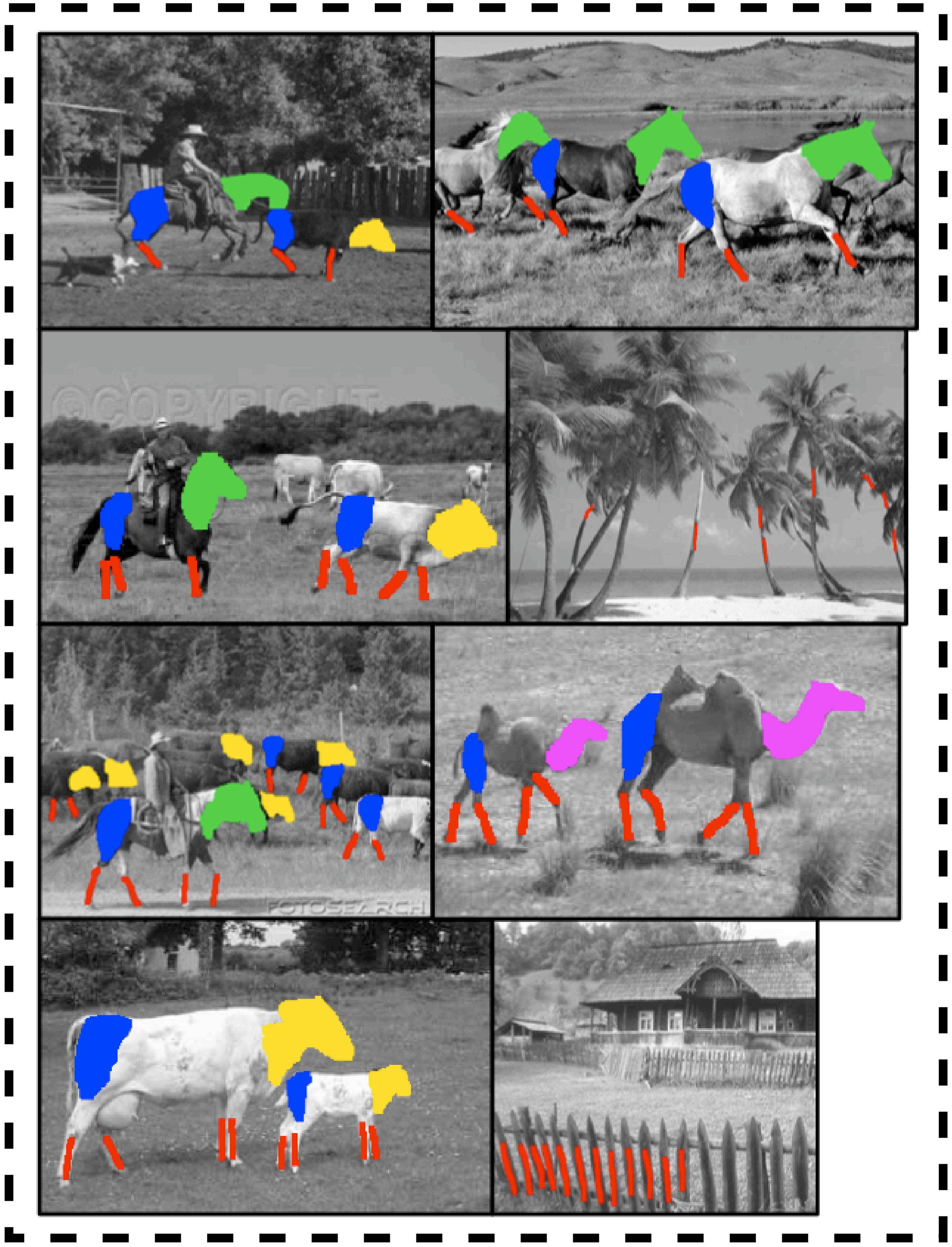
Unsupervised Learning of Categories Appearing in Images

Sinisa Todorovic
sintod@uiuc.edu

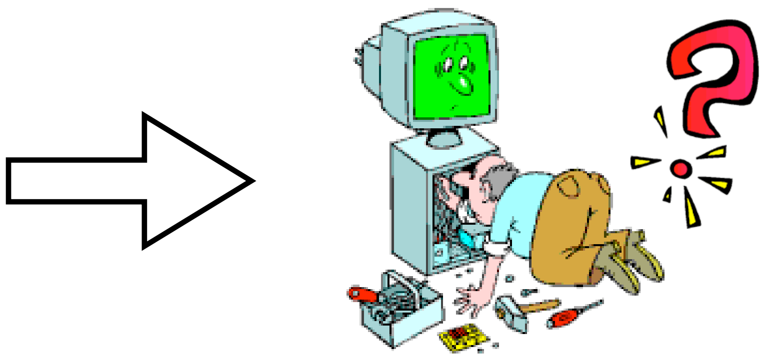
September 2007

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Any similar 2D objects?



arbitrary images



What is similar?

Where are they?

Any shared properties?

Any spatial relationships?

Category = Set of Similar 2D Objects

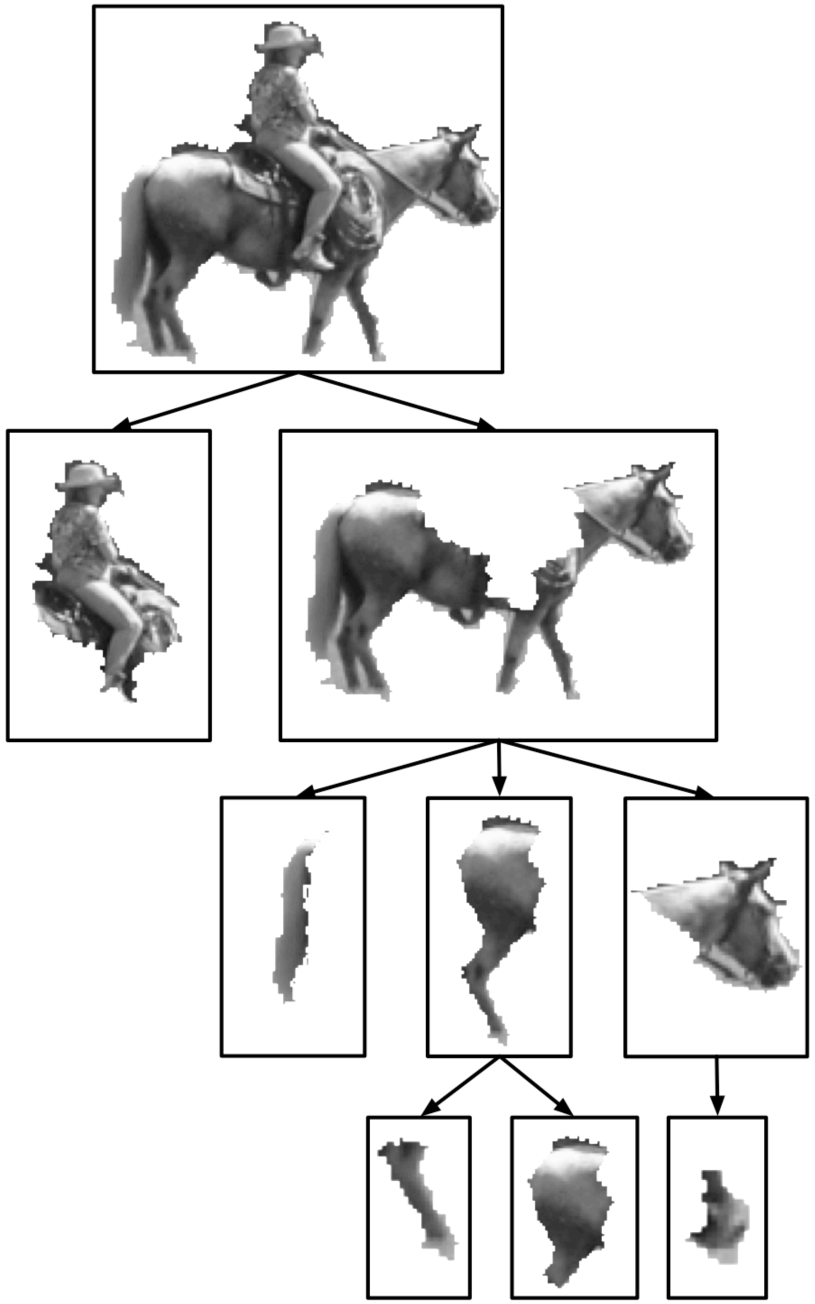
Definition of SIMILARITY
in terms of region properties:

(1) Photometric:
- **color**
- **texture**

(2) Geometric:
- **area**
- **boundary shape**

(3) Topological:
- **spatial layout of subcategories**
- **containment of subcategories**

finite recursive definition



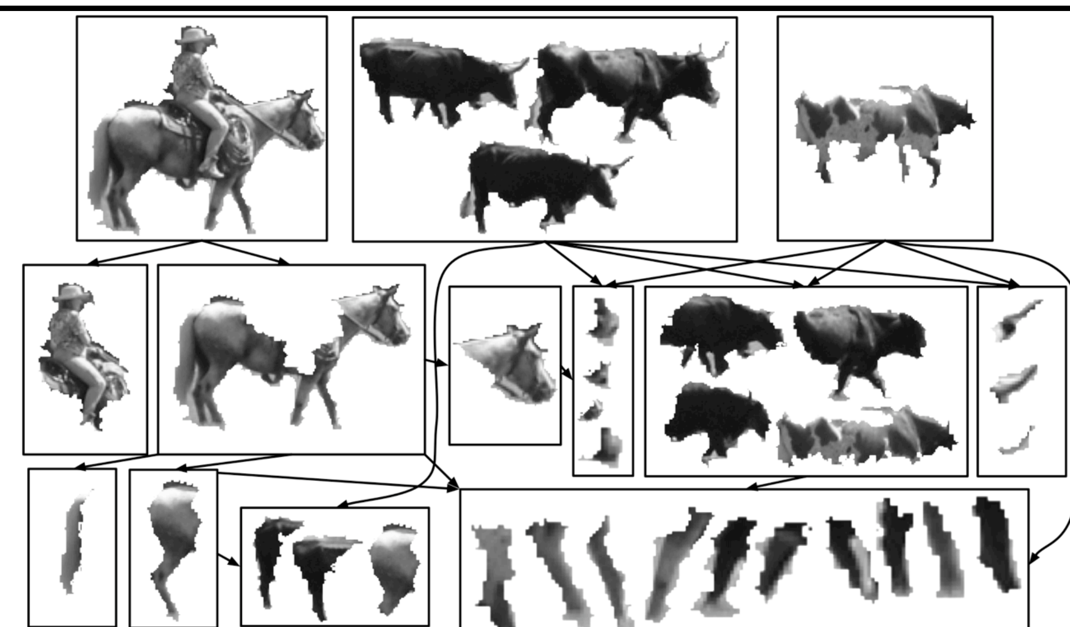
Objective: Given Arbitrary Images ...

DISCOVER all categories present

LEARN the models and spatial relationships of discovered categories

DETECT
RECOGNIZE
SEGMENT
all category instances

EXPLAIN
recognized categories
via identified
simpler subcategories



Object Recognition System

- **Amount of supervision in training**
- **Feature extraction**
- **Object representation and learning**
- **Evaluation**

Prior Work: Training

Caltech-101

Supervised:

- Categories defined by the user -- labeled images
- Manually segmented objects

Fischler&Elschlager 73, Winston 75, Leibe et al. 04, Winn et al. 05, Opelt et al. 06

Weakly supervised:

- Images must contain a pre-selected category
- ``Background” category
- Large inter-category differences
- Some require many training images

Weber et al. 00, Fergus et al. 03, Fei-Fei et al. 04, Forsyth et al. 02, Sivic et al. 05, Lazebnik et al. 06

Our Approach: Unsupervised Training

Categories not defined by the user -- unlabeled images

**many riders,
horses, cows**

**scale
viewpoint
illumination**

**articulation,
(self-)occlusion,
clutter**

no cows

- Small inter-category differences

- No ``background” images

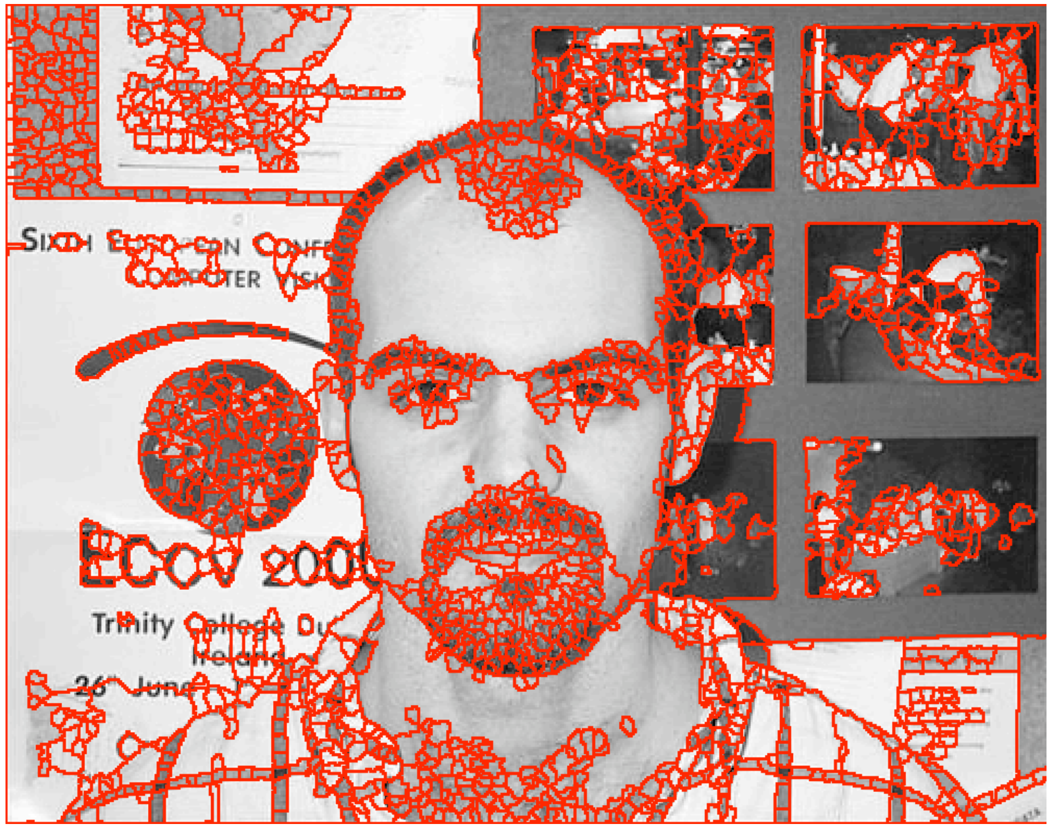
- Small training sets

Sivic et al. 05; Russell et al. 06; Todorovic&Ahuja 06,07

Prior Work: Feature Extraction

- **Key-points (e.g., Harris-Laplacian, Kadir-Brady):**
  - **Fergus et al. 03, Lowe 04, Fei-Fei et al. 04, Torralba et al. 04, Grauman&Darell 05, Mokolajczyk&Schmid 05, Sivic et al. 05, Sudderth et al. 05, Lazebnik et al. 06**
- **Edges (e.g., Canny):**
  - **Rosenfeld 72, Shotton et al. 05, Fergus et al. 05, Ren et al. 05, Opelt et al. 06, Leordeanu et al. 07**
- **Regions (e.g., Mean-shift, N-cuts, Scale-space):**
  - **Hanson&Riseman 78, Nevatia 89, Basri&Jacobs 97, Keselman&Dickinson 05, Weiss&Ray 05, Shokoufandeh et al. 06, Russell et al. 06, Pantofaru&Herbert 07**

