From Hierarchies of Regions to Image Understanding

Prof. Sinisa Todorovic

OSU Oregon State University

Acknowledgment

UIUC:

Prof. Narendra Ahuja

Oregon State University:

William Brendel

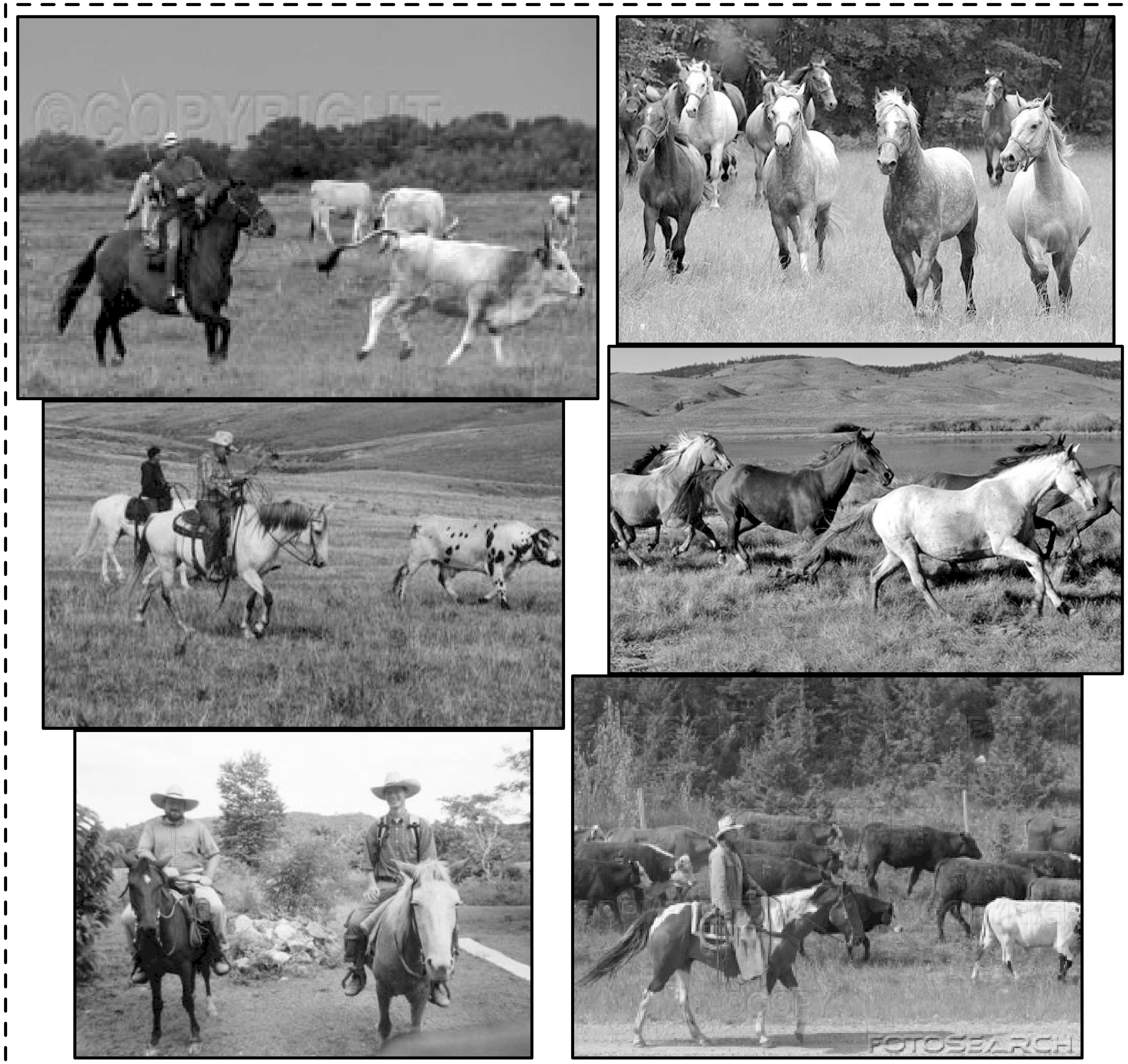
Nadia Payet

Muhamed Amer

Prof. Eugene Zhang

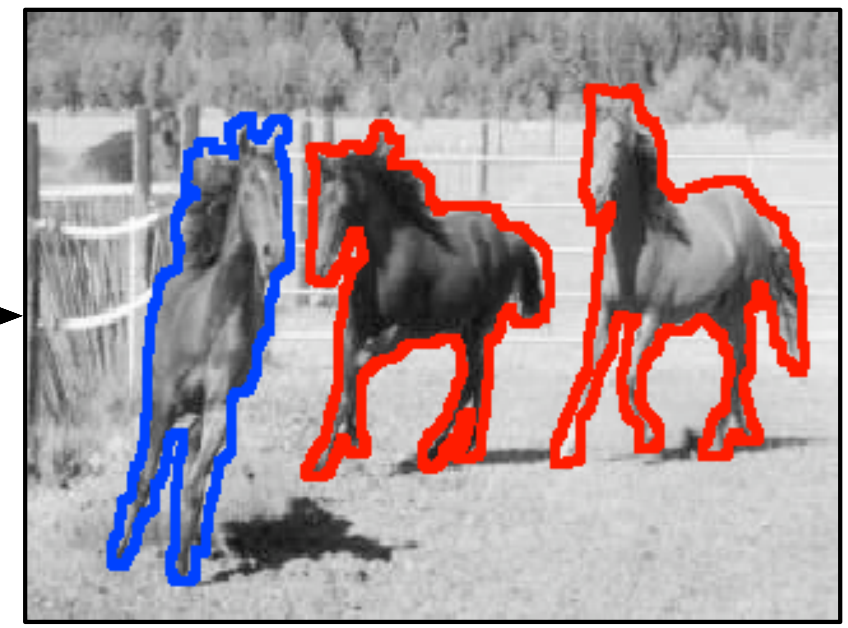
Goals: Object Recognition

input image set



discover and learn all objects present

new image



detect
segment
explain
***all* occurrences**
of the learned objects

Goal: Video Painterly Rendering

video sequence enhanced with multiple painting styles
-- one per each object

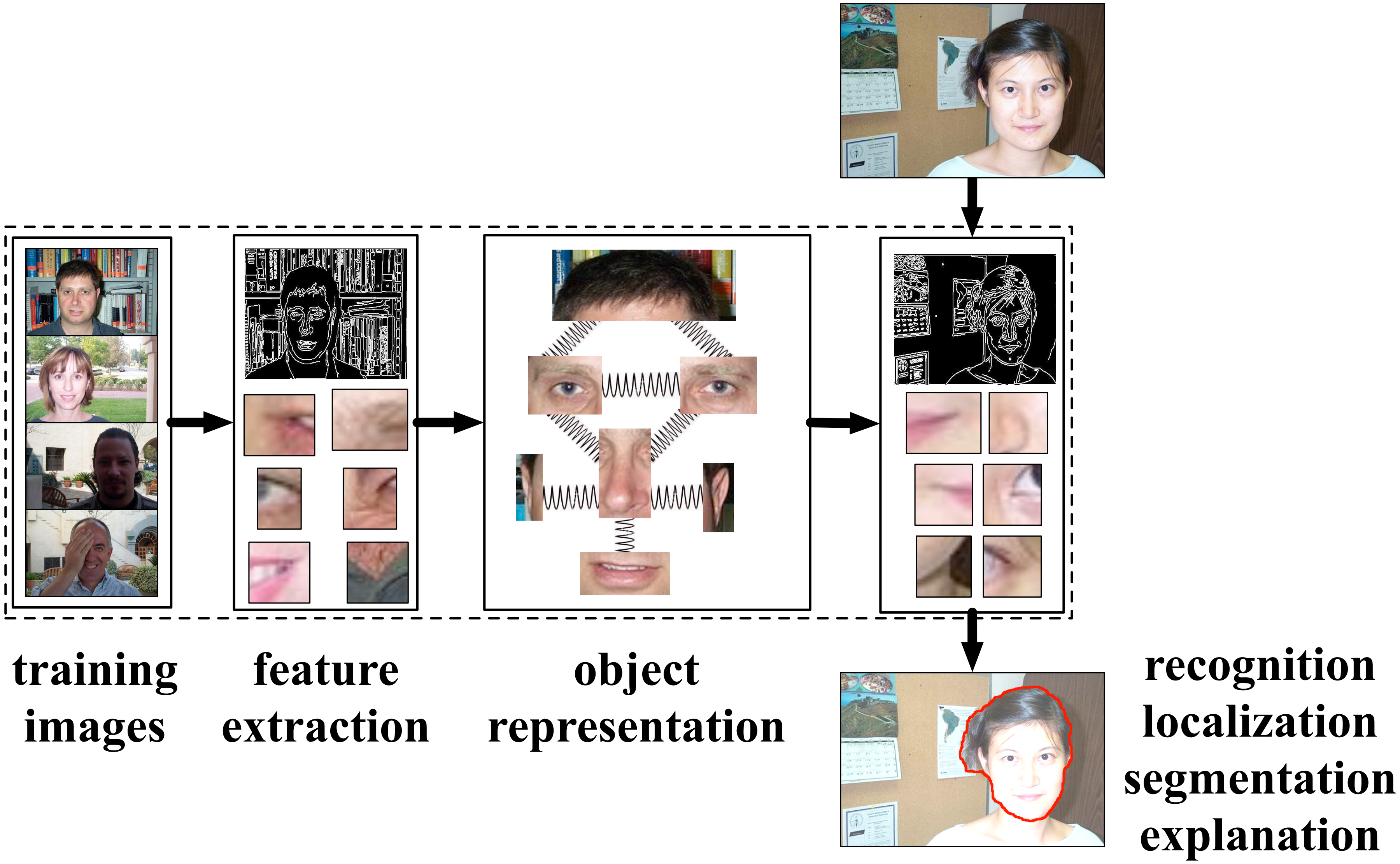
flower petals = van Gogh
stamens = expressionism
background = pointilism

Goals: Texel-based Texture Segmentation

Many applications require unsupervised partitioning of the image into textured and non-textured subimages

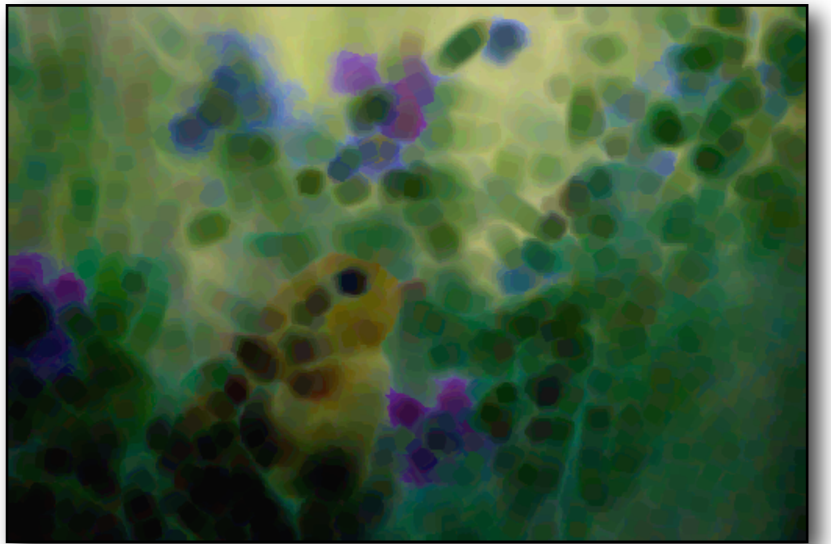
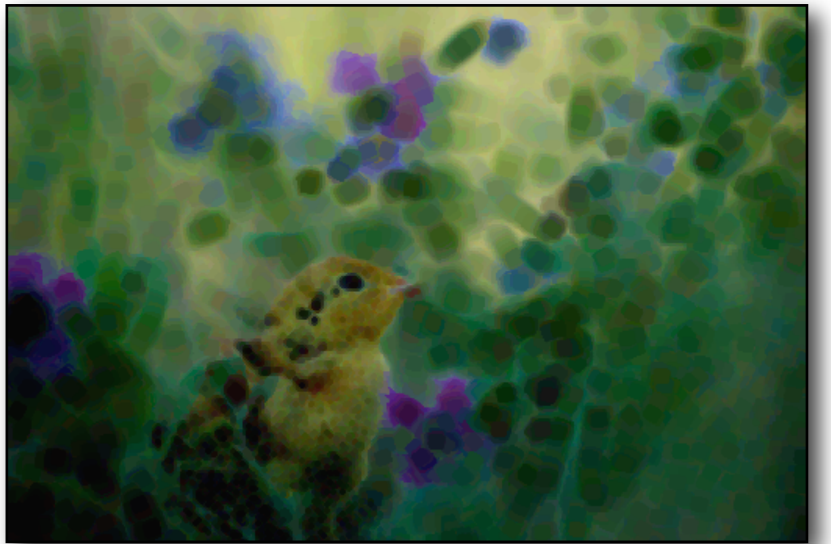
Prior Work: Object Recognition

PRIOR WORK	OUR EXTENSIONS
high degree of supervision	relaxing supervision requirements
predominance of keypoint features	using richer features: regions
ignoring the spatial info	accounting for multiscale spatial info
limited goals	unified framework for many goals



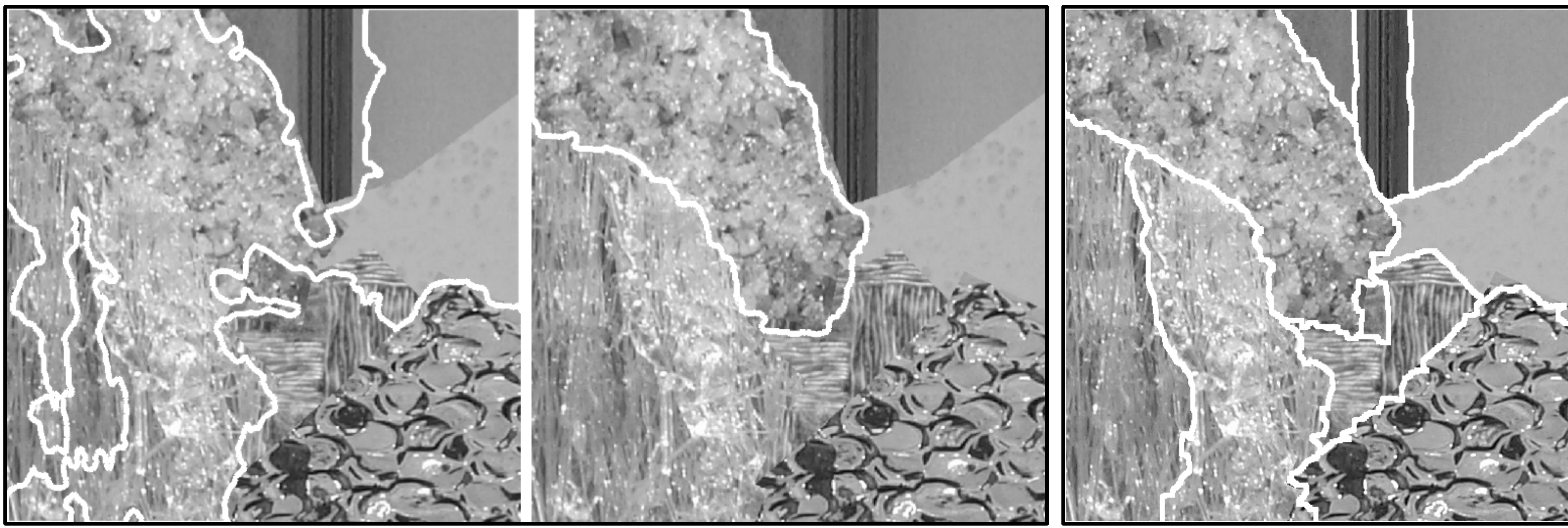
Prior Work: Painterly Rendering

PRIOR WORK	OUR EXTENSIONS
Uses only a single style	Object-based multiple styles
Unrealistic, poor artistic expression	Rich artistic expression

Prior Work: Texture Segmentation

PRIOR WORK	OUR EXTENSIONS
Uses a pre-specified bank of filters	Extraction of texels
Assumptions: smoothness, scale	Relaxing the assumptions



meanshift active contours our results

WHAT IS AN OBJECT?

Properties of Objects

3D objects in the scene $\longrightarrow$ 2D objects in the image

cohesive	occupy regions
form characteristic spatial configurations with other objects	context
have parts	subregions
parts have characteristic spatial layout	spatial layout of subregions

Rationale for Learning – Like a Small Child

input images

It is likely to be meaningful:

- **If some parts repeat in the set of images**
- **If some configurations of the learned parts repeat in the set**

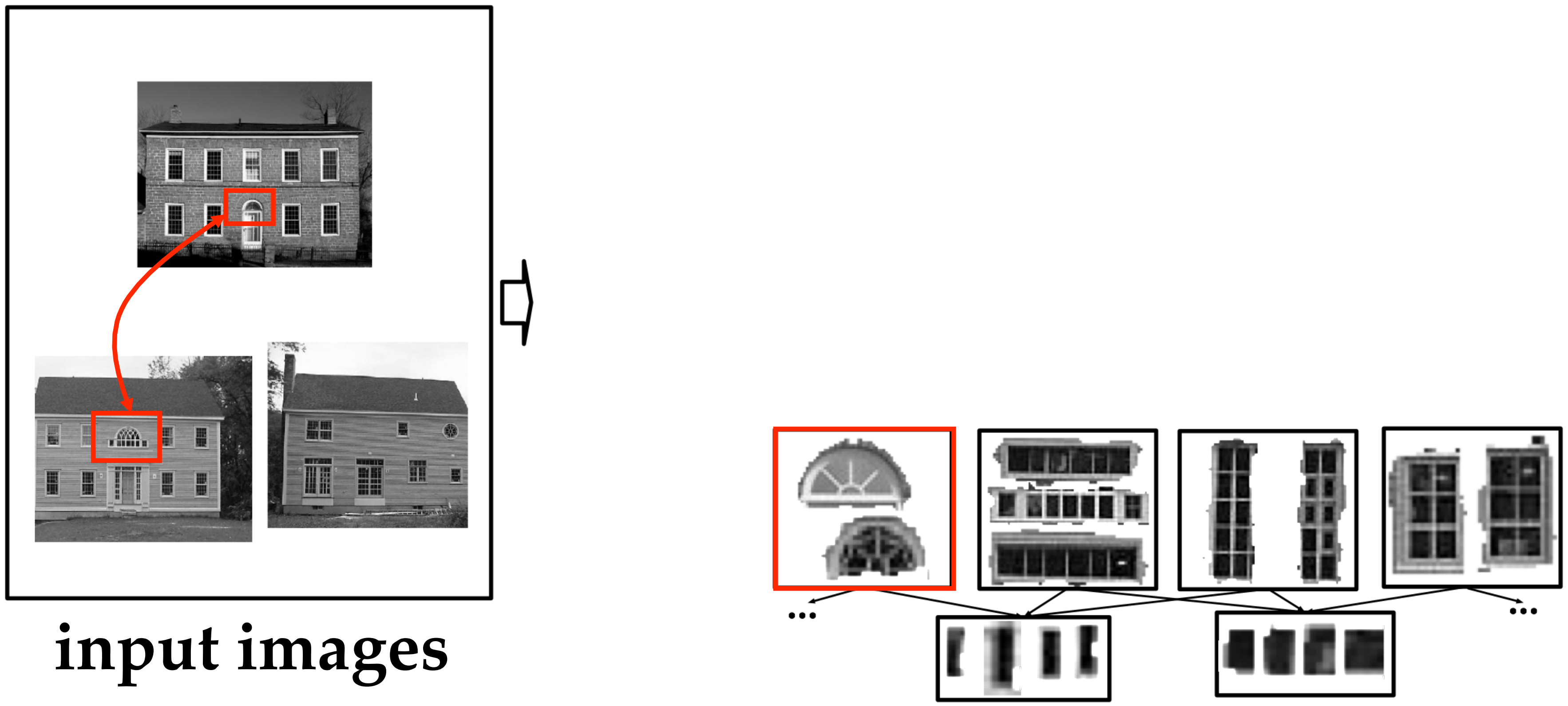
Rationale for Learning – Like a Small Child

input images

It is likely to be meaningful:

- **If some parts repeat in the set of images**
- **If some configurations of the learned parts repeat in the set**

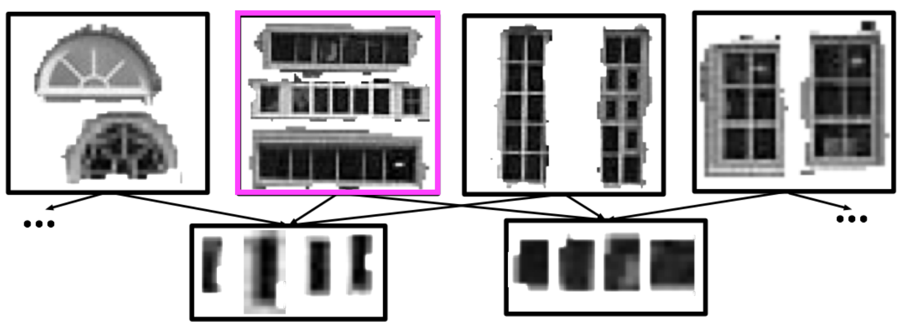
Rationale for Learning – Like a Small Child



It is likely to be meaningful:

- **If some parts repeat in the set of images**
- **If some configurations of the learned parts repeat in the set**

