

Supplementary Material

for

Towards Unsupervised Whole-Object Segmentation: Combining Automated Matting with Boundary Detection

Andrew N. Stein* Thomas S. Stepleton Martial Hebert
The Robotics Institute, Carnegie Mellon University
Pittsburgh, Pennsylvania 15213, USA
anstein@cmu.edu

In this supplementary document, we provide additional example results (Figures 1-38) of the same type as the examples in the main paper, but over a larger range of desired consistency. On individual examples, either BSE or the Multiscale NCuts approach sometimes outperforms our method, but neither does so consistently. Our matting-based approach offers solid performance, if not marked improvement, more consistently. We also show examples of “easy” objects, for which all methods work well, and “hard” (often small) objects, which prove challenging for all methods.

At the top of each figure is the bar graph displaying the number of segments required by each method to achieve the desired consistency on the x -axis. Below this graph is a comparison of the segmentations produced by each method which achieve the desired consistency indicated at the left of each row, while using the minimum number of segments. For visualization, the segments used in computing the actual per-object consistency and efficiency values (displayed beneath each segmentation) are colored using a red-yellow colormap, while the non-object segments are colored with a blue-green colormap. Thus, all of these figures are best viewed in color. Note that a separate figure exists for each object of scenes with multiple objects (the Ground Truth Object Mask will differ in these cases).

In addition, we provide in Figure 39 the same overall performance summary as in the final figure of the paper, but at the full set of desired consistencies for completeness.

*Partial support provided by National Science Foundation (NSF) Grant IIS-0713406.

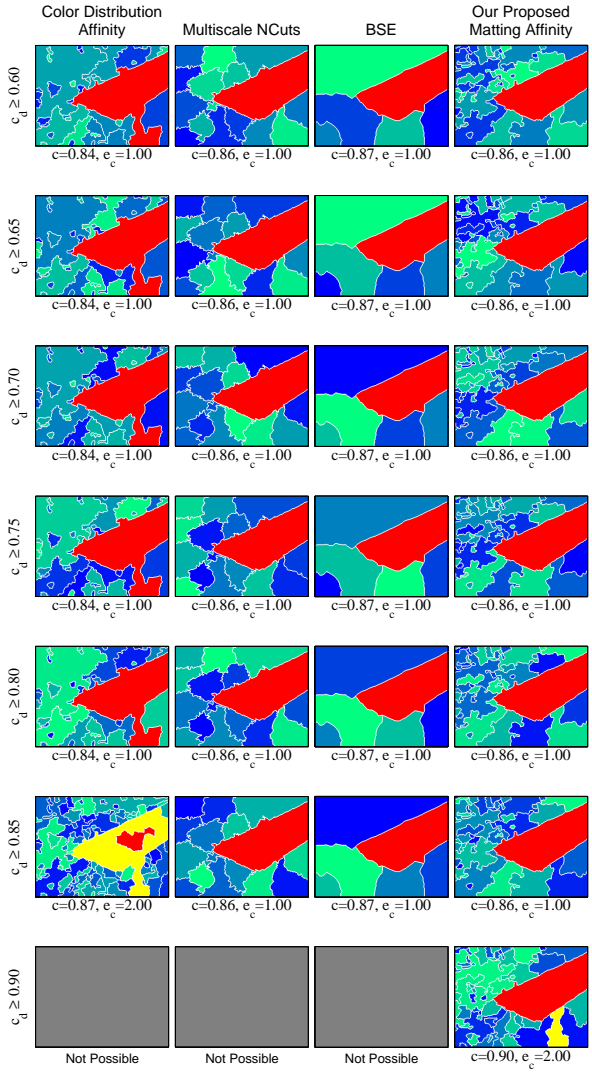
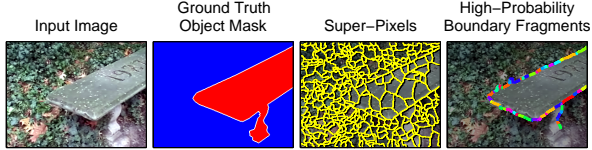
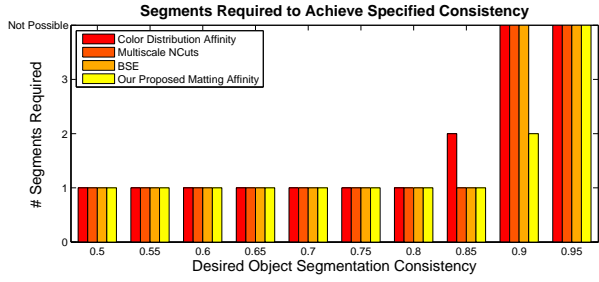


Figure 1. Bench. An “easy” example for which all methods work quite well.

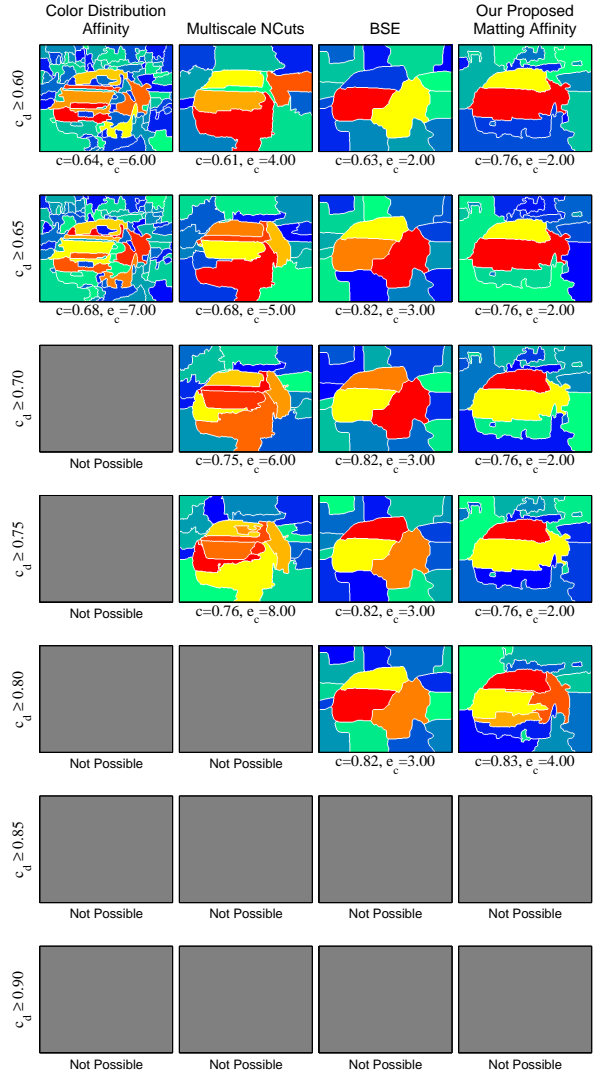
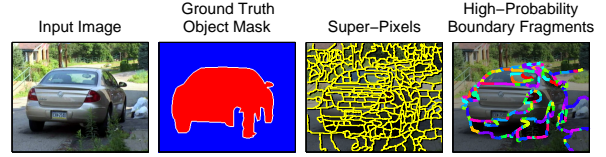
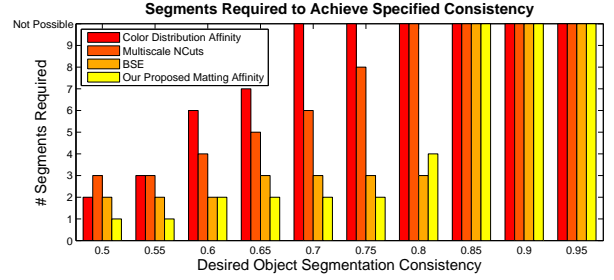


Figure 2. Car

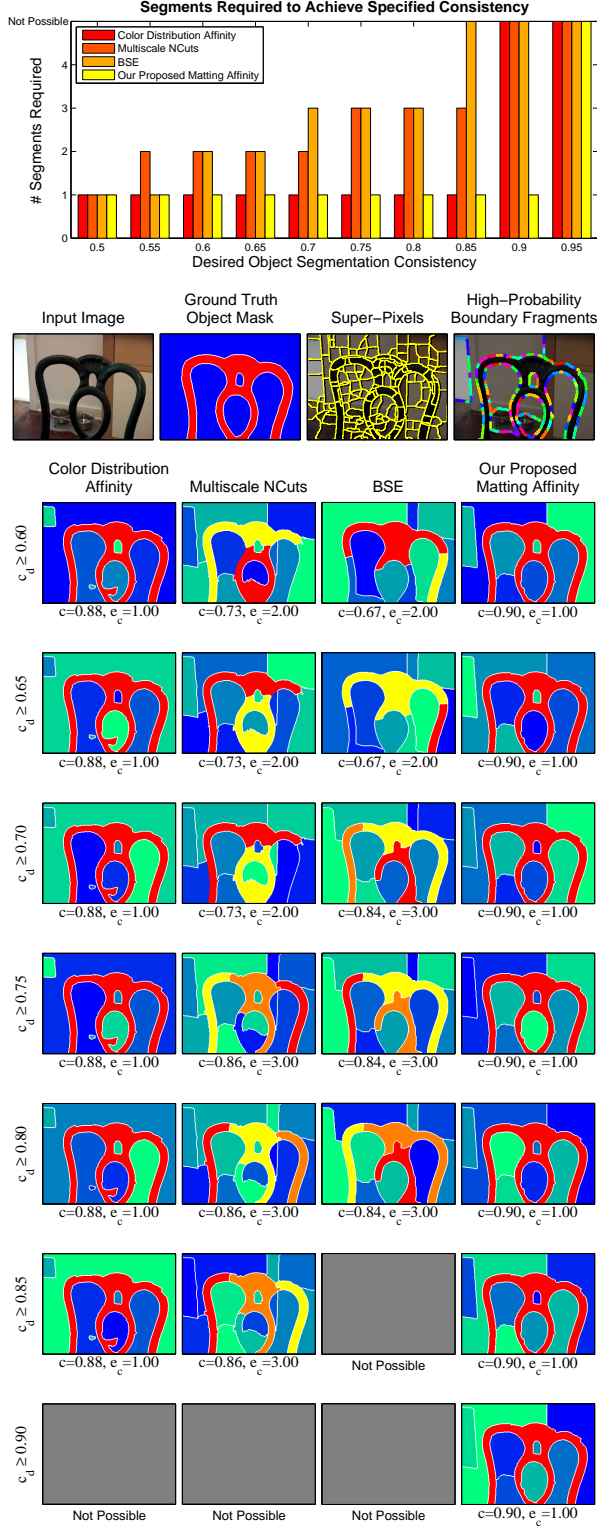


Figure 3. Chair. Both super-pixel-based methods perform well on this (and other) objects with narrow structures, unlike the pixelwise approaches which tend to break such objects apart with “cheap” cuts across narrow parts. Not surprisingly, color alone is sufficient for this object, though it is worth noting that the matting-based approach does not *hurt* (and actually helps a little).

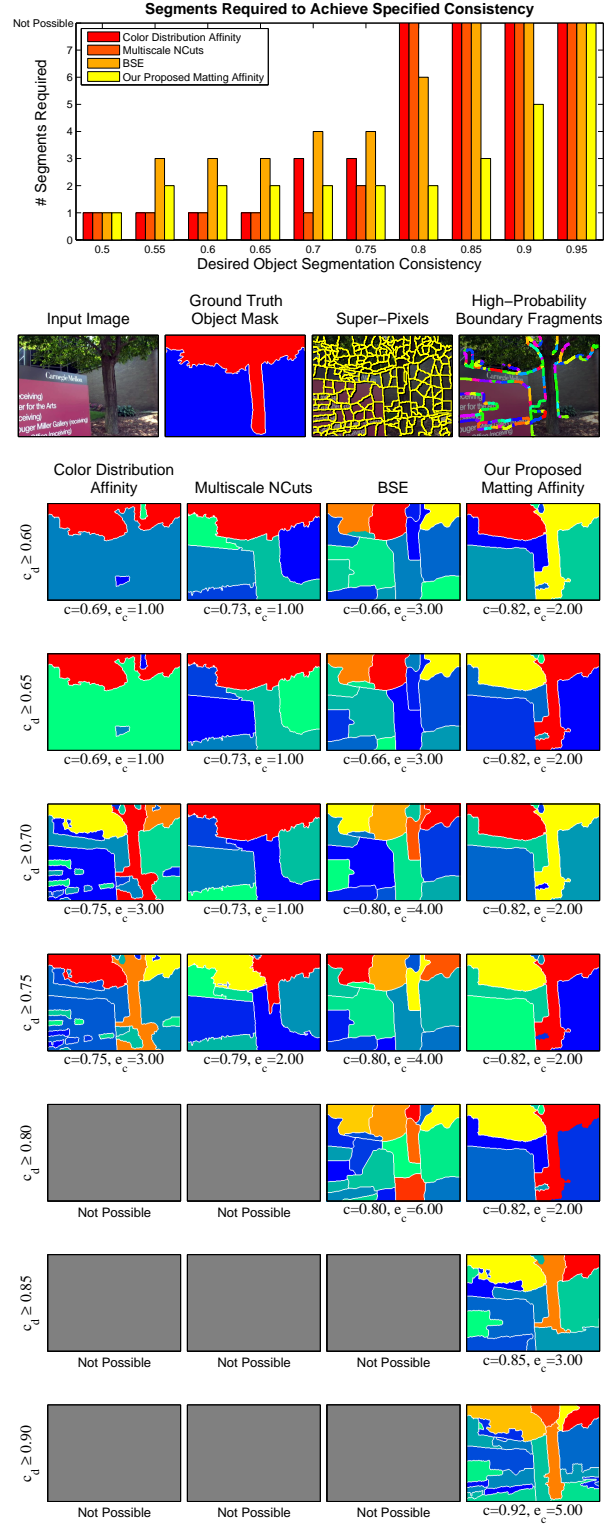


Figure 4. Tree

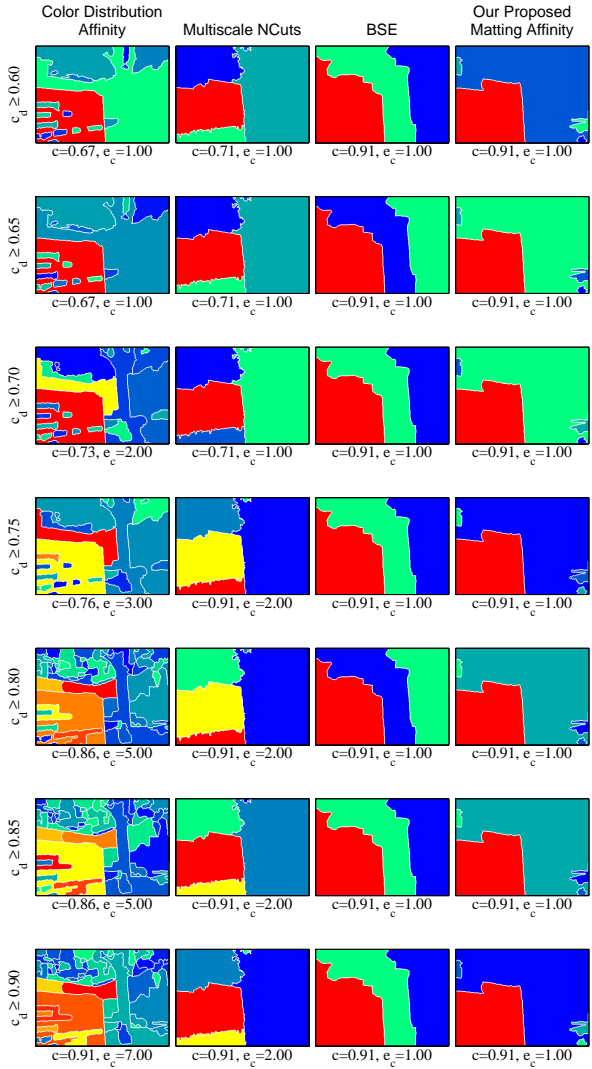
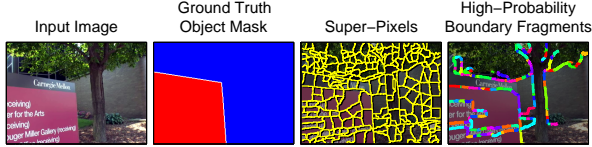
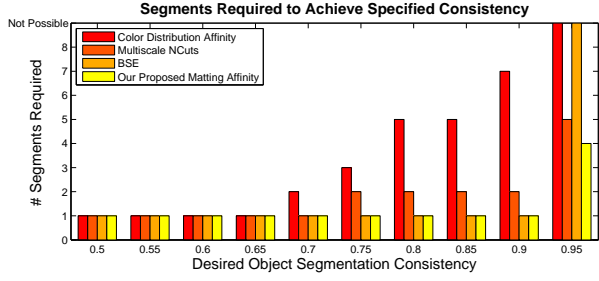


Figure 5. Sign. As discussed in the paper, evaluating segmentation is difficult. Numerical results do not always tell the whole story: our approach and BSE offer the same consistency (0.91) with only one segment, but qualitatively, our segmentation seems “better.”

