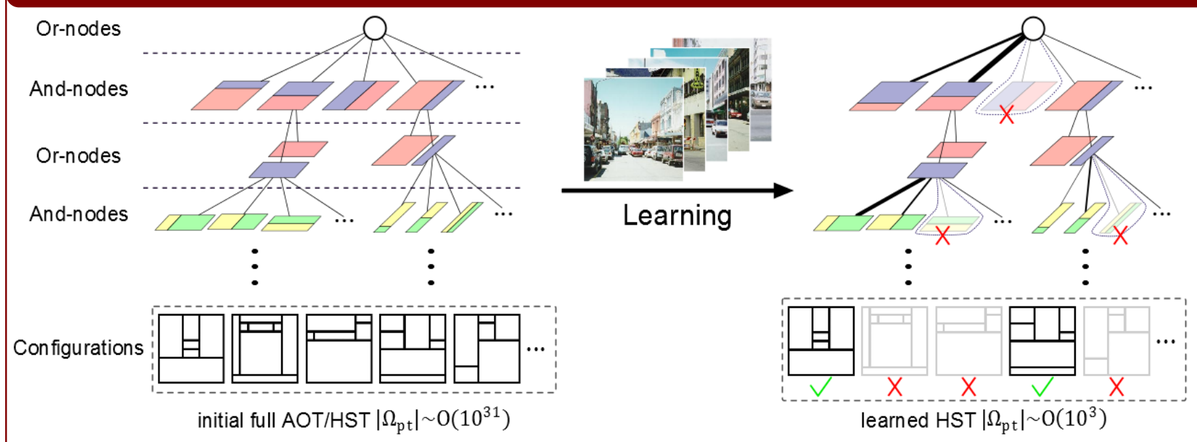
Hierarchical Space Tiling for Scene Modeling

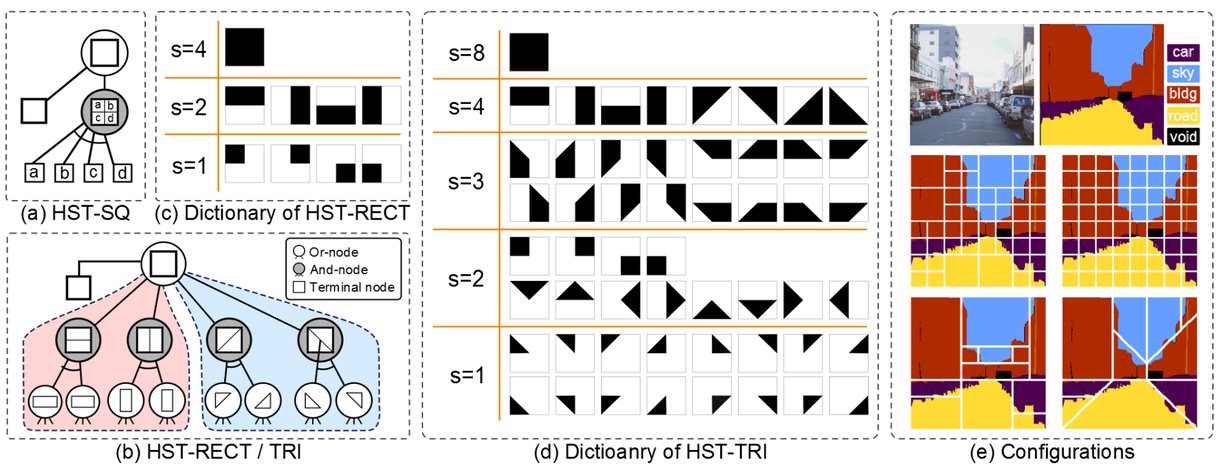
Shuo Wang[1,2], Yizhou Wang[1] and Song Chun Zhu[2]
[1]Nat'l Engineering Lab for Video Technology, Peking University
[2]Center for Vision, Cognition, Learning and Arts, UCLA

Introduction



A typical scene category, e.g., street and beach, contains an enormous number ($O(10^4)$ ~ $O(10^5)$) of distinct scene configurations that are composed of objects and regions of varying shapes in different layouts. A well-known representation that can effectively address such complexity is the family of compositional models. The objective of this paper is to present an efficient method for learning such models from a set of scene configurations. We start with an over-complete representation named *Hierarchical Space Tiling (HST) which quantizes the huge and continuous scene configuration space*. Then estimate the HST model through a learning-by-parsing strategy.

Representation



- We *define HST on an And-Or tree (AOT)*. An *And-node* represents a way of decomposing a region; an *Or-node* represents alternative decompositions with branching probabilities; and a *terminal node* is an element, such as squares and rectangles.
- By selecting the branches at Or-nodes, a parse tree pt can be derived, the energy of pt is:

$$E(pt|C;\Theta,\Delta) = \sum_{v\in V_{pt}^{OR},v_i\in Ch(v)} E^{OR}(v_i|v) + \lambda \sum_{v\in V_{pt}^{T}} E^{T}(C_v|v)$$

- *Three types of HST:* square tiling (e.g., Quadtree and Spatial Pyramid), rectangular tiling (HST-RECT), triangular tiling (HST-TRI)