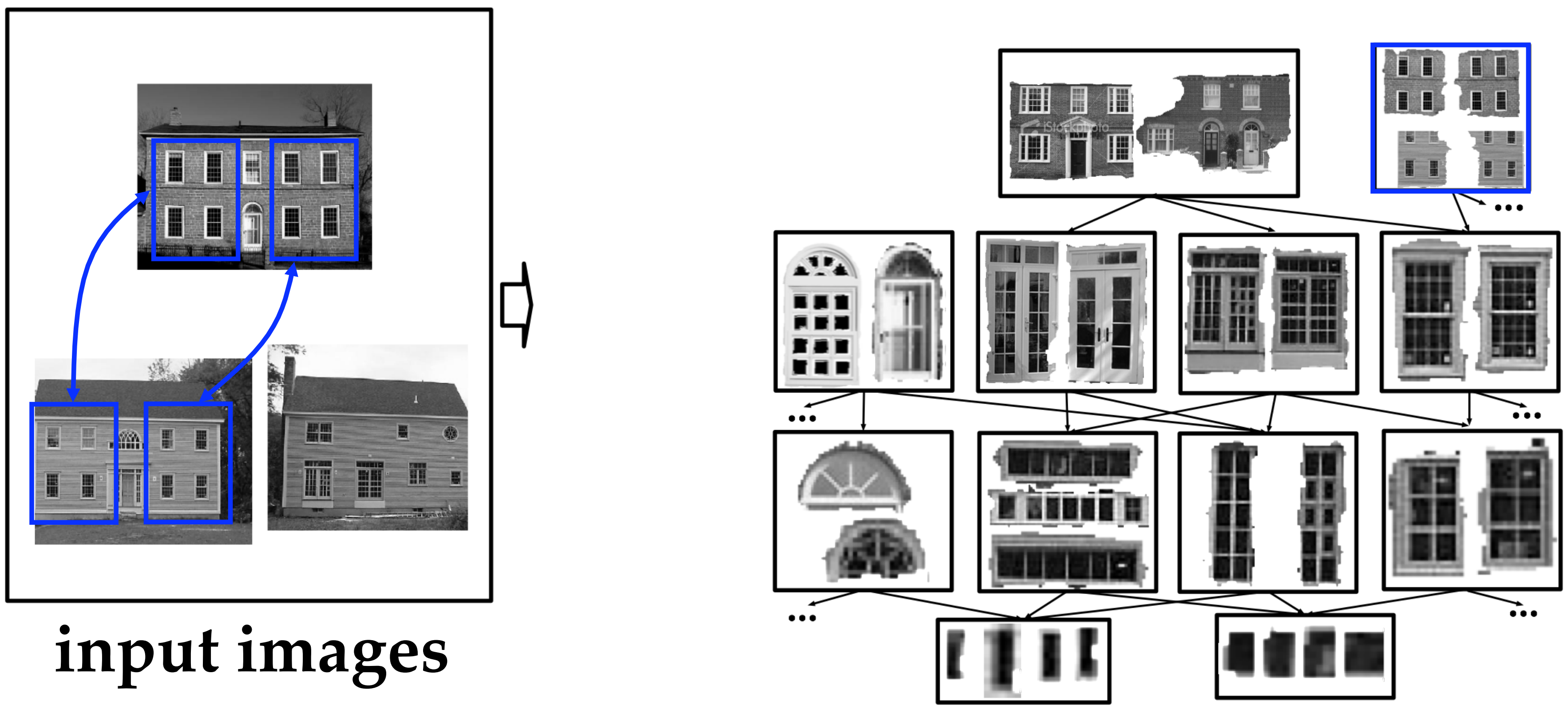
Rationale for Learning – Like a Small Child



It is likely to be meaningful:

- **If some parts repeat in the set of images**
- **If some configurations of the learned parts repeat in the set**

Rationale for Learning – Like a Small Child



It is likely to be meaningful:

- **If some parts repeat in the set of images**
- **If some configurations of the learned parts repeat in the set**

Category = Set of Recurring Similar 2D Objects

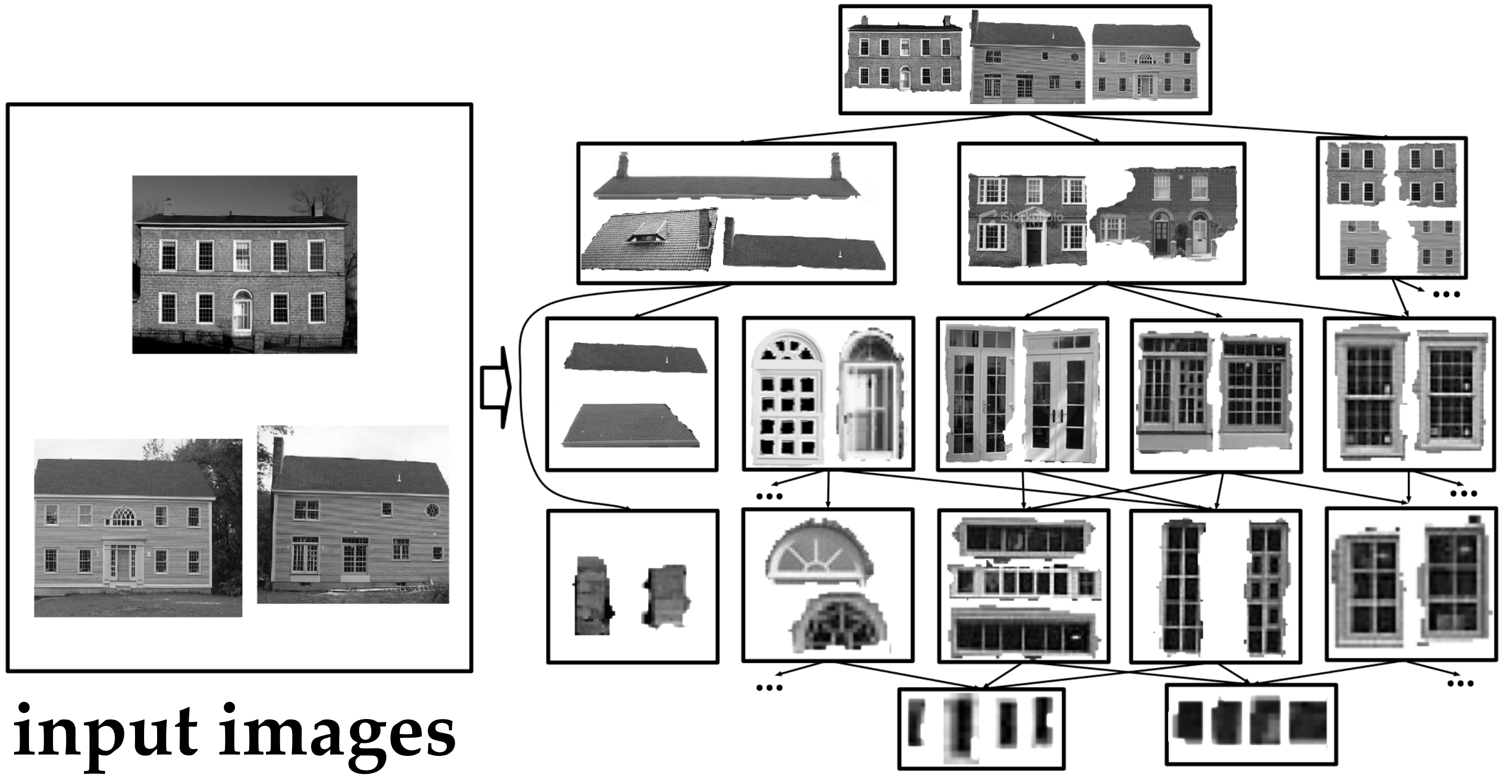
(1) Photometric (e.g., color)

(2) Geometric (e.g., area, shape)

(3) Structural:

spatial layout of subcategories

containment of subcategories



input images

Category = Set of Recurring Similar 2D Objects

(1) Photometric (e.g., color)

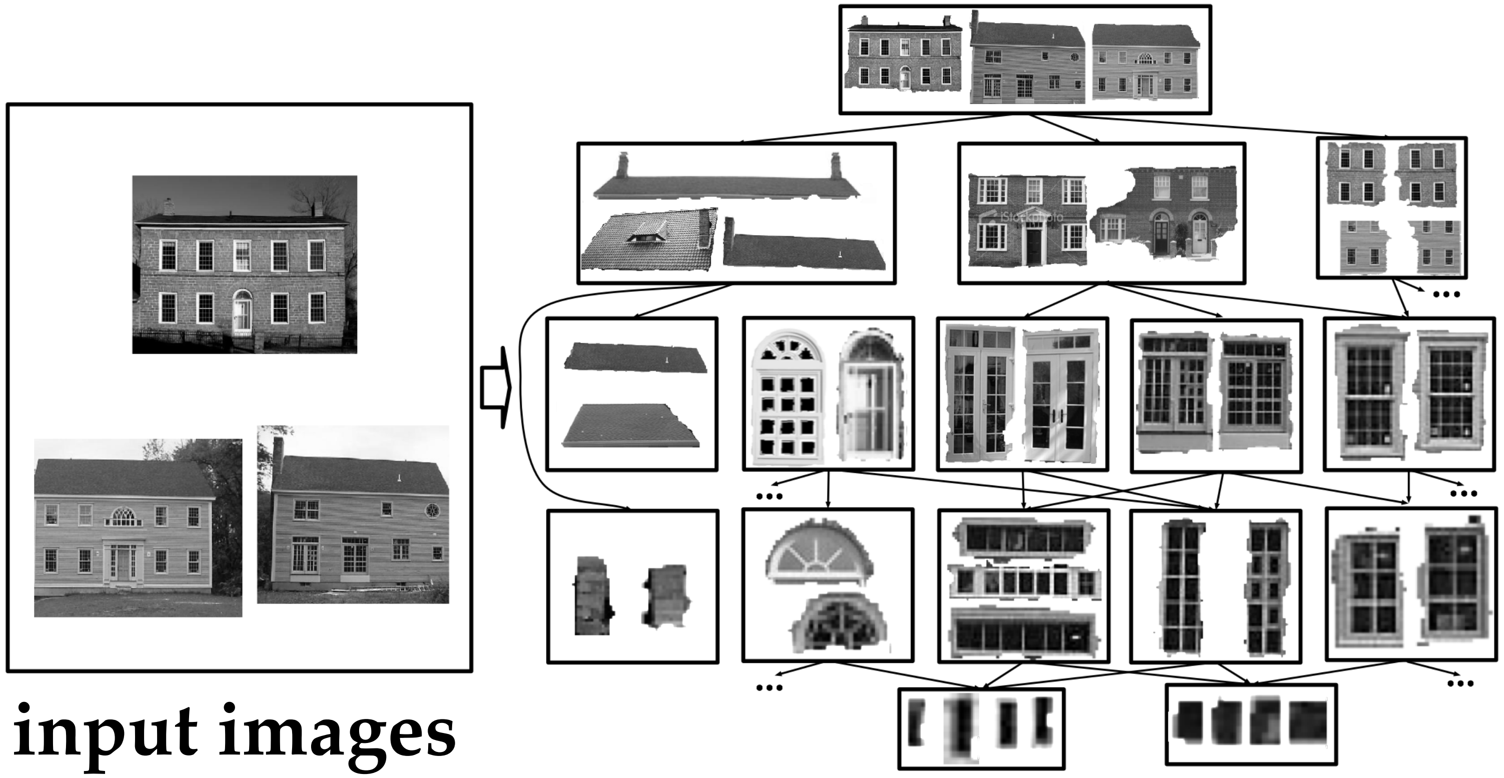
(2) Geometric (e.g., area, shape)

(3) Structural:

spatial layout of subcategories

containment of subcategories

recursive definition



input images

Problem Statement

Problem Statement

- **Given a set of images**

Problem Statement

- **Given a set of images**
- **Discover frequently occurring 2D *objects***
  - **Under illumination and scale changes**
  - **Amidst background clutter**
  - **Under partial occlusion**

Problem Statement

- **Given a set of images**
- **Discover frequently occurring 2D *objects***
  - **Under illumination and scale changes**
  - **Amidst background clutter**
  - **Under partial occlusion**
- **Learn their generative, statistical models**

Problem Statement

- **Given a set of images**
- **Discover frequently occurring 2D *objects***
  - **Under illumination and scale changes**
  - **Amidst background clutter**
  - **Under partial occlusion**
- **Learn their generative, statistical models**
- **Use the models for**
  - **Object recognition**
  - **Object-based painterly rendering and synthesis**
  - **Texel-based texture segmentation**

Rest of the Talk

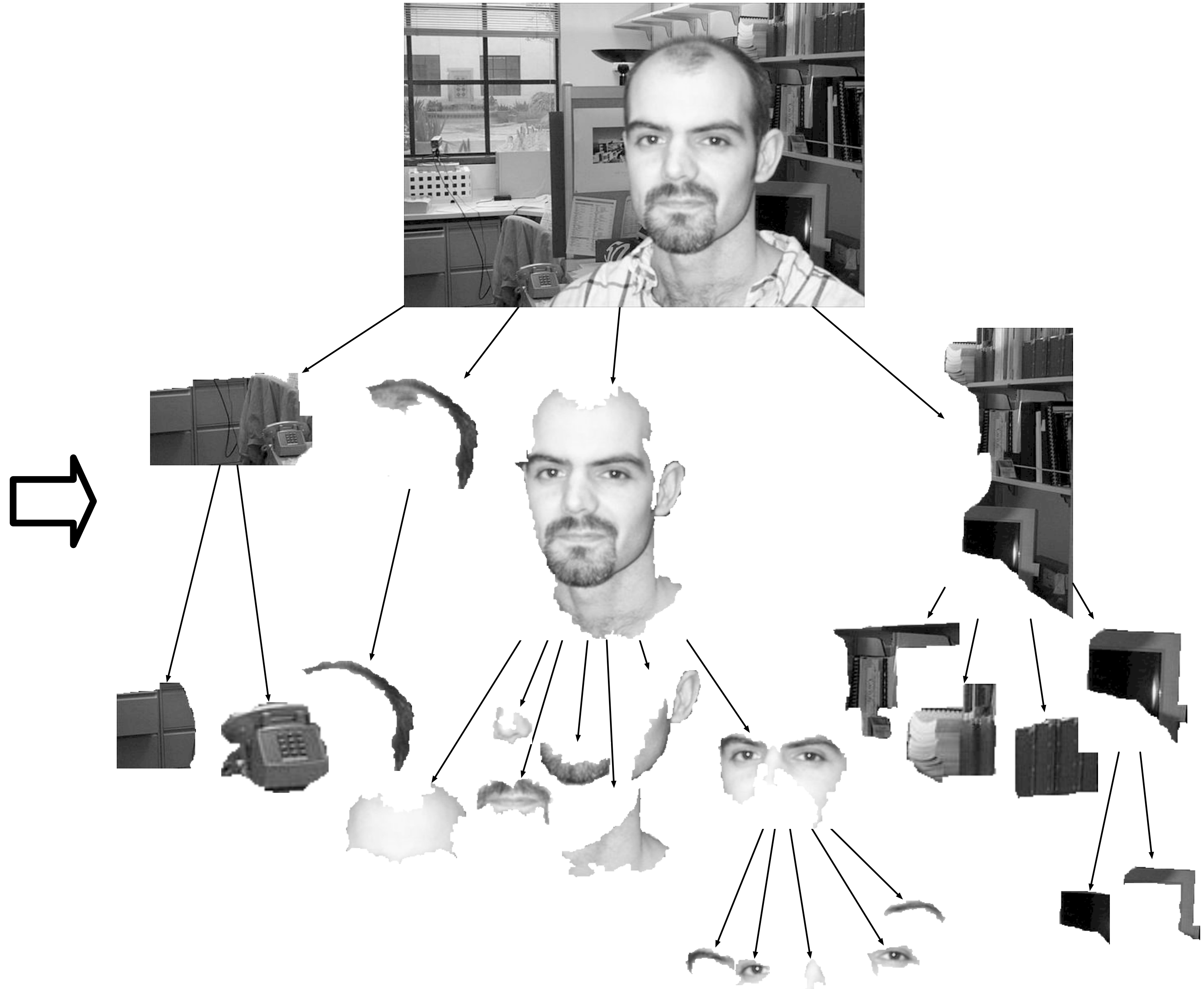
1. **Image representation = Hierarchy of regions**

2. **Region matching**

3. **Applications and results**

Image = Tree ⇒ Object = Subtree

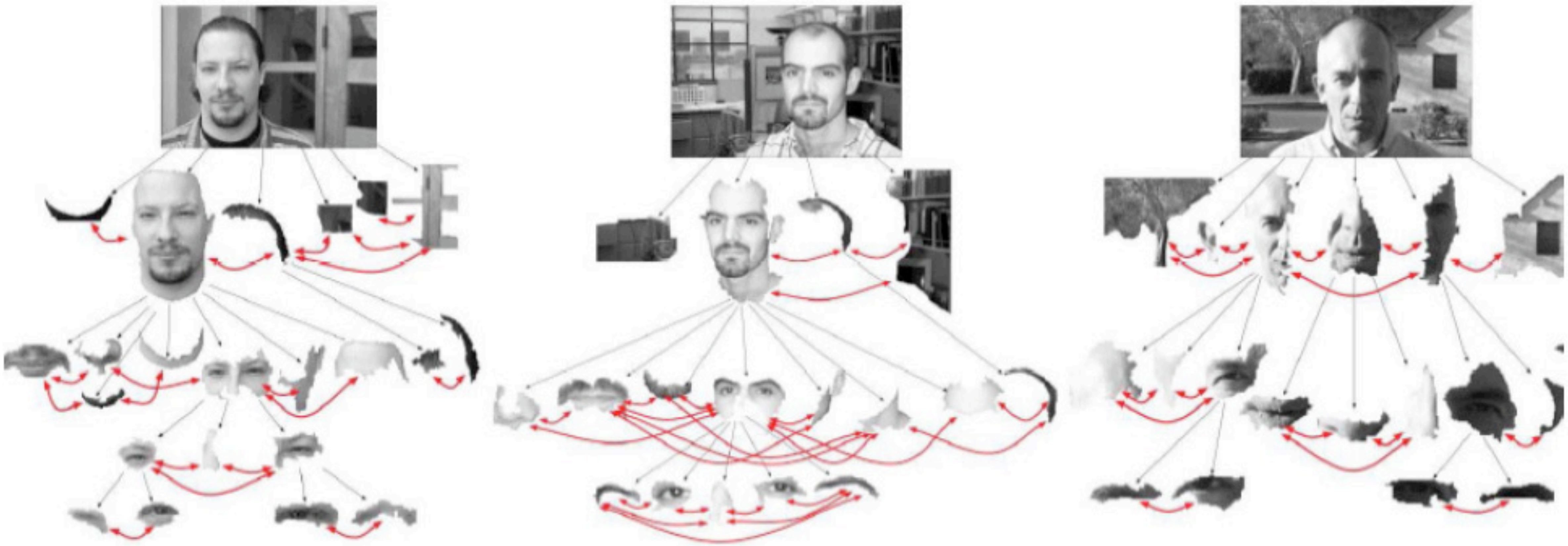
multiscale segmentation



segmentation tree

Ahuja PAMI96, Tobb & Ahuja TIP97, Arora&Ahuja ICPR06

Connected Segmentation Trees



Lateral links = Region neighbor relations

Hierarchical links = Region embedding

Ahuja&Todorovic CVPR08

Region Properties Associated with Each Node

- **Gray-level contrast with surround**
- **Boundary shape**
- **Displacement of centroids**
- **Orientation**

$\vdots$

Properties relative wrt parent $\Rightarrow$ Scale and in-plane rotation invariance

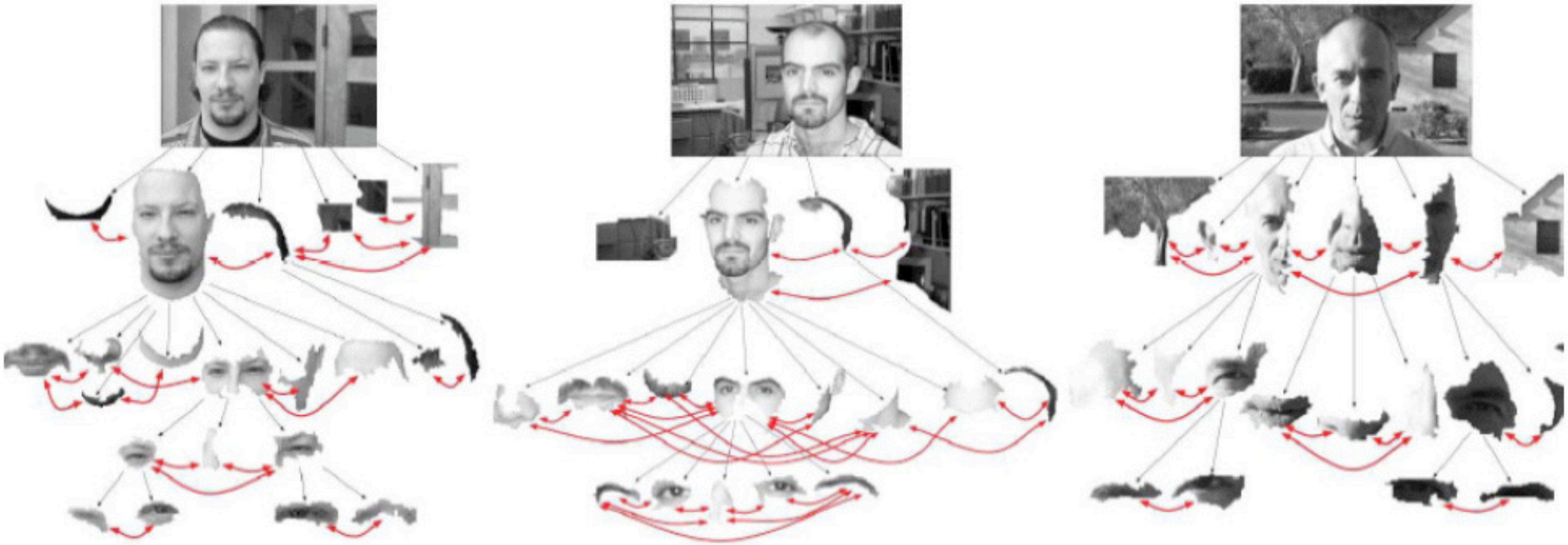
Rest of the Talk

1. **Image representation = Hierarchy of regions**

2. **Region matching**

3. **Applications and results**

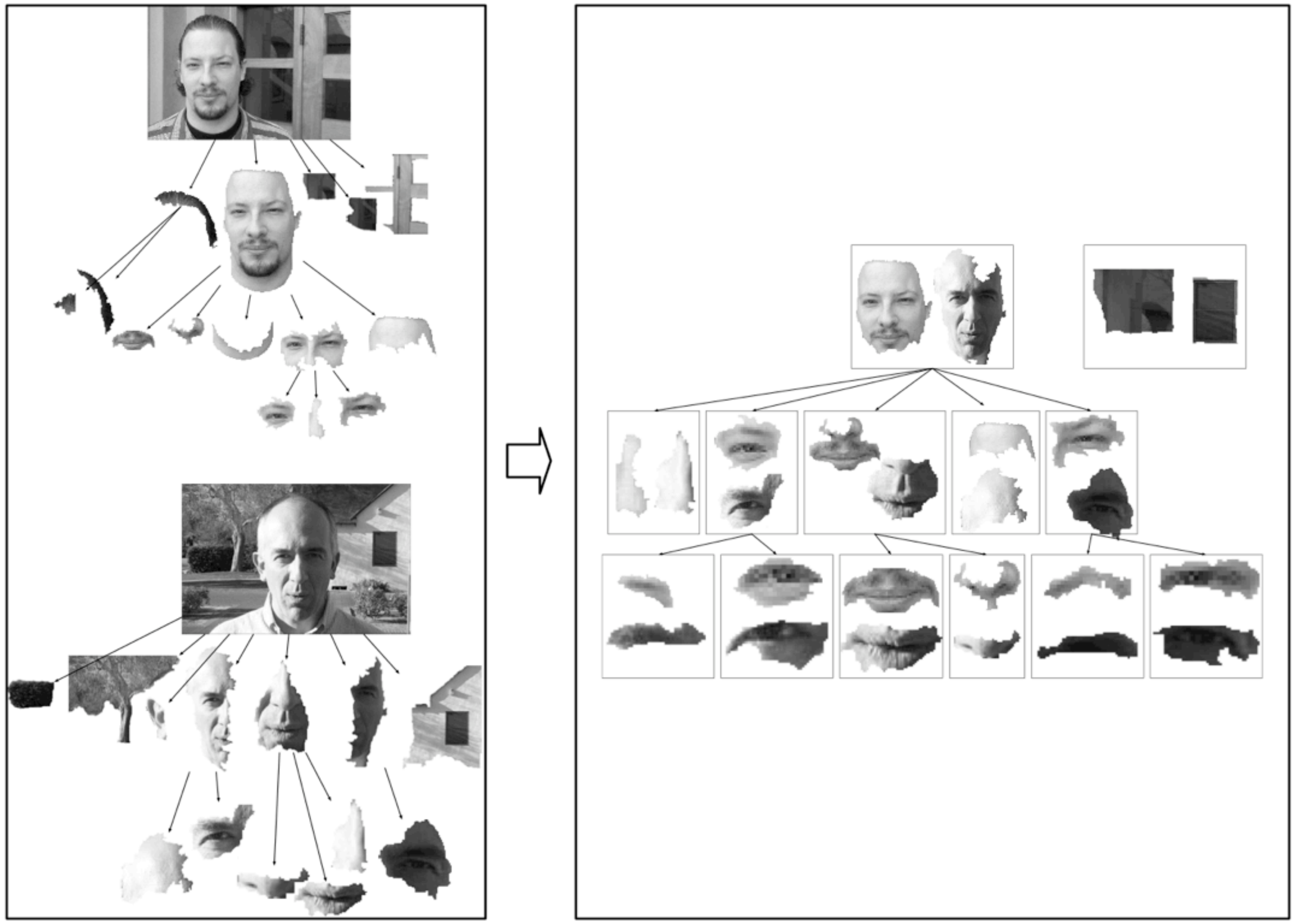
How to Discover Repeating Image Parts?



