Recognition by Association

ask not "What is this?"
but "What is it ***like***?"

Tomasz Malisiewicz
joint work with Alyosha Efros
May 12, 2008
CMU VASC Seminar

Carnegie Mellon
THE ROBOTICS INSTITUTE

Understanding an Image

slide by Fei Fei, Fergus & Torralba

Object naming

Object naming / Object categorization

Object naming / Object categorization

sky

building

flag

face

banner

wall

street lamp

bus

bus

cars

Classical View of Categories

- Dates back to Plato & Aristotle
  - Categories are defined by a list of properties shared by all elements in a category
  - Category membership is binary
  - Every member in the category is equally the same

Classical View of Categories

- Dates back to Plato & Aristotle
  - Categories are defined by a list of properties shared by all elements in a category
  - Category membership is binary
  - Every member in the category is equally the same
- Humans don't do this!
  - People don't rely on abstract definitions (Rosch 1973)
  - Is an olive a fruit? Are curtains furniture?
  - Different cultures have different categories
    - e.g. "Women, Fire, and Dangerous Things" category is Australian aboriginal language (Lakoff 1987)

Categorization in Psychology

- Prototype Theory (Rosch 1973)
  - Single summary representation (prototype) for each category
  - Humans compute similarity between input and prototypes

Categorization in Psychology

- Exemplar Theory (Medin & Schaffer 1978, Nosofsky 1986, Krushke 1992)
  - categories represented in terms of remembered objects (exemplars)
  - Similarity is measured between input and all exemplars
  - *think* non-parametric density estimation

Problems with Visual Categorization

- Categorization is anchored on <u>words</u>
- Words don’t always correspond to visual phenomena
- Visual Polysemy
  - Same category, different visual properties
- Visual Synonyms
  - Same object, multiple correct categories

