Supplementary Material For rCRF: Recursive Belief Estimation over CRFs in RGB-D Activity Videos

Ozan Sener
School of Electrical & Computer Eng.
Cornell University

Ashutosh Saxena
Department of Computer Science
Cornell University

I. Re-defining the belief

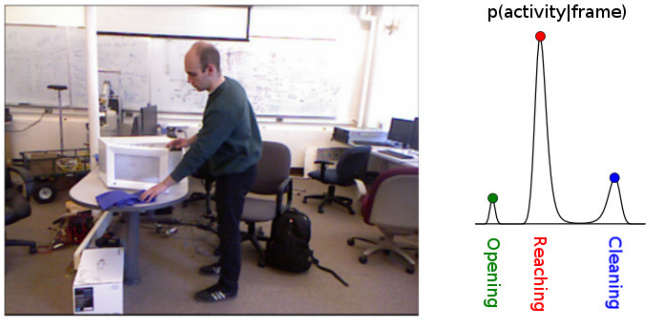

Fig. 1: **Typically there are only a few plausible explanations per scene**. For this scene, only *cleaning the table*, *reaching to the cloth*, *reaching to the microwave*, *opening the microwav* are plausible.

CRF-likelihood over a natural scene concentrates on a few plausible samples!

We compute the belief for only the plausible samples as;

$$\text{approx\_bel}^t(\mathbf{y}) = \begin{cases} \frac{bel^t(\mathbf{y})}{\sum_{\mathbf{y}' \in \mathbf{Y}^t} bel^t(\mathbf{y}')} & \text{if } \mathbf{y} \in \mathbf{Y}^t \\ 0 & \text{o.w.} \end{cases} \quad (1)$$

where the set of plausible solutions at time t is $\mathbf{Y}^t = \mathbf{y}^{t,1}, \dots, \mathbf{y}^{t,M}$.

II. How to find these samples?

Observations:

- These samples have high likelihood since they are plausible
- These samples are diverse because there are likely solutions around MAP solutions (eg. small perturbation of the MAP solution)

Formulation: We solve the following optimization problem in order to get the plausible solutions (we use hamming distance as Δ in our experiments);

$$\begin{aligned} \mathbf{y}^{t,i} &= \arg\max_{\mathbf{y}} bel^t(\mathbf{y}) \\ &s.t.\ \Delta(\mathbf{y}, \mathbf{y}^{t,j}) \geq \delta \quad \forall\, j < i \end{aligned} \quad (2)$$