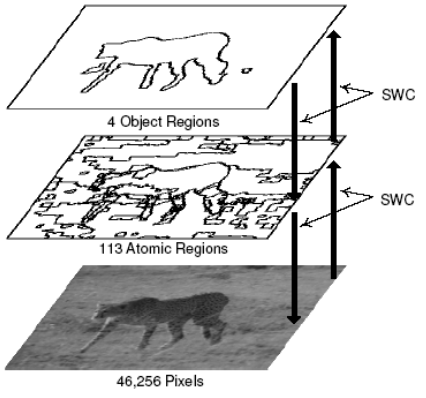
Ch 8-9: Integrated method and parsing in And-Or graphs:

Top-Down/Bottom-up Inference with hierarchical models

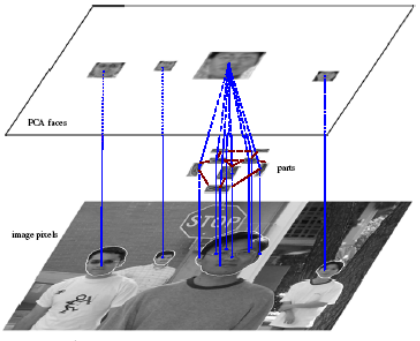
Based on the work by T.F. Wu et al. "A Numeric Study of the Bottom-up and Top-down Inference Processes in And-Or Graphs", *Int'l Journal of Computer Vision*, Vol. 93, No.2, pp226-252, 2011

Two Basic Computing Mechanisms: Bottom-up vs. Top-down

Some objects can be computed more effectively by bottom-up while others by top-down



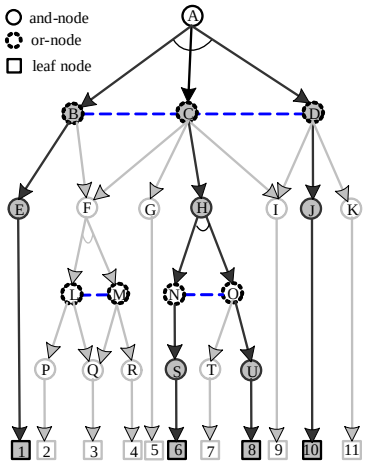
(a) bottom-up graph construction



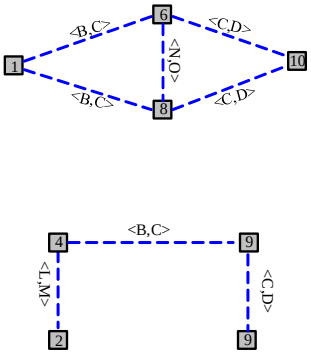
(b) Top-down graph construction

How to formulate this problem ?

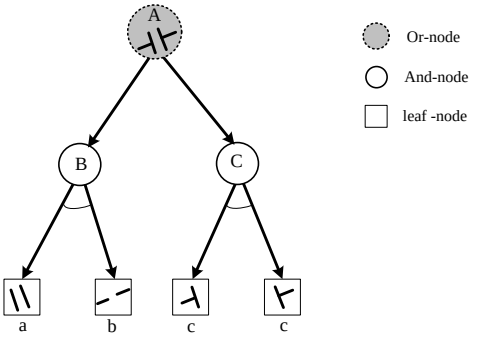
Embedding the integrated models into an And-Or graph



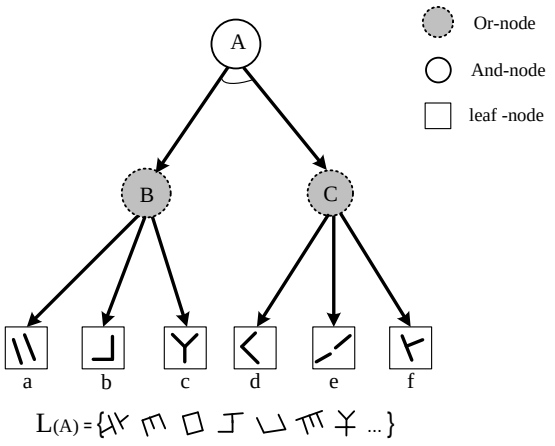
some graph configurations generated by the And-Or graph



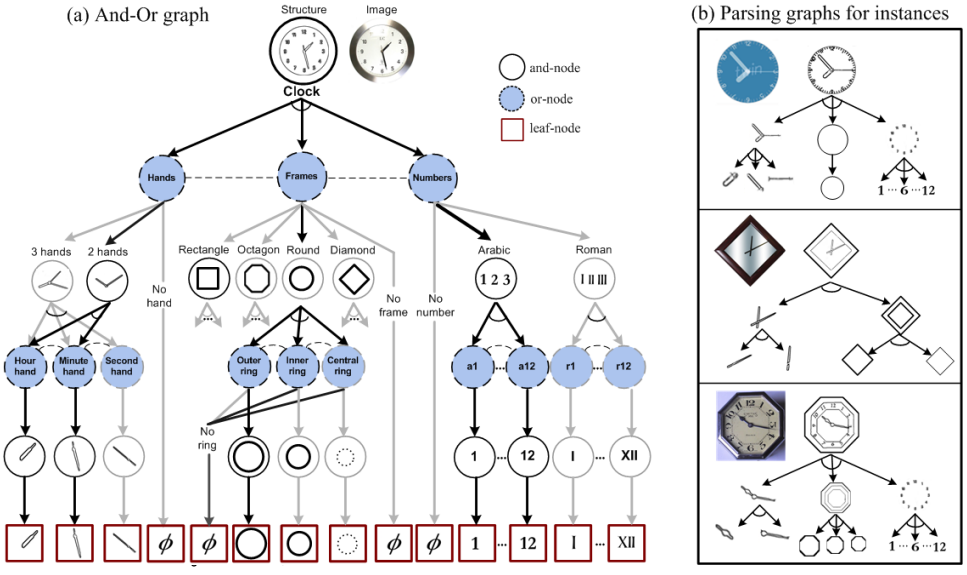
Representing a grammar by and-or graph



Representing a grammar by and-or graph



An example: the clock category



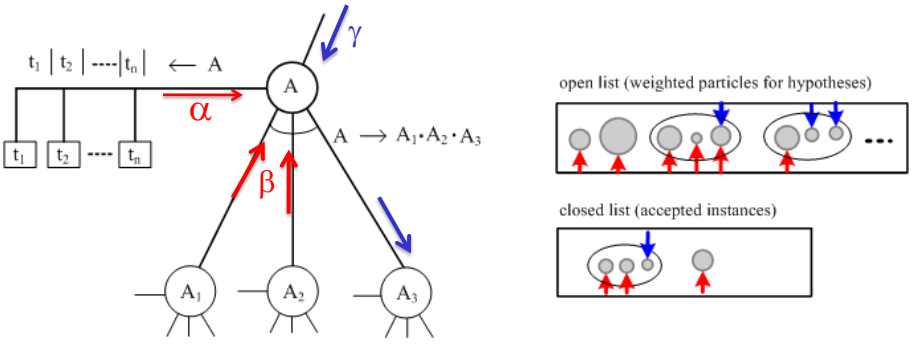
α, β and γ computing processes in AoG

The And-Or graph is a recursive structure. So, consider a node A.

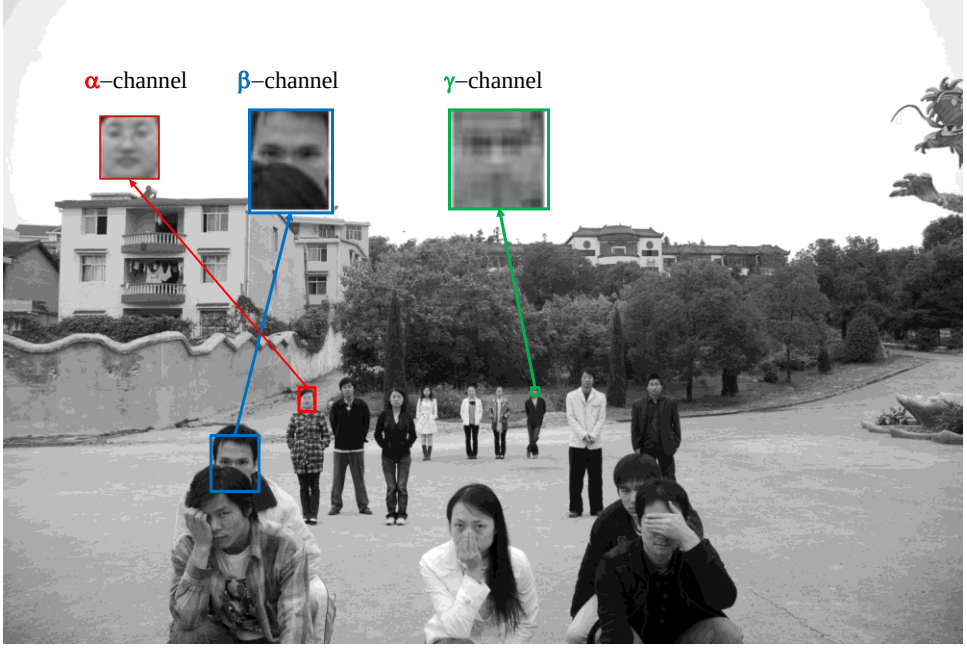
1, any node A terminate to leaf nodes at a coarse scale (ground).

2, any node A is connected to the root.