Object category is present = Many similar subgraphs

Discovering objects = Graph matching

Graph Matching = Subgraph Isomorphism



Match two regions

- **If their immediate properties are similar**
- **AND the same holds for their subregions**
- **AND the same holds for their neighbors**

Graph Matching: Formulation

Given two graphs: $G = (V, E)$ **and** $G' = (V', E')$

Find the mapping $f = \{(v, v')\} \subset V \times V'$

which minimizes their cost of matching:

$$COST_{GG'} = \min_f \left[\sum_{(v,v') \in f} \psi_{vv'} + \sum_{(v,v',u,u') \in f \times f} \phi_{vv'uu'} \right]$$

unary potential
function of region properties

pairwise potential
function of spatial relationships

Graph Matching: Formulation

Linearization by introducing an indicator vector

$$\boldsymbol{X} = [0\ 0\ 1\ 0\ 0\ 1\ \ldots\ 0\ 1\ 0]^T$$

matched pair (v, v') **unmatched pair** (u, u')

$$\Downarrow$$

Discrete problem

$$\min_{\boldsymbol{X}} \left[\boldsymbol{\Psi}^T \boldsymbol{X} + \boldsymbol{X}^T \boldsymbol{\Phi} \boldsymbol{X} \right]$$

$$\textbf{s.t.} \quad x_{vv'} \in \{0, 1\}$$

Graph Matching: Formulation

Relaxation of the discrete problem

$$\min_{\boldsymbol{X}} \left[\boldsymbol{\Psi}^T \boldsymbol{X} + \boldsymbol{X}^T \boldsymbol{\Phi} \boldsymbol{X}\right]$$

$$\textbf{s.t.} \quad \forall \boldsymbol{x}_{vv'} \geq \mathbf{0}, \quad \sum_{v} \boldsymbol{x}_{vv'} = \mathbf{1}, \quad \sum_{v'} \boldsymbol{x}_{vv'} = \mathbf{1}$$

