

Occlusion-Aware Object Localization, Segmentation and Pose Estimation - Supplementary Material

Samarth Brahmbhatt
<http://www.cc.gatech.edu/~sbrahmbh>

Heni Ben Amor
<http://henibenamor.weebly.com>

Henrik Christensen
<http://www.cc.gatech.edu/~hic>

School of Interactive Computing
 Georgia Institute of Technology
 Atlanta, GA USA

1 Detection of Multiple Objects

In this section we show that Algorithm 1 presented in the paper is a special case of the α -expansion procedure presented in Boykov et al. [1]. Table 1 shows the values assigned to the various terms mentioned in the table on page 7 of Boykov et al. [1] by Algorithm 1 in our paper. Since we run only one iteration of this algorithm, $\mathcal{P}_\alpha = \emptyset$ for every α . The assignment of $V_{\{p,q\}}(f_p, f_q)$ to 0 implies that the pairwise potentials are not ‘transferred’ like the unary potentials. Similarly, the higher-order potentials are not transferred too. This makes the algorithm sub-optimal. However, this simplification allows us to reuse the graph structure described in the paper, and works well in practice.

Next, we present results (excluded from the paper due to space constraints) that show the effectiveness of Algorithm 1. Since the CMU Kitchen Occlusion dataset does not have images with multiple objects, we evaluate this algorithm on two datasets of 32 images each, captured in our laboratory while a robot manipulator interacted with a pasta box. The manipulation scenario induces occlusions of varying degrees for both the pasta-box and robot manipulator. We consider two approaches: (1) Running the SD-HOP detectors separately without response-transfer, called ‘Separate’, and (2) Running the SD-HOP detectors sequentially with response-transfer i.e. Algorithm 1, called ‘Joint’. Both approaches give the same object localization performance. However, the joint approach outperforms the separate approach in terms of mean segmentation error as shown in Table 2.

Table 1: Connection between Algorithm 1 to α -expansion

$D_p(f_p)$ for $p \notin \mathcal{P}_\alpha$	$B_i(\mathbf{p})$
$D_p(\alpha)$	$F_i(\mathbf{p})$
$V_{\{p,q\}}(f_p, \alpha)$	W
$V_{\{p,q\}}(f_p, f_q)$	0
$V_{\{p,q\}}(\alpha, f_q)$	W

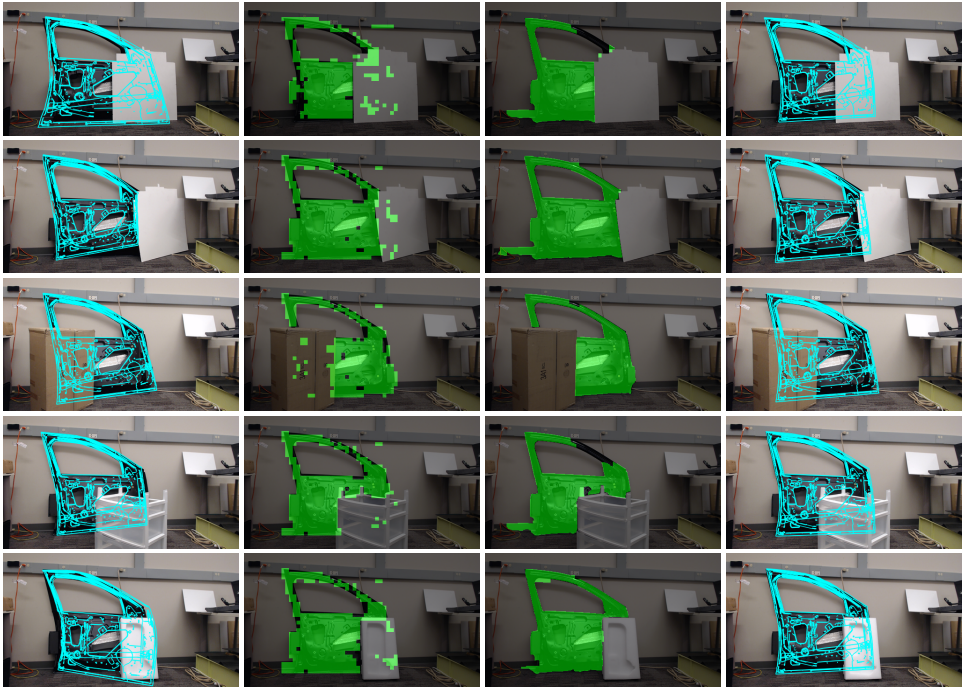


Figure 1: 3D pose estimation results on the dataset. Left: IRLS pose estimation, Center-left: Raw mask from SD-HOP, Center-right: Refined mask from SD-HOP, Right: OR-IRLS pose estimation

2 3D pose estimation results

Figure 1 shows example 3D pose estimation results from our car-door dataset.

Table 2: Mean object segmentation error for separate vs. joint detection

Dataset	Robot manipulator		Pasta-box	
	Separate	Joint	Separate	Joint
Robot-1	0.1980	0.1344	0.2989	0.2792
Robot-2	0.2728	0.2719	0.3099	0.2975

References

[1] Yuri Boykov, Olga Veksler, and Ramin Zabih. [Fast approximate energy minimization via graph cuts](#). *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11): 1222–1239, 2001. URL <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=969114>.