Generative models

Song-Chun Zhu

Modeling

One fundamental problem, in my opinion, that we are facing today is to

represent the enormous amount of visual knowledge

needed at all levels of vision.

More specifically, it is to *distill* data (raw images, video) into visual knowledge in terms of various kind of models.

modeling = visual knowledge representation

Visual Knowledge

There are two kinds of visual knowledge

1. Representational knowledge:

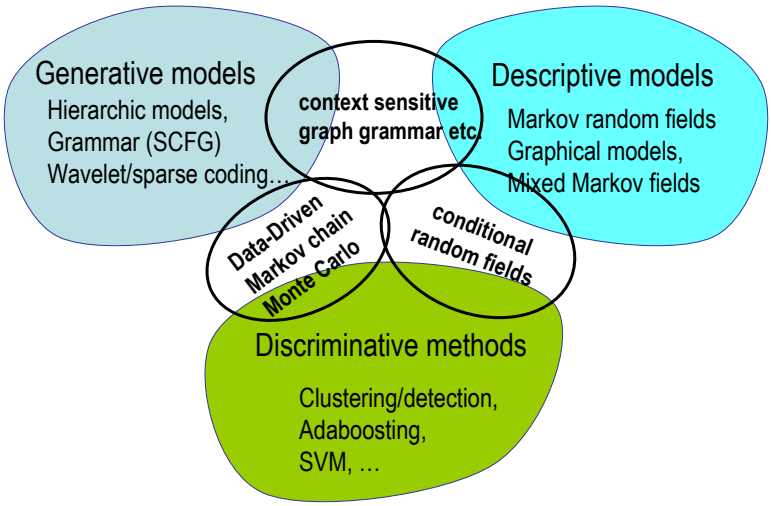
   (a) Visual vocabulary ---hierarchical generative models
   --- dictionaries at all levels of vision

   (b) Spatial relations and context --- descriptive models, such as MRF

2. Computational knowledge: --- discriminative models

   discriminative features for various variables
   ordering of bottom-up tests …

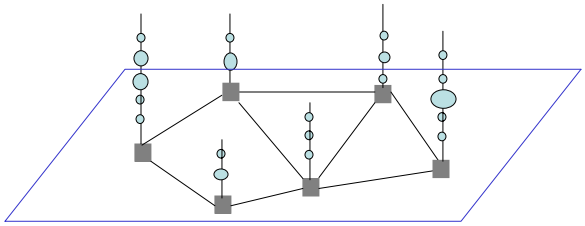
Three families of models in vision



Descriptive models

1. A ***flat graph*** representation G=<V, E>.
   G could be directed, undirected, such as chain, tree, DAG, lattice, etc.
   the nodes in the graph are at the same semantic level.

2. hard constraints or soft "energy" between vertices for ***regularity*** and ***context***



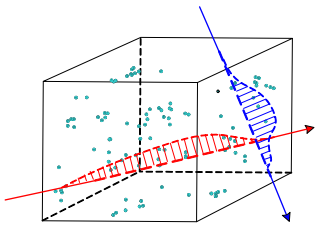
Examples,
constraint-satisfaction, line drawing interpretation, scene labeling, deformable templates, image restoration, segmentation, graph partition/coloring, shape from stereo/motion/shading

Learning descriptive models

Conceptualization and Modeling with Descriptive models

$$X \sim \Omega(h) = \{X : E_p[h_i(X)] = \mu_i, i = 1, 2, ..., K\}.$$

$$X = (x_1, x_2, ..., x_N)$$

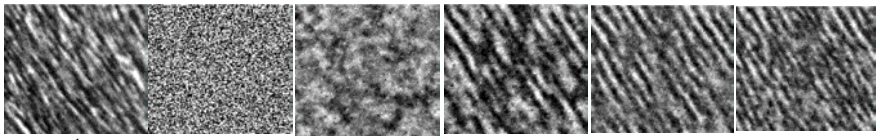


Examples:
Gibbs,
MRF,
FRAME,
Mixed Markov model

Model pursuit: choose informative features and statistics to minimize the log-volume of the set or Shannon entropy of the model --- minimax entropy principle.

