Learning Top-Down Grouping of Compositional Hierarchies for Recognition

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Björn Ommer and Joachim M. Buhmann
Inst. of Computational Science
ETH Zurich

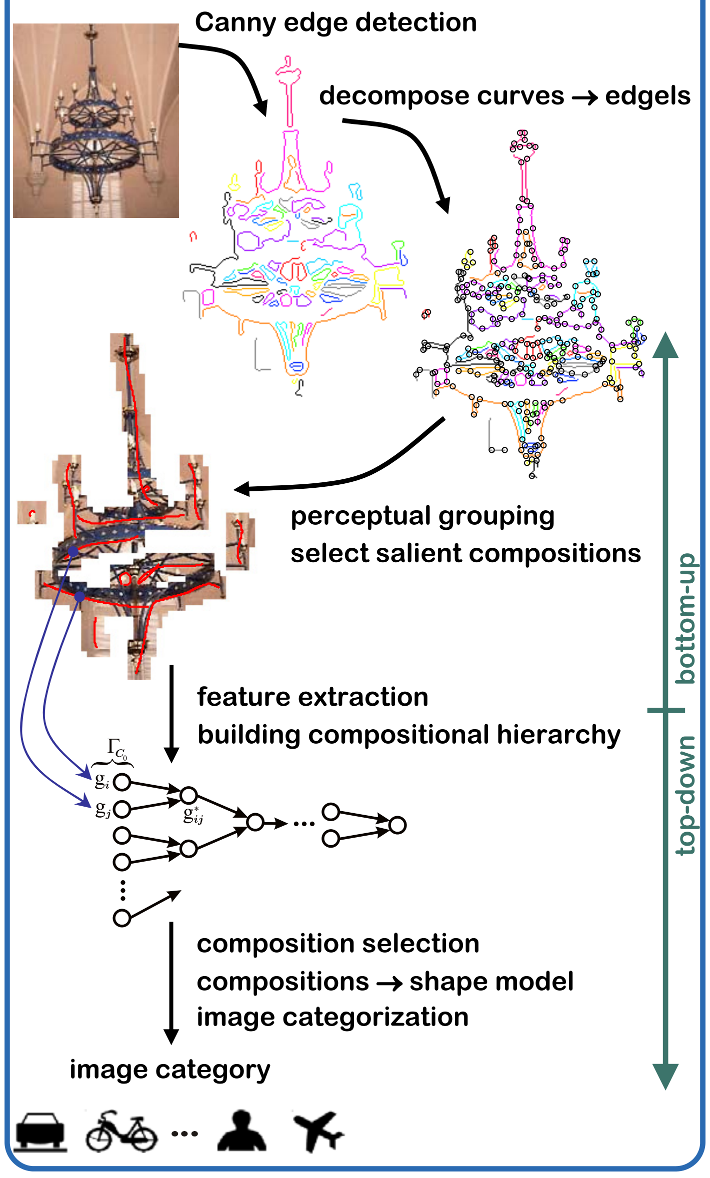
1 Compositional Approach — Goals

- Compositions establish an intermediate image representation
- Automatic construction of full hierarchies of compositions
- Learning of compositional models without hand segmentations (only train images + overall labels)
- Dealing with large intra-category variations in Caltech 101 database

Methodology of the compositional approach:

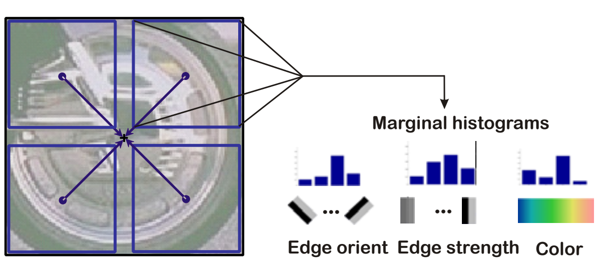
I) Perceptual grouping yields initial set of salient compositions & limits representation complexity
II) Top-down grouping forms category distinctive composition hierarchies
III) Automatic learning of top-down grouping probabilities without information on compositions in training images
IV) Spatial coupling of compositions using probabilistic shape model

