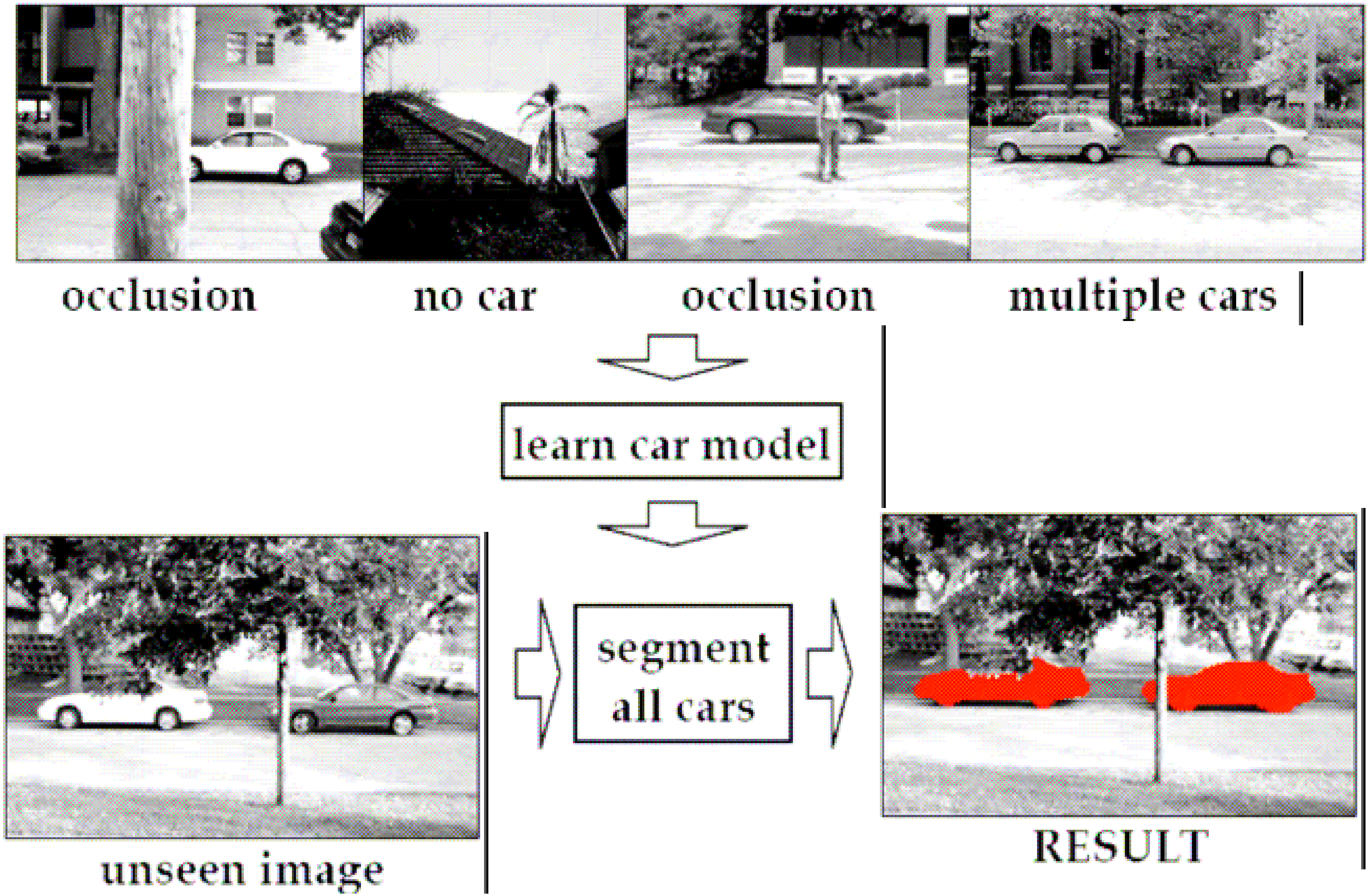
Extracting Subimages of an Unknown Category from a Set of Images

Sinisa Todorovic and Narendra Ahuja

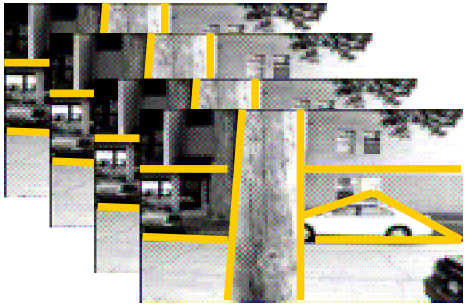
Beckman Institute, UIUC

Presented by Tingfan Wu

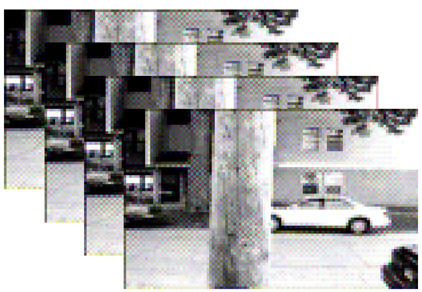
Objective


occlusion
no car
occlusion
multiple cars
learn car model
segment
all cars
unseen image
RESULT

General Steps



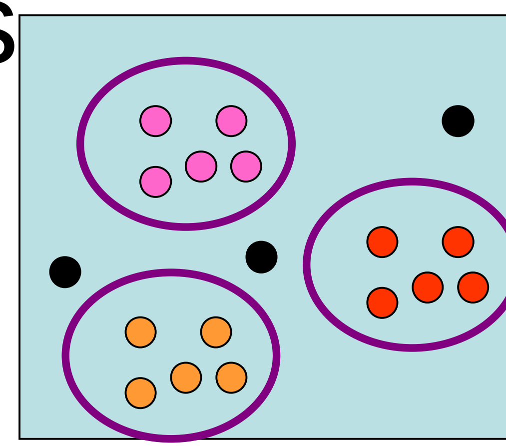
Random segments



Training Images

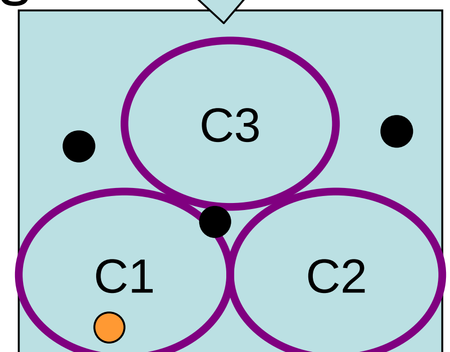
F1=(x1,x2….xn)
F2=(x1,x2….xn)
F3=(x1,x2….xn)
F4=(x1,x2….xn)

feature vectors

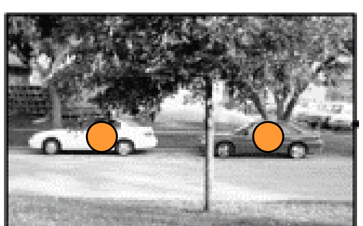


Clusters

- Varieties
  - Segmetation Methods
  - Feature Spaces
  - Clustering Methods



Models

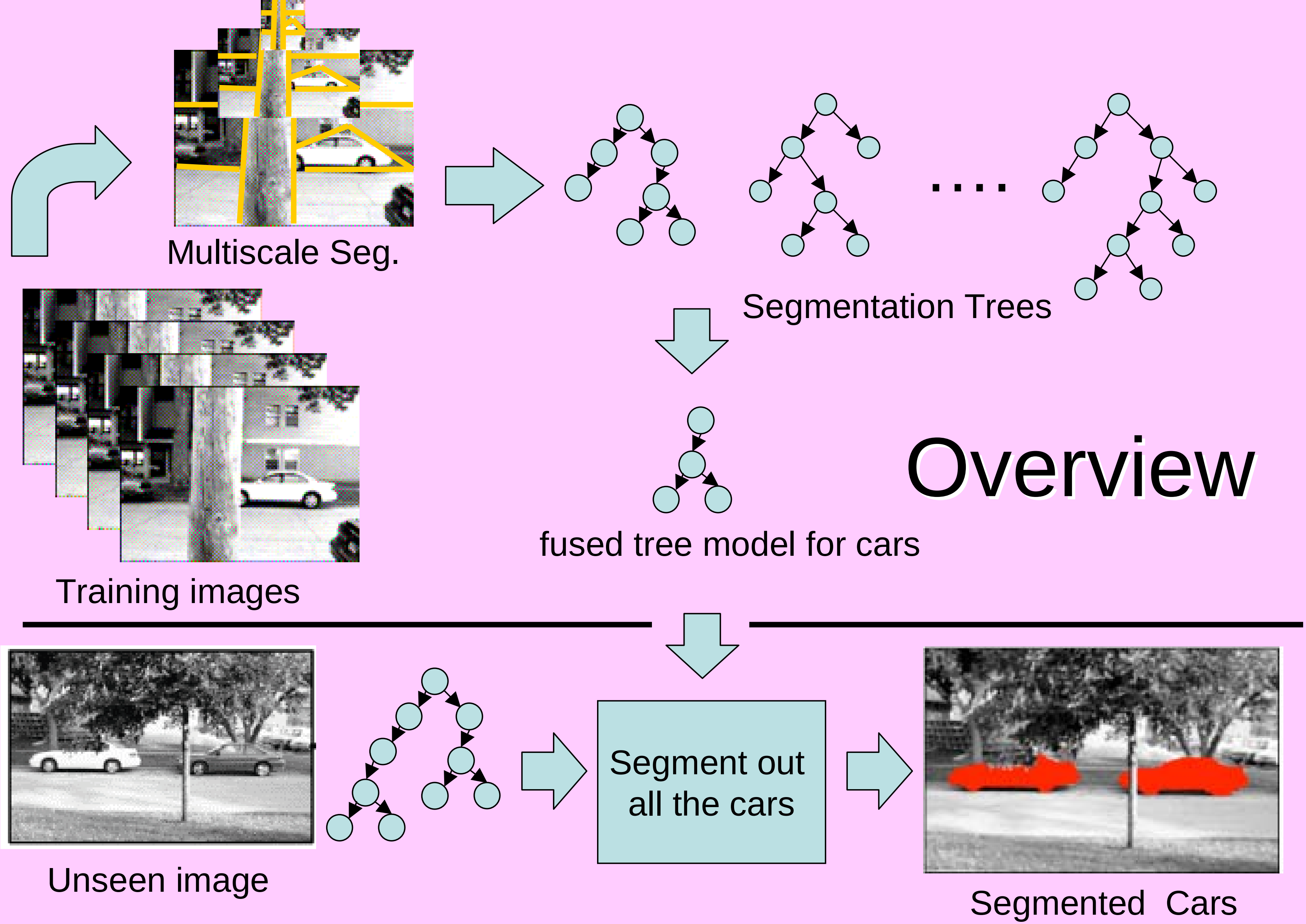


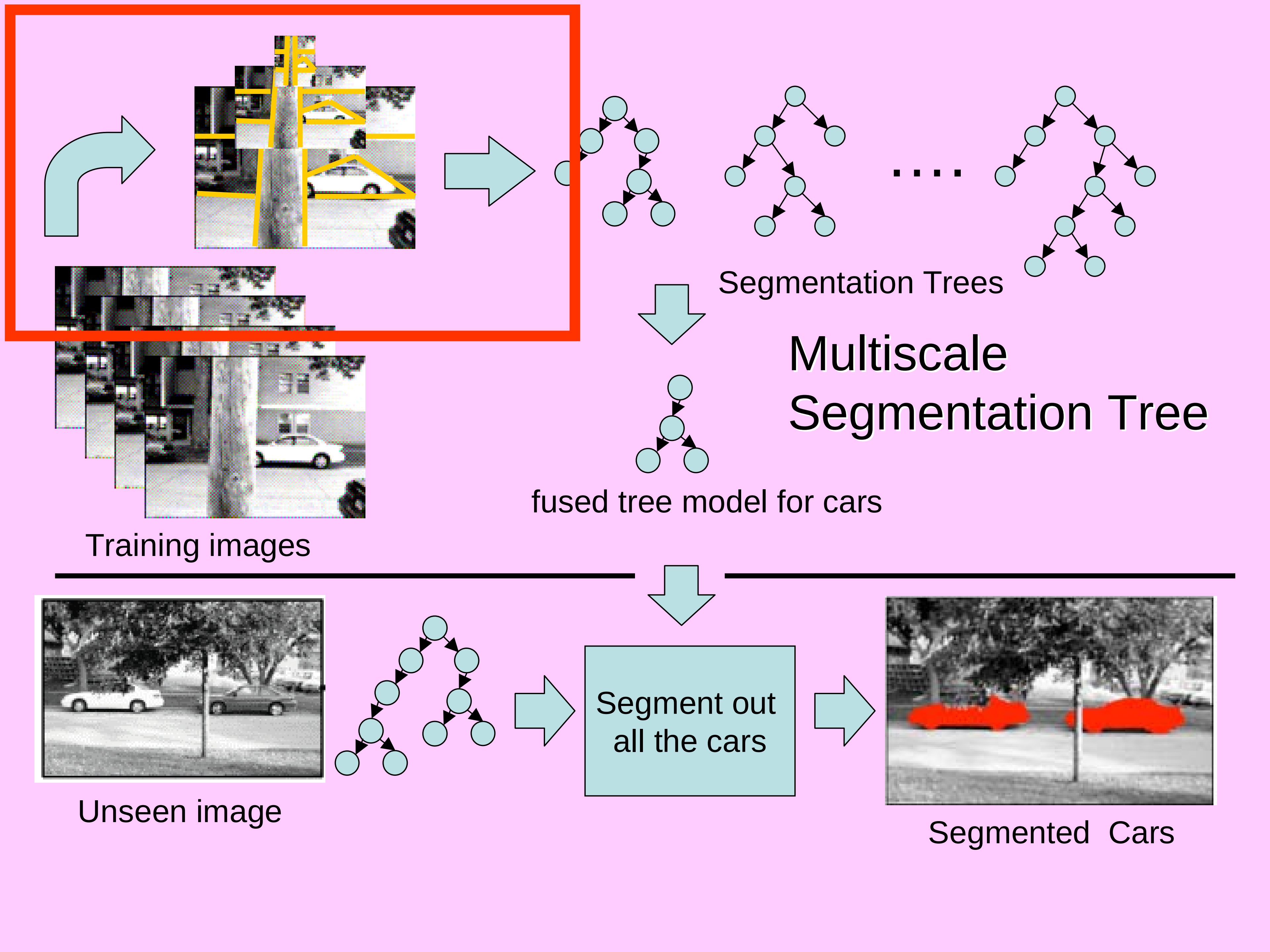
Unseen image

Ft1=(x1,x2….xn)
Ft2=(x1,x2….xn)
Ft3=(x1,x2….xn)
Ft4=(x1,x2….xn)

feature vectors

= C1


Multiscale Seg.
....
Segmentation Trees
Overview
fused tree model for cars
Training images
Unseen image
Segment out all the cars
Segmented Cars