Example

Markov random fields and Gibbs models on pixels



I^{obs} $I^{syn} \sim \Omega(h)$ k=0 $I^{syn} \sim \Omega(h)$ k=1 $I^{syn} \sim \Omega(h)$ k=3 $I^{syn} \sim \Omega(h)$ k=4 $I^{syn} \sim \Omega(h)$ k=7

The same procedure has been applied to various of levels of context models:

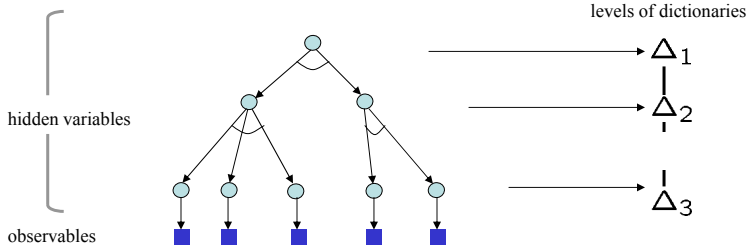
textures, texton process, simple shapes, scene context, …

Generative model

A ***hierarchic graph*** representation

the nodes in the graph are at different semantic levels.

the edges show the decomposition.



Examples,

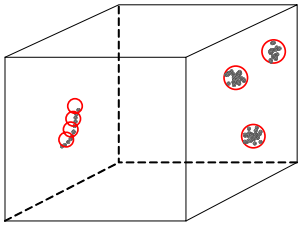
wavelets, sparse coding, stochastic context free grammar, …,

Learning generative models

A generative model is a joint probability with a series of dictionaries

$$p(I, X; \Delta) = p(I|X_k; \Delta_k) \cdots p(X_2|X_1; \Delta_1) p(X_1; \beta)$$

It has to be a descriptive model



Key concepts:

A dictionary is a set of epsilon balls (manifolds)

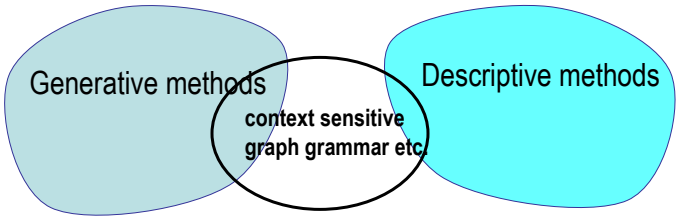
A hidden variable is an index to the dictionary.

Model pursuit: choose optimal dictionary (*epsilon-balls*) to cover the maximum probability mass or minimize the Kolmogorov entropy

--- again, minimax entropy

It is closely related to binding with maximum mutual information.

Integrated generative model



Necessity is obvious:

we need both hierarchic *composition* and *context*.

But how?

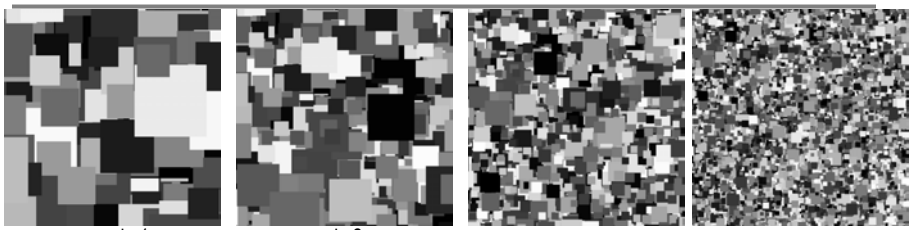
1. When do we represent regularity by a *constraint* (descriptive) or by a *production rule* (generative)?
2. How to handle the continuous transition between texture (descriptive) and structures/shapes (generative)?

