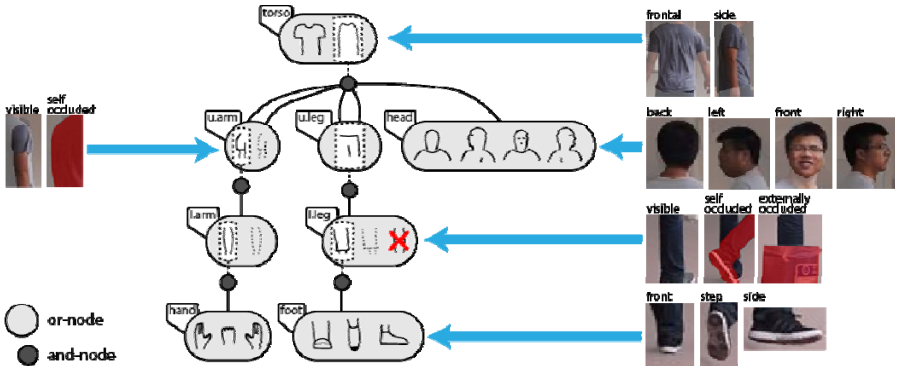
Grammars for Articulated Pose Estimation



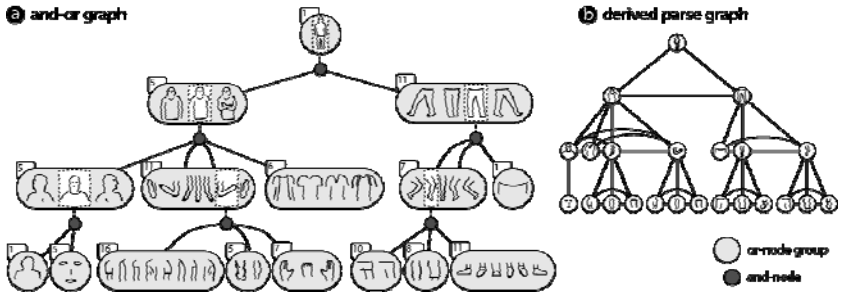
Chapter 7 Image parsing algorithms

1*. Top-down / bottom-up parsing of attributed grammar
2*. Alpha-beta-gamma processes
3*. Discussions on scheduling and decision policy
4. Example on human pose parsing
5. Multi-Armed Bandit problem: exploration vs. exploitation

* Topics discussed in previous chapters.

Parse Graph Derivation

A grammar derivation is created by making a selection for each OR node, starting from the root. Each part encountered is instantiated into a parse graph, which gets its own state variables.



Probability on Parses

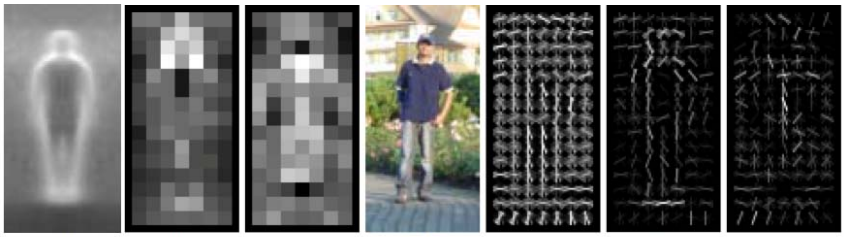
$$p(pt|I) \propto p(pt)p(I|pt) = \frac{1}{Z}\exp\{-\mathcal{E}(pt, I)\}$$

$$\mathcal{E}(pt, I) = \sum_{v \in V(pt)} f^{a}_{\omega}(v, I) + f^{g1}_{\omega}(v) + f^{c1}_{\omega}(v) + \sum_{(v_i, v_j) \in E(pt)} f^{g2}_{\omega_i}(v_i, v_j) + f^{c2}_{\omega_i}(v_i, v_j).$$

$f^{a}(v, I)$	appearance score
$f^{g1}(v)$	geometry orientation score
$f^{g2}(v_i, v_j)$	geometry articulation score
$f^{c1}(v)$	production bias
$f^{c2}(v_i, v_j)$	production compatibility score

Part Appearance

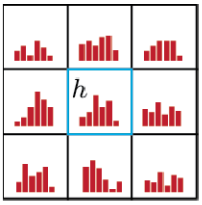
Histogram of Oriented Gradients (Dalal and Triggs, CVPR 2005)

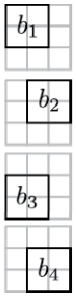


(a) (b) (c) (d) (e) (f) (g)

