UCLA

Human Parsing using Stochastic And-Or Grammars and Rich Appearances

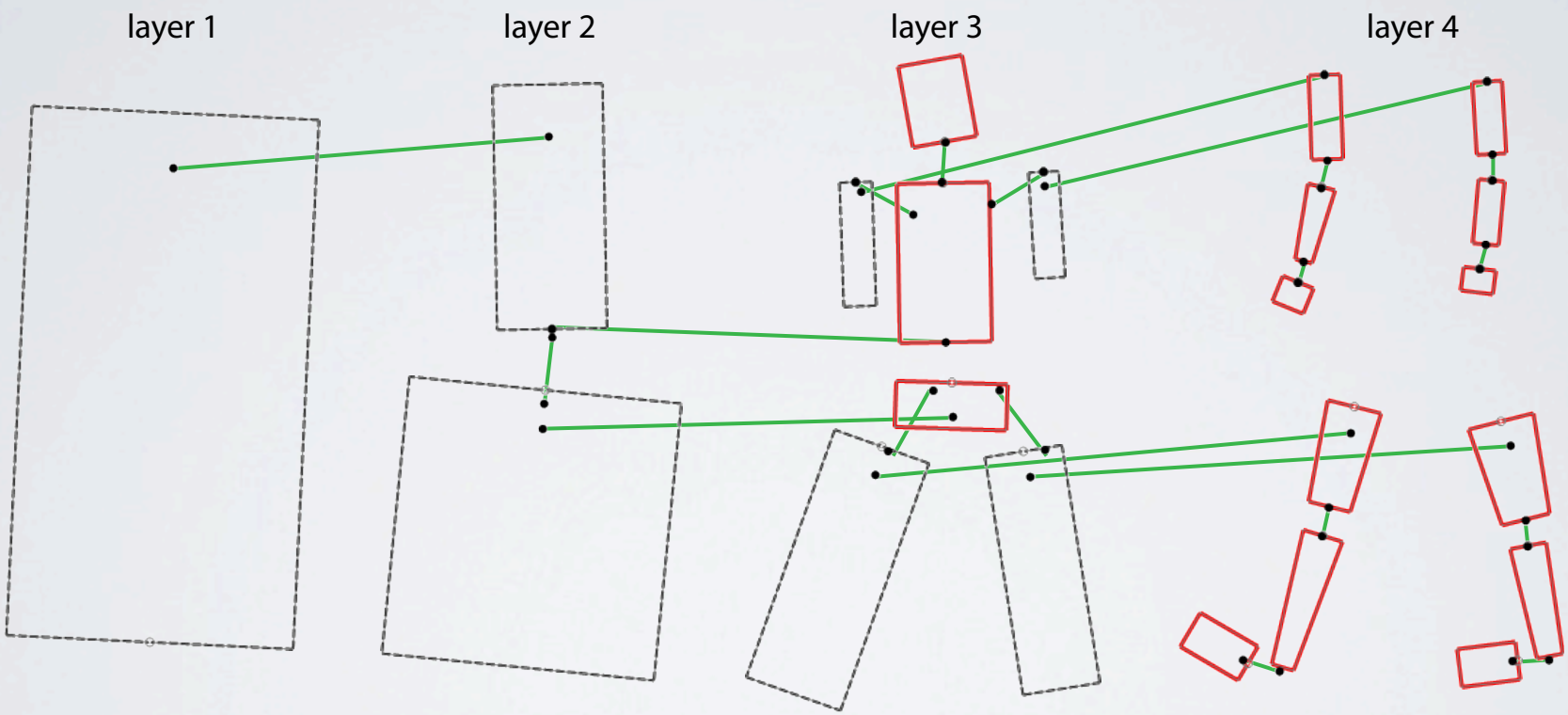
Brandon Rothrock and Song-Chun Zhu
UCLA Dept. of Computer Science

ICCV 2011

MOTIVATIONS

- Rich appearance representation
- Semantic parts
- Part sharing
- Reconfigurable parts

