8. THE USER STUDY TO VALIDATE THE ROBOT USER

We conducted a user study with 6 participants that were familiar with computer vision but had no background knowledge about the tested image segmentation algorithms.

Every user was shown the input image with some initial scribbles (see user interface in figure 7) and an initial object outline computed with these scribbles shown as a line of marching ants. Right next to the image we displayed the same image after cutting out the foreground object using the ground truth segmentation.

The user was asked to refine the initial segmentation result such that it matches this object outline. For refining the object outline the user could place circular brushes strokes on the image (the radius of the circle was determined as in the robot user). Additionally, we automatically switched between fg and bg (red and blue brush) by using the underlying ground truth segmentation information. Hence, switching between the two brushes was not penalised.

The user could place a maximum of 20 brushes per image. If he was satisfied with the result before, he could press the "Next" button to go to the next image (see figure 7).

For each image the segmentation outline was computed with one of the three different segmentation algorithms (GCA, GC or GCS) chosen randomly for each image (to avoid any bias of the user towards an algorithm). We presented every image three times such that every algorithm was applied once per image. Parameter settings for the three algorithms were $w_i = 0$, $w_\beta = 1$, and $w_c = 0.03$ (GCS), $w_c = 0.24$ (GC), $w_c = 0.07$ (GCA). These are reasonable settings and for w_c the same as the learned values in table 2a and 2b.

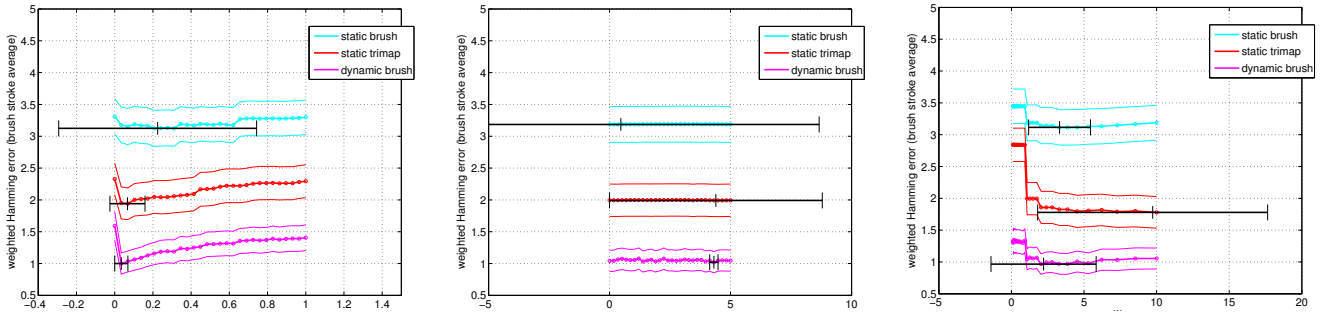
The segmentation results as plotted in figure 3 show a strong similarity between the robot and the human user in that the relative ordering of the performances of the segmentations systems is preserved.

We observed that users tend to ignore whether an erroneous region has a very large or just a large error during the first brush strokes. This means that often the user concentrated on fully correcting one part of the object until they move on with correcting a different part of the object. In the graphs, we see that the error of the human user, compared to the robot user, is higher for the first brush stroke, in contrast to the final brush strokes. After 20 brush strokes both reach quite similar error rates. Note, this is correlated with our motivation of using a *weighted* Hamming error where very big errors have less impact, since a big error (independent of how big it is) has to be corrected by the user.

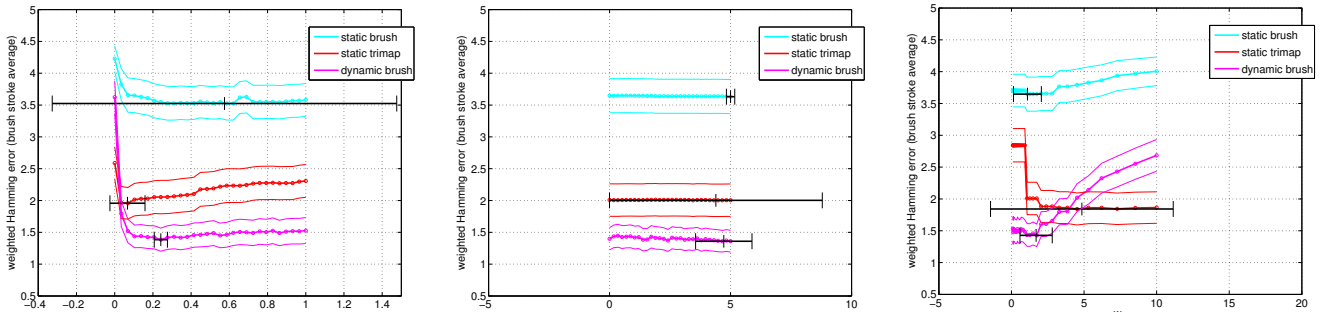
Figure 7: The user interface.