Local (block) normalization

Each histogram is normalized with L2-norm of neighboring cells





n_1

n_2 $\quad n = \frac{1}{\sqrt{||b||_2^2 + \epsilon^2}}$

n_3 $\quad h \rightarrow \frac{1}{2} h(n_1 + n_2 + n_3 + n_4)$

n_4

Effect of Local Normalization

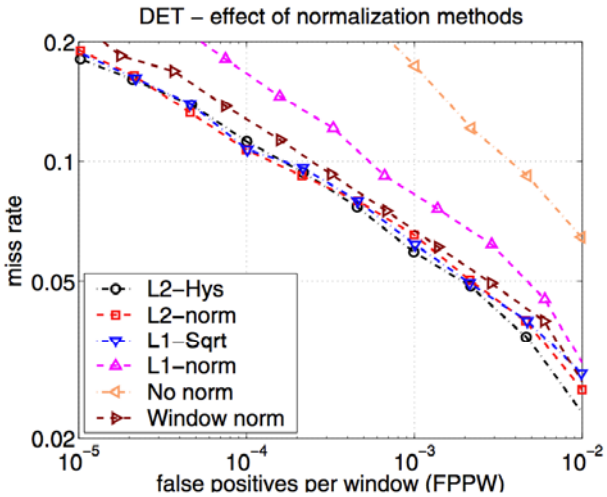


Image Segmentation

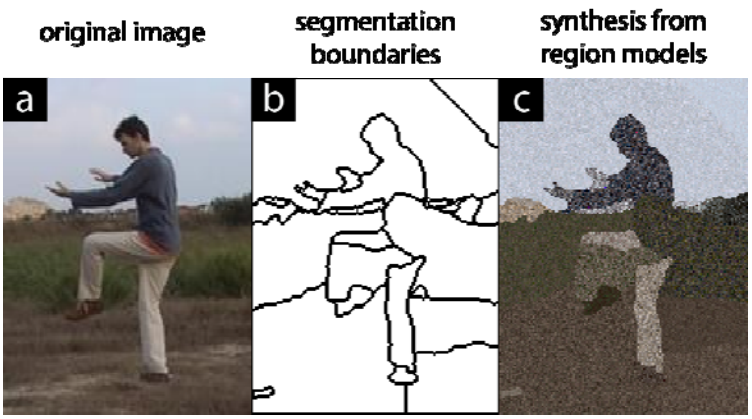


Image Segmentation (DDMCMC)

Prior

boundary smoothness

region model type

few regions

region model complexity

$$p(W) \propto p(K) \prod_{i=1}^{K} p(R_i) p(\ell_i) p(\Theta_i | \ell_i)$$

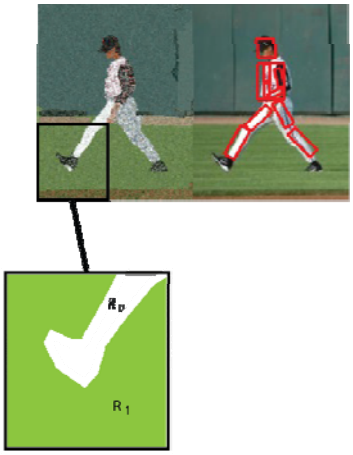
$$\propto \exp\left\{ -\lambda_0 K - \sum_{i=1}^{K} \left[\mu \oint_{\partial R_i} ds + \gamma |R_i|^c + \nu |\Theta_i| \right] \right\}.$$

Likelihood

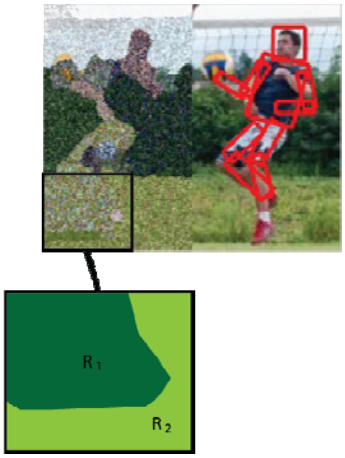
$$p(\mathbf{I}_{\Lambda_i} | R_i; \ell_i, \Theta_i) = \mathcal{N}(\mu_i, \Sigma_i)$$