Our Approach: Features = Regions



Advantages:

- **Higher-dimensional ⇒ richer descriptors, more discriminative**
- **Coincide with object(-part) boundaries**
- **Facilitate modeling of: spatial cohesiveness, smoothness, containment, contiguity, adjacency, etc.**

Our Approach: Feature Extraction

N. Ahuja 96, Tobb & Ahuja 97, Arora&Ahuja 06

input image

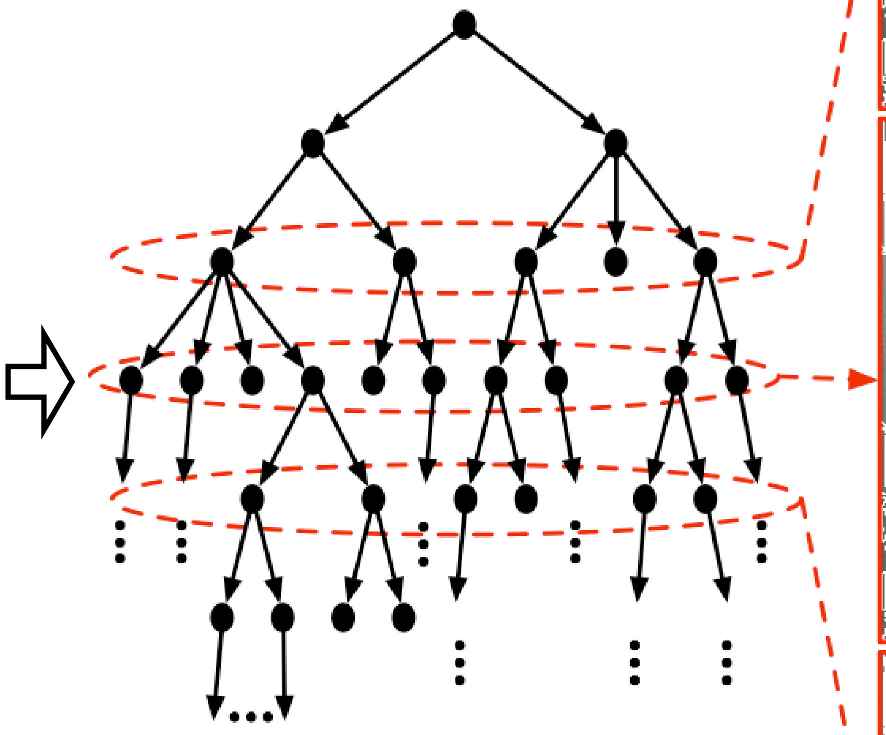
multiscale segmentation

Homogeneous regions at ALL contrasts regardless of shape, size, and context.

Our Approach: Image = Segmentation Tree

Example segmentations

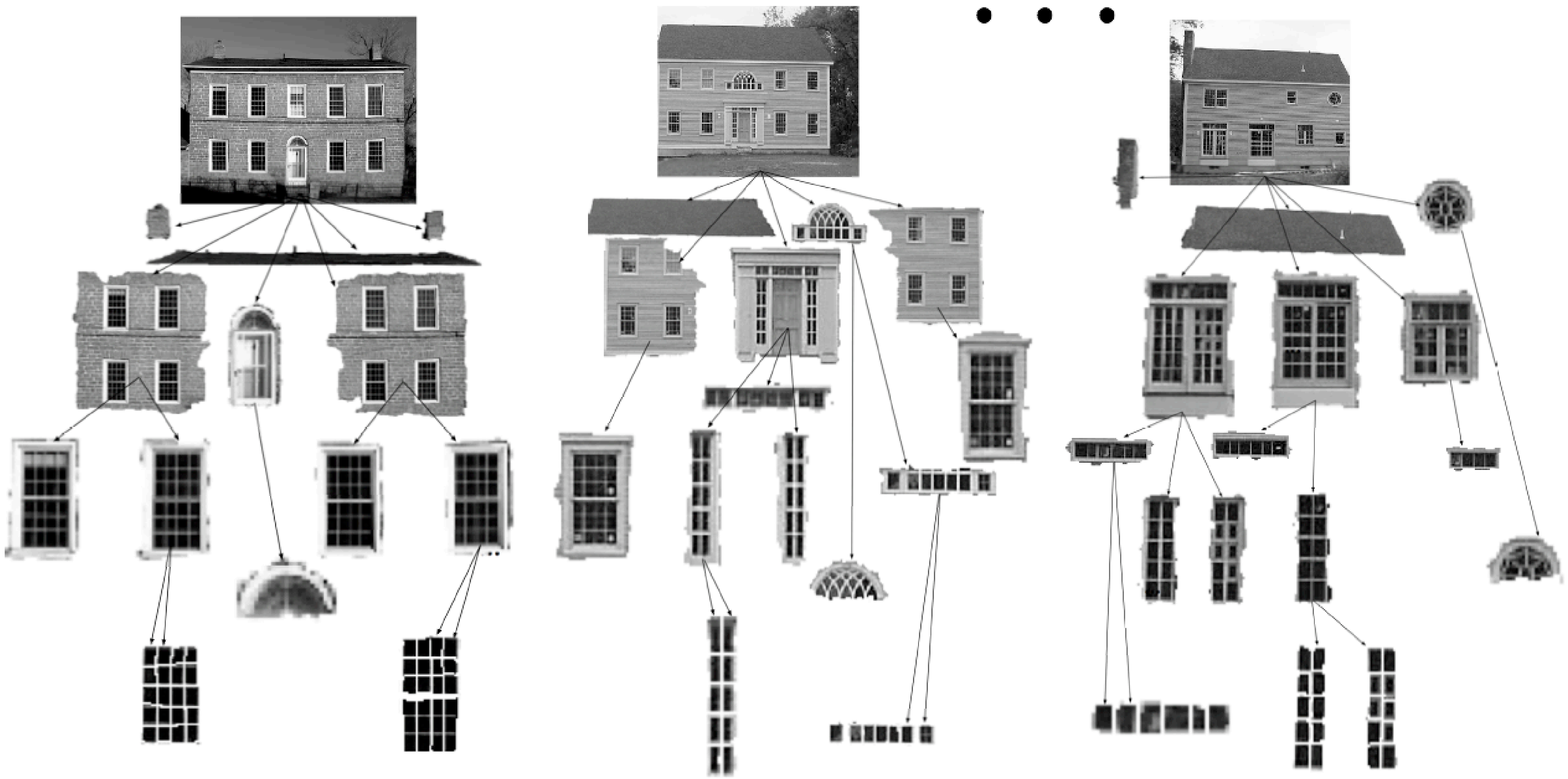
Segmentation tree



Cutsets

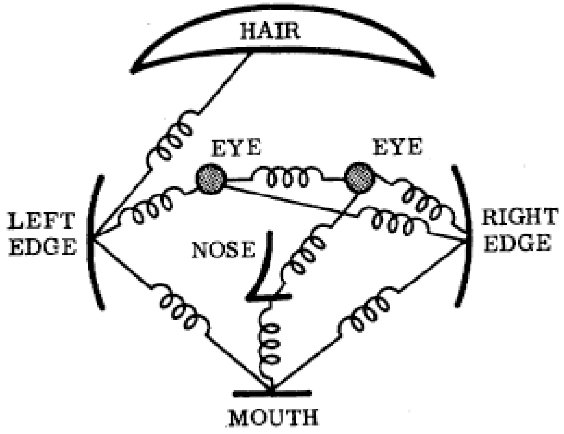
Number of nodes (~100)
Hierarchy depth (~10)
Branching factor (0-10)

Prior Work: Efficient Object Representation



- **Compositionality:**
  - **Objects = Configuration of parts**
  - **Efficient: Parts have smaller variations, and occur more frequently**
- **Sharing:**
  - **Parts occur in the definitions of multiple objects**
  - **Efficient: Sub-linear complexity in the number of objects**

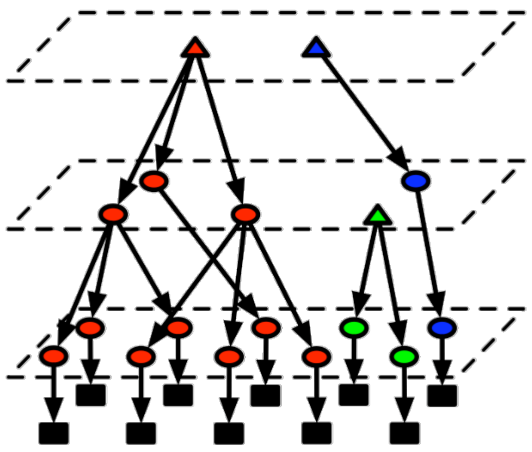
Prior Work: Object Representation - Compositionality



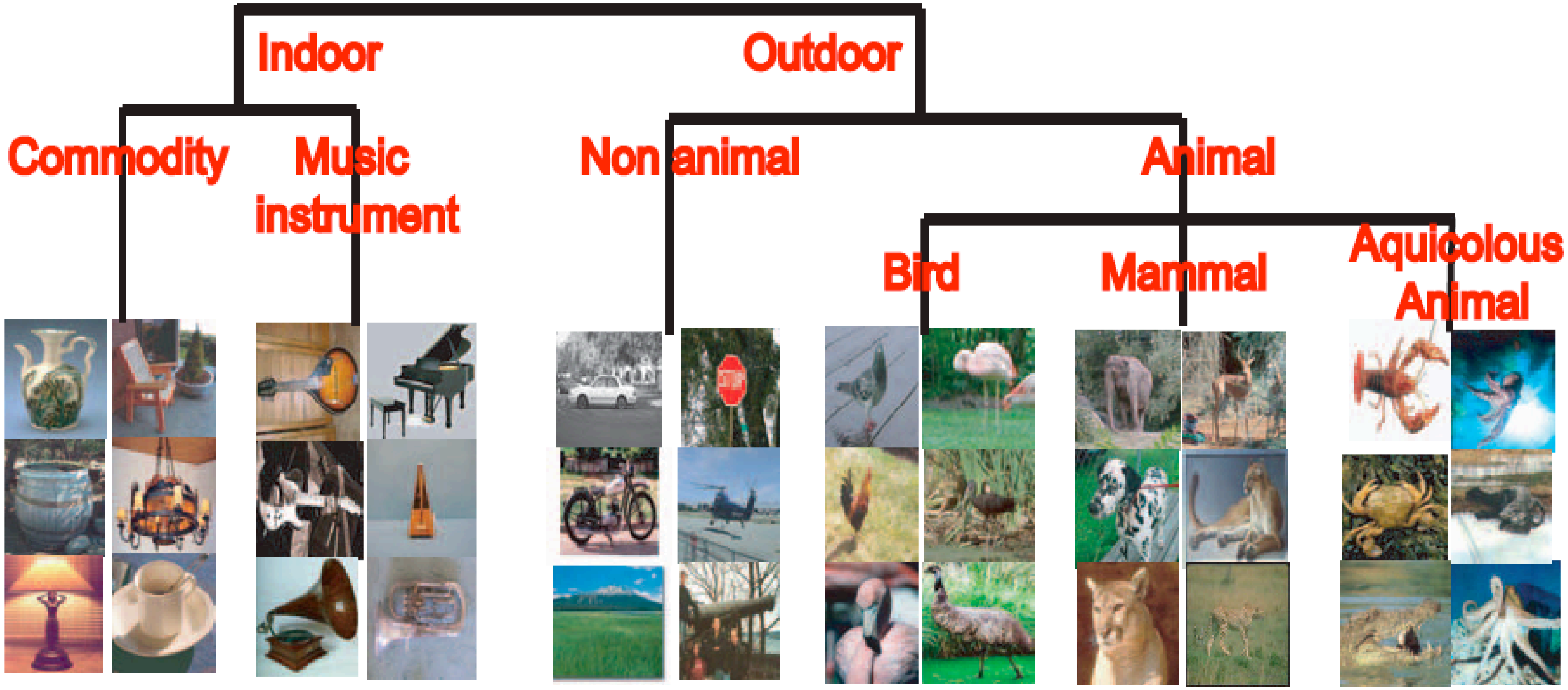
- **Planar graph representations:**
  - **Pictorial structures --** Fischler&Elschlager 73, Felzenszwalb&Huttenlocher 05
  - **Constellation model --** Fergus et al. 03

- **Hierarchical graph representations:**
  - Crowely&Sanderson 87, Ettinger 88, Utans 92, Nishida&Mori 93, Bouman&Shapiro 94, Perrin&Ahuja 98, Bretzner&Lindenberg 99, Shokoufandeh et al. 99, Storkey&Williams 03, Geman-Leonardis-Buhmann 00-07, Todorovic&Nechyba 05, Todorovic&Nechyba 07

- **Computationally infeasible -- approximate inference**
  - The user specifies:
    - Number of parts
    - Model structure
    - Hierarchy depth
    - Branching factor



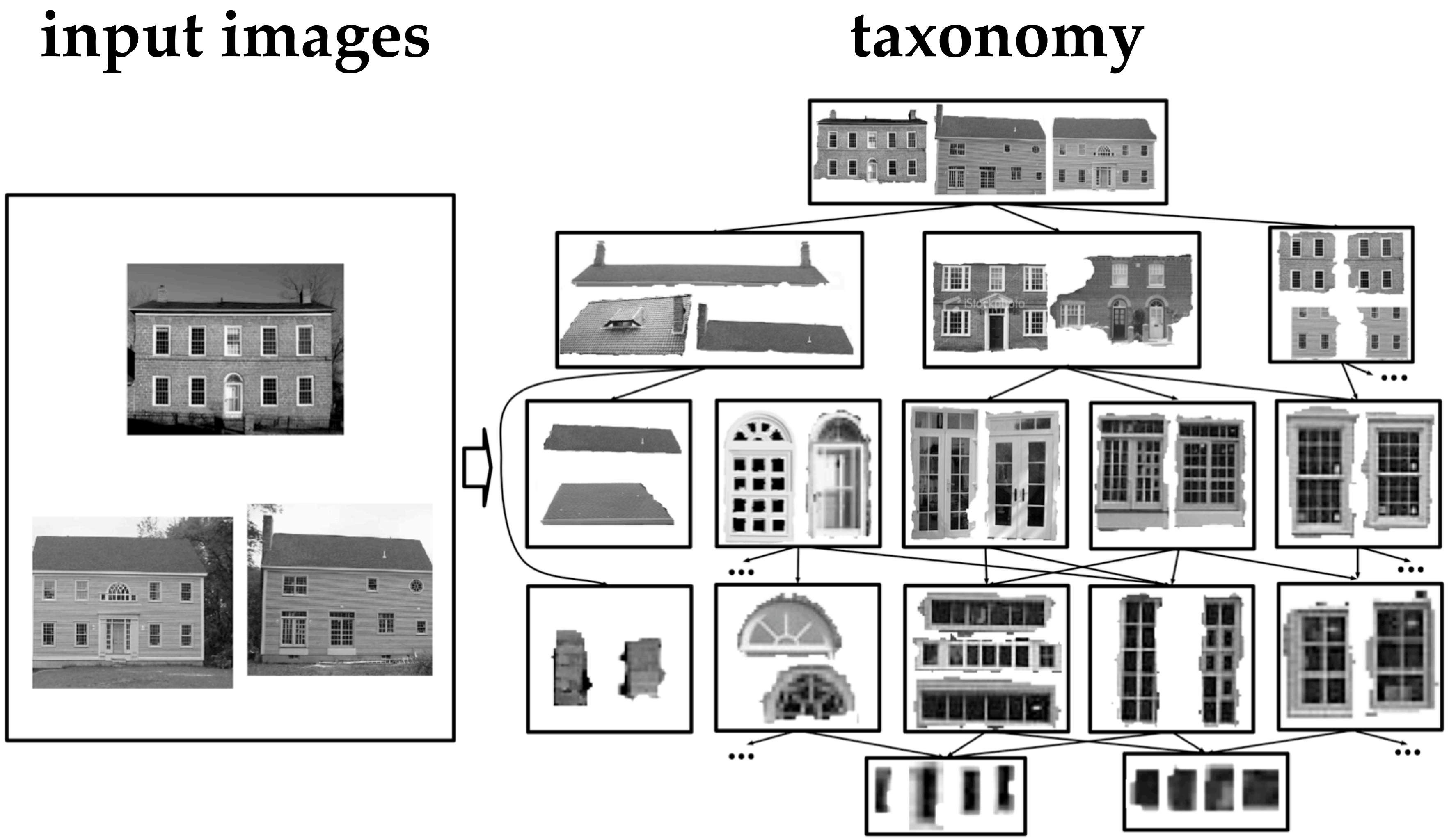
Prior Work: Relationships Among Categories