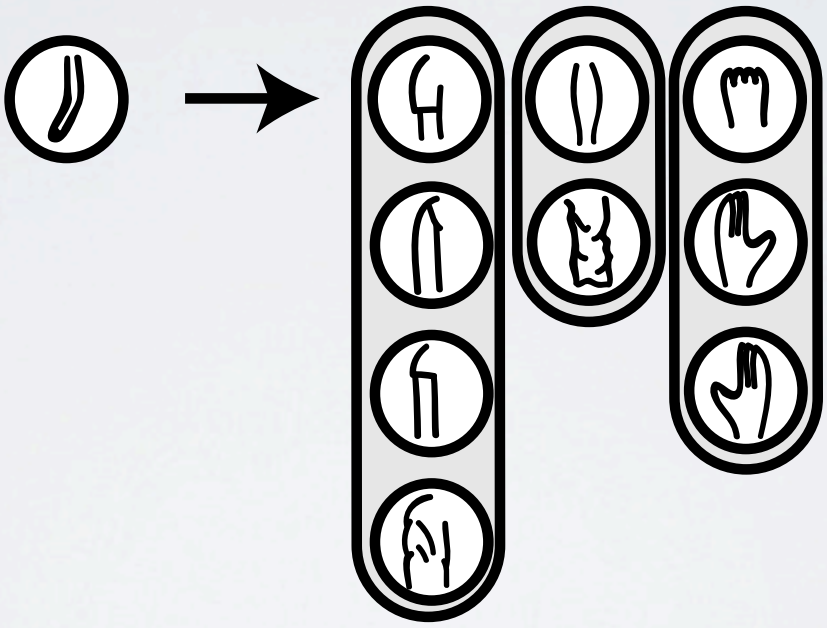
BODY REPRESENTATION



COMPOSITION RULES

and-node

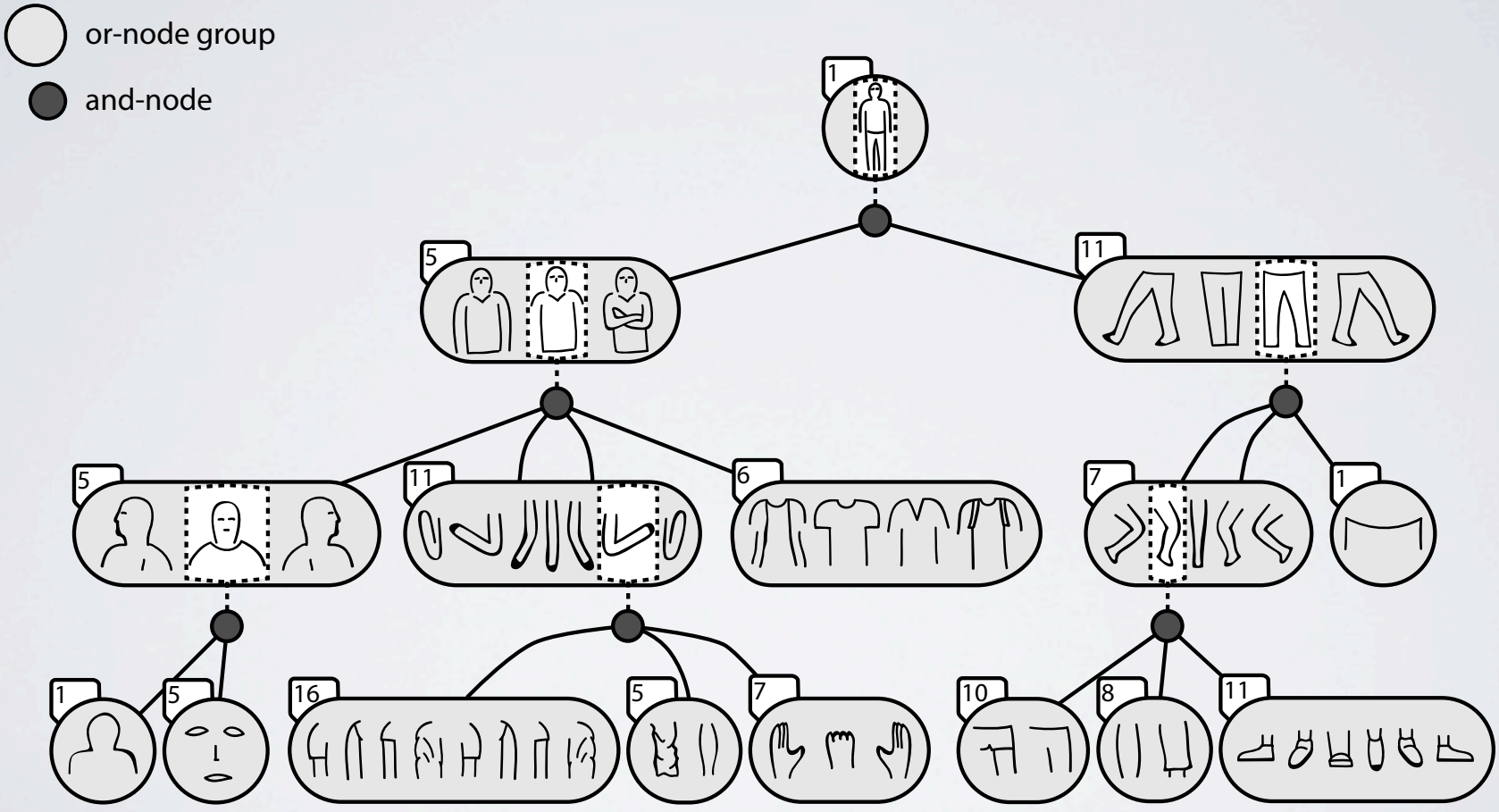
or-nodes



Part form

Part types

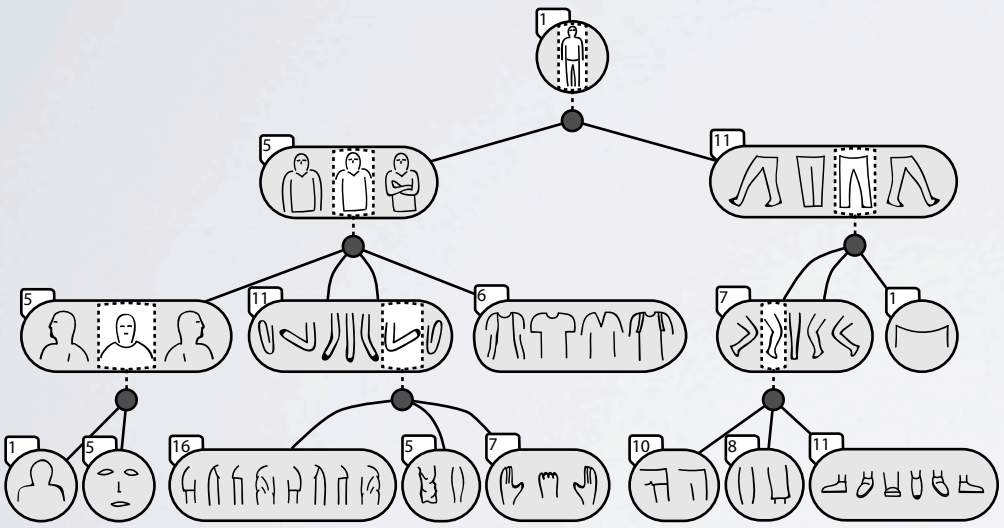
AND-OR GRAPH GRAMMAR



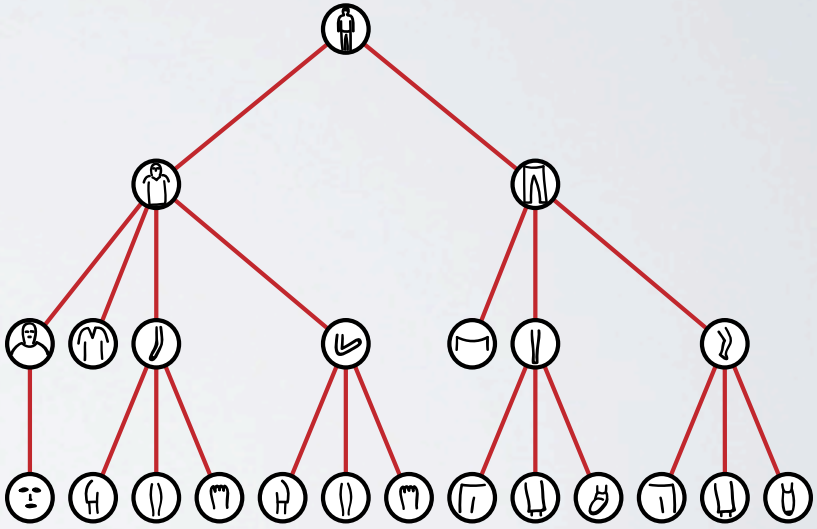
AND-OR GRAPH GRAMMAR

or-node group

and-node



Grammar

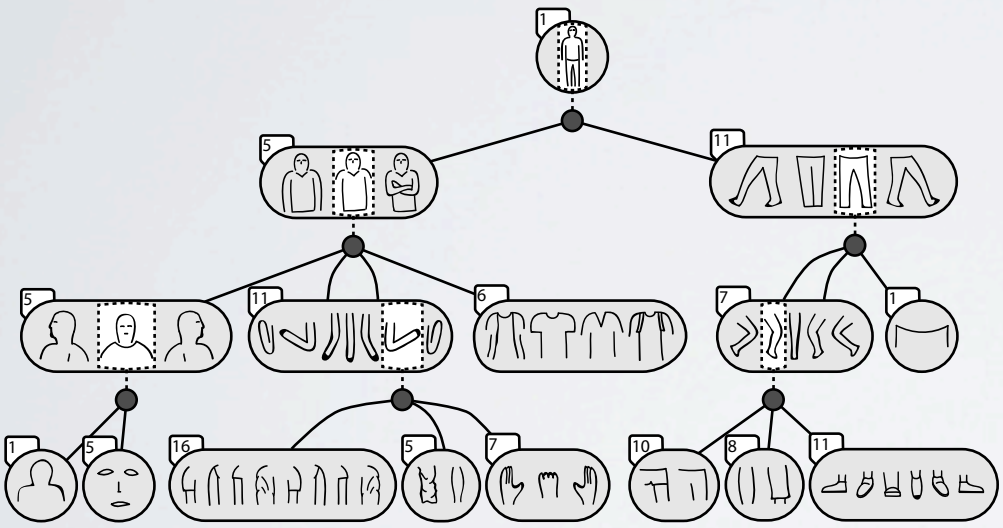


Derivation tree

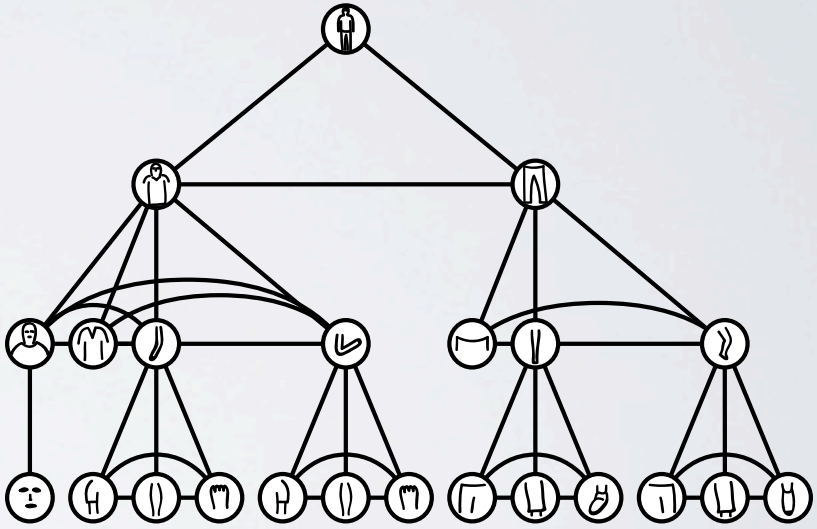
AND-OR GRAPH GRAMMAR

or-node group

and-node



Grammar



Parse Graph

PROBABILITY MODEL

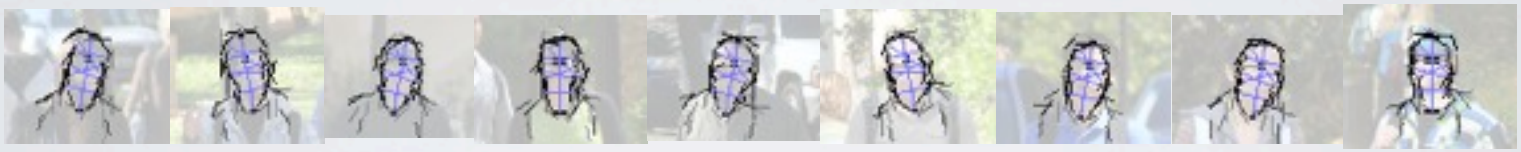
$$p(pg|\mathbf{I};\Theta) \propto p(\mathbf{I}|pg;\Theta)p(pg;\Theta)$$

$$= \frac{1}{Z(\Theta)} \exp\{-\underbrace{E_a(pg|\mathbf{I})}_{\text{appearance}} - \underbrace{E_d(pg)}_{\substack{\text{derivation}\\ \text{(part forms)}}} - \underbrace{E_g(pg)}_{\text{geometry}}\}$$

APPEARANCE MODEL

- Adapted from Active Basis model

(IJCV 2010) Learning Active Basis Model for Object Detection and Recognition (Wu, Si, Gong, Zhu)
(CVPR 2009) Learning Mixed Image Templates for Object Recognition (Si, Gong, Wu, Zhu)



Template learned by pursuit

$$q \rightarrow p_1 \rightarrow \ldots \rightarrow p_k = p(\mathbf{I}_{\Lambda_{v_i}}|v_i) \sim f$$