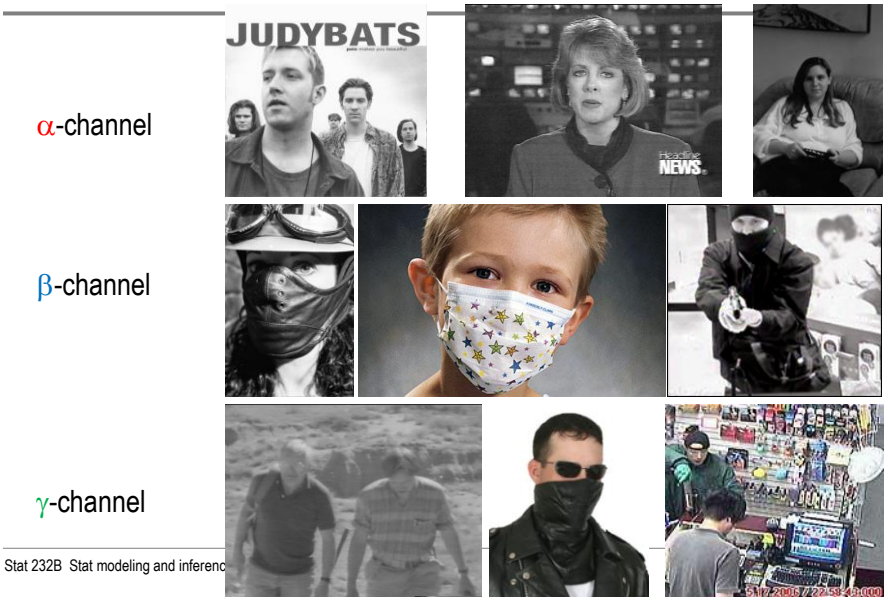
Starting the α/β/γ channels when they are applicable ---an optimal scheduling problem



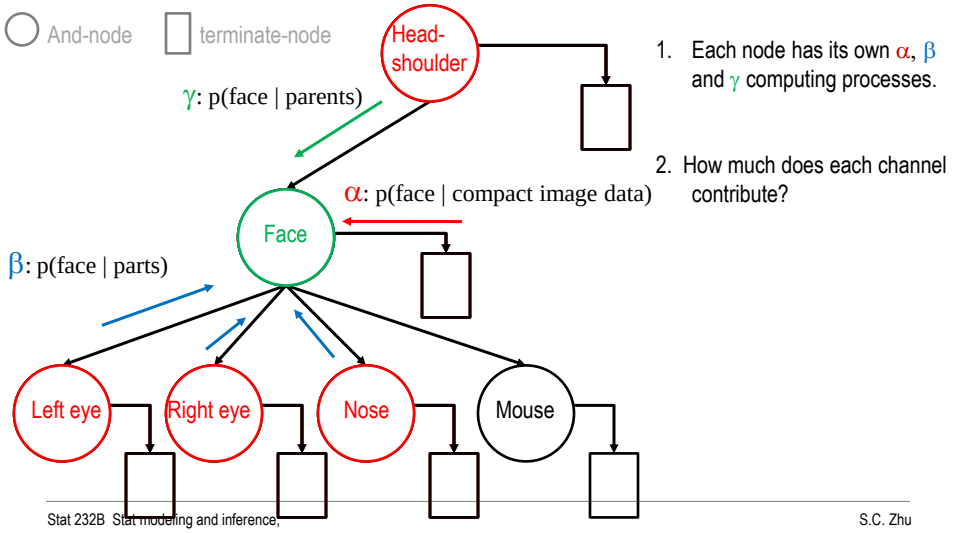
An example: human faces are computed in



Human faces in real scenarios

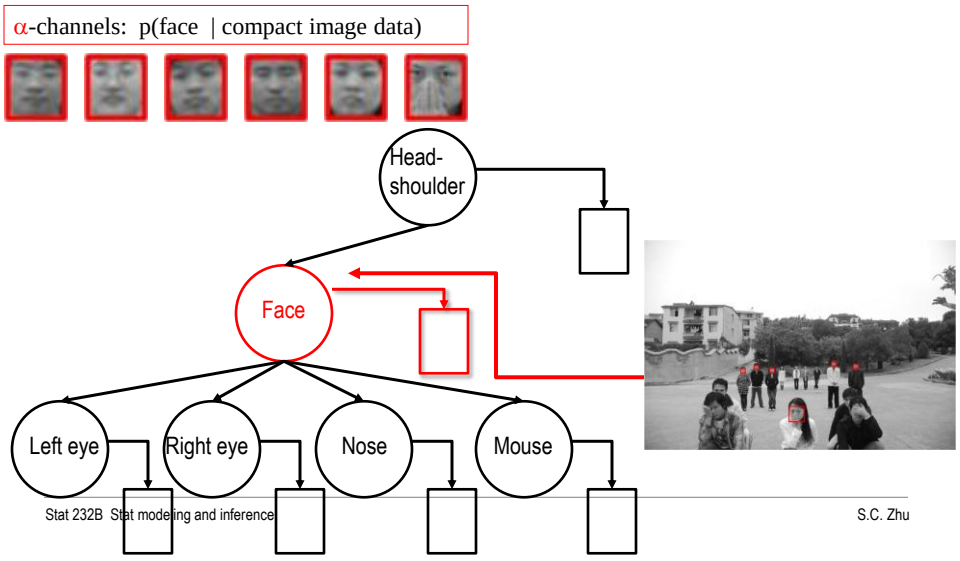


Hierarchical modeling and α, β and γ computing

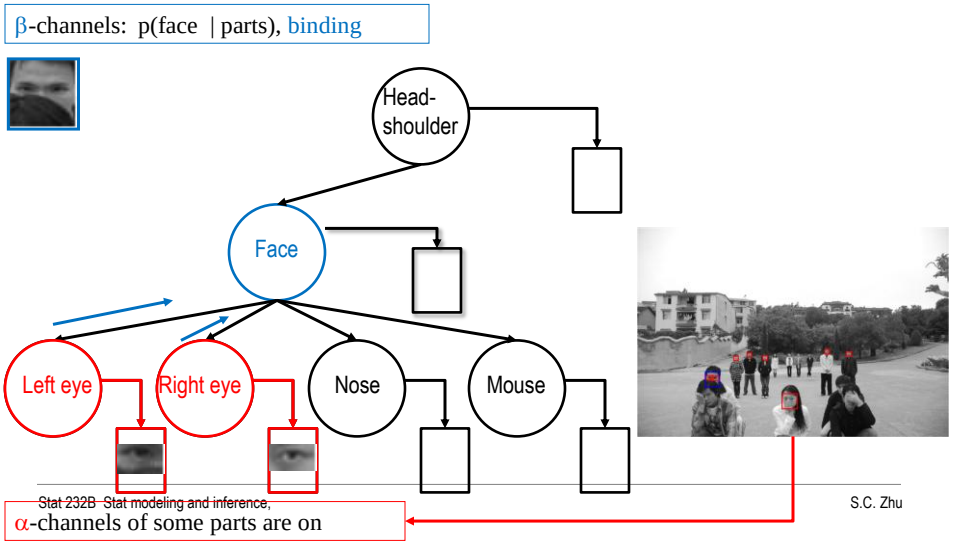


1. Each node has its own α, β and γ computing processes.
2. How much does each channel contribute?

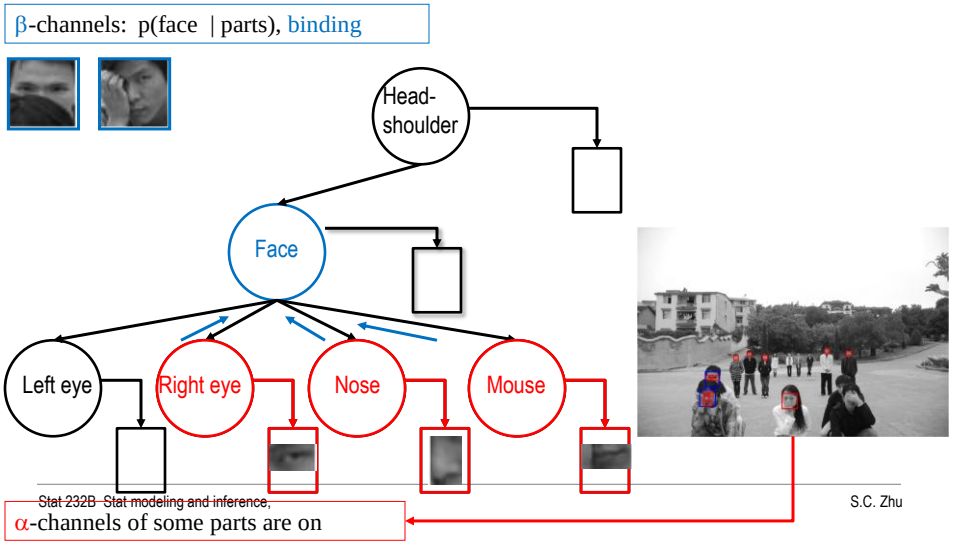
α processes for the face node



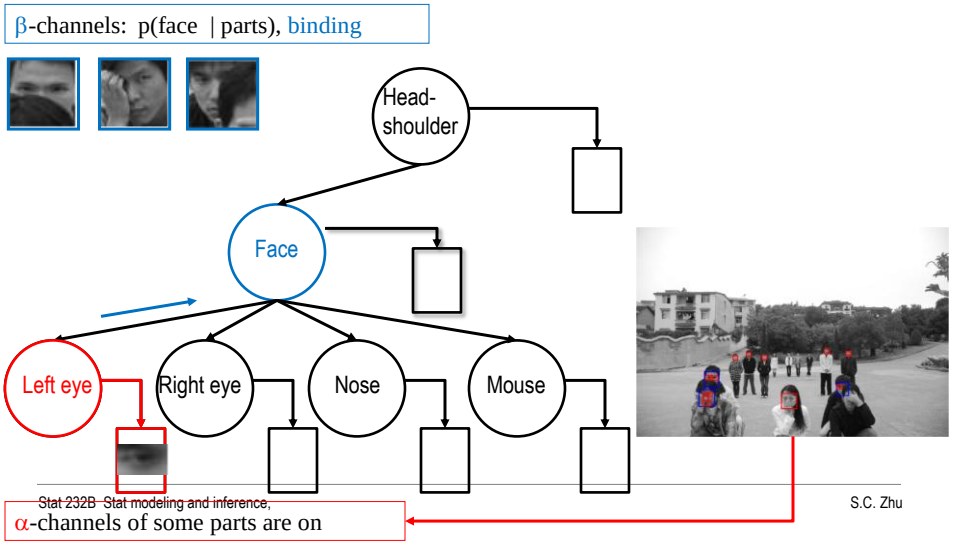
β processes for the face node(when its α is off)