- Learn only sharing of features, not entire categories
- Similarity = Number of shared features
- Dendogram taxonomy vs. Spatial taxonomy of categories

Torralba et al. 04; Opelt et al. 06; Fei-Fei et al. 05, 06, 07

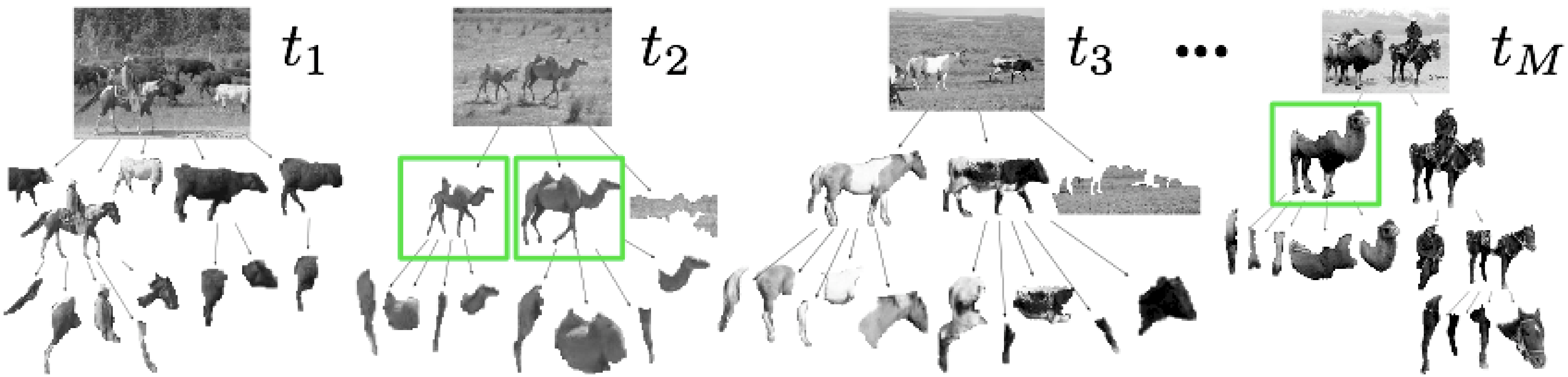
Our Approach: Representation = Taxonomy



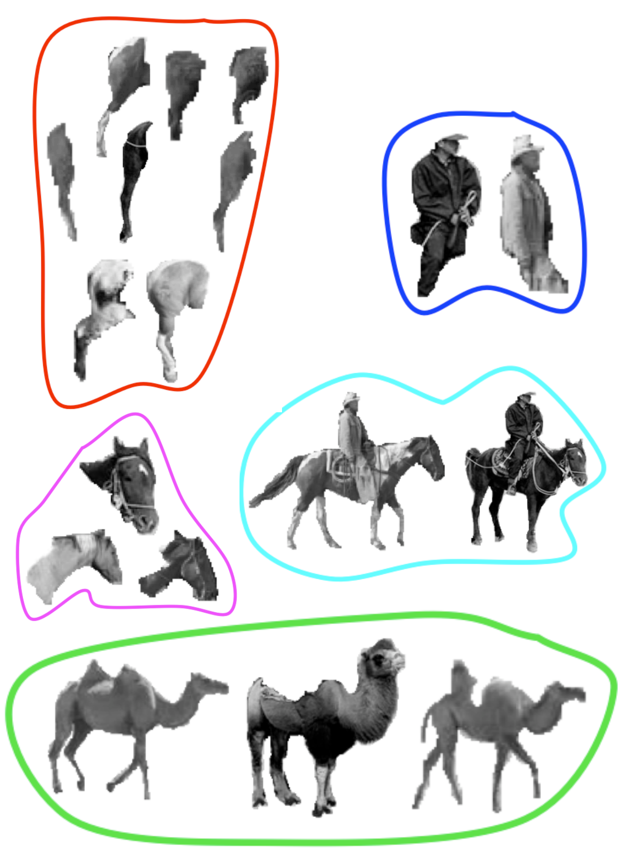
- **Spatial taxonomy:**
  - Complex categories = Configurations of subcategories
  - Co-occurrence category
- **Modeling arbitrarily structured categories**
  - No fixed number of nodes, hierarchy depth, branching factor
- **Exact learning -- no need for approximate inference**

Overview of Our Approach

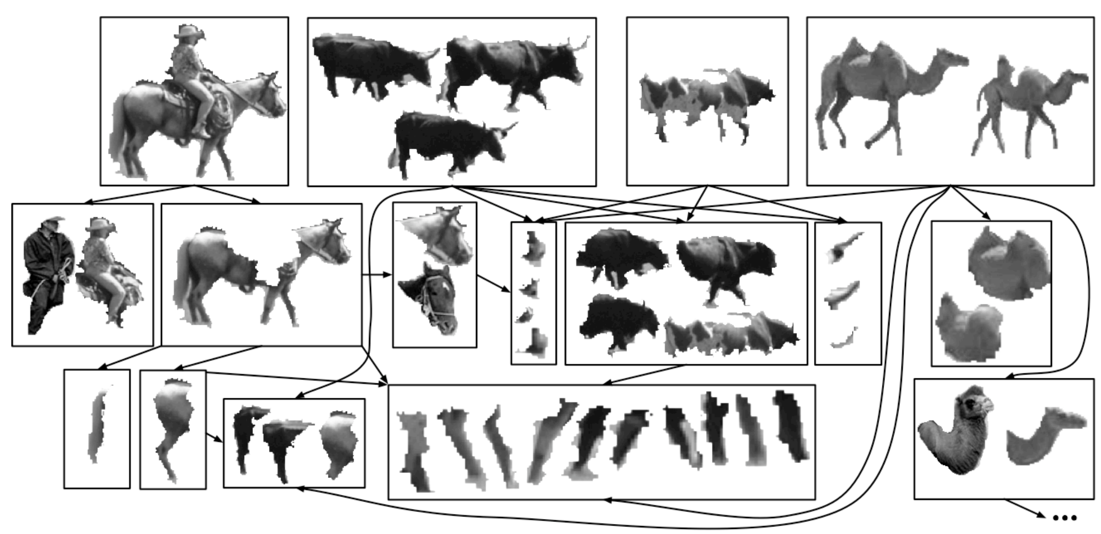
1. TREE MATCHING



2. CLUSTERING



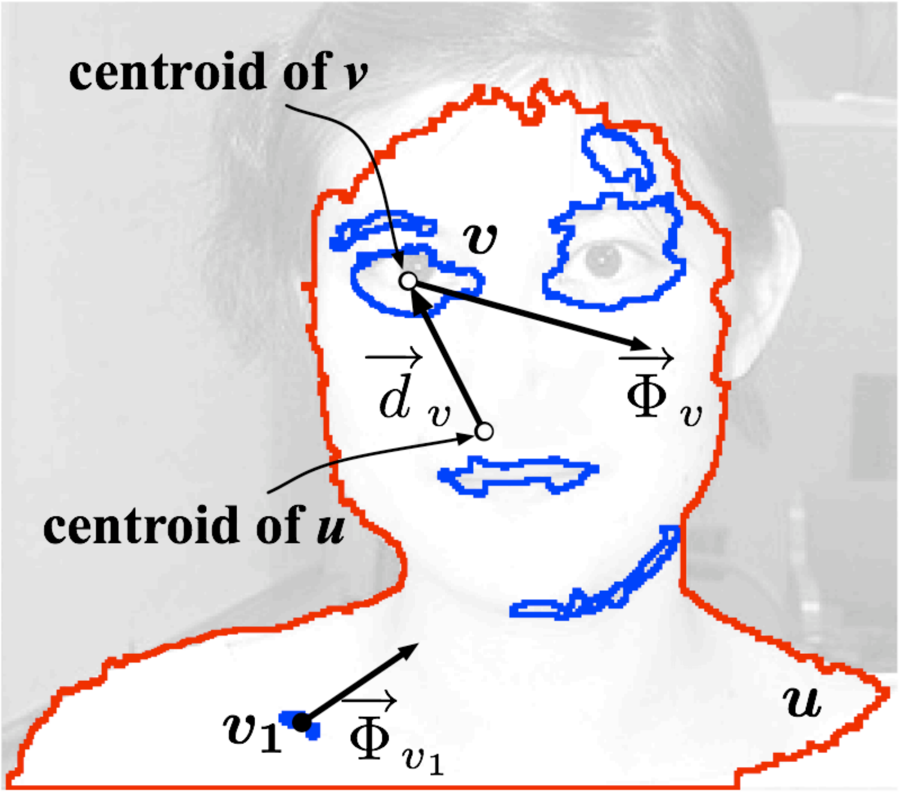
3. CONSTRUCTING TAXONOMY



4. RECOGNIZE SEGMENT EXPLAIN

Tree Matching: Region Properties and Saliency

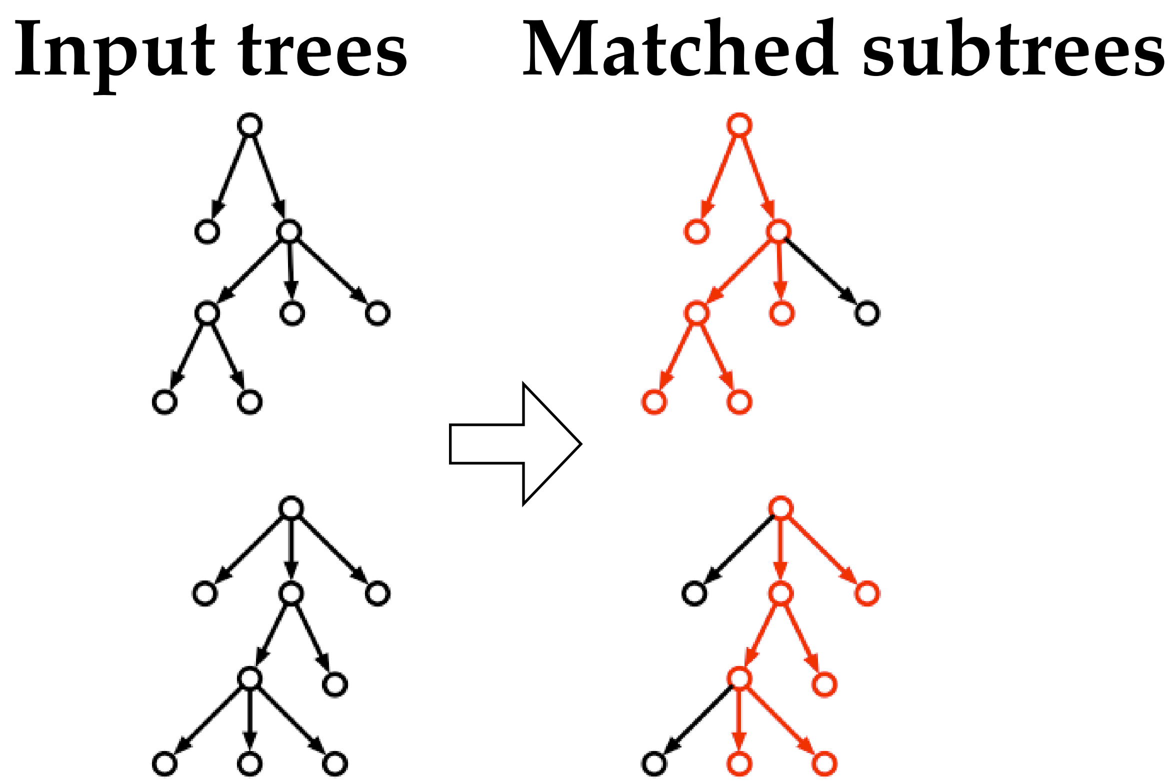
- **Relative to the parent ⇒ Rotation and scale invariance**



- **Normalized properties: ψ_v**
  - Intensity
  - Area
  - Central moments
  - Perimeter over area
  - Displacement of centroids
  - Spatial distribution of siblings

$$\underbrace{r_v}_{\text{saliency}} = \underbrace{\boldsymbol{\xi}}_{\text{relative significance to recognition}}{}^{\mathrm{T}} \underbrace{\boldsymbol{\psi}_v}_{\text{region properties}}, \qquad \boldsymbol{\xi} \geq 0, \quad \boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{\xi} = 1$$

Tree Matching: Consistent Subtree Isomorphism



- **Match regions whose appearance and topology are similar, and the same holds for their subregions**
- **Preserve original topology**
- **Approaches:**
  - **Spectral:** Siddiqi et al. 99, Shokoufandeh et al. 05
  - **Edit-distance:** Eshera&Fu 86, Bunke&Allermann 83, Sebastian&Kimia 05
  - **Max-clique:** Pelillo et al. 99, Torsello&Hancock 03, Todorovic&Ahuja 07