Todorovic&Ahuja IJCV08, PAMI08, CVPR06-08, ICCV07, ICPR06-08

Rest of the Talk

1. Image representation = Hierarchy of regions

2. Region matching

3. **Applications and results**

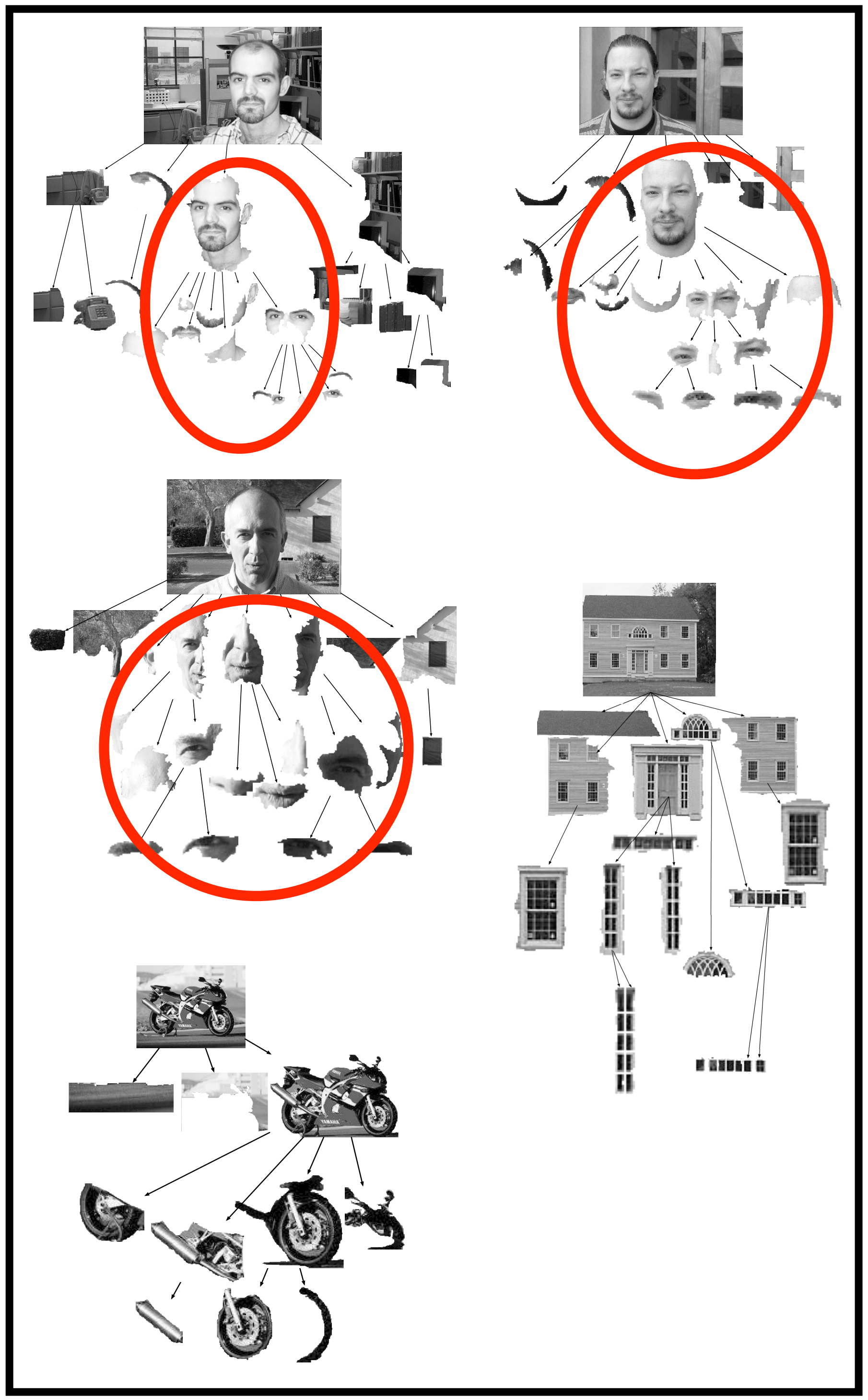
   a. **Object recognition**

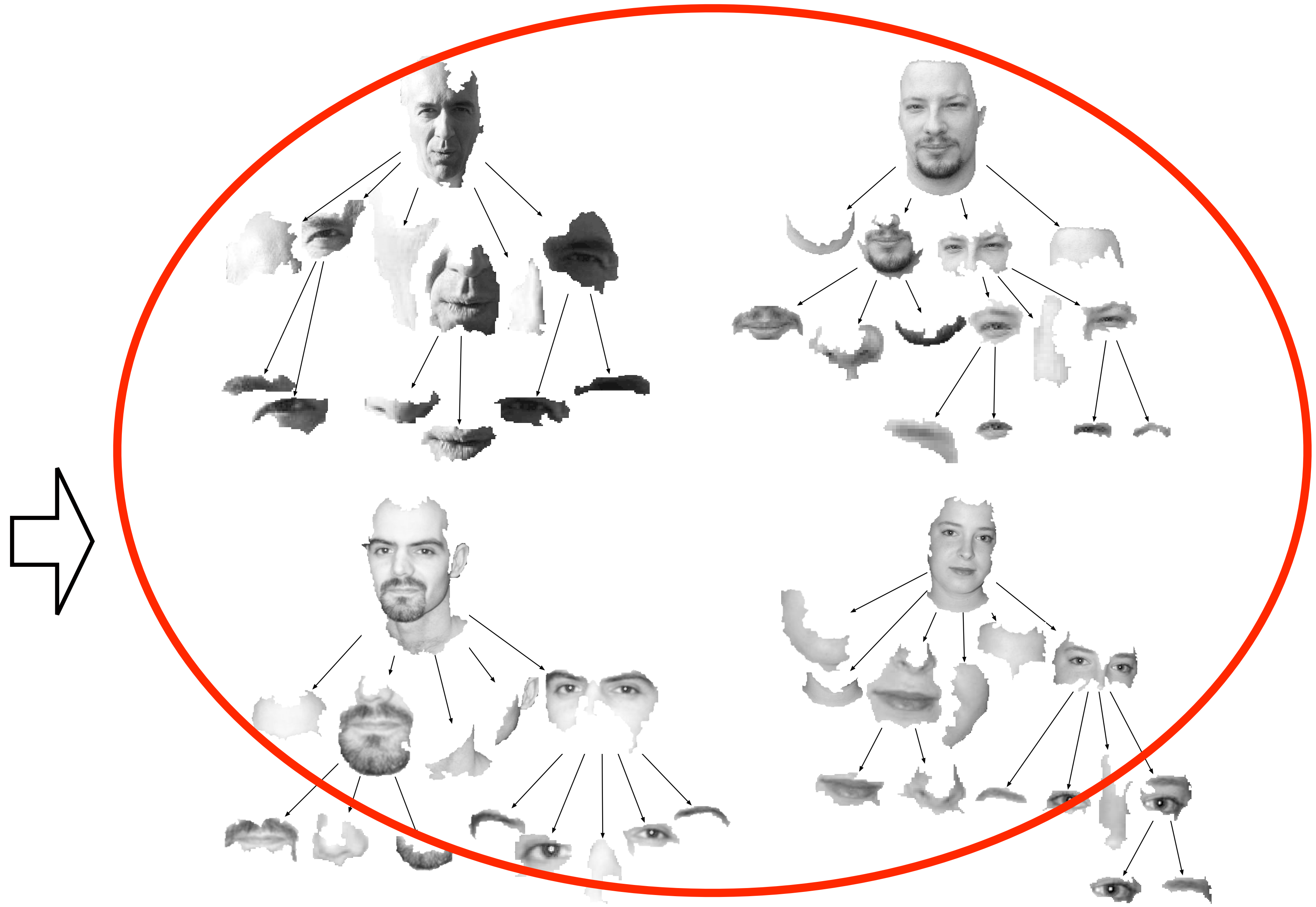
   b. Painterly rendering

   c. Texture segmentation

Discovering Objects = Matching + Clustering

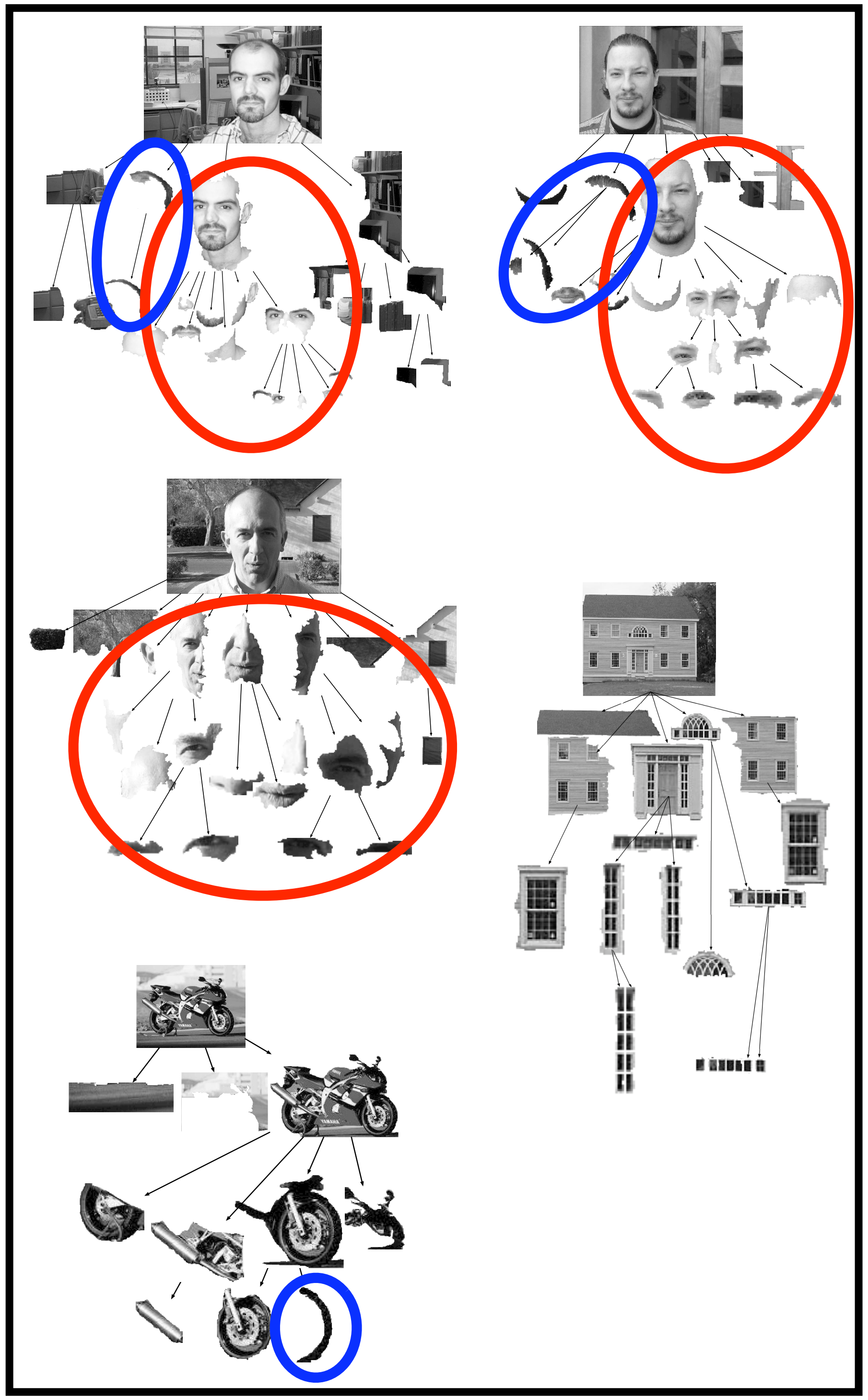
training images

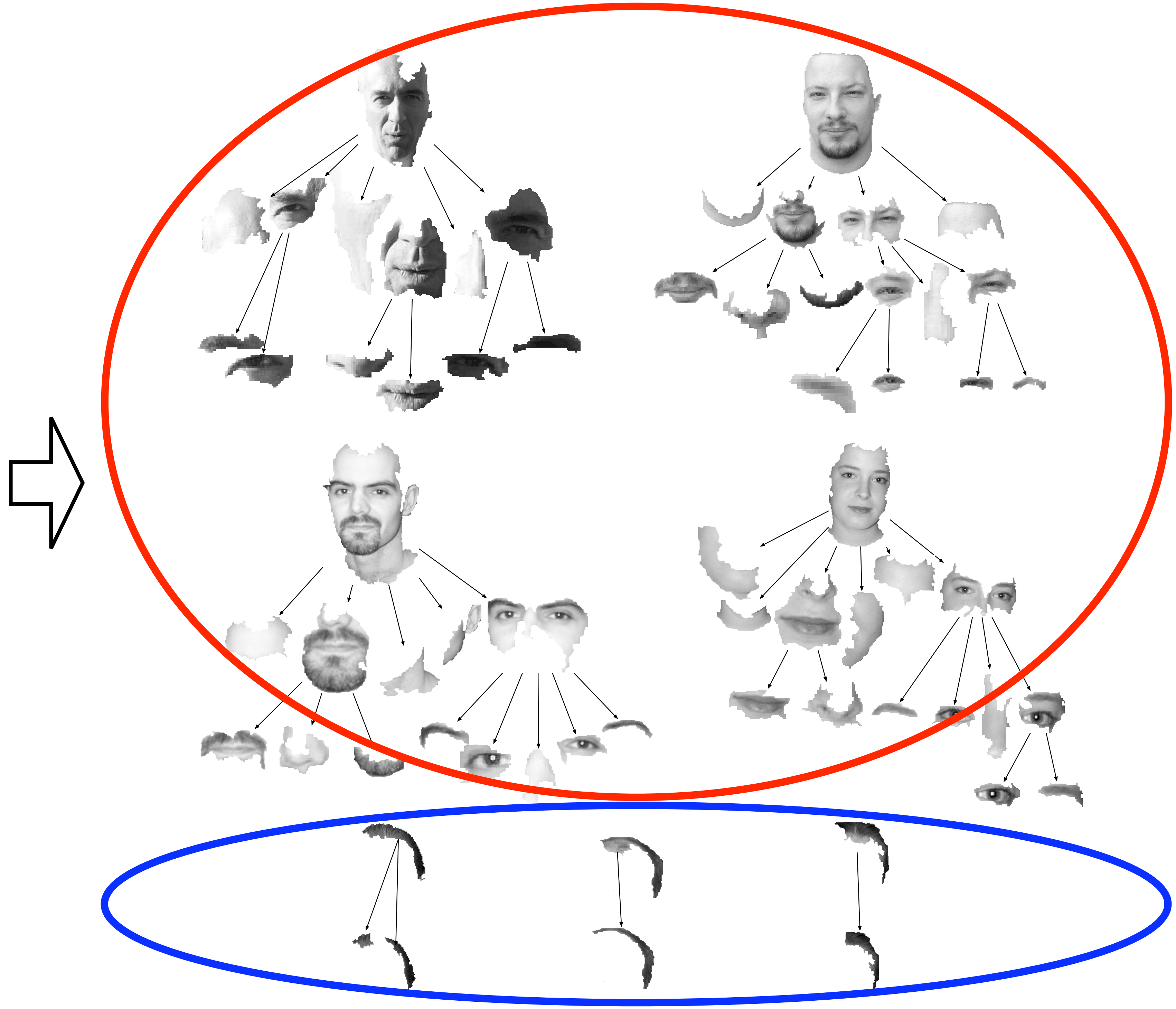




Discovering Objects = Matching + Clustering

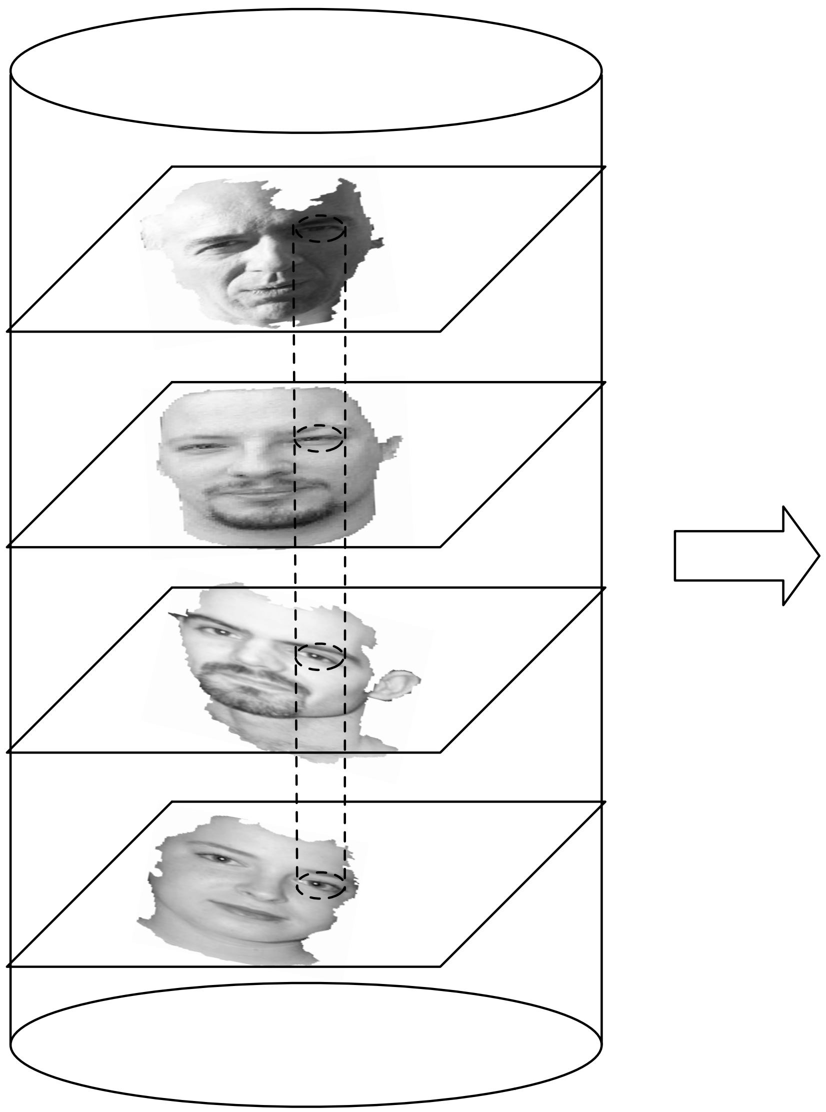
training images



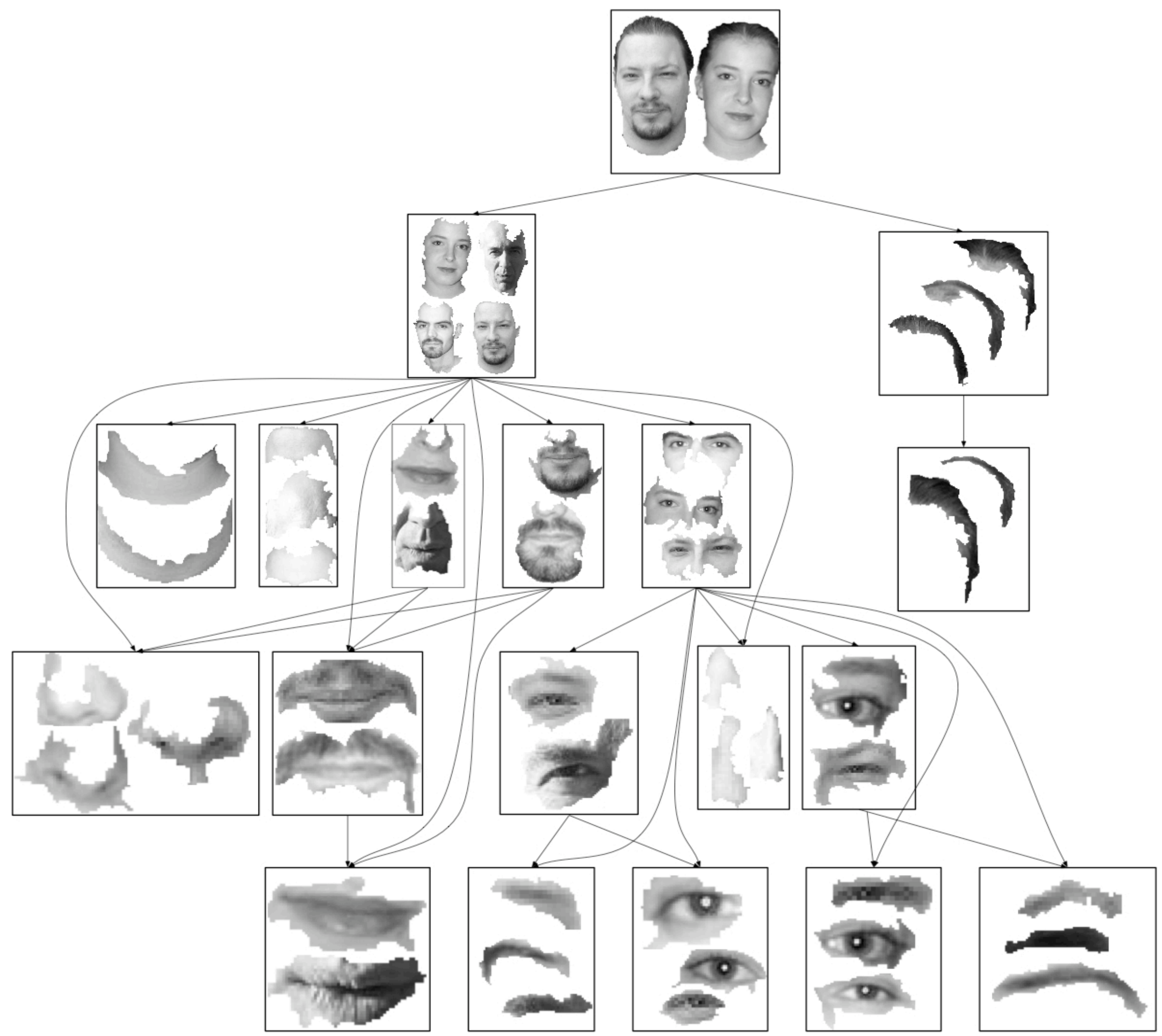


Each cluster = Distinct Object

Learning a Model of Each Cluster = Structural EM



matched subgraphs

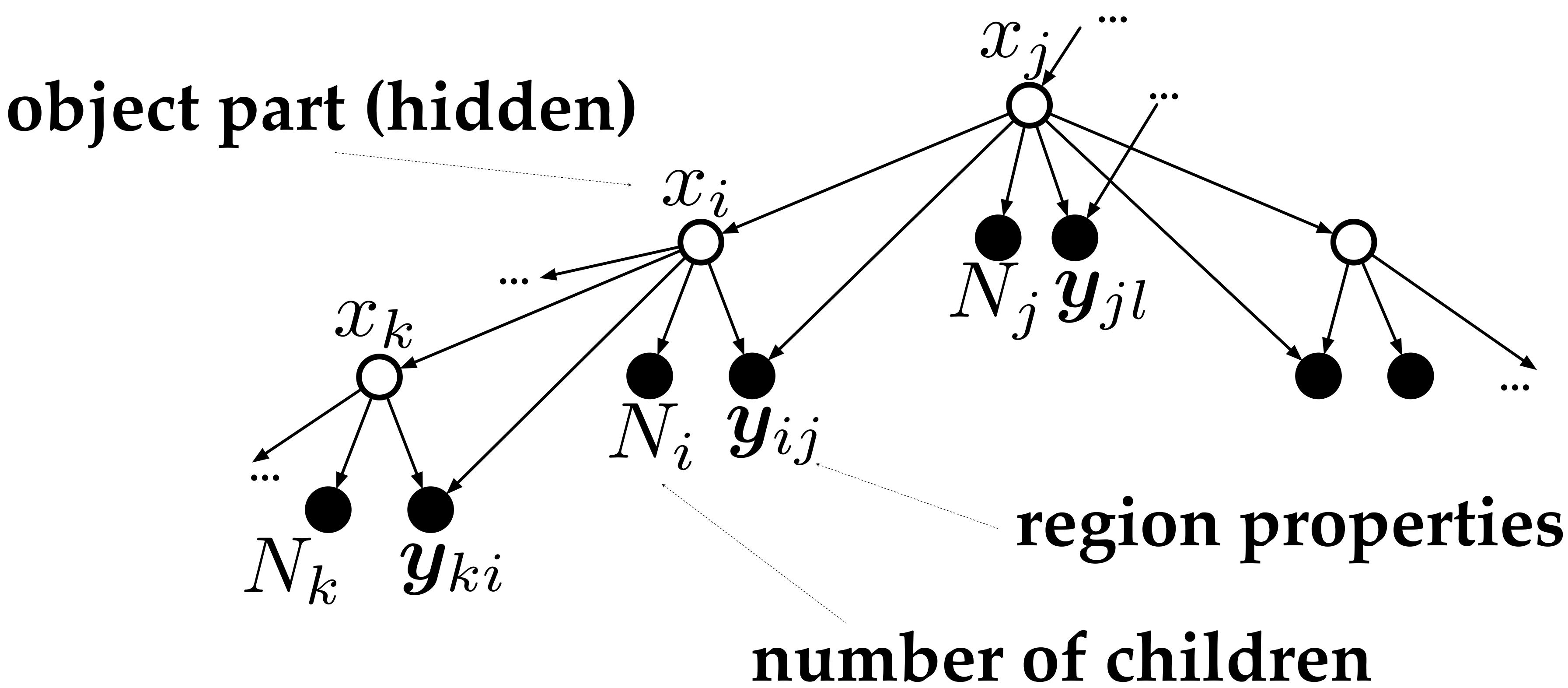


hierarchical object model

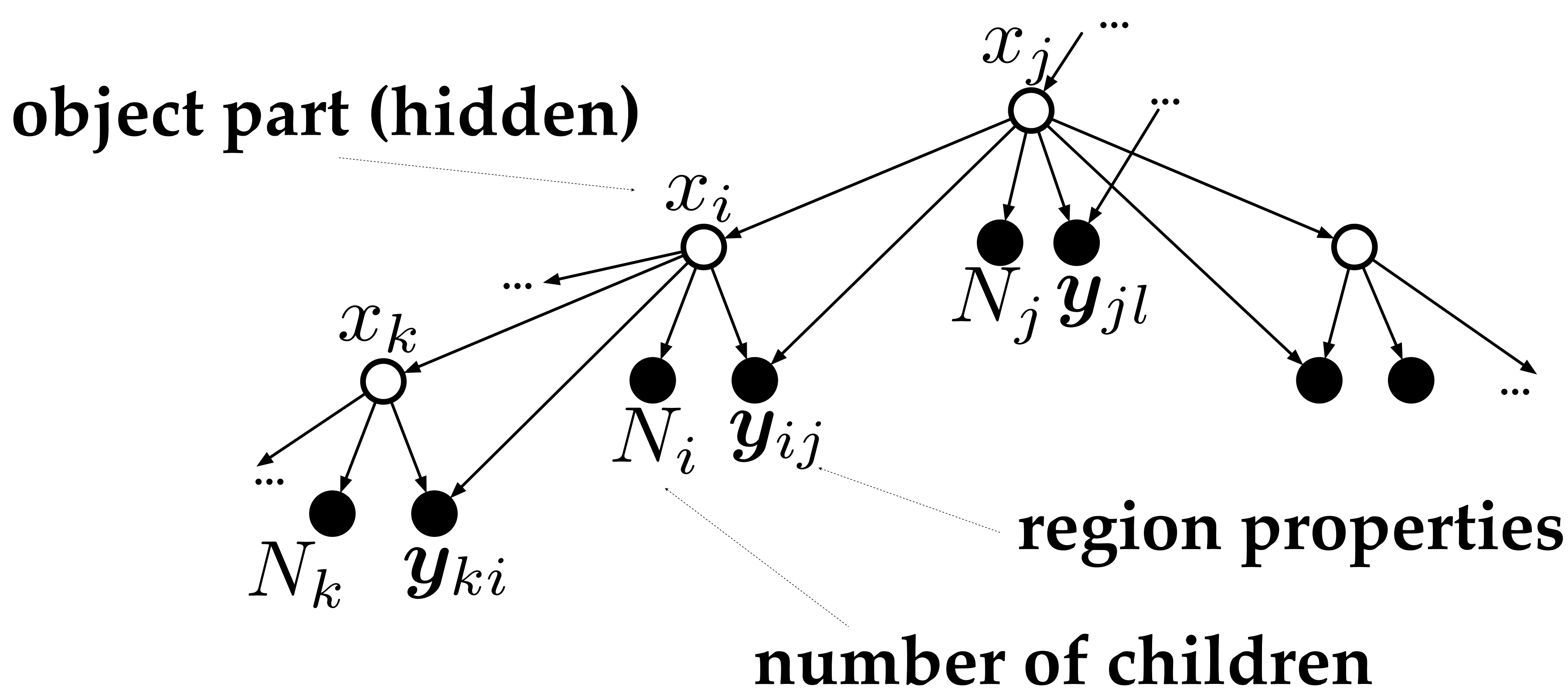
model structure ?

model parameters ?

Category Model = Bayesian Net