Model derived by minimizing KL

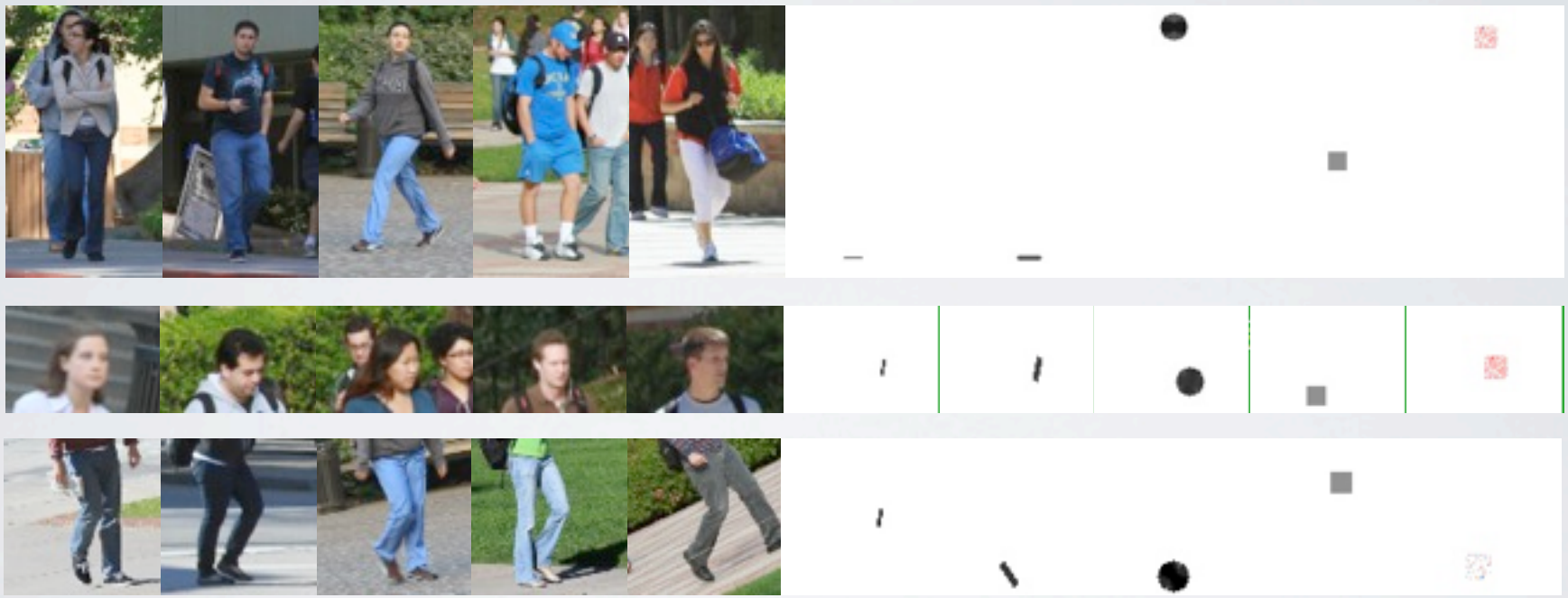
$$p_k^* = \arg\min KL(p_k||p_{k-1}) \quad s.t.\ E_{p_k}[r_k] = E_f[r_k]$$

Resulting in the following form:

$$p_k(\mathbf{I}_\Lambda) = q(\mathbf{I}_\Lambda) \exp\left\{\sum_{i=1}^{k} \lambda_i r_i(\mathbf{I}_{\Lambda_i}) - \log z_i\right\}$$

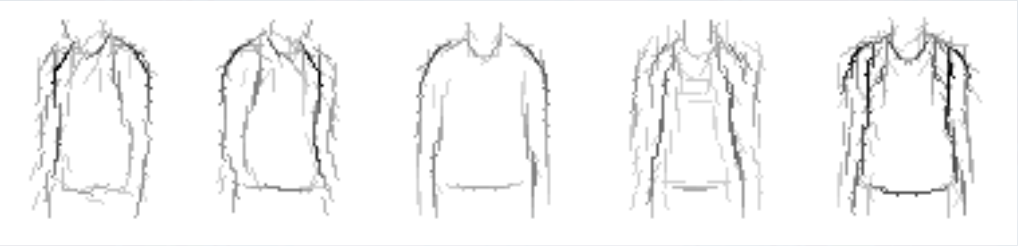
APPEARANCE MODEL

APPEARANCE MODEL



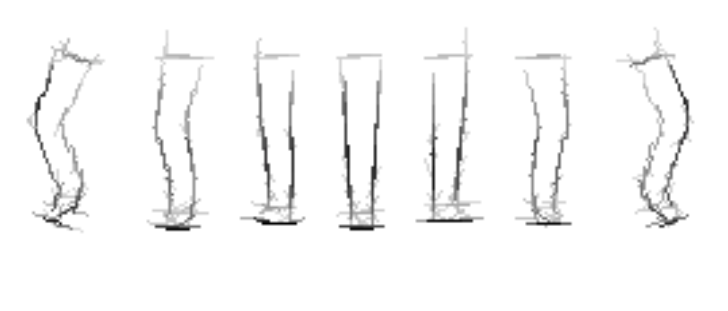
APPEARANCE MODEL

upper body



lower body

leg



upper arm

APPEARANCE MODEL

$$p(\mathbf{I}|pg) = q(\mathbf{I}_{\overline{\Lambda_{pg}}}) \prod_{v_i \in pg} p(\mathbf{I}_{\Lambda_{v_i}}|v_i)$$

$$= q(\mathbf{I}) \prod_{v_i \in pg} \exp\left\{ \sum_{j=1}^{k} \lambda_j r_j(\mathbf{I}_{\Lambda_j}) - \log z_j \right\}$$

