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

\[
p(y|I) \propto p(y)p(I|y) = \frac{1}{Z} \exp(-E(y, I))
\]

\[
E(y, I) = \sum_{w \in V(y)} f^a(v, I) + f^d(v) + f^i(v) + \sum_{(v_i, v_j) \in E(y)} f^{a2}(v_i, v_j) + f^{cd}(v_i, v_j)
\]

- \(f^a(v, I)\): appearance score
- \(f^d(v)\): geometry orientation score
- \(f^{a2}(v_i, v_j)\): geometry articulation score
- \(f^{cd}(v_i, v_j)\): production compatibility score
- \(f^i(v)\): production bias
Part Appearance

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

Local (block) normalization

Each histogram is normalized with L2-norm of neighboring cells

\[
\begin{align*}
\mathcal{N}_1 &= n = \frac{1}{\sqrt{||b||^2 + \epsilon^2}} \\
\mathcal{N}_2 &= h \rightarrow \frac{1}{2} h (n_1 + n_2 + n_3 + n_4)
\end{align*}
\]
Effect of Local Normalization

Image Segmentation
Image Segmentation (DDMCMC)

Prior

\[ p(W) \propto p(K) \prod_{i=1}^{K} p(R_i) p(\ell_i) p(\Theta_i | \ell_i) \]
\[ \propto \exp \left\{ -\lambda_k K - \sum_{i=1}^{K} \left[ \mu \int_{aR_i} ds + \gamma |R_i| \kappa + \nu |\Theta_i| \right] \right\} \]

Likelihood

\[ p(I_{\lambda_k} | R_i; \ell_i, \Theta_i) = N(\mu_i, \Sigma_i) \]

Using Segmentation to Compute Part Contrast

Foreground correctly segmented

Foreground incorrectly segmented
Distance Measure Between Regions

\[ d(\mu_0, \mu_1, \Sigma_0, \Sigma_1) = (\mu_0 - \mu_1)^T \Sigma_1^{-1} (\mu_0 - \mu_1) \]

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.

![Edge vs. Region Scores](image)

---

**Influence of Region Features on Parsing Results**

[Images showing the 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.

\[
\begin{bmatrix}
 dx \\
 dy 
\end{bmatrix} = T_{ω_3}^p(v_3) - T_{ω_1}^d(v_1)
\]

Inference

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

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

\[ M(v_i|\omega_i) = f_{\omega_i}^q(v_i, I) + f_{\omega_i}^1(v_i) + f_{\omega_i}^c \]
+ \[ \sum_{(v_i, v_j) \in R_{\omega_i}} \max_{v_j} \left[ f_{\omega_i}^{g_2}(v_i, v_j) + f_{\omega_i}^{c_2}(v_i, v_j) + M(v_j|\omega_j) \right]. \]

\[ M(x_1, y_1, \theta_1, z, a|x_0) = \]
\[ \max_{(x_2, y_2, \theta_2, x_3, y_3)} \left[ f_{x_1}(\omega_1) + f_{x_1}(x_1, x_2) + \sum_{i=2}^{n} \left( f_{x_i}(\theta_i, \theta_{i-1}) + f_{x_i}(x_i, x_{i+1}) \right) + M(x_n, y_n, \theta_n, x_3, y_3) \right] \]

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

\[ D_g(x) = \min_{x'}((x - x')^2 + g(x')) \] 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):

\[
\min_{\lambda} \frac{1}{2} ||w||^2 + \frac{C}{|D|} \sum_{i=1}^{|D|} \xi_i \\
\text{s.t. } \lambda^T [\phi(\tilde{p}t_i, I_i) - \phi(pt, I_i)] \geq L(pt, \tilde{p}t_i) - \xi_i \\
\forall pt \in \Omega_\Omega, \forall i.
\]

SO-SVM dual objective (Wolfe dual):

\[
\alpha^* = \arg \max_\alpha \left( \langle L, \alpha \rangle - \frac{1}{2} \langle \alpha, H \alpha \rangle \right) \\
\text{s.t. } \sum_{i \in I_k} \alpha_i = \frac{C}{m}, \forall k, \\
\alpha_i \geq 0, \forall i
\]

\[
w = \sum_i \alpha_i z_i \\
z_i = \phi(\tilde{p}i, I) - \phi(pt, I)
\]

Can be solved by one of many freely available QP solvers.
Adding Constraints

Separation oracle:

\[
\hat{p}_t = \arg \max_{p_t} \lambda^T \phi(p_t, I_t) + L(p_t, \tilde{p}_t)
\]

Problem: what happens if the parameter weight for the $\alpha^2$ terms are negative?