Category Model = Bayesian Net



$$P(X, Y, N|\mathcal{T}, \Omega) = \prod_{j \in \mathcal{T}} P(N_j|x_j) \prod_{i=1}^{N_j} P(x_i|x_j) P(y_{ij}|x_i x_j)$$

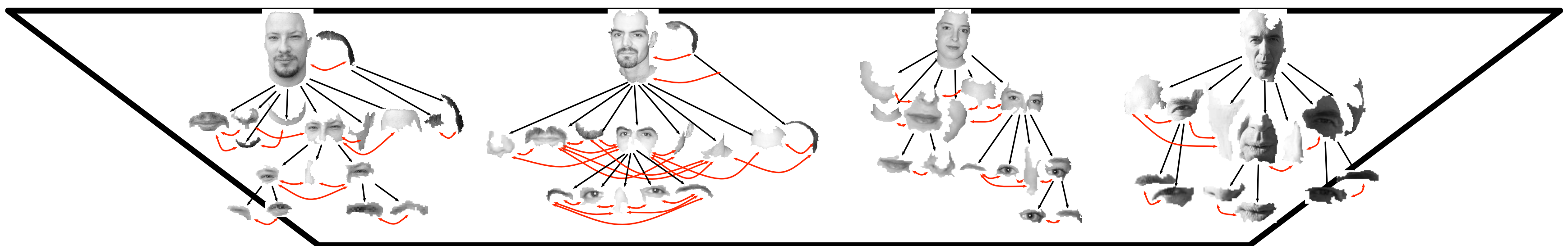
structure + parameters

Markovian chain

Exponential

Gaussian

Learning a Model = Structural EM



Given $\mathcal{T}$

Belief propagation $\Rightarrow \Omega$

Given Ω

Graph matching $\Rightarrow \mathcal{T}$

Category model

$\mathcal{T}$ Ω

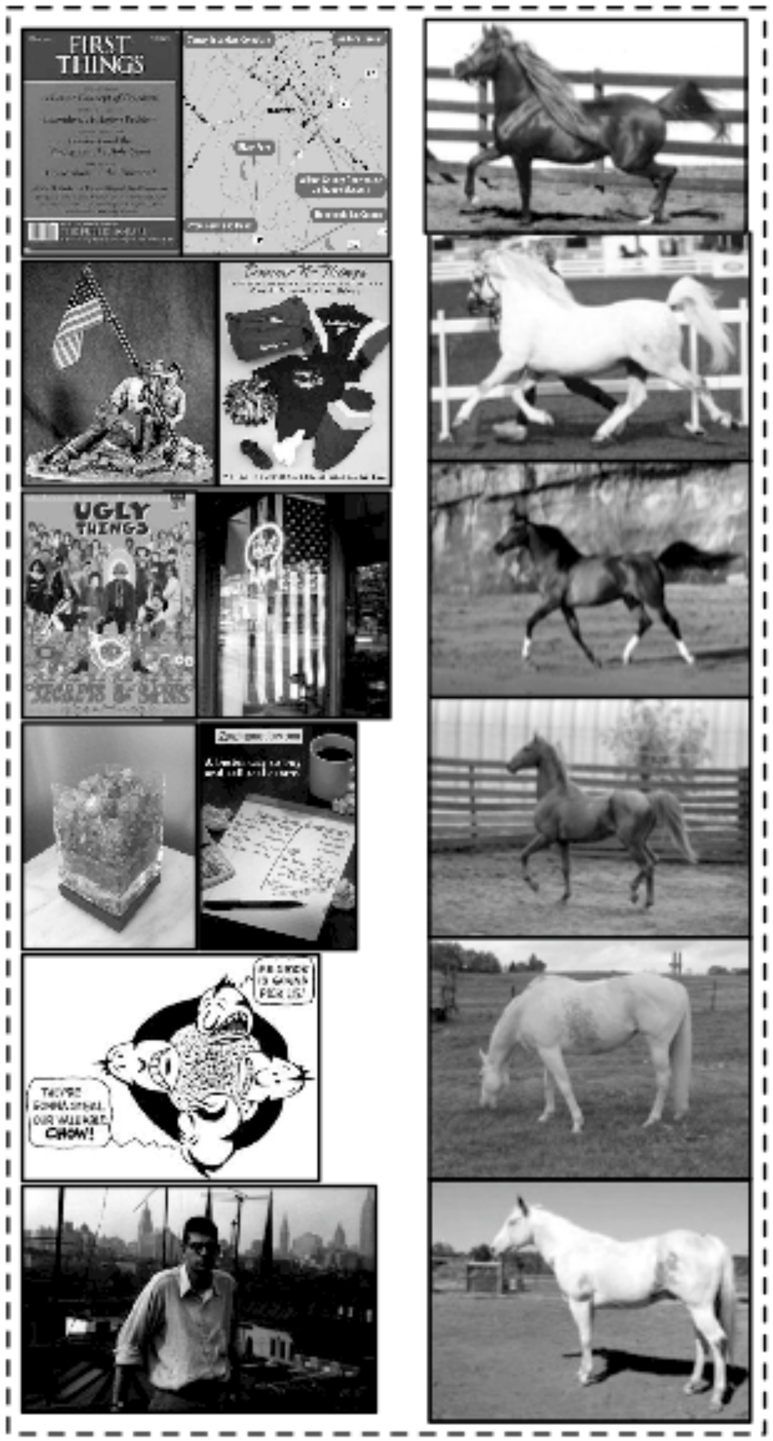
$\mathcal{T}$ **model structure**

Ω **model parameters**

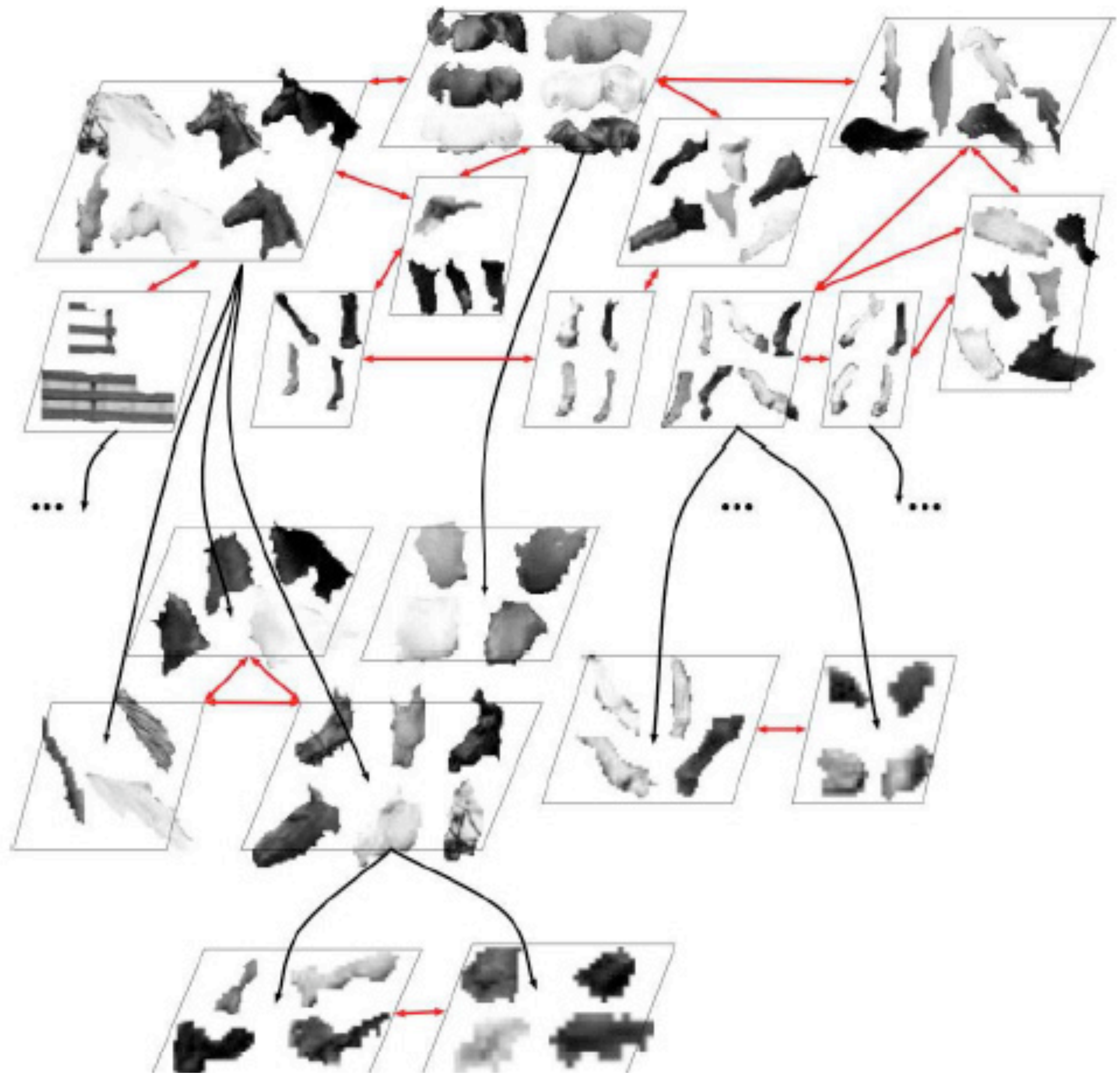
Todorovic&Ahuja ICCV07

Results: Weizmann Horses

training images

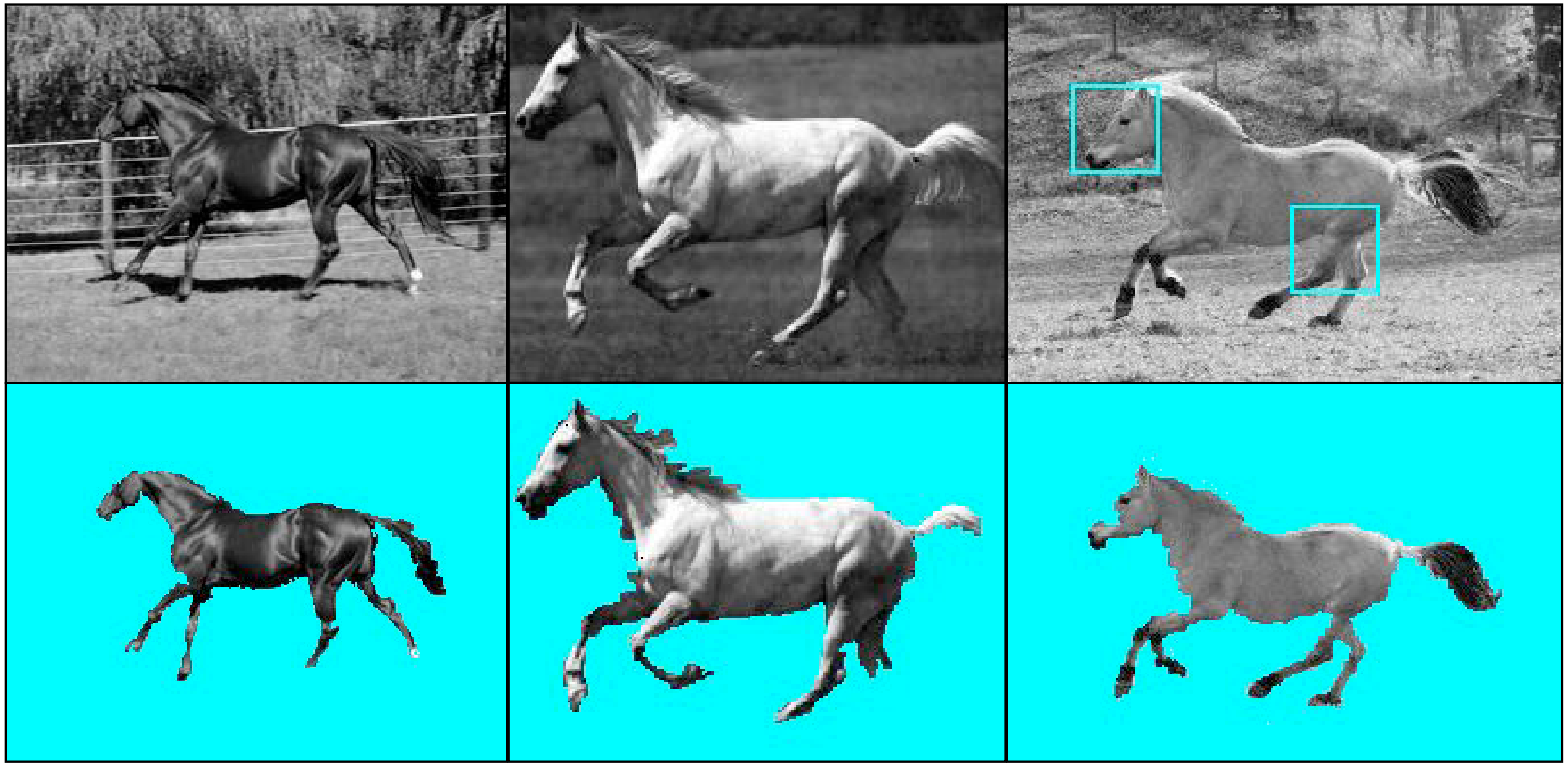


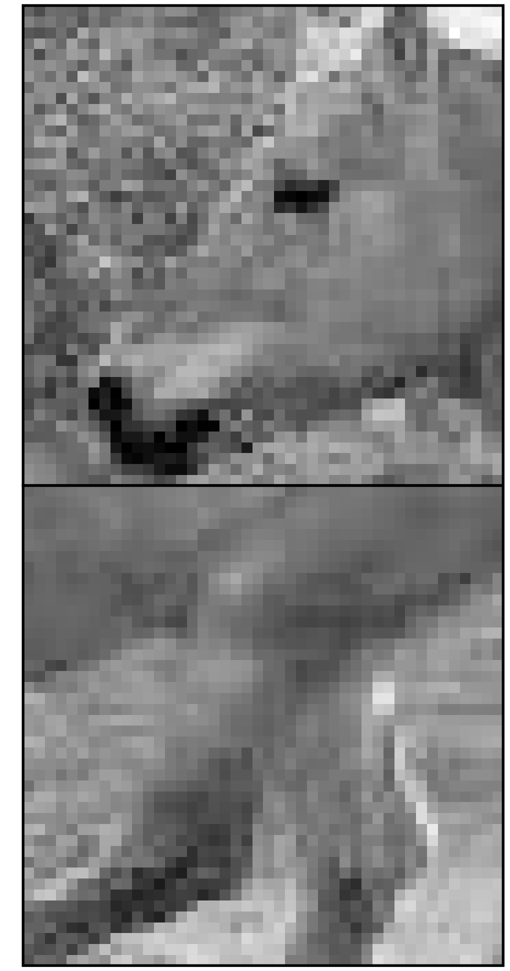
category model



Todorovic&Ahuja PAMI08

Results: Weizmann Horses





- Object segmentation is good on contours that are:
  - Jagged
  - Blurred
  - Form complex patterns
- Low-contrast regions merge with background

