Extracting Subimages of an Unknown Category from a Set of Images

Sinisa Todorovic and Narendra Ahuja

CVPR 2006

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

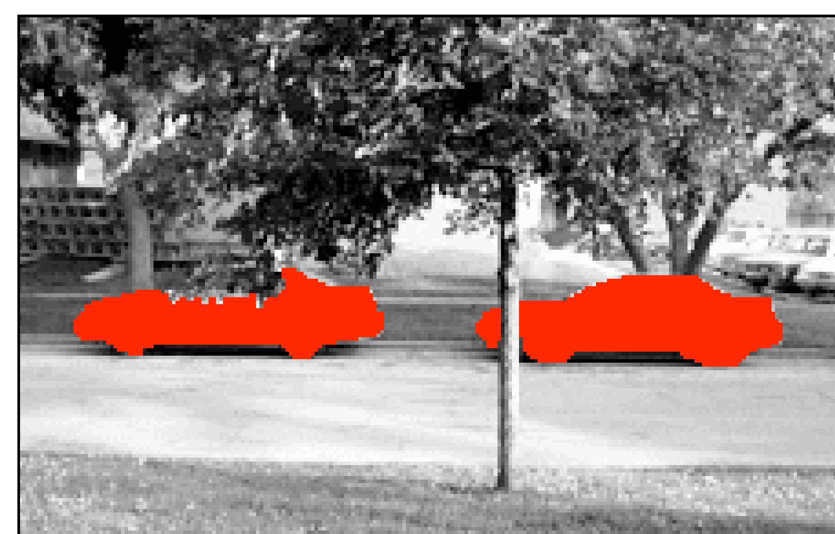
Objective: Car Category Example

occlusion　　no car　　occlusion　　multiple cars

learn car model

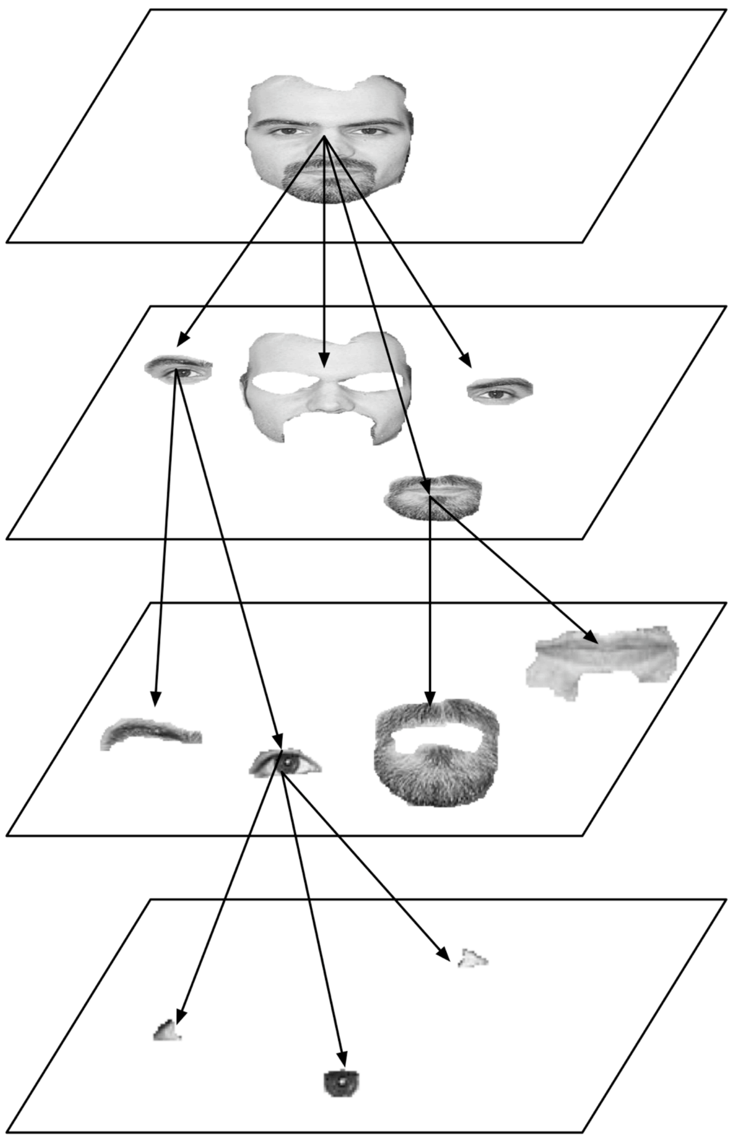
unseen image

segment all cars



RESULT

Category Modeling is Extremely Difficult



- Recursive embedding of object subparts
- Regions vs. local features open questions:
  - More informative?
  - More stable and robust to noise?
- Regions allow:
  - simultaneous object detection and segmentation
  - explicit representation of the recursive embedding property

Objective

GIVEN

Images possibly containing objects from a category

DETERMINE

If a category is present

AND IF YES LEARN

Model of the category

Training

GIVEN

An unseen image

SEGMENT

All occurrences of the category

Testing

What is Category?

CATEGORY ⇔ SET OF SUBIMAGES comprised of

REGIONS having

SIMILAR properties:

(1) Photometric -> brightness, contrasts

(2) Geometric -> area, boundary shape

(3) Topological -> layout and recursive embedding

UNSUPERVISED LEARNING OF A CATEGORY !

Rationale

CATEGORY PRESENT IN THE SET

find ?

$\Downarrow$

MANY SIMILAR SUBIMAGES

image matching

$\Downarrow$

ABUNDANT DATA

do ?

$\Downarrow$

ROBUST LEARNING IS FEASIBLE

structural learning

Prior Work Dominated By:

- Statistical modeling of local features
- Object detection ⇔ Image classification
- Object segmentation ⇔ Object localization (e.g. probabilistic map)
- A training image must contain a category
- Modeling background
- Discriminative approaches require hundreds of training images

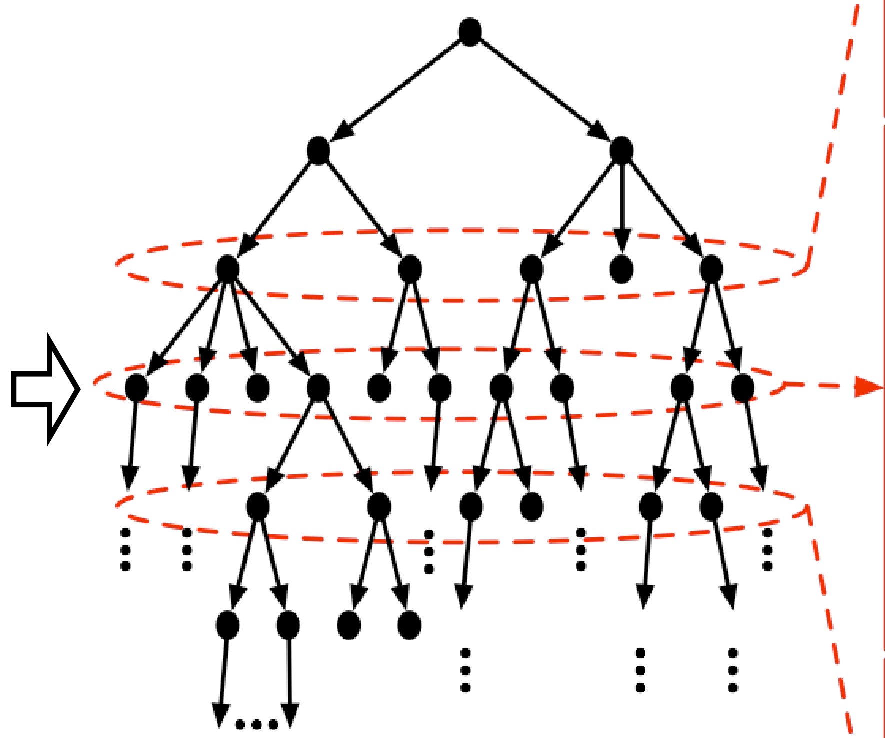
Image = Tree ⇒ Object = Subtree

[N. Ahuja TPAMI '96, Tobb & Ahuja TIP '97, Arora & Ahuja ICPR '06]