Positivity Constraints on Parameters

Primal constraints are of the form:

\[
\langle w, z_i \rangle \geq b_i - \zeta_k
\]

Introduce this artificial example for each constrained parameter, for each example
1-slack vs. N-slack

1-slack primal objective:

\[
\min_\lambda \frac{1}{2} \|w\|^2 + C \xi \\
\text{s.t.} \quad \frac{1}{\eta} \lambda^T \left[ \sum_i (\phi(p\tilde{t}_i, I_i) - \phi(p_i, I_i)) \right] \geq \frac{1}{\eta} \sum_i L(pt_i, \tilde{t}_i) - \xi_i \\
\forall pt \in \Omega_g, \forall (p_{t_1}, \ldots, p_{t_m}).
\]

1-slack dual objective:

\[
\max_{\alpha \geq 0} \langle L, \alpha \rangle - \frac{1}{2} \langle \alpha, H\alpha \rangle \\
\text{s.t.} \quad \sum_i \alpha_i = C
\]
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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>torso</th>
<th>head</th>
<th>u.leg</th>
<th>l.leg</th>
<th>u.arm</th>
<th>l.arm</th>
<th>avg</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PARSE</strong></td>
<td>J Ea [13] (2010)</td>
<td>85.4</td>
<td>76.1</td>
<td>73.4</td>
<td>65.4</td>
<td>64.7</td>
<td>46.9</td>
<td>66.2</td>
</tr>
<tr>
<td></td>
<td>TZN [21] (2012)</td>
<td>97.1</td>
<td>92.2</td>
<td>85.1</td>
<td>76.1</td>
<td>71.0</td>
<td>45.1</td>
<td>74.4</td>
</tr>
<tr>
<td></td>
<td>FMP [25] (2011)</td>
<td>97.6</td>
<td>93.2</td>
<td>83.9</td>
<td>75.1</td>
<td>72.0</td>
<td>48.3</td>
<td>74.9</td>
</tr>
<tr>
<td></td>
<td>DR [8] (2012)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>77.4</td>
</tr>
<tr>
<td></td>
<td>Ours (AG)</td>
<td>99.5</td>
<td>95.6</td>
<td>81.8</td>
<td>67.0</td>
<td>74.3</td>
<td>54.6</td>
<td>75.0</td>
</tr>
<tr>
<td></td>
<td>Ours (AOG)</td>
<td>100.0</td>
<td>96.2</td>
<td>87.0</td>
<td>75.3</td>
<td>73.2</td>
<td>53.9</td>
<td>77.5</td>
</tr>
<tr>
<td></td>
<td>Ours (AOG+BG)</td>
<td>99.5</td>
<td>97.4</td>
<td>88.4</td>
<td>78.0</td>
<td>74.1</td>
<td>56.1</td>
<td>79.0</td>
</tr>
<tr>
<td></td>
<td>Ours (AOG+BG)^1</td>
<td>99.5</td>
<td>97.4</td>
<td>89.2</td>
<td>78.3</td>
<td>74.6</td>
<td>56.9</td>
<td>79.5</td>
</tr>
<tr>
<td><strong>Leeds</strong></td>
<td>TZN [21] (2012)</td>
<td>95.8</td>
<td>87.8</td>
<td>69.9</td>
<td>60.0</td>
<td>51.9</td>
<td>32.9</td>
<td>61.3</td>
</tr>
<tr>
<td></td>
<td>JBb [14] (2011)</td>
<td>88.1</td>
<td>74.6</td>
<td>74.5</td>
<td>66.5</td>
<td>53.7</td>
<td>37.5</td>
<td>62.7</td>
</tr>
<tr>
<td></td>
<td>Ours (AG)</td>
<td>98.4</td>
<td>92.8</td>
<td>81.2</td>
<td>69.8</td>
<td>61.9</td>
<td>38.2</td>
<td>69.3</td>
</tr>
<tr>
<td></td>
<td>Ours (AOG)</td>
<td>98.8</td>
<td>92.7</td>
<td>83.9</td>
<td>74.4</td>
<td>64.0</td>
<td>41.1</td>
<td>71.8</td>
</tr>
<tr>
<td></td>
<td>Ours (AOG+BG)</td>
<td>98.3</td>
<td>92.7</td>
<td>83.7</td>
<td>73.1</td>
<td>66.0</td>
<td>41.4</td>
<td>71.9</td>
</tr>
<tr>
<td></td>
<td>Ours (AOG+BG)^1</td>
<td>98.3</td>
<td>92.7</td>
<td>86.8</td>
<td>78.2</td>
<td>70.2</td>
<td>45.1</td>
<td>75.2</td>
</tr>
</tbody>
</table>