Visual Polysemy

- A lot of categories are functional

Chair

Visual Polysemy

Chair

- A lot of categories are functional

Car

- Different views of same object can be visually dis-similar

Visual Synonyms

- Multiple levels of categories

Asphalt

Road

Visual Synonyms

- Multiple levels of categories

Asphalt

Road

- Multiple good category names

Player 1: purse

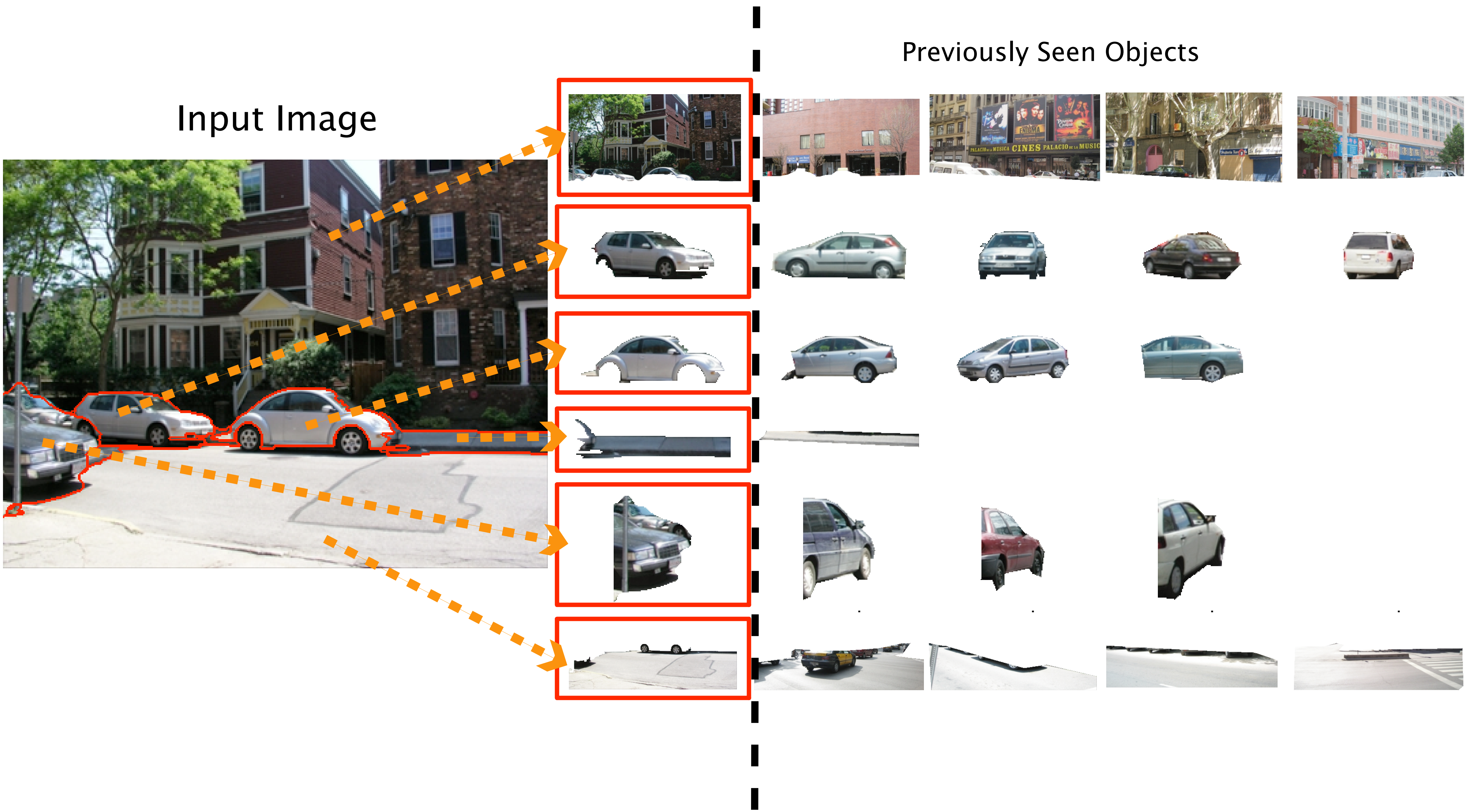
Player 2: handbag

*Luis von Ahn's ESP Game

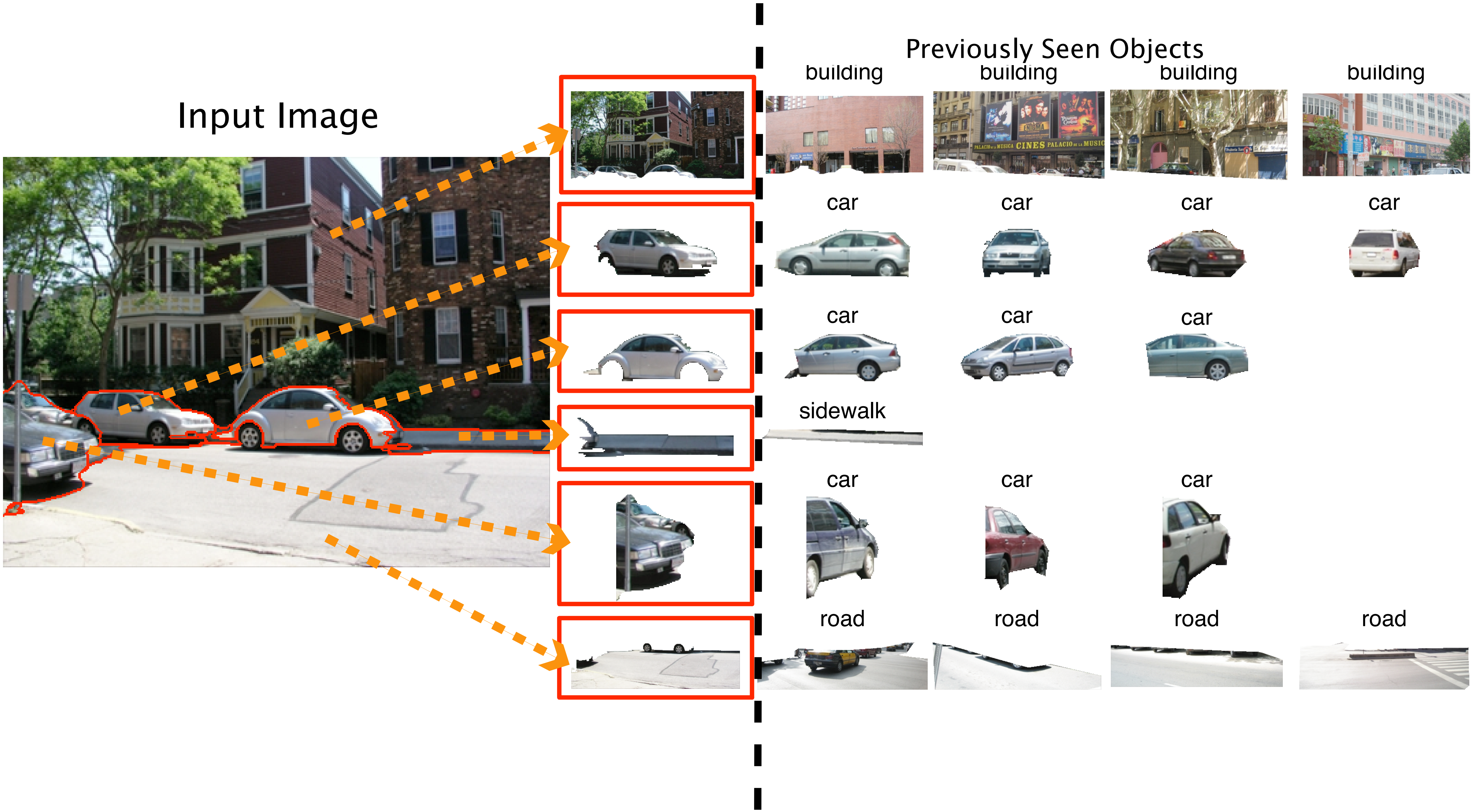
Different way of looking at recognition

Input Image

Different way of looking at recognition



Different way of looking at recognition



Our Contributions

- Posing Recognition as Association
  - Use large number of object exemplars

Our Contributions

- Posing Recognition as Association