Cutsets

Example segmentations

Segmentation tree



Contrast level ≠ Tree level

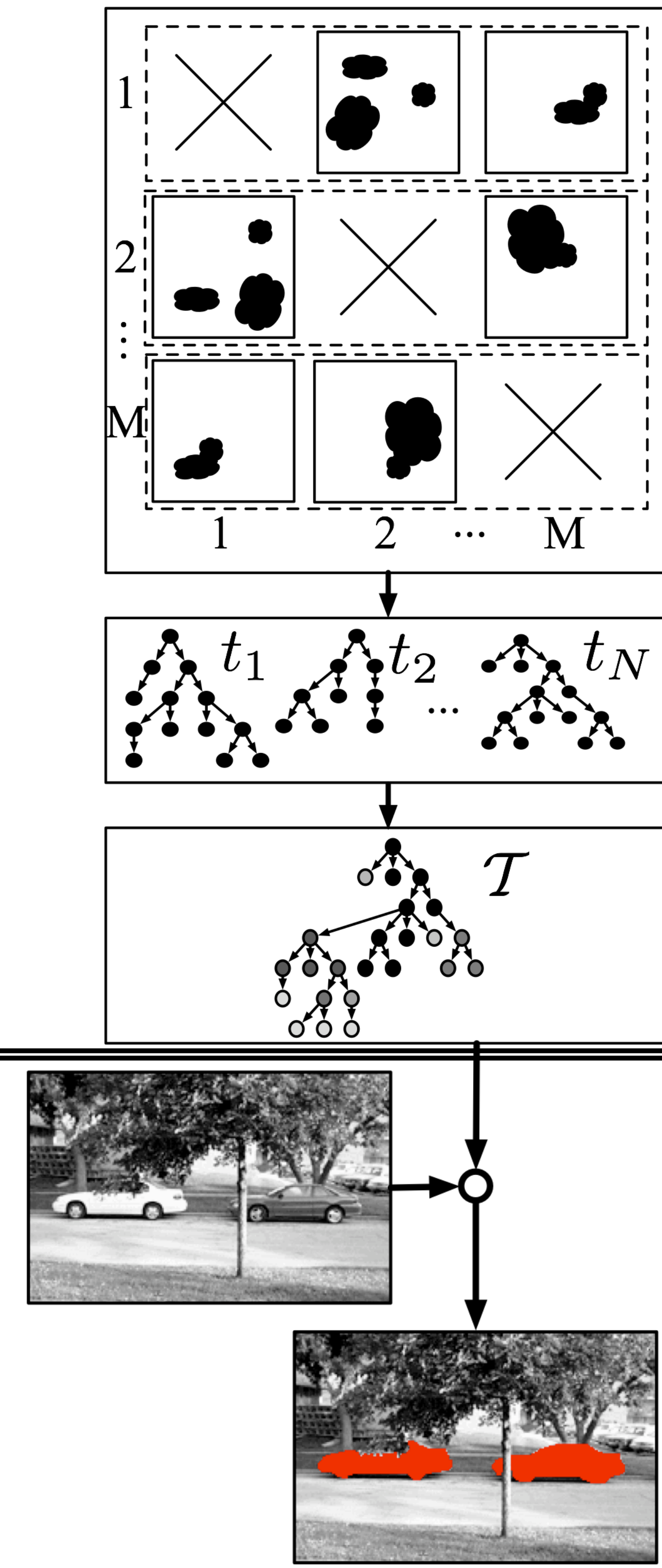
Outline of Our Approach

Images = Trees

Category present = Many similar subtrees

Extracting similar subtrees = Tree matching

Category model = Union of similar subtrees

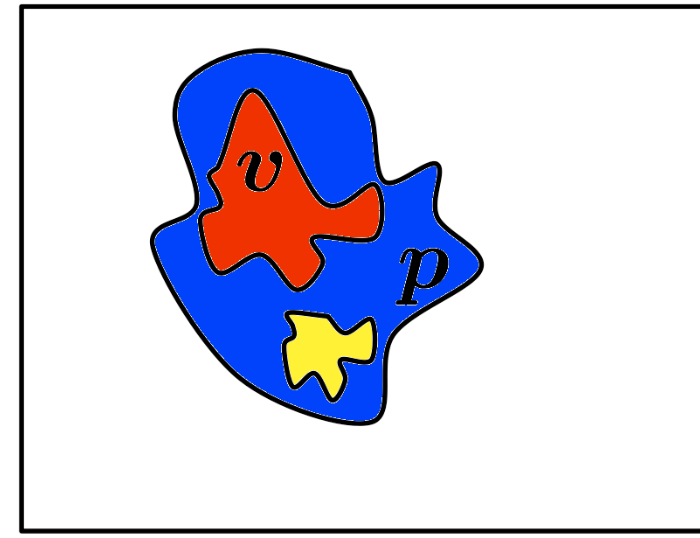


Simultaneous detection and segmentation of ALL category instances

=

Matching model with image

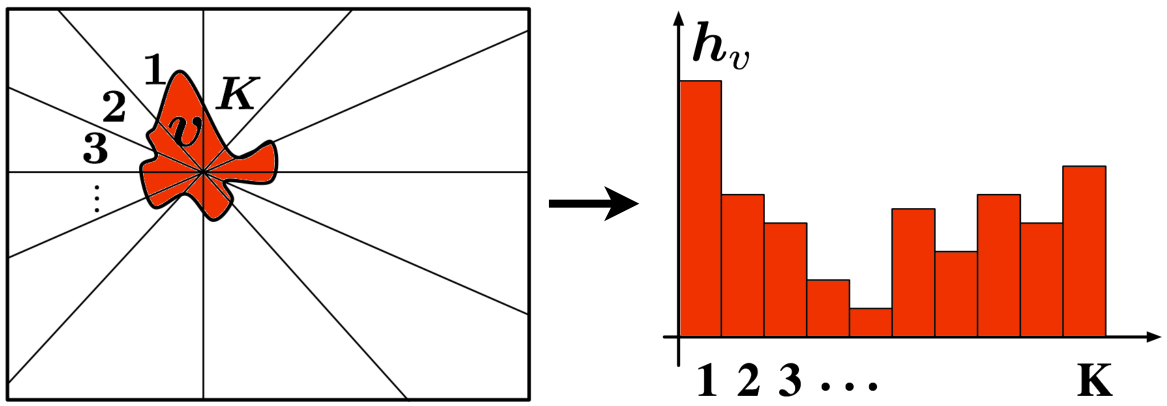
Intrinsic Region Properties



- **Gray-level mean, variance, and area**

  $\mu_v, \ \sigma_v^2, \ a_v$

- **Rotation invariant boundary shape context**



$$\boldsymbol{h}_v = \{h_v(1), h_v(2), \ldots, h_v(k)\}$$

- **Derived quantity: region saliency**

$$w_v \triangleq \lambda\left[\frac{|\mu_v - \mu_p|}{\max(\mu_v, \mu_p)} + \frac{|\sigma_v^2 - \sigma_p^2|}{\max(\sigma_v^2, \sigma_p^2)}\right] + (1-\lambda)\left[\frac{a_v}{a_p} + H_v\right],$$