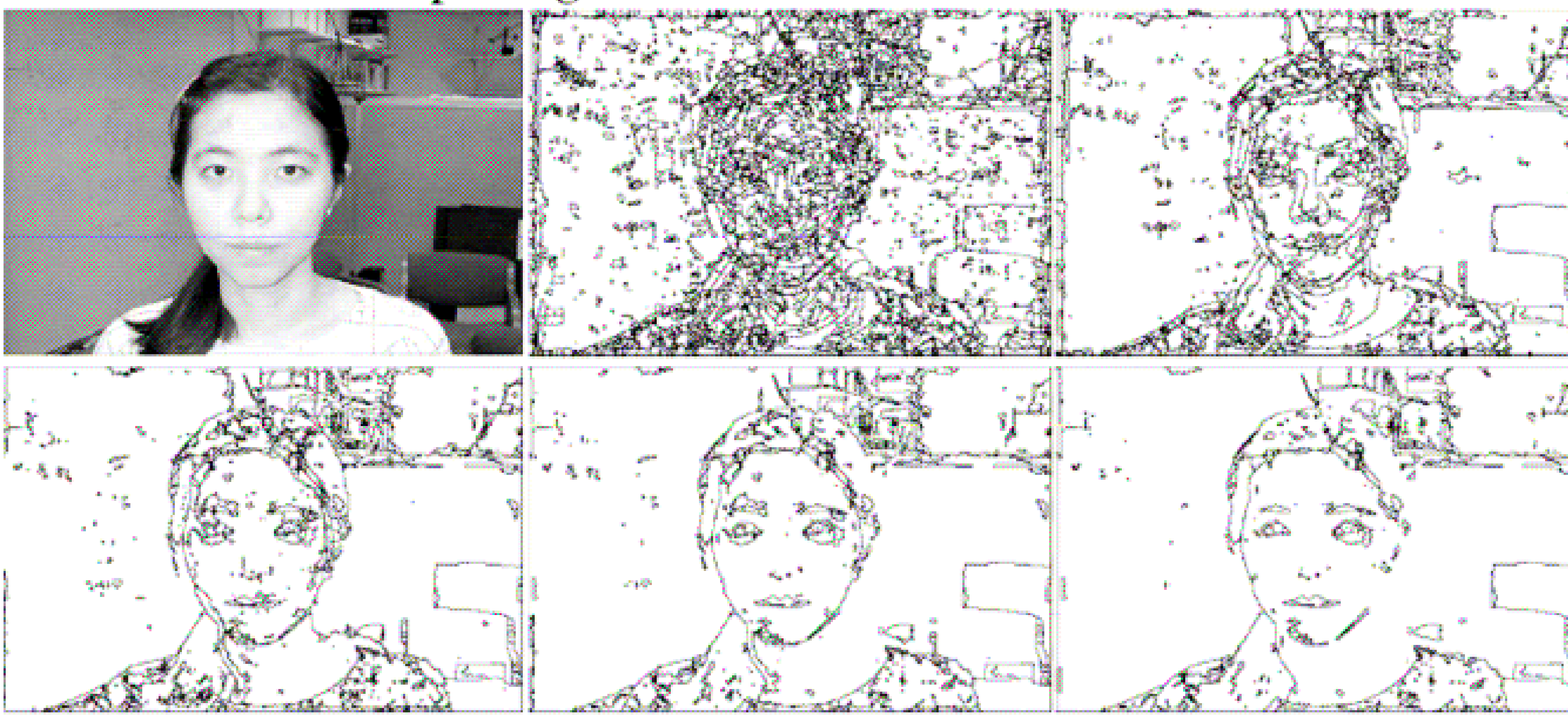
Segmentation Trees
....
Multiscale
Segmentation Tree
fused tree model for cars
Training images
Unseen image
Segment out
all the cars
Segmented Cars

Feature Extraction = Image Segmentation

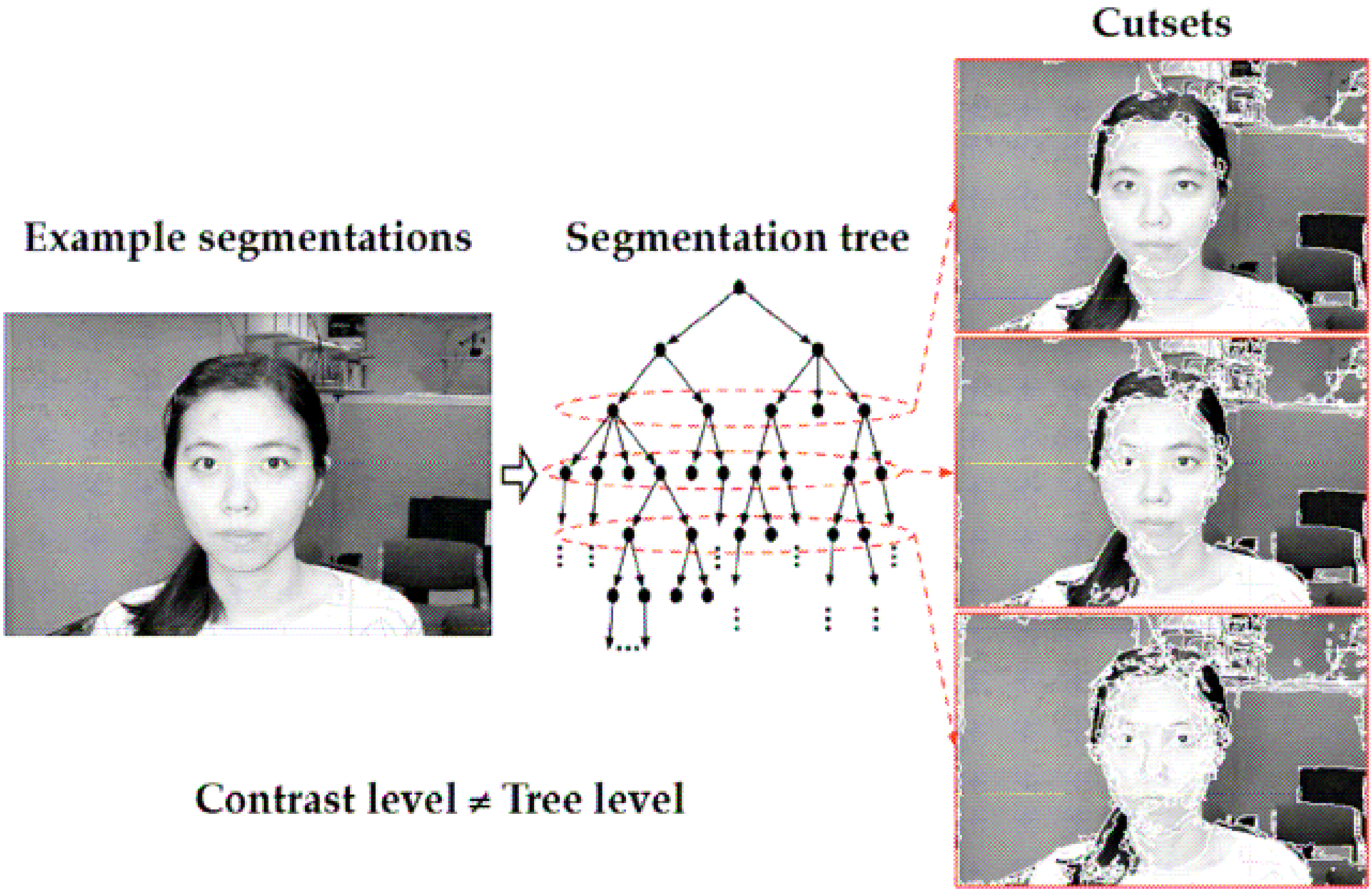
Image $\xrightarrow{\text{segmentation}}$ Homogeneous regions at ALL contrasts and $\boxed{\text{sizes}}$

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

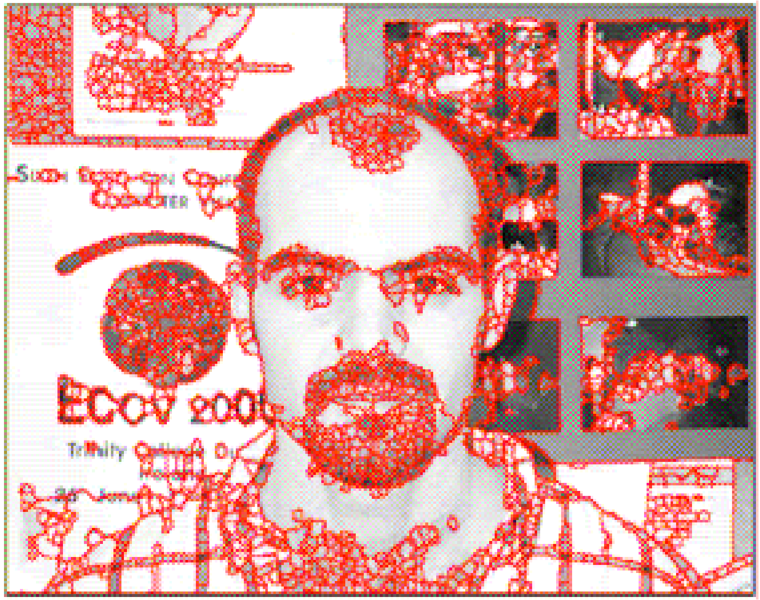
Example segmentations for several contrasts