Leaves at a range of scales

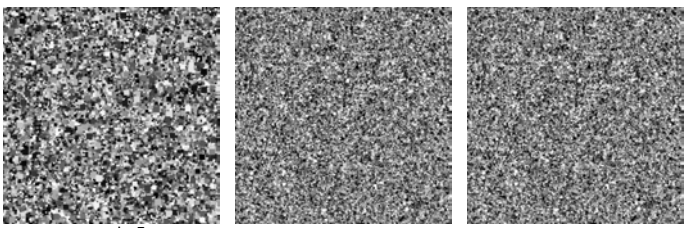
This picture contains trees/leaves at four ranges of distance, over which our perception changes.

A: see individual leaves with sharp edge/boundary (occlusion model)
B: see leaves but blurry edge (additive model)
C: see a texture impression (MRF)
D: see constant area (iid Gaussian)

Model regime transitions in scale space



scale 1 scale 2 scale 3 scale 4



scale 5 scale 6 scale 7

We need a seamless transition between different regimes of models

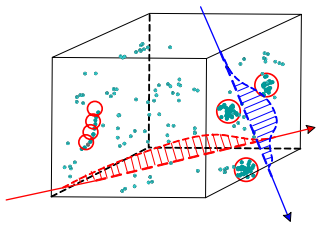
By analogy: A picture of the universe

At different temperatures we observe different entropy patterns.

At different scales, different forces rule the systems.

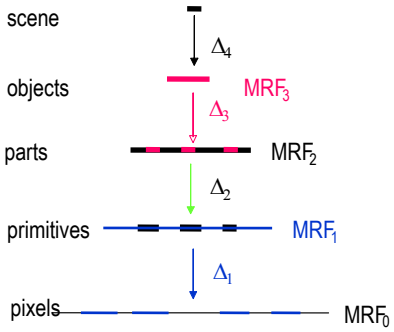
A photo from Cosmology. Our image space is very much like this, it contains patterns of wide range of entropy regimes.

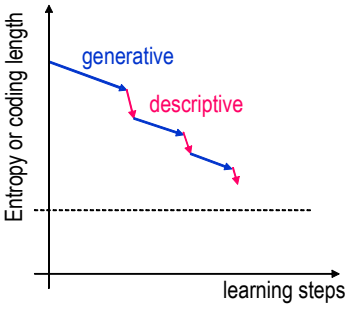
Integrated learning



Augmentation of the Integrated generative model

The pursuit of an integrated model is to minimize the Shannon and Kolmogorov entropy in turns,