Tree Matching: Formulation

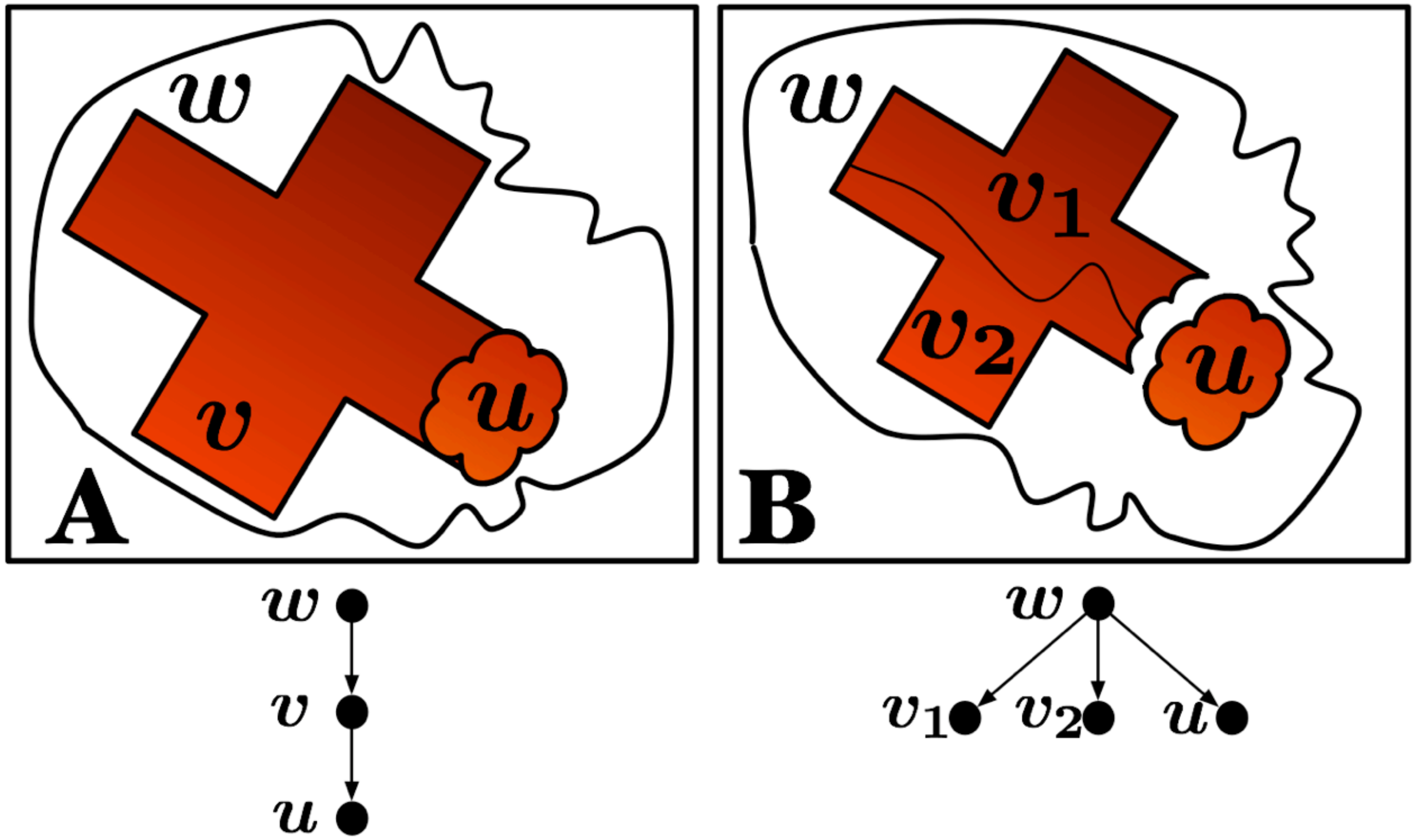
GIVEN two trees: $t, \ t'$

FIND legal bijection $f : (v, v'), \ v \in t, \ v' \in t'$

which MAXIMIZES the similarity measure:

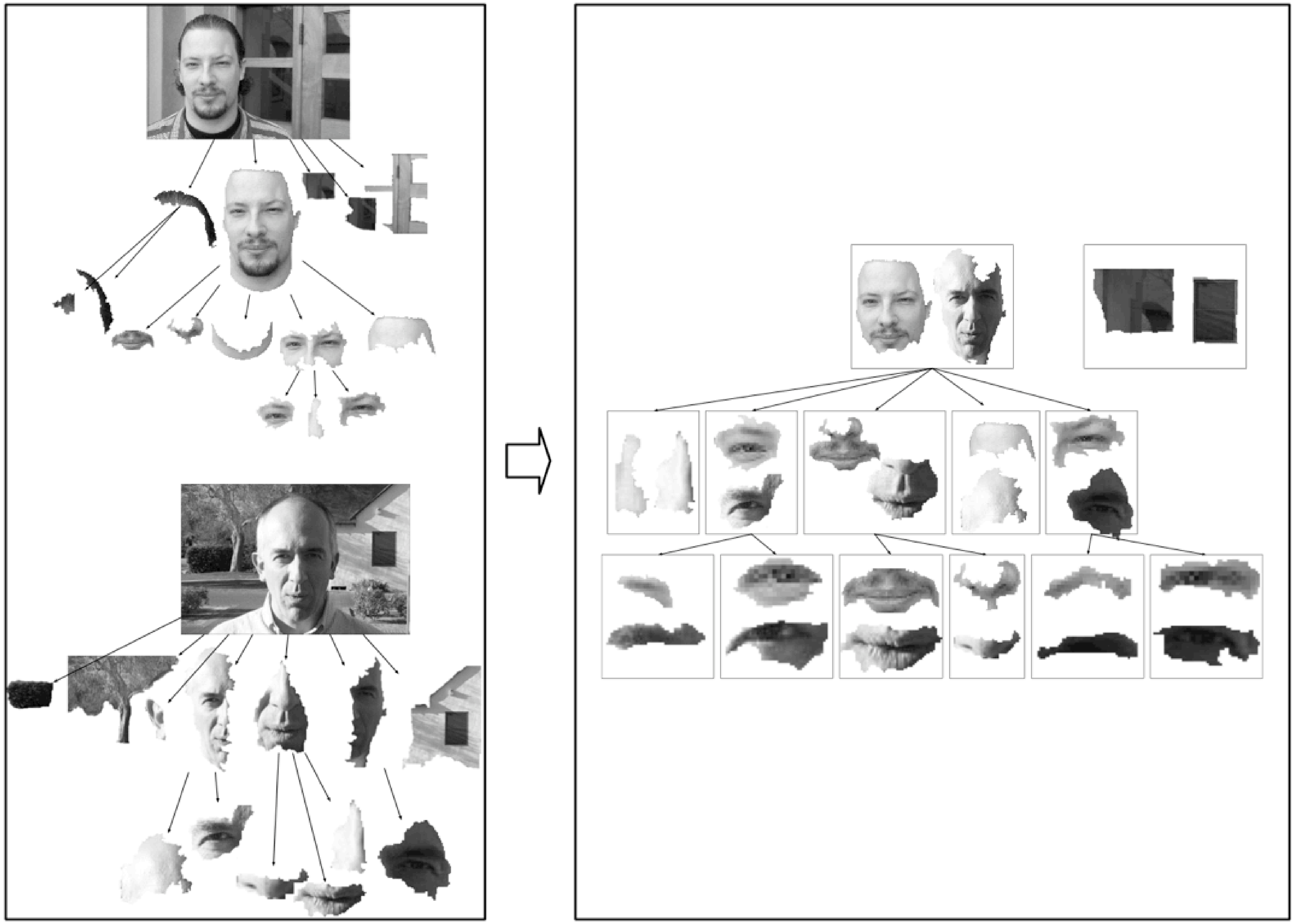
$$\mathcal{S}_{tt'} = \max_{f} \sum_{(v,v') \in f} \Big[\underbrace{\min(r_v, r_{v'})}_{\textbf{saliency}} - \underbrace{[\max(r_v, r_{v'}) - \min(r_v, r_{v'})]}_{\textbf{cost of region matching}} \Big]$$

Tree Matching: Instability of Segments



- Low-contrast regions may split or merge
- Similar regions may appear at different hierarchy depths

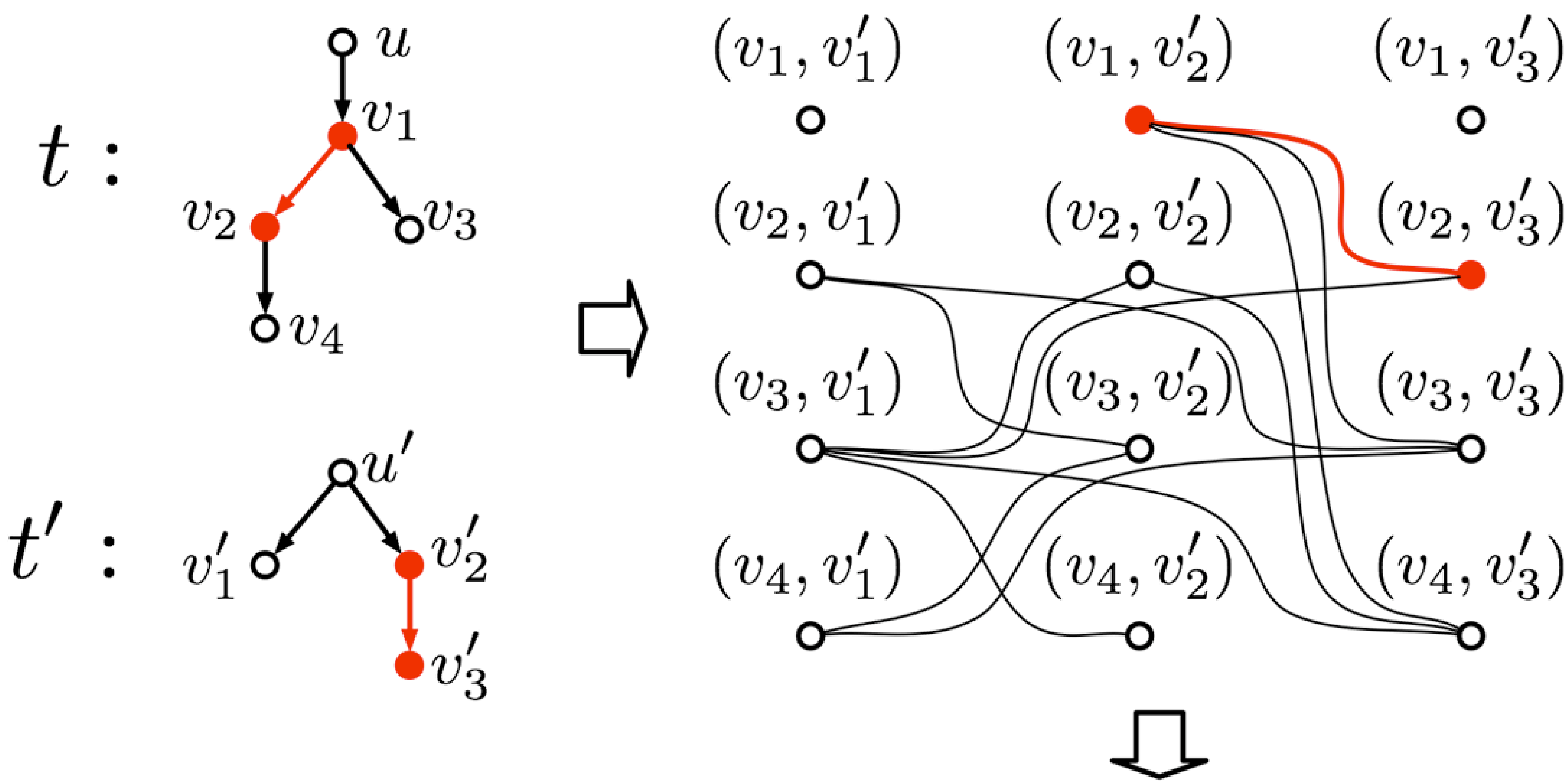
Tree Matching: Achieving Robustness



- **Consider:**
  - **Many-to-many, many-to-one, one-to-one correspondences**
  - **Matching of all descendants under a visited node**

Tree Matching: Solution = Relaxation

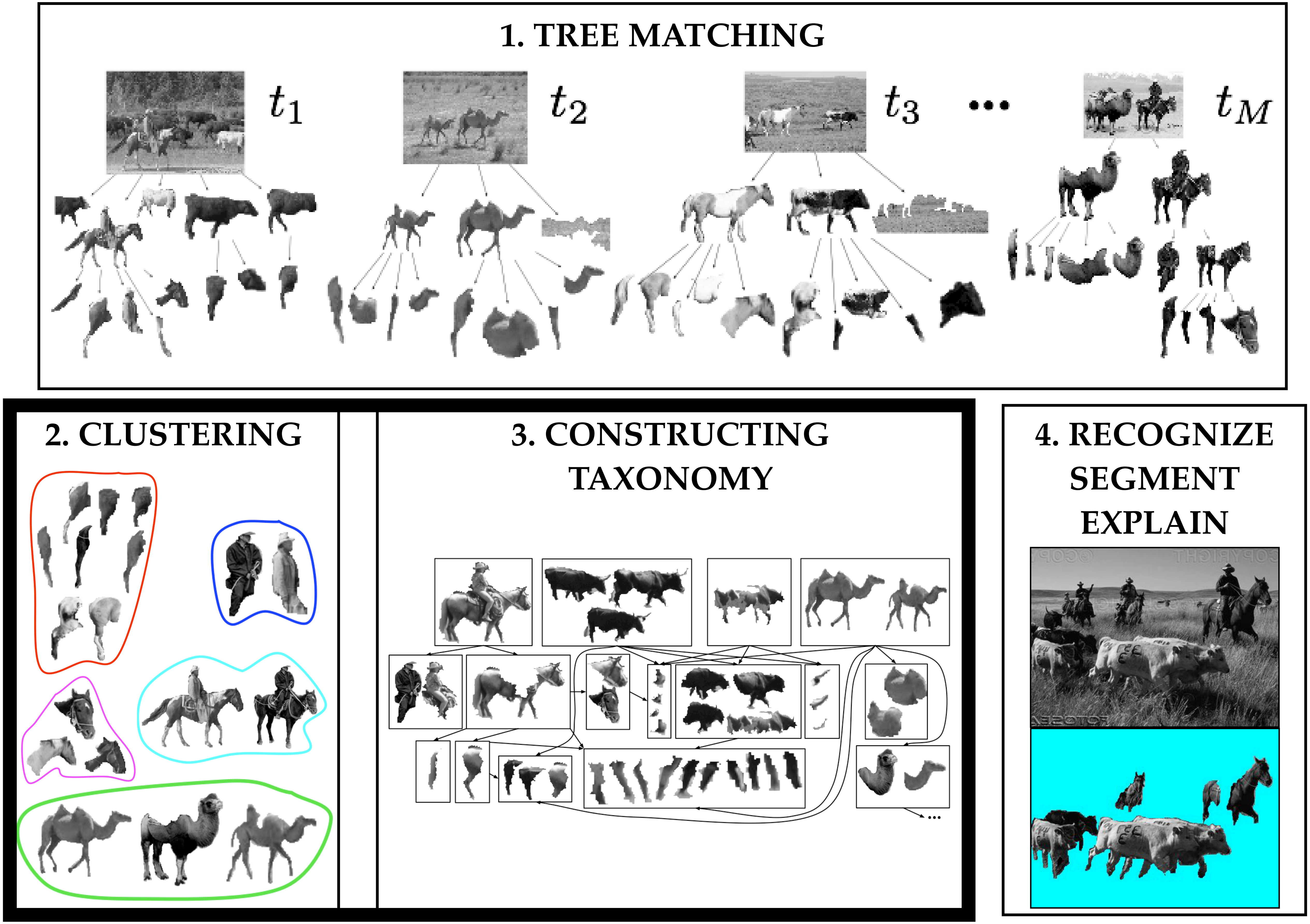
$$\mathscr{A}_{uu'} = \{(v, v') : v \in t, v' \in t'\}$$