Using Segmentation to Compute Part Contrast

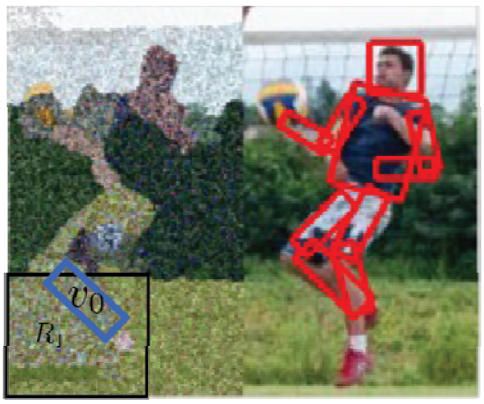
Foreground correctly segmented



Foreground incorrectly segmented



Distance Measure Between Regions



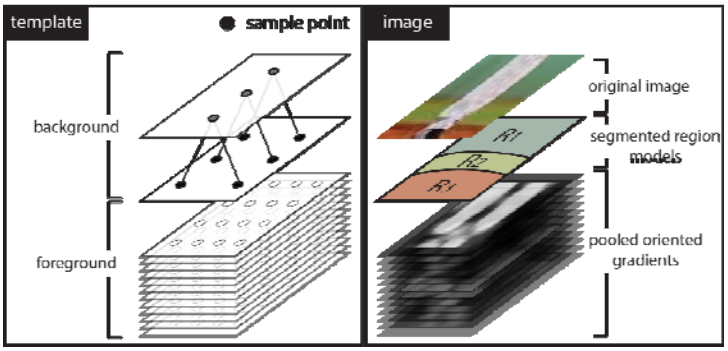
Sample mean inside part boundary

$$d(\mu_{v_0}, \mu_1, \Sigma_1) = (\mu_{v_0} - \mu_1)^\top \Sigma_1^{-1} (\mu_{v_0} - \mu_1)$$

Region model

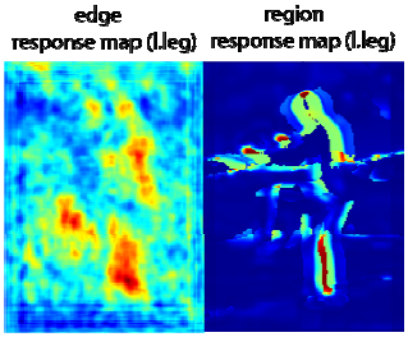
Part Template

Unknown number of adjoining regions is handled by placing region features uniformly around the part boundary

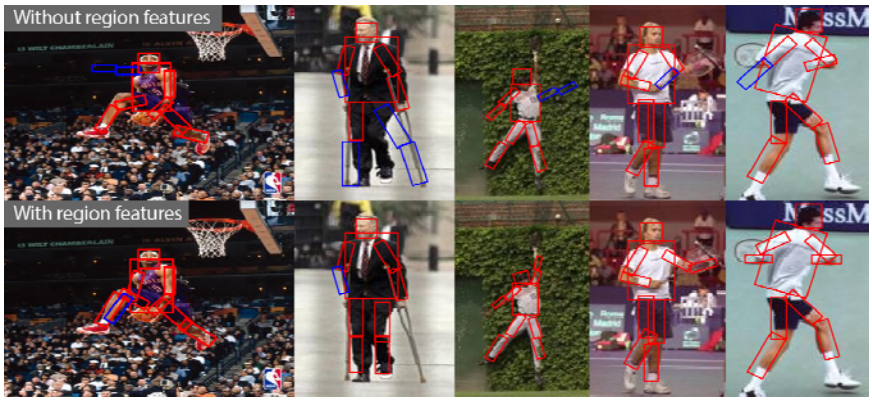


Edge vs. Region Scores

Largely due to local normalization, edge response has many spurious peaks. Region response, however, is more stable.

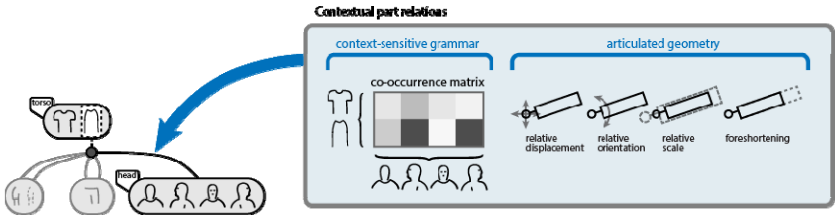


Influence of Region Features on Parsing Results



Potential Functions and Contextual Relations

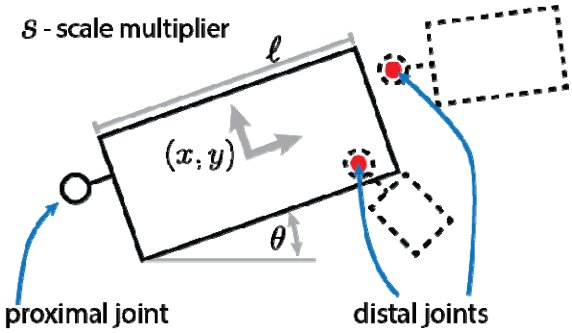
Contextual relations control the geometric, and syntactic compatibility between parts.