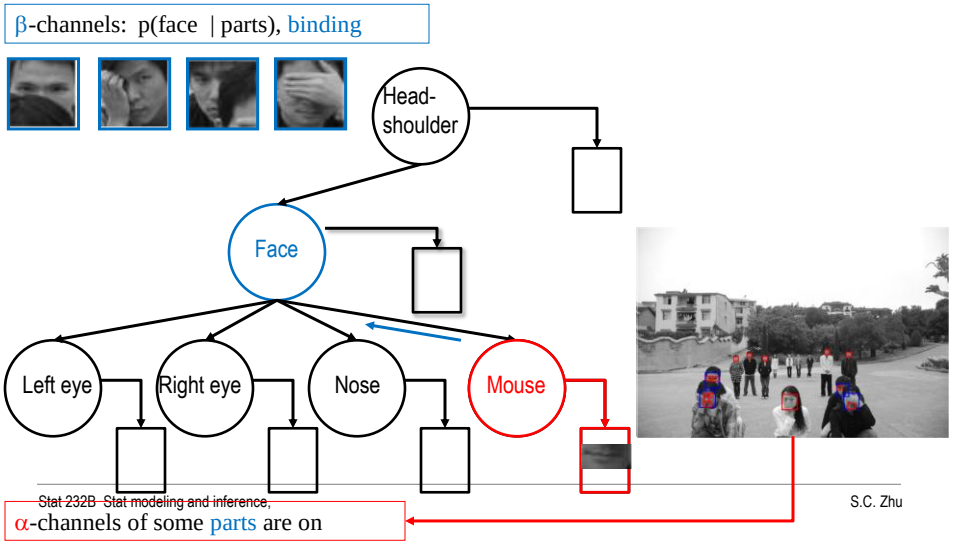
β processes for the face node(when its α is off)



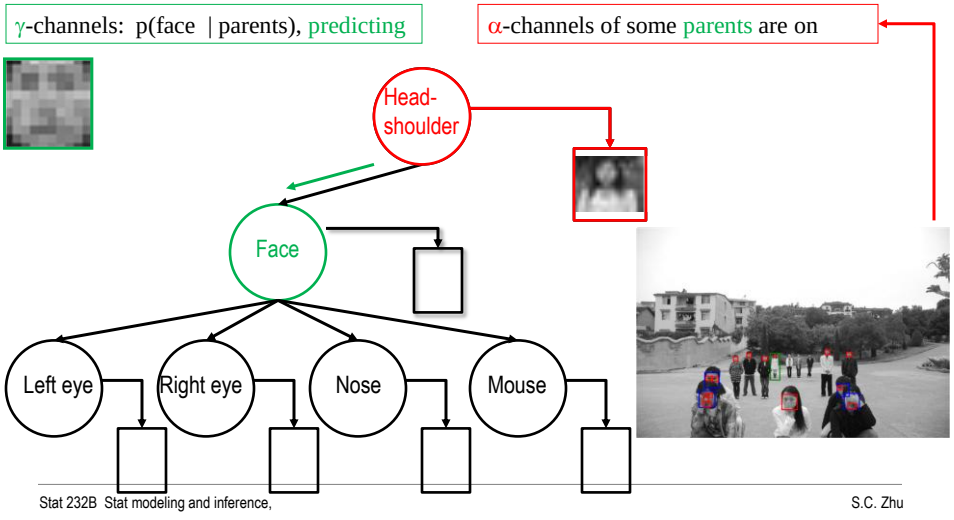
β processes for the face node(when its α is off)



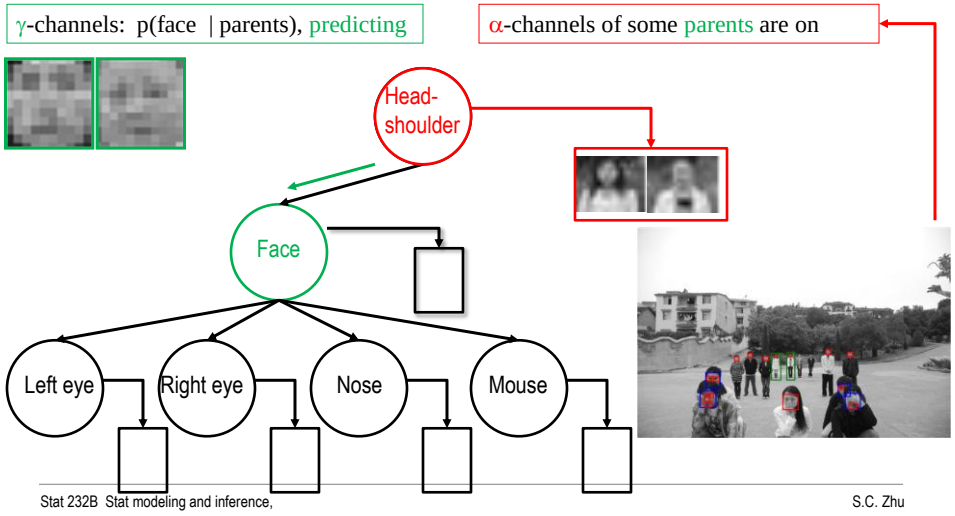
β processes for the face node(when its α is off)



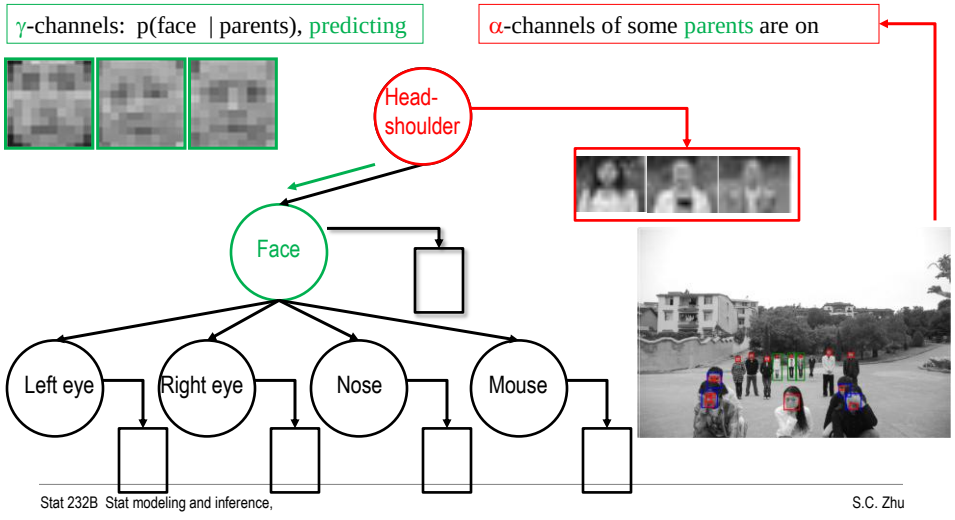
γ processes for the face node(when it's α and β is off)



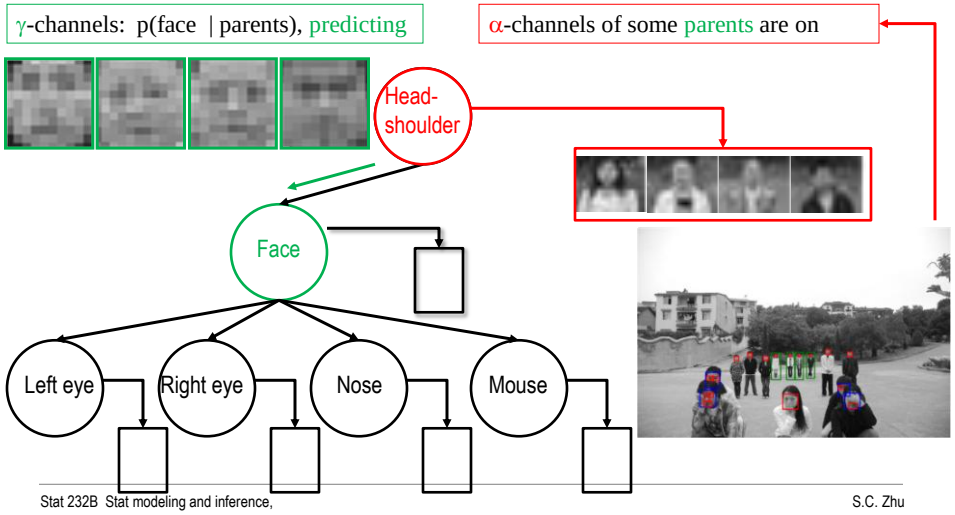
γ processes for the face node(when it's α and β is off)



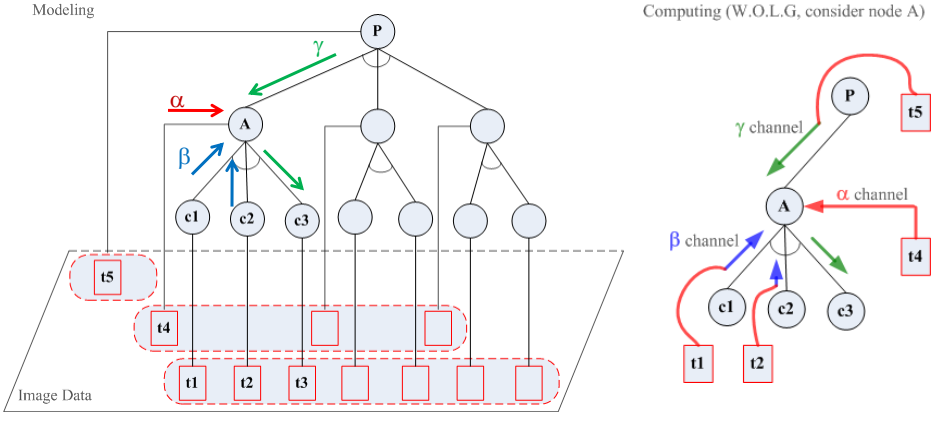
γ processes for the face node(when it's α and β is off)