$$H_v \triangleq -\sum_{k=1}^{K} h_v(k) \log h_v(k)$$

Chosen to make recognition invariant to rotation and scale changes

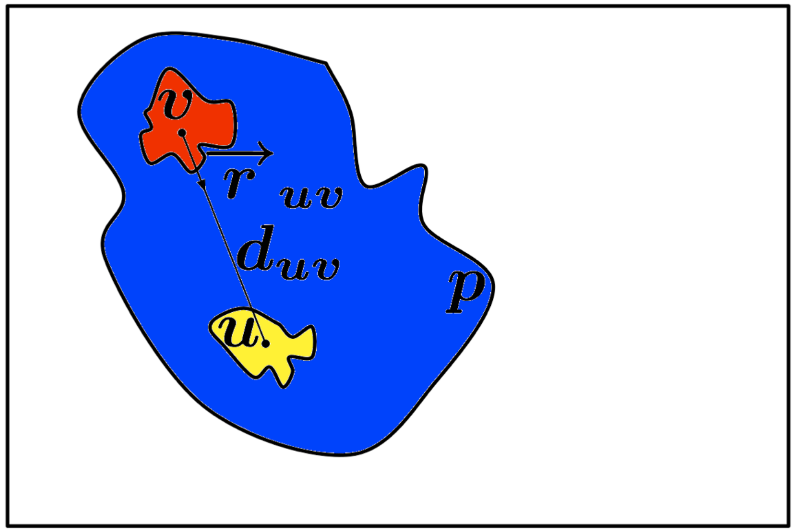
Relative Region Properties

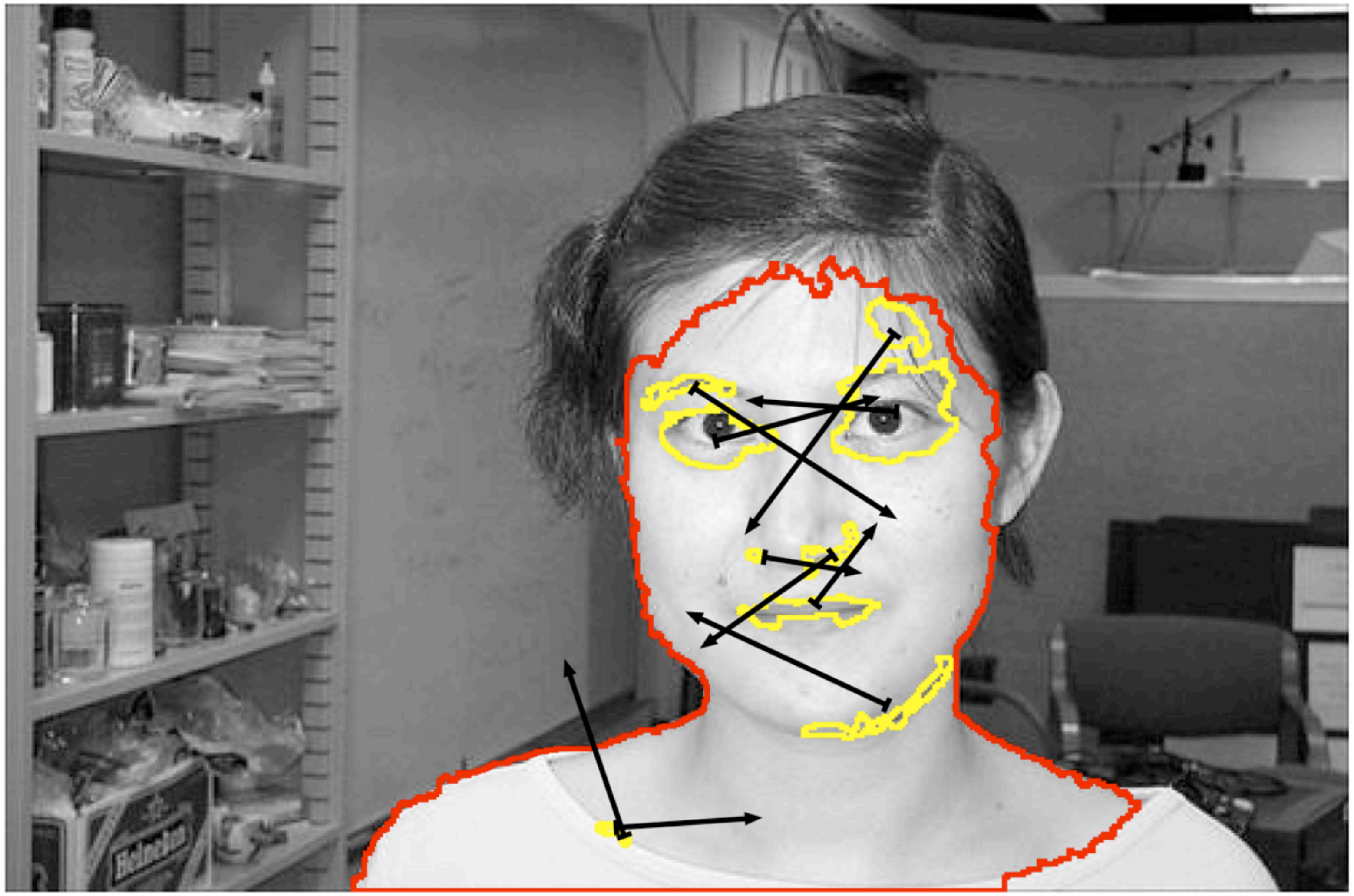
- **Context vector: attraction field at the centroid of a region**

$$\overrightarrow{\Phi}_v = \sum_{u \in \mathcal{N}_v} \frac{w_u}{d_{uv}^2} \overrightarrow{r}_{uv} = \{|\overrightarrow{\Phi}_v|, \phi_v\}$$