Todorovic&Ahuja PAMI09

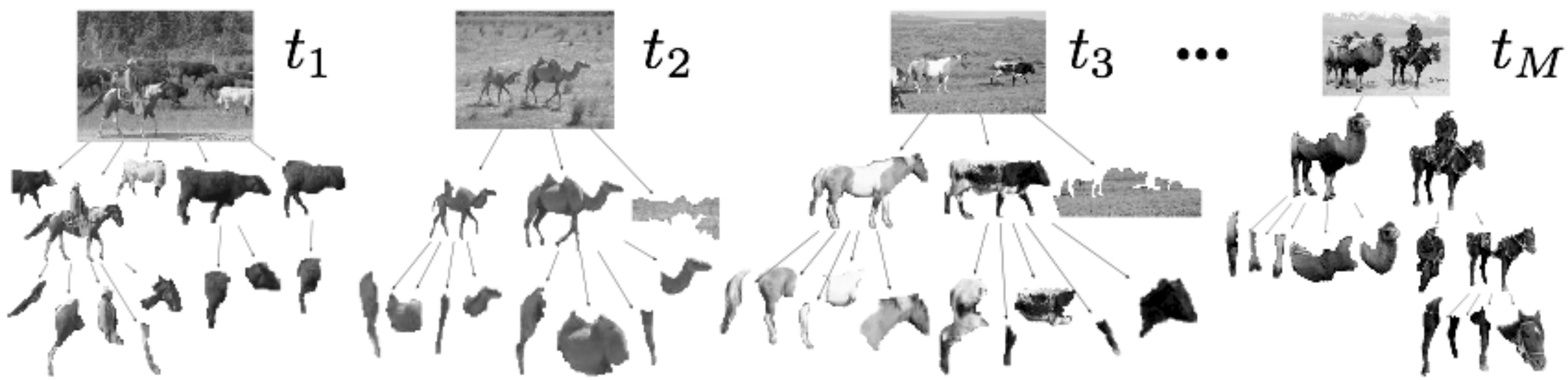
UIUC Hoofed Animals Dataset

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

training images

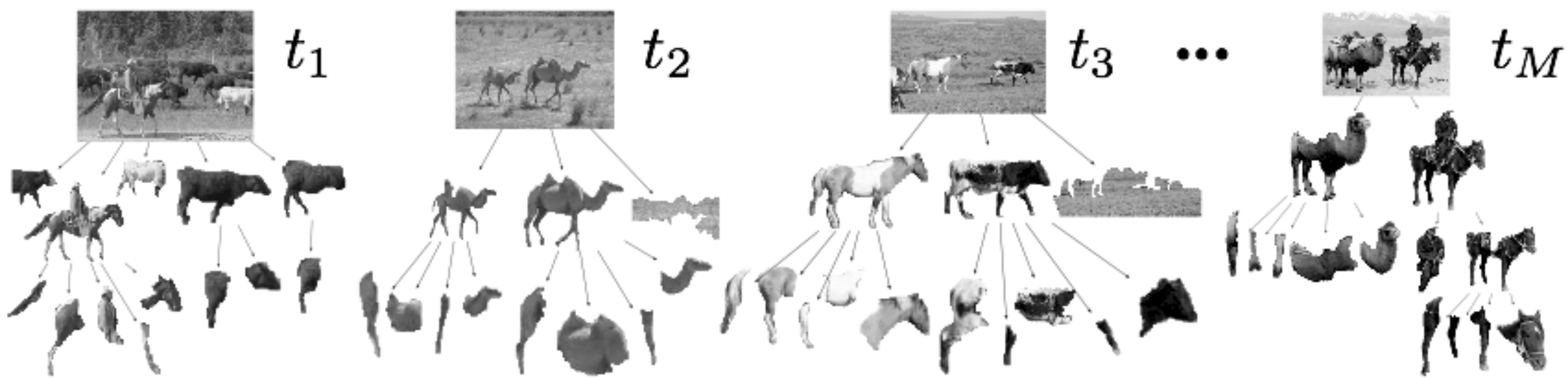
Multi-Object Recognition

1. TREE MATCHING



Multi-Object Recognition

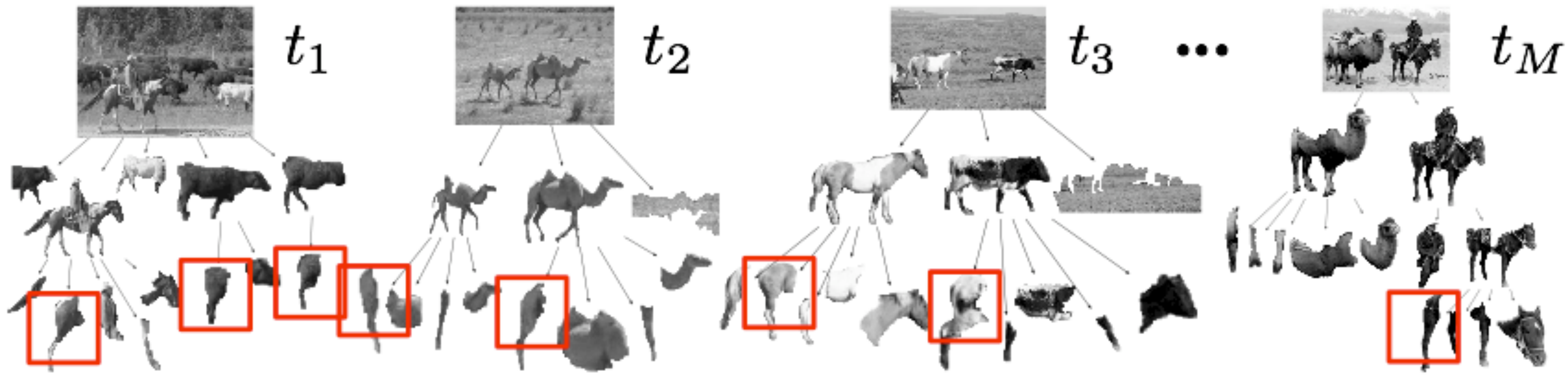
1. TREE MATCHING



2. CLUSTERING

Multi-Object Recognition

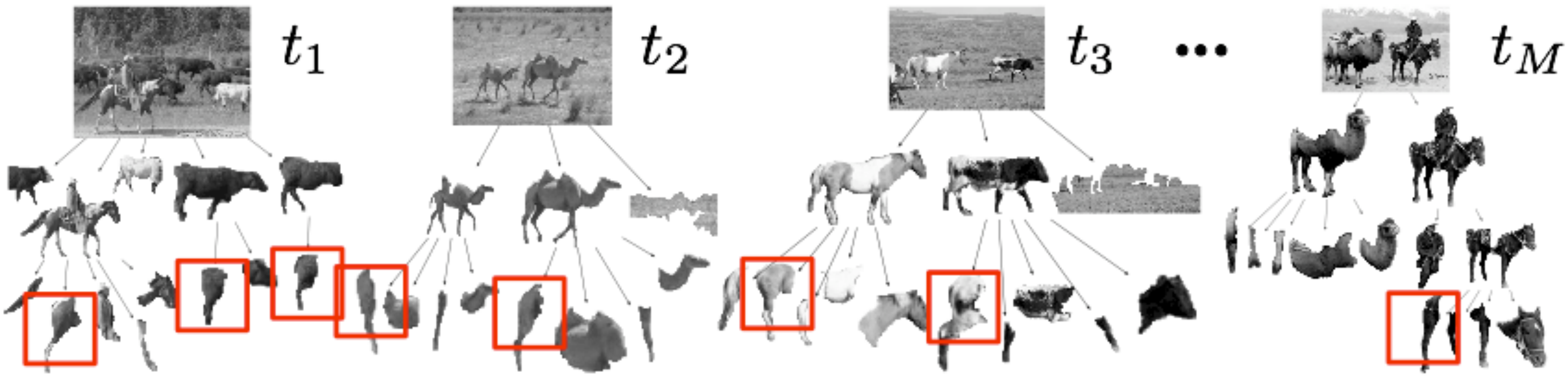
1. TREE MATCHING



2. CLUSTERING

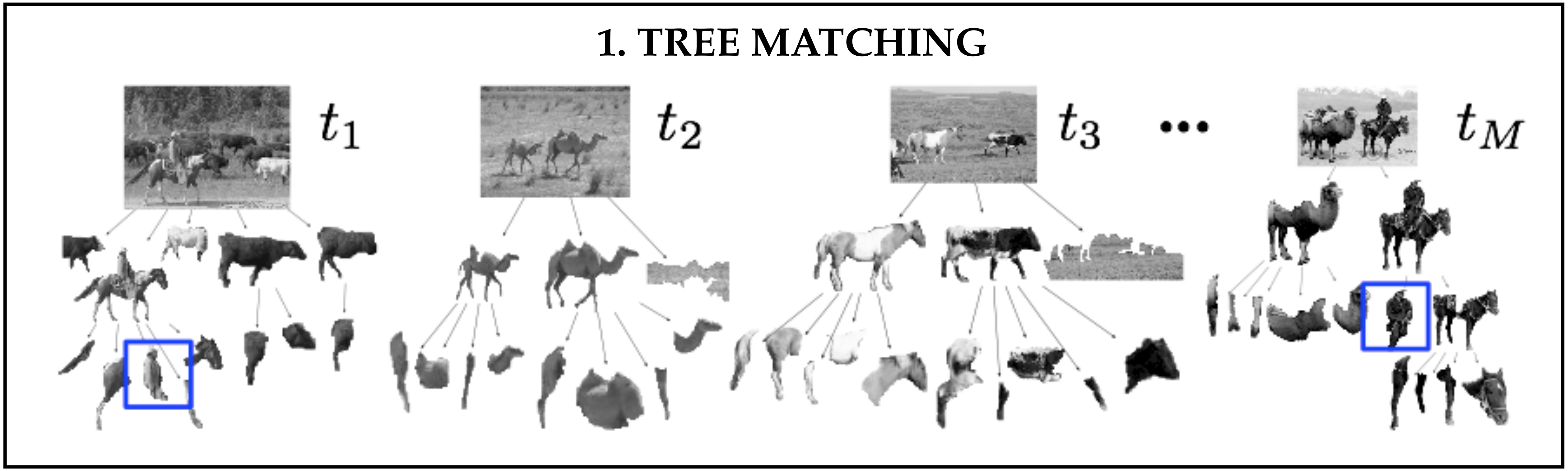
Multi-Object Recognition

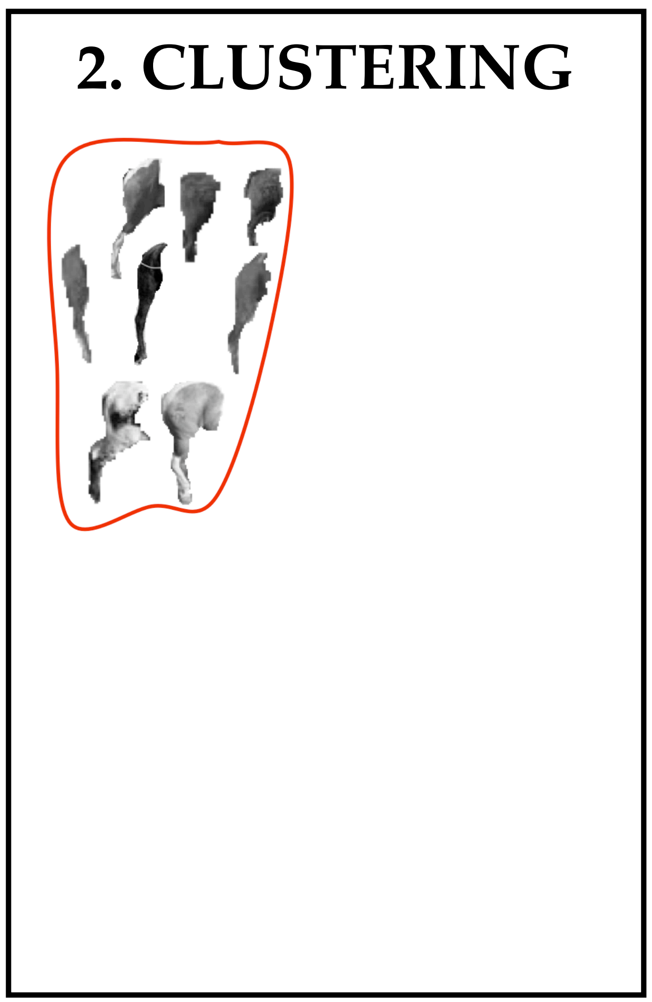
1. TREE MATCHING



2. CLUSTERING

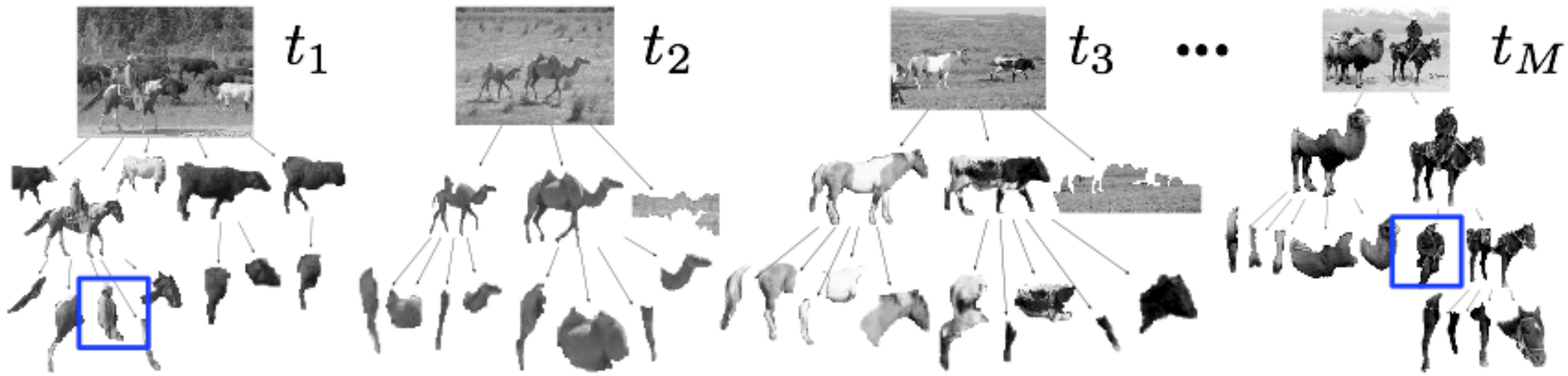
Multi-Object Recognition



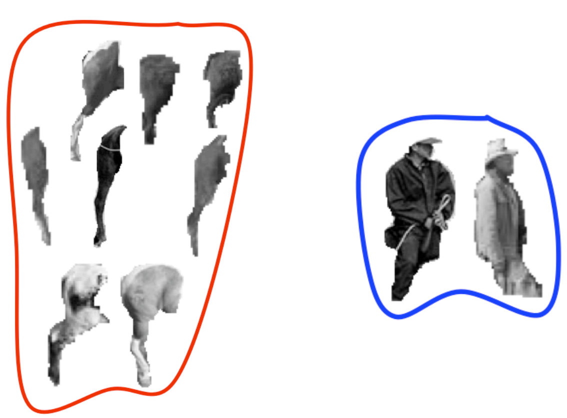


Multi-Object Recognition

1. TREE MATCHING

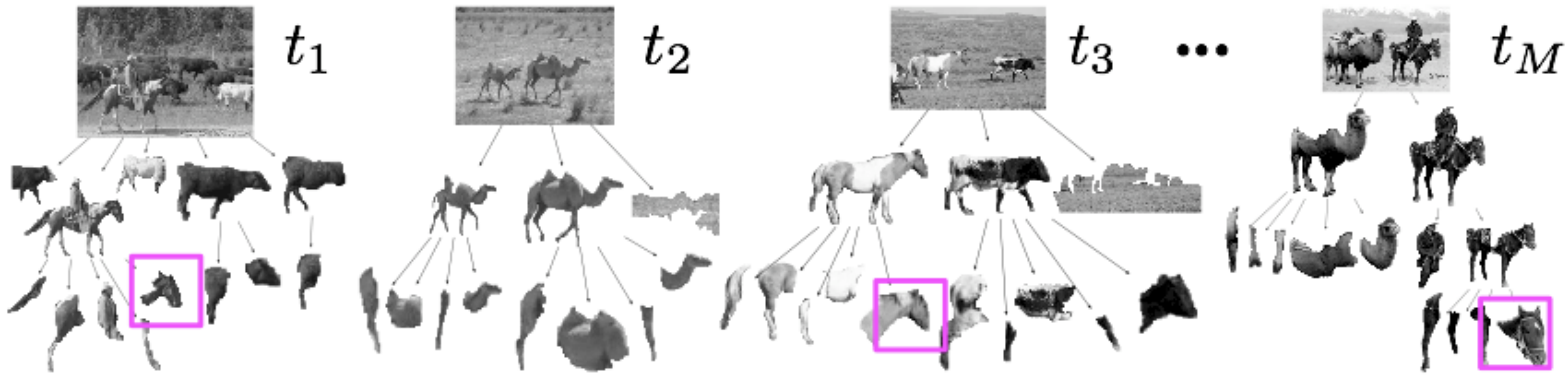


2. CLUSTERING

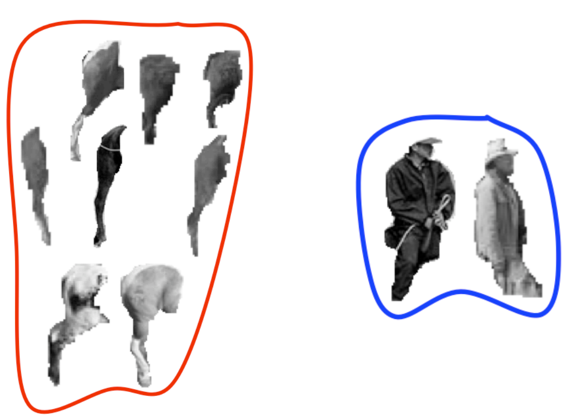


Multi-Object Recognition

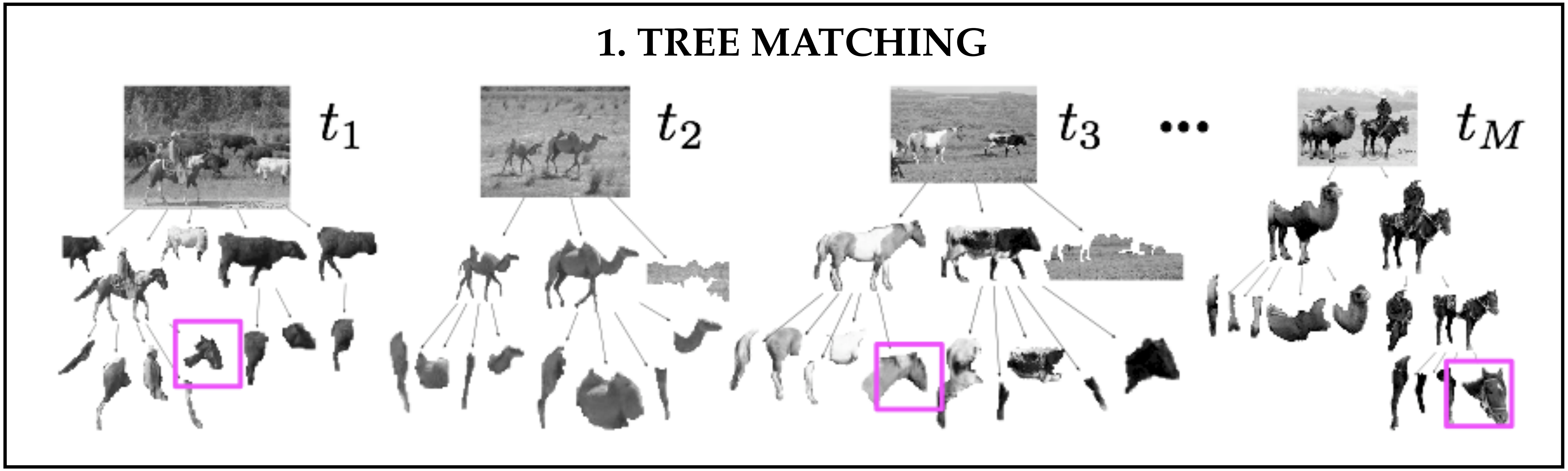
1. TREE MATCHING

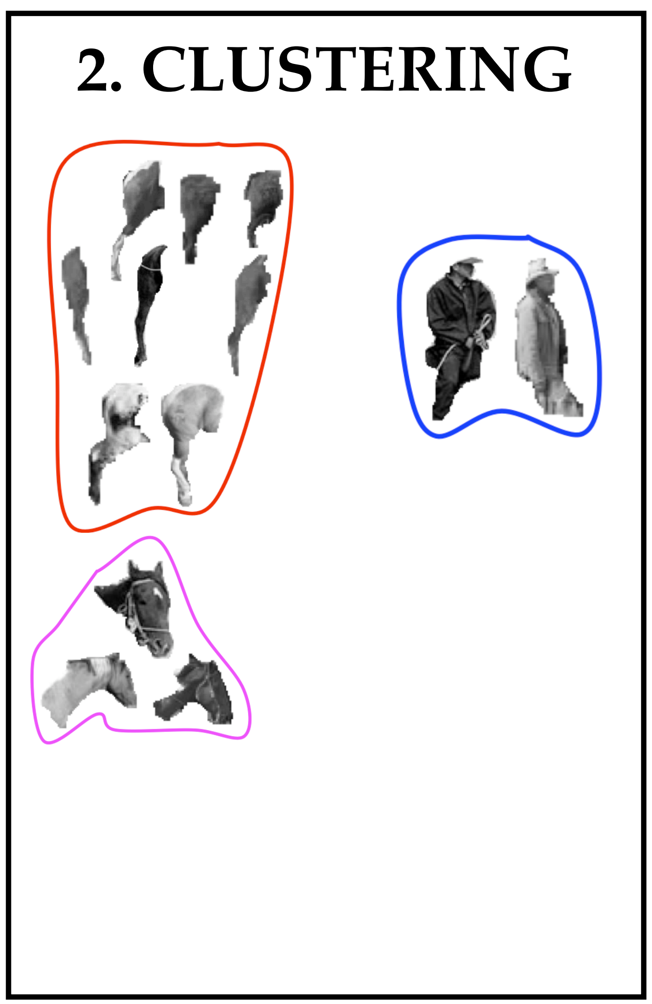


2. CLUSTERING



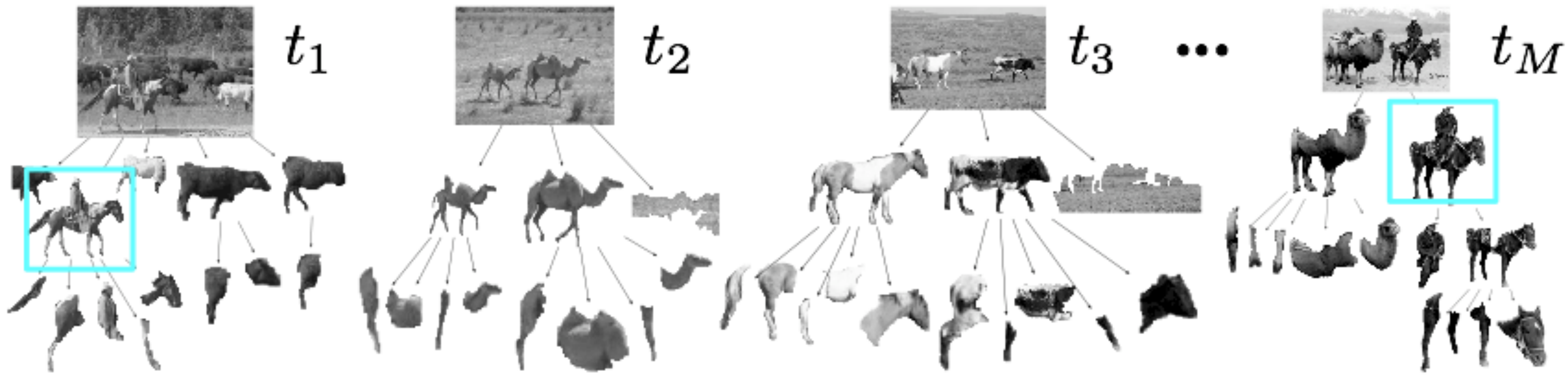
Multi-Object Recognition



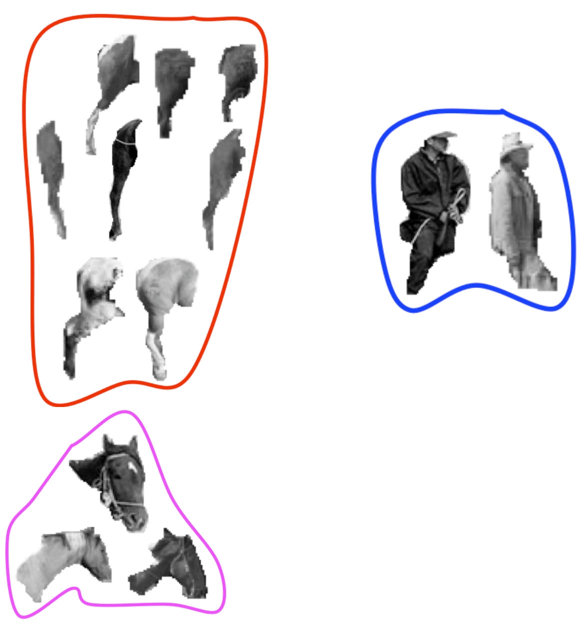


Multi-Object Recognition

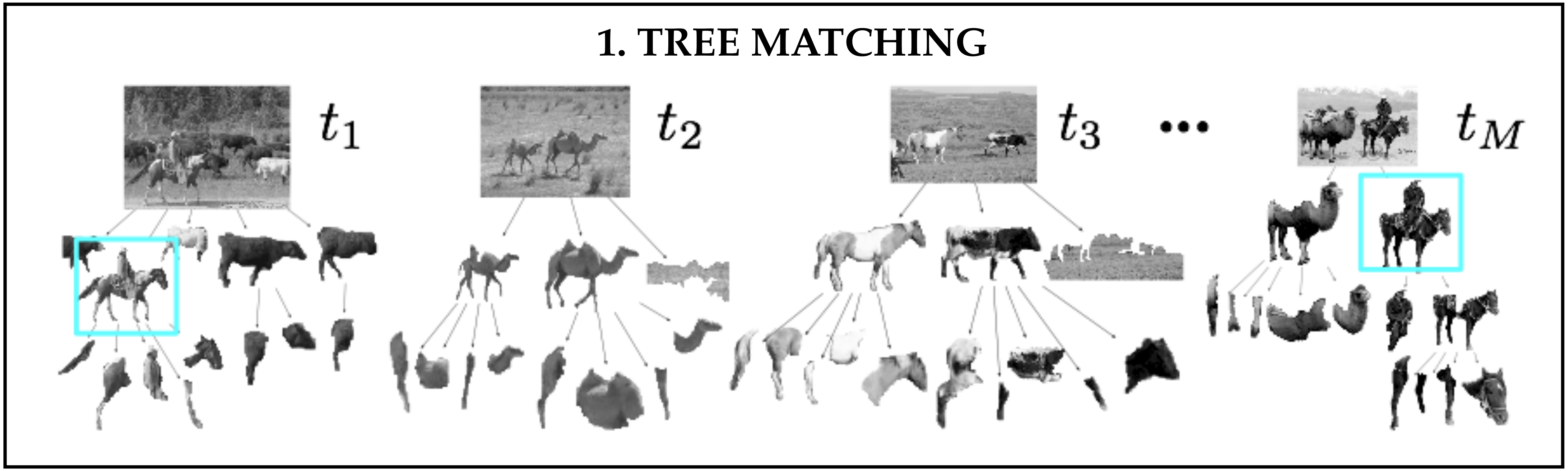
1. TREE MATCHING

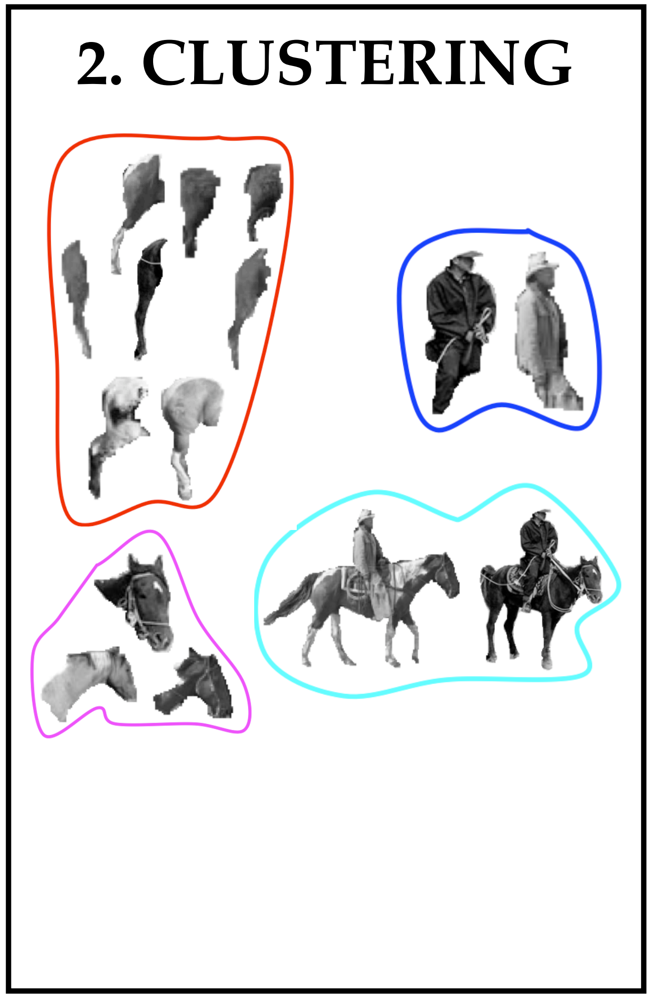


2. CLUSTERING

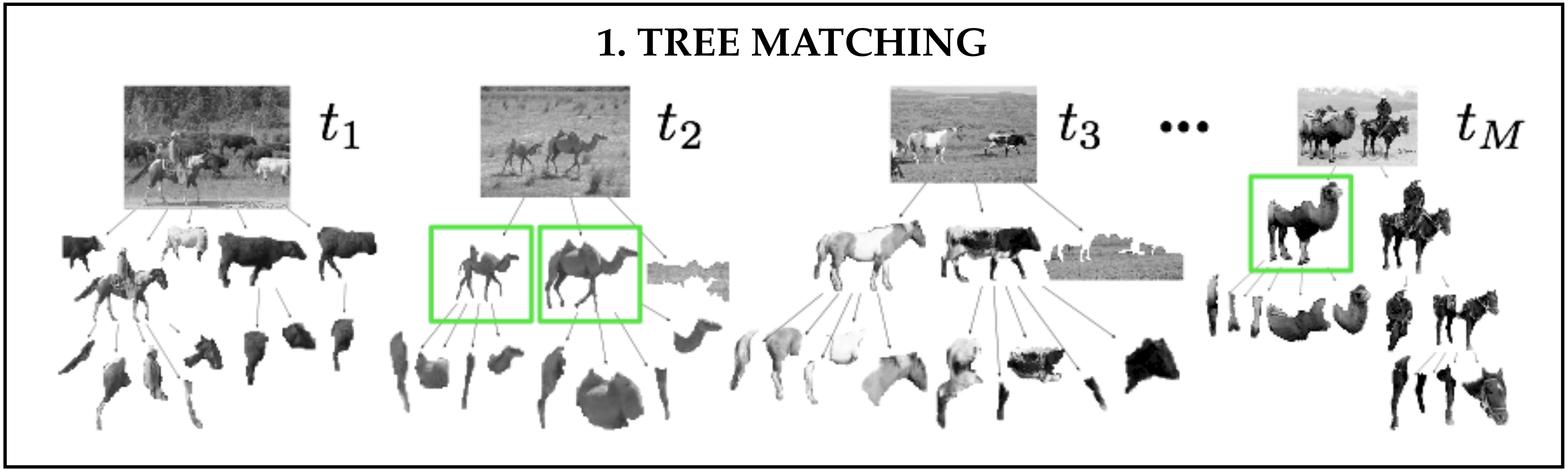


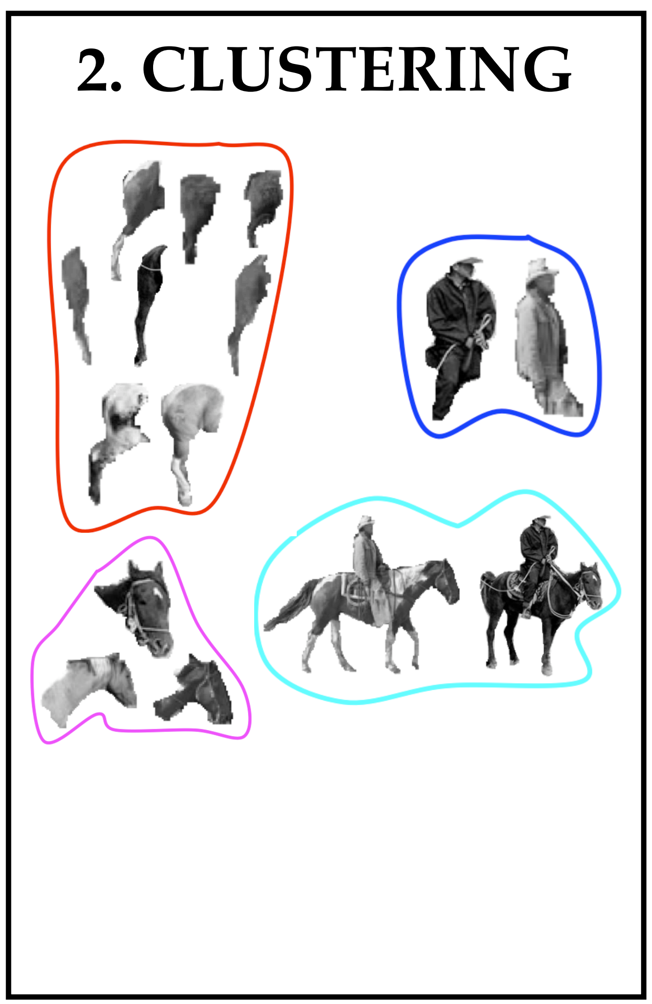
Multi-Object Recognition



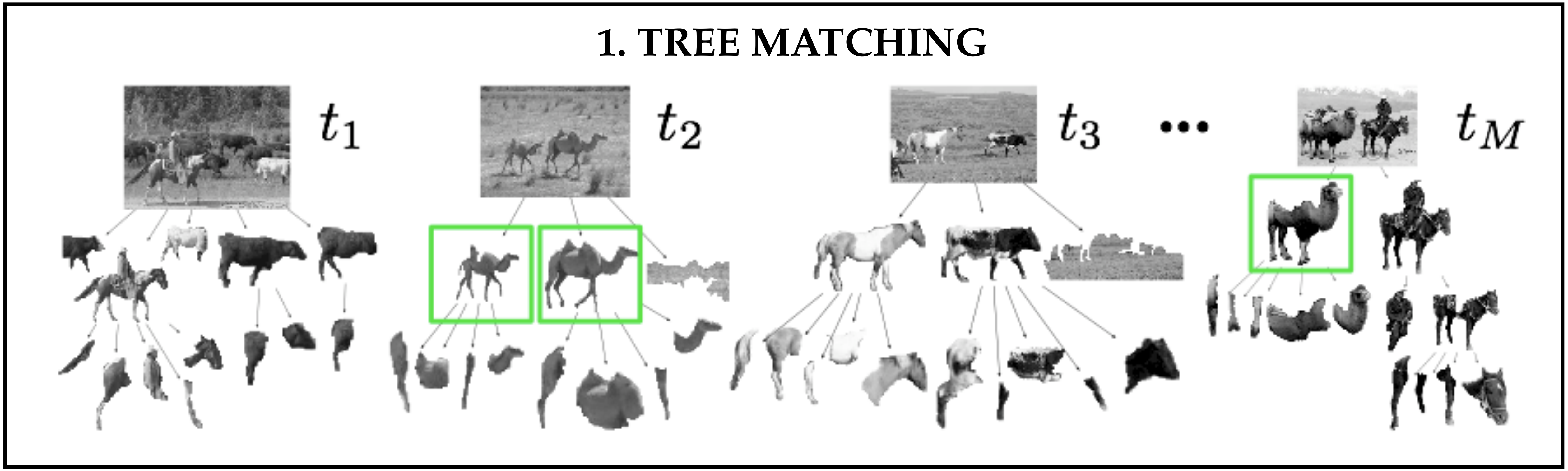


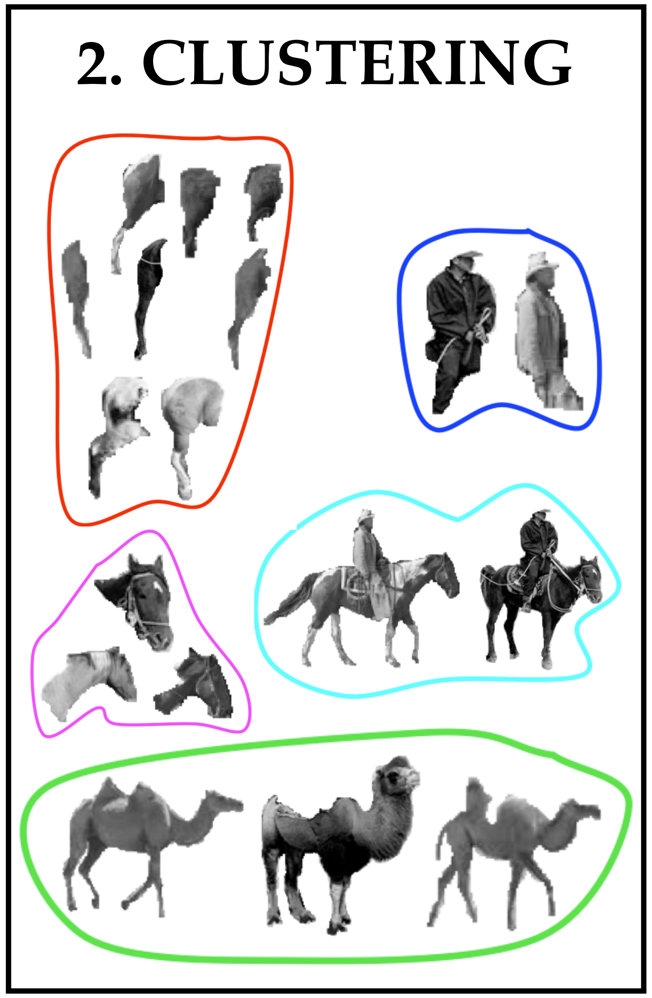
Multi-Object Recognition



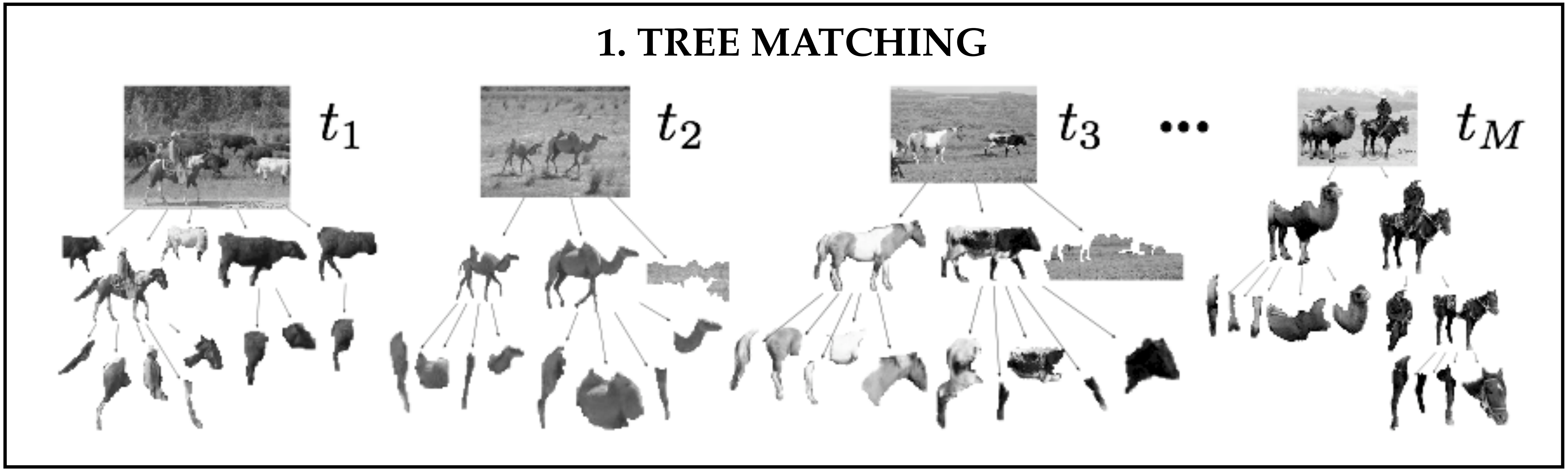


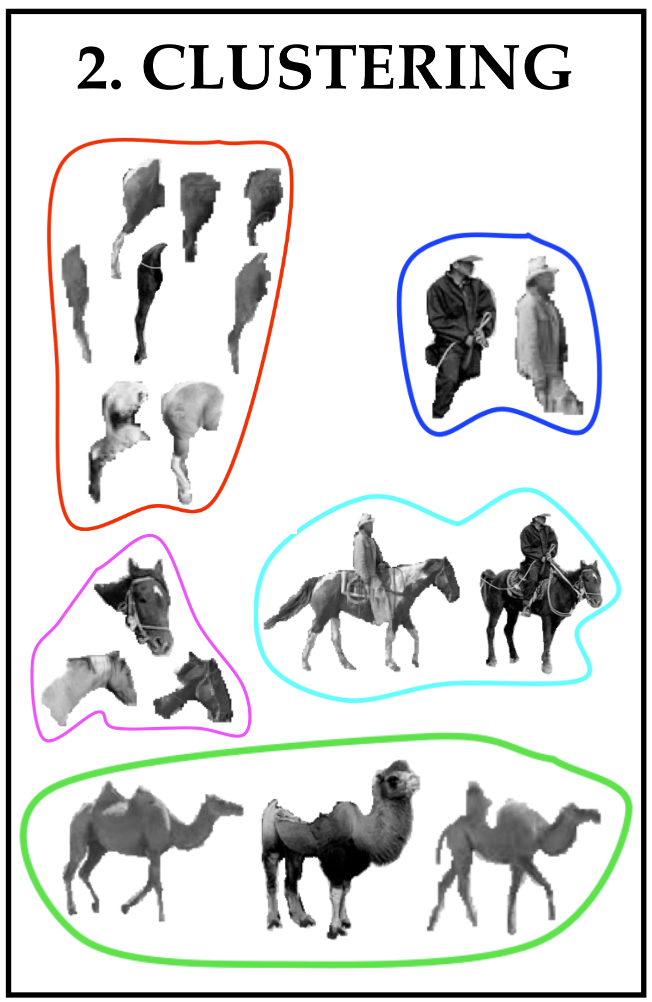
Multi-Object Recognition



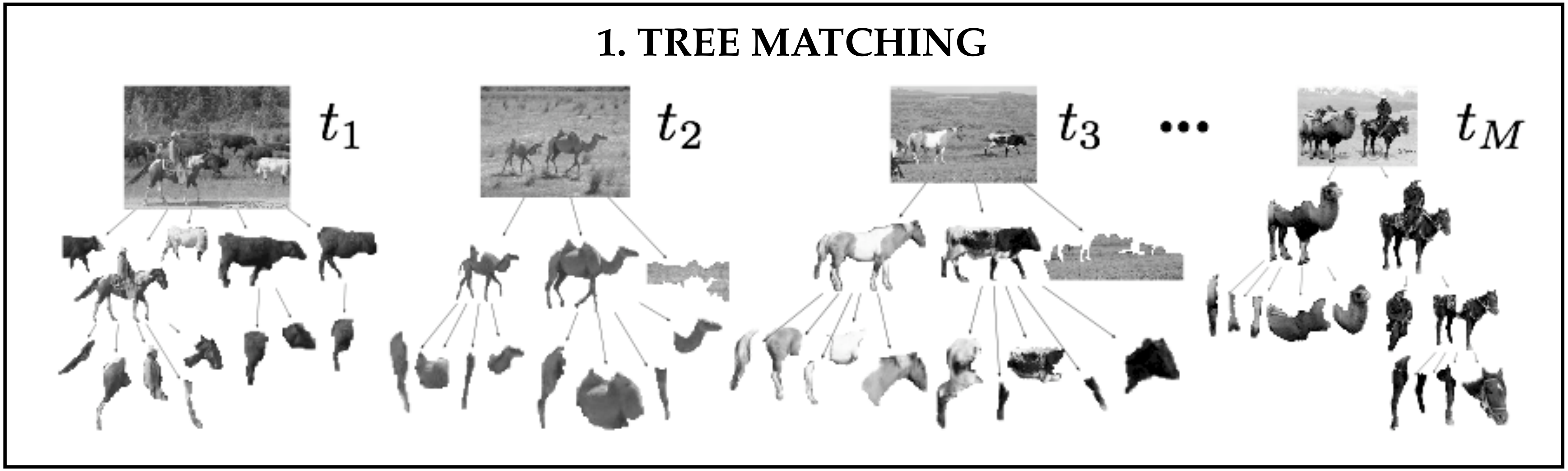


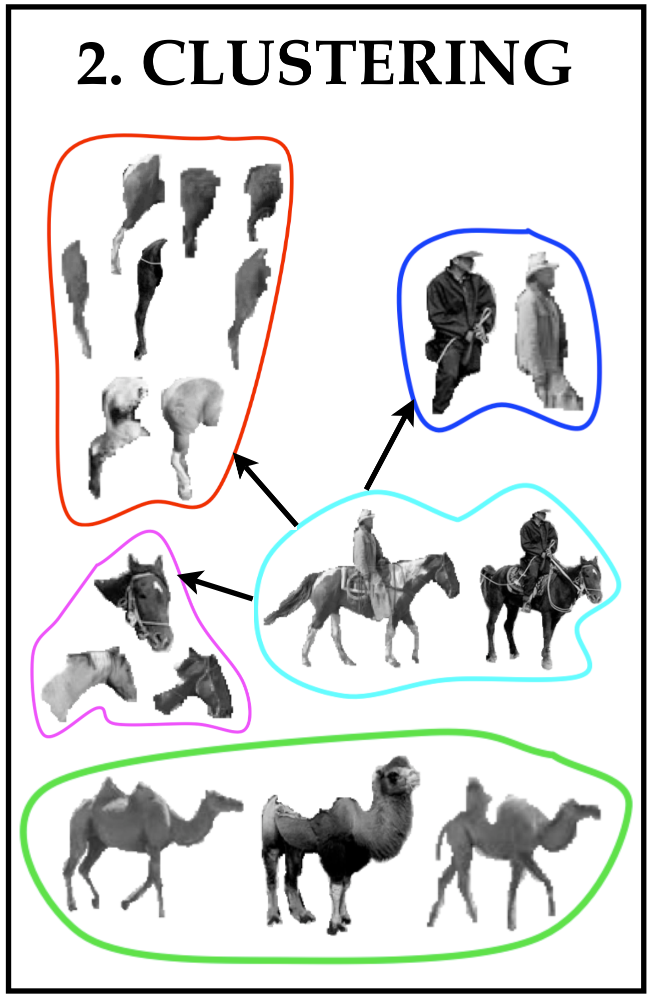
Overview of Multi-Category Recognition



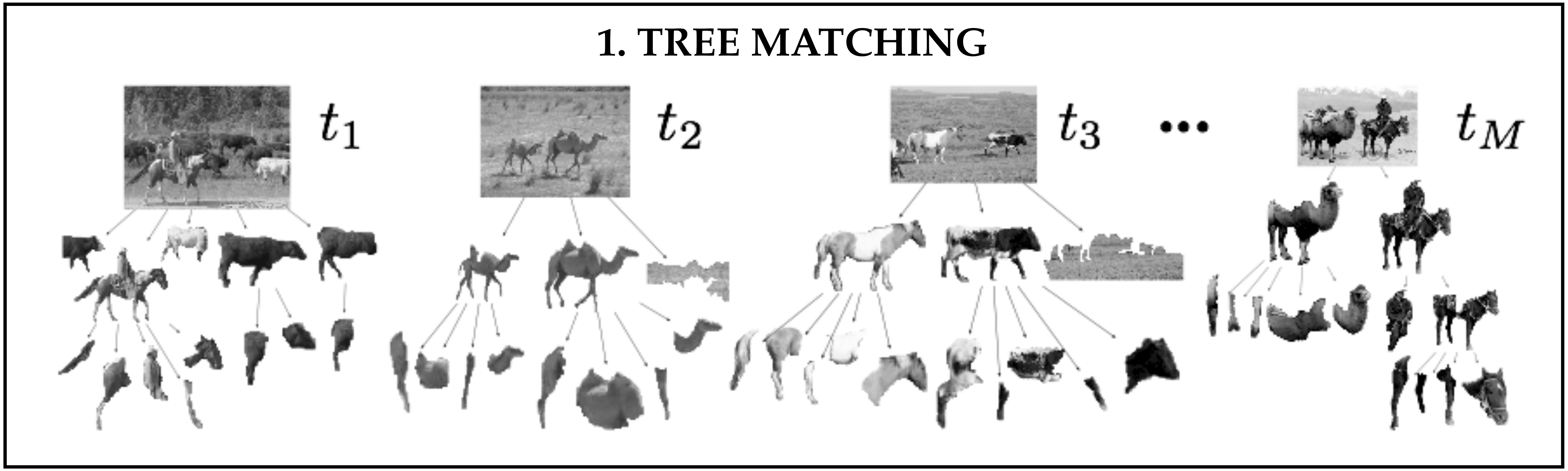