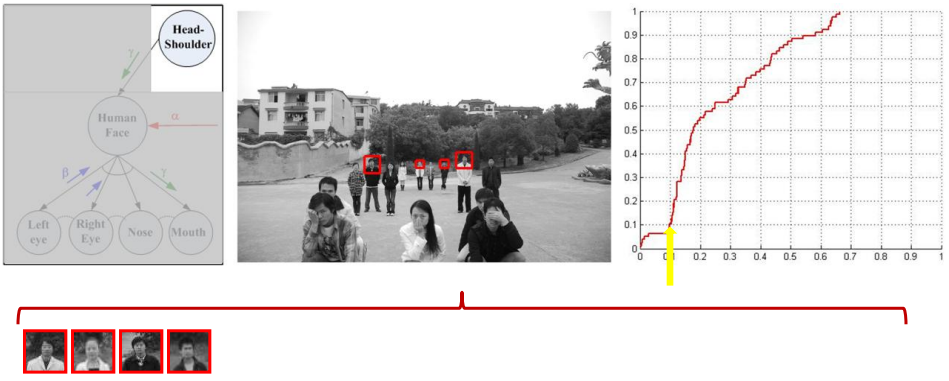
γ processes for the face node(when it's α and β is off)



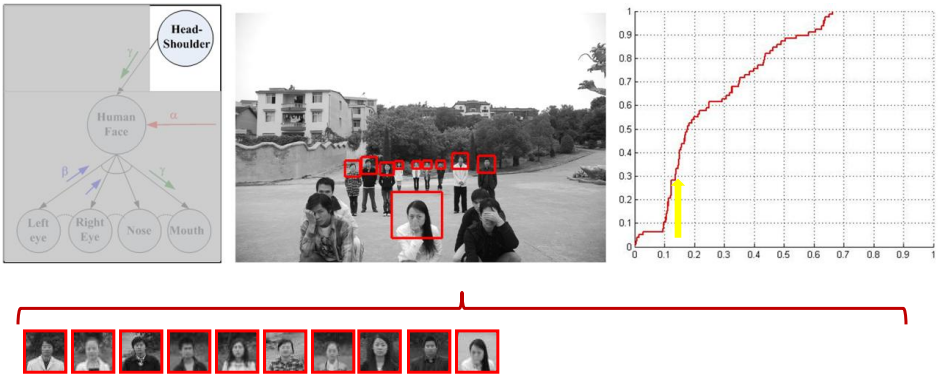
In general: recursive α, β and γ channels