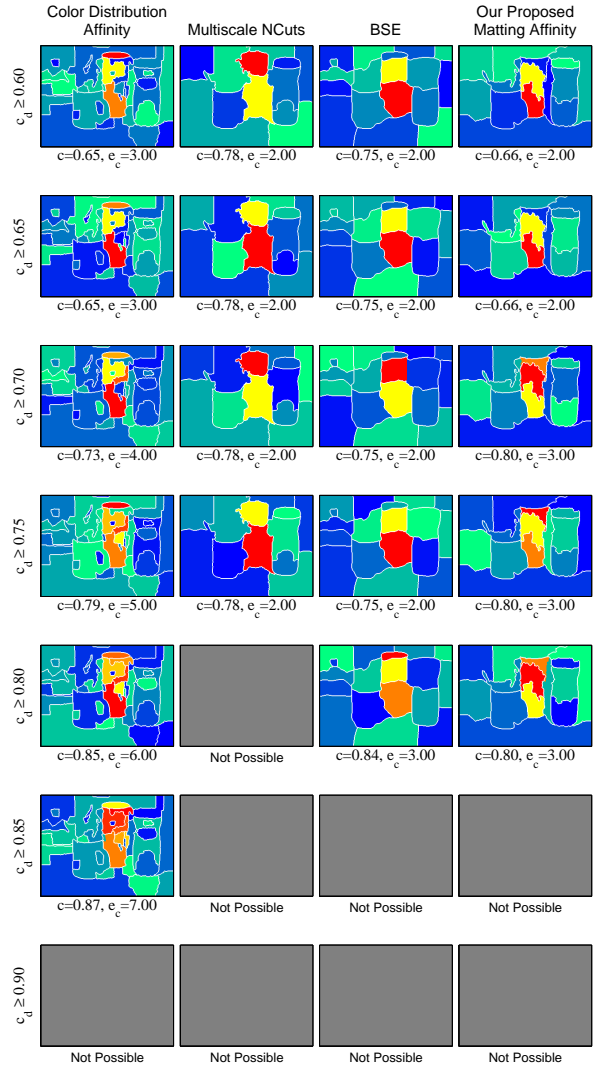
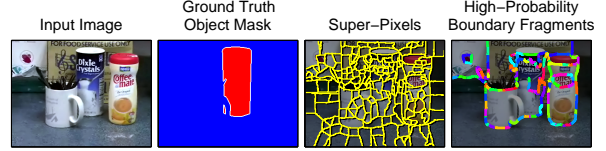
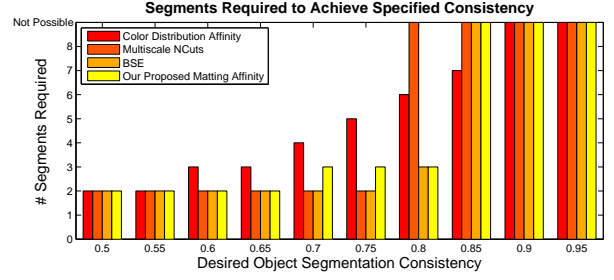


Figure 6. Coffee sugar.

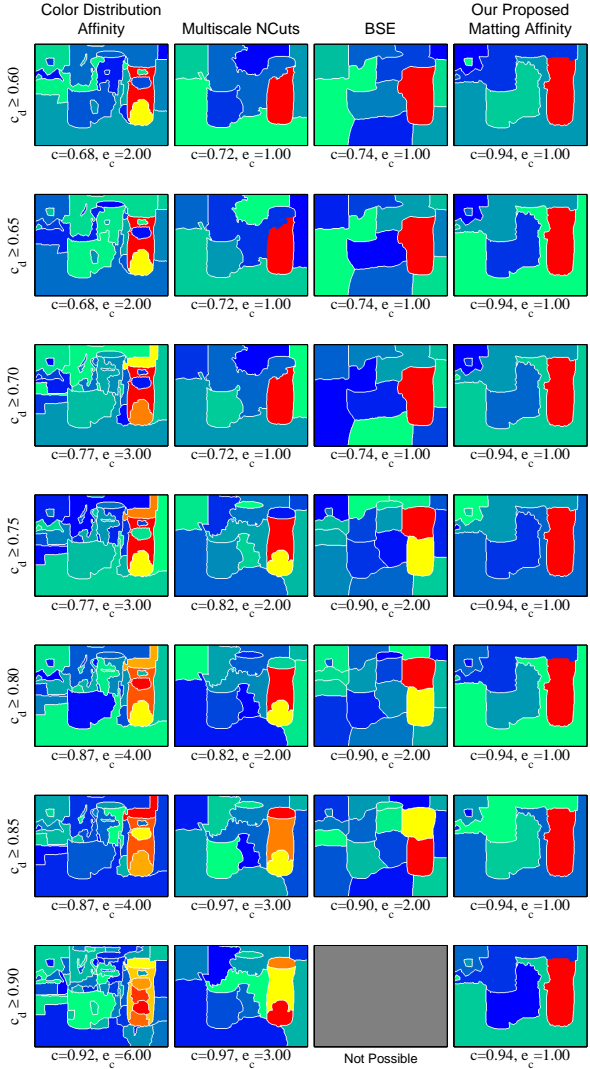
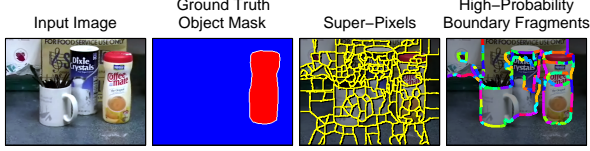
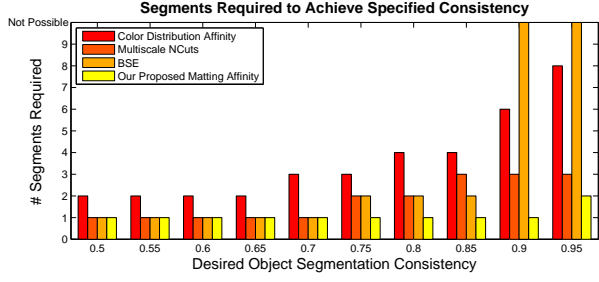


Figure 7. Coffee creamer.

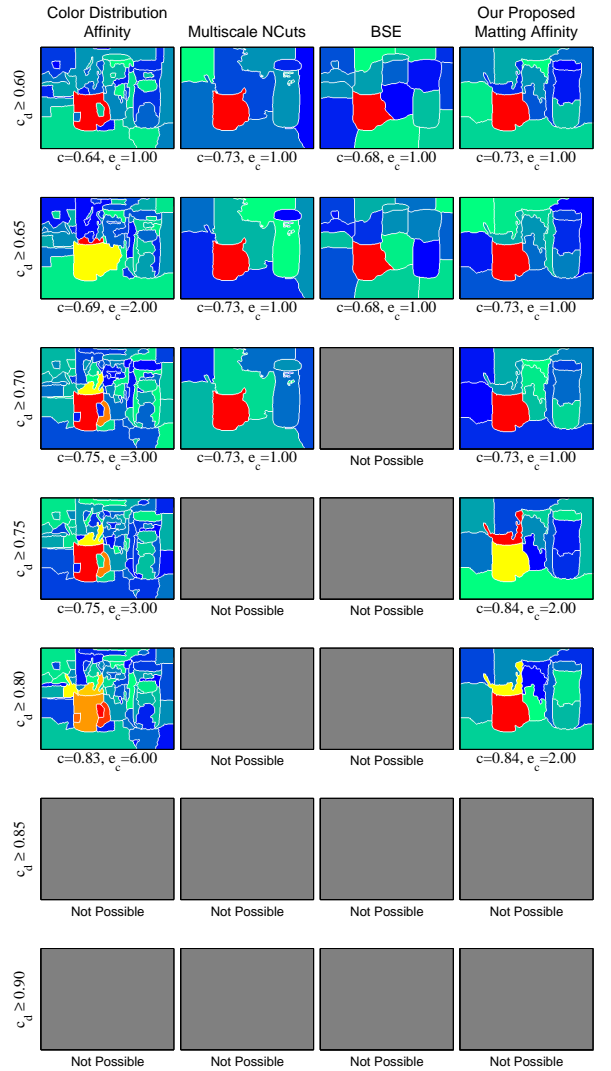
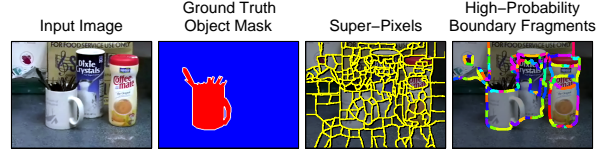
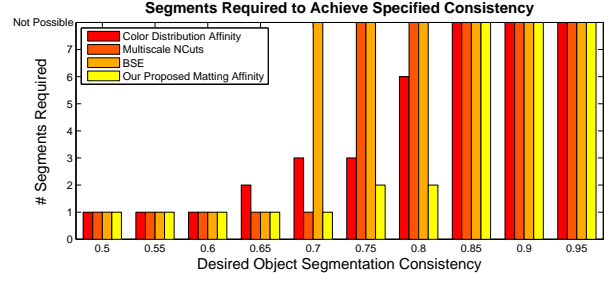


Figure 8. Coffee mug. Some subjectivity in selecting “whole objects” remains: the mug has been labeled together with straws and plastic-ware inside it, but each method clearly – and justifiably – attempts to separate the white mug from the black contents.

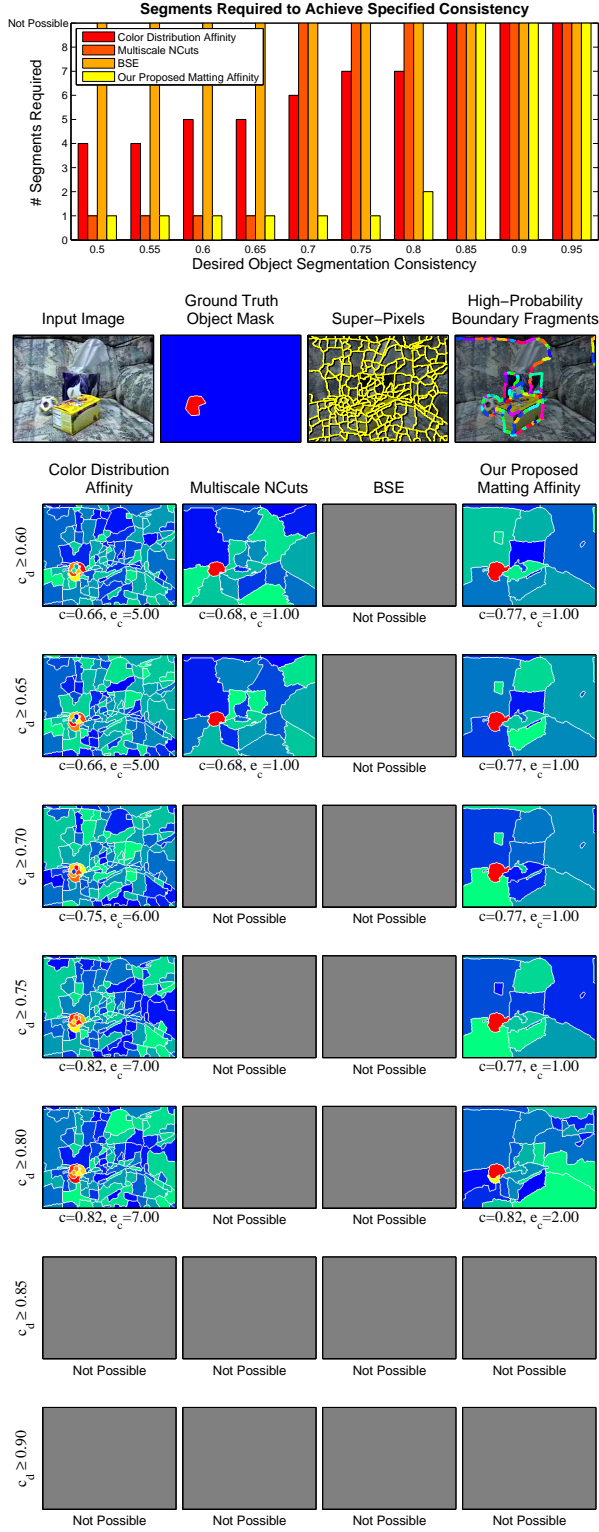


Figure 9. Origami ball. Both pixel-based methods have trouble with this small object. The super-pixel approaches fare better, but the matting-based approach achieves superior efficiency, despite the ball's multiple colors.

