# 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. Recursive Conditional Random Field

*Definition 1:* Let $\mathcal{G}^t = (V^t, E^t)$ be set of graphs indexed by the temporal variable $t$ and $\mathbf{y}^t$ is indexed by the vertices of $\mathcal{G}^t$ as $\mathbf{y}^t = (y_v^t)_{v \in V^t}$. Then, $(\mathbf{x}^{1 \dots T}, \mathbf{y}^{1 \dots T})$ is a ***Recursive Conditional Random Field*** with dynamics $p_v(\cdot|\cdot)$ when

1) For each $t$, $(\mathbf{y}^t, \mathbf{x}^t)$ is a CRF over $\mathcal{G}^t = (V^t, E^t)$
2) $p(\mathbf{y}^t|\mathbf{y}^1, \dots, \mathbf{y}^{t-1}) = p(\mathbf{y}^t|\mathbf{y}^{t-1}) \quad \forall t$ (Markov)
3) $p(\mathbf{x}^t|\mathbf{y}^1, \dots, \mathbf{y}^t, \mathbf{x}^1, \dots, \mathbf{x}^{t-1}) = p(\mathbf{x}^t|\mathbf{y}^t) \quad \forall t$
4) $p(\mathbf{y}^t = \mathbf{y}|\mathbf{y}^{t-1} = \mathbf{y}') = p_v(\mathbf{y}|\mathbf{y}')$ (stationarity)

■

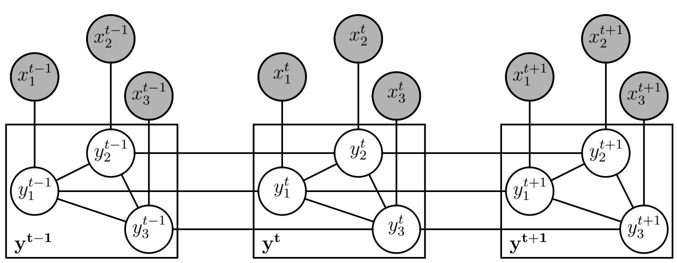

*Fig. 1:* The graphical model, we use within the rCRF, is a temporal CRF with additional constraints.

We are interested in the belief over state variables at $t$ as $bel^t(\mathbf{y}) = p(\mathbf{y}^t = \mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_T)$. Following the independence properties, it can be defined as.

$$bel^t(\mathbf{y}) \propto \underbrace{p(\mathbf{y}^t = \mathbf{y}|\mathbf{x}^1, \dots, \mathbf{x}^t)}_{\alpha^t(\mathbf{y})} \underbrace{p(\mathbf{x}^{t+1}, \dots, \mathbf{x}^T|\mathbf{y}^t = \mathbf{y})}_{\beta^t(\mathbf{y})} \tag{1}$$

with recursive definition of $\alpha^t$ and $\beta^t$ as;

$$\begin{aligned} \alpha^t(\mathbf{y}^t) &= p(\mathbf{x}^t|\mathbf{y}^t) \sum_{\mathbf{y}^{t-1}} \alpha^{t-1}(\mathbf{y}^{t-1}) p(\mathbf{y}^t|\mathbf{y}^{t-1}) \\ \beta^t(\mathbf{y}^t) &= \sum_{\mathbf{y}^{t+1}} p(\mathbf{x}^{t+1}|\mathbf{y}^{t+1}) \beta^{t+1}(\mathbf{y}^{t+1}) p(\mathbf{y}^{t+1}|\mathbf{y}^t) \end{aligned} \tag{2}$$

with initializations $\alpha^1(\mathbf{y}^1) = p(\mathbf{x}^1|\mathbf{y}^1)$ and $\beta^T(\mathbf{y}^T) = 1$.

### A. Challanges:

- CRF is modelling $p(\mathbf{y}^t|\mathbf{x}^t)$ instead of $p(\mathbf{x}^t|\mathbf{y}^t)$.
- Messages requires a summation over the entire output space which has an exponential dimension.