2 Recognition Phase



3 Localized Feature Histograms

Along grouped curve segments, features are extracted as local part descriptors:



Local descriptor is Gibbs distrib. over codebook:

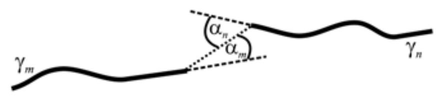
$$P(F_i = \nu | \mathbf{e}_i) := Z(\mathbf{e}_i)^{-1} \exp\left(-d_\nu(\mathbf{e}_i)\right)$$

$$Z(\mathbf{e}_i) := \sum_\nu \exp\left(-d_\nu(\mathbf{e}_i)\right)$$

Compositions represented as bags of parts:

$$\mathbf{g}_j \propto \sum_{i=1}^{m} \Big(P(F_i = 1|\mathbf{e}_i), \dots, P(F_i = k|\mathbf{e}_i) \Big)^T$$

4 Perceptual Bottom-Up Grouping



Group decomposed curves $\gamma_m, \gamma_n \rightarrow \gamma_g$ recursively according to the Gestalt laws of good continuation, proximity, and convexity:

Maximize criterion function $\tilde{\zeta}(\gamma) := \left| 4\pi \frac{A(\gamma)}{l^2(\gamma)} - \frac{1}{2} \right|$ under the constraint

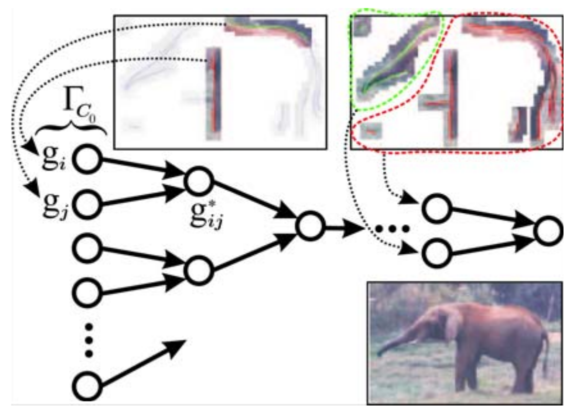
$$\text{discard } \gamma_g \Leftrightarrow \begin{cases} \min\{\alpha_m, \alpha_n\} > 90° \vee \\ \text{gap}(\gamma_m, \gamma_n) > \min\{l(\gamma_m), l(\gamma_n)\} \vee \\ \tilde{\zeta}(\gamma_g) < \min\{\tilde{\zeta}(\gamma_m), \tilde{\zeta}(\gamma_n)\} \end{cases}$$

5 Applying Top-Down Grouping

1. Start with set Γ_C of salient compositions from perceptual bottom-up grouping
2. Recursive grouping of compositions using previously learned grouping statistics:

$$\mathbf{g}^\star_{ij} = \underset{\mathbf{g}_{ij}:\mathbf{g}_i,\mathbf{g}_j \in \Gamma_C}{\operatorname{argmax}} \max_{c \in \mathcal{L}} P(c|\mathbf{g}_{ij})$$

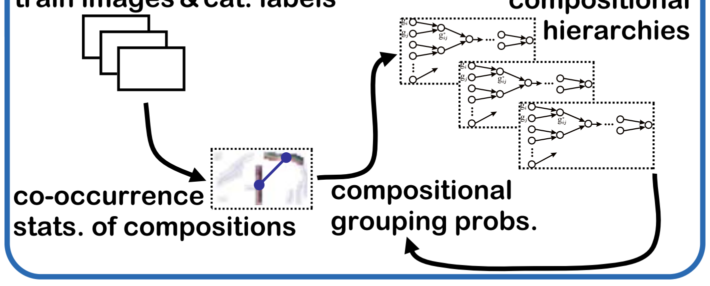
$$\Gamma_C \leftarrow \Gamma_C \cup \{\mathbf{g}^\star_{ij}\} - \{\mathbf{g}_i, \mathbf{g}_j\}$$