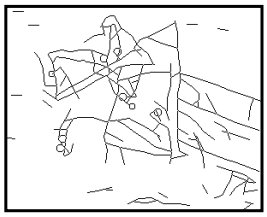




A low level example: primal sketch

org image

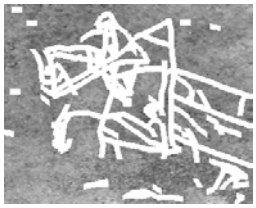
sketching pursuit process



sketches

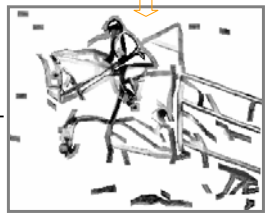
syn image

=



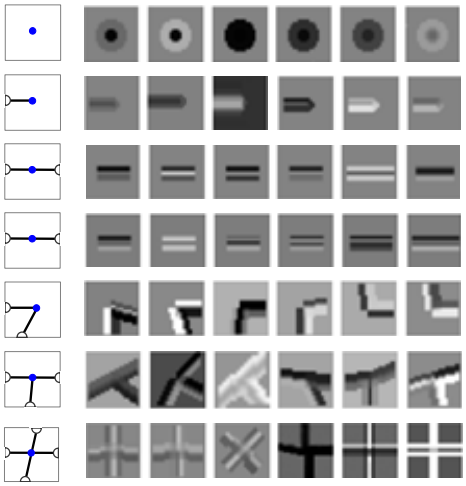
synthesized textures

+

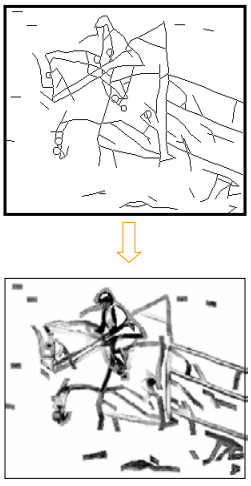


sketch image

Examples of the dictionary of image primitives

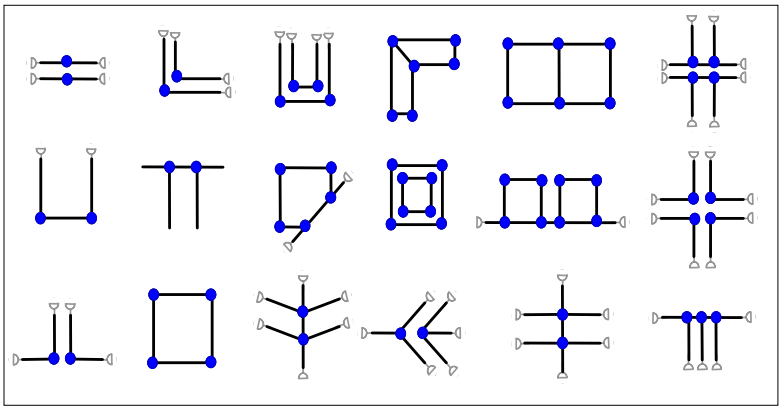


Example: primal sketch

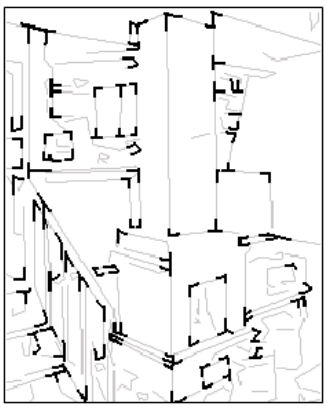


Middle-level example: the dictionary of graphlets

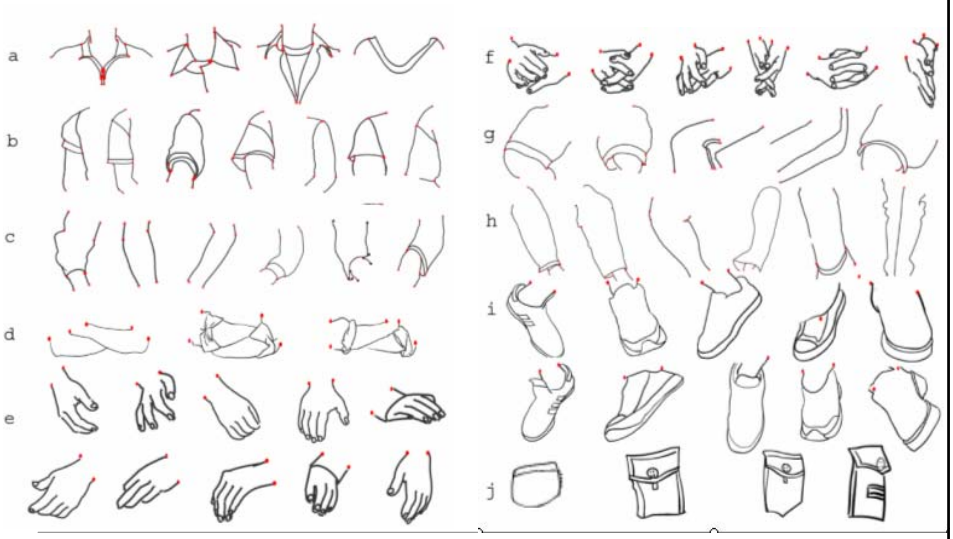
A graphlet is a graph composed of 2-8 image primitives with open bonds.