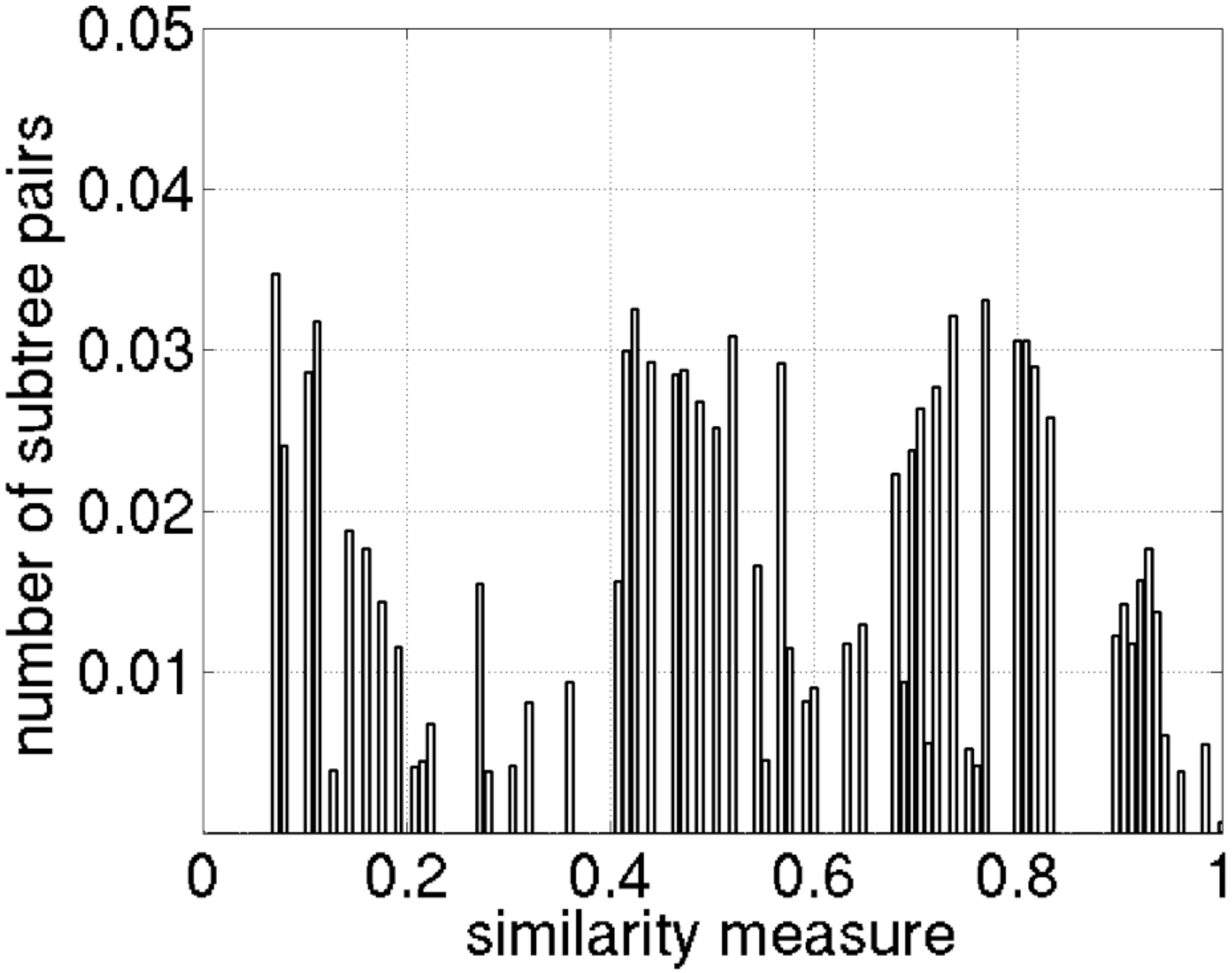
subtree isomorphism $f = \mathcal{C}_{uu'} = \{(v_1, v_2'), (v_2, v_3')\}$

- **Theorem:** [Todorovic&Ahuja 07]
  Consistent subgraph isomorphism = Maximum weight clique
- **Complexity:** $O(N^2)$, **N - number of nodes**

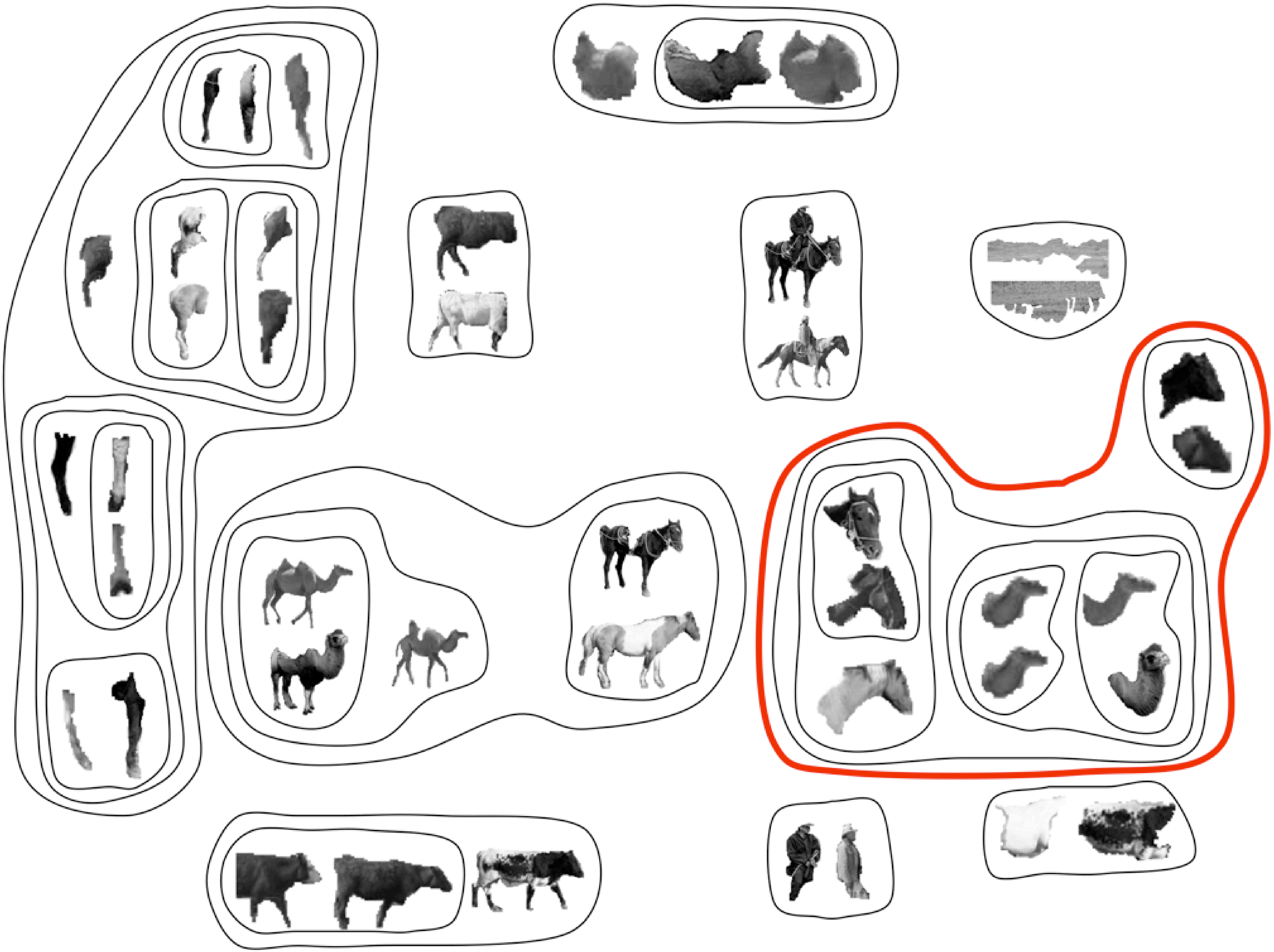
Overview of Our Approach



Similarity Measure of Regions

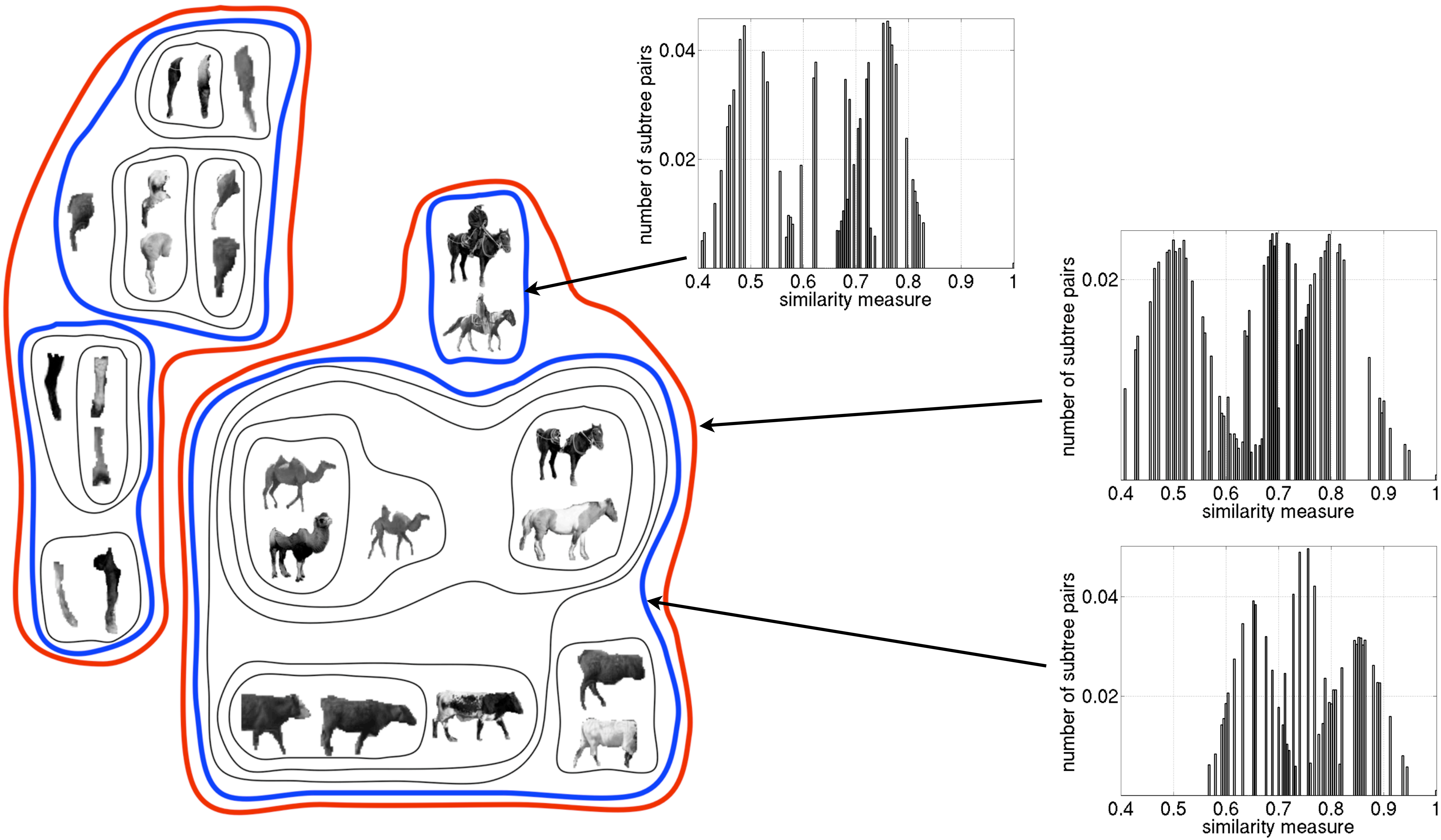


Agglomerative Clustering

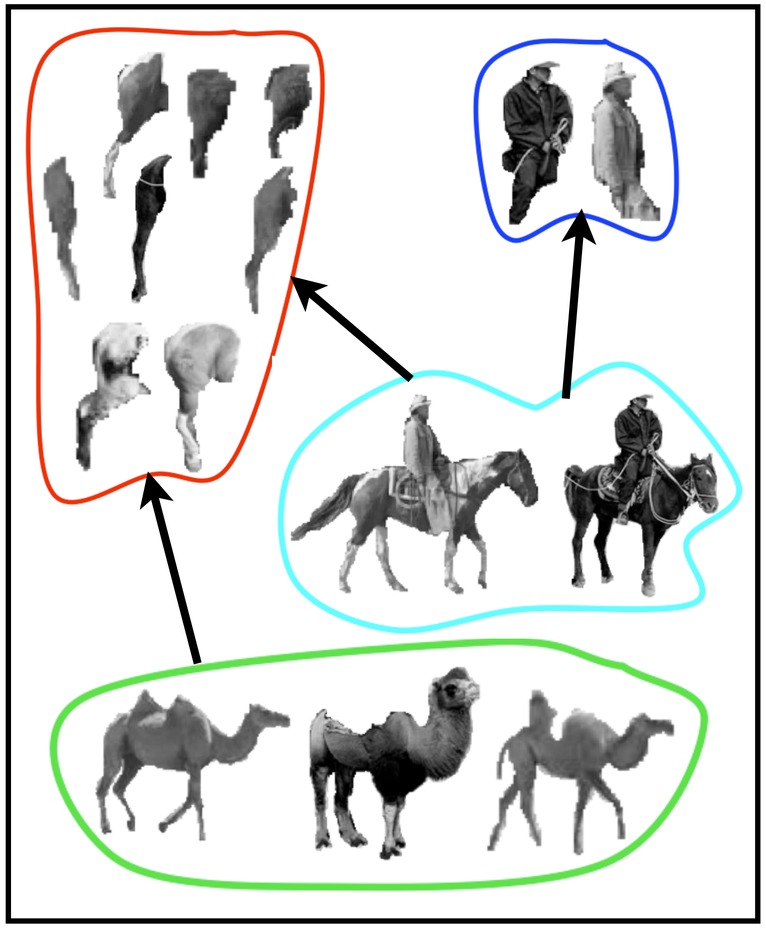


From Clusters to a Particular Categorization: KS-Test

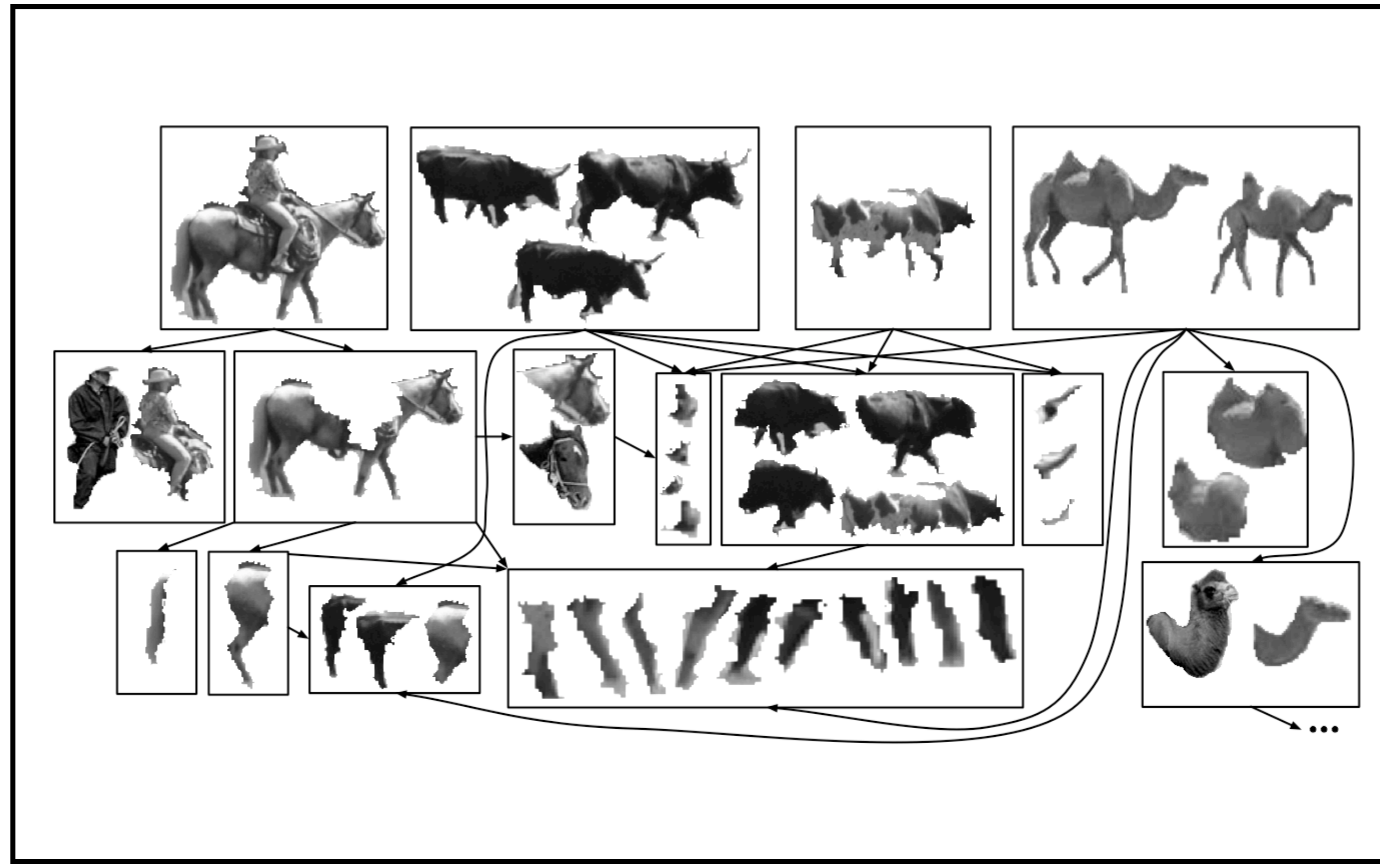
Cluster that passed the KS-test with $\alpha = 5\% \Leftrightarrow$ Discovered Category