*1) From $p(\mathbf{y}^t|\mathbf{x}^t)$ to $p(\mathbf{x}^t|\mathbf{y}^t)$:* Since $(\mathbf{x}^t, \mathbf{y}^t)$ is a CRF, the posterior of the label given the observation follows;

$$p(\mathbf{y}^t|\mathbf{x}^t) \propto \exp\left( \sum_{i \in V^t} \theta_{x_i^t}(y_i^t) + \sum_{i,j \in E^t} \theta_{x_i^t, x_j^t}(y_i^t, y_j^t) \right) \tag{3}$$

where $\theta$ is the energy function defined over the node set $v \in V^t$ as $\theta_v$ and over the edge set $(u,v) \in E^t$ as $\theta_{u,v}$.

$$p(\mathbf{y}^t) = \sum_{\mathbf{x}^t} \exp\left( \sum_{i \in V^t} \theta_{x_i^t}(y_i^t) + \sum_{i,j \in E^t} \theta_{x_i^t, x_j^t}(y_i^t, y_j^t) \right) p(\mathbf{x}^t) \tag{4}$$

For tractability, we approximate the $p(\mathbf{y}^t)$ with its lower bound after applying the Jensen inequality as;

$$p(\mathbf{y}^t) \approx \exp( \sum_{i \in V^t} \underbrace{\sum_{\mathbf{x}^t} \theta_{x_i^t}(y_i^t) p(\mathbf{x}^t)}_{\tilde{\theta}(y_i^t)} + \sum_{i,j \in E^t} \underbrace{\sum_{\mathbf{x}^t} \theta_{x_i^t, x_j^t}(y_i^t, y_j^t) p(\mathbf{x}^t)}_{\tilde{\theta}(y_i^t, y_j^t)}) \tag{5}$$

We then estimate the inner summations $\tilde{\theta}(\cdot)$ using Monte Carlo as $\tilde{\theta}(\cdot) = \frac{1}{N} \sum_{i=1}^{N} \theta_{\mathbf{x}^{(i)}}(\cdot)$ where $\mathbf{x}^{(i)}$ is the $i^{th}$ training sample. The final observation likelihood is: $p(\mathbf{x}^t|\mathbf{y}^t) \propto$

$$\exp\left( \sum_{i \in V^t} \theta_{x_i^t}(y_i^t) - \tilde{\theta}(y_i^t) + \sum_{i,j \in E^t} \theta_{x_i^t, x_j^t}(y_i^t, y_j^t) - \tilde{\theta}(y_i^t, y_j^t) \right) \tag{6}$$

*2) Belief is a CRF:* The belief function can be computed as follows:

$$\begin{aligned} bel(\mathbf{y}^t) \propto \exp \Bigg[ & \sum_{i,j \in E^t} \left( \theta_{x_i^t, x_j^t}(y_i^t, y_j^t) - \tilde{\theta}(y_i^t, y_j^t) \right) \\ & \sum_{i \in V^t} \Bigg( \theta_{x_i^t}(y_i^t) - \tilde{\theta}(y_i^t) + \sum_{\mathbf{y}^{t-1}} \alpha^{t-1}(\mathbf{y}^{t-1}) \log p(y_i^t|y_i^{t-1}) \\ & + \frac{1}{\gamma} \sum_{\mathbf{y}^{t+1}} \beta^{t+1}(\mathbf{y}^{t+1}) p(\mathbf{x}^{t+1}|\mathbf{y}^{t+1}) \log p(y_i^{t+1}|y_i^t) \Bigg) \Bigg] \end{aligned} \tag{7}$$

This function is graph decomposable hence the posterior belief in rCRF is also a CRF. Hence, we can use the existing learning, inference and structured diversity algorithms developed for CRFs by just modifying the energy functions.