9. THE LINE-SEARCH: TRAINING

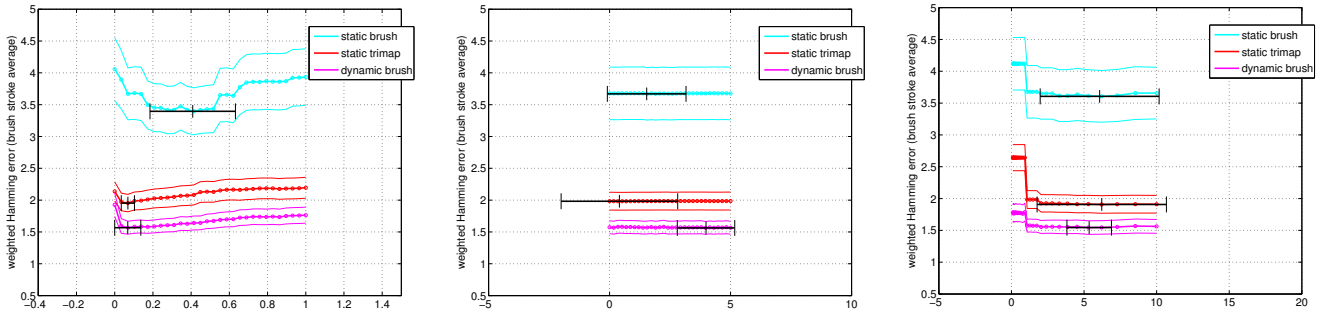
To complete the picture, we provide all plots for the line-search experiments for the systems GCA and GC, parameters w_c, w_i, w_β, and weighting functions $f(e)$.



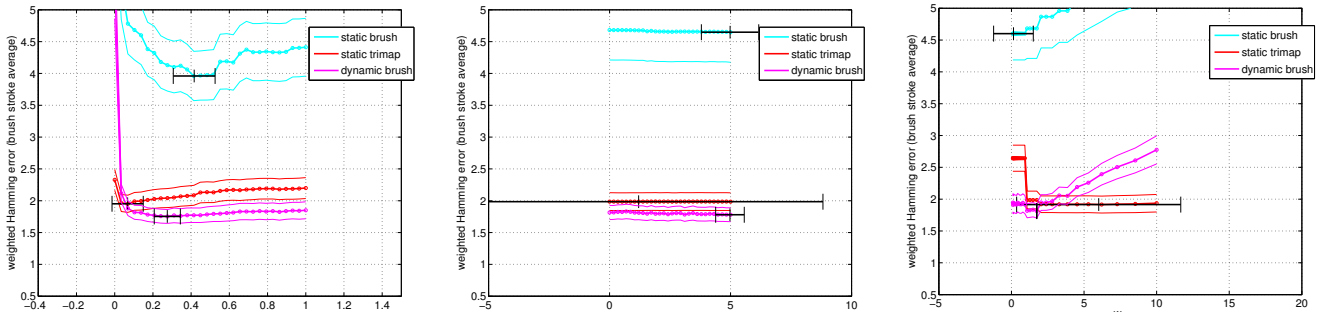
(a) GCA, contrast weight w_c (b) GCA, Ising weight w_i (c) GCA, β-scale w_β