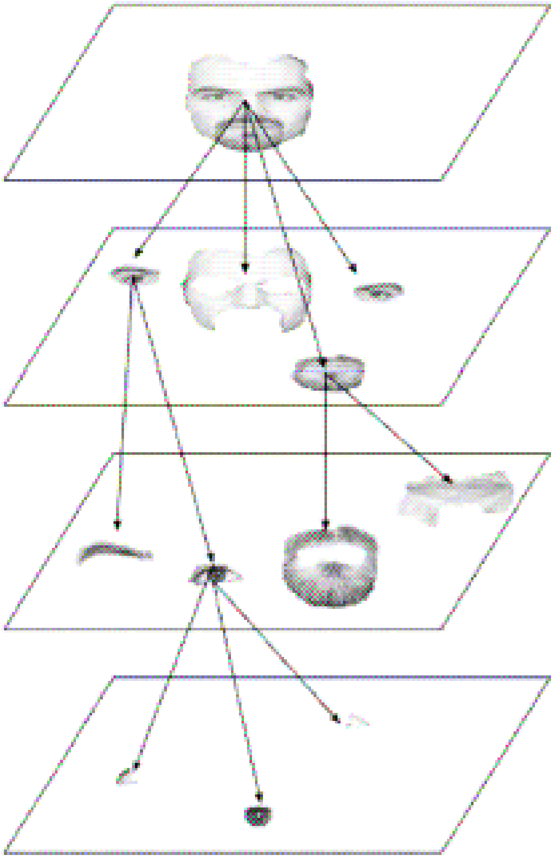


Multiscale Segmentation Tree



Region Descriptor on Tree Node

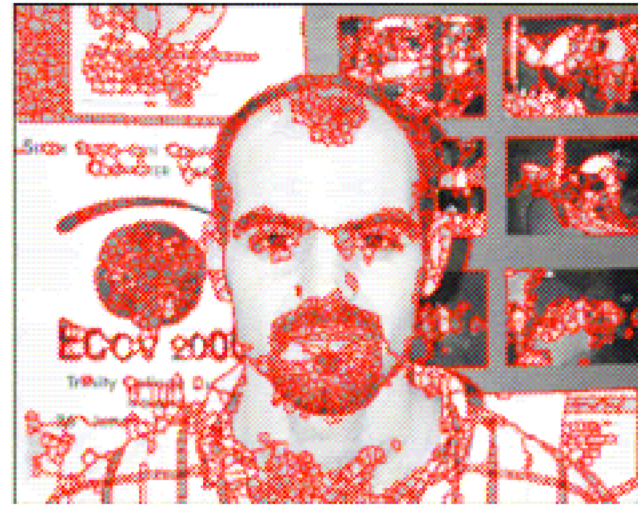


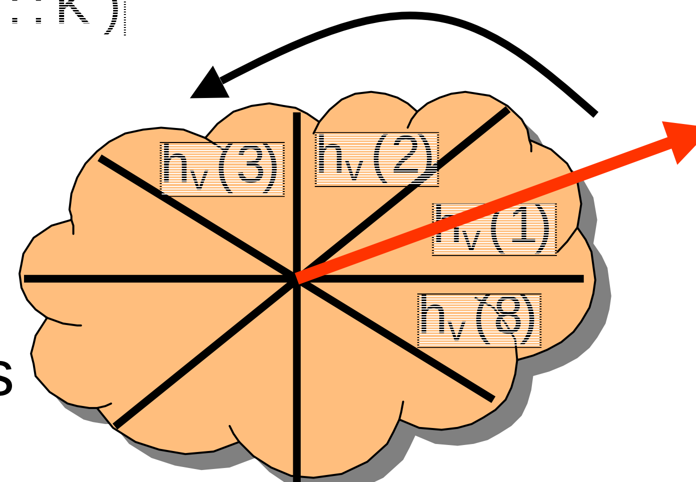


Attr(Node) = Description of the region

What are good region descriptors?

- Photometric
  - Gray level (μ_v, σ_v^2)
- Geometric (rotation invariant)
  - Area (a_v)
  - C.M. (x_v, y_v)
  - Boundary Shape Histogram $h_v(1 \ldots K)$
- Hybrid
  - Salient descriptor $(©_v)$
- Topology
  - Recursive containment of regions



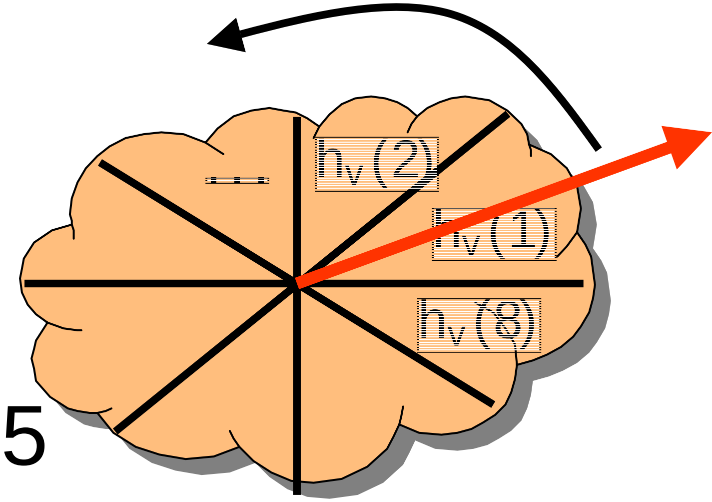


Can be rotation invariant

Salient Descriptor for a Region

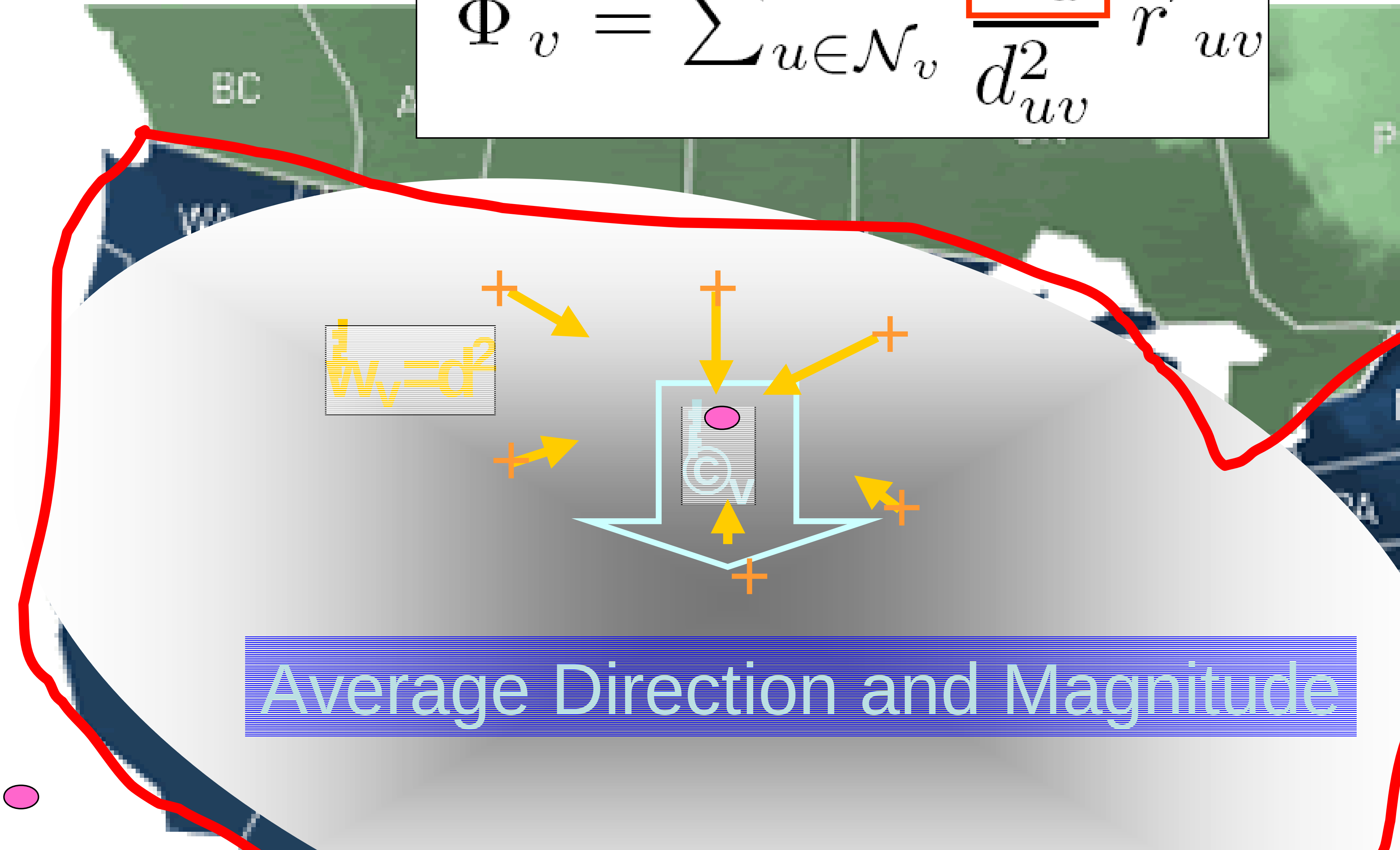
$$w_v \triangleq \lambda\Big[\underbrace{\boxed{\frac{|\mu_v - \mu_p|}{\max(\mu_v, \mu_p)}} + \boxed{\frac{|\sigma_v^2 - \sigma_p^2|}{\max(\sigma_v^2, \sigma_p^2)}}}_{\text{Photometric}}\Big] + (1-\lambda)\Big[\underbrace{\boxed{\frac{a_v}{a_p}} + \boxed{H_v}}_{\text{Geometric}}\Big]$$

- An outstanding region among siblings?
  - Brighter/darker?
  - Noisier /more homogenous
  - Larger/Smaller
  - Higher/lower entropy on boundary shape
- Empirical result: best λ =0.5



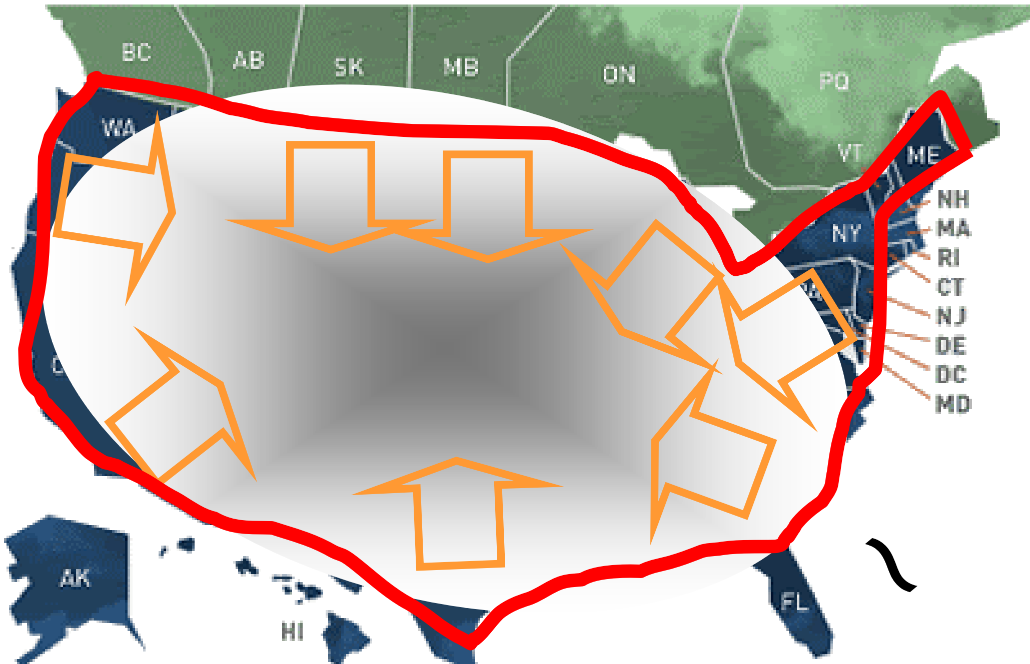
Salience Contract Flow(microview)

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



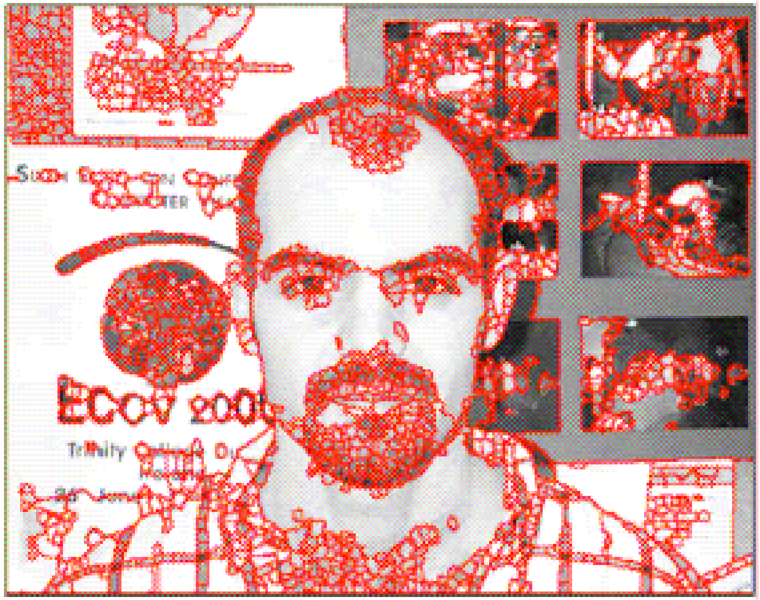
Salience Contract Flow(macroview)