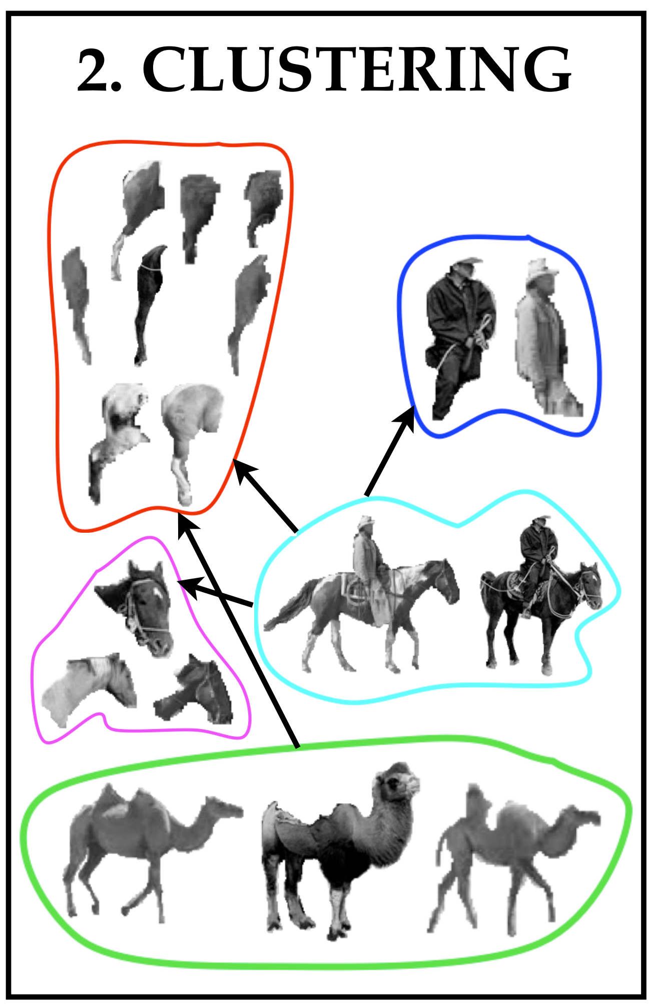
Overview of Multi-Category Recognition





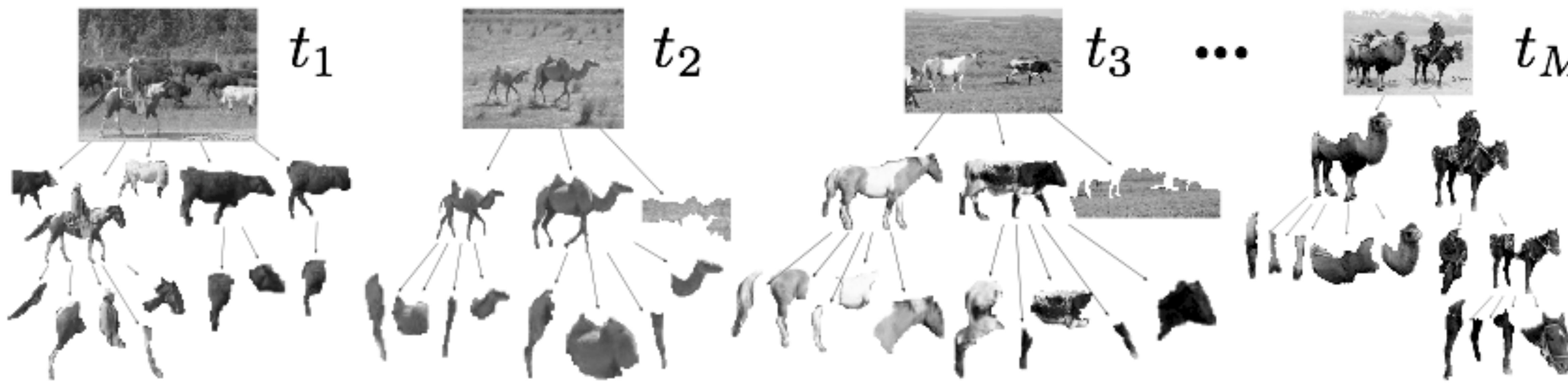
Overview of Multi-Category Recognition



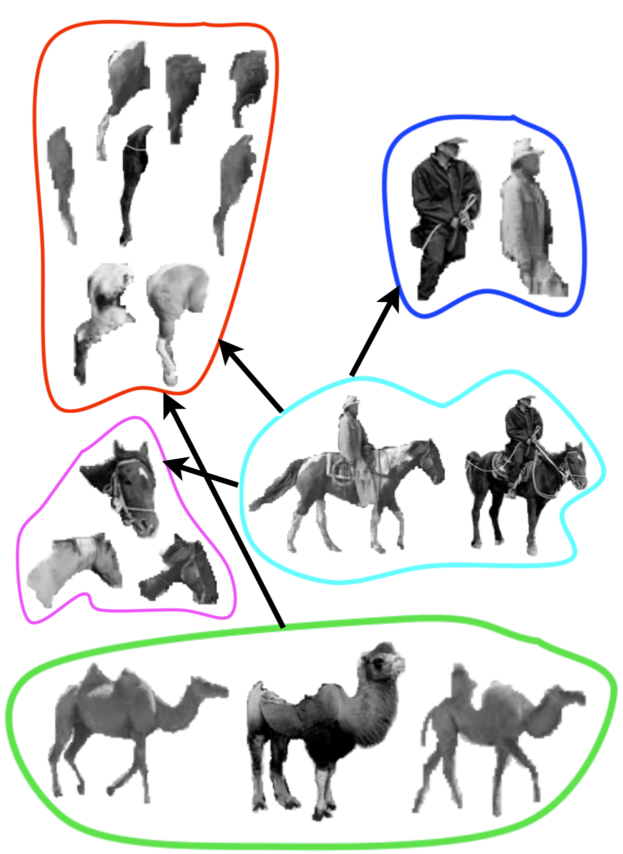


Overview of Multi-Category Recognition

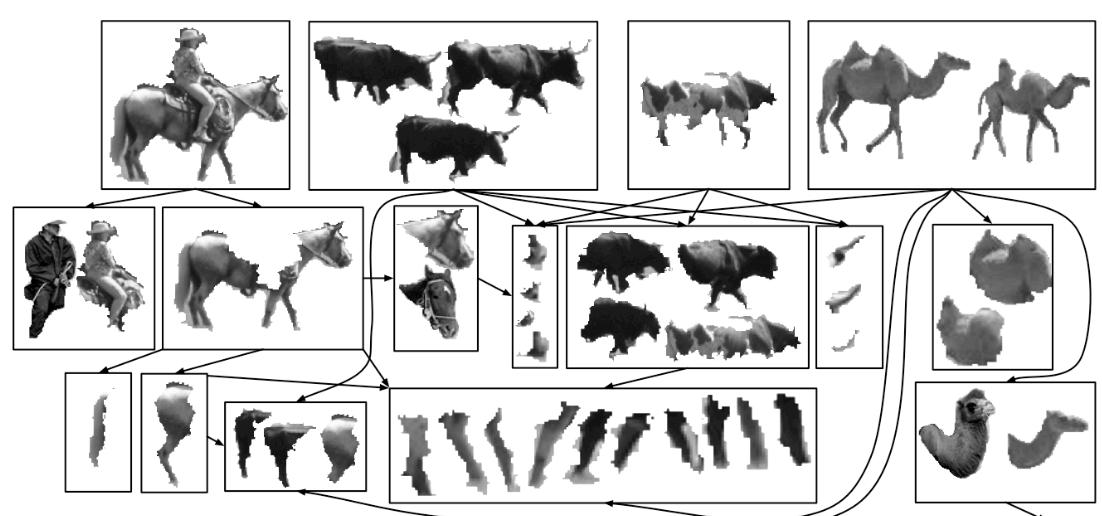
1. TREE MATCHING



2. CLUSTERING

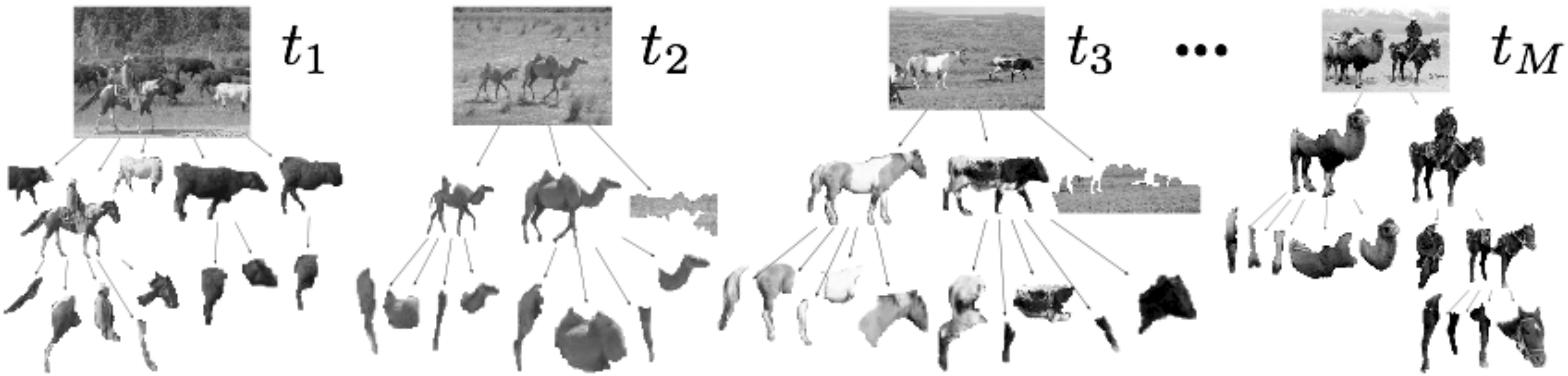


3. TAXONOMY OF ALL DISCOVERED CATEGORIES WITH DIFFERENT COMPLEXITIES

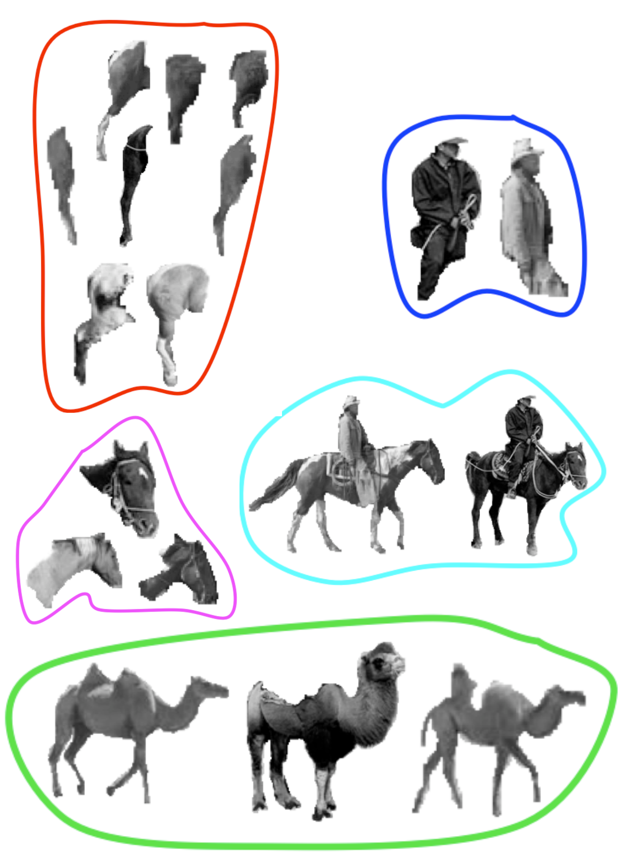


Overview of Multi-Category Recognition

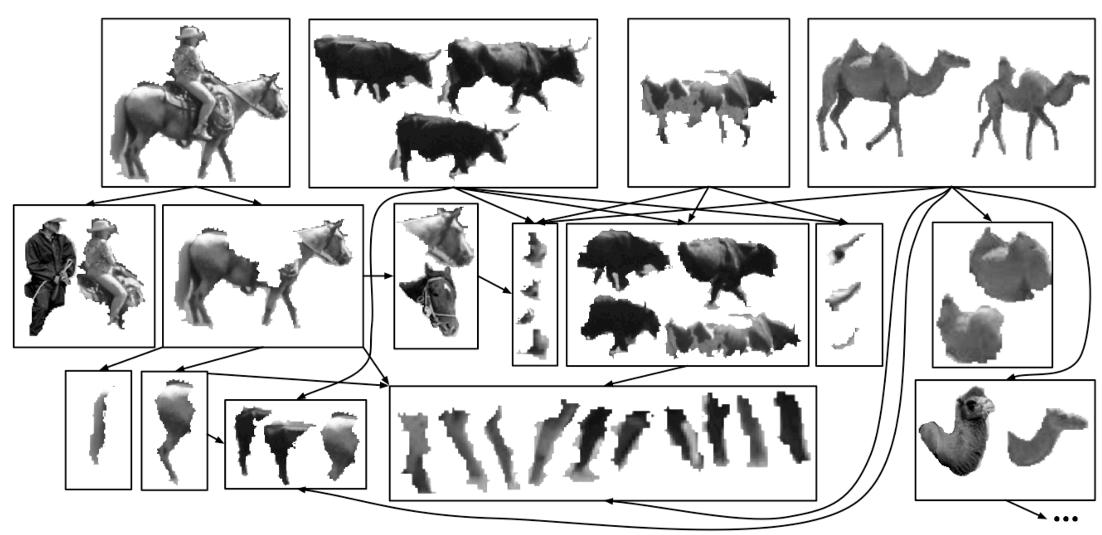
1. TREE MATCHING



2. CLUSTERING

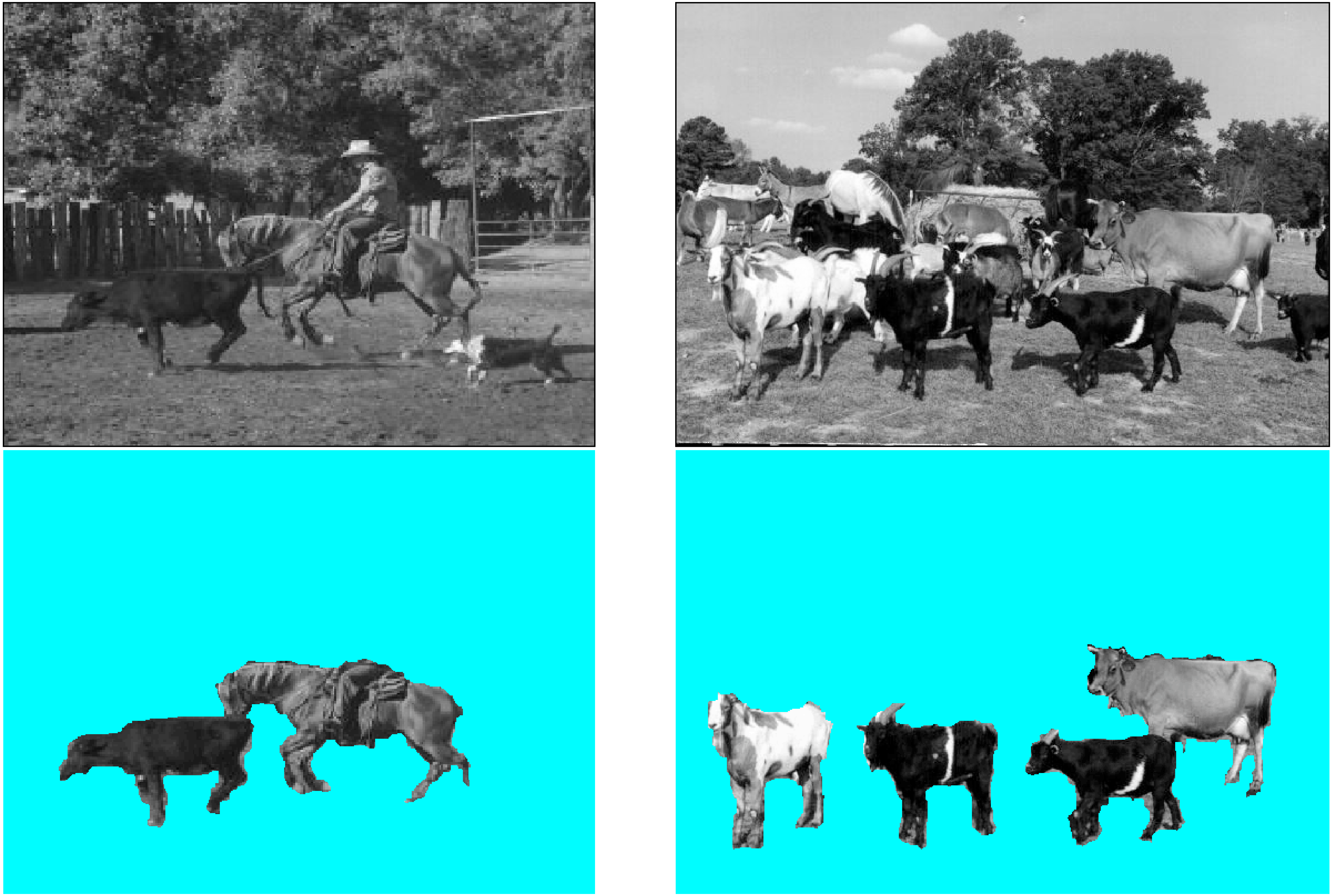


3. TAXONOMY OF ALL DISCOVERED CATEGORIES WITH DIFFERENT COMPLEXITIES

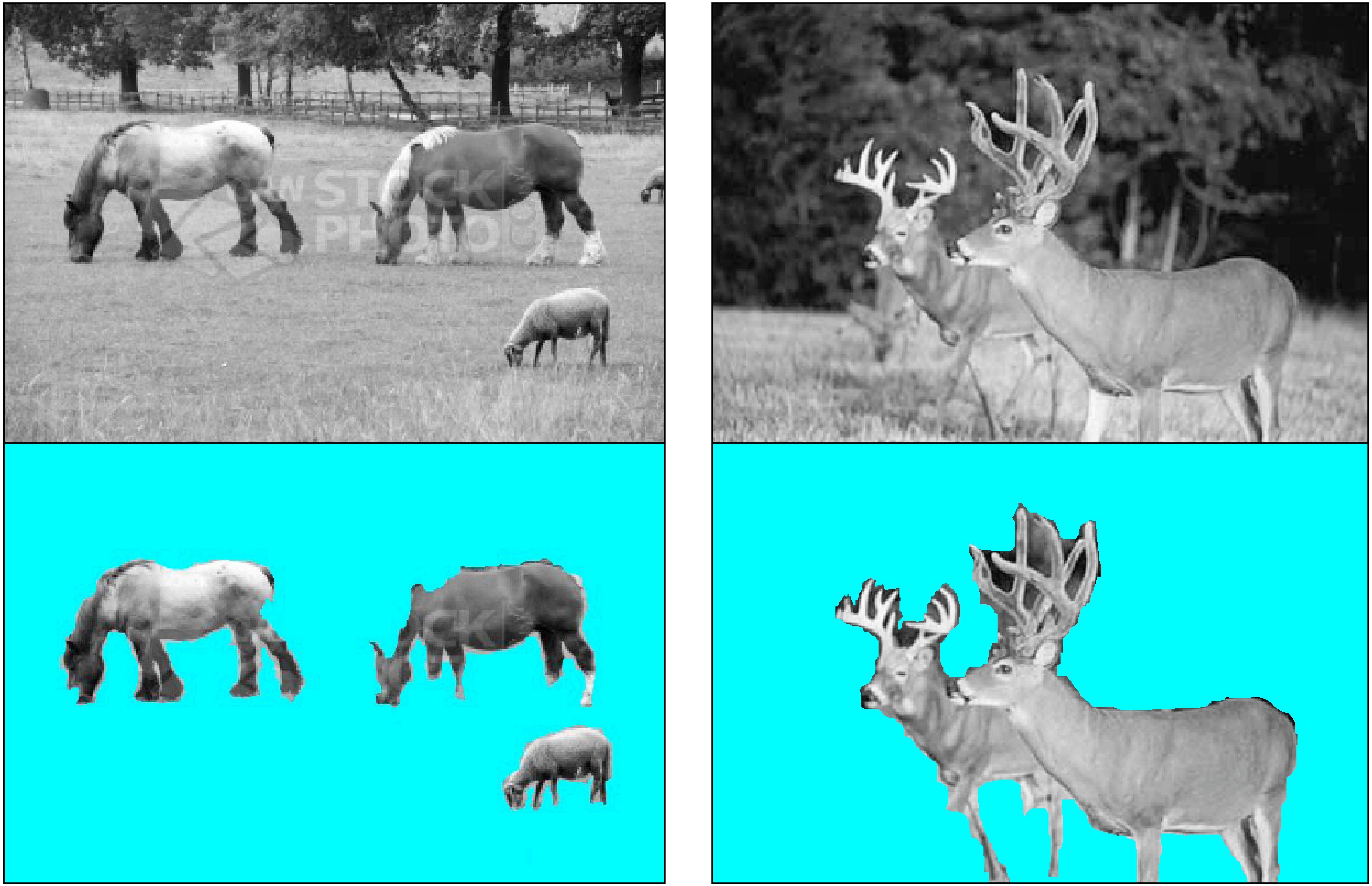


4. RECOGNIZE SEGMENT EXPLAIN

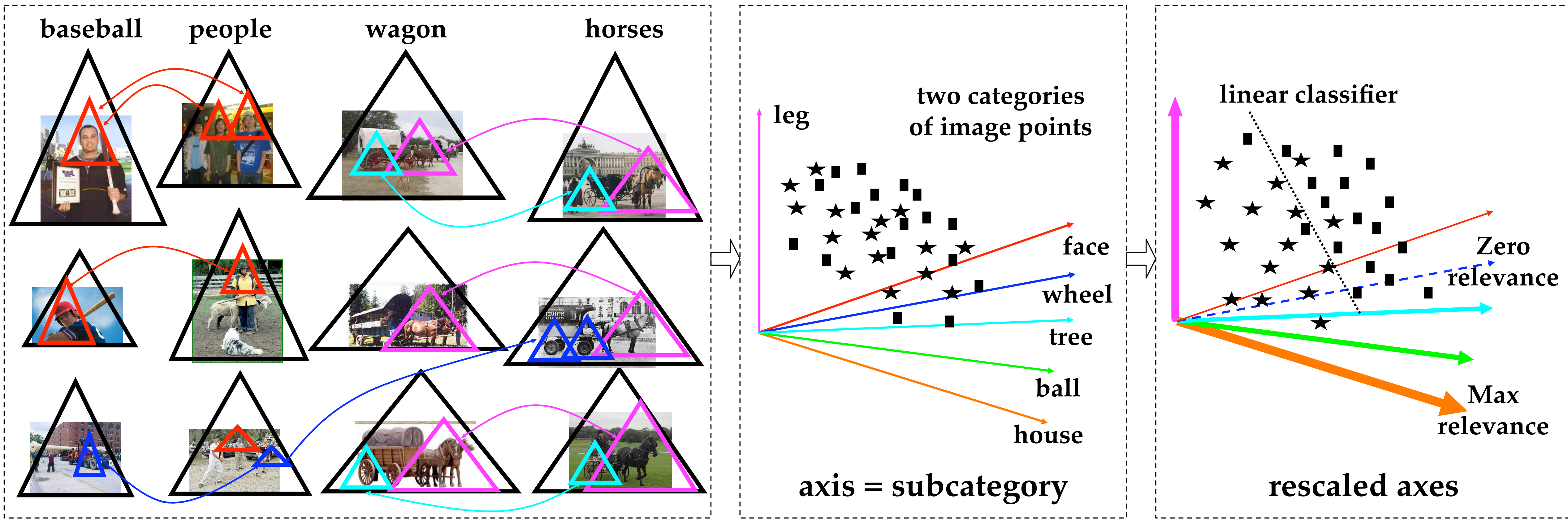
Simultaneous Recognition and Segmentation



Simultaneous Recognition and Segmentation



Discriminative Learning of Object Parts

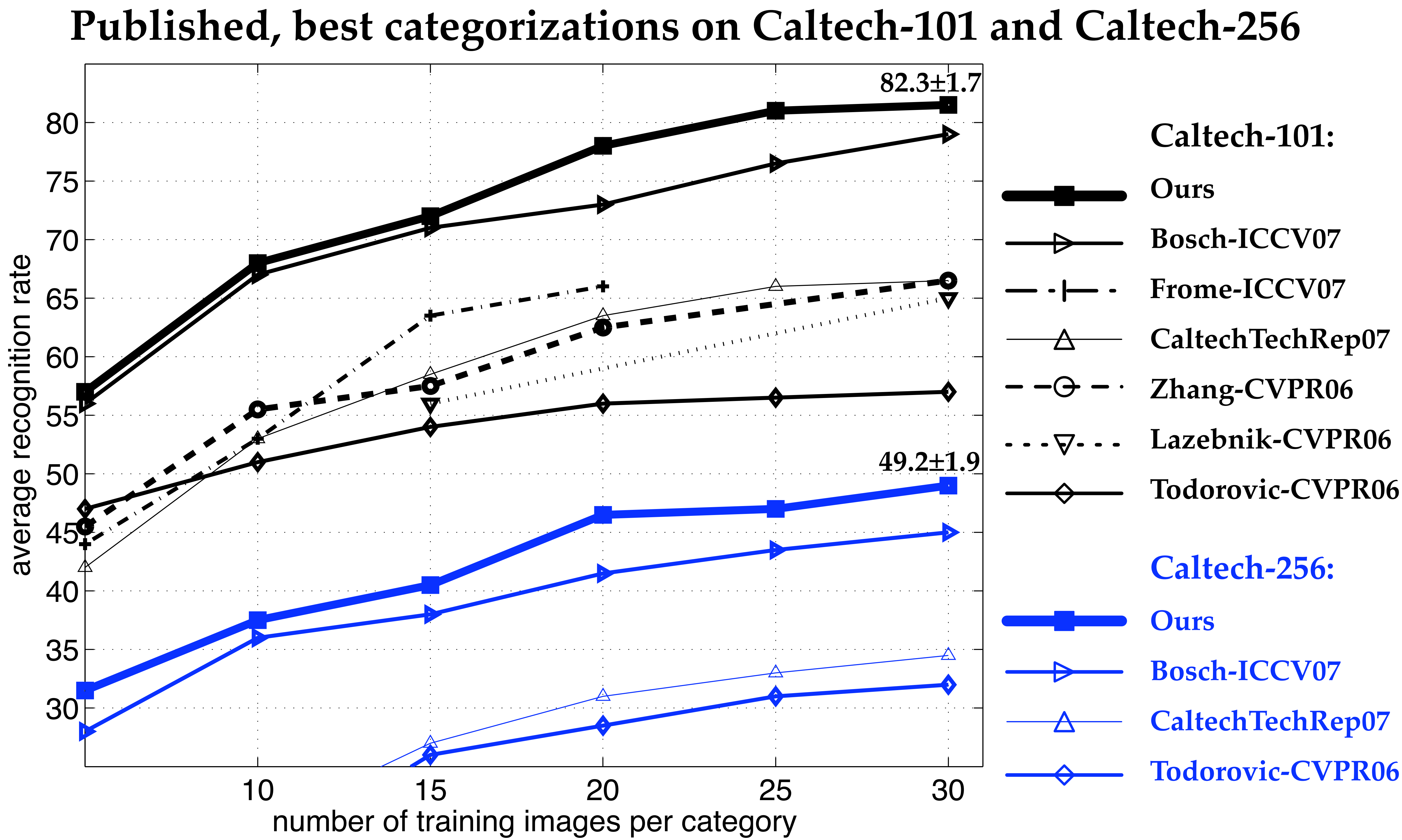


discovery of subcategories in segmentation trees

trees = points in the feature space of subcategories

Todorovic&Ahuja CVPR08

CVPR 2008: Results on Caltech-256



Rest of the Talk

1. **Image representation = Hierarchy of regions**

2. **Region matching under unstable segmentations**

3. **Applications and results**

    a. **Object recognition**

    b. **Painterly rendering**

    c. **Texture segmentation**

Multi-style Painterly Rendering

Collaboration with Prof. Eugene Zhang at Oregon State University

Results: Video Object Segmentation

Brendel&Todorovic ICCV09

Results: Multi-style Painterly Rendering

Collaboration with Prof. Eugene Zhang at Oregon State University

Rest of the Talk

1. Image representation = Hierarchy of regions