Constructing the Taxonomy



clusters of subtrees

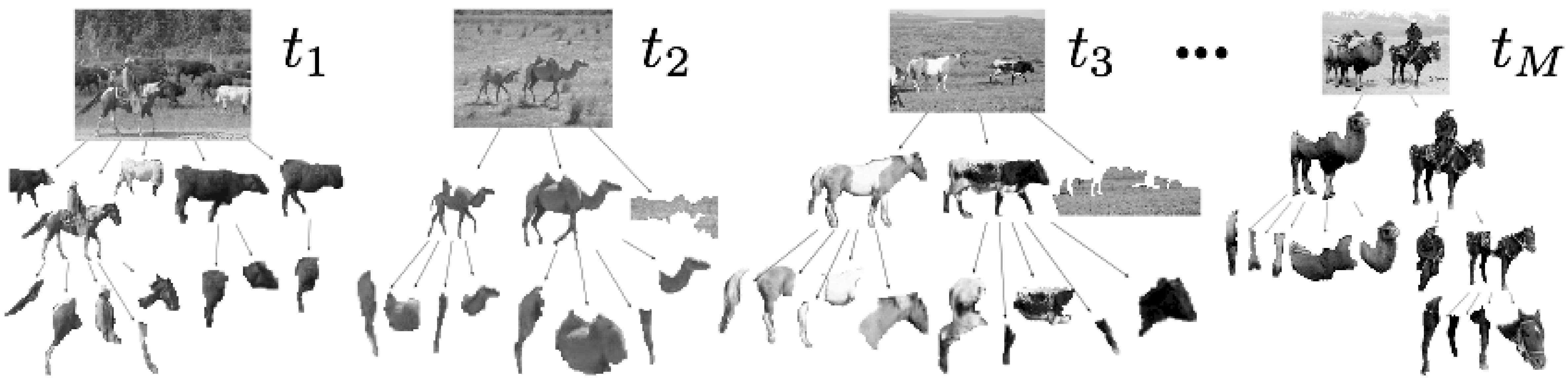


taxonomy of discovered categories

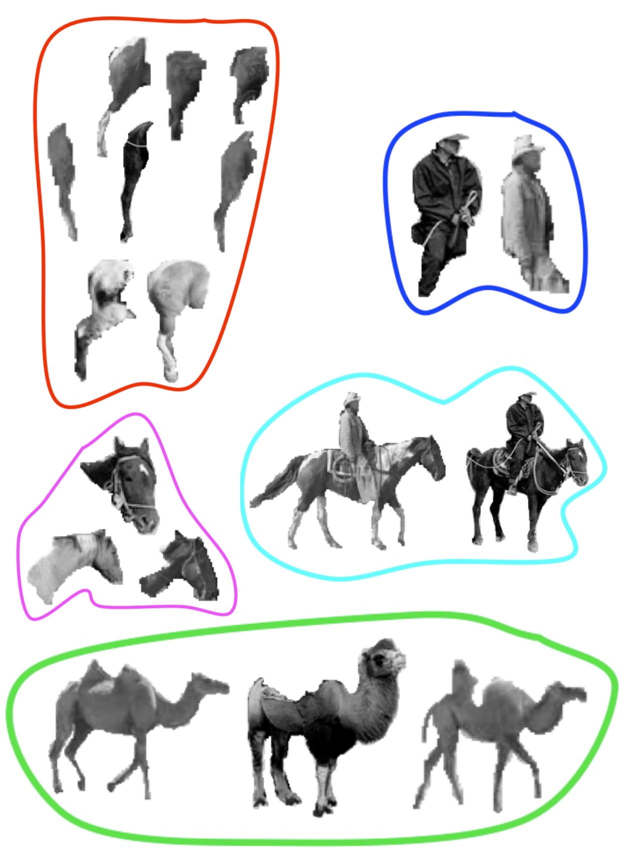
- Spatial relationships of subtrees are extended to the clusters
- Co-occurrence category = Forest of disjoint subtrees
- Typically, 20 nodes per category