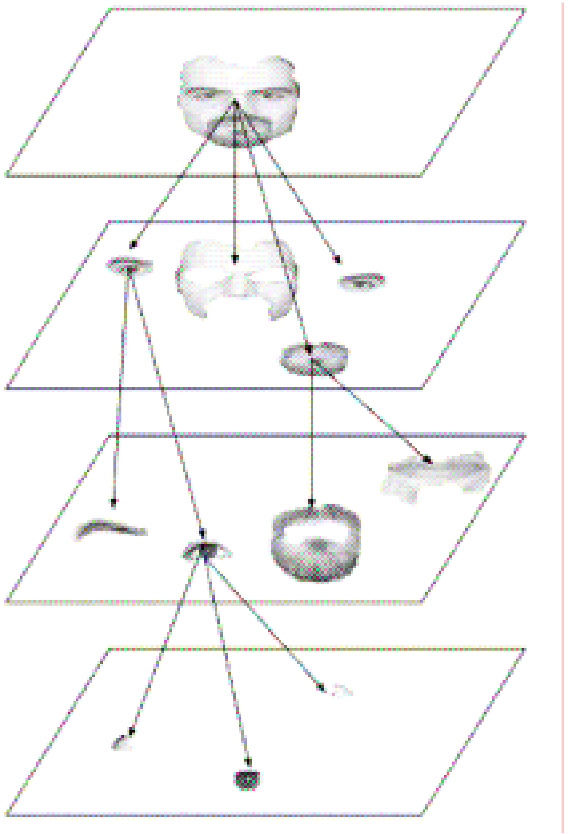
Match salience contract flow



©$_1$ ¼ ©$_2$

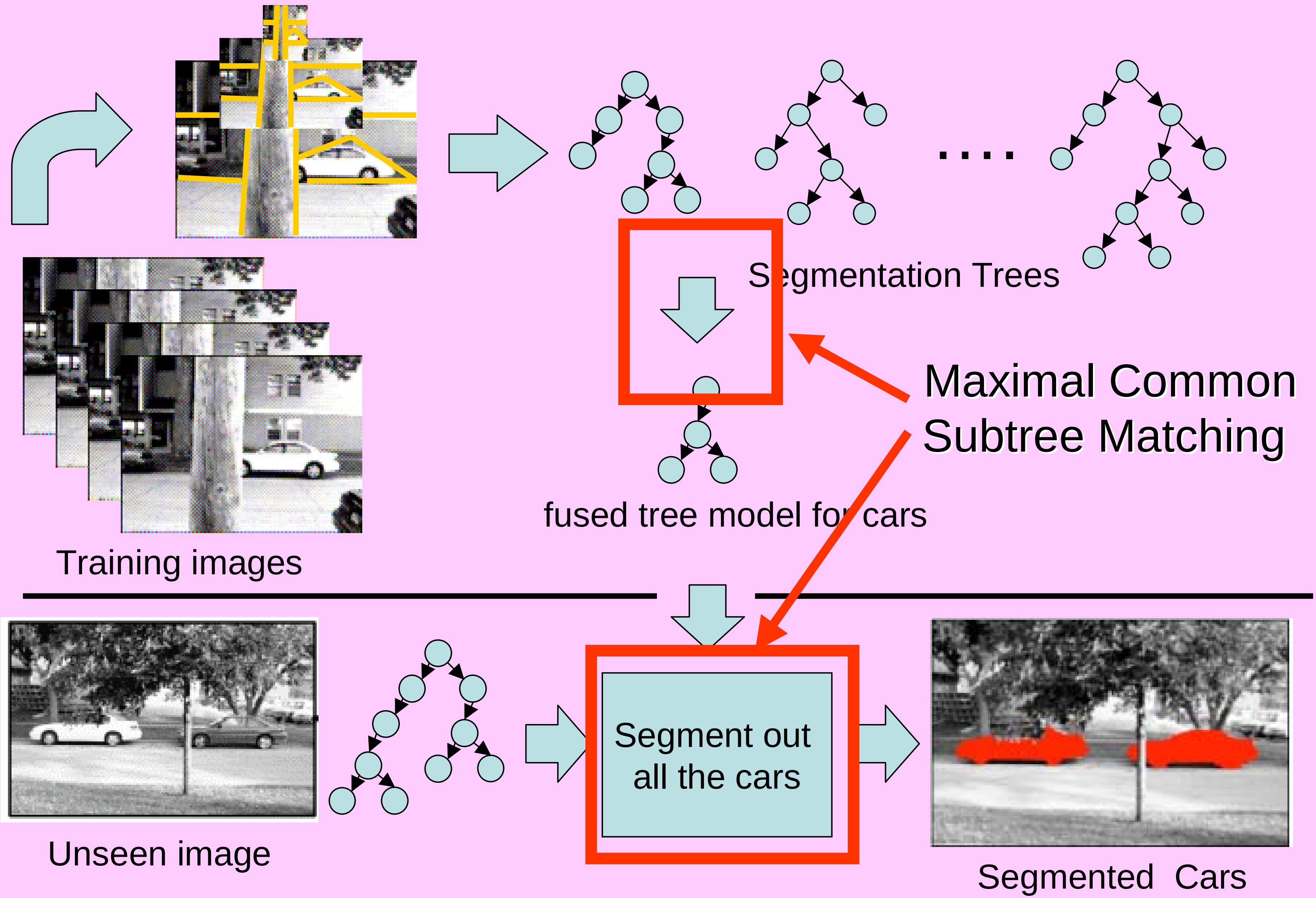
Store Regional Descriptor on Treenode





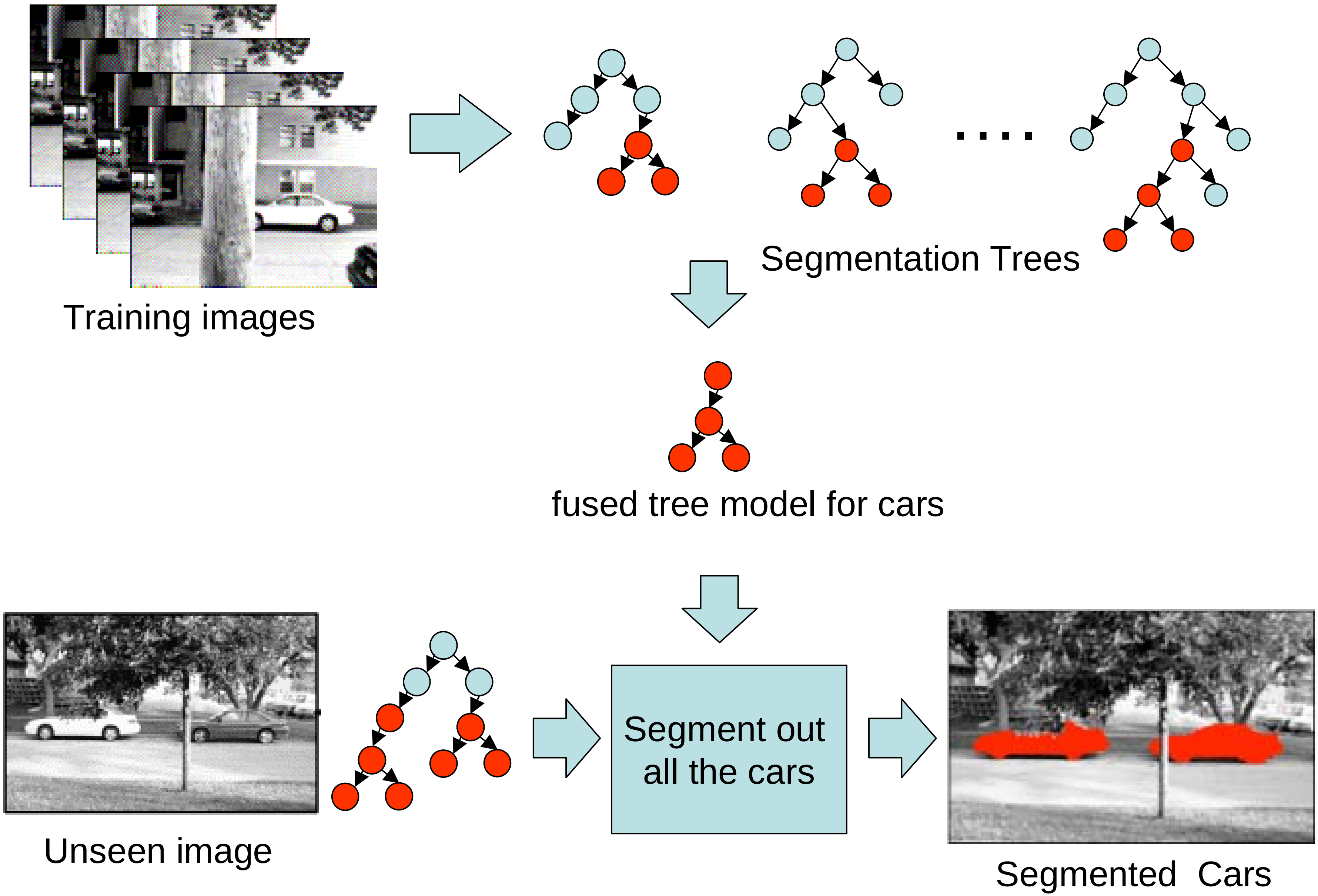
$$\Omega_v = [\mu_v, \sigma_v^2, a_v, x_v, y_v, h_v(1), \ldots, h_v(K), \overrightarrow{\Phi}_v]$$

Photometric	Geometric	Salient

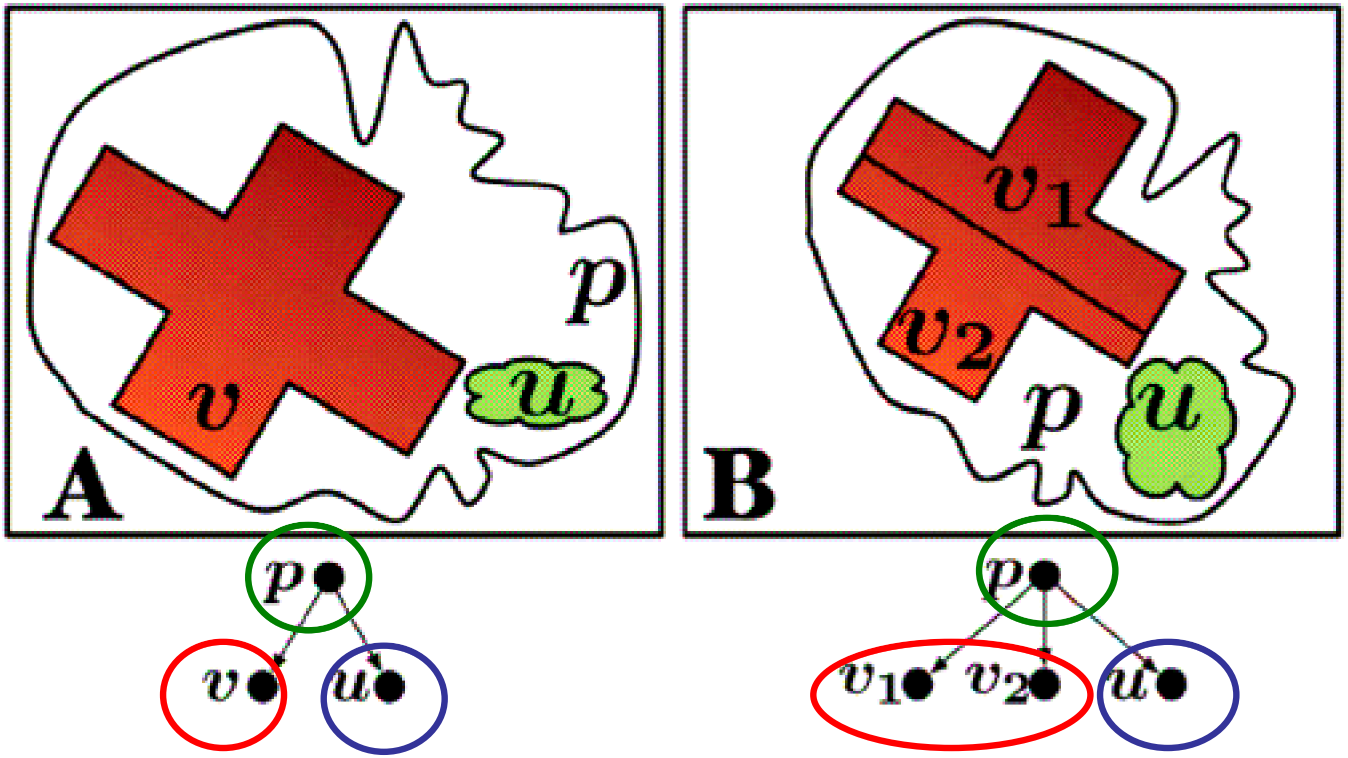


Segmentation Trees
....
Maximal Common
Subtree Matching
fused tree model for cars
Training images
Segment out
all the cars
Unseen image
Segmented Cars

How does it works?