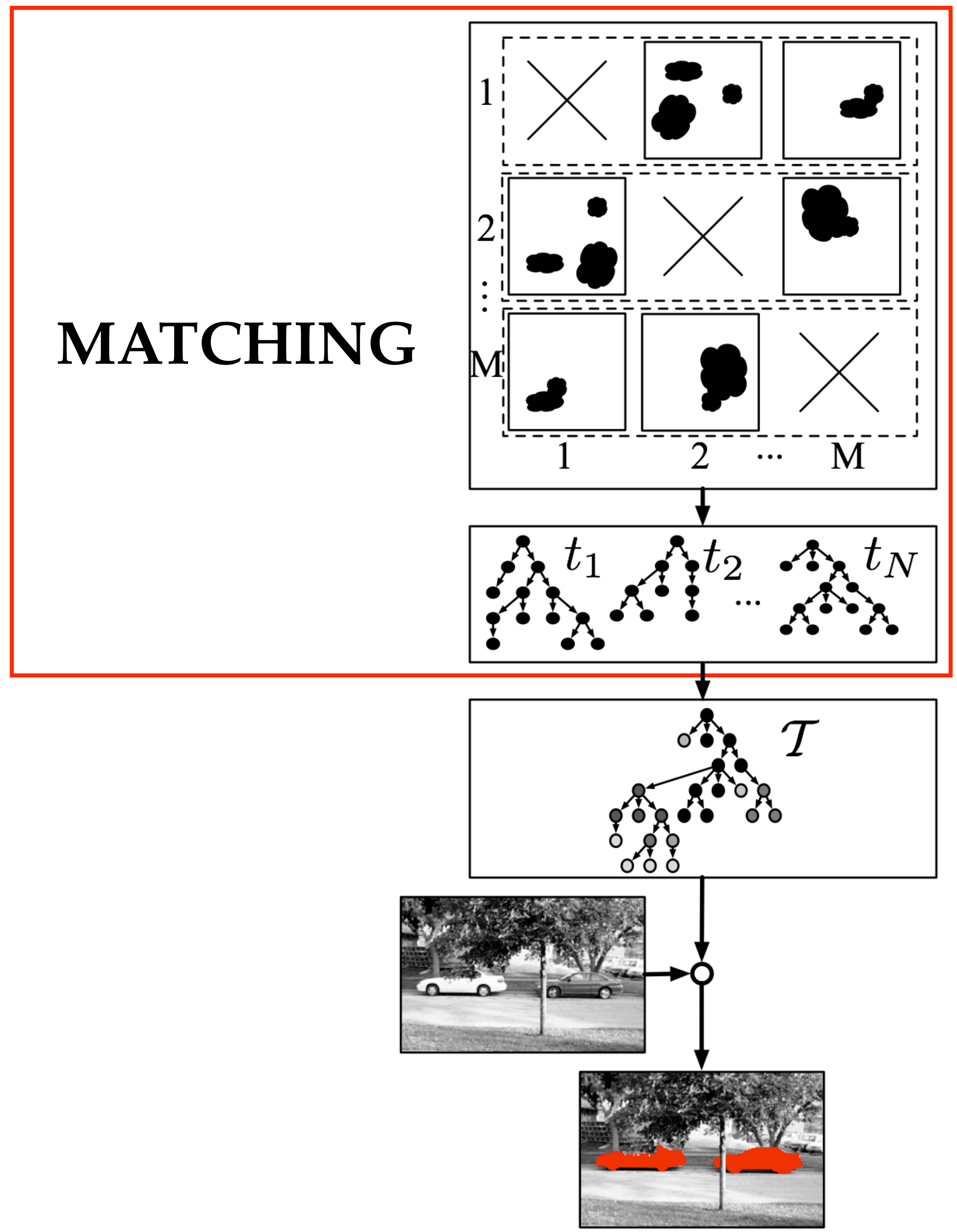
neighborhood

Rotation invariant relative to the parent



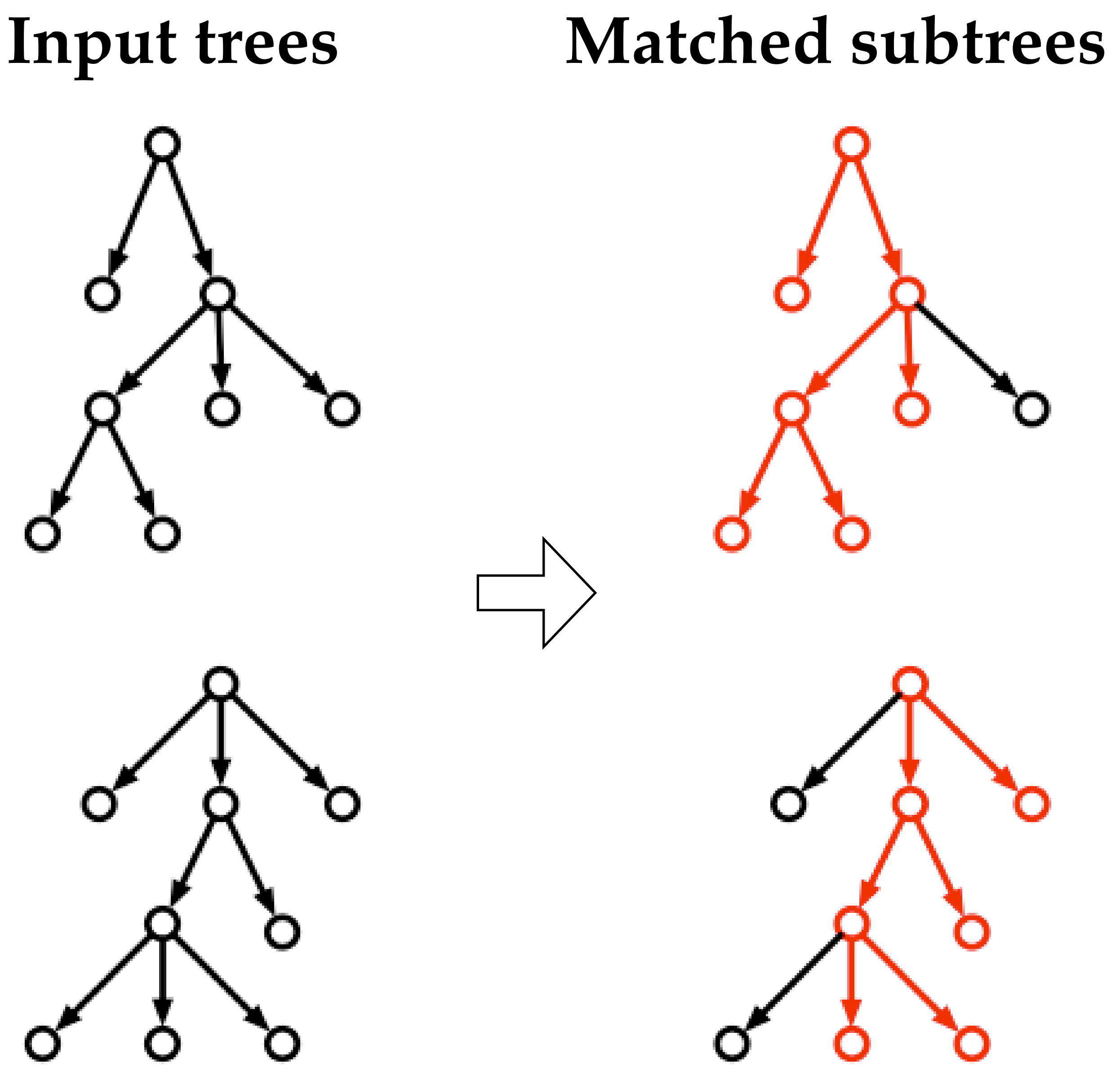


Outline



Matching Algorithm

[Torsello & Hancock ECCV'02, ECCV'04]



Matching Algorithm

GIVEN two trees: $t, \ t'$

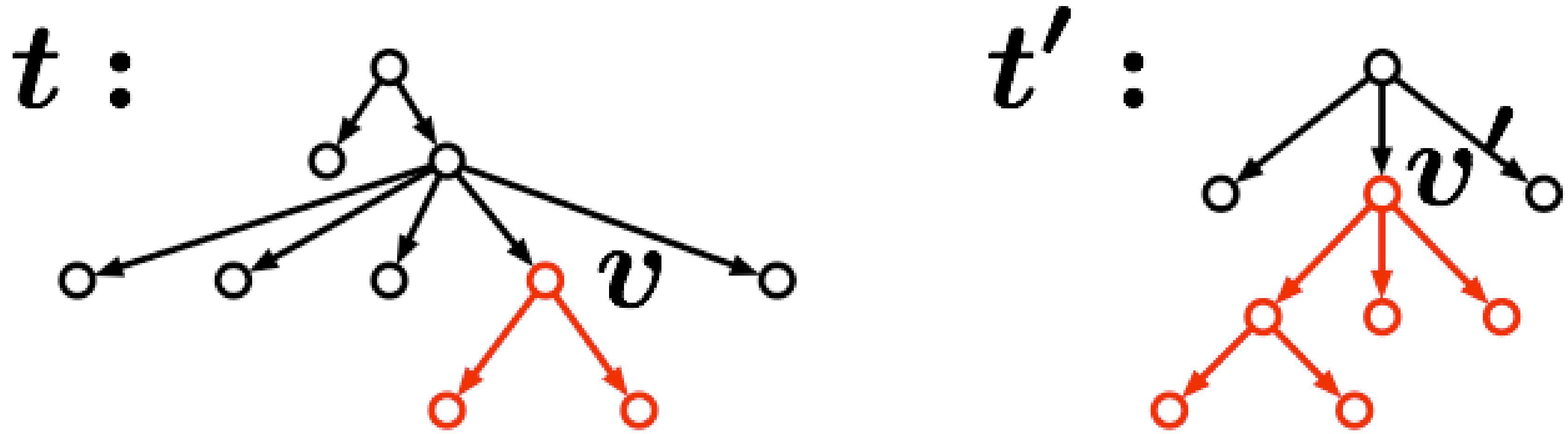
FIND bijection $f : (v, v'), \ v \in t, \ v' \in t'$

which MAXIMIZES the QUALITY OF MATCH

$$\mathcal{U}(t,t') = \sum_{(v,v') \in f} [\underbrace{w_v + w_{v'}}_{\text{node saliency}} - \underbrace{m_{vv'}}_{\text{cost of node matching}}]$$

while PRESERVING ancestor-descendant relationships

Matching Algorithm: Recursive Solution



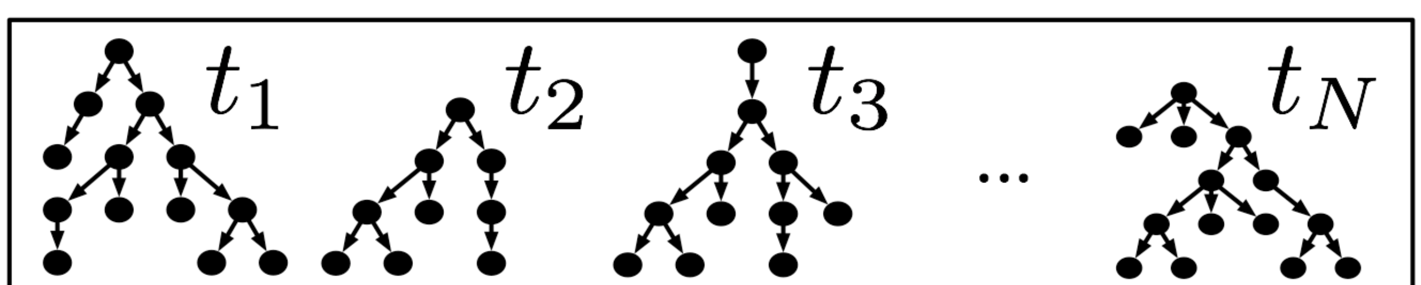
$$\mathcal{U}(t_v, {t'}_{v'}) = w_v + w_{v'} - m_{vv'} + \max_{\mathcal{C}_{vv'}} \sum_{(d,d') \in \mathcal{C}_{vv'}} \mathcal{U}(d, d')$$