Overview of Our Approach

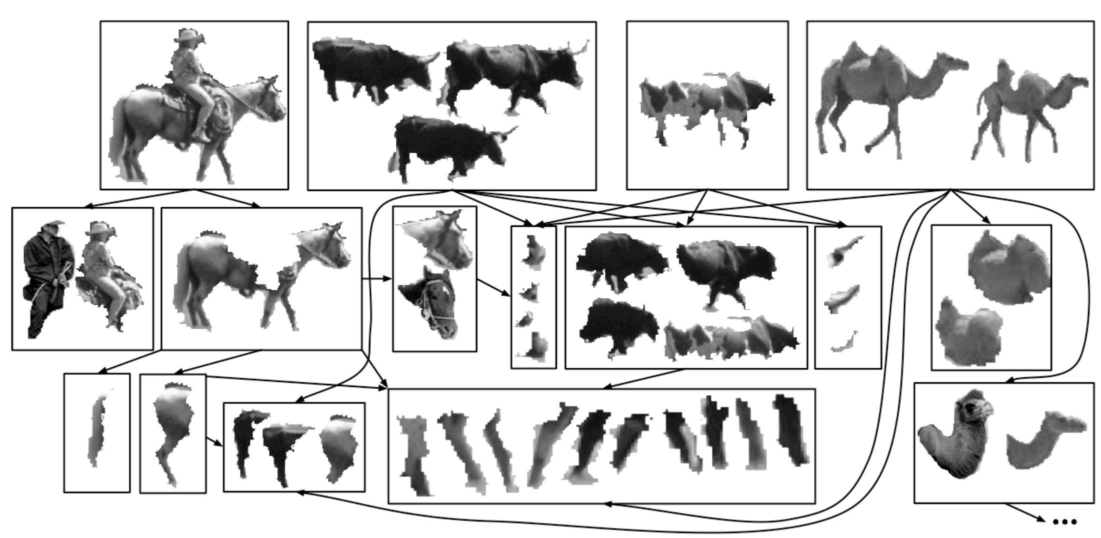
1. TREE MATCHING



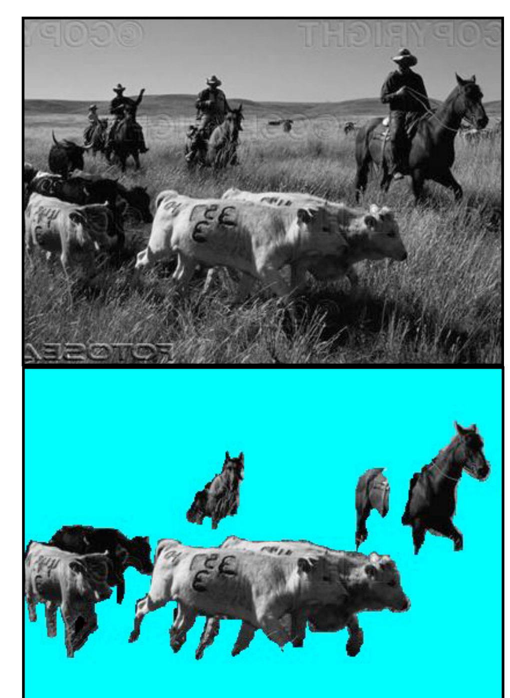
2. CLUSTERING



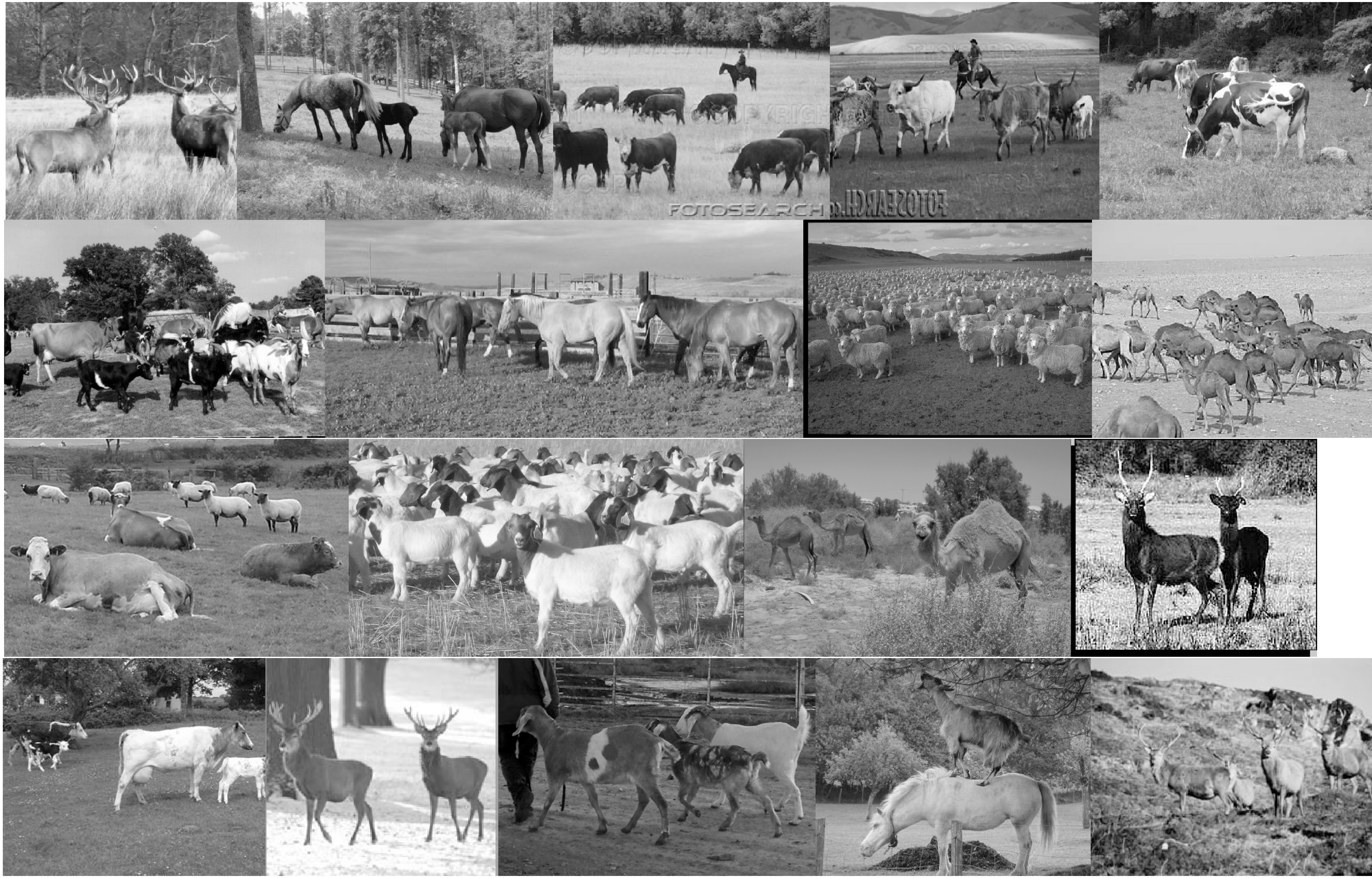
3. CONSTRUCTING TAXONOMY



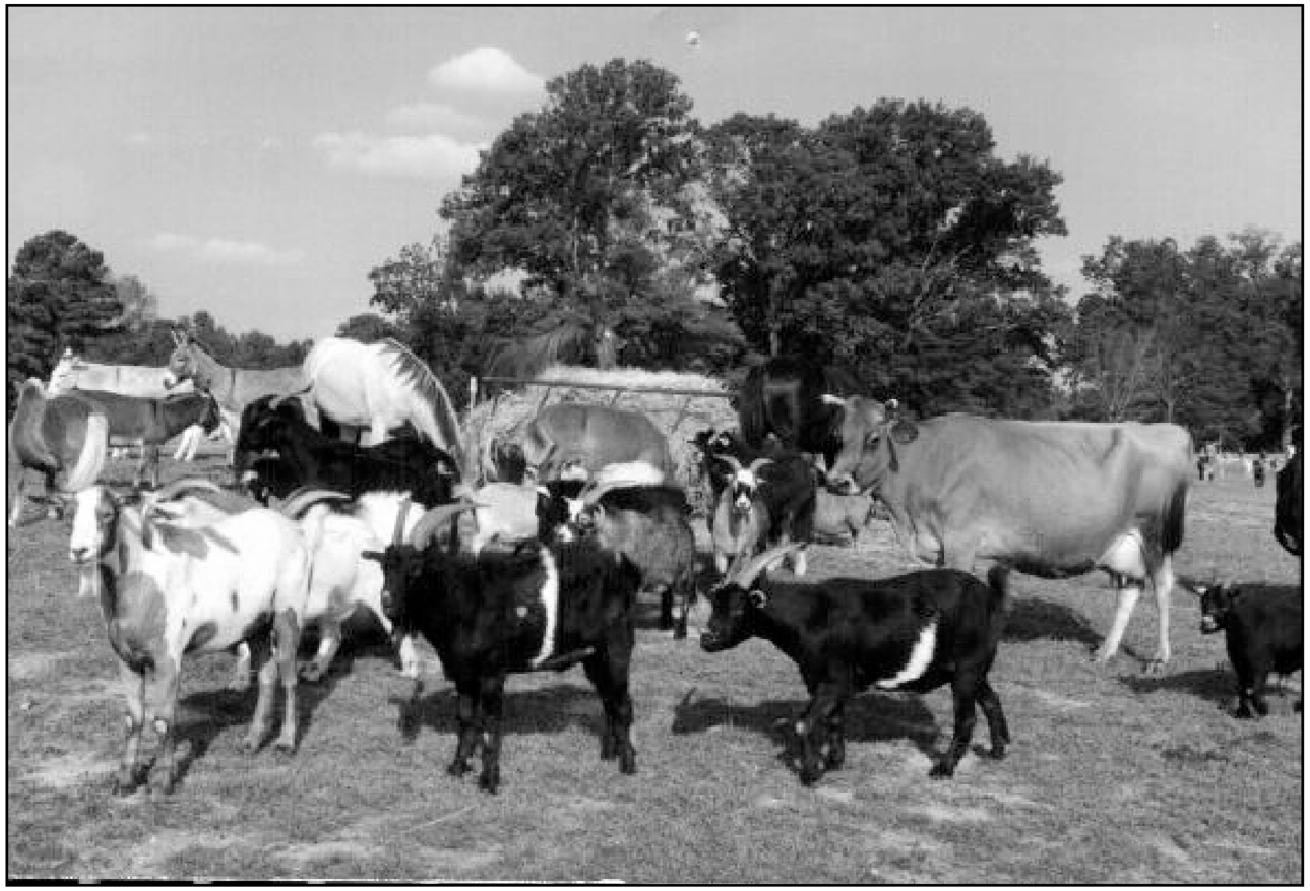
4. RECOGNIZE SEGMENT EXPLAIN

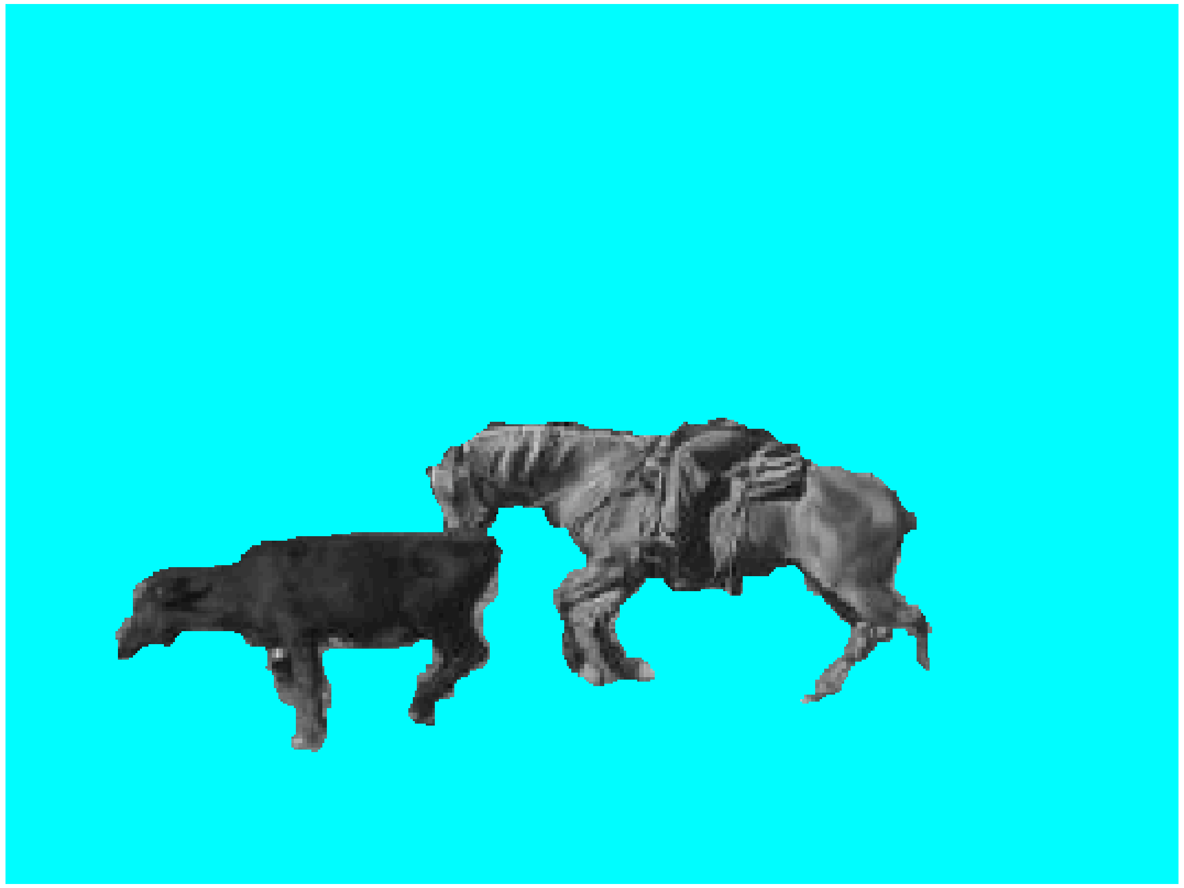


Results: New Dataset – Hoofed Animals

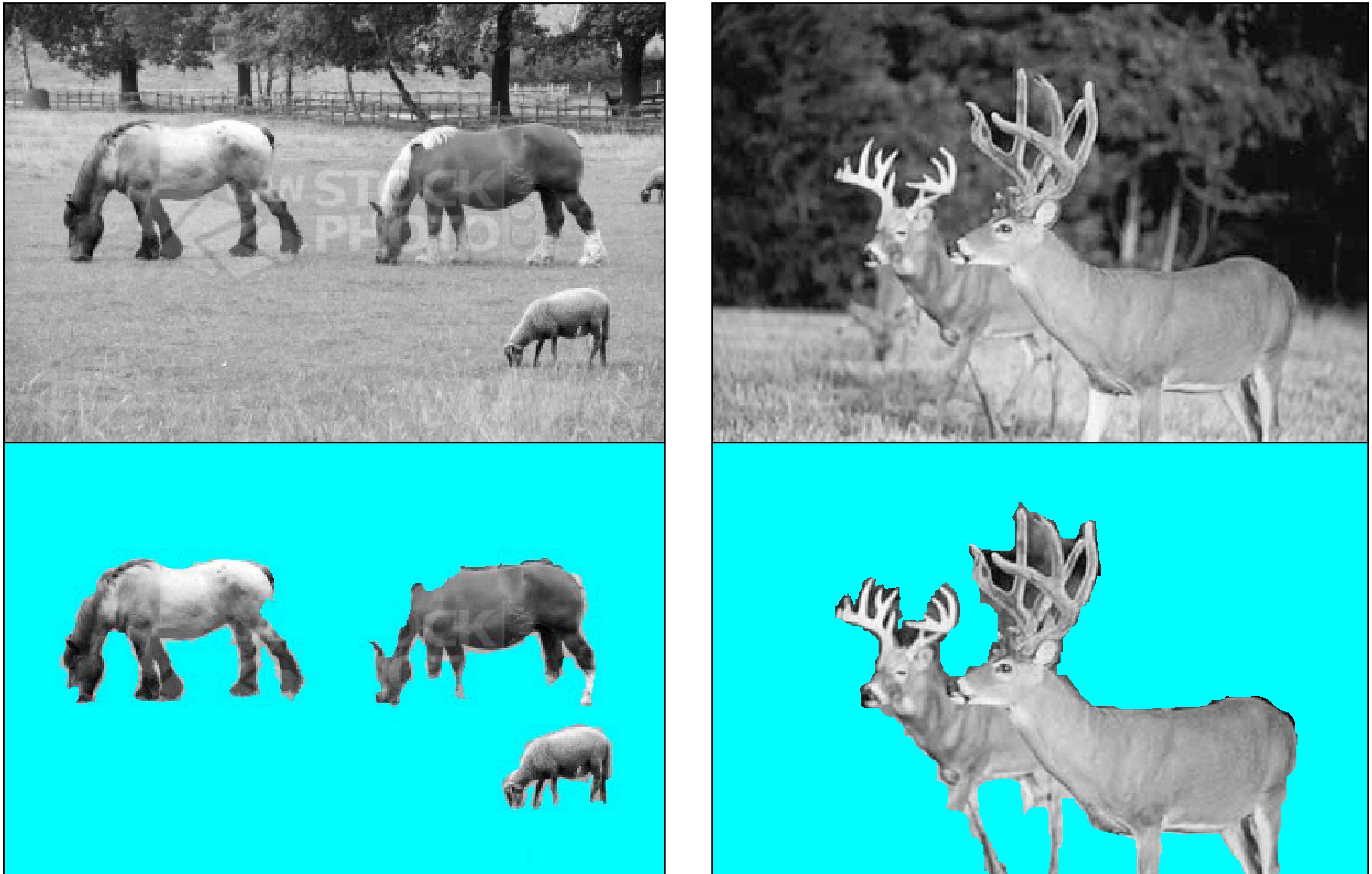


Simultaneous Detection, Recognition, Segmentation





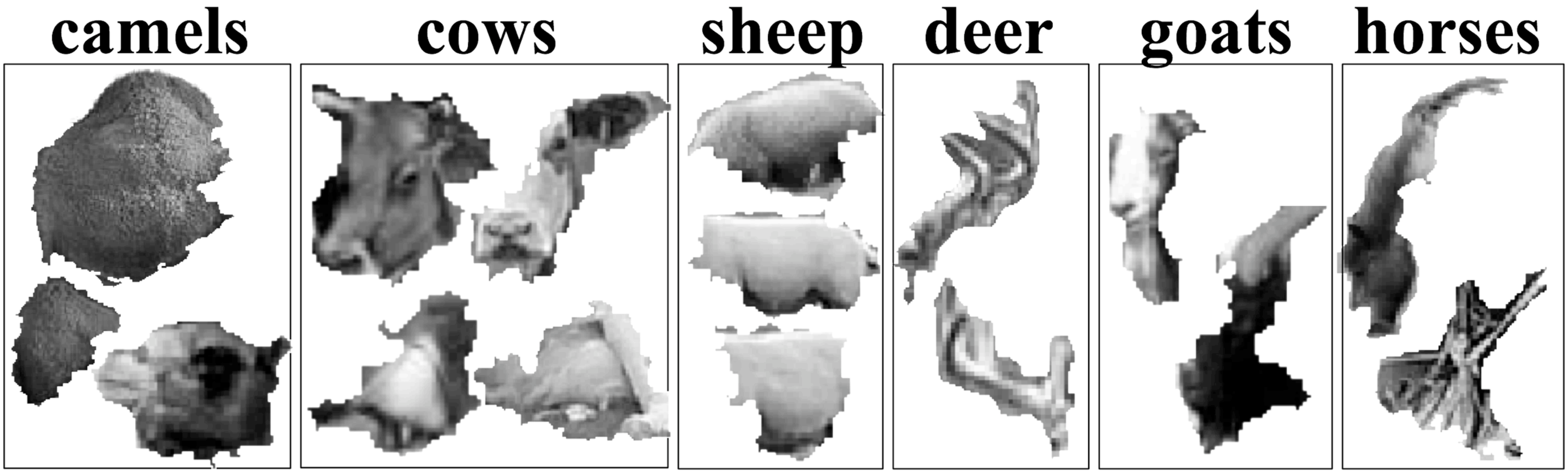
Simultaneous Detection, Recognition, Segmentation



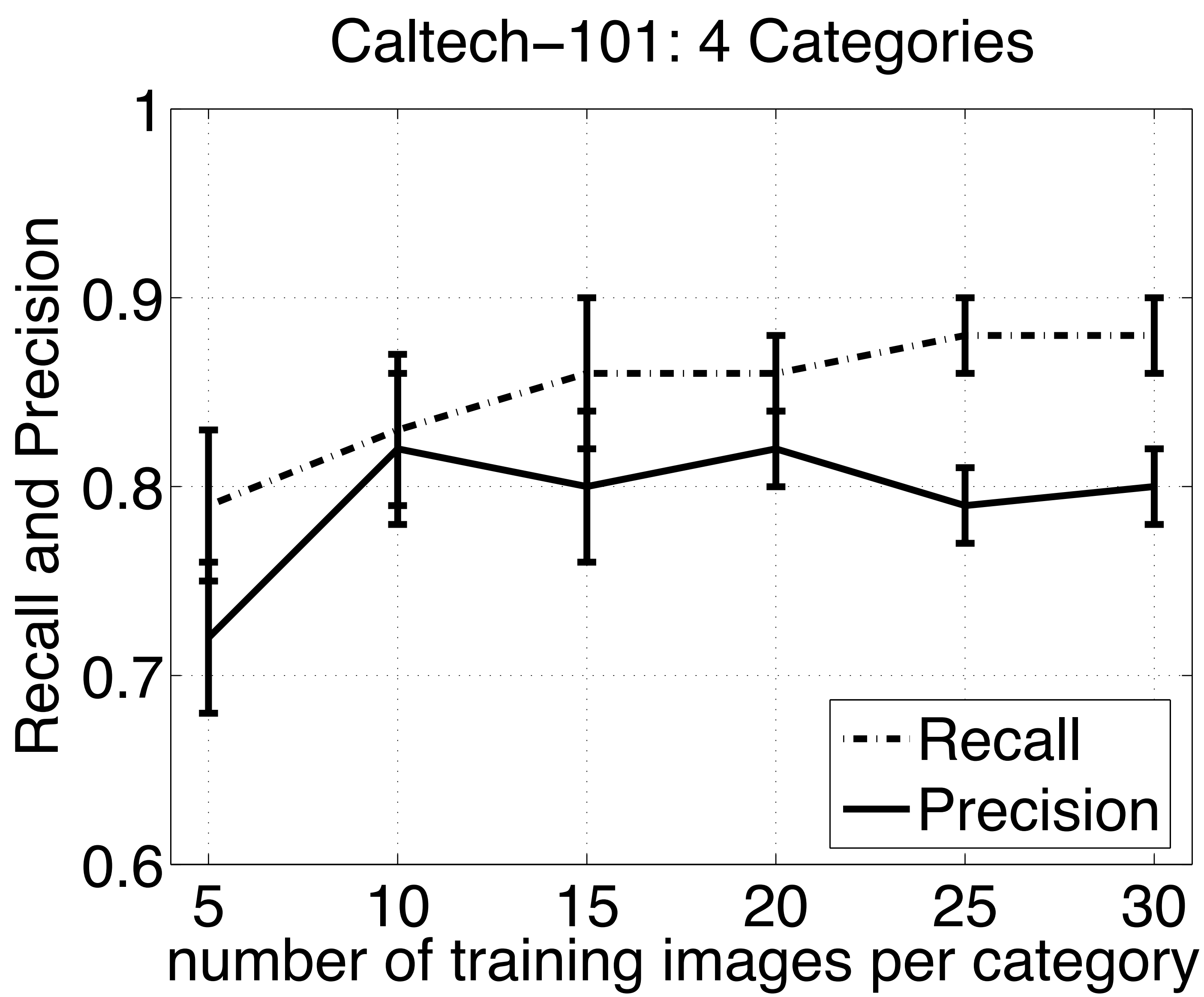
Qualitative Evaluation: Weizmann Horses

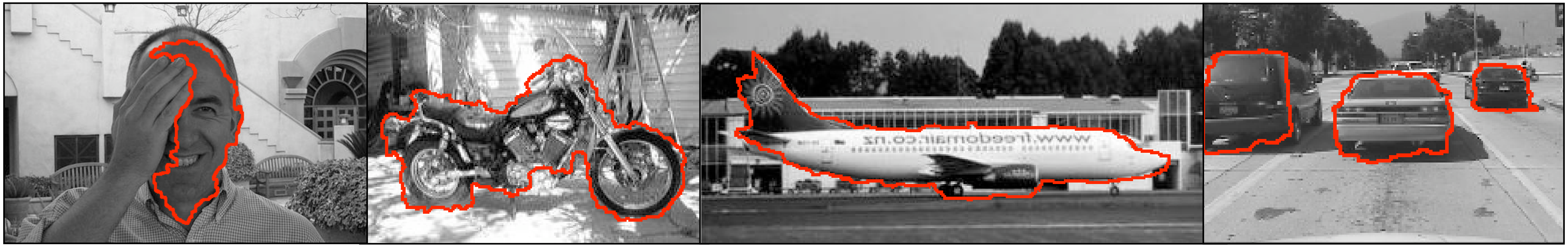
- **We handle:**
  - **Translation, in-plane rotation, scale, articulation, partial occlusion, clutter**
- **Segmentation is good even when object boundaries are:**
  - **Jagged, blurred, form complex topology**
- **Problem: Low-contrast regions that do not form category-specific subtrees within the segmentation trees**

