Supplementary Materials for Object Segmentation using Sptial and Spatio-Temporal Cues

March 8, 2011

Before presenting more qualitative results of our method, we need to clarify that the quantitative results for object segmentation mentioned in the paper reflect the balanced accuracy defined by

$$Acc_B = \frac{1}{2}\left(\frac{tp}{tp+fn} + \frac{tn}{tn+fp}\right) \tag{1}$$

where tp, tn, fp, fn represent the true and false positive and negative detections. Such a measure will compensate for the imbalanced size of the positive and negative segments in images and thus, if the segments are not equally sized, which is the case for most of the sequences, will have smaller values compared to the accuracy defined by

$$Acc = \frac{tp+tn}{tp+tn+fp+fn} \tag{2}$$

To emphasize the difference, we also represent the accuracy (2) of the segmentations in Table 1. It is evident from the table that the Acc measures have higher values compared to Acc_B as there are many sequences that contain much fewer positive pixels compared to negative pixels or vice versa. It is interesting to note that using either measure, all the arguments about the quantitative evaluation of the method remain the same as was mentioned in Section 3.4 in the paper e.g. the use of object boundary detector robustly increases the accuracy, use of motion consistently improves the accuracy, the optimal parameters for each configuration remain exactly as before. The only differences is that using two frames, in comparison with one frame, leads to 3 percent increment of the accuracy reflected by Acc measure while the improvement is 5.5 percent in case of Acc_B. Similarly, using the Acc measure, we are able to classify 94.9 percent of the pixels in our dataset using two frames and with the extra feedback to the method(in terms of specifying the parameters) this number goes up to 97.1 percent. We will clarify this in the paper.

Figure 1 depicts the apparent motion estimated between two frames of 16 sequences in our dataset. It can be observed that while the motion feature

Feature	Detector	Parameters	Mean Acc
Color	Not used	$\lambda = 2$, $h = 1$	0.9198
Color	Used	$\lambda = 5$, $h = 0.5$	0.9317
Color+Motion	Not used	$\lambda = 2$, $h = 1$	0.9387
Color+Motion	Used	$\lambda = 5$, $h = 0.5$	**0.9490**
Color	Not used	Tuned	0.9436
Color	Used	Tuned	0.9551
Color+Motion	Not used	Tuned	0.9578
Color+Motion	Used	Tuned	**0.9715**

Table 1: The mean accuracy of the segmentations(the *Acc* measure) using color and motion features. Tuned parameters means that the best performing parameters from a set of parameters(see Section 3.4 in the paper) were selected individually for each sequence.

clearly holds information about the geometry of the 3D world, it is noisy and in some cases erroneous and thus, requires further processing. Figure 2 depicts more qualitative results of our object boundary detector. Extra qualitative results for interactive segmentation methods and our method are given in Figures 3 and 4. Please refer to section 3.4 in the paper for more information regarding these two figures.

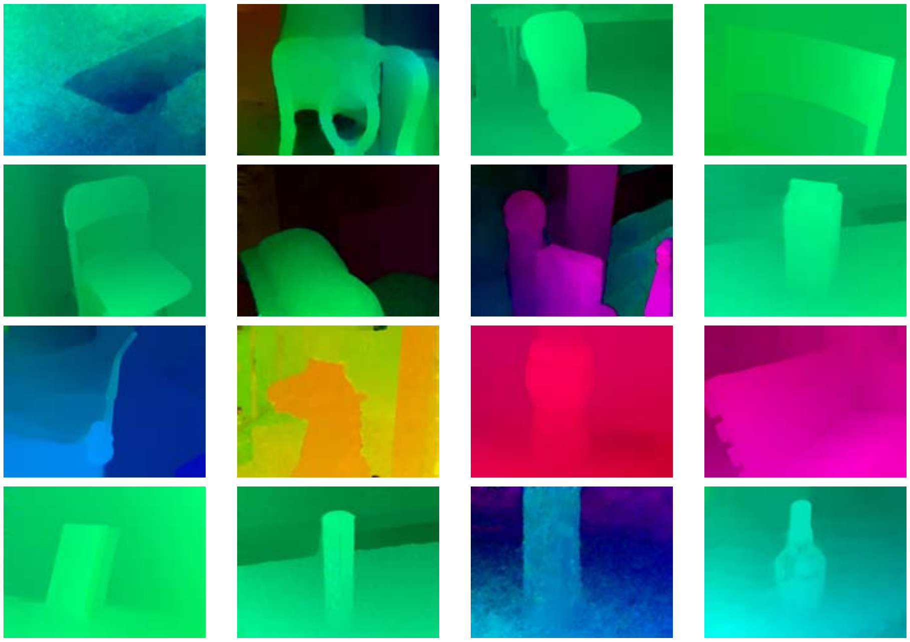

Figure 1: The apparent motion(forward flow) estimated between the two frames of 16 sequences in our dataset. The sequences are bench, chair1, Chair1, Chair2, Chair3, couch_corner, fencepost, Juice, Pipe1, rocking_horse, Salt1, Sofa1, Speaker1, Spray, tree, Whisky1. Notice the over-regularization in weakly textured areas(e.g. the Pipe1 and chair1).