Maximum clique over all descendant pairs

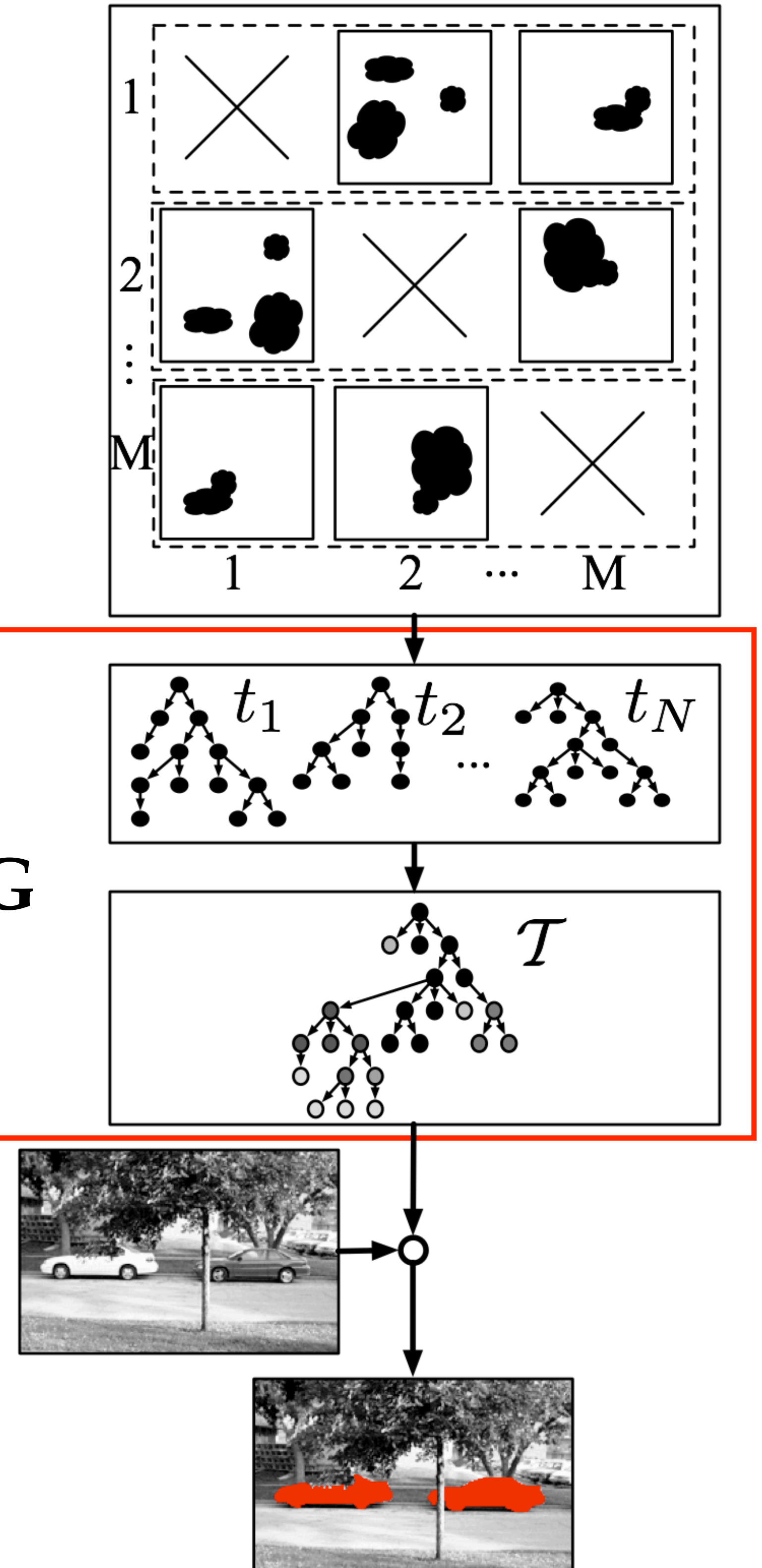
descendants

SOLUTION

Select all pairs (v, v') with $\mathcal{U}(t_v, {t'}_{v'}) >$ threshold.