  - Use large number of object exemplars

- Learning Object Similarity
  - Different distance function per exemplar

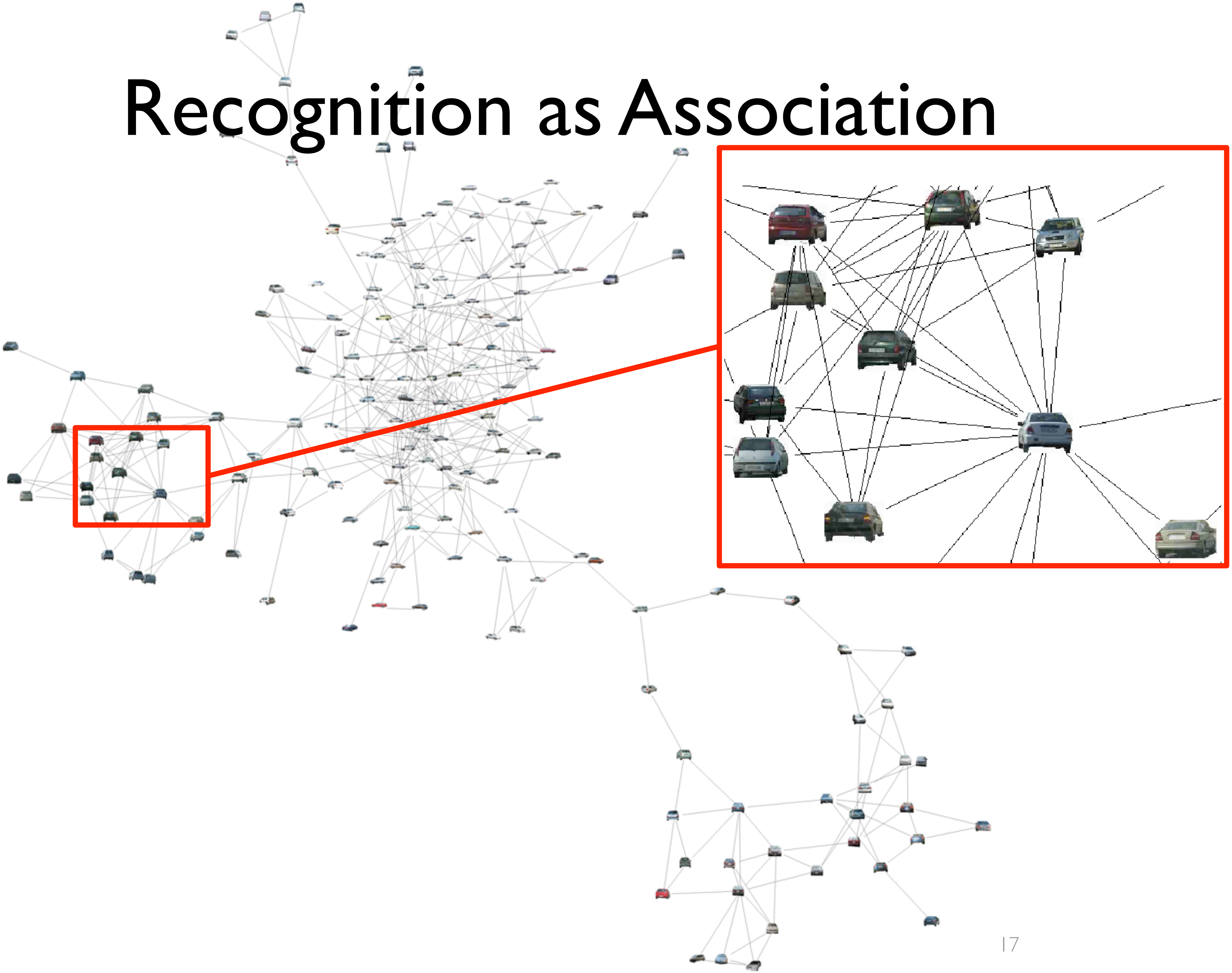
Our Contributions

- Posing Recognition as Association
  - Use large number of object exemplars

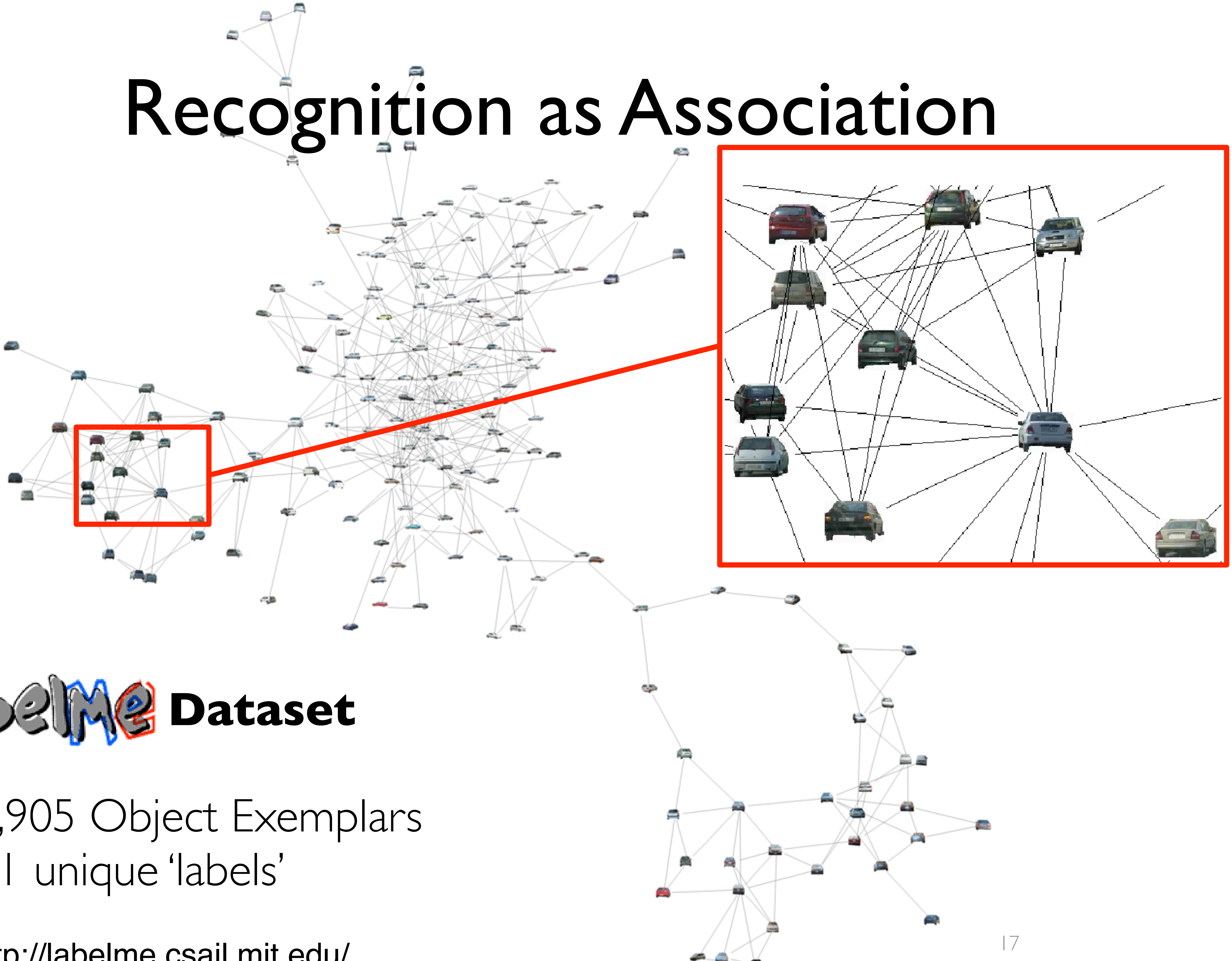
- Learning Object Similarity
  - Different distance function per exemplar