DERIVATION MODEL

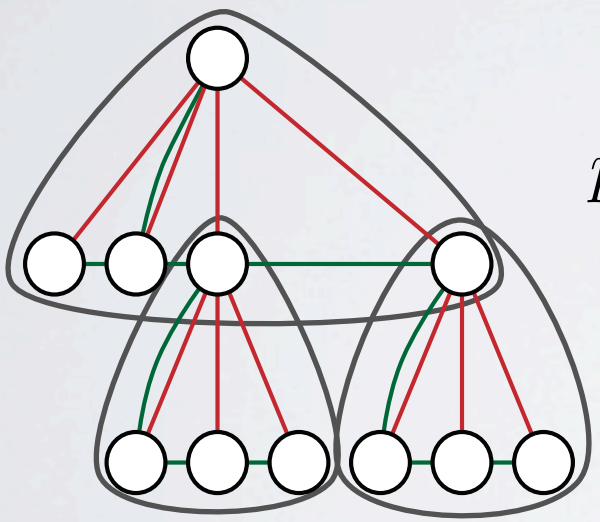
SCFG case is indifferent to neighboring forms

form index

$$p(pg) = \prod_{v_i \in V_{pg}} p(\omega(v_i))$$

DERIVATION MODEL

Allow child forms to depend on parent form



children

$$p_d(pg) = \prod_{v_i \in V_{pg}} p(\omega(C(v_i))|\omega(v_i))$$

$$\approx \prod_{v_i \in V_{pg}} \prod_{(jk) \in E(v_i)} p(\omega(v_j), \omega(v_k)|\omega(v_i))$$

GEOMETRY MODEL

$$p(v_j, v_k | \omega(v_i)) = \mathcal{N}(T_{kj}(\mathbf{x}_k) - T_{jk}(\mathbf{x}_j), 0, \Sigma_{ij})$$

$$p_g(pg) \propto \prod_{v_i \in V_{pg}} \prod_{(jk) \in E(v_i)} p(v_j, v_k | \omega(v_i))$$

SAMPLING

constrained samples

unconstrained samples

INFERENCE

Recursive scoring function

$$s(\rho|\mathbf{I}) = \overbrace{s_a(\rho|\mathbf{I})}^{appearance} + \overbrace{s_g(\rho)}^{geometry} + \overbrace{s_d(\rho)}^{derivation} + \overbrace{\sum_{\rho_i \in C(\rho)} s(\rho_i|\mathbf{I})}^{children}$$

$$s_a(\rho|\mathbf{I}) = \log p(\mathbf{I}_{\Lambda_\rho}|\rho) = \sum_{i=1}^{n} [\lambda_i r_i(\mathbf{I}_{\Lambda_i}) - \log z_i]$$

$$s_d(\rho) = \log \frac{\prod_{(ik)\in E(\rho)} p(\omega(\rho_j), \omega(\rho_k)|\omega(\rho))}{\prod_{\rho_j \in V(\rho)} p(\omega(\rho_j)|\omega(\rho))^{d(\rho_j - 1)}}$$

$$s_g(\rho) = \sum_{(ij)\in E(v)} \log p(\rho_i, \rho_j)$$

INFERENCE

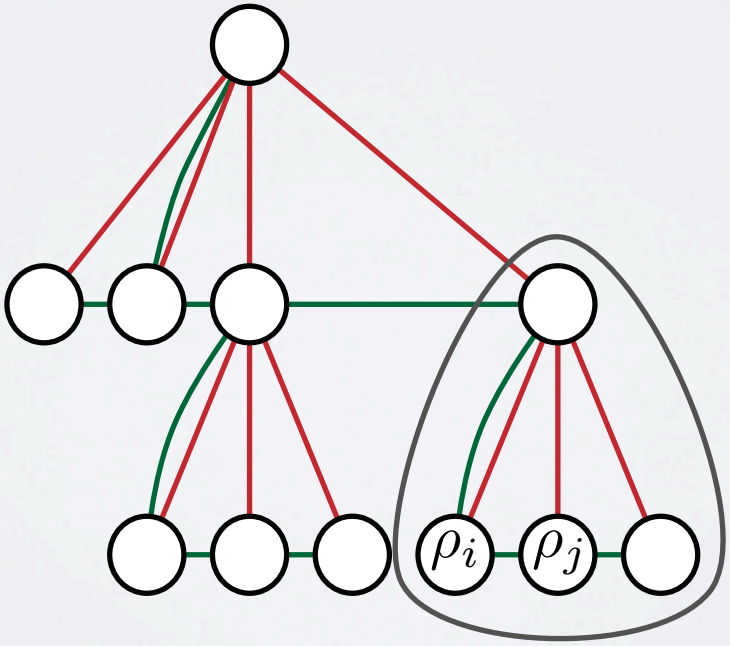
$$s^*(\rho, \mathbf{I}) = s(\rho, \mathbf{I}) + \max_{\mathbf{x}_i, \omega(\rho_i) \forall \rho_i \in C(\rho)} \left(\sum_{\rho_i \in C(\rho)} s^*(\rho_i, \mathbf{I}) \right)$$

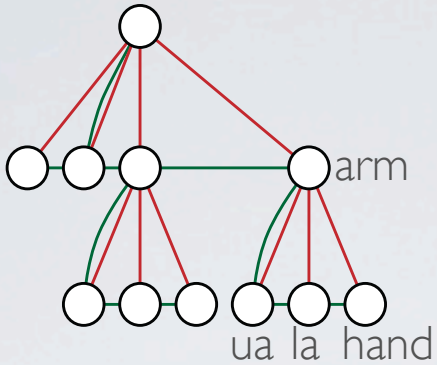
$$\rho_{\mathcal{S}}^{MAP} = \max_{\mathbf{x}, \omega(\rho_{\mathcal{S}})} s^*(\rho_{\mathcal{S}}, \mathbf{I})$$