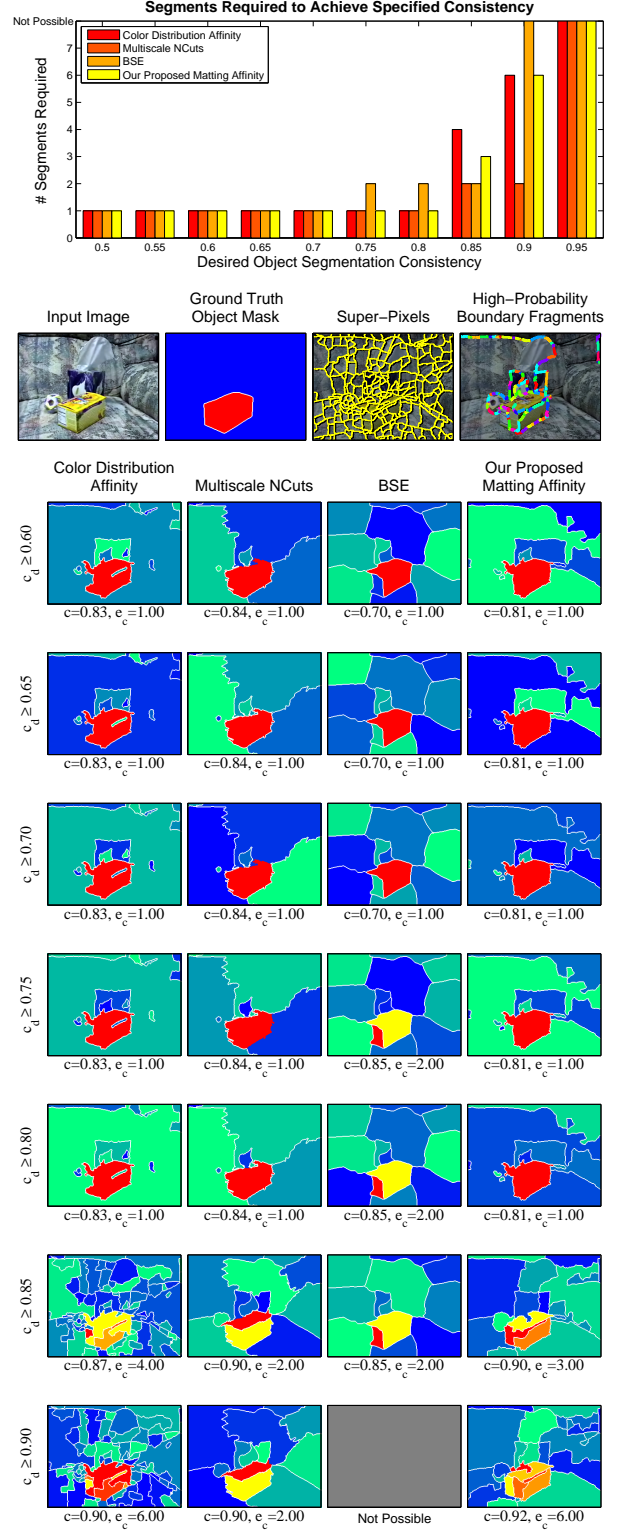


Figure 10. Tea Box.

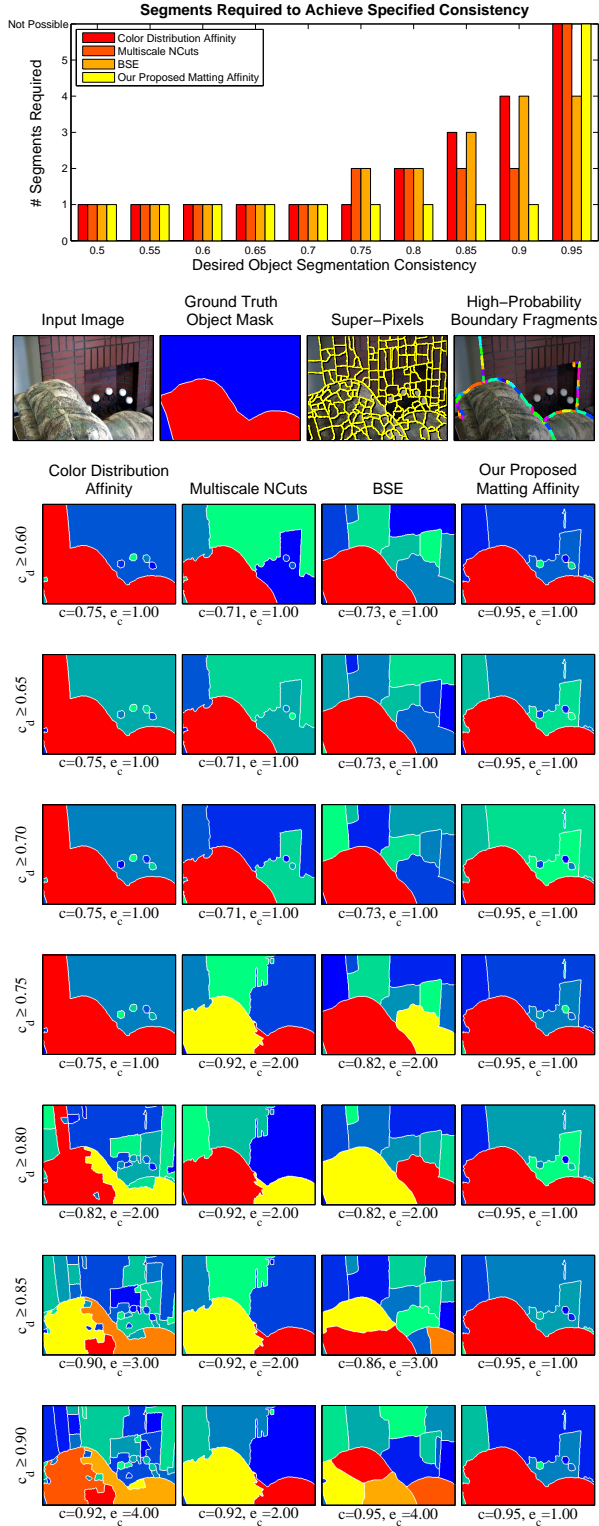


Figure 11. Despite the hypothesized boundary fragments between the two parts of the couch, the appearance reasoning of the matting puts the whole couch in one segment for our approach. Also, the use of super-pixels seems to help avoid a cheap cut through the object to the bottom of the image.

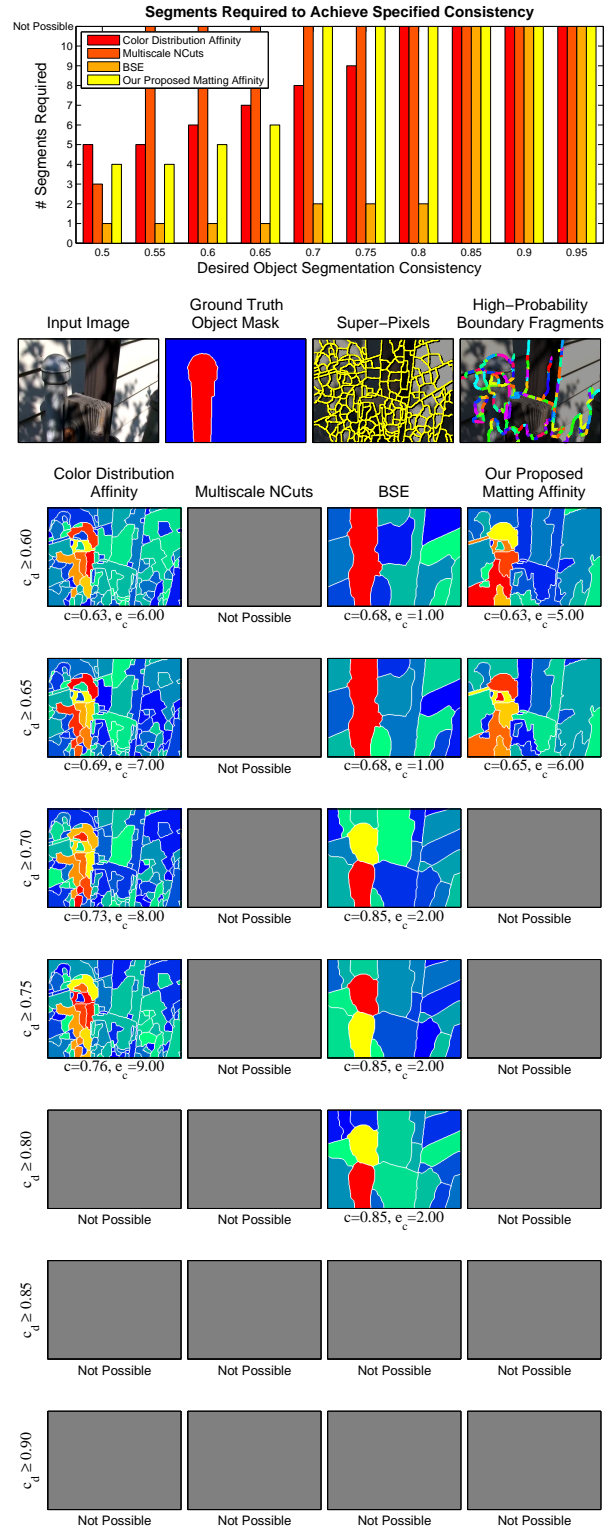


Figure 12. The high-contrast illumination and shadows in this scene make it very difficult. They seem to confuse the boundary-hypothesis step and, in turn, our matting-based approach. (See also Figures 13-14.)

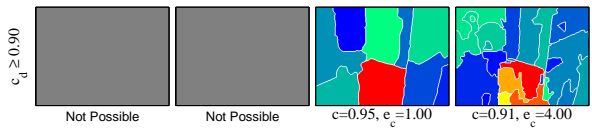
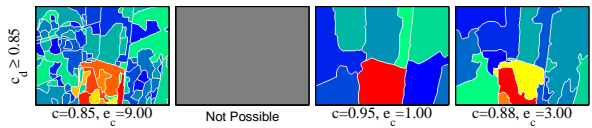
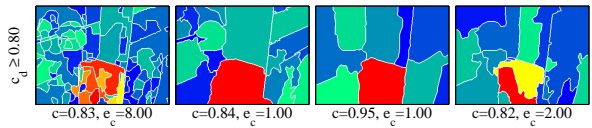
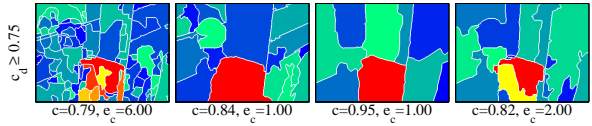
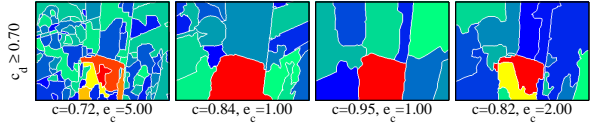
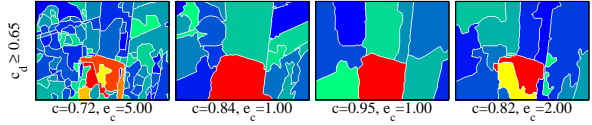
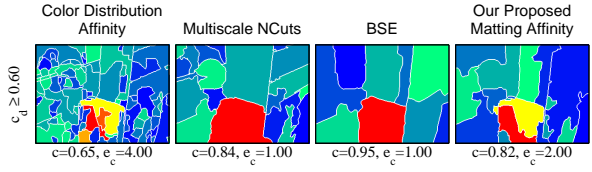
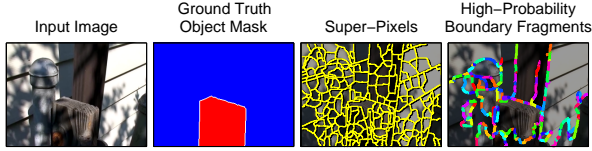
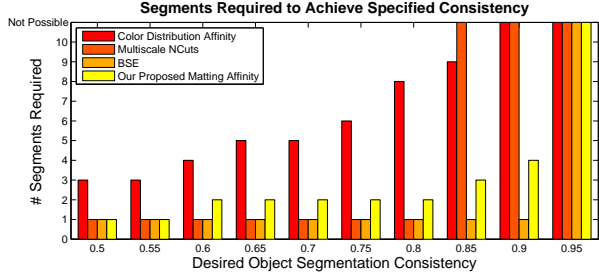


Figure 13. Difficult fencepost scene.

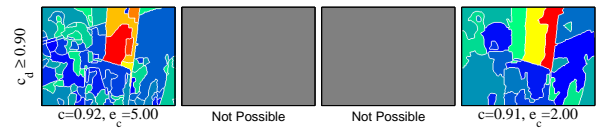
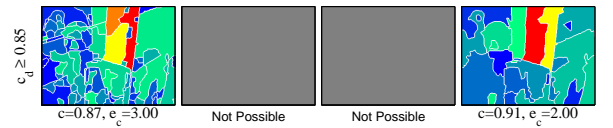
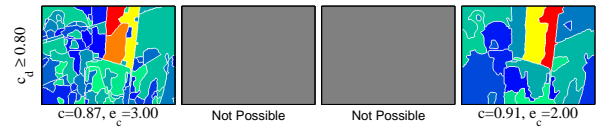
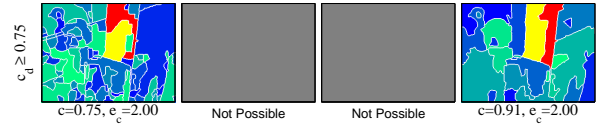
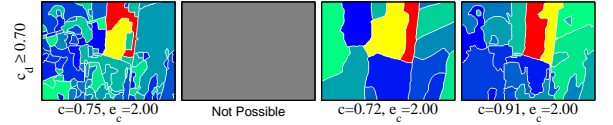
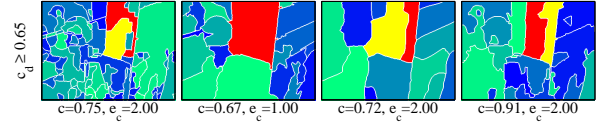
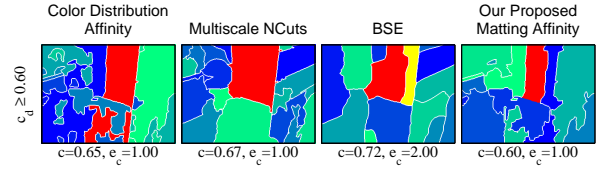
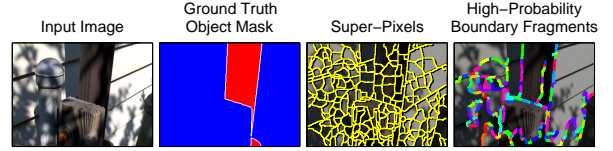
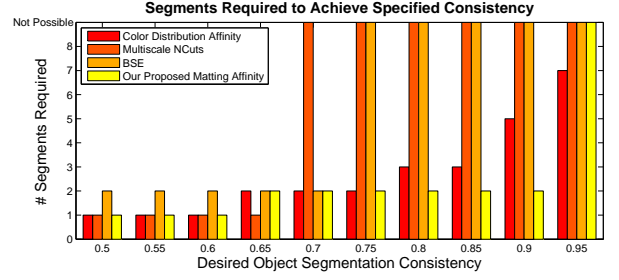


Figure 14. Difficult fencepost scene.