Qualitative Evaluation: Discriminative Unshared Parts



Quantitative Evaluation: Detection





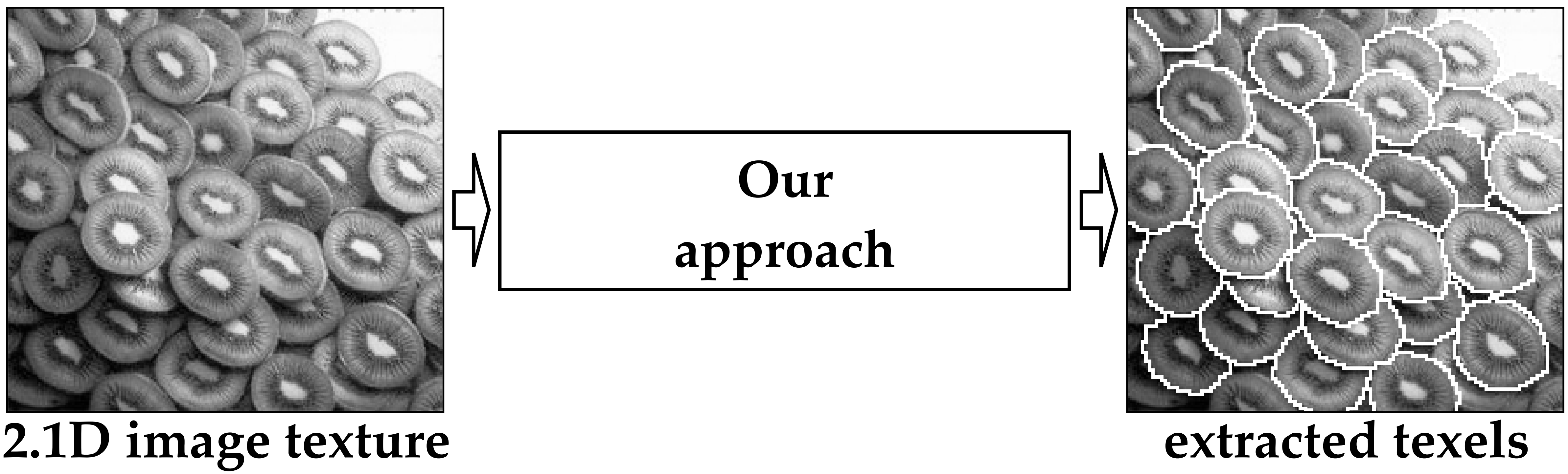
Quantitative Evaluation: Detection, Segmentation, Recognition

	Horses	Cows	Deer	Sheep	Goats	Camels
Total number	88	166	82	135	136	108

	Horses	Cows	Deer	Sheep	Goats	Camels
Recall %	78.9±12.3	75.6±14.8	84.3±5.9	78.2±10.4	72.1±9.5	86.6±8.1
Precision %	82.8±7.5	79.9±11.7	82.2±4.9	78.1±7.2	78.8±5.3	86.2±7.2
Seg. error %	16.1±7.3	18.1±4.2	12.2±7.24	25.9±8.2	21.3±11.2	12.1±4.2
Rec. error %	8.6±3.2	7.2±4.1	9.2±2.4	9.2±6.1	15.9±6.4	3.6±4.9

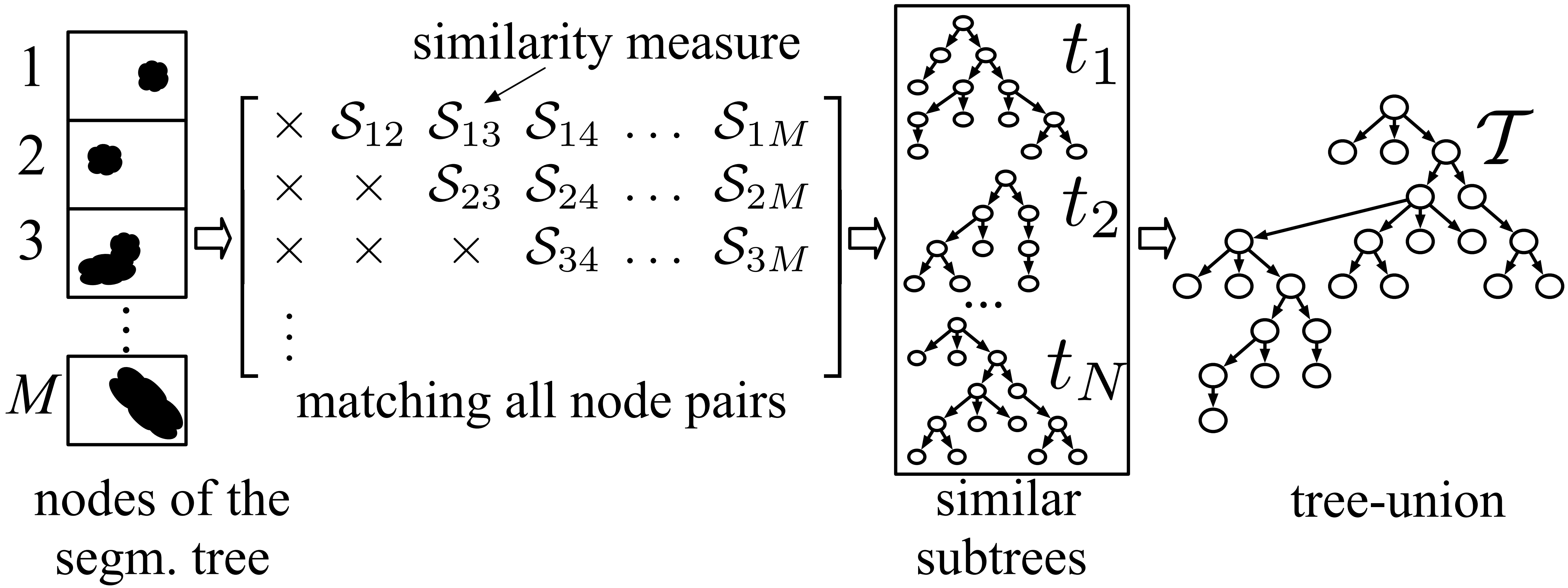
Table 1: Average recall, precision, segmentation, and recognition errors (in %)

Generality of Our Approach: Texture Modeling



- **Texels = Images of spatially recurring physical texture elements**
- **2.1 D texture: thin patches overlaying one another along the surface**

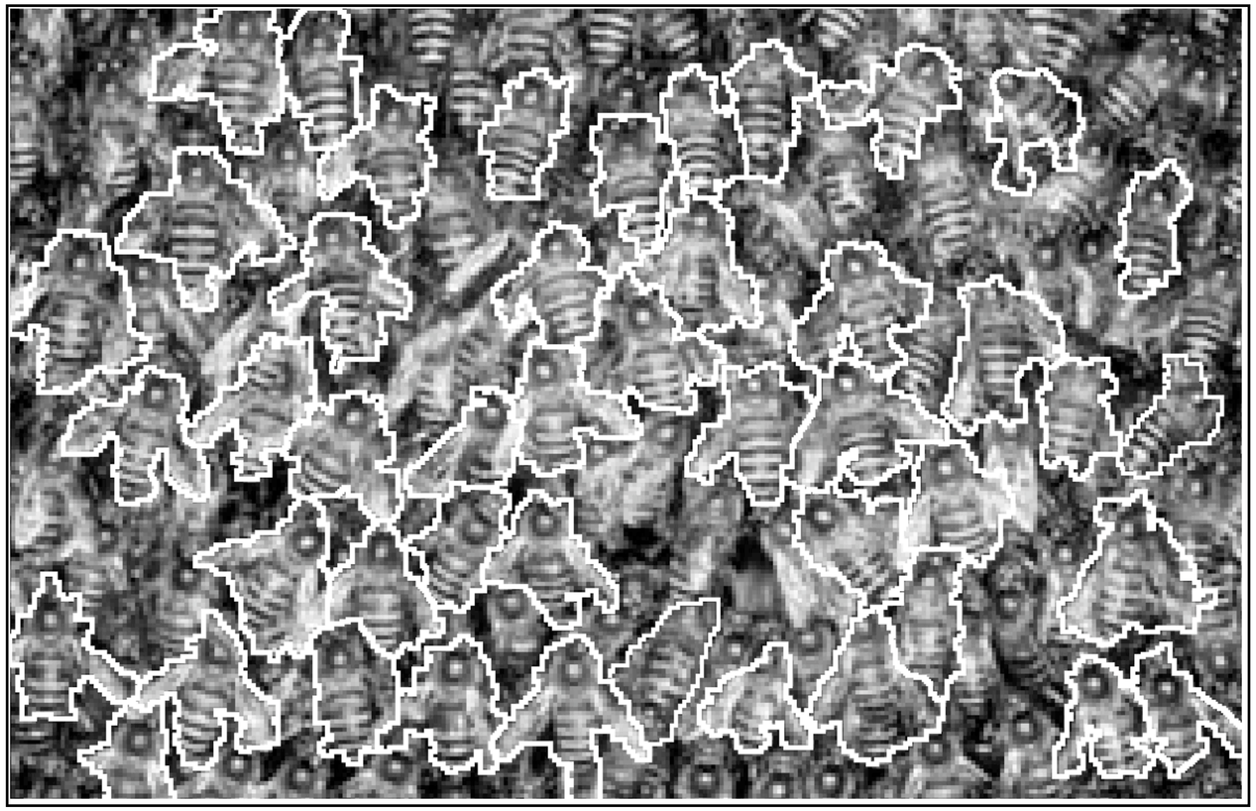
Texel Extraction: Outline of Our Approach

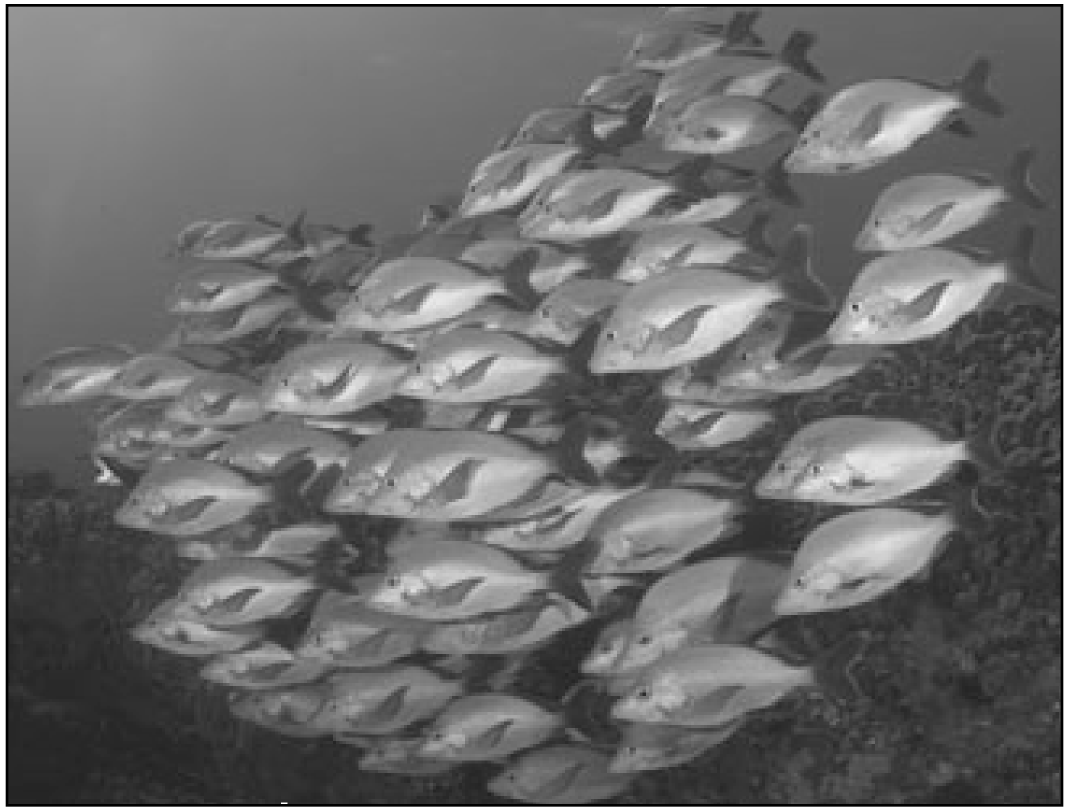


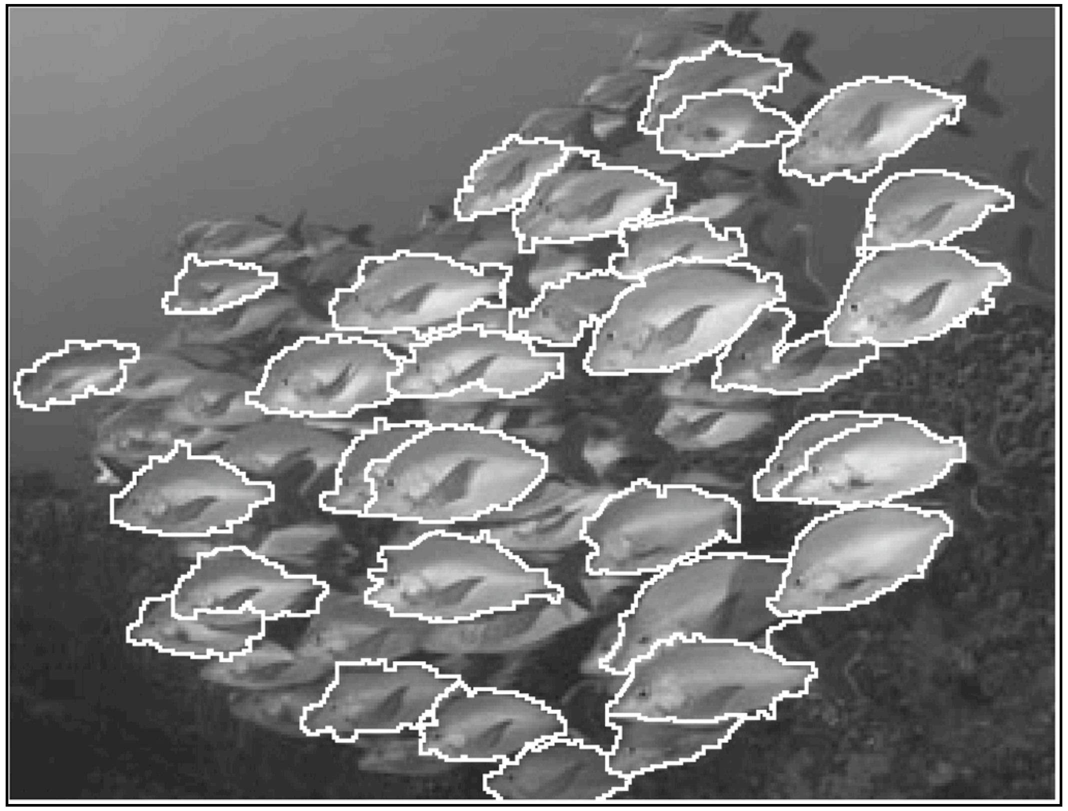
- **Texels = Largest similar subtrees within the segmentation tree**
- **Extracted similar subtrees may represent only partial views of texels**
- **Tree-union = Model of the entire (unoccluded) texel**
- **New learning algorithm for a probabilistic characterization of texels**

Ahuja&Todorovic 07

Texel Extraction Results







Summary

- **Formulation of two new problems and their first solutions**
- **Definition of a category**
- **<u>Unsupervised, simultaneous</u> learning of the taxonomy and models of multiple categories present in an arbitrary image set**
- **Region-based, structural approach**
- **<u>Simultaneous</u> detection, recognition, and segmentation**
- **Providing a semantic basis of recognition**
- **Small number of training images; no approximate inference**
- **Complexity and accuracy comparable with existing methods**

Acknowledgment

Narendra Ahuja

sintod@uiuc.edu

http://vision.ai.uiuc.edu/~sintod