A Synthetic Video Results

A.0.1 Tracking through Occlusion

We first construct simple synthetic videos to demonstrate tracking objects through occlusion (via the learned object models). The first three synthetic videos are each 200 frames and contain a number of colored squares of different (and potentially time-varying) sizes moving at varied speeds and trajectories over a black background. The synthetic videos contain instances of occlusion (of one or more squares) and objects with time-varying appearances, as these notoriously decrease the accuracy of detection and tracking.

The first two videos contain a red and a blue square, which begin at opposite sides of the scene at frame $f = 1$ and occlude (blue over red) in the center of the frame at $f = 100$ (Figure 3(a)). In the first video, both squares then continue in the same direction, and in the second they reverse directions and end at their initial starting positions. Data extraction yields identical spatial features in both videos; hence, successful tracking depends fully on the model of object color. The third video is the same as the first, except a third green square is added (Figure 3(b)). This square is 50 pixels/side at $f = 1$, travels across the frame while linearly shrinking to 10 pixels/side at $f = 100$ (where it is occluded), and grows back to 50 pixels/side to end at the opposite side of the frame.

The DDPMO successfully tracked the squares via the SMC inference algorithm in all three occlusion scenarios and correctly modeled the time-varying square in the third video (Figure 4 (a)-(c)).

A.0.2 Learning Diverse Object Models

In this experiment, we constructed a video from images of four physically different objects (in shape, scale, and appearance)—a truck, human, cat, and beetle—which randomly walk around the scene. The DDPMO fully tracks all four objects, and accurately learns a diverse set of object models. Inferred results for each object are shown in Figure 4(d)-(g).
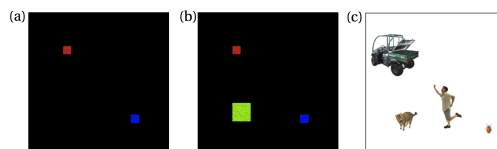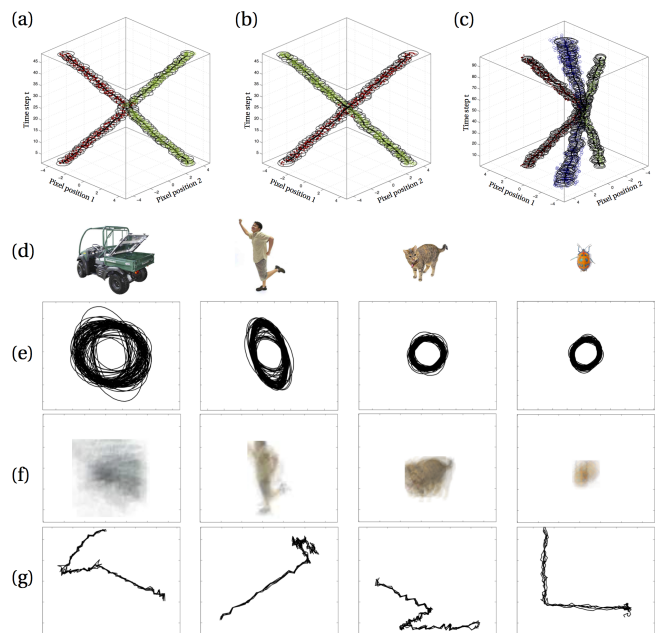

Figure 7: Inference results are plotted for synthetic experiments 1-3 (a-c). We show the four objects in synthetic experiment 4 (d), samples from the posterior over their spatial parameters (e), averages over extracted color observations for each inferred object (f), and samples from the posterior over the objects' tracks (g).

Figure 6: Still video frames from synthetic experiments 1 and 2 (a), 3 (b), and 4 (c).