2. Region matching under unstable segmentations

3. **Applications and results**

    a. Object recognition

    b. Video object segmentation

    c. Painterly rendering

    d. **Texture segmentation**

What is image texture?

...Repeated occurrence of image texture elements (or texels)...

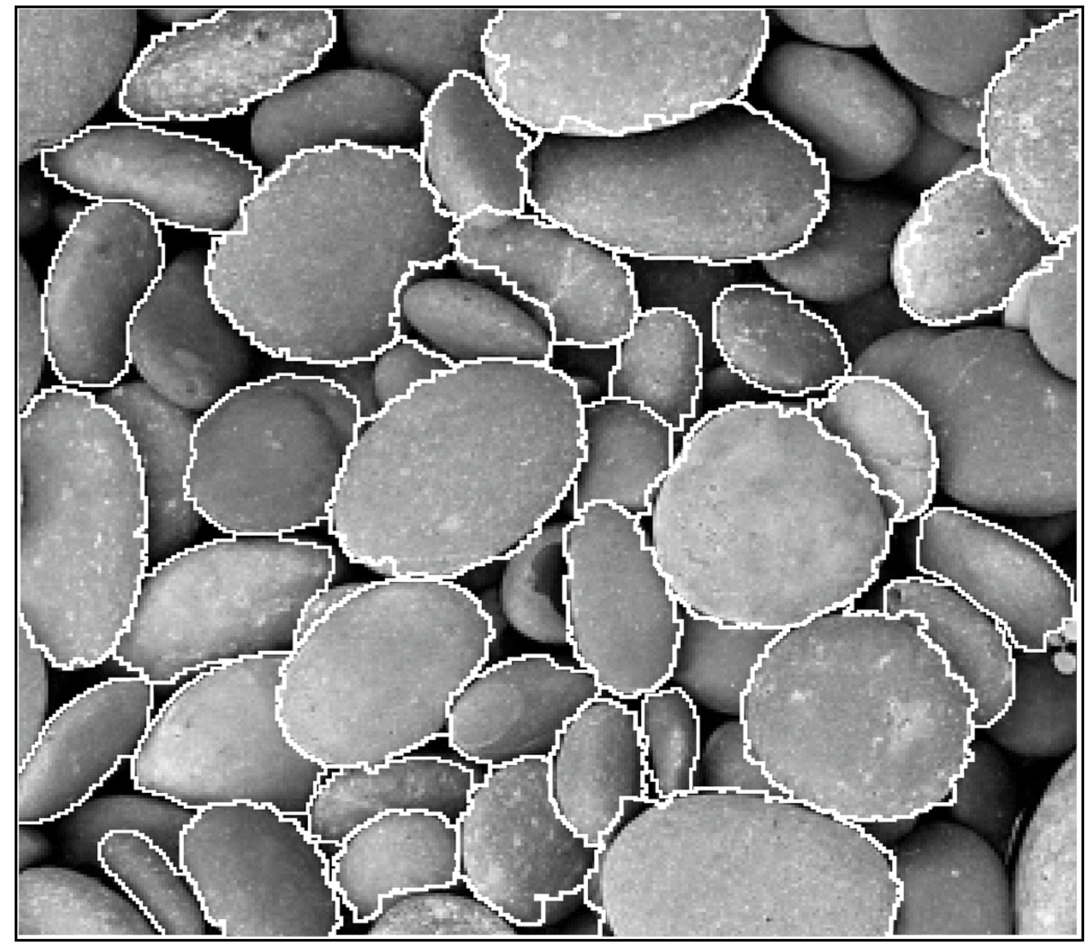
[Beck ‘82]

Texture = Spatial Repetition of Texels

- Image texels = Images of physical texture elements
- Texels are not identical, only statistically similar
- Texel placement is not regular


Results: UIUC Texture Dataset

original image

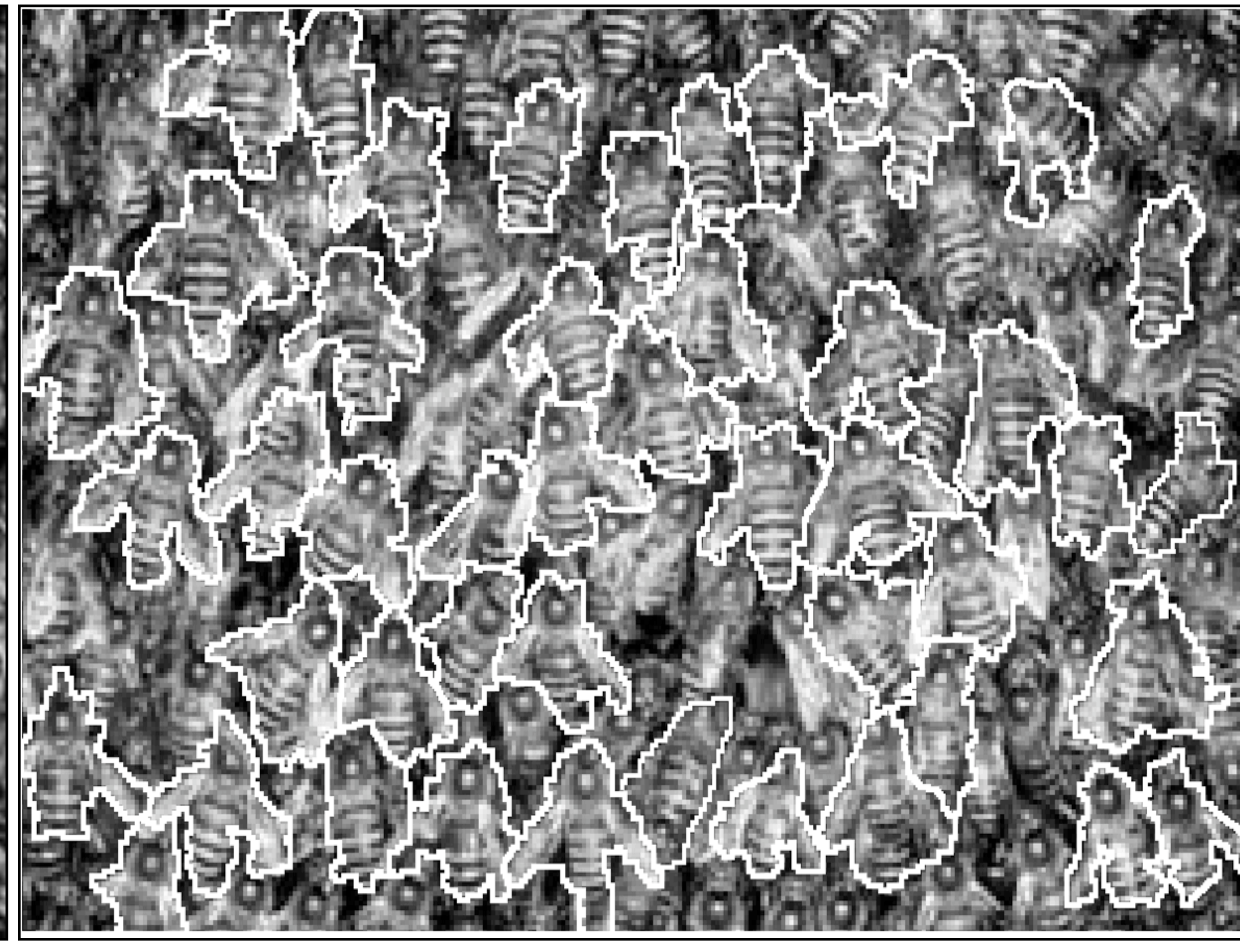


extracted texels

Ahuja&Todorovic ICCV07

Results: UIUC Texture Dataset

original image



texel segmentation

Ahuja&Todorovic ICCV07

Results: Texture Segmentation

Todorovic&Ahuja ICCV09

Results: Texture Segmentation

original image

texel-based
Todorovic&Ahuja ICCV09

filter-based
Galun et al ICCV03

Summary

Summary

- **Hierarchical region-based image representation**

Summary

- **Hierarchical region-based image representation**
- **Robust matching of regions**

Summary

- **Hierarchical region-based image representation**
- **Robust matching of regions**
- **Operative definition of an object category**

Summary

- Hierarchical region-based image representation
- Robust matching of regions
- Operative definition of an object category
- Hierarchical taxonomy of shared categories

Summary

- Hierarchical region-based image representation
- Robust matching of regions
- Operative definition of an object category
- Hierarchical taxonomy of shared categories
- The framework allows:
  - Simultaneous recognition and segmentation
  - Semantic basis of recognition
  - Space-time coherent video object segmentation
  - Texel-based texture analysis

Thank you!