Inexact Matching: Structural Noise

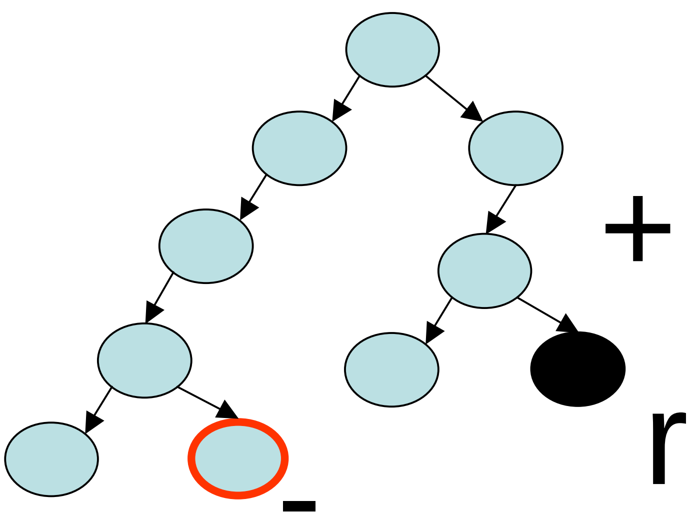


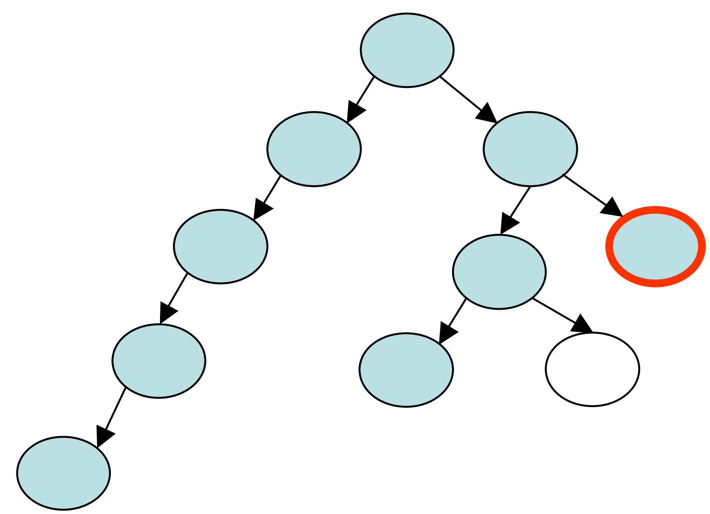
Allow: (1) one-to-one, (2) many-to-one, (3) many-to-many node correspondences

Use tree edit distance instead

Tree Edit Distance

- Editor Operations : costs ~ Dissimilarity(x,y)
  - Remove a node
  - Add a node
  - Replace a node





Metaphor: String Edit Distance

- Unifying Editor Operations
  - Remove a node
  - Add a node → (removal on partner)
  - Replace a node → (paired removal on both string/tree)

AABBBBCC ⟺ AABBYBBCC

Edit : Add Y

Edit : Remove Y

AABBXBBCC ⟺ AABBYBBCC

Edit : Replace X with Y

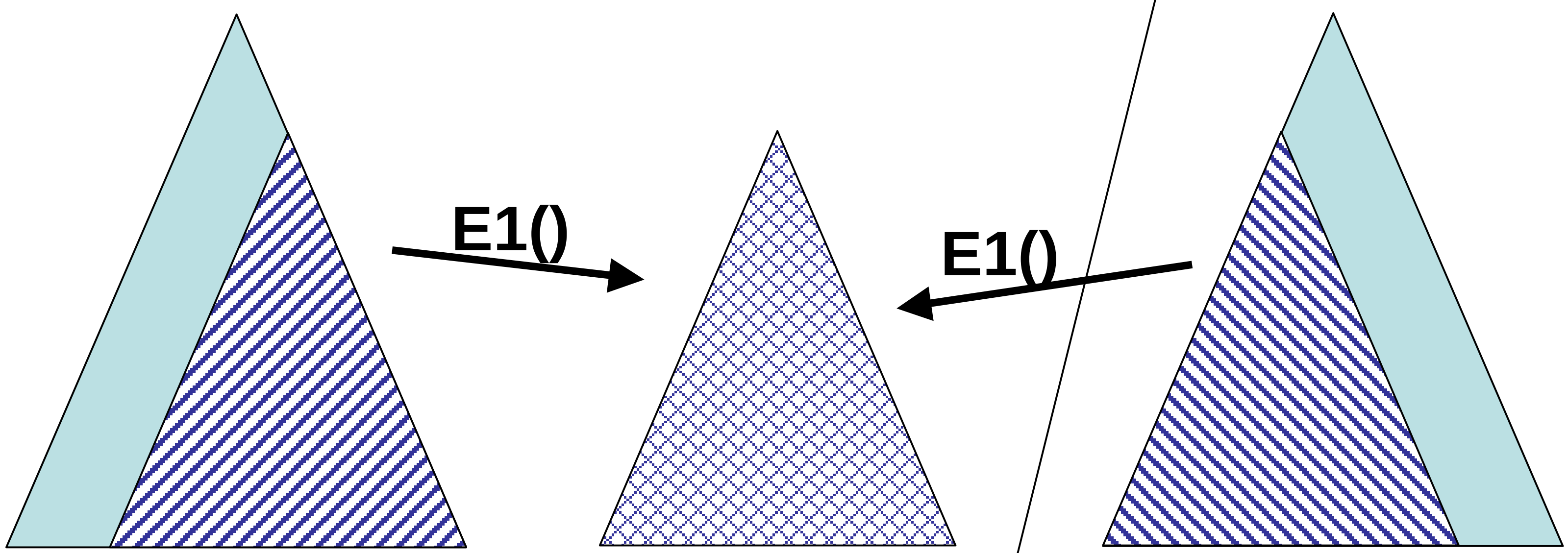
Edit : Remove X

Edit : Remove Y

Tree Edit Distance

- Editor Operation (with costs)
  - Two way removal only

E():Sequence of removal

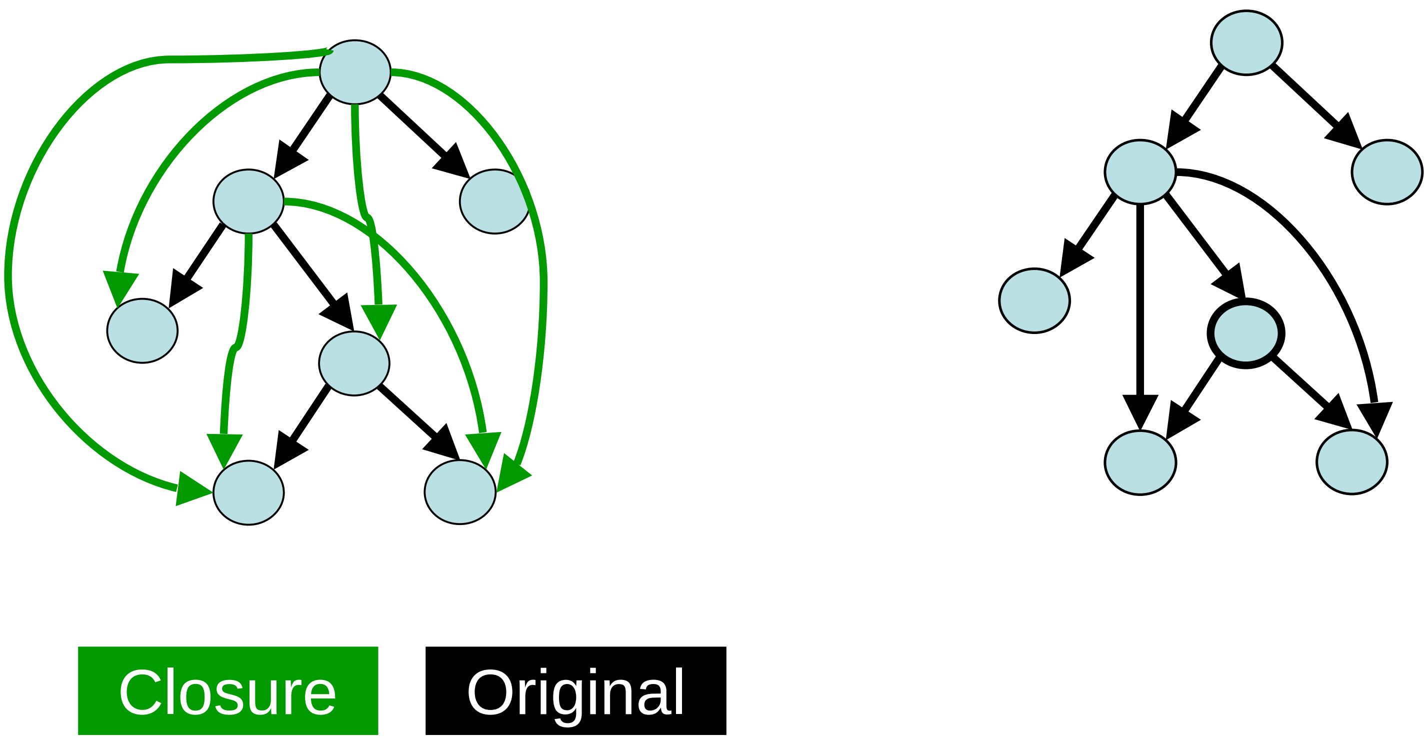


t $\quad u=E1(t) \cap E2(t')$ $\quad$ t’

$$\text{Dist.}(t, t') = \text{Dist.}(t, u) + \text{Dist}(u, t')$$

Reduce Edit-Distance matching to Non-edit matching

- Transitive Closure
- (see animation)



Matching Criteria

GIVEN two trees: $t,\ t'$

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

which MAXIMIZES the QUALITY OF MATCH

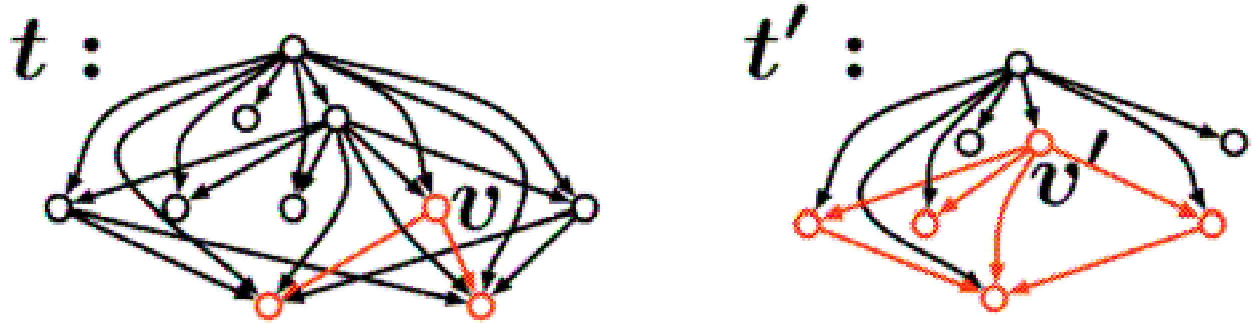
$$\mathcal{U}(t, t') = \sum_{(v,v') \in f} [w_v + w_{v'} - m_{vv'}]$$

node saliency

cost of node matching

while PRESERVING ancestor-descendant relationships

Divide and Conquer



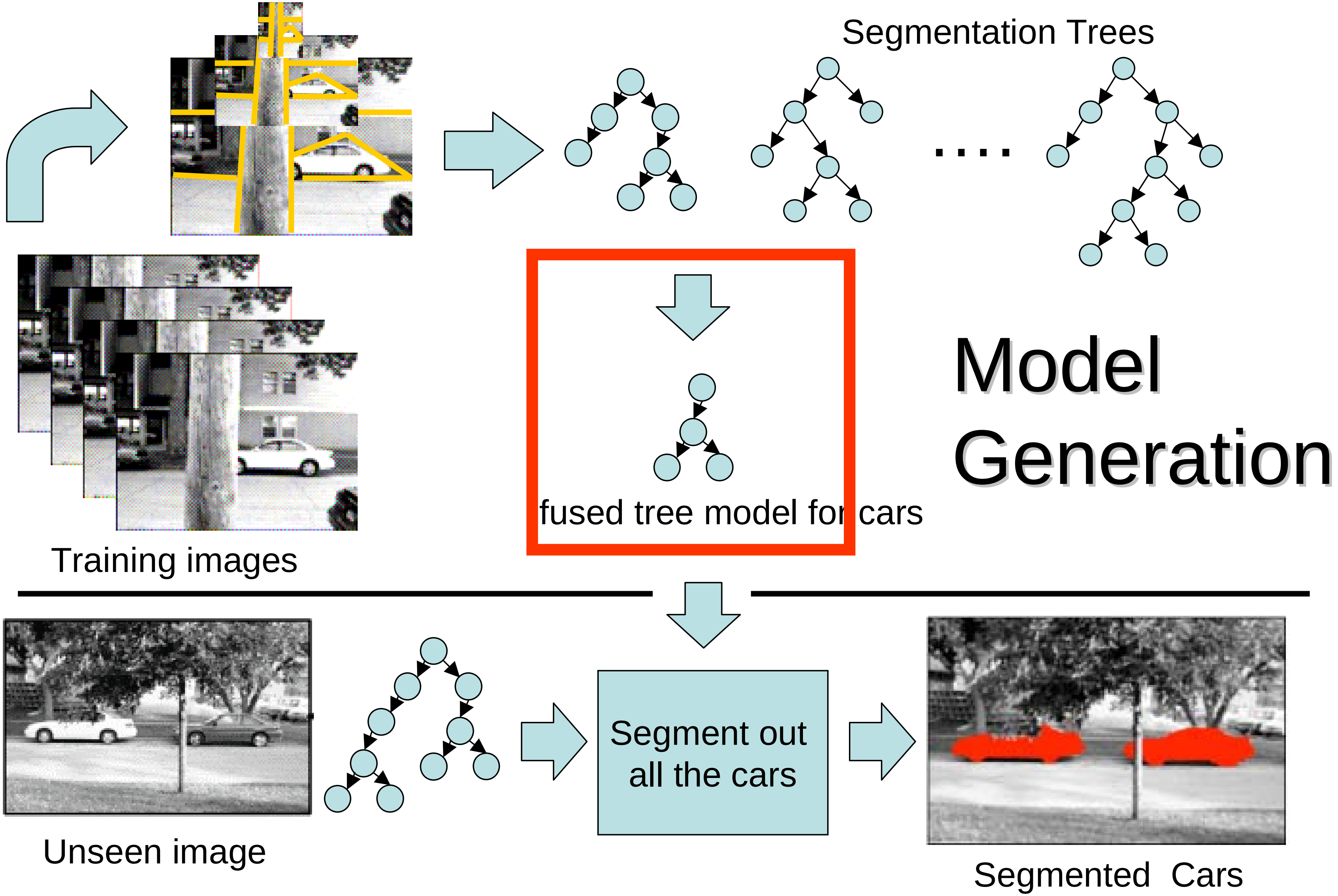
$$\mathcal{U}(t_v, {t'}_{v'}) = w_v + w_{v'} - m_{vv'} + \max_{\mathcal{C}_{vv'}} \sum_{(d,d') \in \mathcal{C}_{vv'}} \mathcal{U}(d, d')$$