Part Parameterization

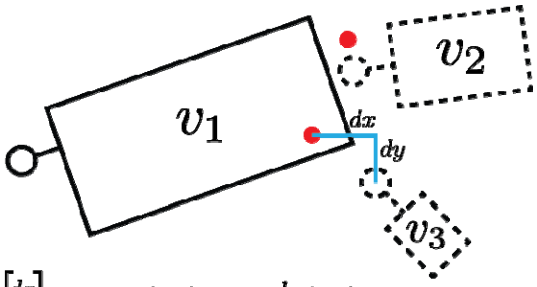
Part state is 6-dimensional, for type, position, scale, orientation, and aspect ratio

$$v = (\omega, x, y, \theta, \ell, s)$$



Coordinate Transforms to Proximal and Distal Joints

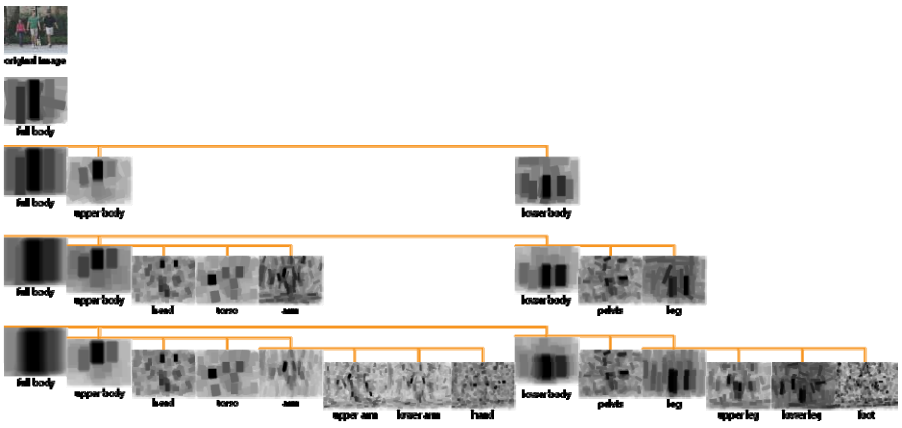
Relations between parts are transformed to the coordinate system of the joint that connects them.



$$\begin{bmatrix} dx \\ dy \end{bmatrix} = T^p_{\omega_3}(v_3) - T^d_{\omega_1}(v_1)$$

Inference

Exact inference can be computed using message-passing / dynamic programming



Inference

Recursive relation to compute the maximal scoring (minimal energy) parse.

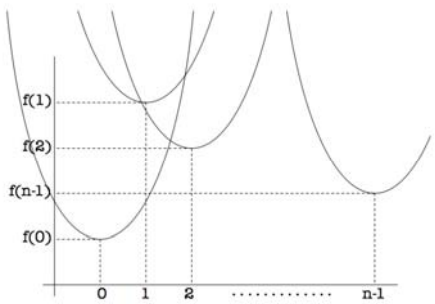
$$M(v_i|\omega_i) = f^a_{\omega_i}(v_i, I) + f^{g1}_{\omega_i}(v_i) + f^{c1}_{\omega_i} + \sum_{(v_i,v_j)\in R_{\omega_i}} \max_{v_j} \left[f^{g2}_{\omega_i}(v_i, v_j) + f^{c2}_{\omega_i}(v_i, v_j) + M(v_j|\omega_j) \right].$$

$$M(x_i, y_i, \theta_i, \ell_i, s_i|\omega_i) = \max_{(\omega_j, x_j, y_j, \theta_j, \ell_j, s_j)} f_{\omega_i}(\omega_j) + f_{\omega_i}(x_i, x_j) + f_{\omega_i}(y_i, y_j) + f_{\omega_i}(\theta_i, \theta_j) + f_{\omega_i}(\ell_i, \ell_j) + f_{\omega_i}(s_i, s_j) + M(x_j, y_j, \theta_j, \ell_j, s_j|\omega_j)$$

Inference

Distance transforms of sampled functions (Felzenszwalb and Huttenlocher, 2004)

$\mathcal{D}_g(x) = \min_{x'}((x - x')^2 + g(x'))$ can be computed in linear time!