(d) GC, contrast weight w_c (e) GC, Ising weight w_i (f) GC, β-scale w_β

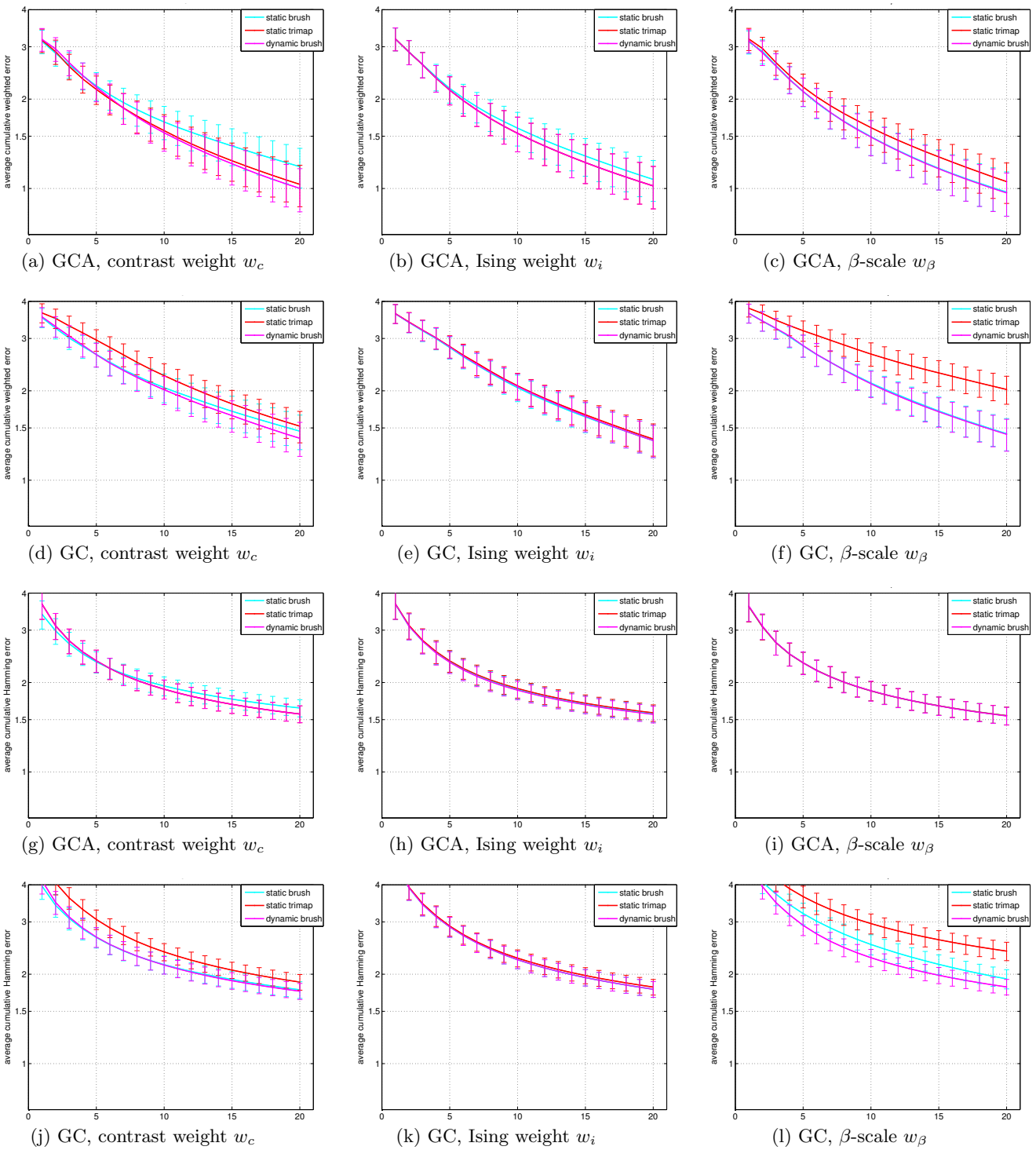


(g) GCA, contrast weight w_c (h) GCA, Ising weight w_i (i) GCA, β-scale w_β



(j) GC, contrast weight w_c (k) GC, Ising weight w_i (l) GC, β-scale w_β

Figure 8: Learning with line-search. Training results. (a-f) uses the weighted Hamming error Er where $f(e)$ is defined as in equation 2. (g-l) uses a non-weighted Hamming error, i.e. Er with $f(e) = e$



(a) GCA, contrast weight w_c (b) GCA, Ising weight w_i (c) GCA, β-scale w_β

(d) GC, contrast weight w_c (e) GC, Ising weight w_i (f) GC, β-scale w_β

(g) GCA, contrast weight w_c (h) GCA, Ising weight w_i (i) GCA, β-scale w_β

(j) GC, contrast weight w_c (k) GC, Ising weight w_i (l) GC, β-scale w_β

Figure 9: Learning with line-search. Testing results. (a-f) uses the weighted Hamming error Er where $f(e)$ is defined as in equation 2. (g-l) uses a non-weighted Hamming error, i.e. Er with $f(e) = e$