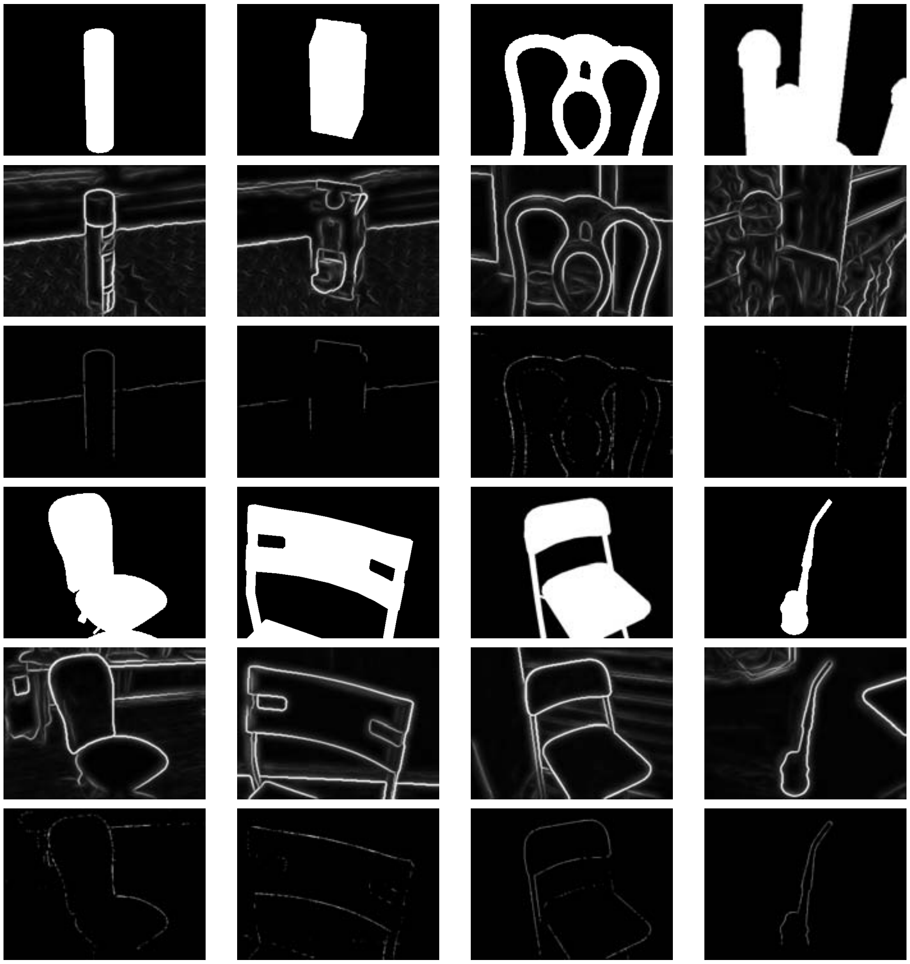

Figure 2: Qualitative results of our object boundary detector. The figure depicts the ground truth segmentation, the gPb detector(thick version) and the detection result of our object boundary detector.