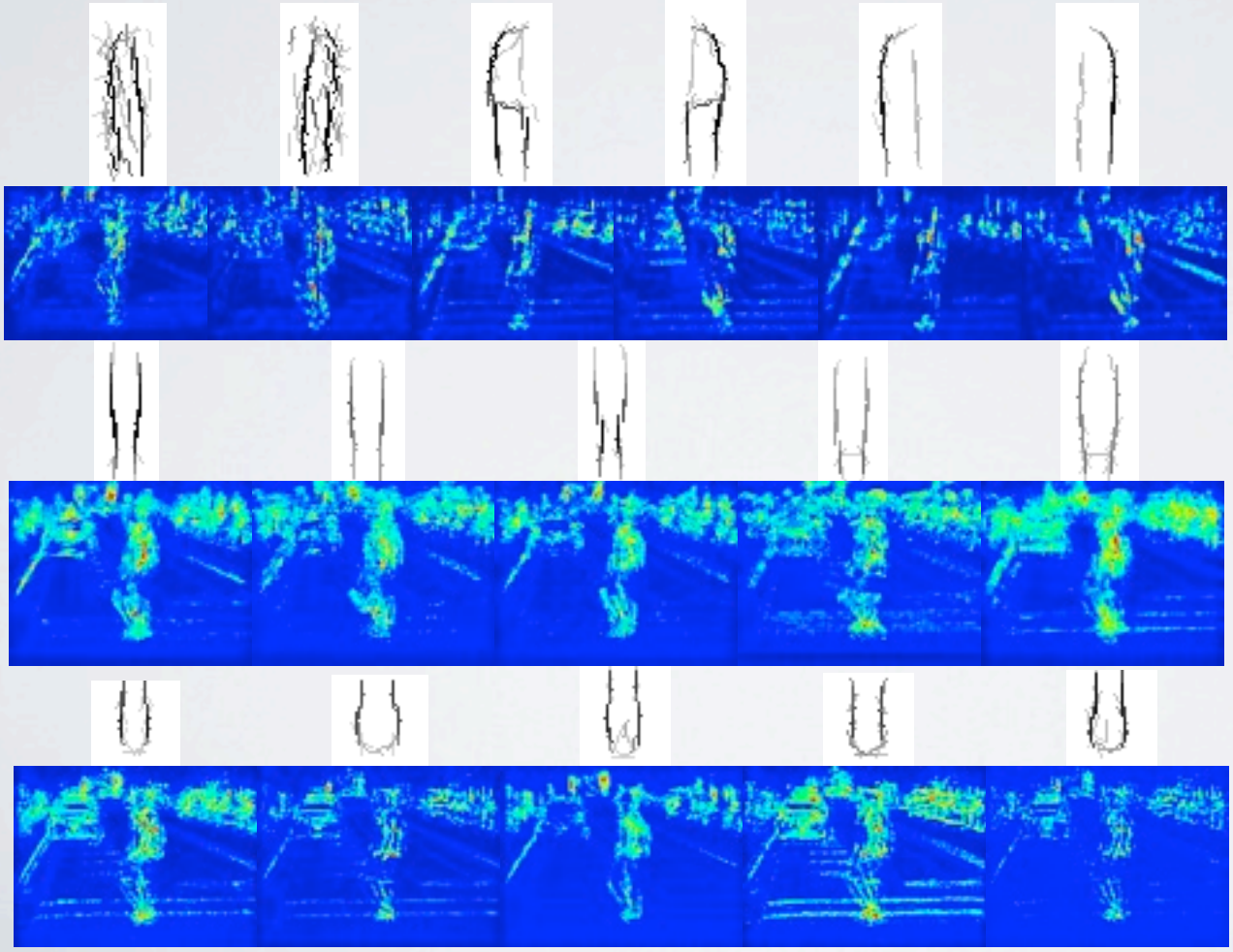
INFERENCE

$$\max_{\mathbf{x}_i, \omega(\rho_i) \forall \rho_i \in \rho_C} \left(\sum_{\rho_i \in \rho_C} s^*(\rho_i, \mathbf{I}) \right) =$$

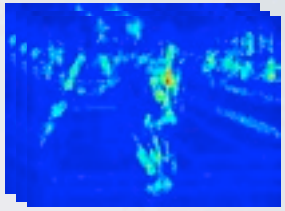
$$B_{\rho_j}(\rho_i) = \max_{\omega(\rho_j)} \left(\max_{\mathbf{x}_j} \left(s^*(\rho_j) + \log \frac{p(\omega(\rho_i), \omega(\rho_j))}{p(\omega(\rho_j))^{d(p_j)-1}} + \log p(\rho_i, \rho_j) + \sum_{\rho_k \in C(\rho_j)} B_{\rho_k}(\rho_j) \right) \right)$$