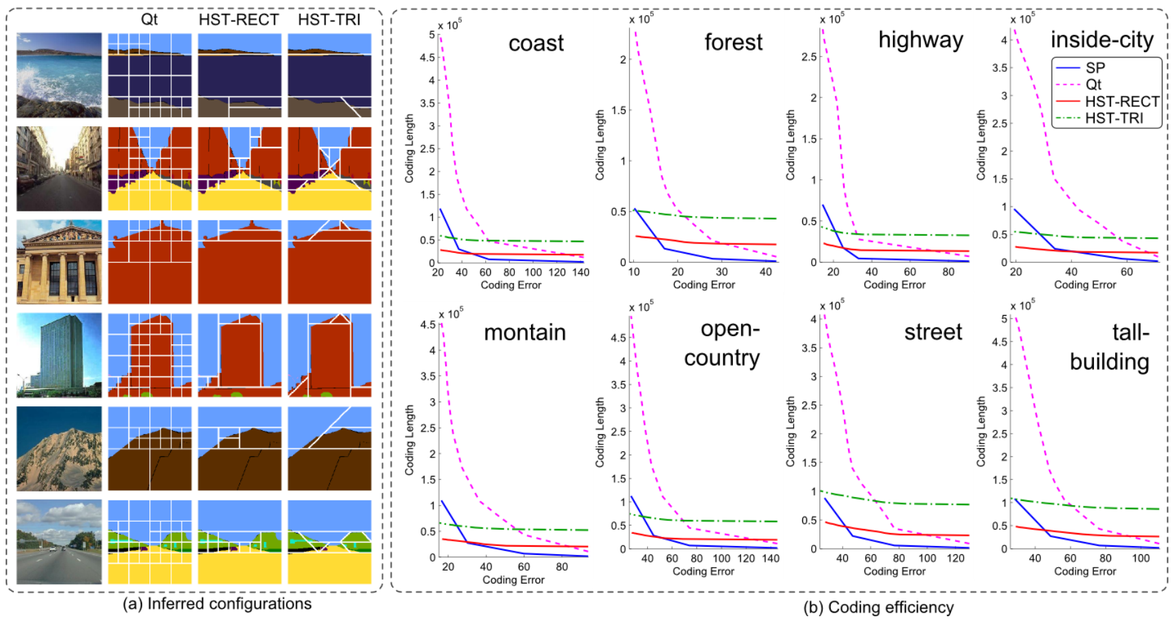
Learning

Given a set of scene configurations (scene label maps), we present a *learning-by-parsing (EM-like) method* to learn the HST representation including the branching probability at Or-nodes and the tiling dictionary.

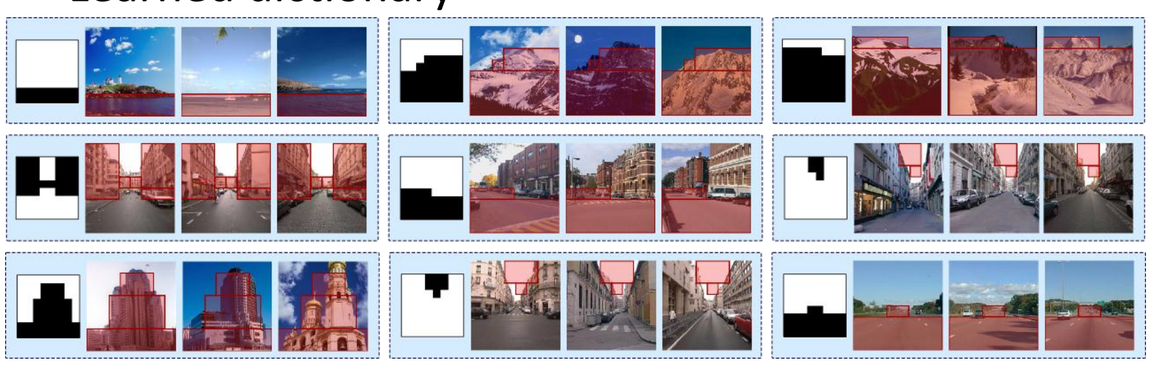
- E-Step: fix HST and infer parse tree by dynamic programming
- M-Step: based on the inferred parse trees, estimate HST by Maximum Likelihood Estimation (MLE)
- Prune branches which has very low branching probability

Experiments

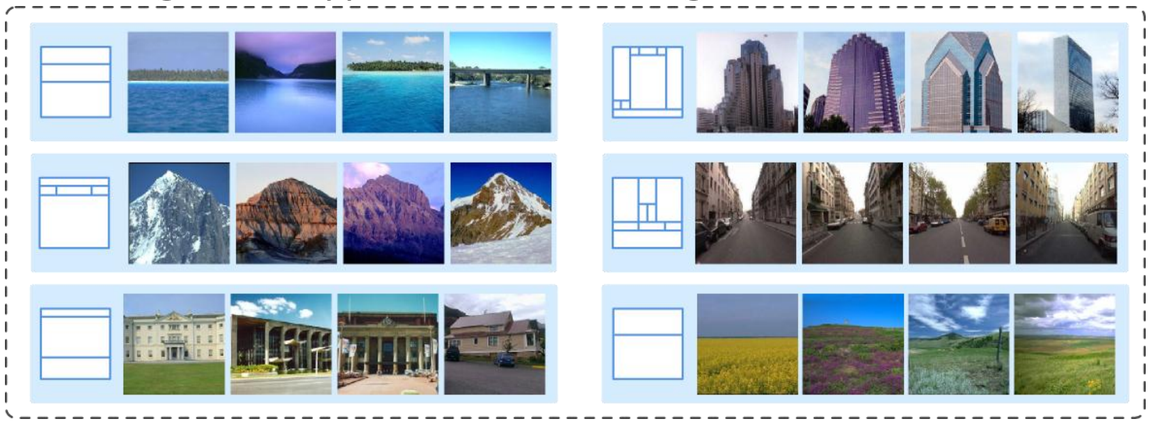
- Coding efficiency: coding error w.r.t coding length



- Learned dictionary



- Categorical typical scene configurations



- Scene classification: 8 categories outdoor scene

Methods	Gist	BoW	SPM	LLC	Tangram	Ours
AP(%)	72.15	84.57	84.92	87.97	86.07	**91.71**