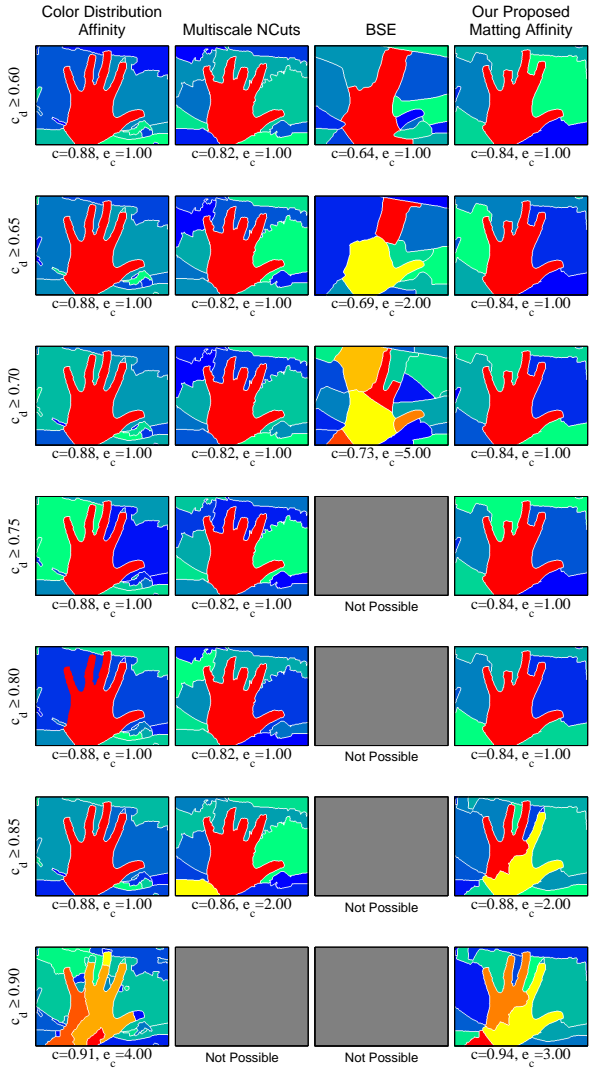
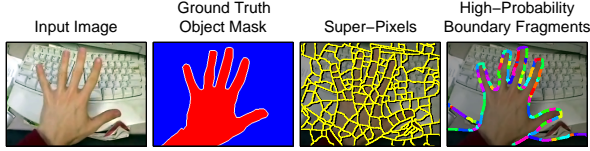
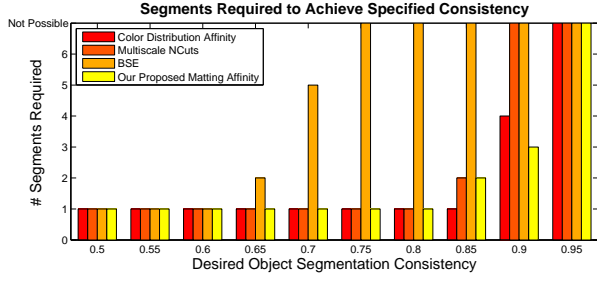


Figure 15. Hand 1. Here, color alone is a strong cue for segmentation, but our approach also performs fairly well. All methods are reluctant to include the differently-colored bit of sleeve labeled in the ground truth.

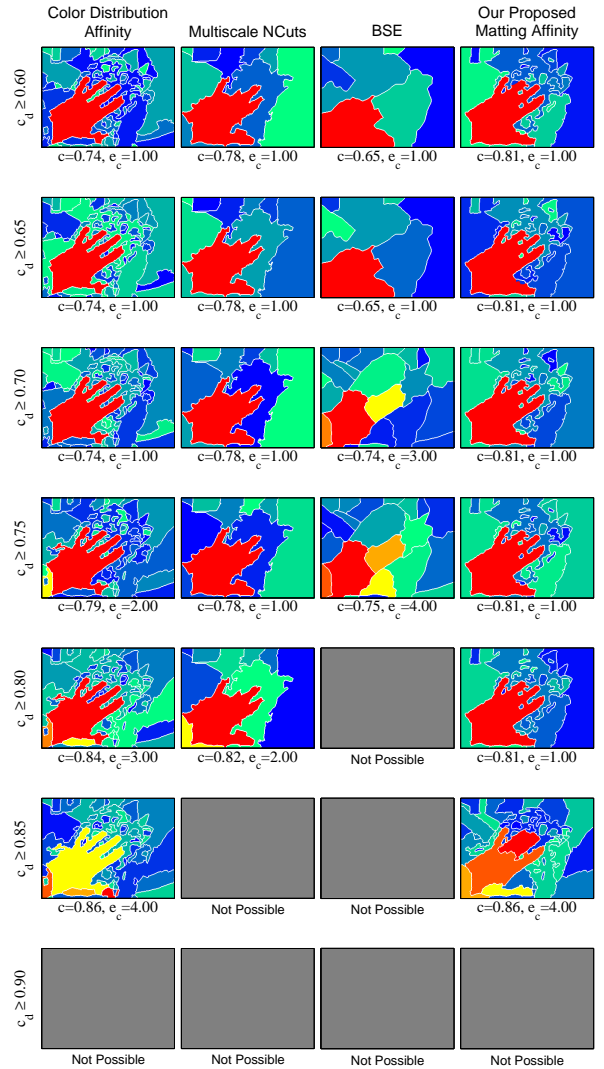
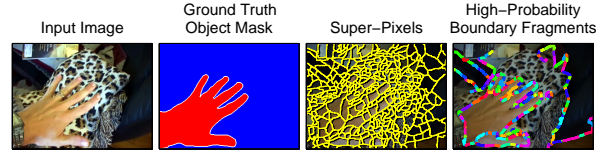
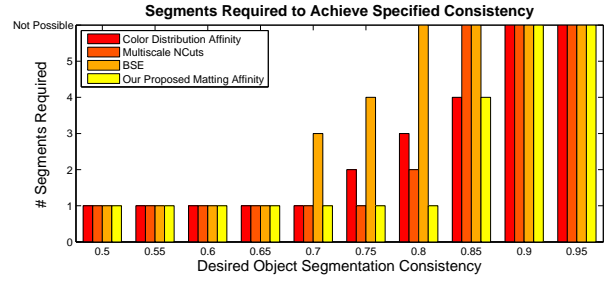


Figure 16. Hand 2. The texture of the blanket behind the hand serves to confuse all methods, but ours does a good job of extracting the fingers. Here again, super-pixels seem to help with the extraction of narrow structures (particularly as compared to BSE).

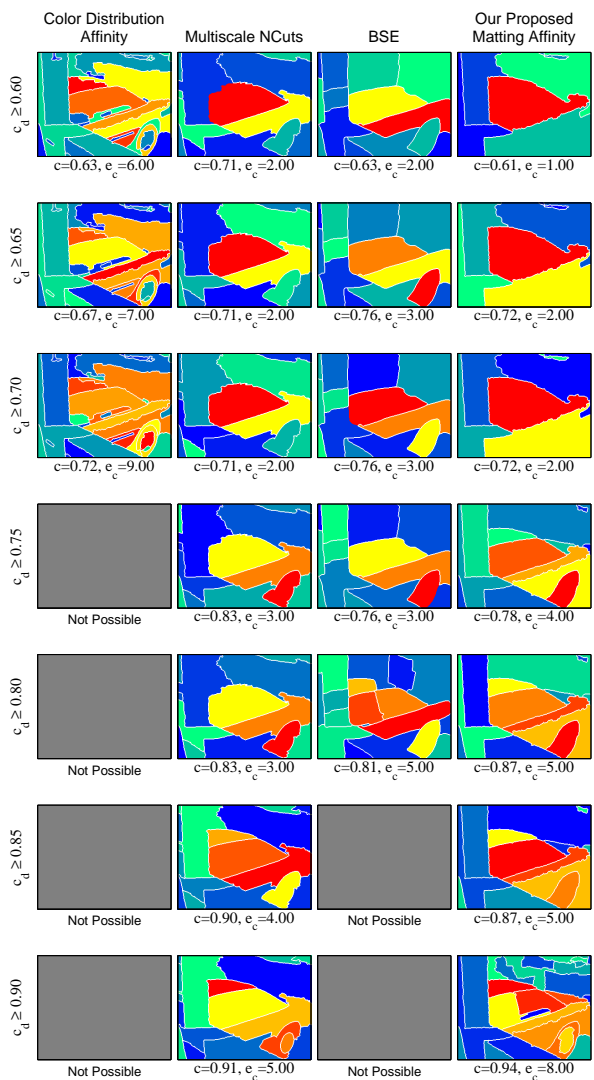
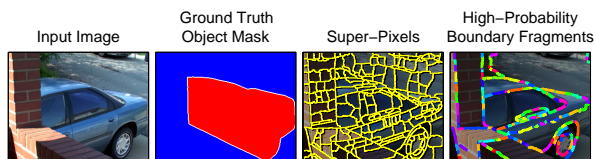
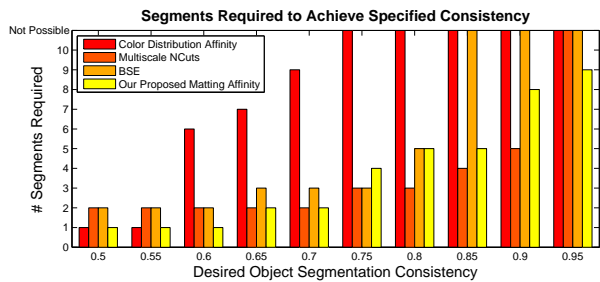


Figure 17. Car.

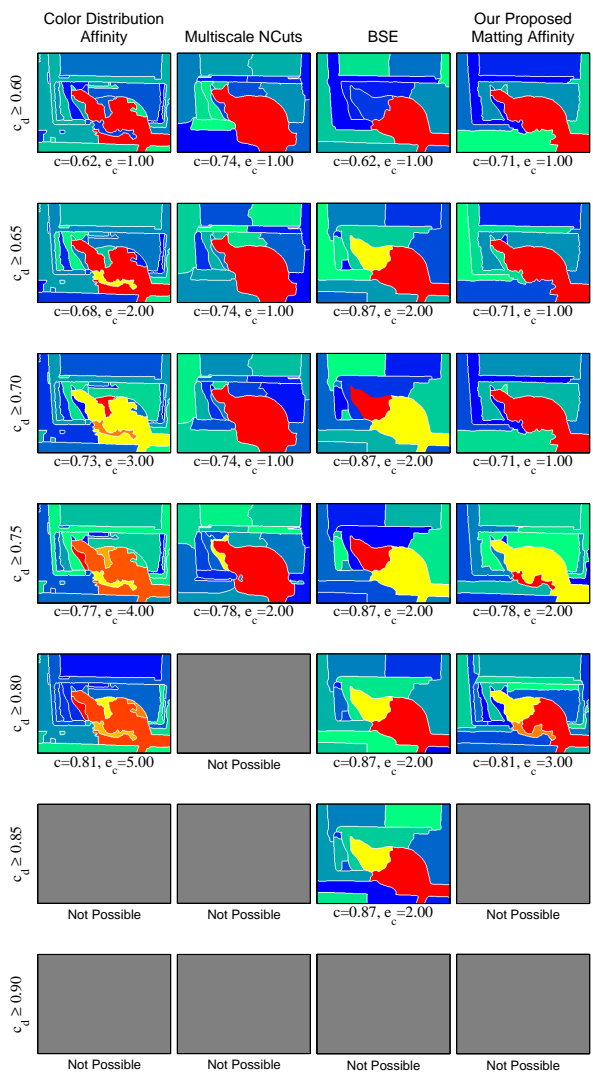
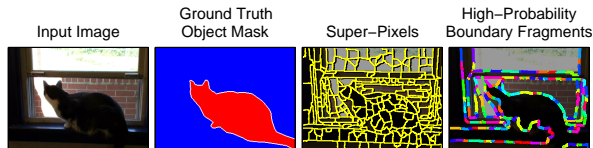
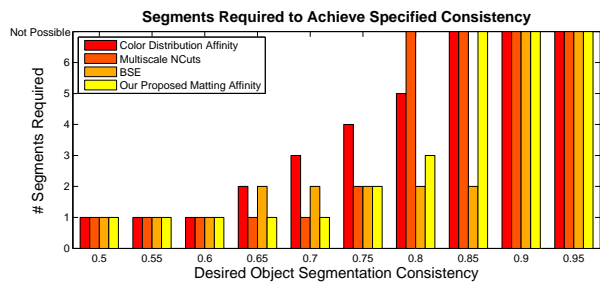


Figure 18. Cat in window. The very low contrast between the cat and the shadowed window frame make this a difficult scene.

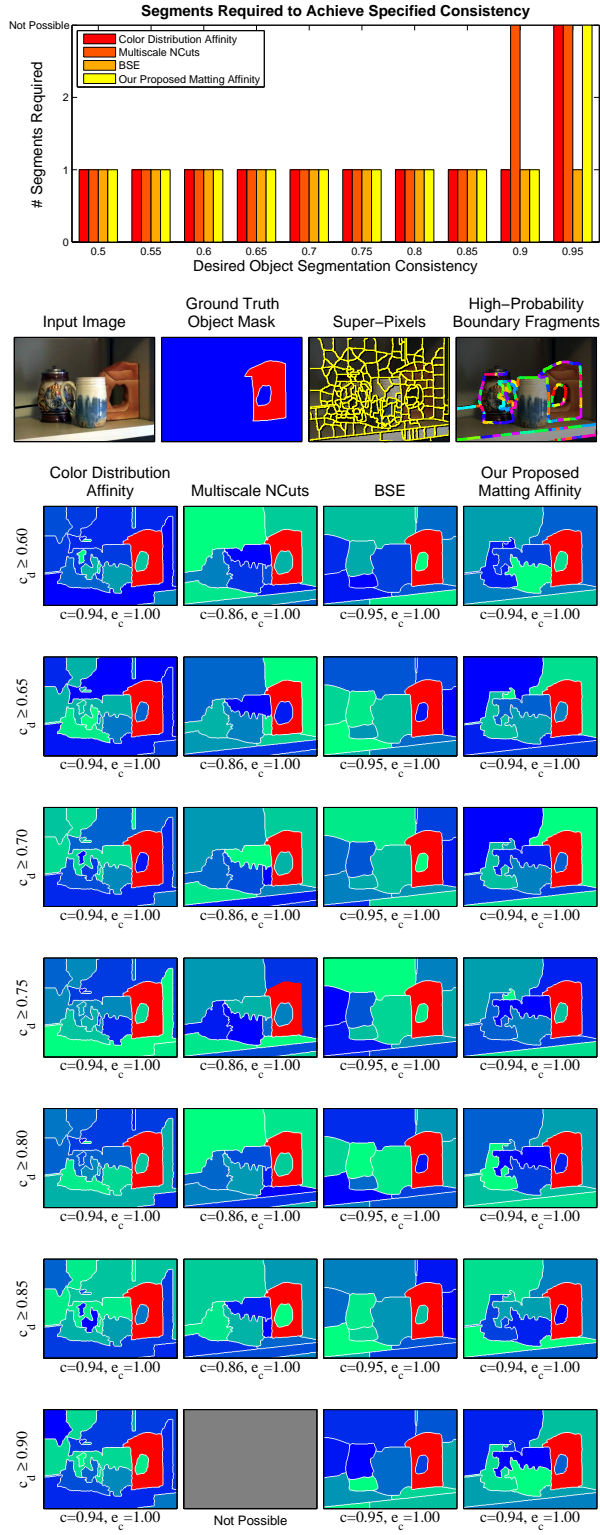


Figure 19. Bookend. Another “easy” object for which all methods perform well.