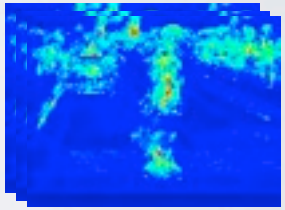




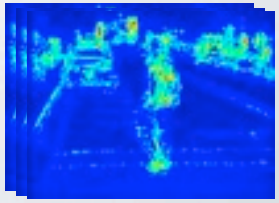
ua

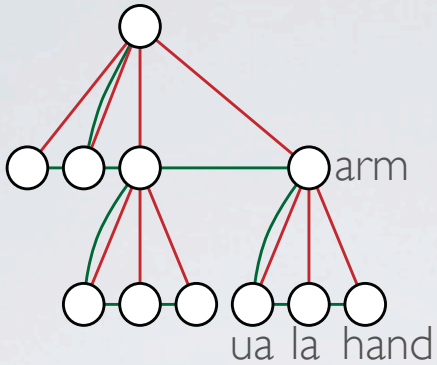


la

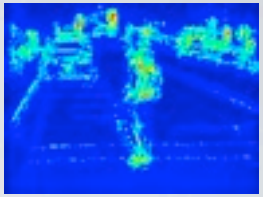


hand

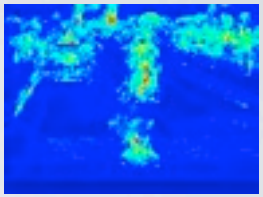




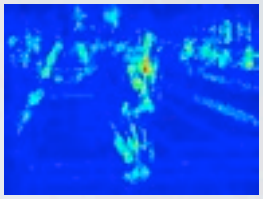
hand



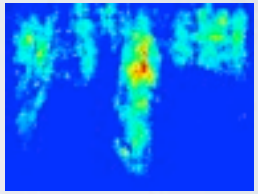
la



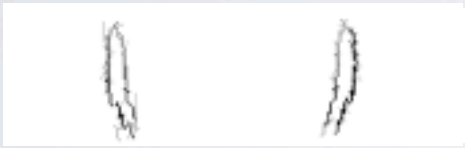
ua

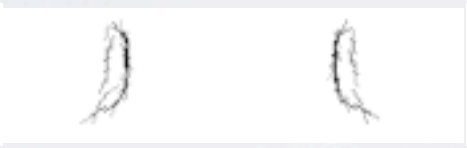


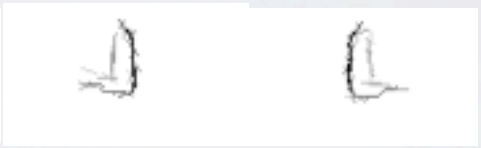
arm

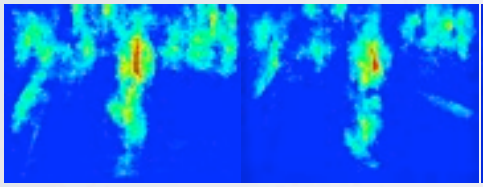


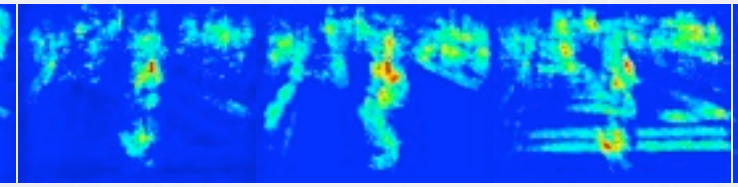
arm

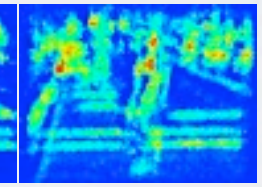




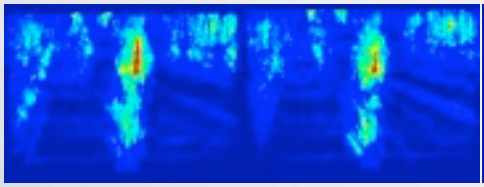


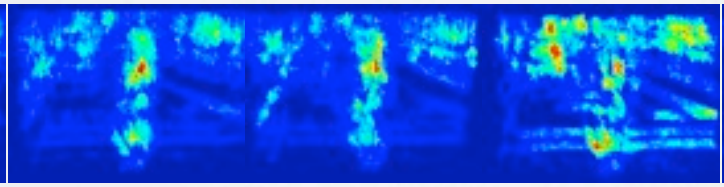


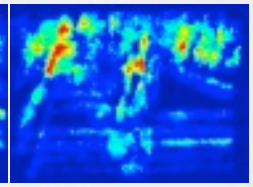


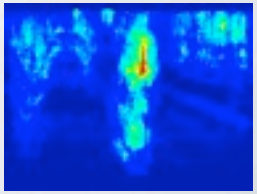


arm+children

