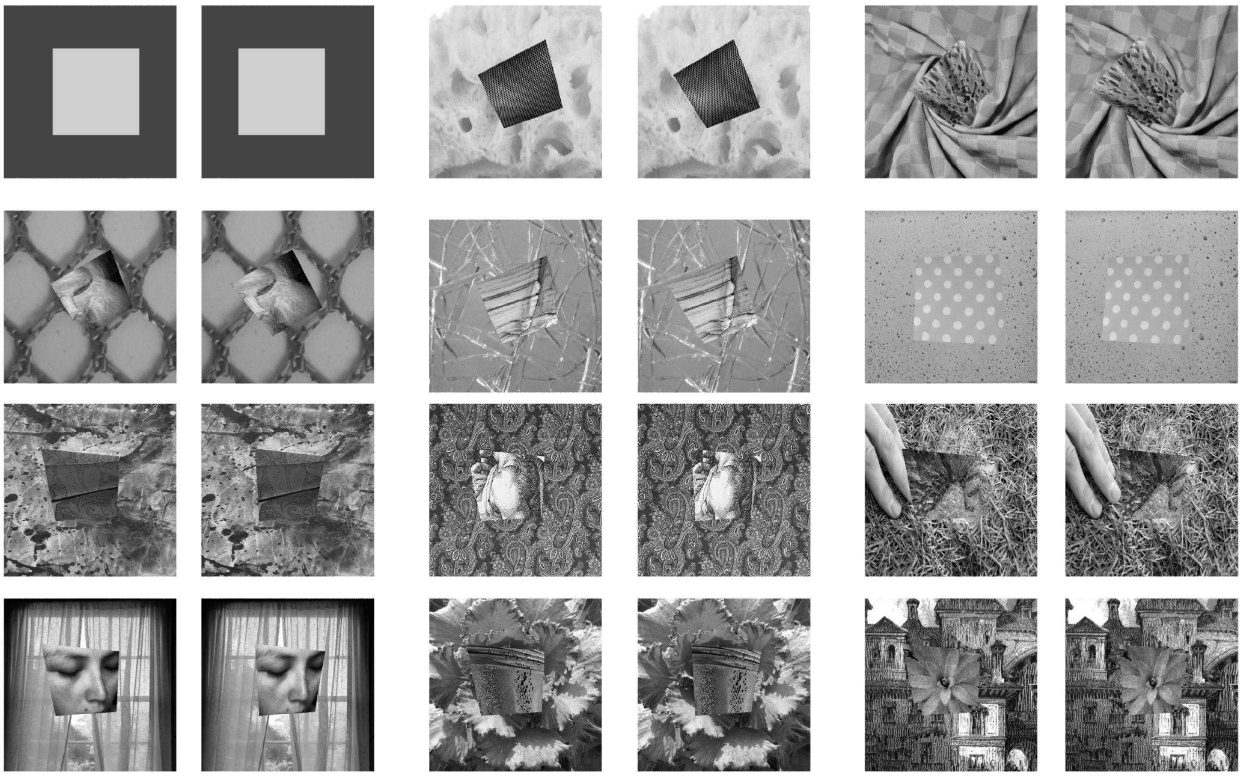

Figure 4. Selected synthetic images we use to train our feedforward stereo occlusion boundary detector. These images are rendered using a grayscale version of the Describable Textures Dataset [4] plus 35 uniform-intensity "textures" that have different intensities. The images are 600×600 resolution.

References

[1] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, and Michael J. Black. A naturalistic open source movie for optical flow evaluation. *European Conference on Computer Vision*, 2012. 2

[2] Ayan Chakrabarti, Ying Xiong, Steven J Gortler, and Todd Zickler. Low-level vision by consensus in a spatial hierarchy of regions. *Computer Vision and Pattern Recognition*, 2015. 3

[3] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. *Computer Vision and Pattern Recognition*, 2018. 3

[4] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Computer Vision and Pattern Recognition*, 2014. 4

[5] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 2008. 3

[6] Daniel Scharstein, Heiko Hirschmüller, York Kitajima, Greg Krathwohl, Nera Nešić, Xi Wang, and Porter Westling. High-resolution stereo datasets with subpixel-accurate ground truth. *German Conference on Pattern Recognition*, 2014. 1

[7] Jialiang Wang, Daniel Glasner, and Todd Zickler. Toward perceptually-consistent stereo: A scanline study. *International Conference on Computer Vision*, 2017. 3