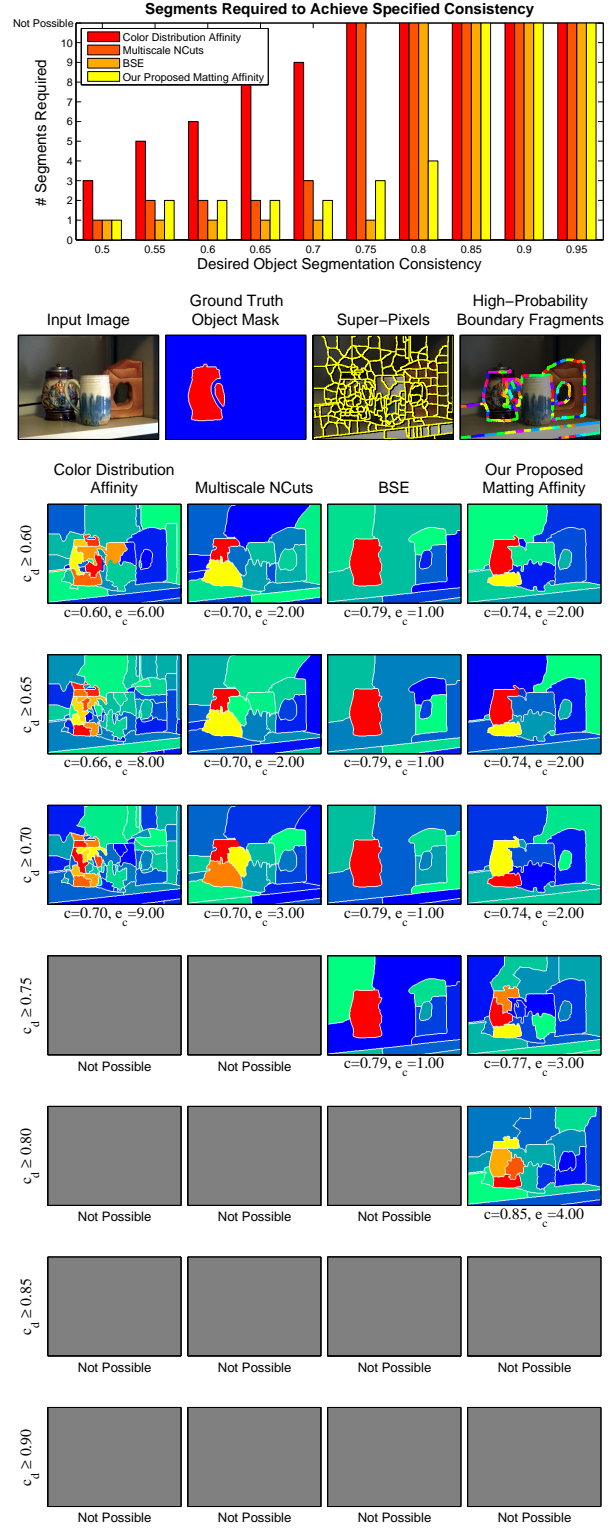


Figure 20. Beer stein.

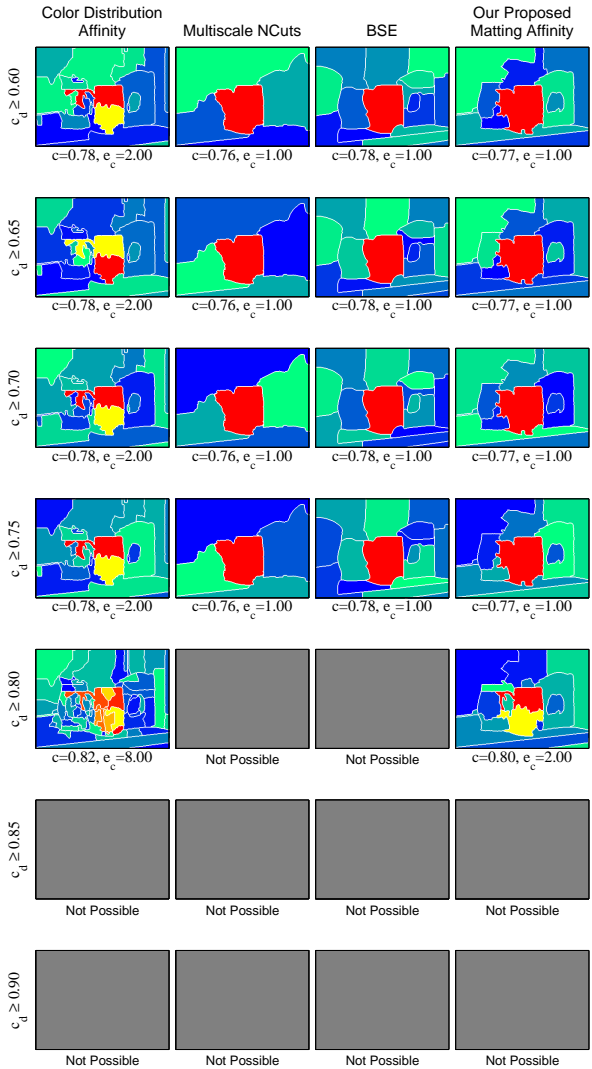
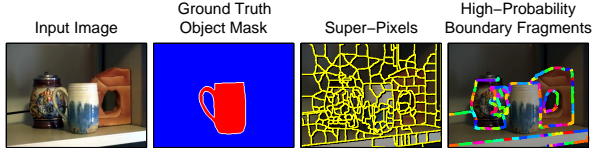
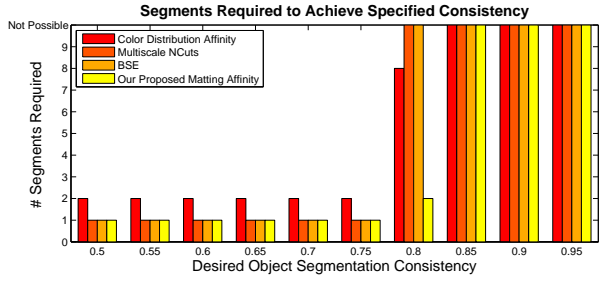


Figure 21. Mug.

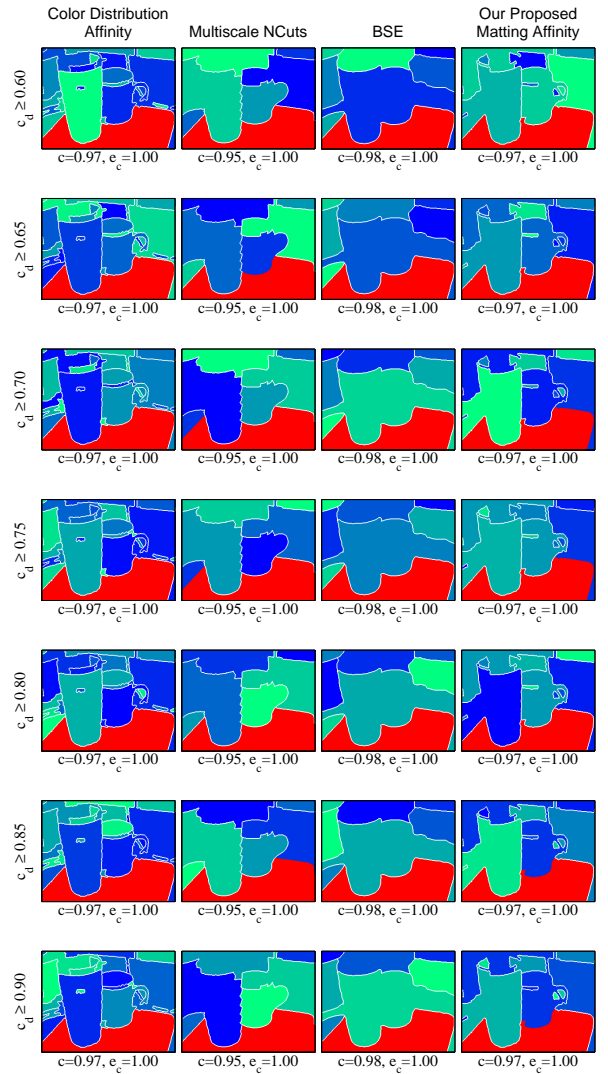
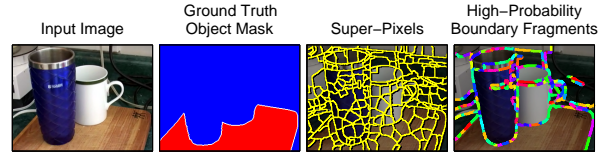
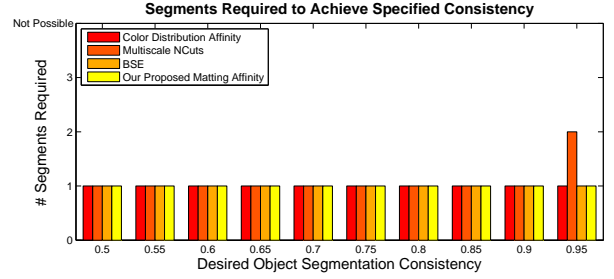


Figure 22. Cutting board. With such uniform appearance, this is not a difficult object for any method. It is somewhat surprising, however, that neither pixelwise NCuts approach split the object with a cheap cut as in other examples, *e.g.* at the very narrow point between the bottom of the blue cup and the image border. Such unpredictable performance is a common side effect of standard, pixelwise affinities.

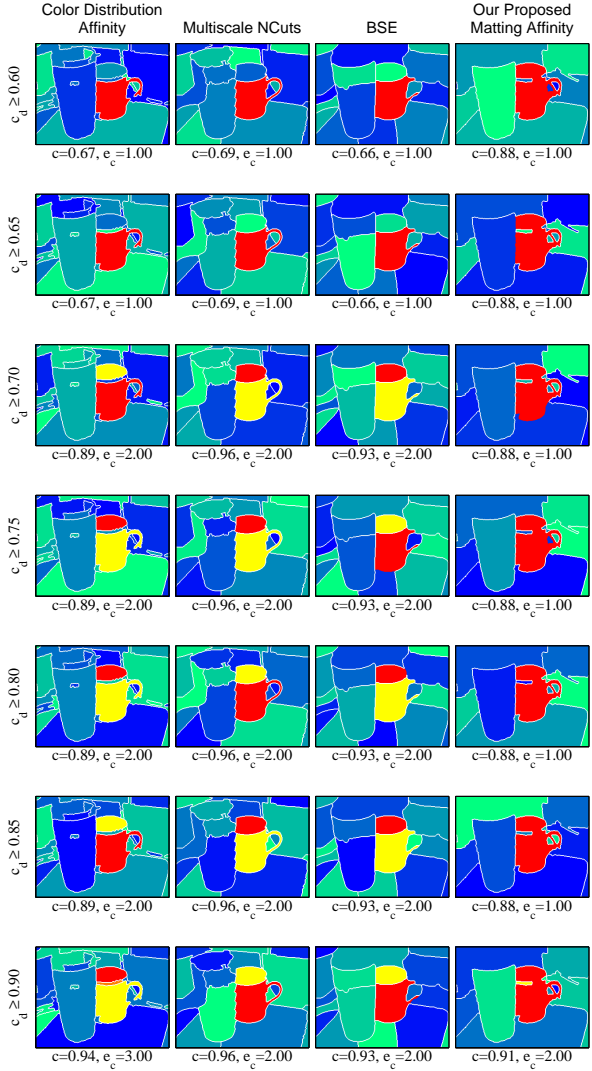
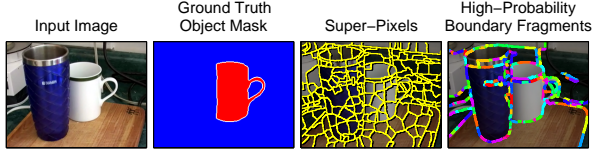
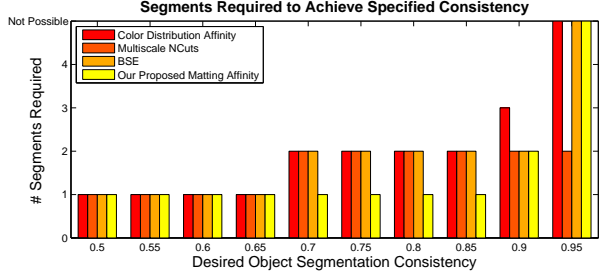


Figure 23. Coffee mug.

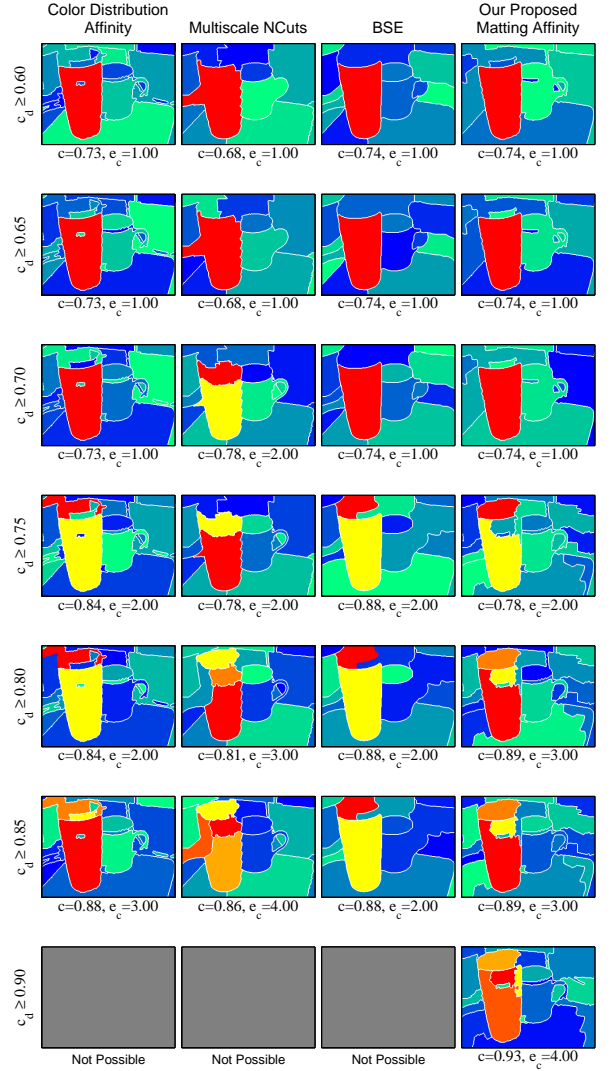
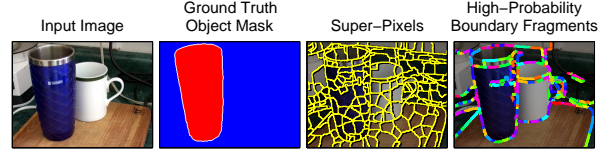
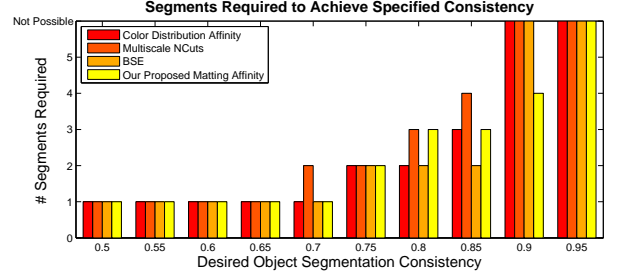


Figure 24. Cup.