Middle level example

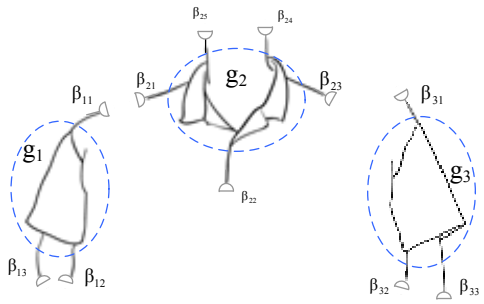


High level examples: the dictionary of human figure



Composing the parts

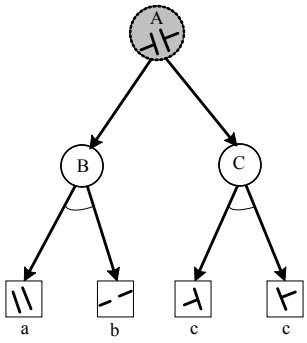
Each sub-template is a vertex in a composite template, and vertices are connected through “bonds”.

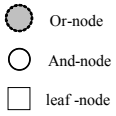


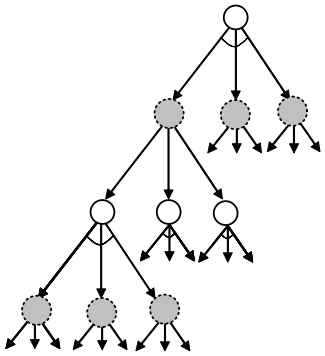
The ideas of binding, composition, hierarchical, and reusable parts

A grammar rule:

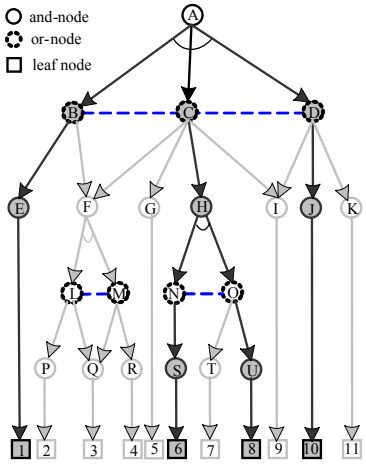
$A \rightarrow ab \mid cc$



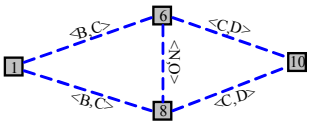


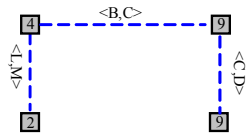


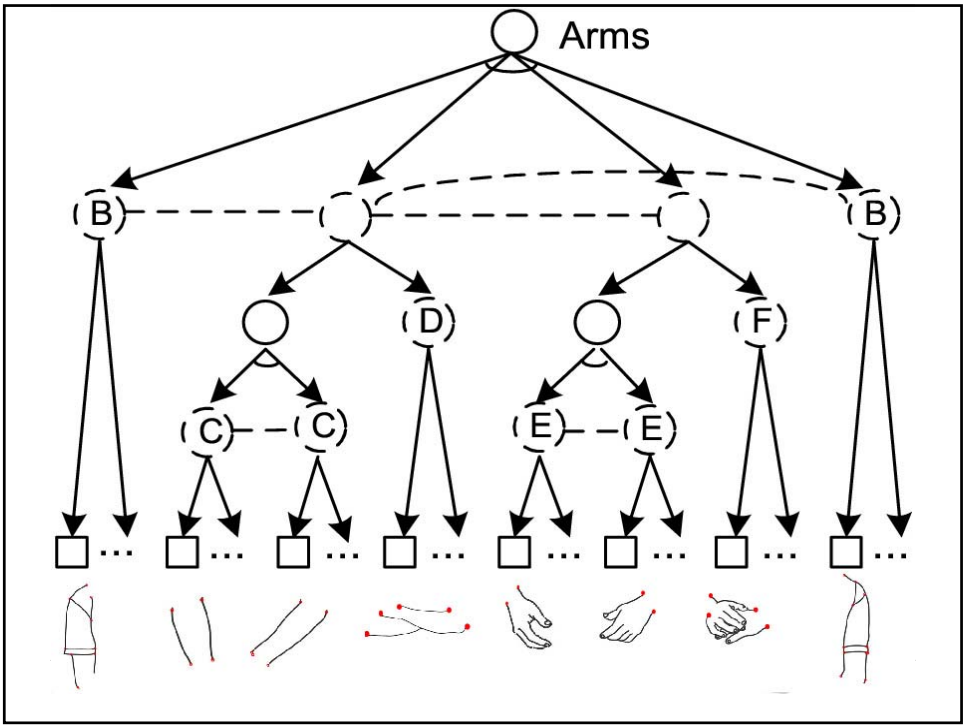
Embedding the integrated models into an And-Or graph