α–channel: head-shoulder

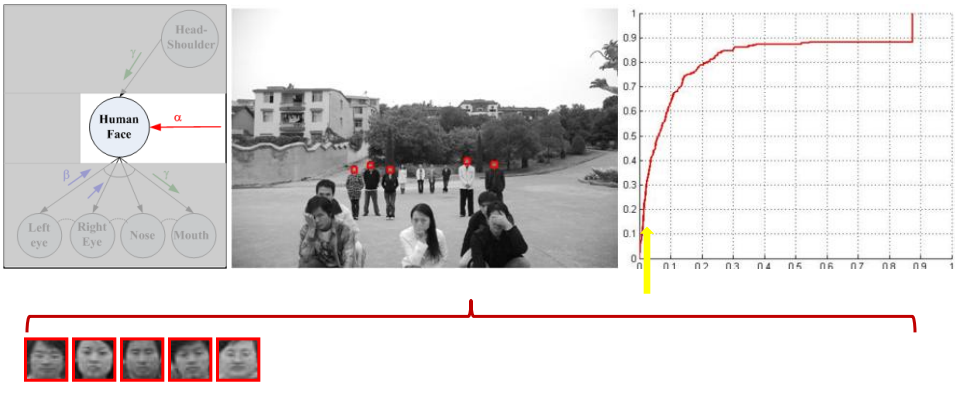


α–channel: head-shoulder

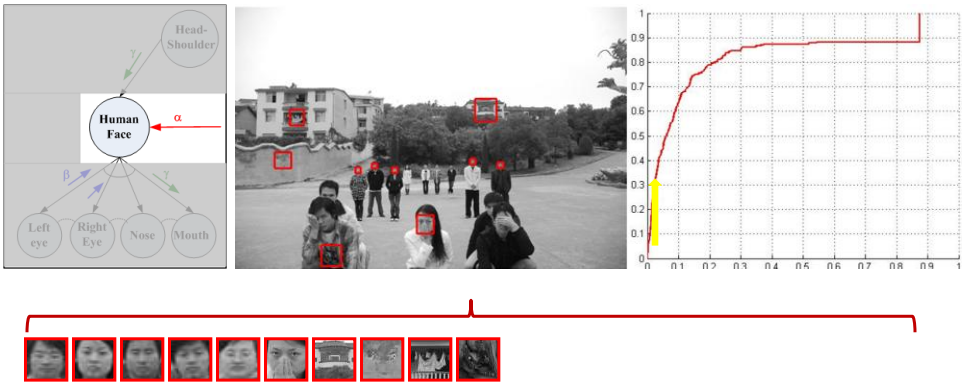


α–channel: head-shoulder

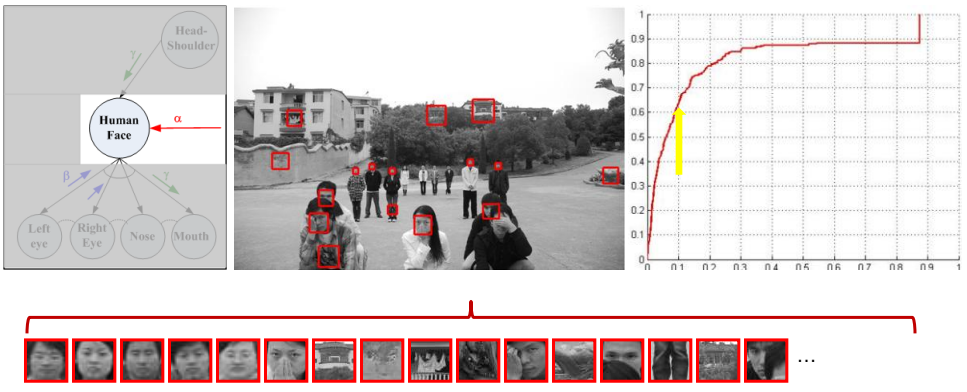
α–channel: face



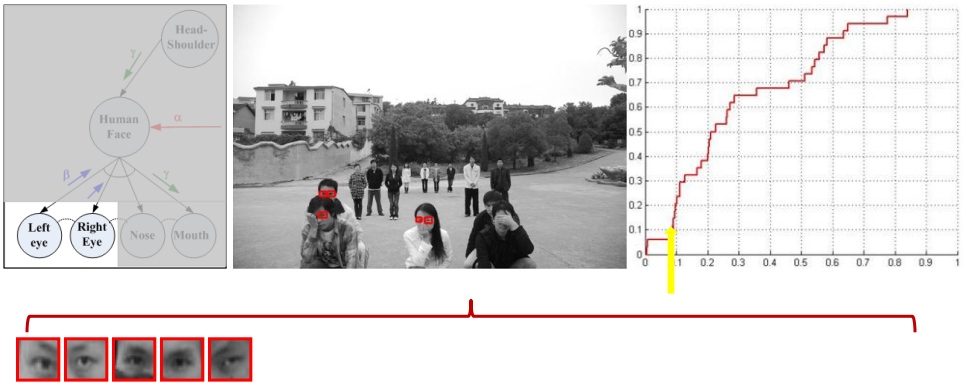
α–channel: face



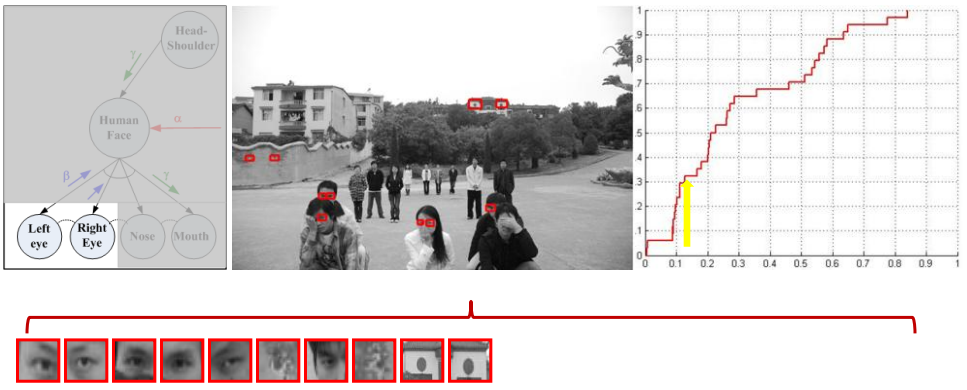
α–channel: face



α–channel: eye

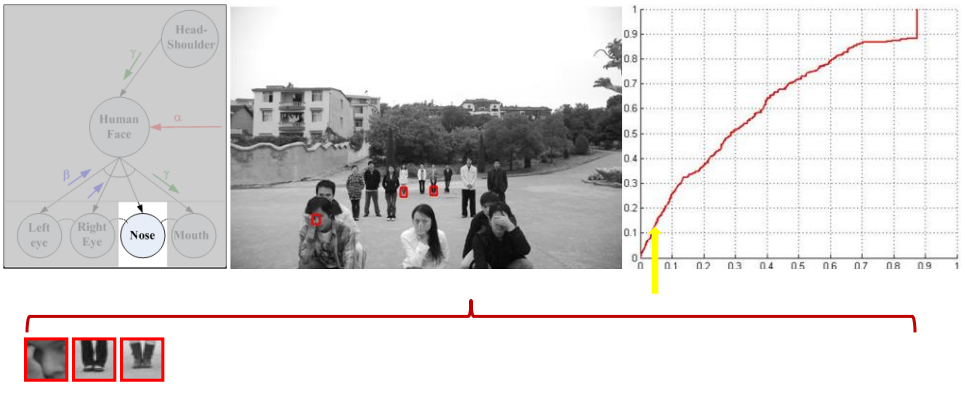


α–channel: eye

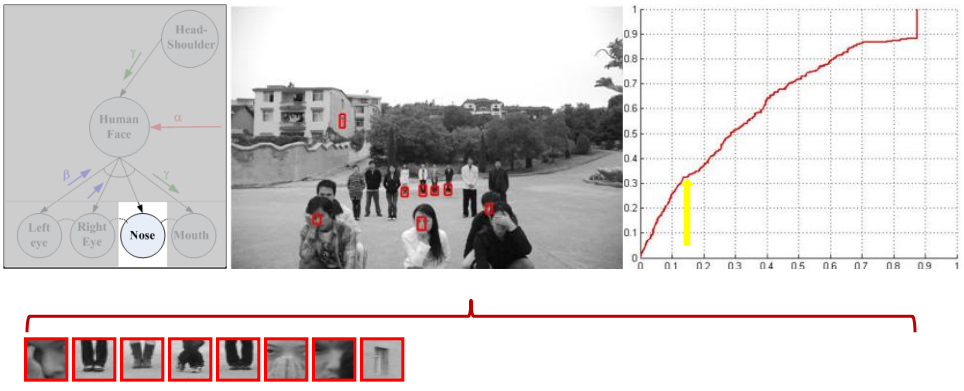


α–channel: eye

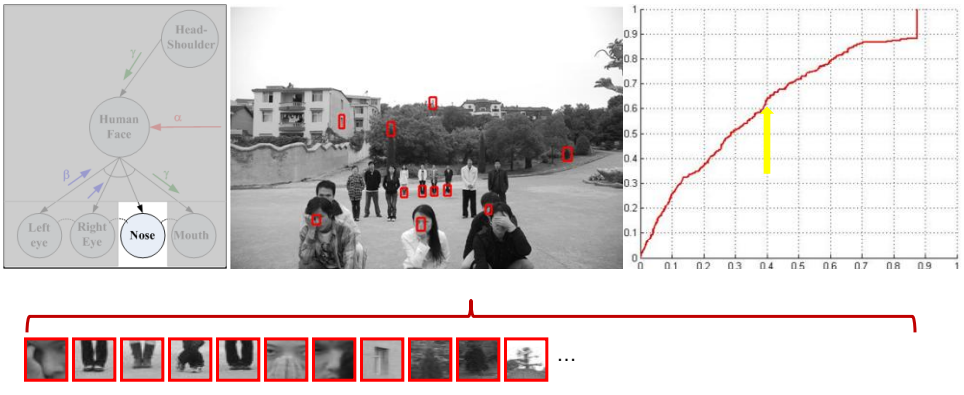
α–channel: nose



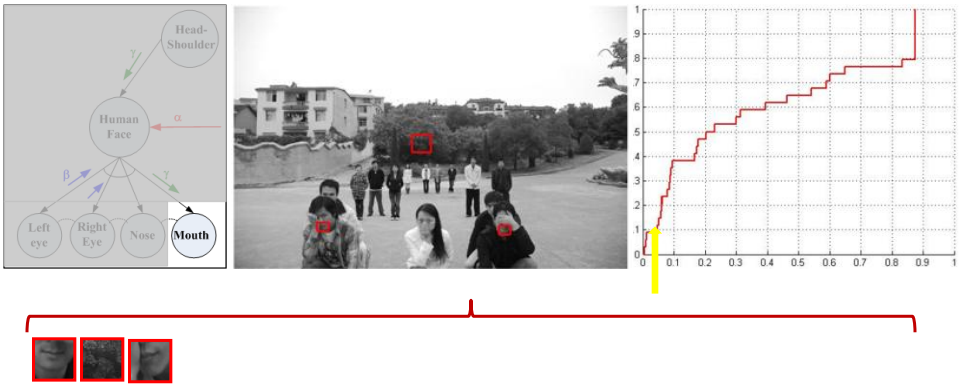
α–channel: nose



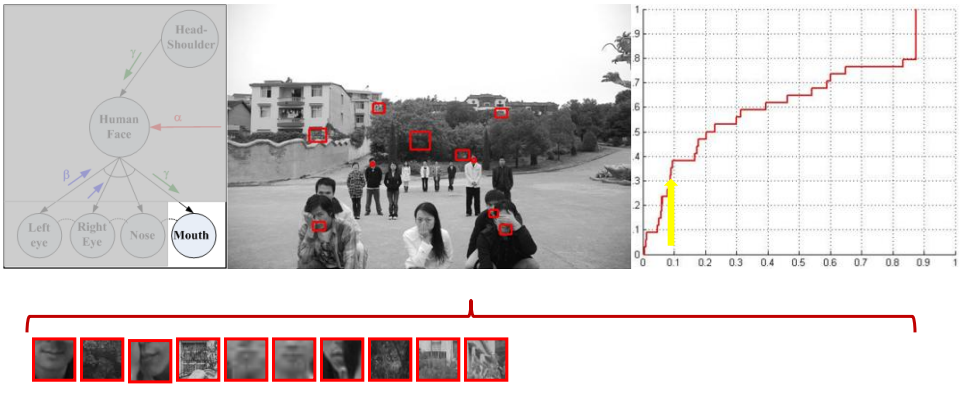
α–channel: nose



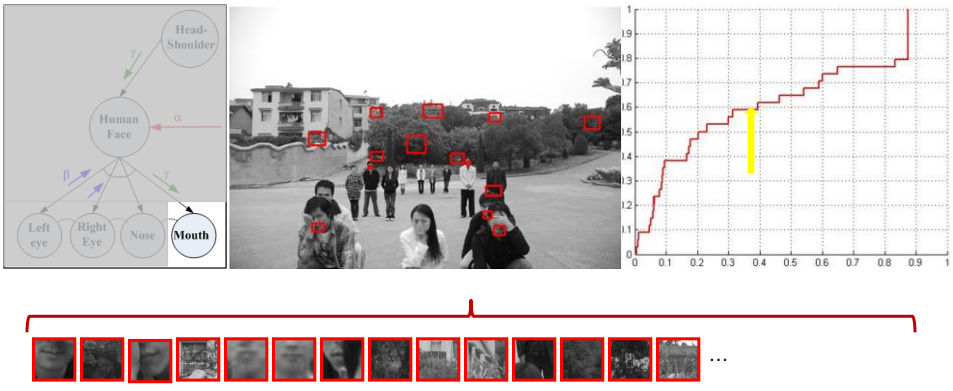
α–channel: mouth



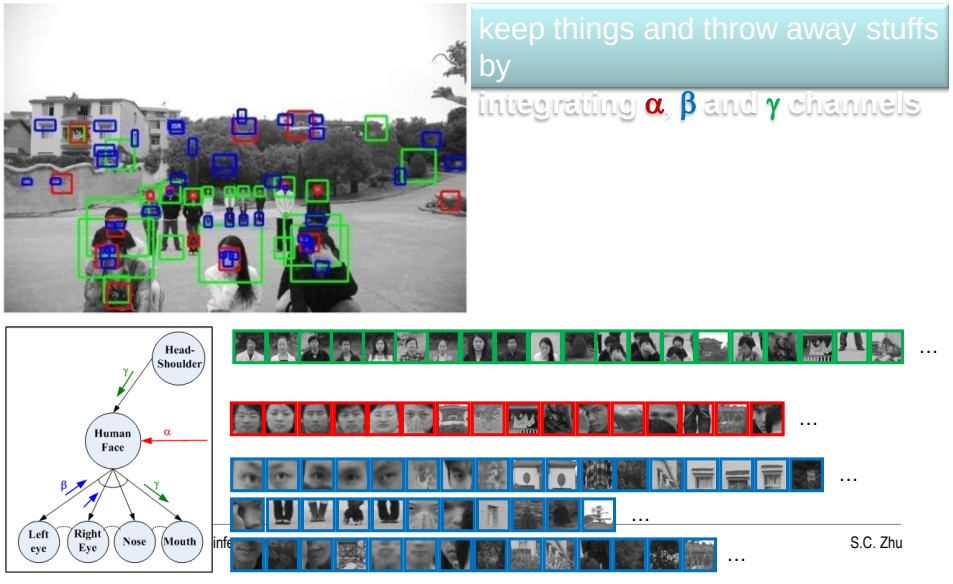
α–channel: mouth



α–channel: mouth



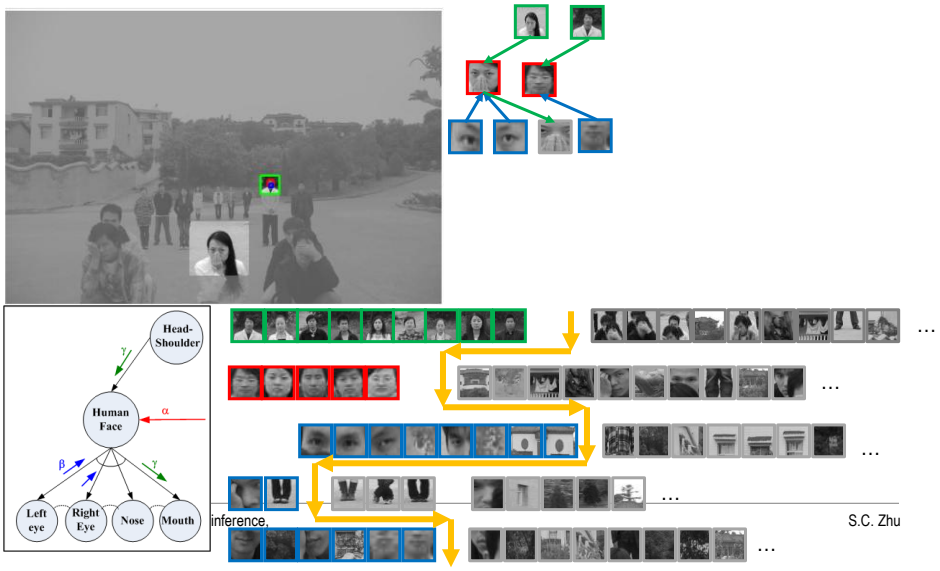
All α channels



Integrating α, β and γ channels