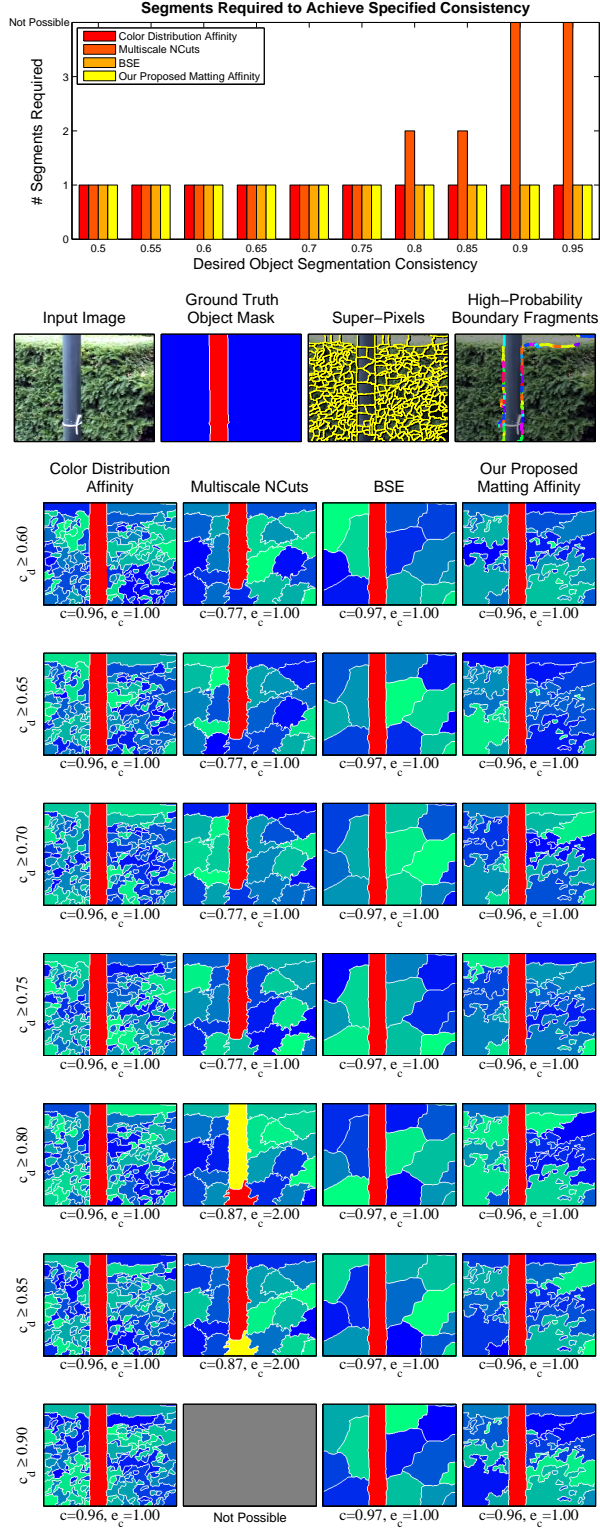


Figure 25. Post.

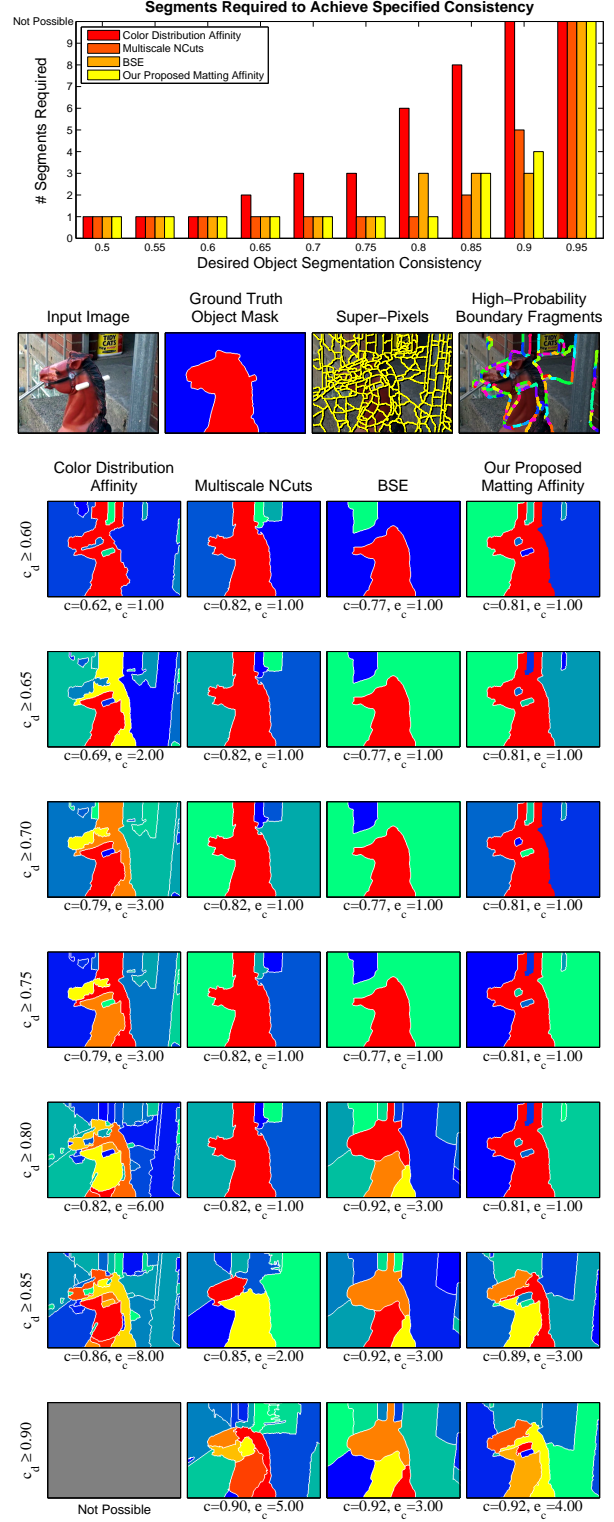


Figure 26. Rocking horse. The low contrast between the top of the horse's head and the brick cause bleeding for many results.

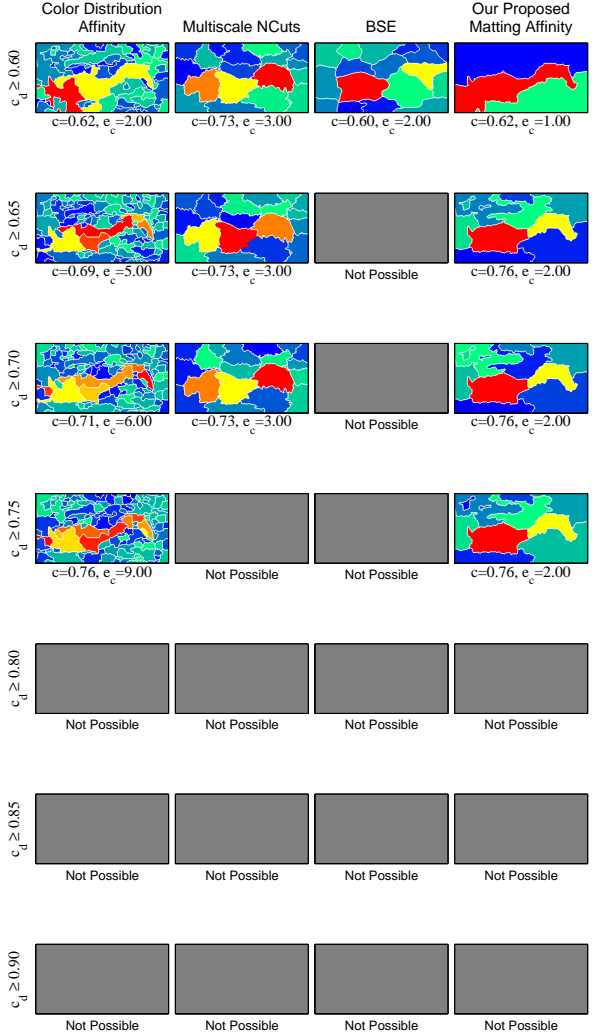
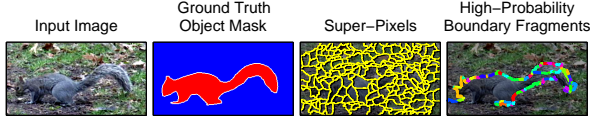
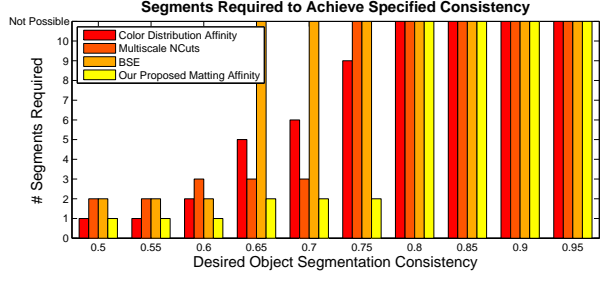


Figure 27. Squirrel. Good boundary hypotheses along with the strong appearance-reasoning offered by the mattes help our method outperform the the other approaches on this difficult ex-ample, similar to the one in the main paper.

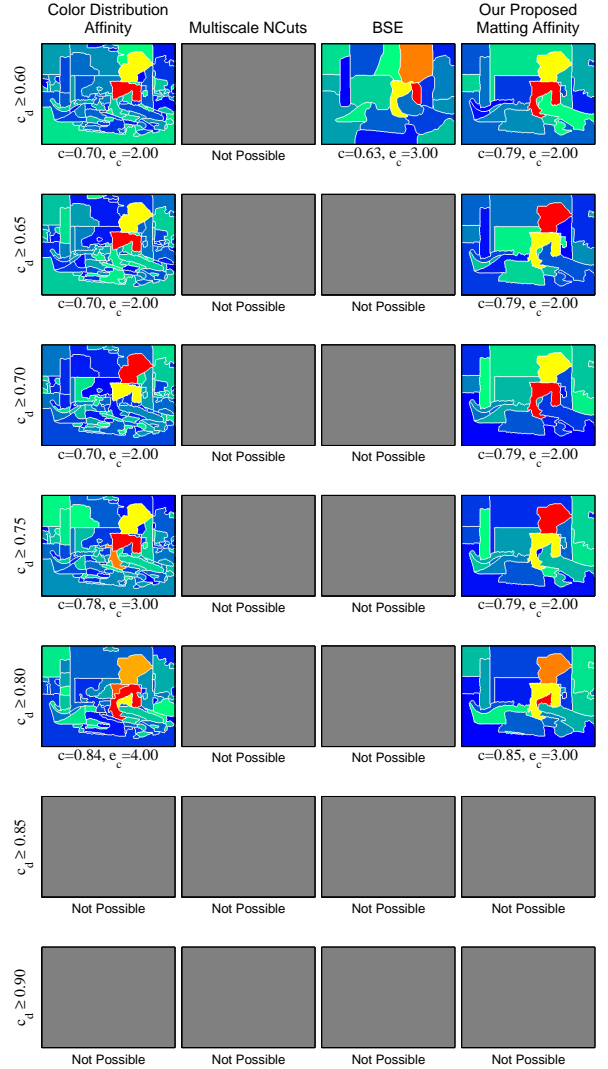
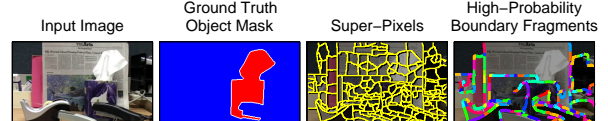
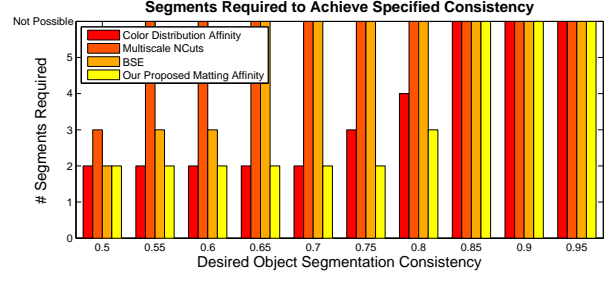


Figure 28. Kleenex box. Again, color alone seems sufficient here, but unlike the other methods, our matting-based approach does not hurt – and even offers some improvement.

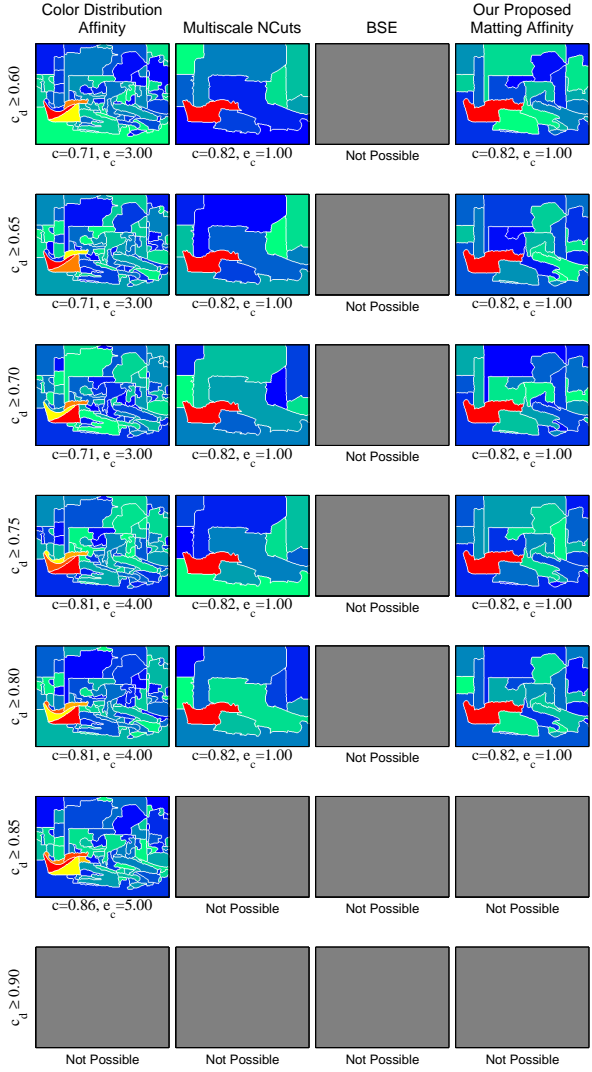
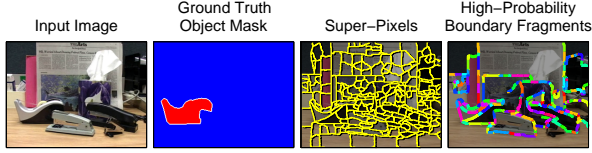
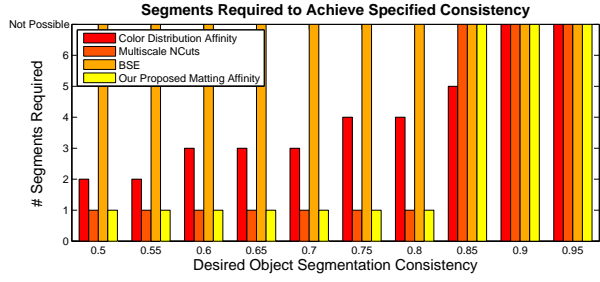


Figure 29. Tape dispenser.

