Learning the Taxonomy and Models of Categories Present in Arbitrary Images

Narendra Ahuja and Sinisa Todorovic
{n-ahuja, sintod}@uiuc.edu

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

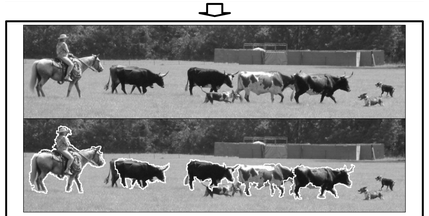
BECKMAN INSTITUTE
AT THE UNIVERSITY OF ILLINOIS

ICCV '07

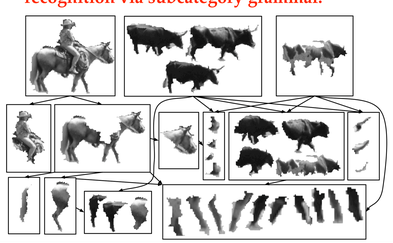
PROBLEM STATEMENT

GIVEN an arbitrary set of images, with:
1) ≥ 0 occurrences of ≥ 0 categories per image
2) Articulation, partial occlusion, clutter
3) Varying illumination
4) Small changes in scale, viewpoint

DISCOVER
all categories present, and
LEARN
the models and spatial taxonomy
of discovered categories.



In a new image
DETECT, RECOGNIZE, SEGMENT
all occurrences of the learned categories, and
EXPLAIN
recognition via subcategory grammar.

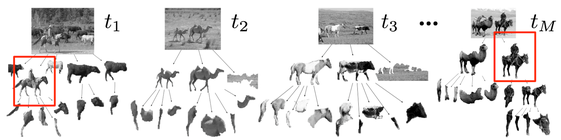


RECURSIVE DEFINITION OF CATEGORY

Category = Set of similar and frequently occurring 2D objects in the training set.

Similarity is defined in terms of subcategory properties:
- Geometric: area, shape
- Photometric: gray-level contrast, variance
- Topological: spatial layout, containment

LOCATING SIMILAR OBJECTS



1) Images = Segmentation trees [Ahuja PAMI 96]

2) Region properties computed relative to parent regions
⇒ Rotation-in-plane and scale invariance

3) Similar objects = Similar subtrees

4) Finding similarity values between subtrees:

GIVEN two trees t and t', FIND consistent subtree isomorphism $f : (v, v'),\ v \in t,\ v' \in t'$ which MAXIMIZES

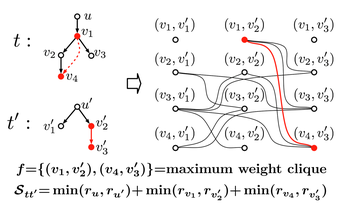
$$\mathcal{S}_{tt'} = \max_{f} \sum_{(v,v')\in f} \left[\underbrace{r_v + r_{v'}}_{\text{region saliency}} - \underbrace{|r_v - r_{v'}|}_{\text{cost of region matching}} \right]$$

similarity measure

$$r_v = \xi^{\mathrm{T}} \psi_v, \quad \xi \geq 0, \quad \xi^{\mathrm{T}} \xi = 1$$

region properties — relative significance to recognition

Example:

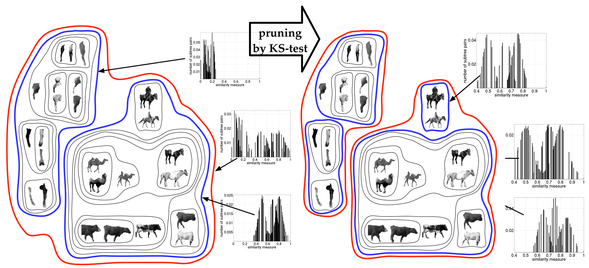


$f = \{(v_1, v'_2), (v_4, v'_3)\}$ = maximum weight clique

$\mathcal{S}_{tt'} = \min(r_u, r_{u'}) + \min(r_{v_1}, r_{v'_2}) + \min(r_{v_4}, r_{v'_3})$

DISCOVERING ALL CATEGORIES OF ALL COMPLEXITIES

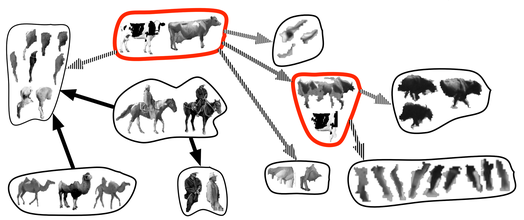
1) Agglomerative binary clustering of all regions from all images
2) KS-test with sensitivity α=5% for selecting valid clusters
3) Valid cluster = Discovered category



TAXONOMY OF ALL DISCOVERED CATEGORIES:

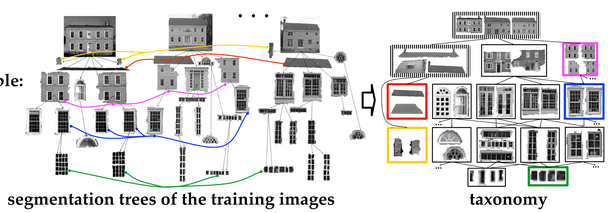
Region containment to category containment

Region co-occurrence to category co-occurrence



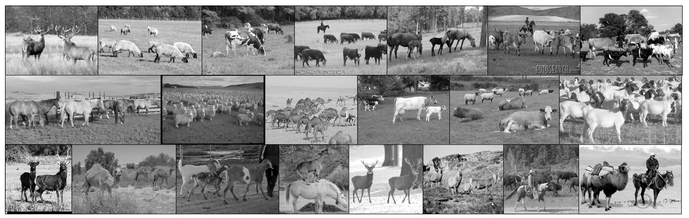
Co-occurrence category ■ = Layout of co-occurring categories

Example:



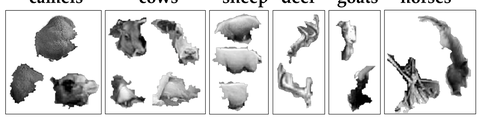
segmentation trees of the training images — taxonomy

EVALUATION ON A NEW DATASET -- HOOFED ANIMALS:
http://vision.ai.uiuc.edu/~sintod/datasets.html

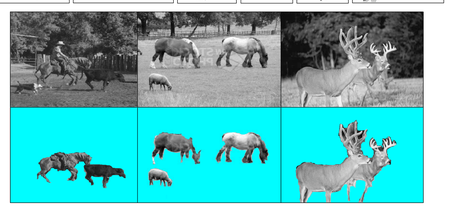


Convenient for evaluating cross-category resolvability, and subcategory sharing

camels cows sheep deer goats horses



Learned discriminative (non-shared) subcategories



Simultaneous detection, recognition, segmentation, and semantic explanation

1) Recognition is invariant to: translation, in-plane rotation, articulation, partial occlusion, clutter, small variations in scale, illumination, and viewpoint.

2) Object segmentation is good even on jagged and blurred object boundaries forming complex topologies.

3) Approach robust because object boudnaries often coincide with segment boundaries.

CONTRIBUTIONS:

1) Recursive definition of a category via the appearance and spatial configuration of subcategories.

2) Some subcategories may be shared by multiple categories
⇒ Subcategory grammar ⇒ Providing a semantic basis for recognition.

3) Joint, completely unsupervised learning of the models and taxonomy of arbitrarily structured categories present in a given image set.

ACKNOWLEDGMENT: ONR under grant N00014-06-01-0101