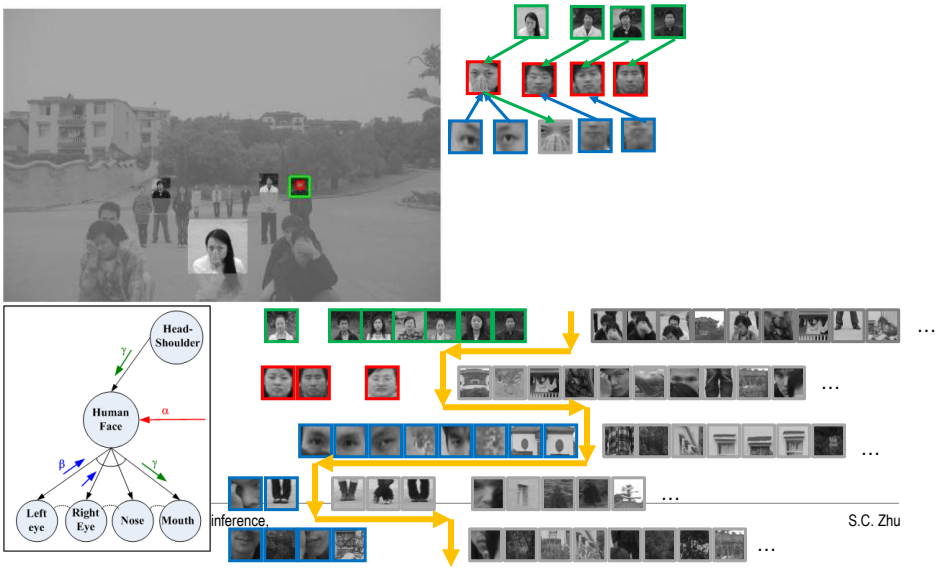
Integrating α, β and γ channels

Integrating α, β and γ channels

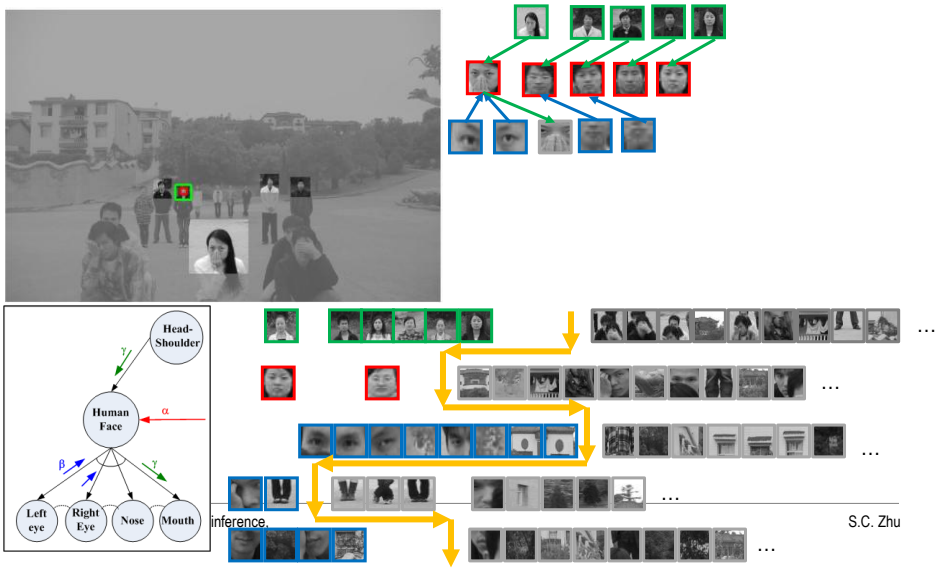


Integrating α, β and γ channels

Integrating α, β and γ channels



Integrating α, β and γ channels



Integrating α, β and γ channels

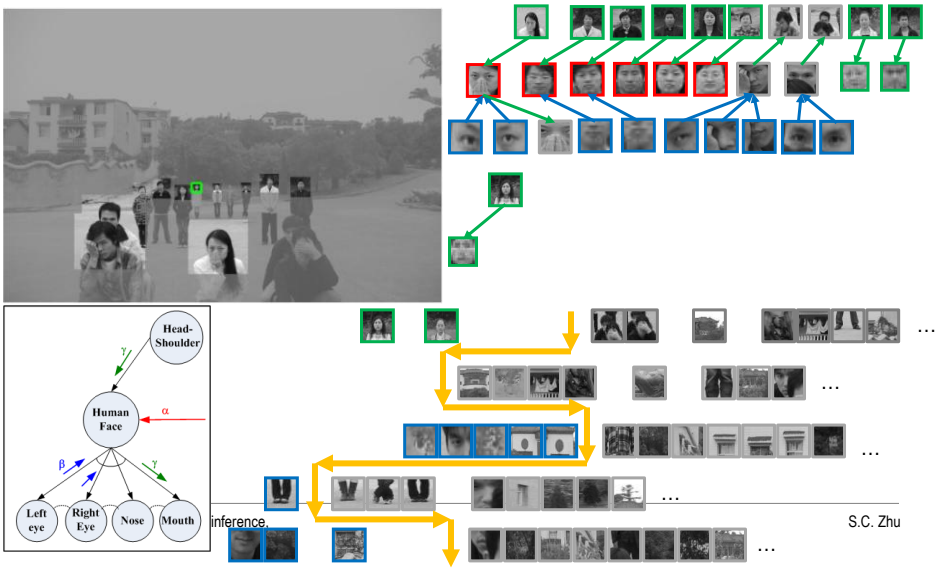
Integrating α, β and γ channels

Integrating α, β and γ channels

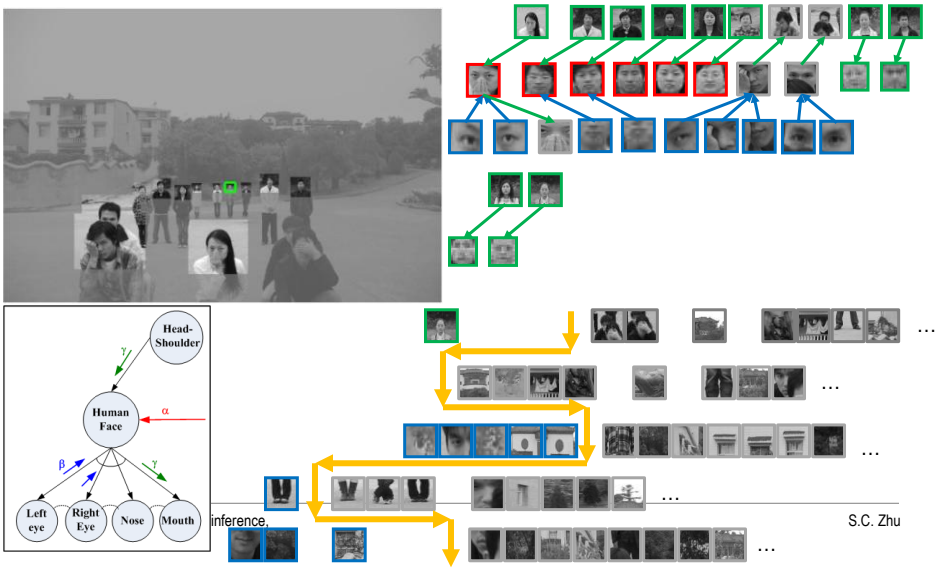
Integrating α, β and γ channels

Integrating α, β and γ channels

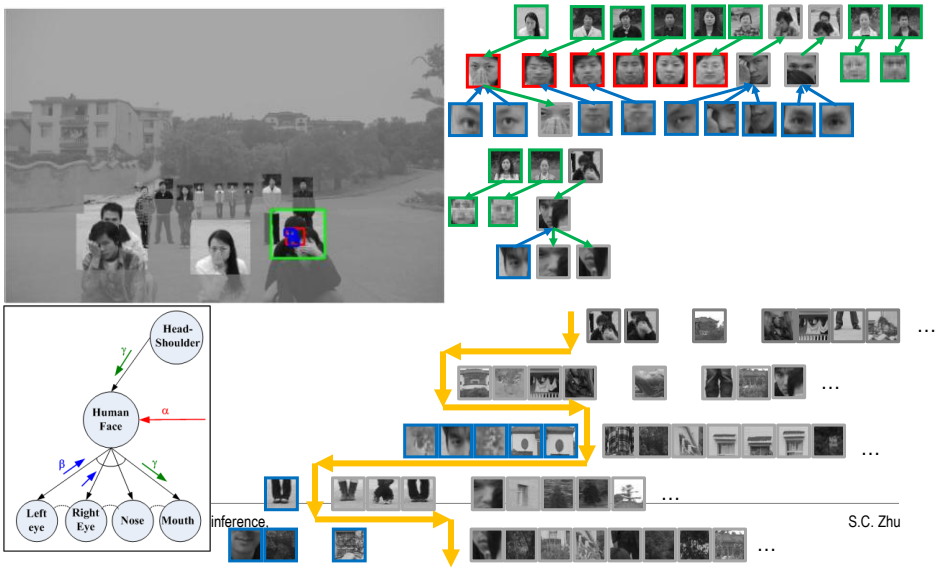
Integrating α, β and γ channels



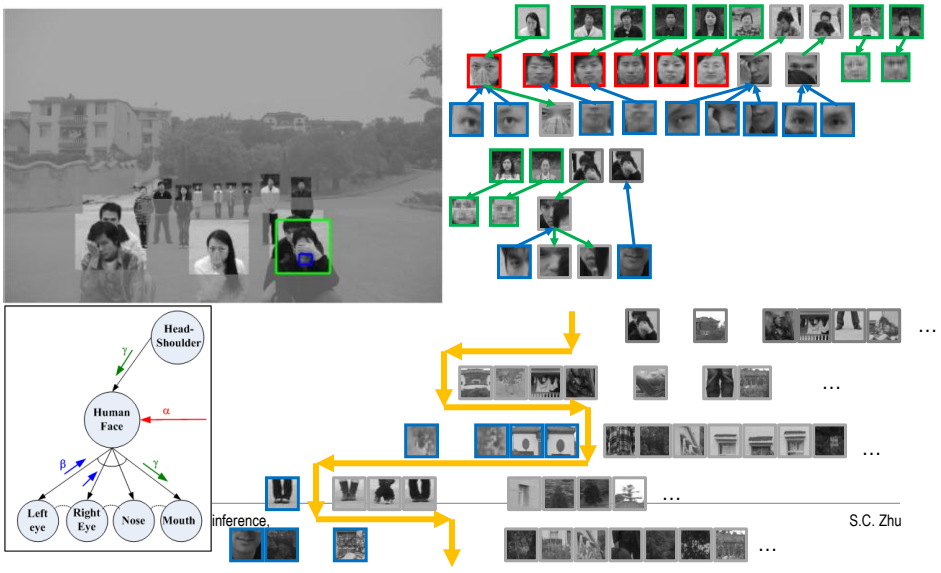
Integrating α, β and γ channels



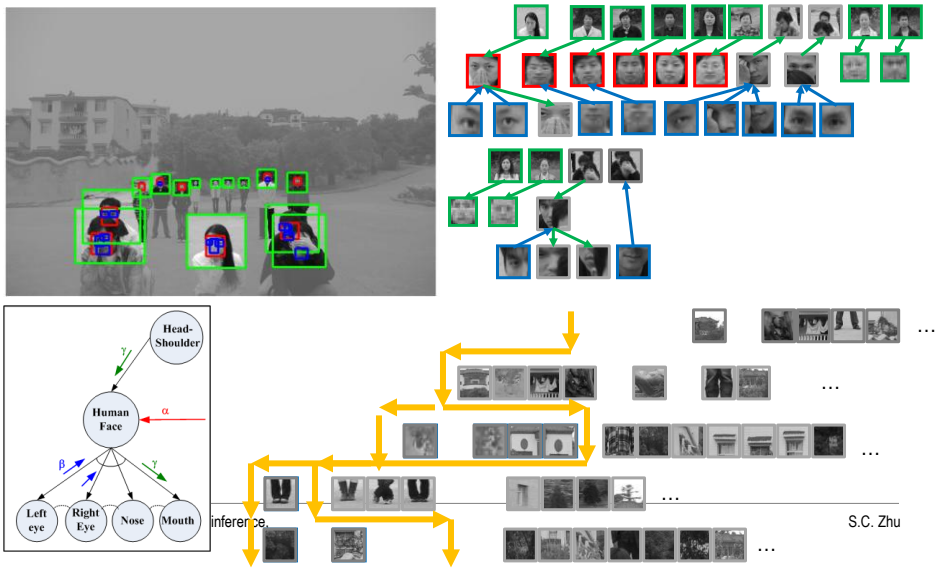
Integrating α, β and γ channels



Integrating α, β and γ channels

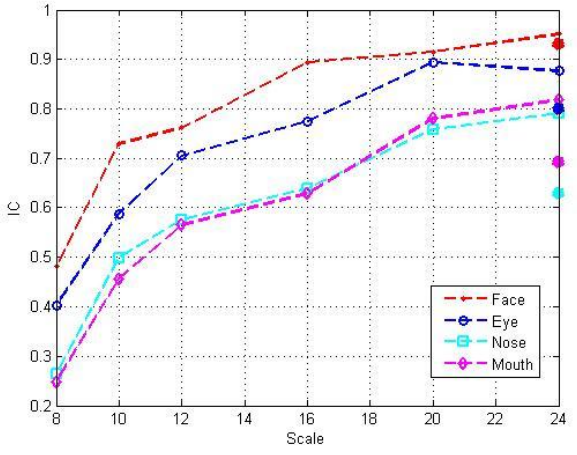


Integrating α, β and γ channels



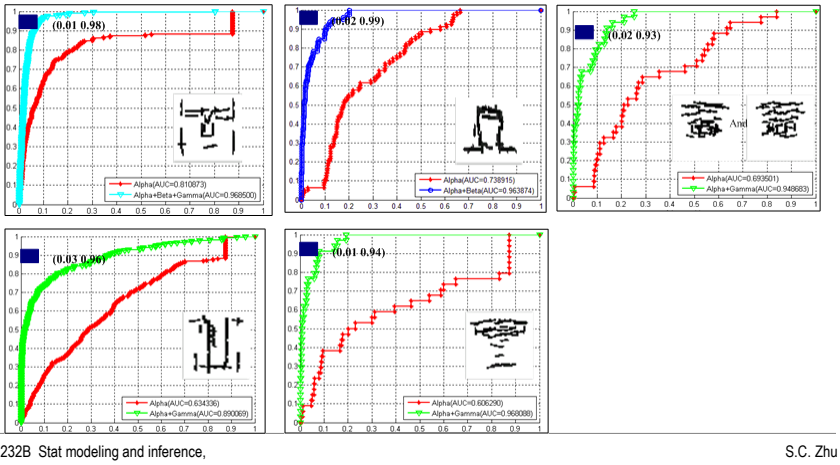
Information contribution