Maximum clique over all descendant pairs

descendants

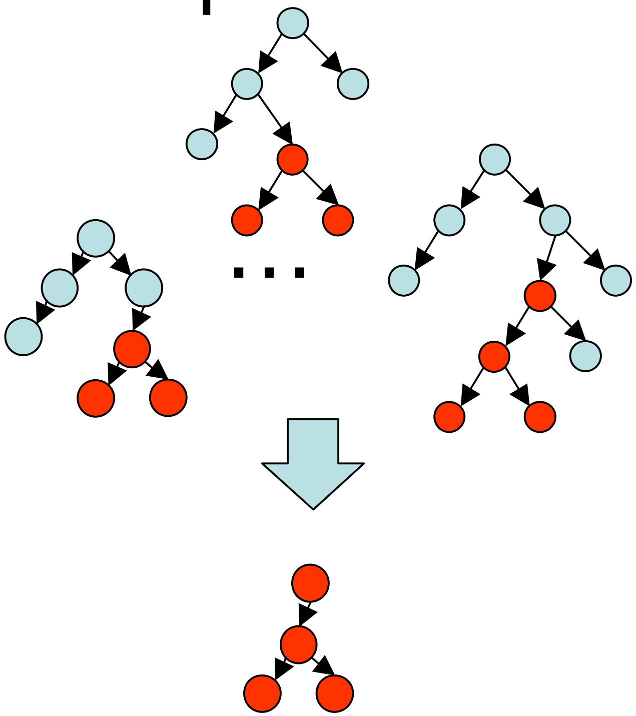
NP-complete → QP approx. O(|C$_{vv'}$|)

Try all pairs of (v, v') combinations = O(|t| + |t'|)


Segmentation Trees
....
Model Generation
fused tree model for cars
Training images
Unseen image
Segment out all the cars
Segmented Cars