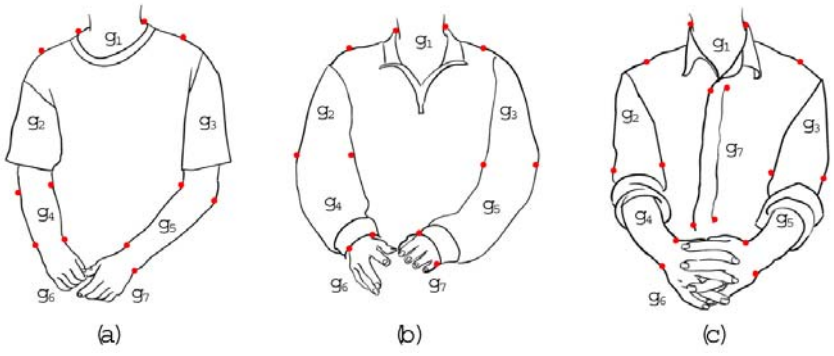
some graph configurations generated by the And-Or graph





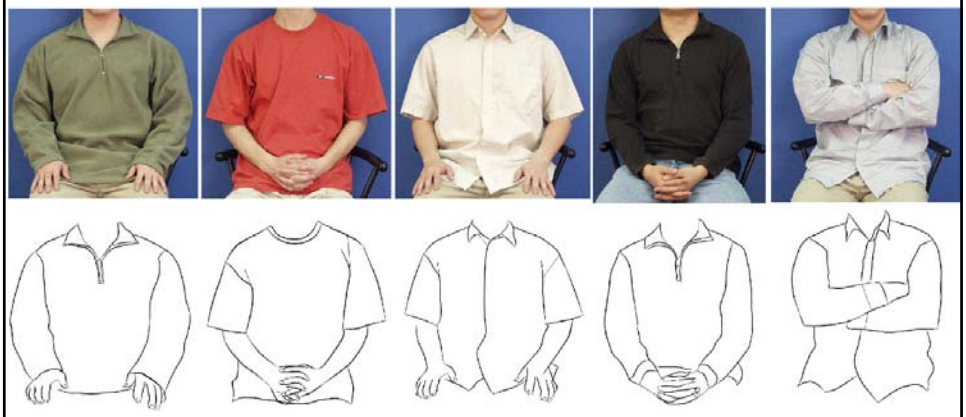


Examples of new configurations (synthesis)



Note that the number of sub-templates and their connections change. Each is a possible "configuration".

Examples of human upper-cloth representation



Examples of human upper-cloth representation

Back to history

In the 1980s, we have a popular model for low and middle level vision
(Geman and Geman, Blake and Zisserman, Koch and Poggio, Mumford and Shah)

$$p(I, B) = \frac{1}{Z} \exp\{-\mu \int \int_{\Lambda \setminus B} |\nabla I|^2 dxdy - \lambda |B|\}.$$

Three questions:

1. *Why is the potential quadratic?*
2. *Why is the gradient operator?*
3. *Where is the concept of edge from?*