We therefore choose the Gaussian potential

$f_{\omega_i}(x_i, x_j) = -dx^2$

Learning

Margin-rescaled structured-output SVM objective function (primal)

$$\min_{\lambda} \frac{1}{2}||w||^2 + \frac{C}{|D|}\sum_{i=1}^{|D|}\xi_i$$

$$s.t.\ \ \lambda^{\top}\left[\phi(\bar{pt}_i, I_i) - \phi(pt, I_i)\right] \geq L(pt, \bar{pt}_i) - \xi_i$$

$$\forall pt \in \Omega_{\mathcal{G}}, \forall i.$$

Learning

SO-SVM dual objective (Wolfe dual):

$$\alpha^* = \arg\max_{\alpha}\left(\langle L, \alpha\rangle - \frac{1}{2}\langle \alpha, H\alpha\rangle\right)$$

$$s.t.\ \ \sum_{i\in\mathcal{I}_k}\alpha_i = \frac{C}{m},\ \forall k,$$

$$\alpha_i \geq 0,\ \forall i$$

$$w = \sum_i \alpha_i z_i$$

$$z_i = \phi(\bar{pt}, I) - \phi(pt, I)$$

Can be solved by one of many freely available QP solvers

Adding Constraints

Separation oracle:

$$\hat{pt}_i = \arg\max_{pt} \lambda^\top \phi(pt, I_i) + L(pt, \bar{pt}_i)$$

Problem: what happens if the parameter weight for the dx^2 terms are negative?

Positivity Constraints on Parameters

Primal constraints are of the form:

$$\langle w, z_i \rangle \geq b_i - \xi_k$$

unit vector to isolate parameter that must be positive

zero

Introduce this artificial example for each constrained parameter, for each example

1-slack vs. N-slack

1-slack primal objective:

$$\min_{\lambda} \frac{1}{2}||w||^2 + C\xi$$

$$s.t. \quad \frac{1}{m}\lambda^\top \left[\sum_i (\phi(\bar{pt}_i, I_i) - \phi(pt, I_i))\right] \geq \frac{1}{m}\sum_i L(pt, \bar{pt}_i) - \xi_i$$

$$\forall pt \in \Omega_g, \forall (pt_1, ..., pt_m).$$

1-slack vs. N-slack

1-slack dual objective:

$$\max_{\alpha \geq 0} \langle L, \alpha \rangle - \frac{1}{2}\langle \alpha, H\alpha \rangle$$

$$s.t. \quad \sum_i \alpha_i = C$$

Constraint Cache with 1-slack

- Keep top-N constraints for each example in a cache, sorted by their inner product with the current model vector
- Synthesize a new 1-slack constraint by summing the top scoring constraints from each example. If this new constraint causes a violation, add it to H and reoptimize the dual.
- This can dramatically reduce the number of times the separation oracle must be called.

Some Results

State-of-art performance on PARSE and Leeds datasets (published in CVPR13)

Performance Evaluation

Dataset	Method	torso	head	u.leg	l.leg	u.arm	l.arm	avg
PARSE	JEa [13] (2010)	85.4	76.1	73.4	65.4	64.7	46.9	66.2
	TZN [21] (2012)	97.1	92.2	85.1	76.1	71.0	45.1	74.4
	FMP [25] (2011)	97.6	93.2	83.9	75.1	72.0	48.3	74.9
	DR [5] (2012)	-	-	-	-	-	-	77.4
	Ours (AG)	99.5	95.6	81.8	67.0	74.3	54.6	75.0
	Ours (AOG)	100.0	96.2	87.0	75.3	73.2	53.9	77.5
	Ours (AOG+BG)	99.5	97.4	88.4	78.0	74.1	56.1	79.0
	Ours (AOG+BG)^f	99.5	97.4	89.2	78.3	74.6	56.9	79.5
Leeds	TZN [21] (2012)	95.8	87.8	69.9	60.0	51.9	32.9	61.3
	JEb [14] (2011)	88.1	74.6	74.5	66.5	53.7	37.5	62.7
	Ours (AG)	98.4	92.8	81.2	69.8	61.9	38.2	69.3
	Ours (AOG)	98.8	92.7	83.9	74.4	64.0	41.1	71.8
	Ours (AOG+BG)	98.3	92.7	83.7	73.1	66.0	41.4	71.9
	Ours (AOG+BG)^f	98.3	92.7	86.8	78.2	70.2	45.1	75.2