- Recognition-Based Object Segmentation
  - Use multiple segmentation approach

Recognition as Association



Recognition as Association



Dataset

12,905 Object Exemplars
171 unique 'labels'

http://labelme.csail.mit.edu/

Measuring Similarity

- How are objects similar?

Measuring Similarity

- How are objects similar?


Measuring Similarity

- How are objects similar?

Measuring Similarity

- How are objects similar?

Measuring Similarity

- How are objects similar?

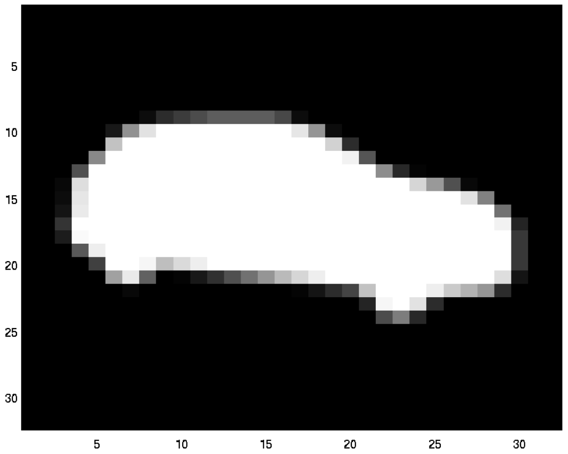
Exemplar Representation

Segment from LabelMe

Shape

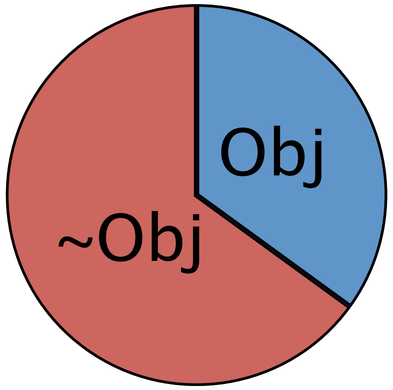
Type	Name	Dimension
Shape	Centered Mask	32x32=1024
	BB Extent	2
	Pixel Area	1
Texture	Right Boundary Tex-Hist	100
	Top Boundary Tex-Hist	100
	Left Boundary Tex-Hist	100
	Bottom Boundary Tex-Hist	100
	Interior Tex-Hist	100
Color	Mean Color	3
	Color std	3
	Color Histogram	33
Location	Absolute Mask	8x8=64
	Top Height	1
	Bot Height	1