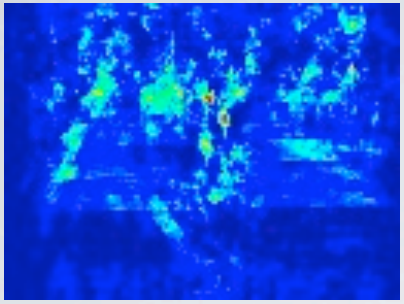




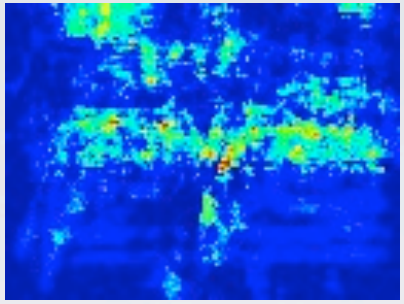




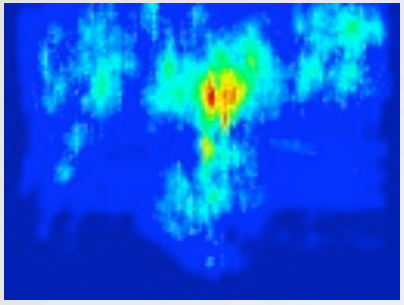
ubody



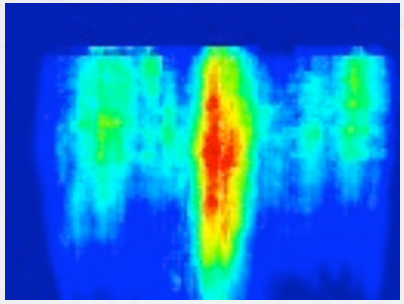
fullbody



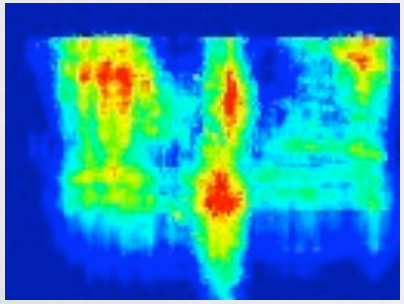
ubody+children

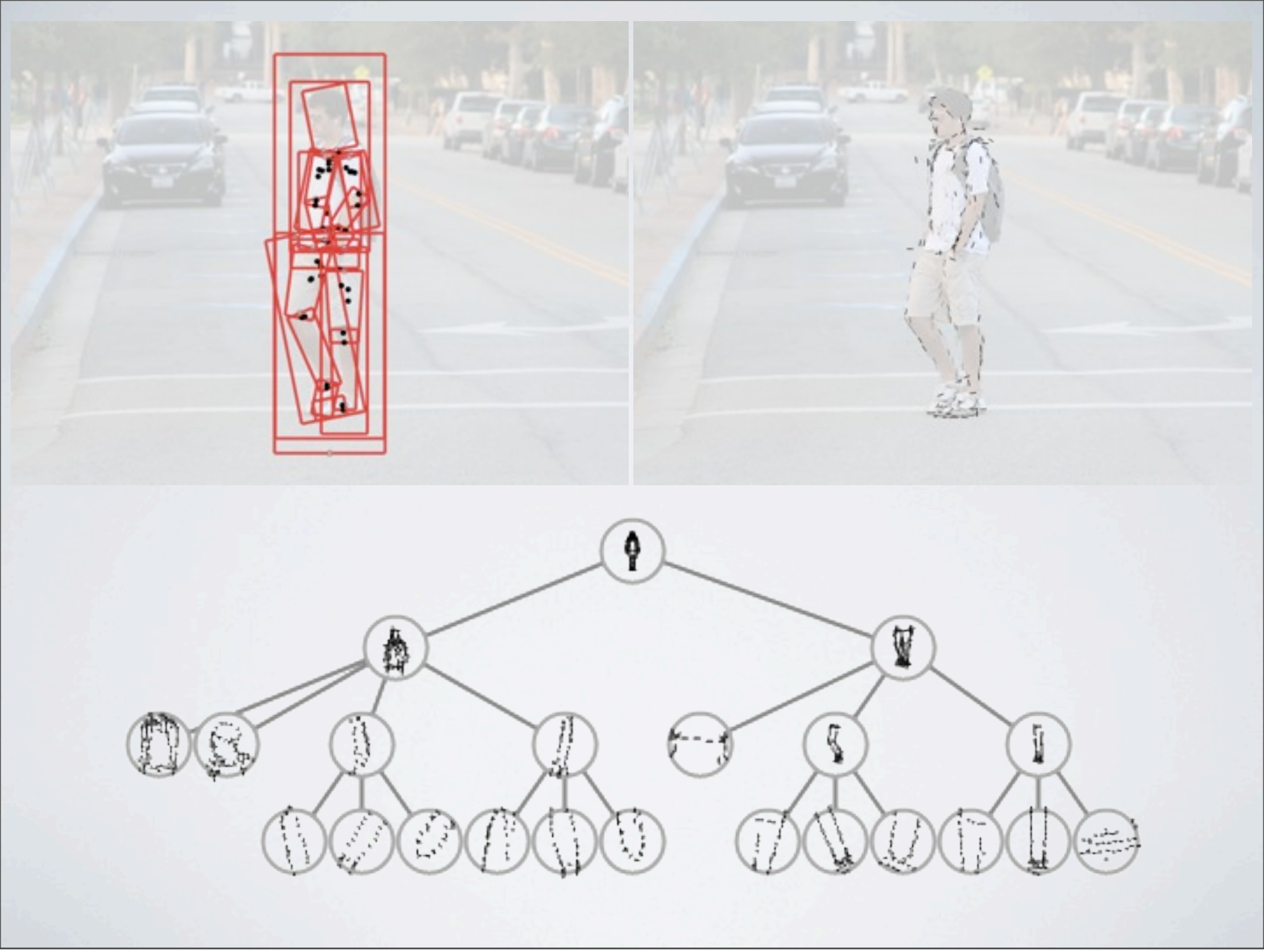


fullbody+children



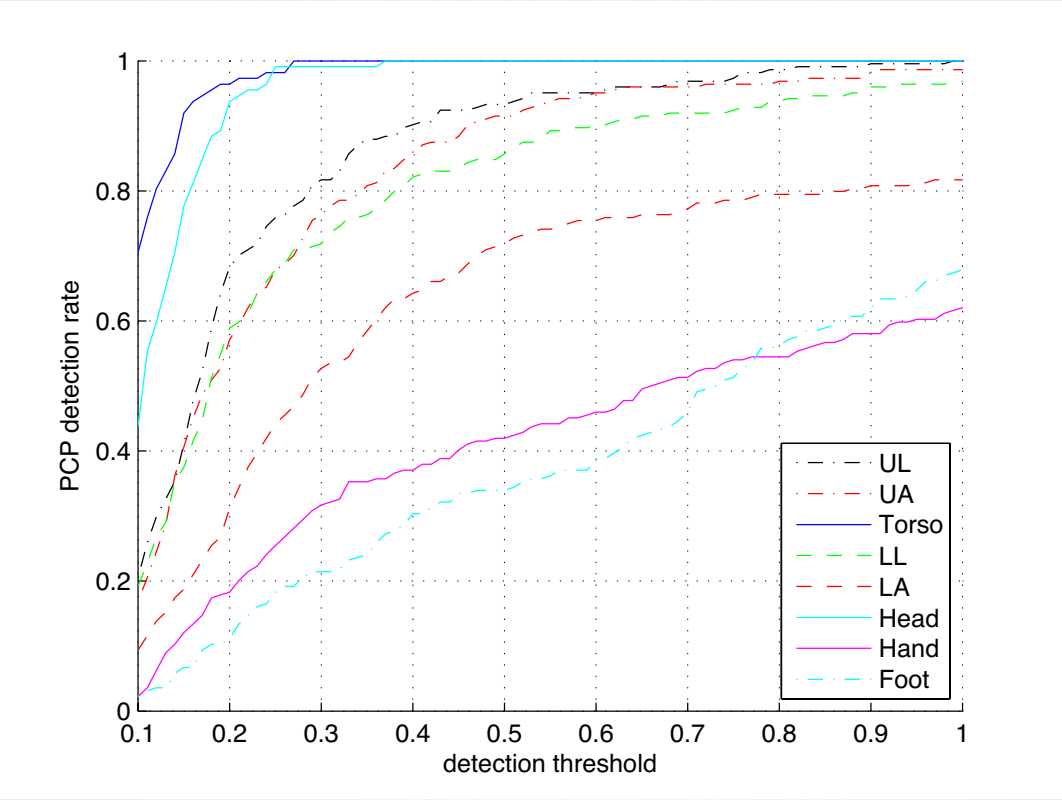
lbody+children





PERFORMANCE

	head	torso	u.leg	l.leg	u.arm	l.arm	hand	foot	avg
Method of Yang et al. (CVPR11)	1.000	1.000	0.975	0.839	0.951	0.577			0.869
Our method	1.000	1.000	0.933	0.857	0.915	0.719	0.420	0.339	0.884



CONCLUSIONS

- Generative model for representing rich appearance of highly deformable objects
- Capture semantic relationships between neighboring part productions
- DP framework for computing exact inference

THANKS!