Hierarchical Object Recognition

Chris Williams

Institute for Adaptive and Neural Computation
School of Informatics, University of Edinburgh, UK

September 2006

Object Recognition as Image Parsing

Is a nose an object? Is a head one? Is it still one if it is attached to a body? What about a man on horseback?
Marr, 1982, p 270

- Metaphor (?) of image interpretation as parsing
- Recursive grouping of lower level features into *parts*
- Structural pattern recognition
- But current methods (constellation model, templates of features) are "flat"; should we be worried about this?

Some references

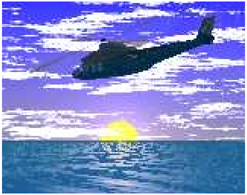
- Marr and Nishihara (1978)
- Biederman (1985): Recognition by Components
- Bienenstock, Geman and Potter (1997): Compositionality, MDL priors and object recognition
- Williams and Adams (1999): Dynamic Trees
- Hinton, Ghahramani and Teh (2000): Credibility nets
- S-C Zhu et al (ICCV 2003, ICCV 2005)
- Bouchard and Triggs (CVPR, 2005)
- Jin and S. Geman (CVPR, 2006)
- Fidler, Berginc, Leonadis (CVPR, 2006)

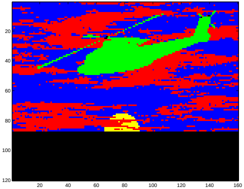
Two Computational Approaches

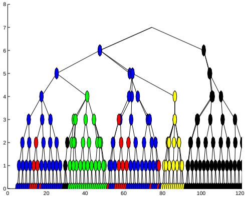
- Tree-like generative models
- Bottom-up grouping of features

Dynamic Trees

Storkey and Williams, PAMI 2003. Inference using structured mean field methods







$$P(Y, X, R', Z | R^L) = P(Z)P(X|Z)P(R'|R^L, Z)P(Y|X)$$

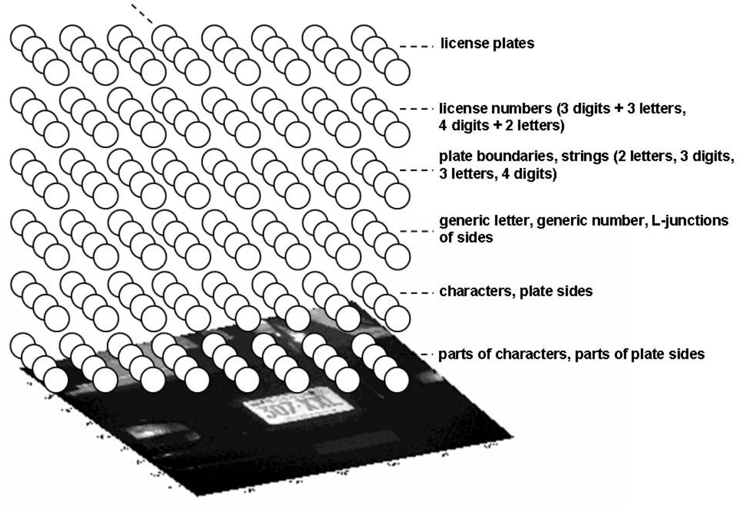
where

- Z denotes the tree structure
- X denotes the labels of the nodes in the tree
- R' denotes the positions of the non-terminal nodes
- R^L denotes the positions of the terminal nodes
- Y denotes the image data

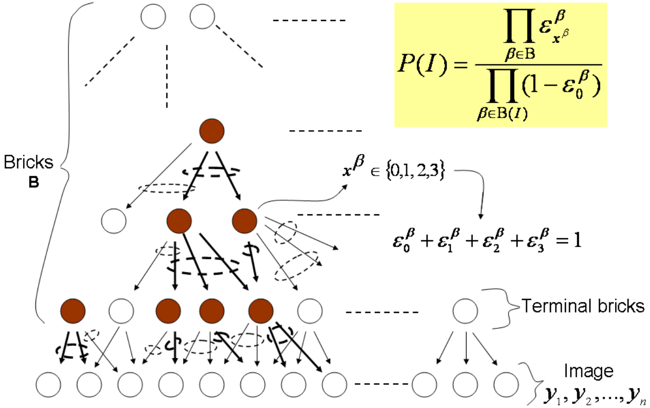
Jin and Geman, CVPR 2006

Context and Hierarchy in a Probabilistic Image Model
Similar to Dynamic Trees, but

- Dense fields of "bricks" which have sparse activation
- Use of a "compositional distribution" to overcome the limitations of the Markov backbone
- Productions of bricks at one level give rise to relative positions and types of bricks at lower levels, e.g. characters $\rightarrow$ parts (of characters or plate sides)
- Note that parts of characters are used in many different character types

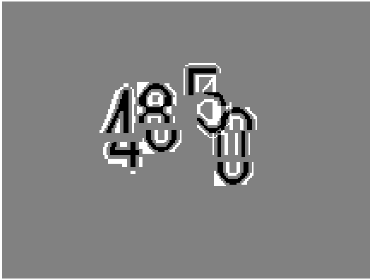


Jin and Geman, 2006



Markov backbone

Jin and Geman, 2006



Jin and Geman, 2006

Jin and Geman, 2006. A typical parse with its top 25 objects: the licence plate, followed by L-junctions, lines, and false positive characters

Bottom-up strategies

- In Jin and Geman (2006), a greedy bottom-up pass is used to initialize the search strategy
- This involves testing each brick in each possible state in a given level, given possible states activated in the level below
- Note that in their system the relationships were hand coded
- Fidler et al (2006) suggest a bottom up method for *learning* to group features from a lower level, looking at what, orientation and where information (Barlow's suspicious coincidences)
- See also Sivic and Zisserman (CVPR, 2004) for discovering configurations of visual words

Pros and Cons of Hierarchical Strategies

Pros

- Neat recursive strategy, appealing on aesthetic grounds
- Re-use of parts when dealing with multiple classes $\Rightarrow$ efficiency
- Computational efficiency; break down the matching problem into subtasks (divide and conquer)
- Utility of deep nets, e.g. for modelling digits

Cons

- "Tower of jelly" problem of deep networks
- Lack of a probabilistic drive for parts unless there is more between-parts variability than within-parts
- Hierarchy is more complex, don't add it unless it is needed
- Is there any psychological or biological evidence for multiple layers?