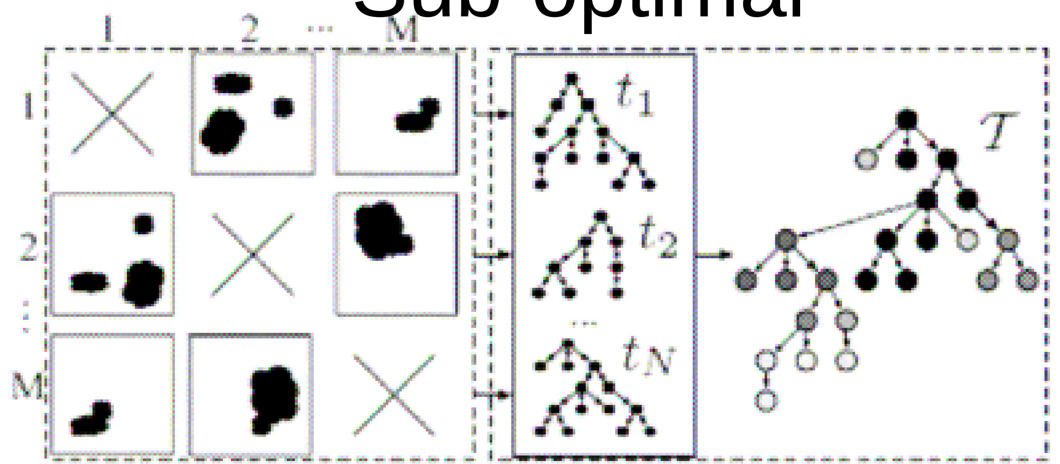
Model: Union of Subtrees

Optimal

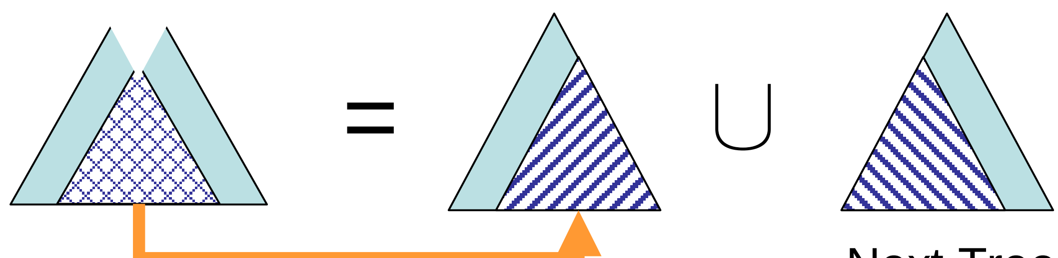


NP-Hard

Sub-optimal



1. Pairwise matching
2. One by one union



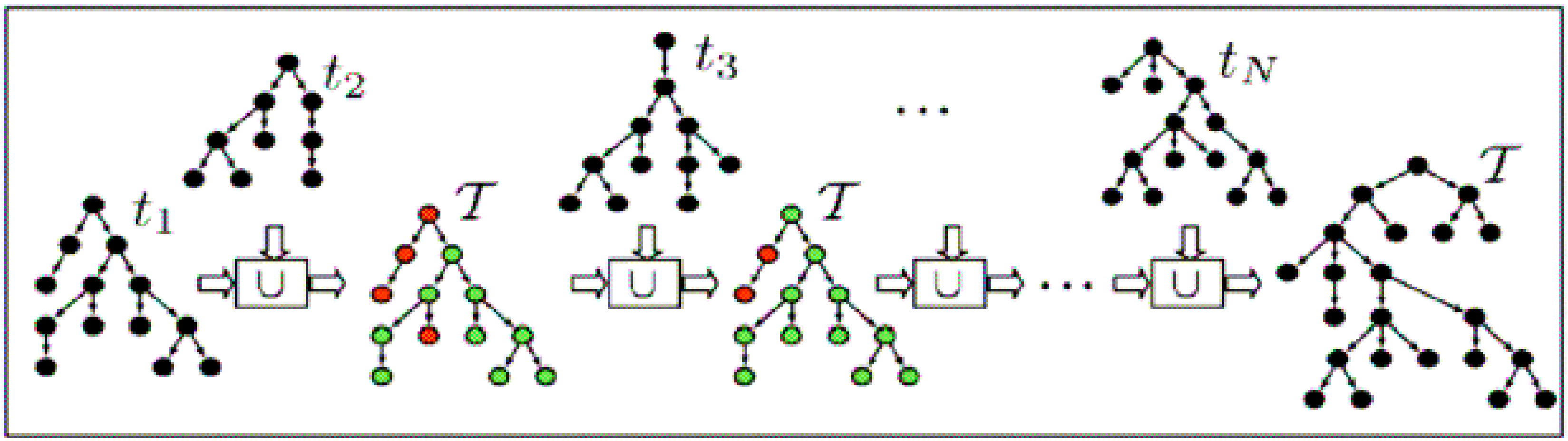
Next Tree

$T = T \cup T_{next}$

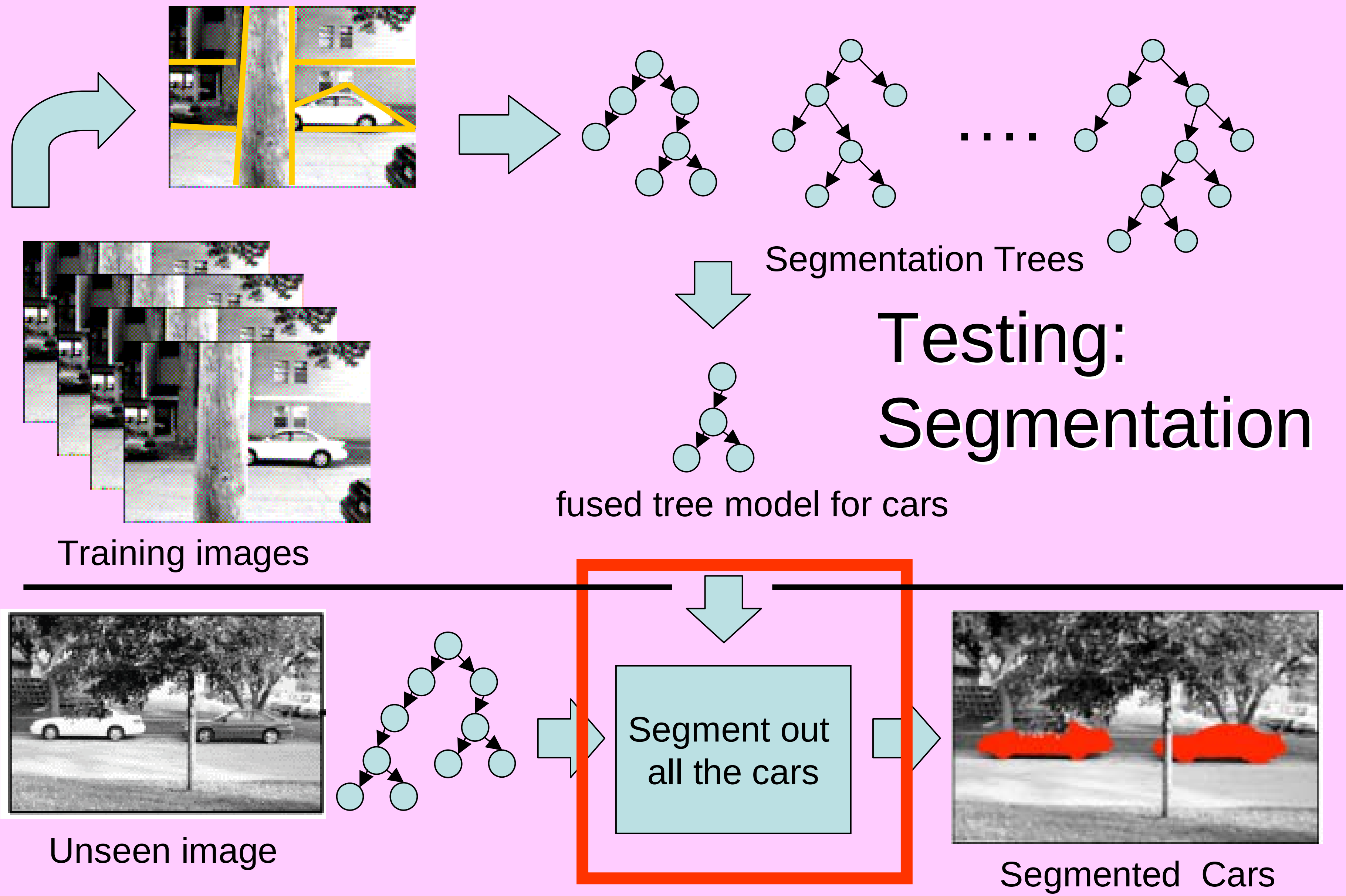
Category Model

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

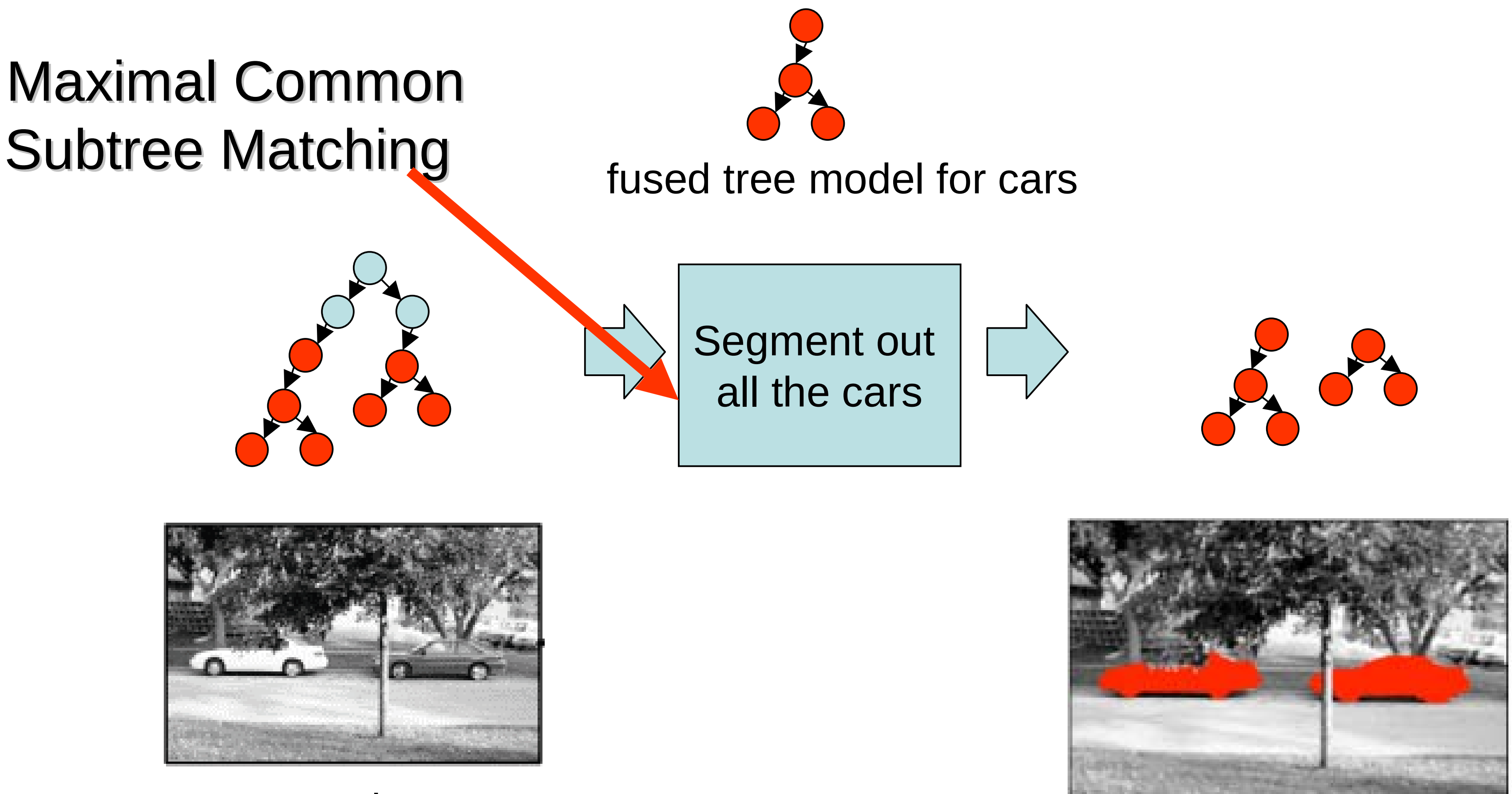
$$\mathcal{T} = \tau \cup t_i \setminus \tau \cup t_{i+1} \setminus \tau$$



Testing: Segmentation



Testing: Detect & Segmentation

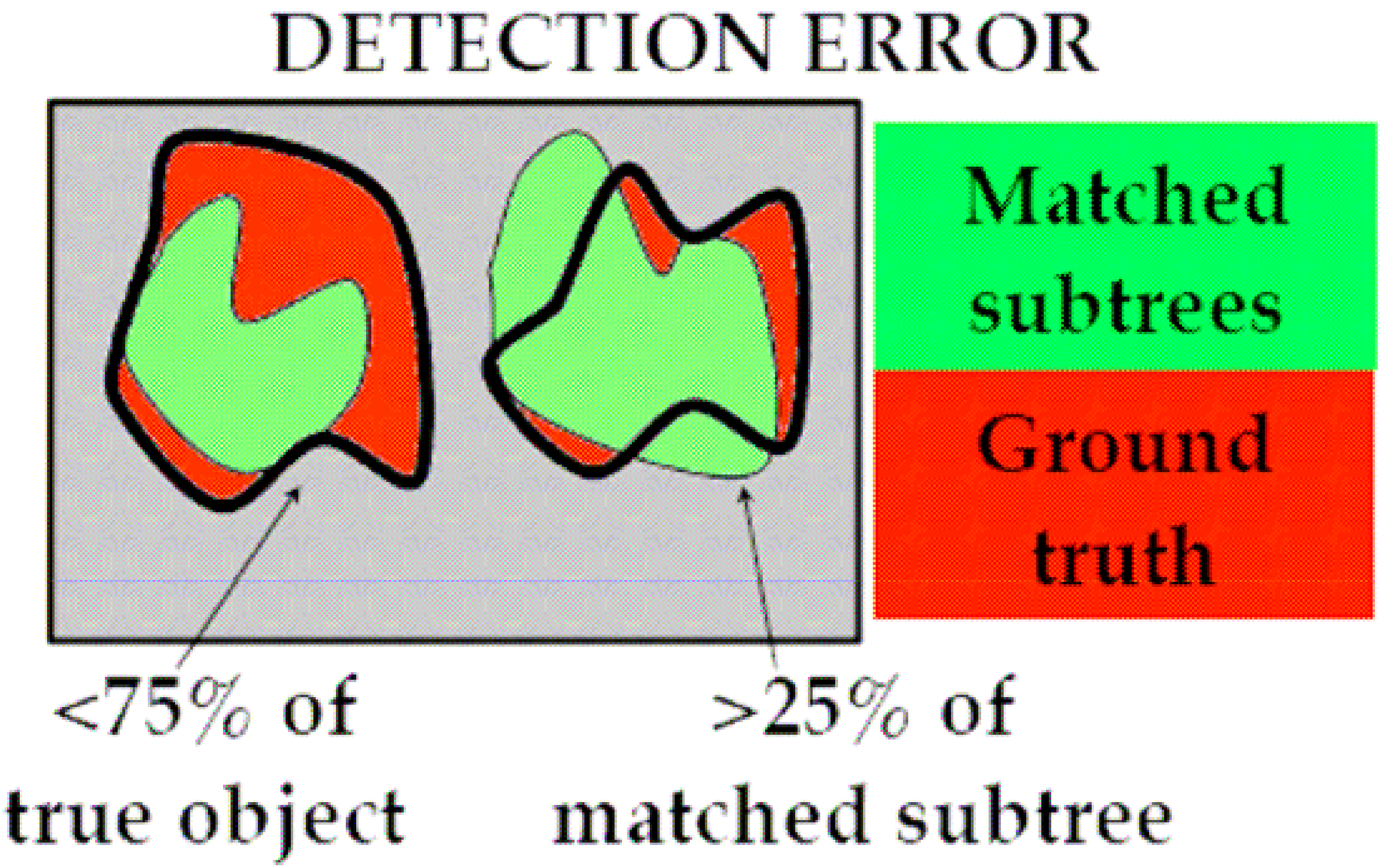


Unseen image

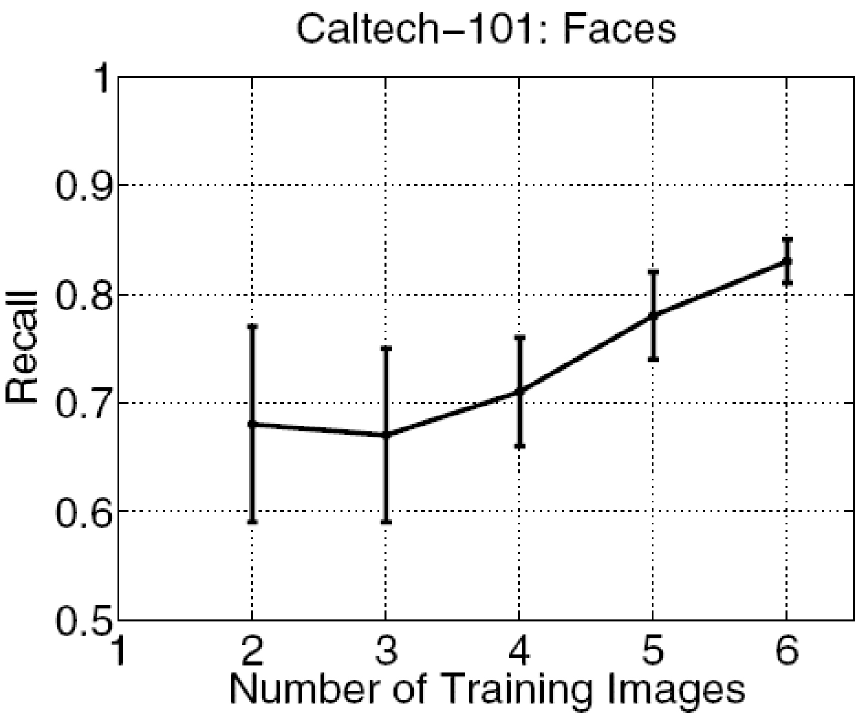
Segmented Cars

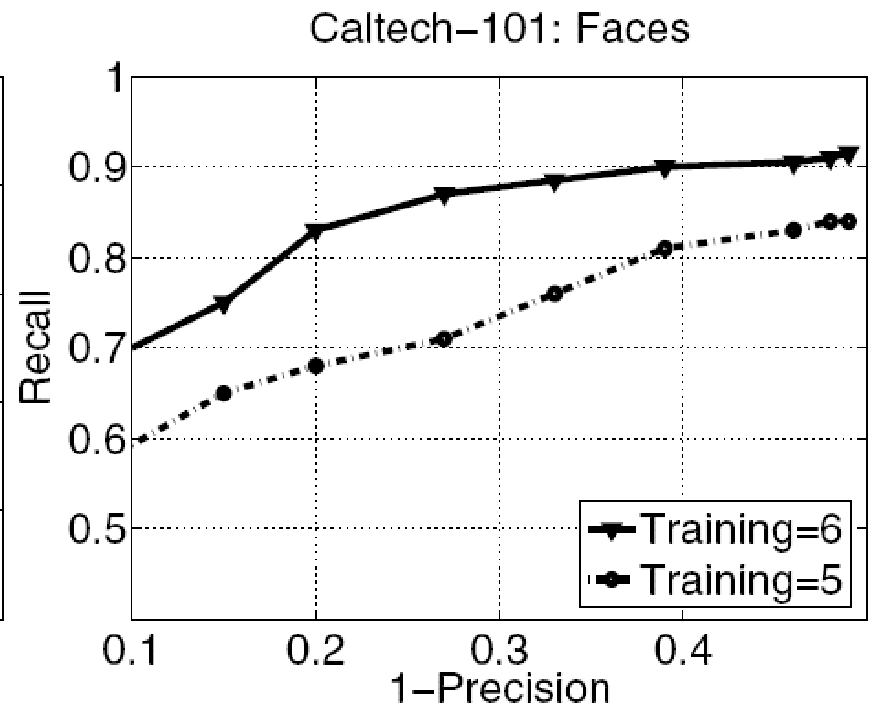
Match : (Similarity > Thresh) → (precision/recall)

Performace Evaluaton



Results (Caltech 101 Face)

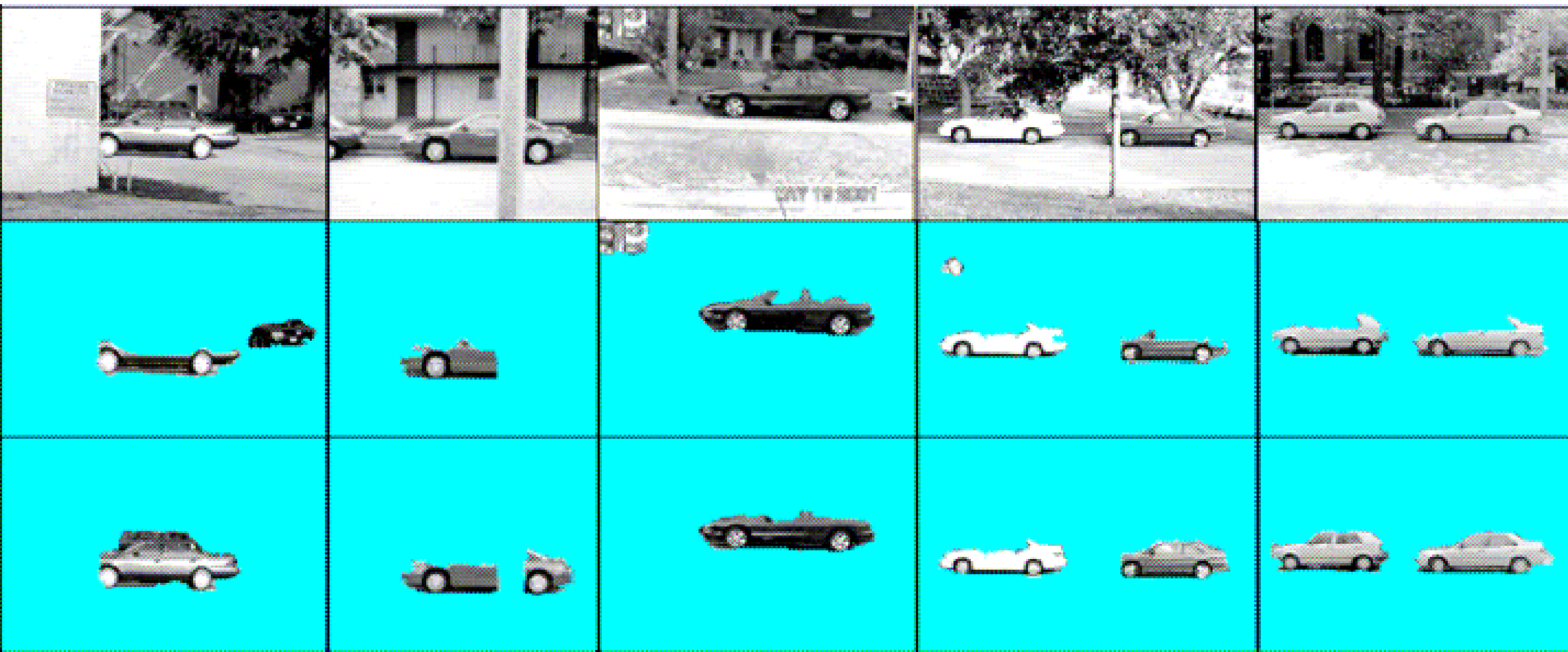




Varying Matching Thresh. →(precision/recall)