Centered Mask



Bounding Box Dimensions

Pixel Area



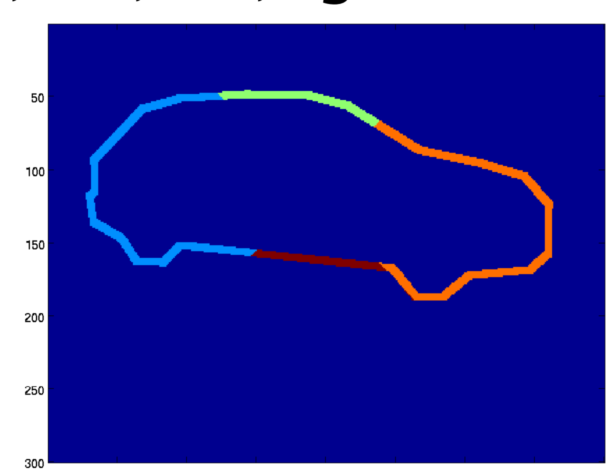
Texture

Type	Name	Dimension
Shape	Centered Mask	32x32=1024
	BB Extent	2
	Pixel Area	1
Texture	Right Boundary Tex-Hist	100
	Top Boundary Tex-Hist	100
	Left Boundary Tex-Hist	100
	Bottom Boundary Tex-Hist	100
	Interior Tex-Hist	100
Color	Mean Color	3
	Color std	3
	Color Histogram	33
Location	Absolute Mask	8x8=64
	Top Height	1
	Bot Height	1