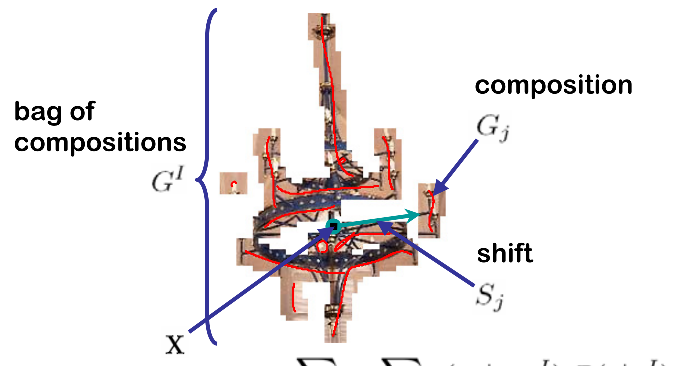
Finding local maxima of compositional hierarchy: Go from each leaf to the root and collect locally optimal compositions ⇨ relevant compositions that enter into shape model

6 Learning Top-Down Grouping

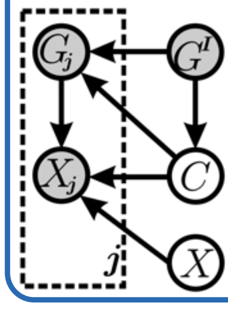
No information about compositional nature of objects in training data ⇨ bootstrap using estimated co-occurrence statistics of bottom-up groupings:



7 Shape Model for Binding Comps.



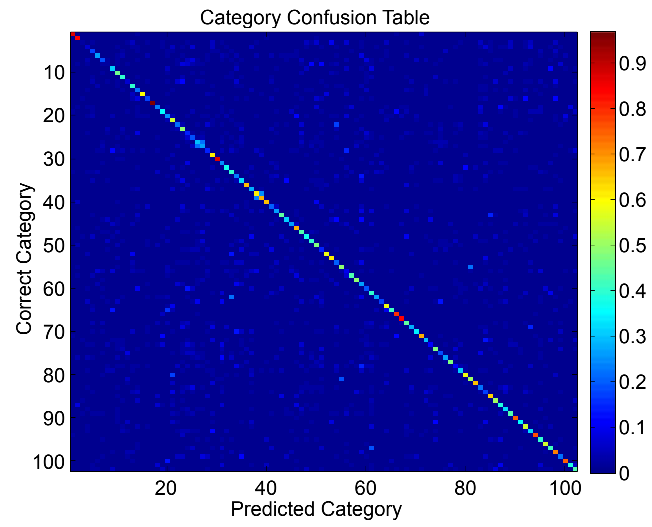
$$\mathbf{x} = \sum_j \mathbf{x}_j \sum_{c \in \mathcal{L}} p(\mathbf{g}_j | c, \mathbf{g}^I) \, P(c|\mathbf{g}^I)$$



$$P(c|\mathbf{g}^I, \mathbf{x}, \{\mathbf{g}_j, \mathbf{x}_j\}_{j=1:|\Gamma_L|}) = \exp\Big[(1 - |\Gamma_L|) \ln P(c|\mathbf{g}^I) + \sum_{\mathbf{g}_j \in \Gamma_L} \ln P\left(c | S_j = \mathbf{x} - \mathbf{x}_j, \mathbf{g}_j, \mathbf{g}^I\right)\Big]$$

8 Performance w/o Compositions

Baseline method using only a bag of features:



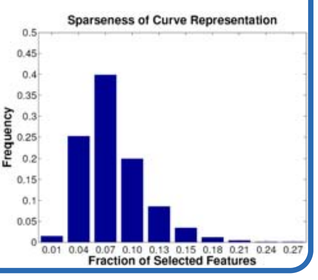
Retrieval rate for 200 prototypes: 41.3 ± 0.38 %

9 Perform. of Compositional Model

Retrieval rate (single-scale): 53 ± 0.5 %

Retrieval rate (multi-scale): 58 ± 0.8 %

Sparseness of induced image representation (fraction of selected features):



10 Establishing Category Similarity Hierarchies by Clustering Confusion Mat.