α channels



Performance improvement

red for α, blue for $\alpha+\beta$, green for $\alpha+\gamma$, cyan for $\alpha+\beta+\gamma$ channels



Top-down / Bottom-up Inference at all levels

Objective: Constructing parse graphs on-line !

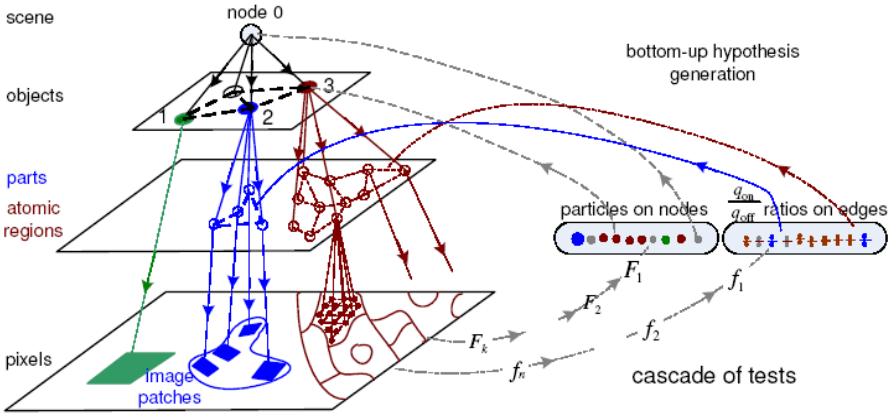
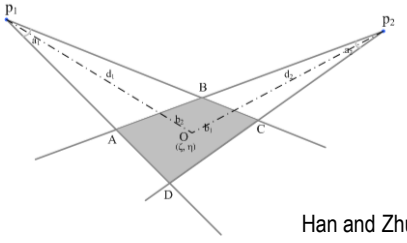


Image parsing by DDMCMC, Tu et al, 2002-05

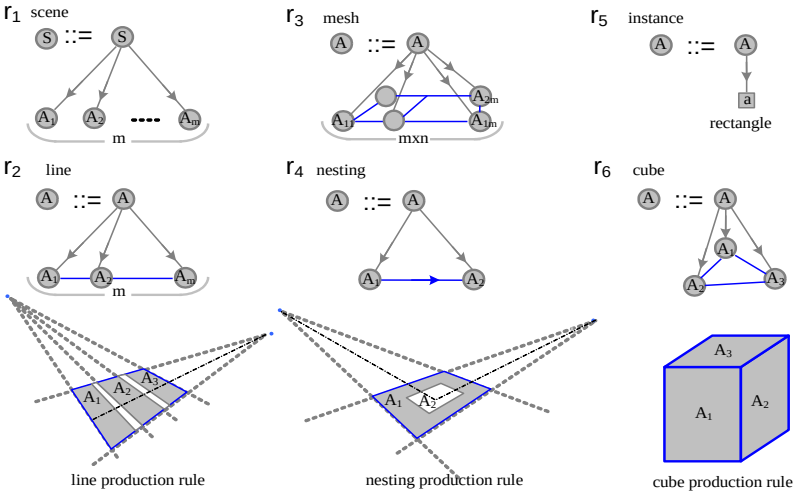
A simpler and more flexible graph grammar

One terminal sub-template
--- a planar rectangle in 3-space

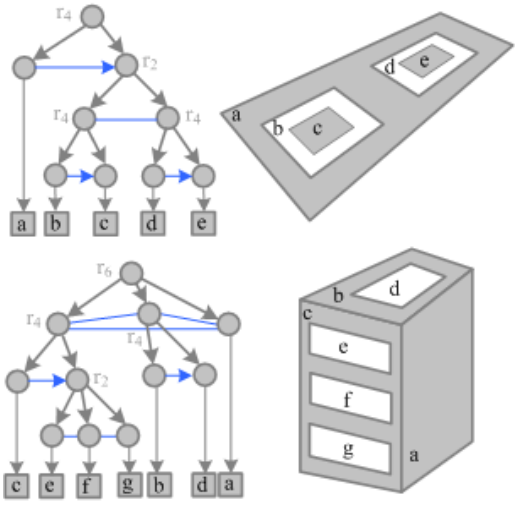


Han and Zhu 2005

Six grammar rules which can be used recursively

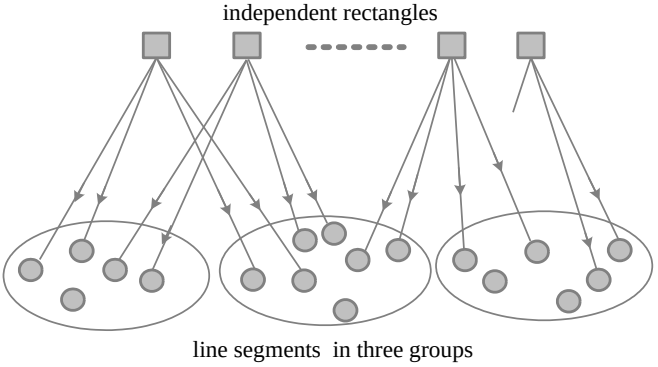


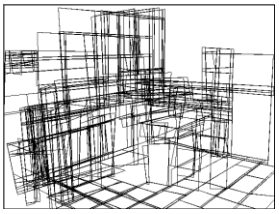
Two configuration examples



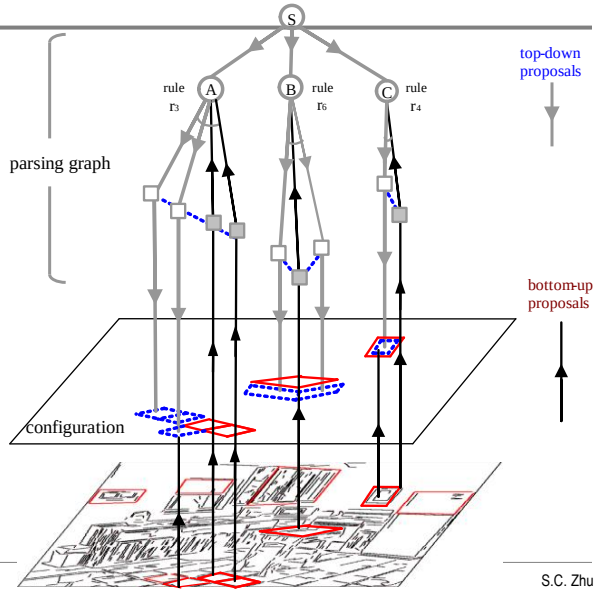
Bottom-up detection (proposal) of rectangles

Each rectangle consists of two pairs of line segments that share a vanish point.