Textons

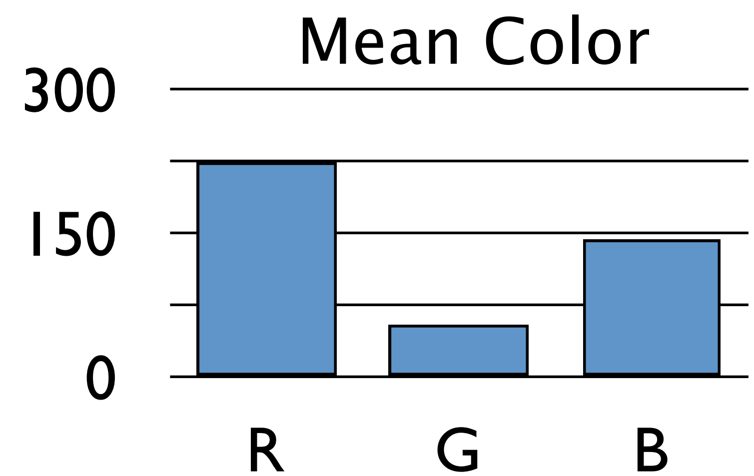
Interior: Bag-of-Words

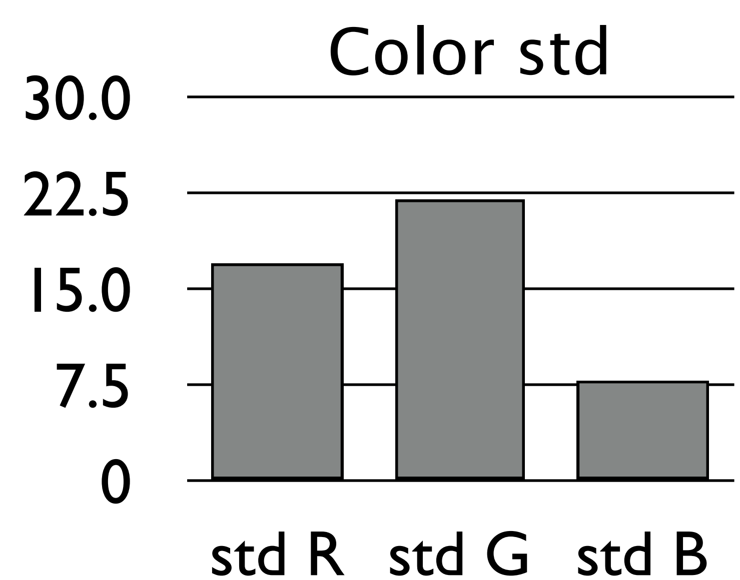
top,bot,left,right boundary

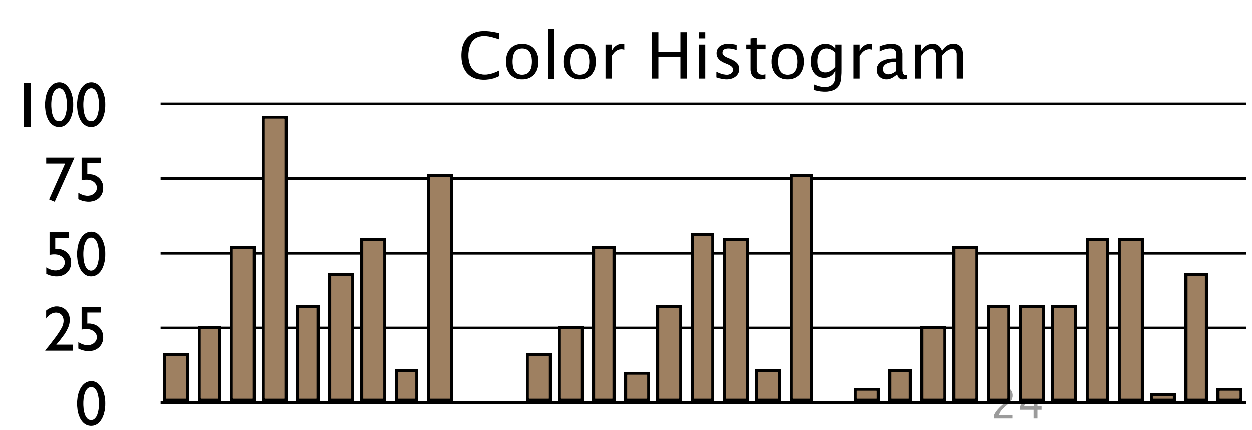


Color

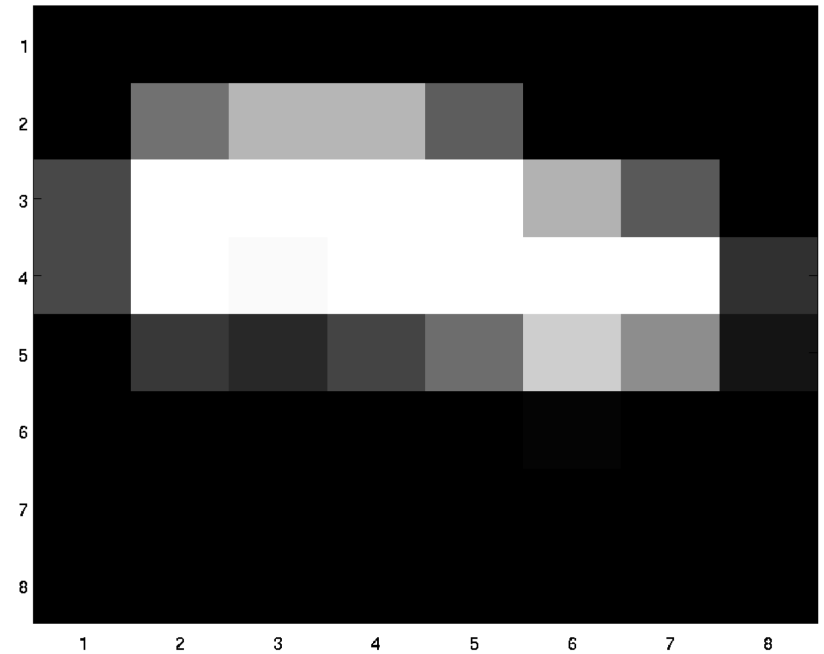
Type	Name	Dimension
Shape	Centered Mask	32x32=1024
	BB Extent	2
	Pixel Area	1
Texture	Right Boundary Tex-Hist	100
	Top Boundary Tex-Hist	100
	Left Boundary Tex-Hist	100
	Bottom Boundary Tex-Hist	100
	Interior Tex-Hist	100
Color	Mean Color	3
	Color std	3
	Color Histogram	33
Location	Absolute Mask	8x8=64
	Top Height	1
	Bot Height	1



