
Hemerson Pistori  
*Biotechnology Dept., Dom Bosco Catholic University*

January, 2012  
*Bristol, UK*

*Most of the pictures used in this presentation were extracted from Zhu's paper*
Grammars as Universal And-Or Trees

Grammar

\[ V_T = \{a, b\} \]
\[ V_N = \{S\} \]
\[ R = \{r_1 : S \rightarrow aS, \ r_2 : S \rightarrow b\} \]

And–Or tree

Fig. 2.2 A very simple grammar, its universal And–Or tree and a specific parse tree in shadow.

- Universal And-Or Tree can have an infinite size (as in the example)
- Rules are explicitly named (r1, r2, …)
- Each or-node A have one child for each rule having A at its left side
- A parsing tree is a sub-graph of a universal and-or tree

“ab” not “abb”
Ambiguity in Visual And-Or Trees

- Cycles or Diamonds
- Overlapping Reusable Parts
- Not dealt with yet (2007)
- Most common causes in images:
  - Ambiguous scenes
  - Multiple partial patterns
  - Joints
  - Occlusion (most common) ... extra leaves, interpolation
Main issues in Visual Grammars unseen in Textual Grammars

- No left-to-right ordering in language
  
  **Solution:** explicitly add horizontal edges to represent adjacency

- Objects appear in arbitrary scales
  
  **Solution:** termination rules at different levels (higher leaves)

- Much wider spectrum of quite irregular local patterns
  
  **Solution:** combine Markov random fields with stochastic grammars
Contextual information

Fig. 2.11 A parser tree for a block world from [22]. The ellipses represent non-terminal nodes and the squares are for terminal nodes. The parse tree is augmented into a parse graph with horizontal connections for relations, such as one object supporting the other, or two adjacent objects sharing a boundary.

Horizontal lines to represent relations and constraints:

- Bonds and connections (more dense)
- Joints and junctions
- Interactions and semantics (less dense). E.g.: person eating an apple
Stochastic information

- Probabilities for rules (stochastic grammars). One local probability at each Or-node to account for the relative frequency of each alternative.

- Probabilities of relations (Markov random fields). Local energies associated with each horizontal link.

- A Configurations is a “word” of the “visual language”.
Visual Vocabulary

- Bonds – topological information
- Three hierarchical and connected levels

Image **primitives**: Textons (blobs, bars, terminators and crosses)

Geometric **groupings**: Graphlets

**Object parts**
Clock example

And-Or Graph (Grammar)  
And-Or Parse Graphs
Main elements to be learned: (1) Vocabulary and And-Or tree, (2) Relations – Horizontal Line and (3) Parameters

What is available (training data): Images and parse trees (manually constructed ground-truths)

Three phases:

- Learning parameters from training data given relations and vocabulary (gradient method)
- Learning news relations given vocabulary and learned parameters (inspired in texture synthesis)
- Learning vocabulary and And-Or tree
Image Parsing

- Iterative Heuristic top-down + bottom-up search

- Bottom-up:
  - Hough transforms, Adaboosting … identify possible terminals from images
  - Bind a number of parts

- Top-down:
  - Expand top level nodes
  - Update the weights of the current hypothesis

Key issue: scheduling of bottom-up and top-down search steps (depends on the problem)
What we can explore

- Learning visual structure from pure textual information
- Top level grammars or constraints based on Santini's work on Semiological Level of Significance
- Define and Adaptive Stochastic And-Or Graph to deal with non-static scenes
- Propose a way to deal with temporal information in the context of Visual Grammars