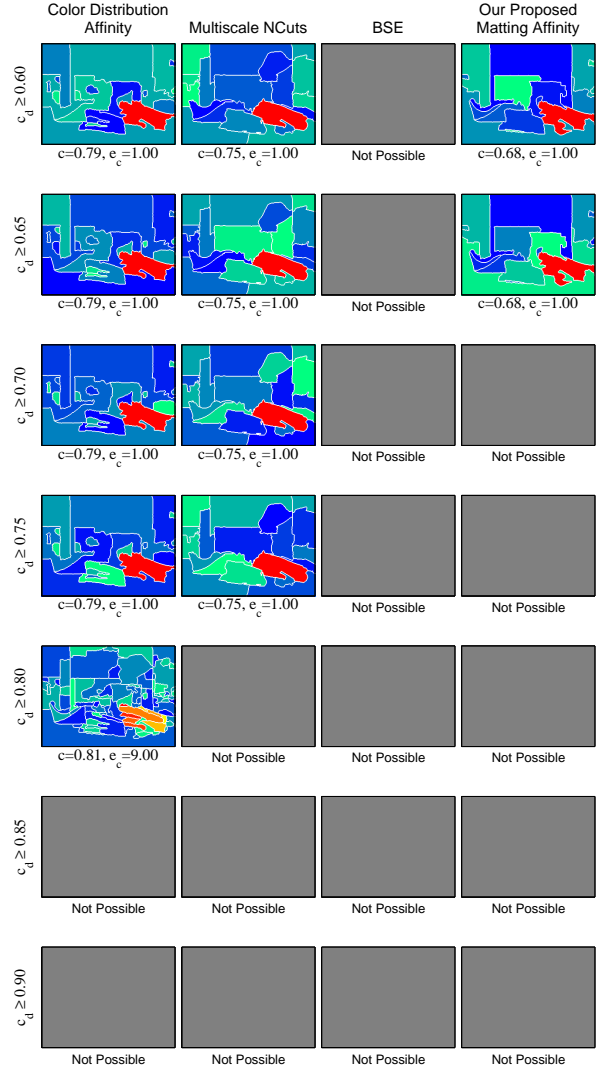
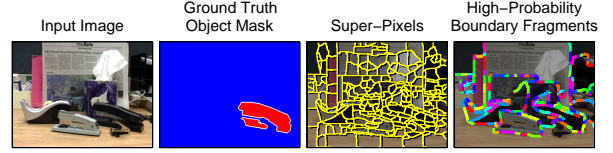
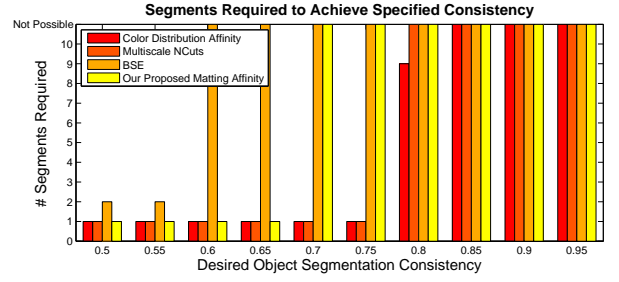


Figure 30. Stapler 1.

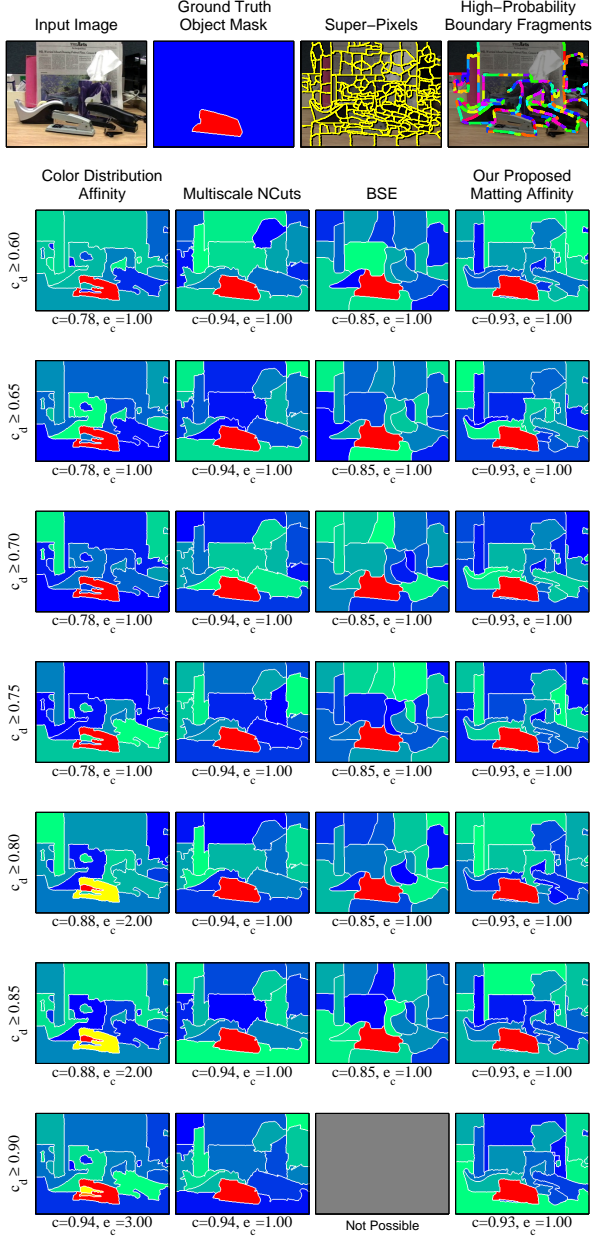
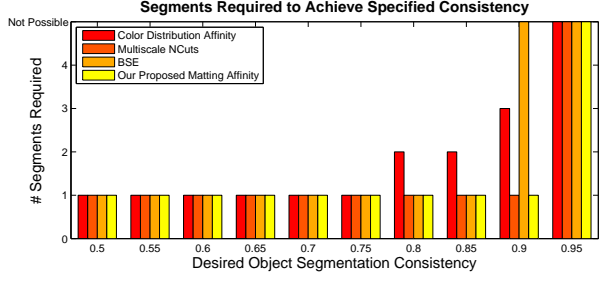


Figure 31. Stapler 2.

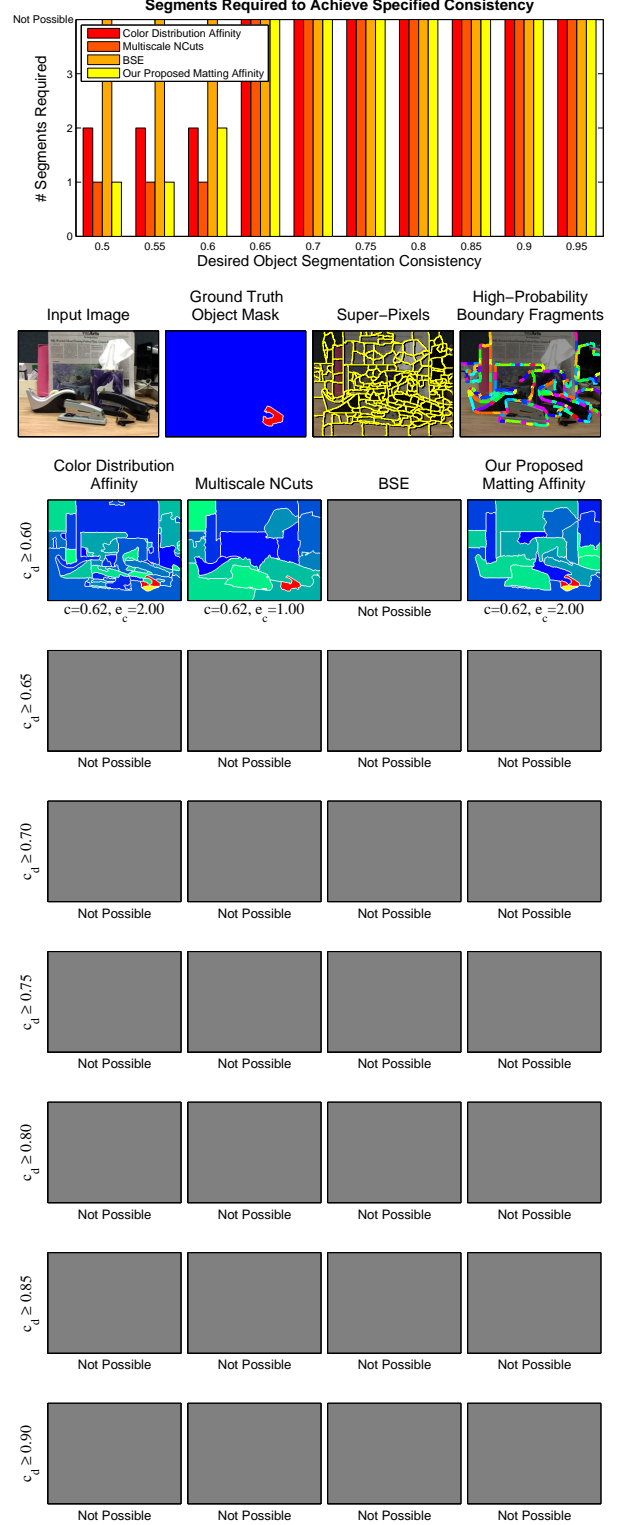


Figure 32. Staple remover. All methods have a difficult time with this small object, particularly since the top is the same color (black) as the stapler behind it (and thus, the underlying over-segmentation is actually incorrect: the correct boundary fragment is not even hypothesized).

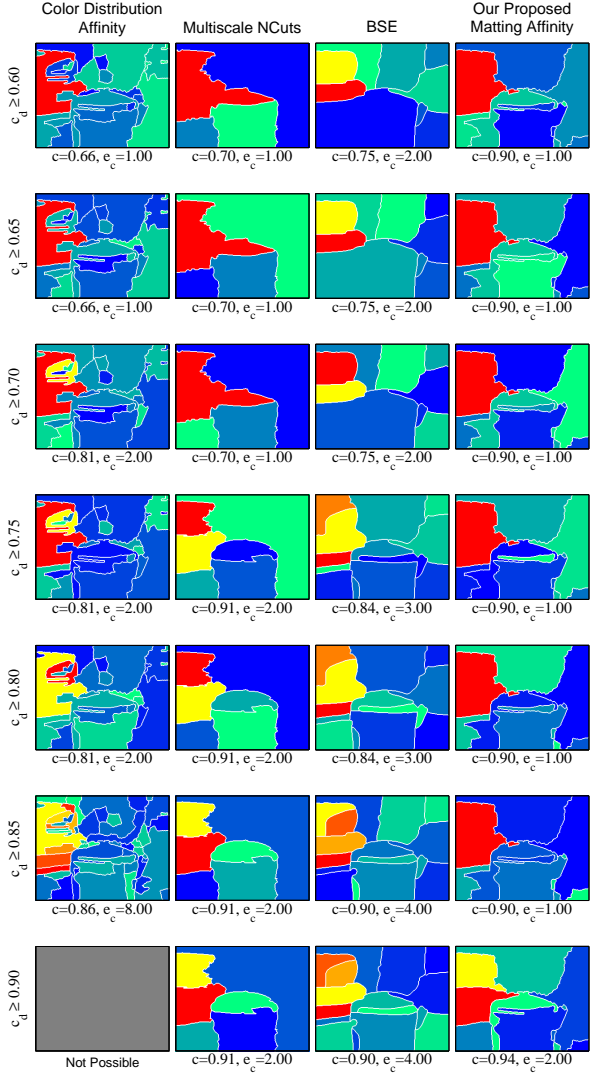
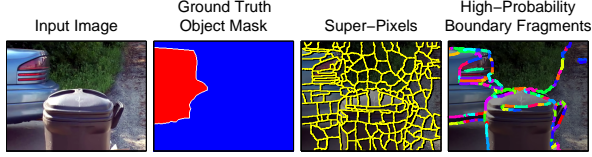
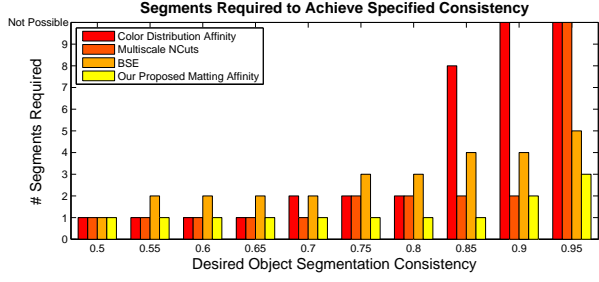


Figure 33. Car, rear.

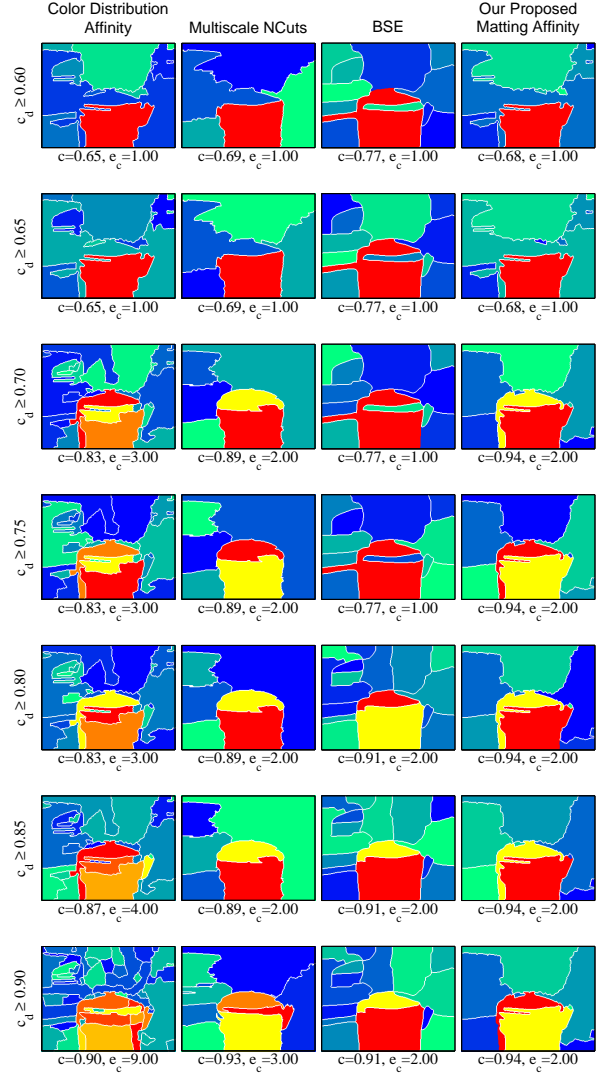
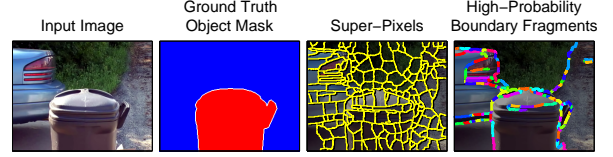
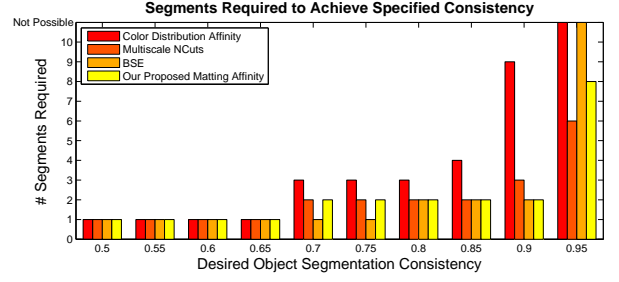


Figure 34. Trash can.