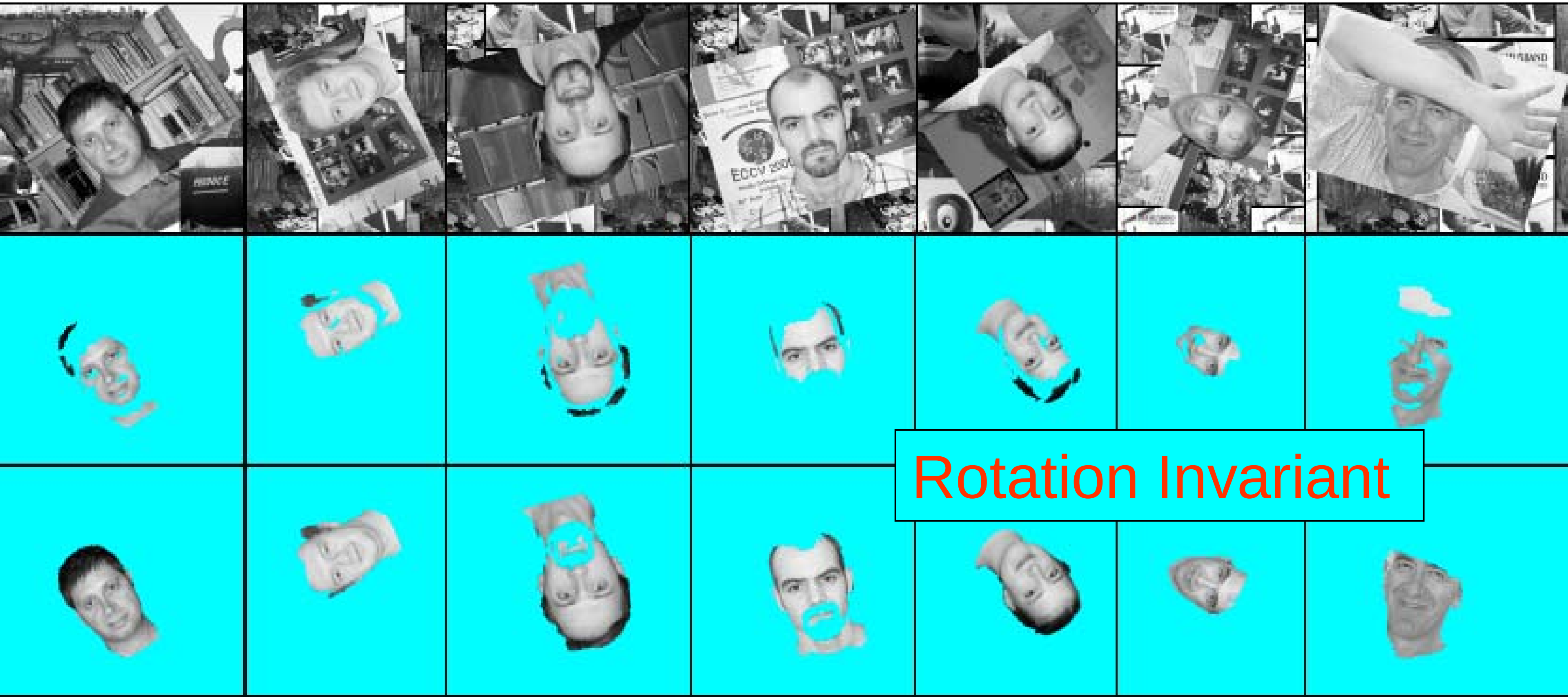
Results (UIUC Car Side View)

#positive/#training: 5/10 vs 10/20(2hr on P4-2.4G/2G)

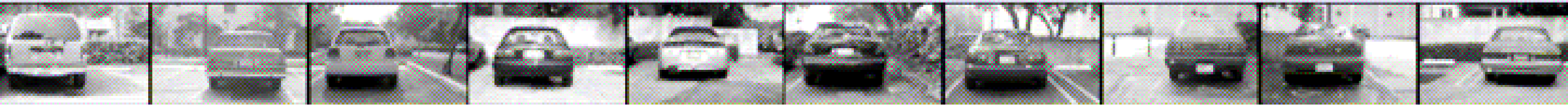


Results (Caltech 101 Face)

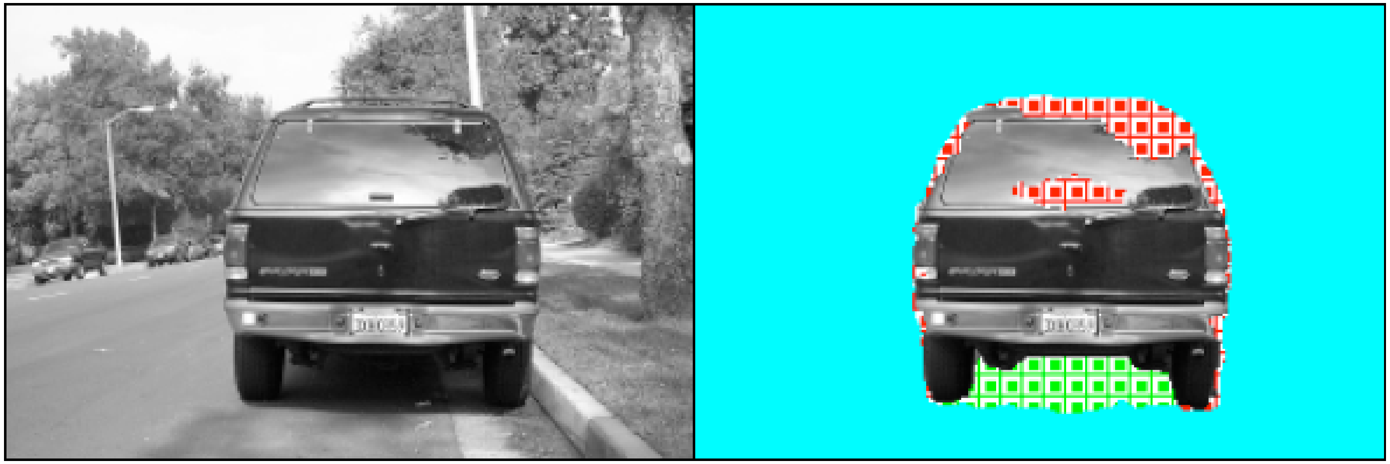
#positive/#training: 3/6 vs 6/12



Caltech (Cars Rear View)



#positive/#training: 10/20



Conclusion

- Contribution
  - Good Image Representation → Seg. Tree
- Small amount of training data
  - Cf. Statistical Learning/Clustering
    - Ex. Visual Words + pLSA
- Allow Non-category Images noise
- Allow occlusion (disconnected regions)

Region Descriptor

	Photmetric	Geometric	Topological
Graylevel	x		
Region Area		x	
Sliced area histogram		x	
Salient Flow	x	x	x
Annotated Recursive Tree		x	x

Thank you

- Quicktopic

Cf. Visual Words+pLSA

Given a large collection of unlabeled images:

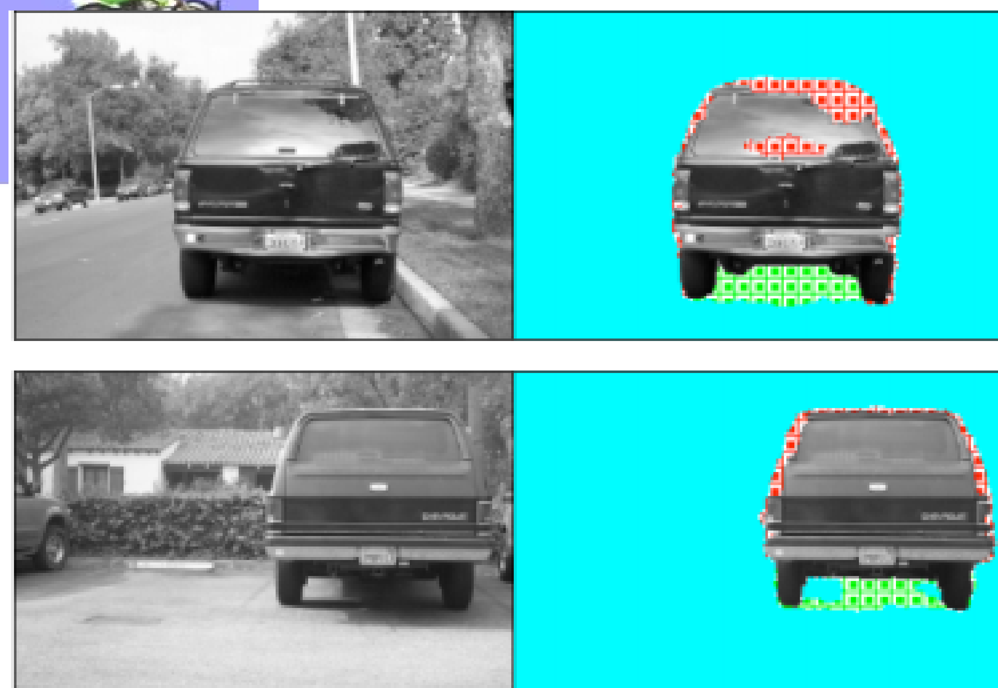
1. For each image, compute multiple candidate segmentations using Normalized-Cuts.
2. For each segment, compute histograms of visual words.
3. Perform topic discovery, treating each segment as a document, using LDA over all segments in the collection.
4. For each topic sort segments using KL divergence.

- Visual Words recognize connected object only
- Tree Matching is more conservative due to intersection

Cf. Visual Words +pLSA

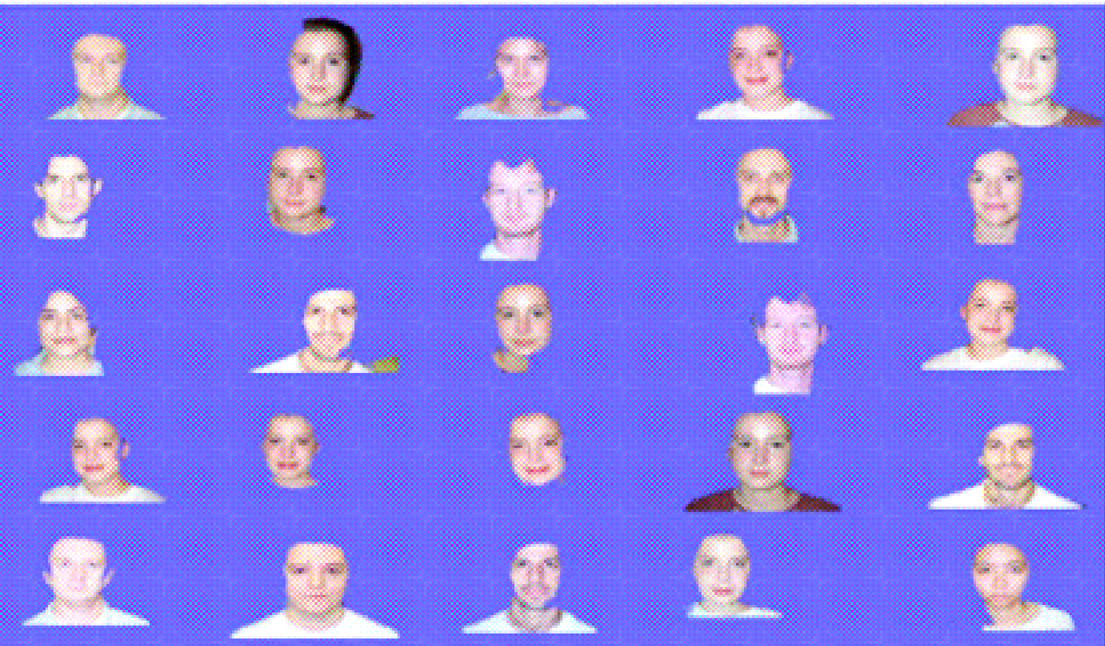
Visual Words/pLSA

Tree matching



Caltech Faces

Visual Words/pLSA



Tree matching

ReSPEC(Use Color Histogram)