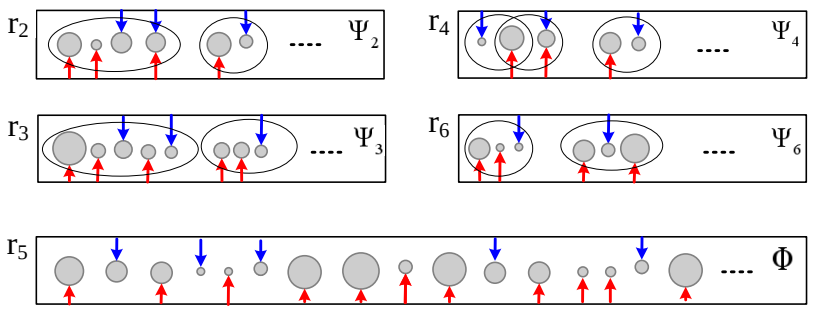


Top-down / bottom-up inference

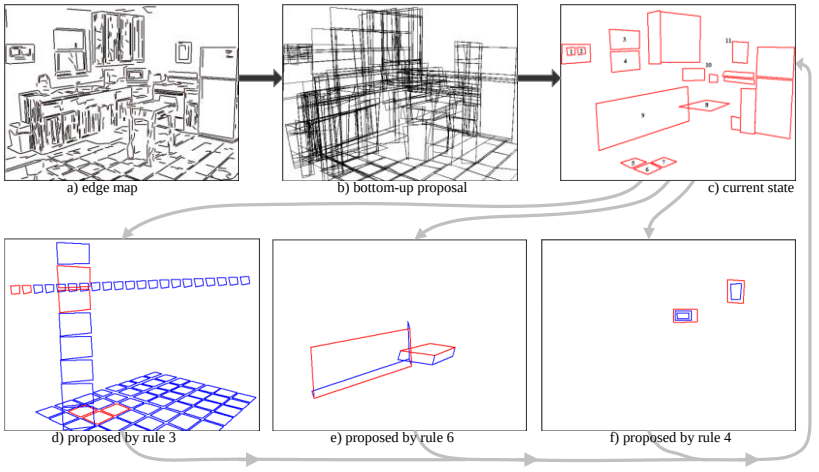


Each grammar rule is an assembly line and maintains an Open-list and Closed-list of particles

A particle is a production rule partially matched, its probability measures an approximated posterior probability ratio.

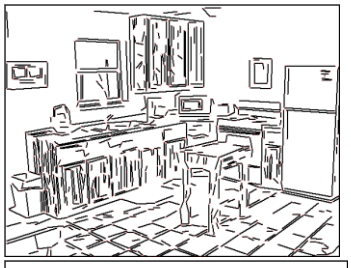


Example of top-down / bottom-up inference

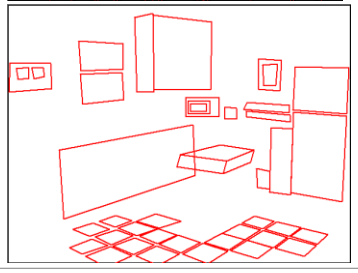


Results

(Han and Zhu, 05)

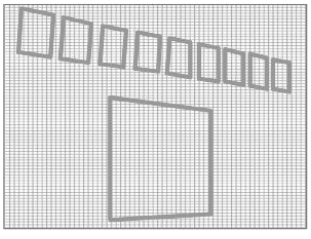


Edge map



Rectangles inferred

Likelihood model based on primal sketch



$$\Lambda = \Lambda_{sk} \cup \Lambda_{nsk},$$

$$\Lambda_{sk} = \cup_{k=1}^{N} \Lambda_{sk,k}$$

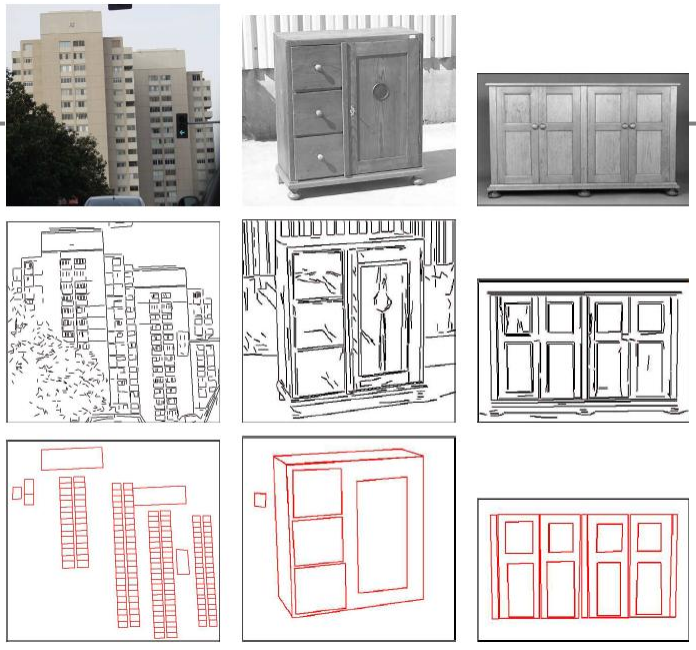
$$\Lambda_{nsk} = \cup_{m=1}^{M} \Lambda_{nsk,m},$$

$$\Lambda_{nsk,m_1} \cap \Lambda_{nsk,m_2} = \emptyset, m_1 \neq m_2$$

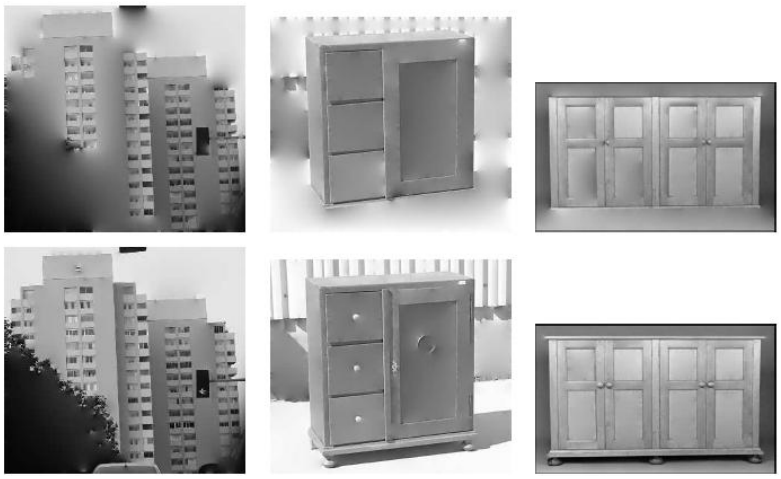
$$p(\mathbf{I}_{sk,k}|C) \propto \exp\{-\sum_{(x,y)\in\Lambda_{sk,k}} \frac{(\mathbf{I}(x,y) - B_k(x,y))^2}{2\sigma^2}\}$$

$$p(\mathbf{I}|C(\mathbf{G})) = \frac{1}{Z}\exp\{-\sum_{k=1}^{N}\sum_{(x,y)\in\Lambda_{sk,k}} \frac{(\mathbf{I}(x,y) - B_k(x,y))^2}{2\sigma^2} - \sum_{m=1}^{M}\sum_{i=1}^{n}\langle\beta_{mi}, h_i(\mathbf{I}_{\Lambda_{nsk,m}})\rangle\}$$

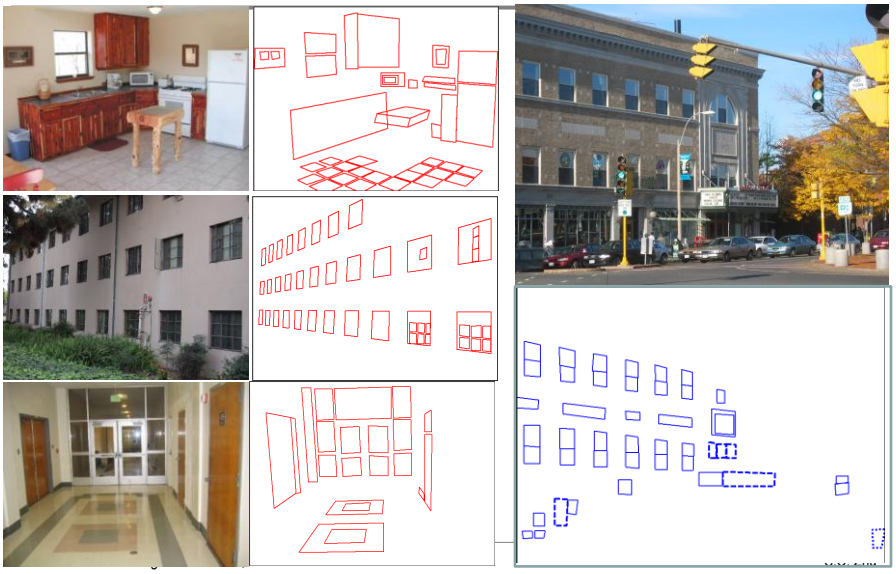
Synthesis based on the parsing model



Synthesis based on the parsing model



Parsing rectangular scenes by grammar



How much does top-down improve bottom-up?

In the rectangle experiments:

Han and Zhu, 2005-07