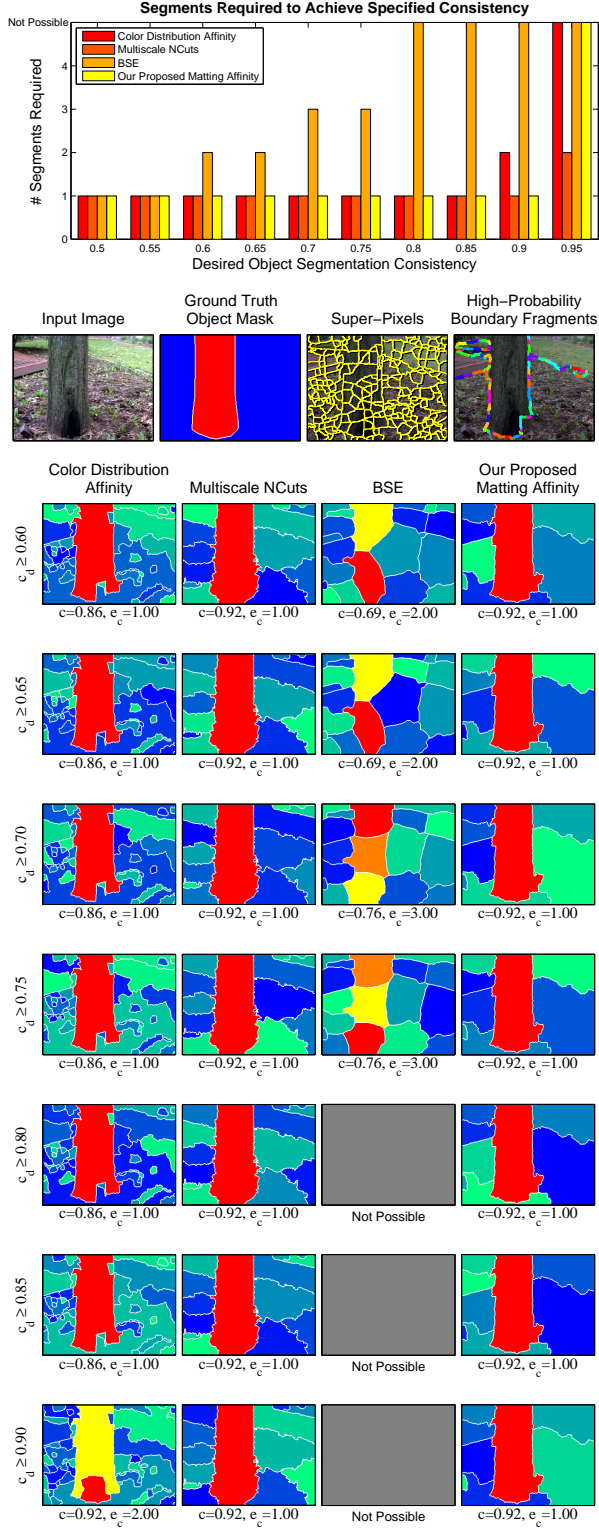


Figure 35. Tree.

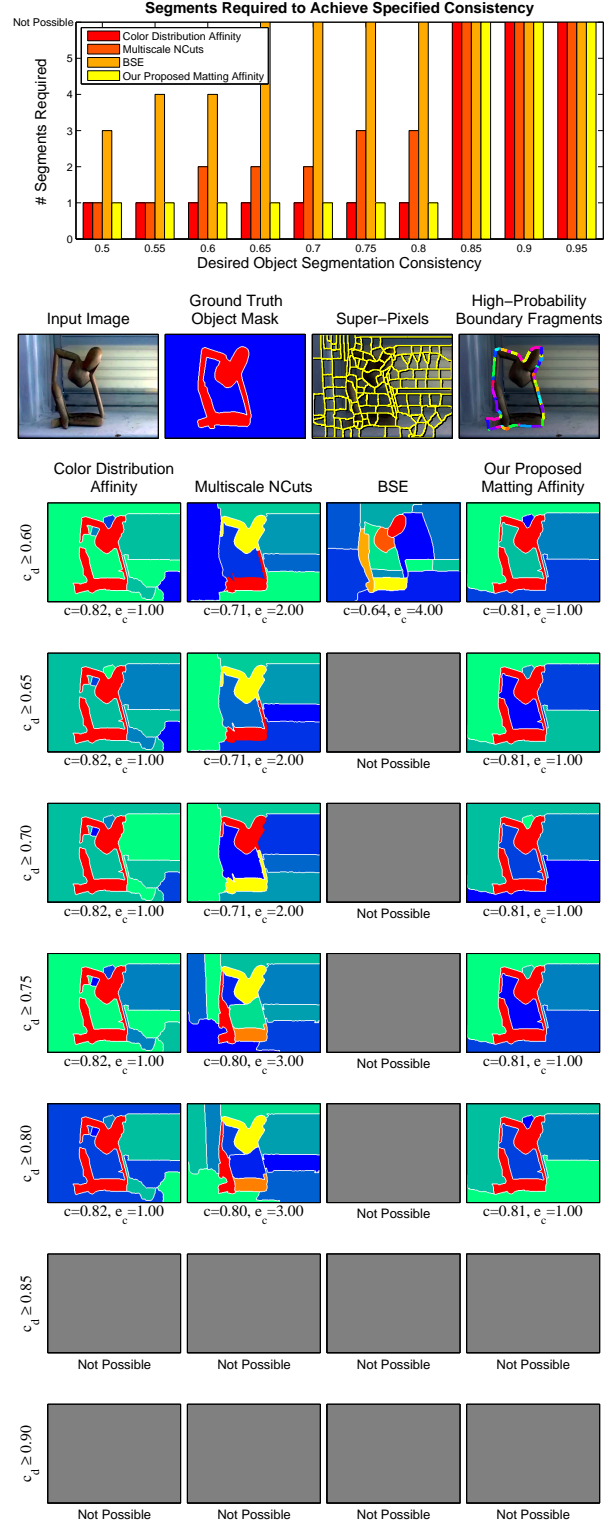


Figure 36. Wooden statue. Again, narrow structures prove challenging for the methods relying on pixelwise affinities.

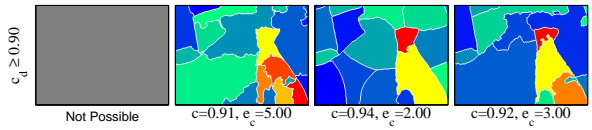
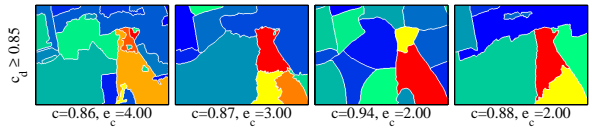
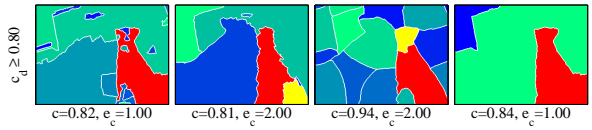
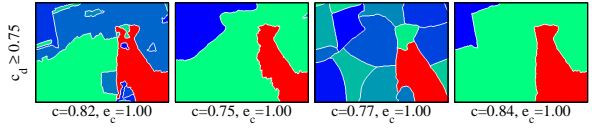
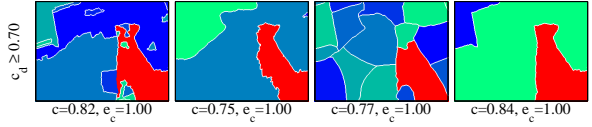
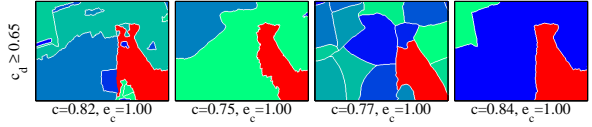
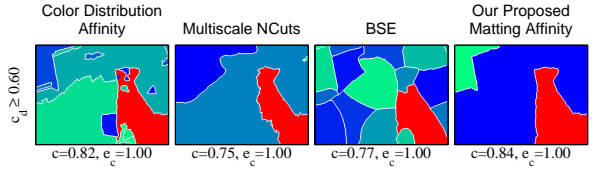
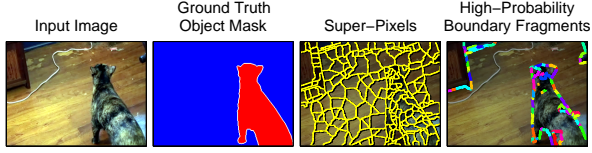
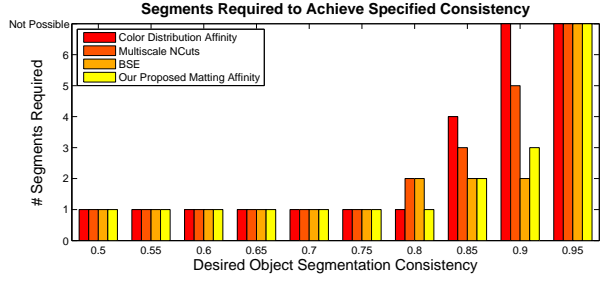


Figure 37. Cat 1.

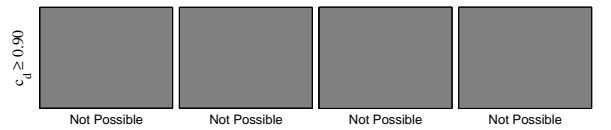
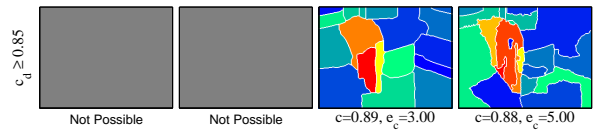
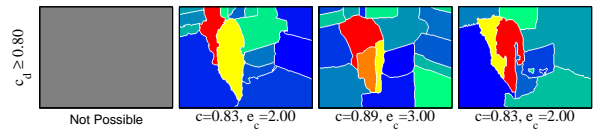
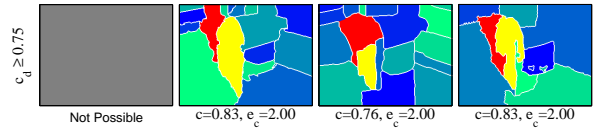
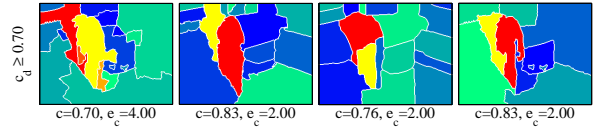
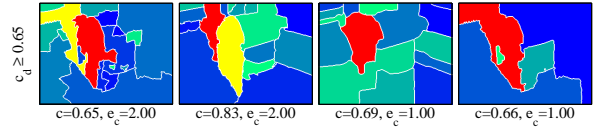
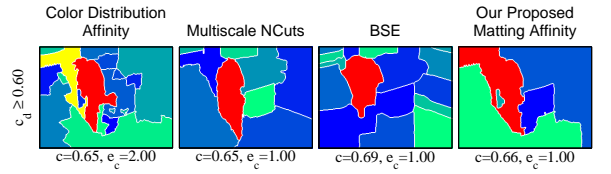
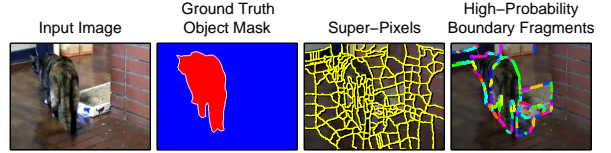
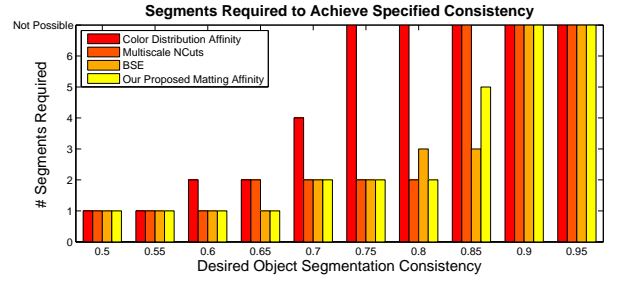


Figure 38. Cat 2.

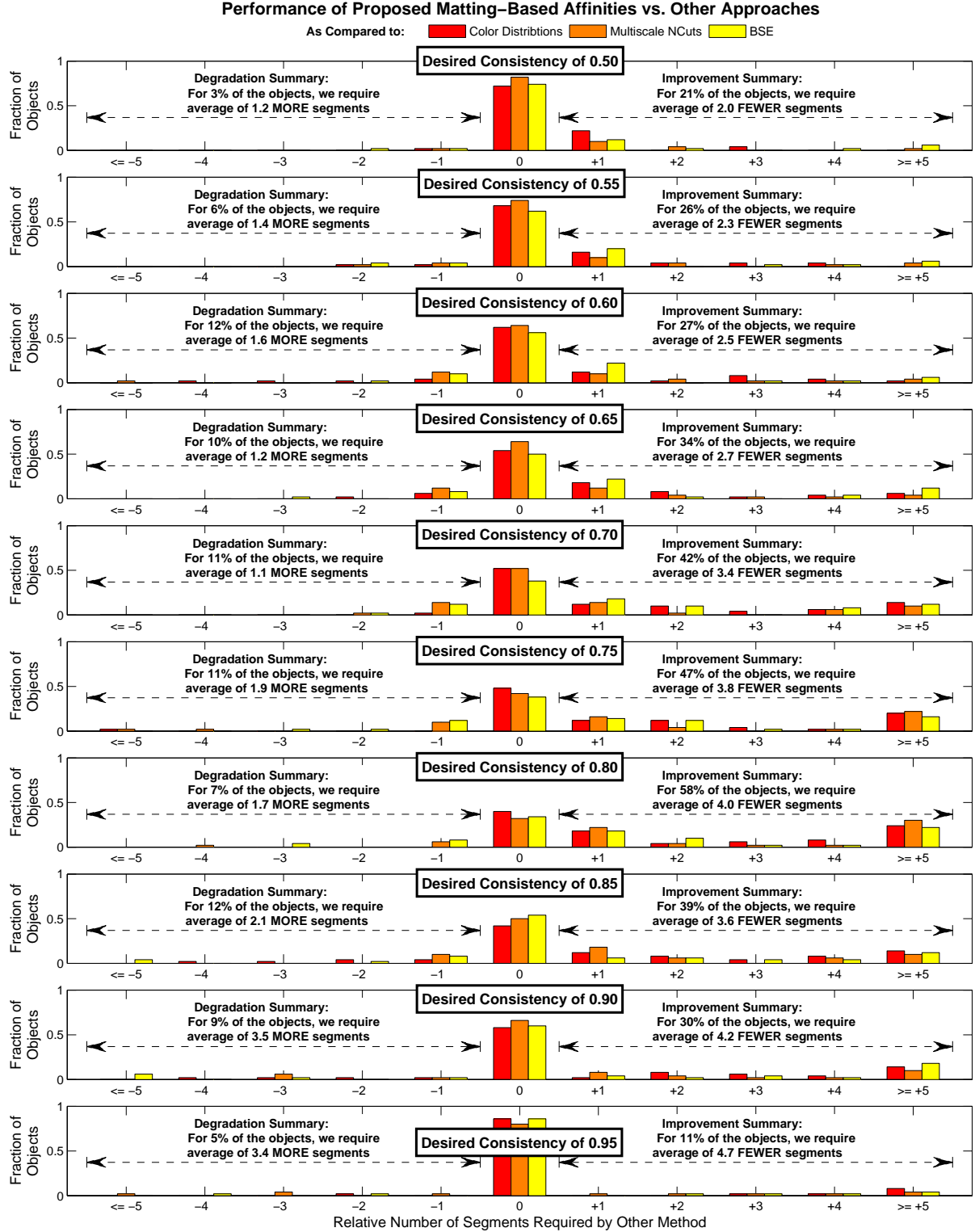


Figure 39. **Overall Performance.** Corresponding to the last figure from the main paper, we provide histograms of the *relative* number of segments required by our approach as compared to the other methods. The height of the bars corresponds to the fraction of the total number of objects for which we achieve the specified relative efficiency on the x -axis. Thus, bars at zero, in the center of the graph, correspond to cases when we perform just as well as the other approaches. Bars to the right (left) correspond to cases where we perform better (*resp.*, worse), using fewer (*resp.*, more) segments than the competition. As indicated, each plot corresponds to a different desired consistency level (increasing down the rows). Bars at the extreme left and right also include complete failure cases.