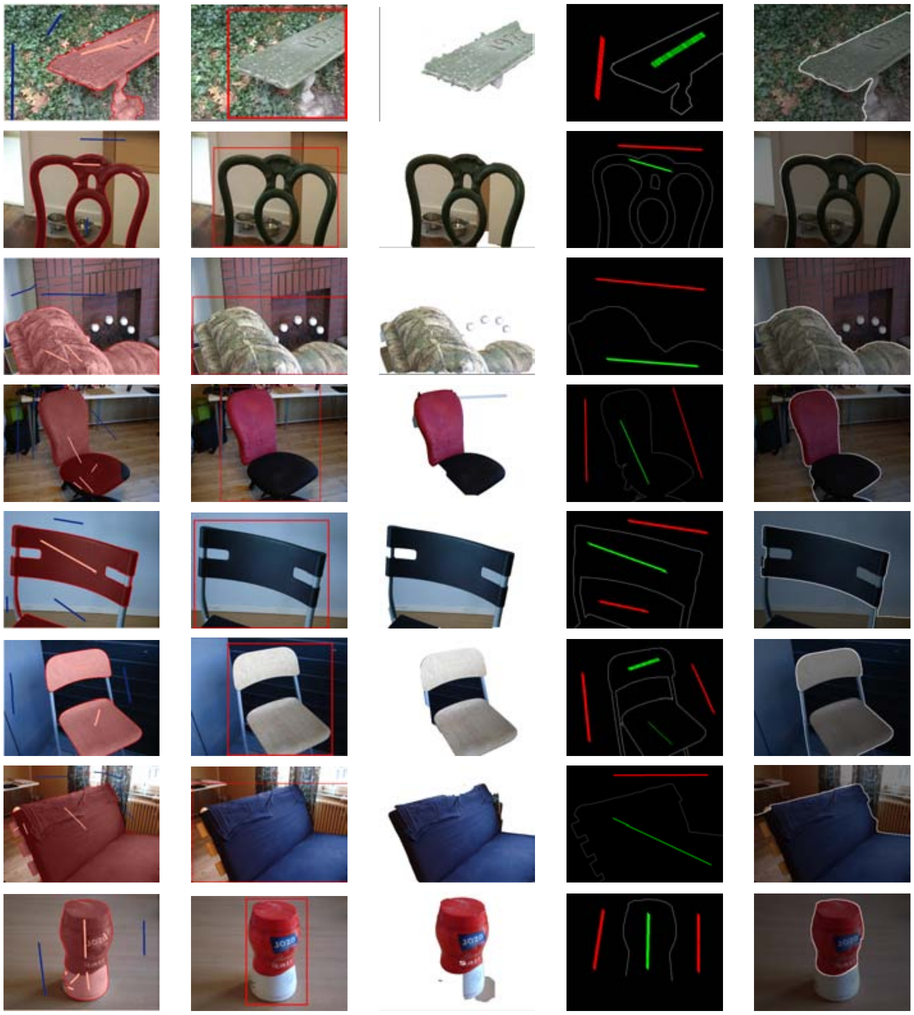

Figure 3: Qualitative results of three interactive methods on eight sequences. See the description of Figure 6 in the paper.

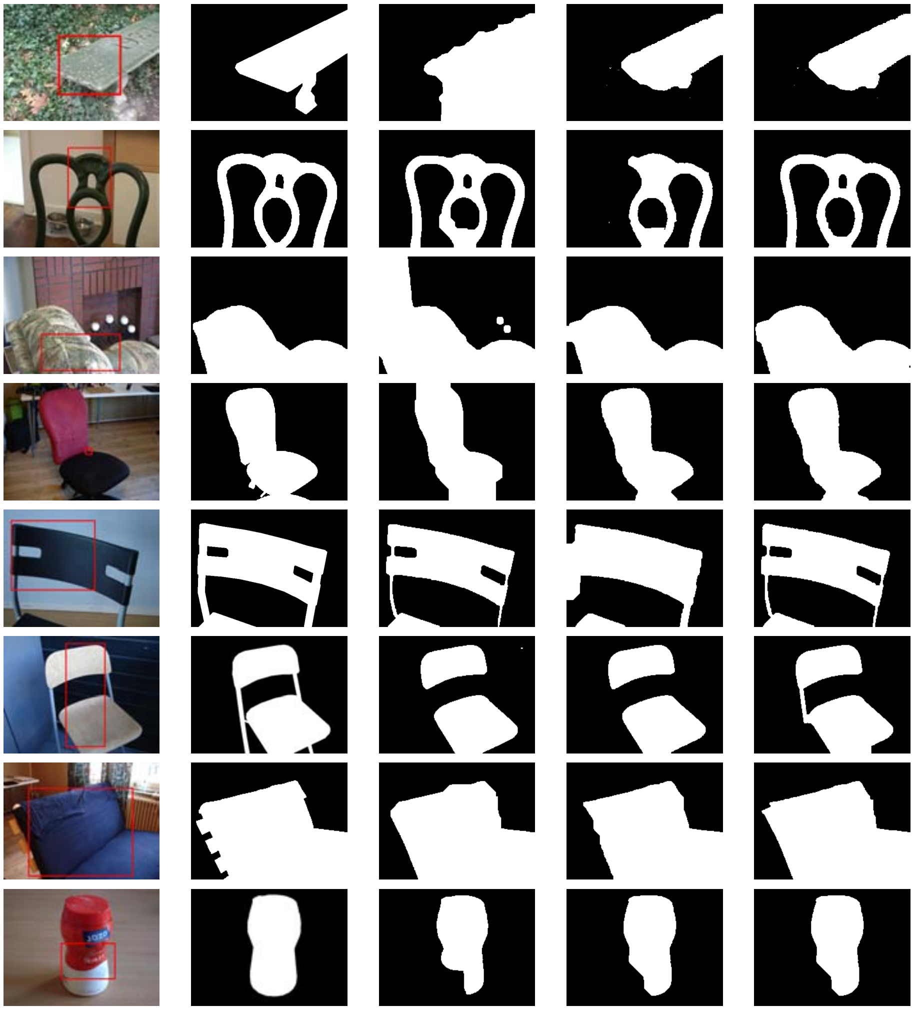

Figure 4: Qualitative results of our method on eight sequences. From left to right: initialization, ground truth segmentation, segmentation using color feature(tuned), segmentation using color and motion features(default parameters: $\lambda = 5, h = 0.5$), segmentation using color and motion features(tuned). See Figure 7 in the paper.