Outline

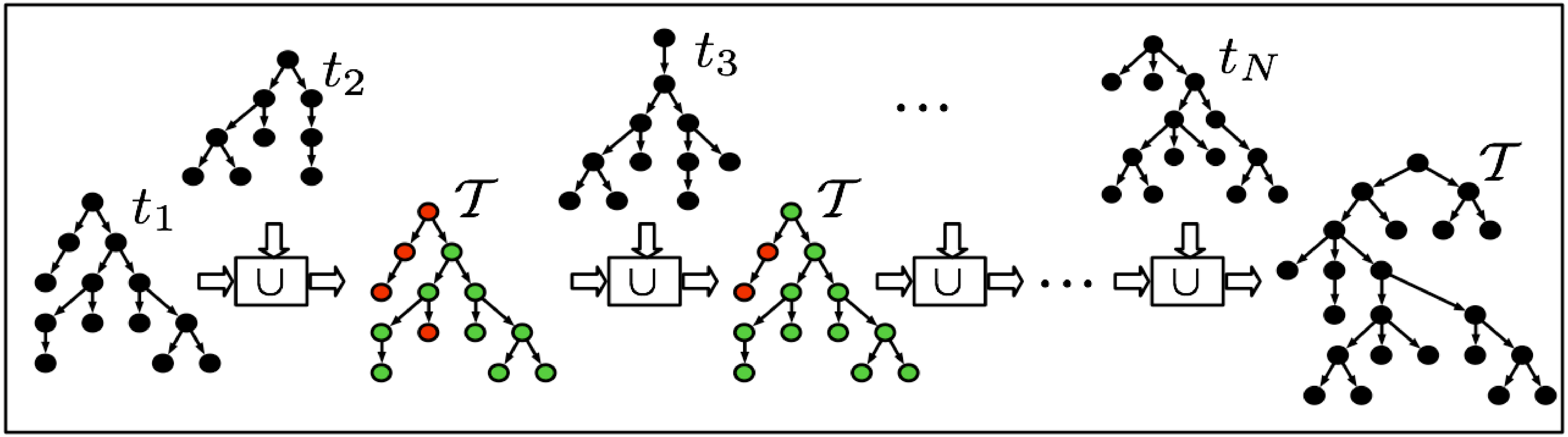


LEARNING

Category Model = Tree Union

$$\tau = t_i \cap t_{i+1}$$

$$\mathcal{T} = \tau \cup t_i \setminus \tau \cup t_{i+1} \setminus \tau$$



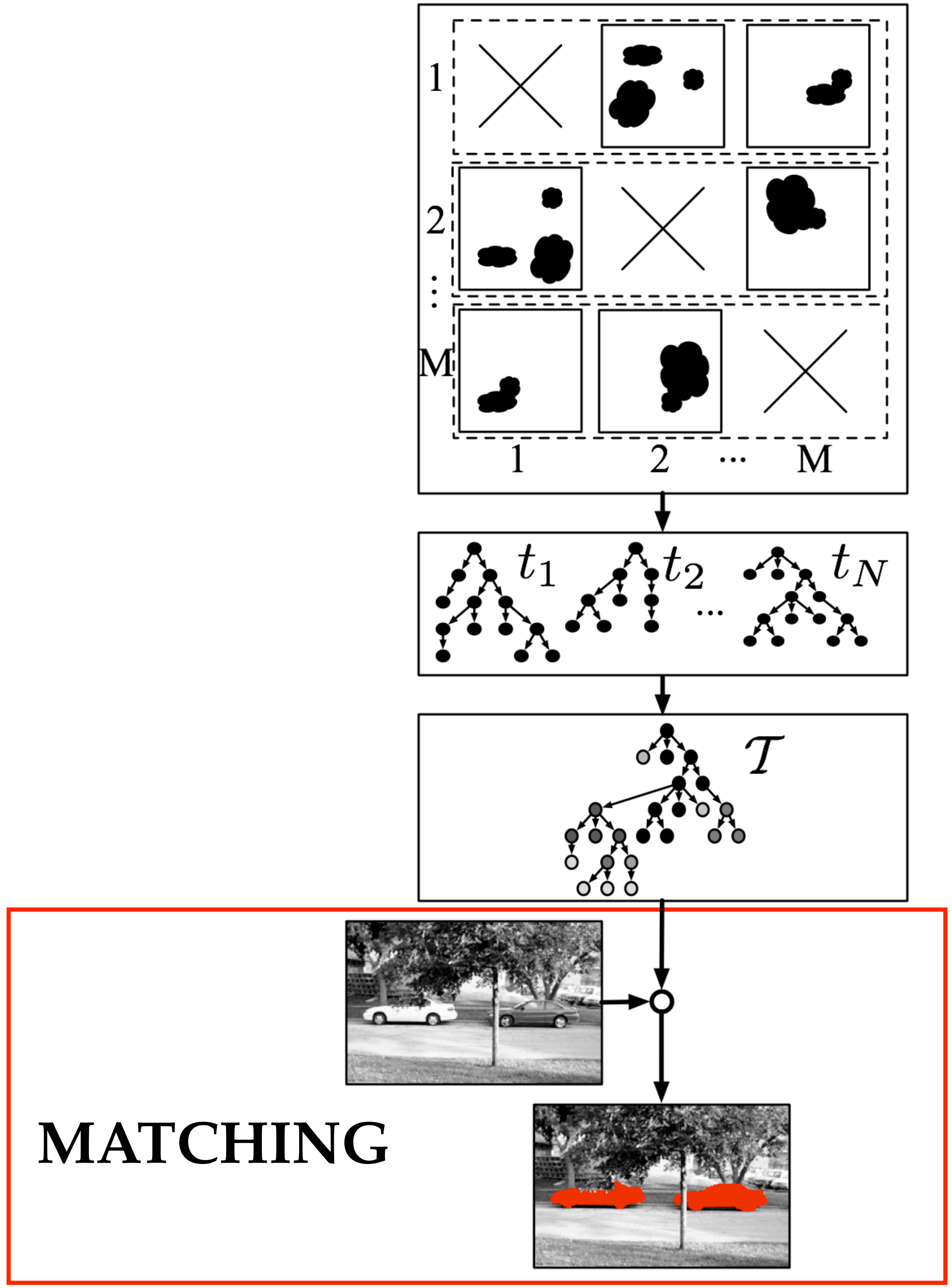
Structural learning estimates:

1) **Data-model correspondence**

2) **Model structure**

3) **Model parameters**

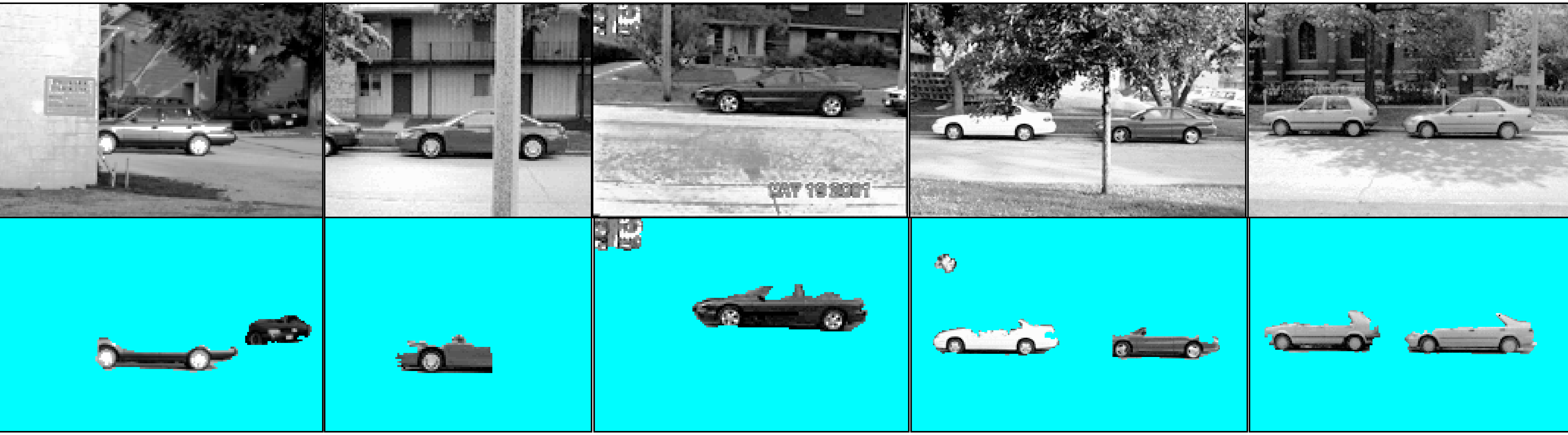
Simultaneous Detection and Segmentation



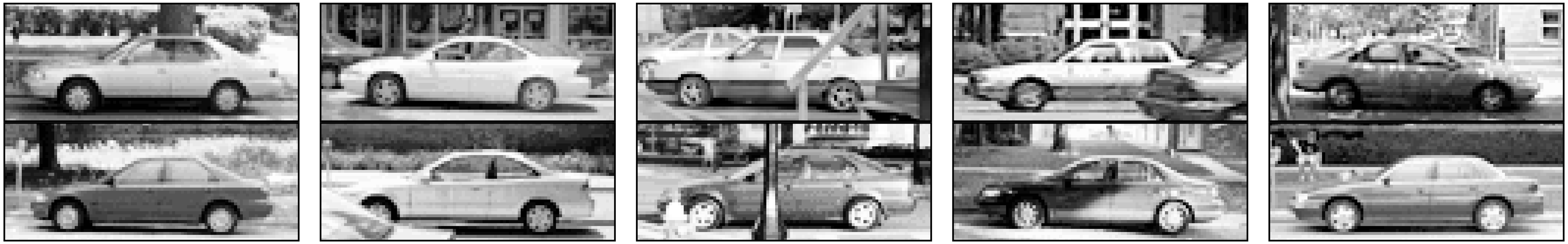
Results: UIUC Cars Side View

5 positive out of 10 training images

Results on test images:

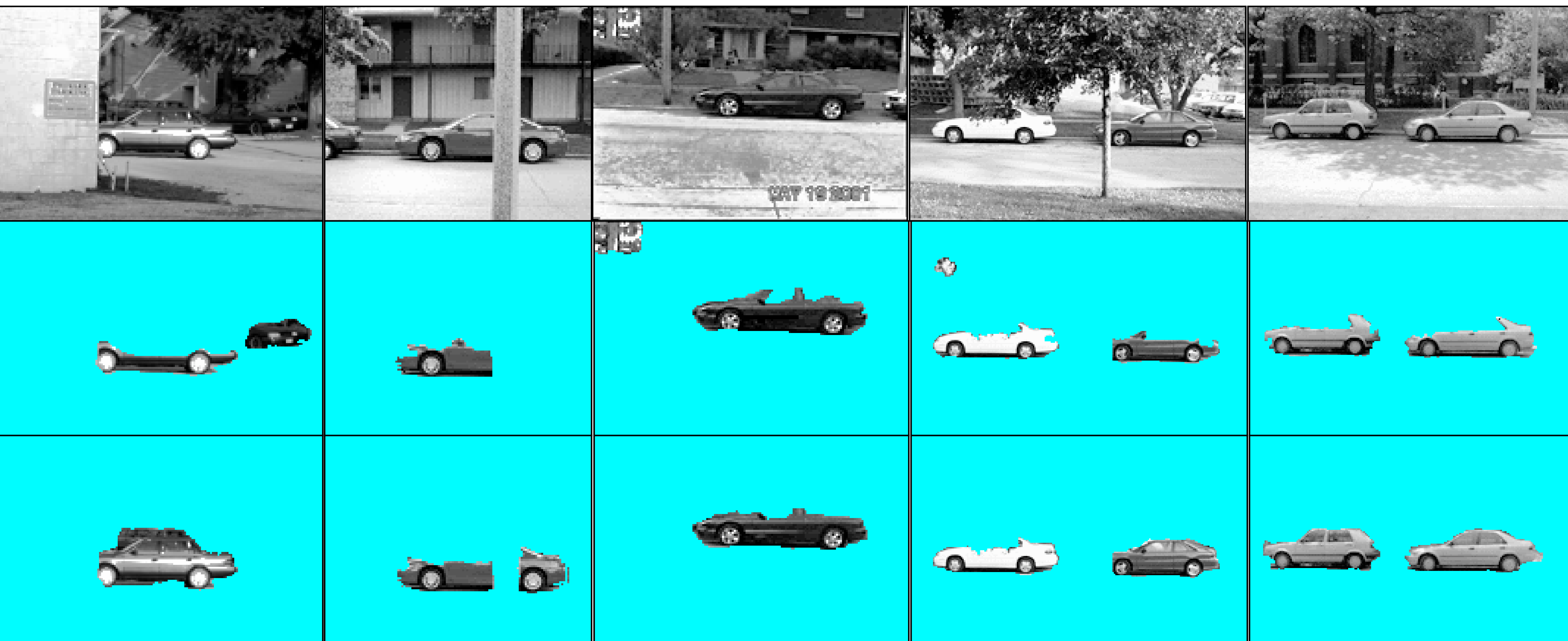


Results: UIUC Cars Side View



10 positive out of 20 training images

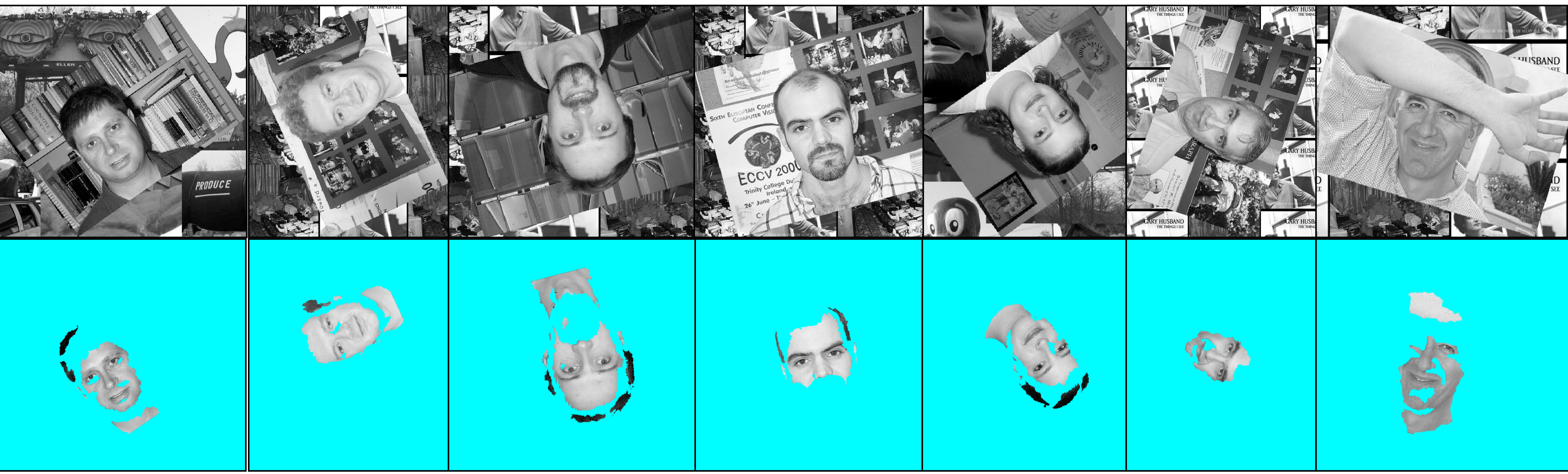
Results on test images:



Results: Faces – Caltech 101 Database

3 positive out of 6 training images

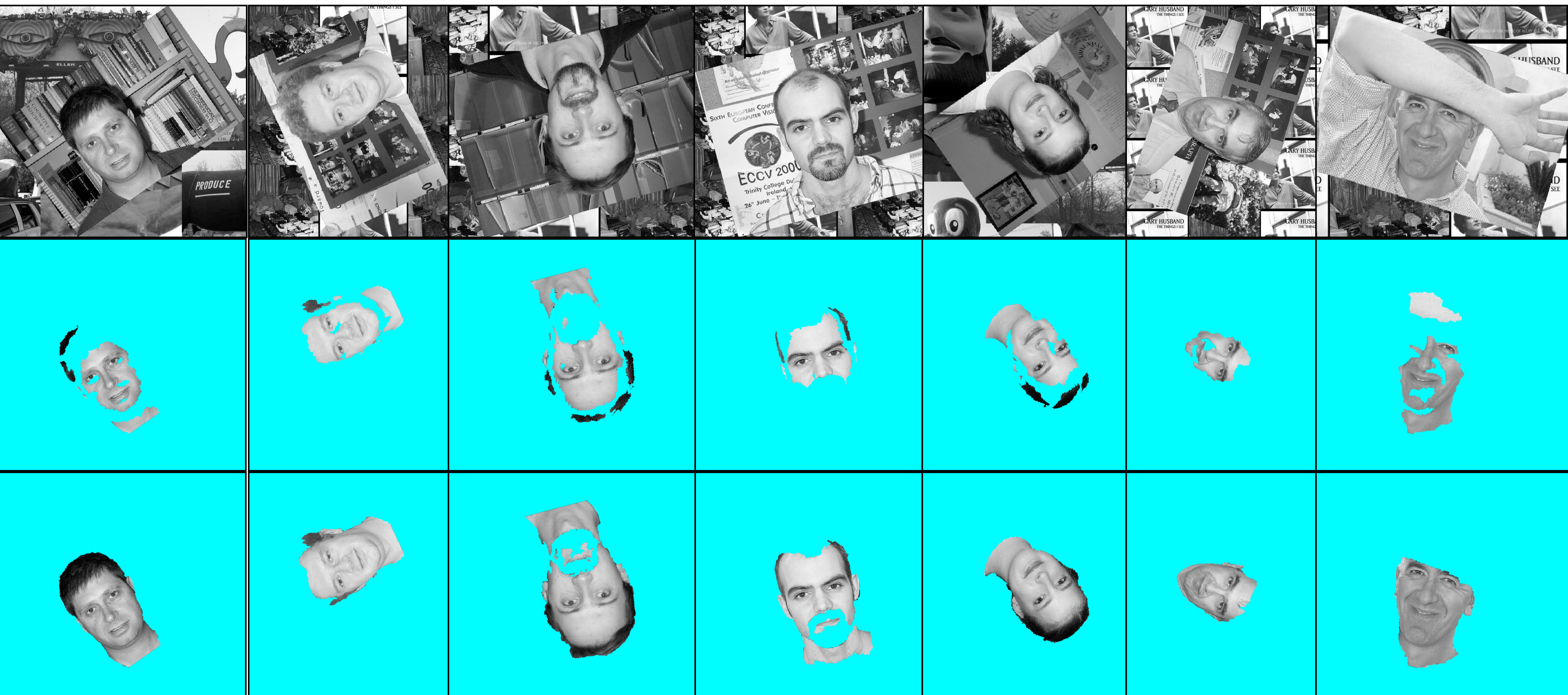
Results on test images:



Results: Faces – Caltech 101 Database

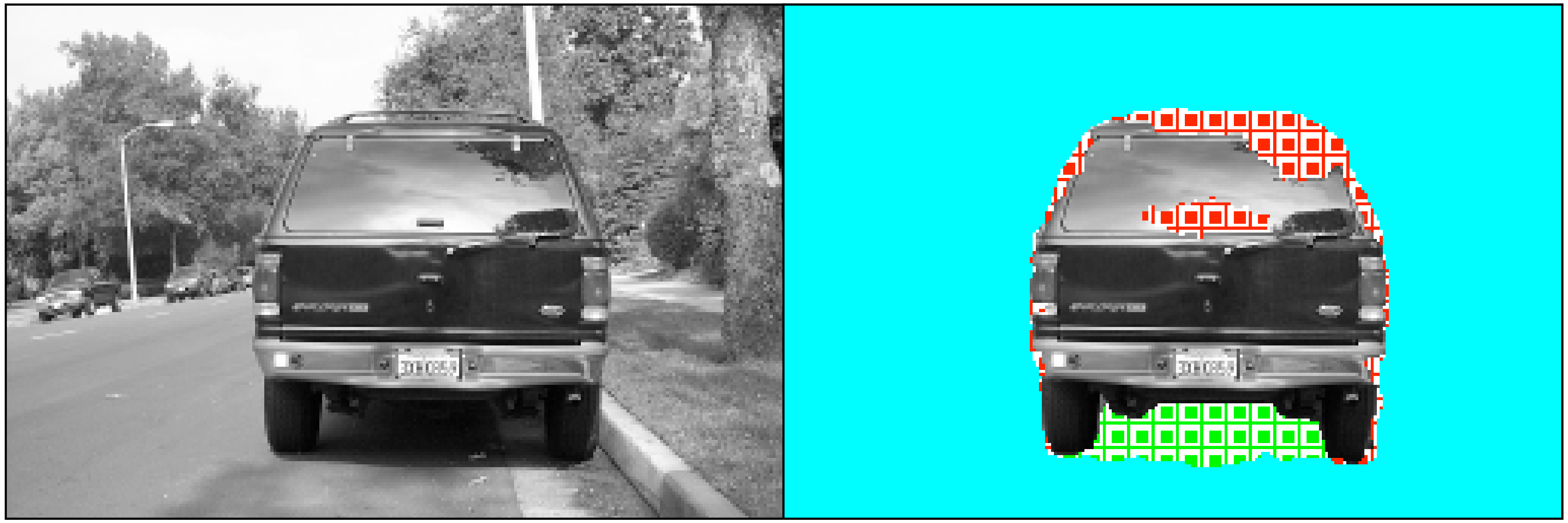
6 positive out of 12 training images

Results on test images:

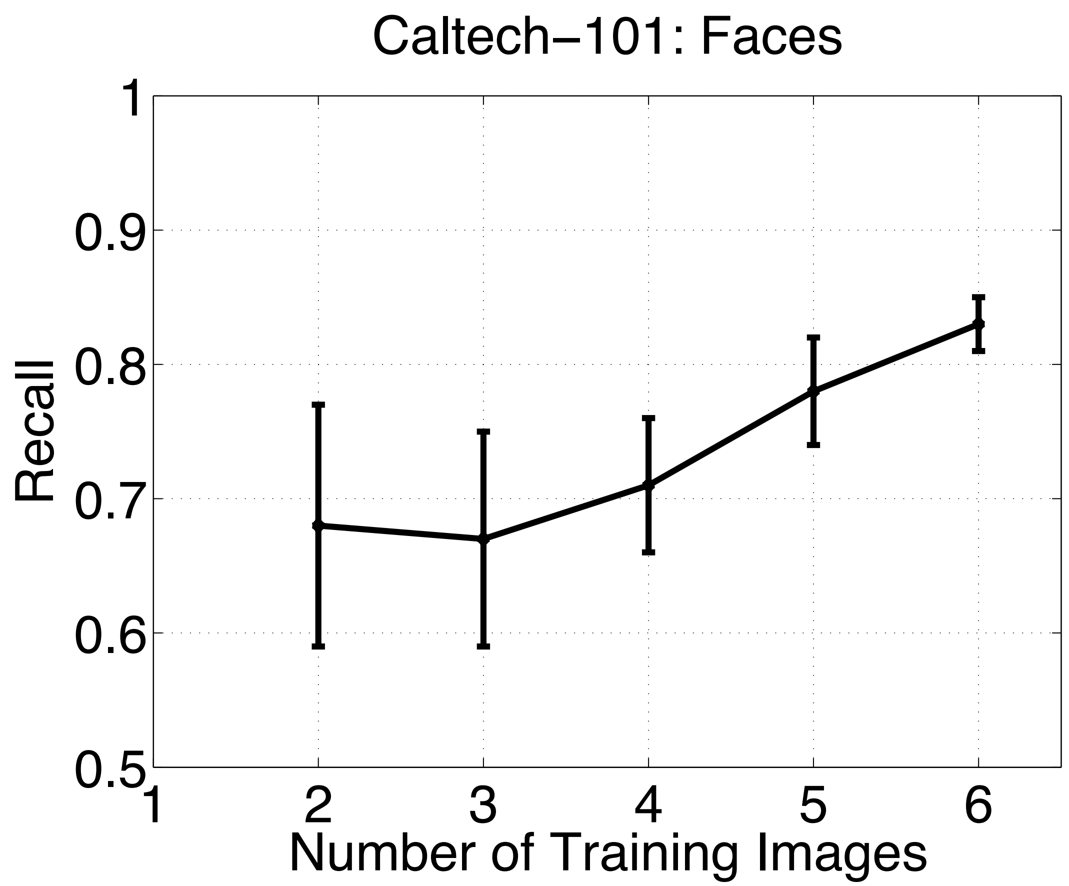


Results: Caltech Cars Rear View

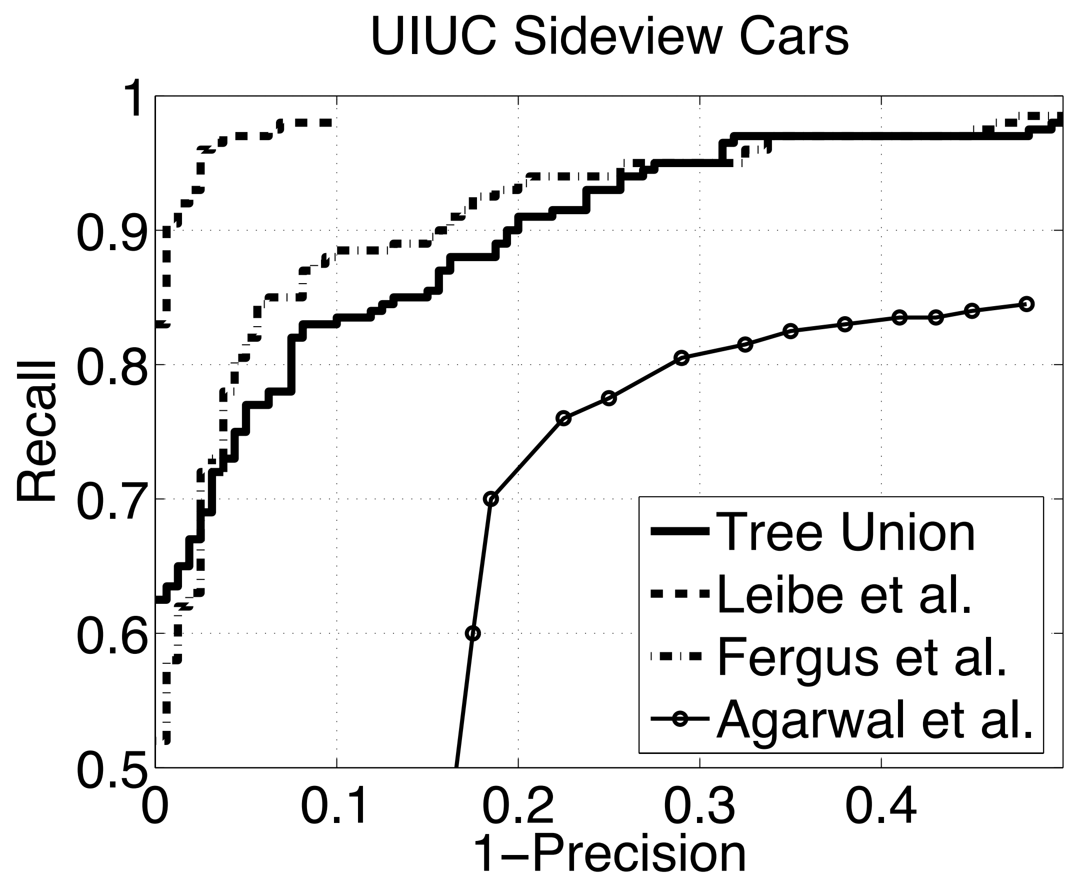
10 positive out of 20 training images



Recall-Precision



Training from a small-size dataset



Varying evaluation criteria

Complexity and Runtime on 2.4GHZ 2GB RAM PC

Extracting similar subtrees: $O(|V|^4)$ **per image pair**

# of tree nodes

Training on 20 images of UIUC CARS: < 2 hours

Learning: $O(|V_s|^4)$

# of subtree nodes

Learning on 32 subtrees extracted for UIUC CARS: < 1 hour

Detection and segmentation: $O(|V_{\mathcal{T}}|^4)$

# of model nodes

Processing time for UIUC CARS: < 10 sec,
regardless of the total number of target objects

Summary and Conclusion

- <u>Unsupervised</u> category detection and learning
- Region-based, <u>structural</u> approach
- <u>Simultaneous</u> detection and segmentation of <u>all</u> objects
- NO multiple detections on the same object
- NO hypothesis on the number of objects and their parts
- Small number of training images
- Complexity comparable with standard methods

Acknowledgment

Himanshu Arora provided the segmentation code

THANK YOU!

{sintod, ahuja}@vision.ai.uiuc.edu

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