Tracking indistinguishable translucent objects over time using weakly supervised structured learning

Luca Fiaschi[1], Ferran Diego[1], Konstantin Gregor[1], Martin Schiegg[1], Ullrich Koethe[1], Marta Zlatic[2], and Fred A. Hamprecht[1]

[1]HCI University of Heidelberg, Germany
[2]HHMI Janelia Farm, USA

Folder contents:

- **this file: main_supplementary.pdf**
- **file: learning_curve.png and Fig. 1**
  Learning curve of the proposed algorithm. The learning curve is obtained as follows: 100 examples of difficult encounters of 2 larvae are selected: 30 are reserved to training and 70 are left for validation. Performance is measured on the validation set with the metric explained in Sec. 4.1 of the main paper (normalized number of identity switches). Error bars correspond to the standard deviation of 3 trials with a randomized subset of training examples.
- **folder: encounters_row**
  Folder containing encounters of 2 and 3 larvae from the test set. Each image shows: downsampled (c.f. Sec. 4.2, main paper) raw data and possible interpretations of the encounter ranked by energy. Each column corresponds to a different frame, the first row from the top is the raw data, other rows are different interpretations (arranged from top to bottom with increasing energy). The interpretation with lowest energy (prediction of our algorithm) is the second row from the top. Only connected components of the foreground that are in the subregion of the counting graph which corresponds to the encounter are shown. Note that, for encounters of more than 2 larvae, some larva may attach to the subgraph at time different from zero, as well as some individuals may leave before the end of the sequence. These examples are sorted dividing fully correct predictions from partial or complete mistakes, for clarity. Examples of encounters of 4 or more larvae are not shown because the space of possible interpretations is too large for this representation.
- **folder: movies_cvpr_interpolated**
  Folder containing the reconstructed tracking for a set of movies. In each movie we start tracking each object as it becomes isolated for the first

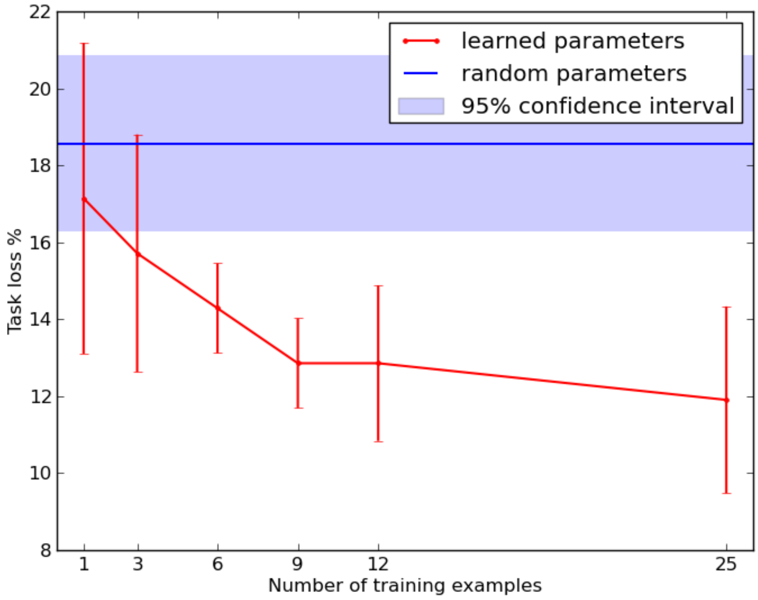

Figure 1: Learning curve of the proposed algorithm over a validation set with 70 among difficult encounters of 2 larvae. Average performance over three independent runs with variable number of training examples (red), as well as average performance with random positive weights (blue). Red error bars correspond to the standard deviation over 3 independent trials while the blue region represents a 95% confidence interval (1.95 times standard deviation). Our approach shows already good predictive power by using random positive weights (initialization of the CCCP procedure), however our learning procedure greatly improves performance with very few training examples and minimal human effort.

time. Tracklets for all isolated individuals are generated by simple nearest neighbor tracking and terminated upon occluded regions (where objects are merged due to undersegmentation mistakes or overlap). We then resolve all occluded regions with our main proposal. This matches each track that enters the occlusion event with one other that leave the occlusion and assigns to both tracklets the same color coded ID. For visualization purpose, we also linearly interpolate the position of the objects during occlusion. Note that this is a rough approximation that is only used for visualization: a linear trajectory does not follow precisely the larvae during prolonged occlusion as they may follow an erratic motion. A more precise trajectory could be obtained by computing the centroid of the colored mass distributions as inferred by our proposal. Also note that objects are tracked only within a control region of 50 pixels from the borders of the image. Objects that exit and the re-enter this area will get assigned a new ID.

$N_L^c = 2$	2 ± 3
$N_L^c = 3$	57 ± 57
$N_L^c \geq 4$	573 ± 412

Table 1: Average running time in min per encounter type

Running time

The breakdown of the running time per encounter type is given in table 1. This table shows that our method can handle interactions with 2-3 objects while only few interactions with $N_L^c \geq 3$ should be present. The average running time per movie is $3.43 \pm 5.4hr$ ($\max 25hr, \min 0.20hr$). The high variance is due to outlier movies where many clusters $N_L^c \geq 3$ are present.