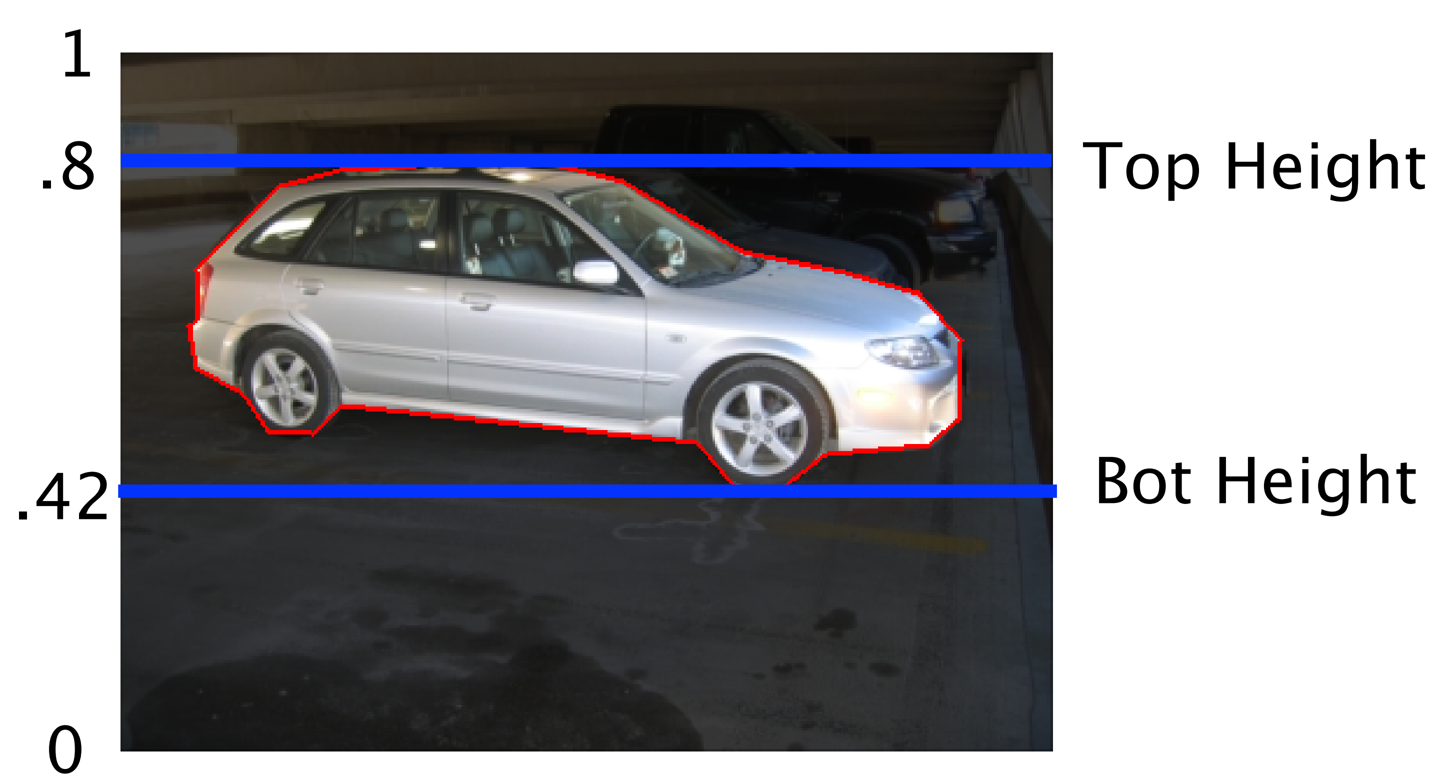




Location



Type	Name	Dimension
Shape	Centered Mask	32x32=1024
	BB Extent	2
	Pixel Area	1
Texture	Right Boundary Tex-Hist	100
	Top Boundary Tex-Hist	100
	Left Boundary Tex-Hist	100
	Bottom Boundary Tex-Hist	100
	Interior Tex-Hist	100
Color	Mean Color	3
	Color std	3
	Color Histogram	33
Location	Absolute Mask	8x8=64
	Top Height	1
	Bot Height	1

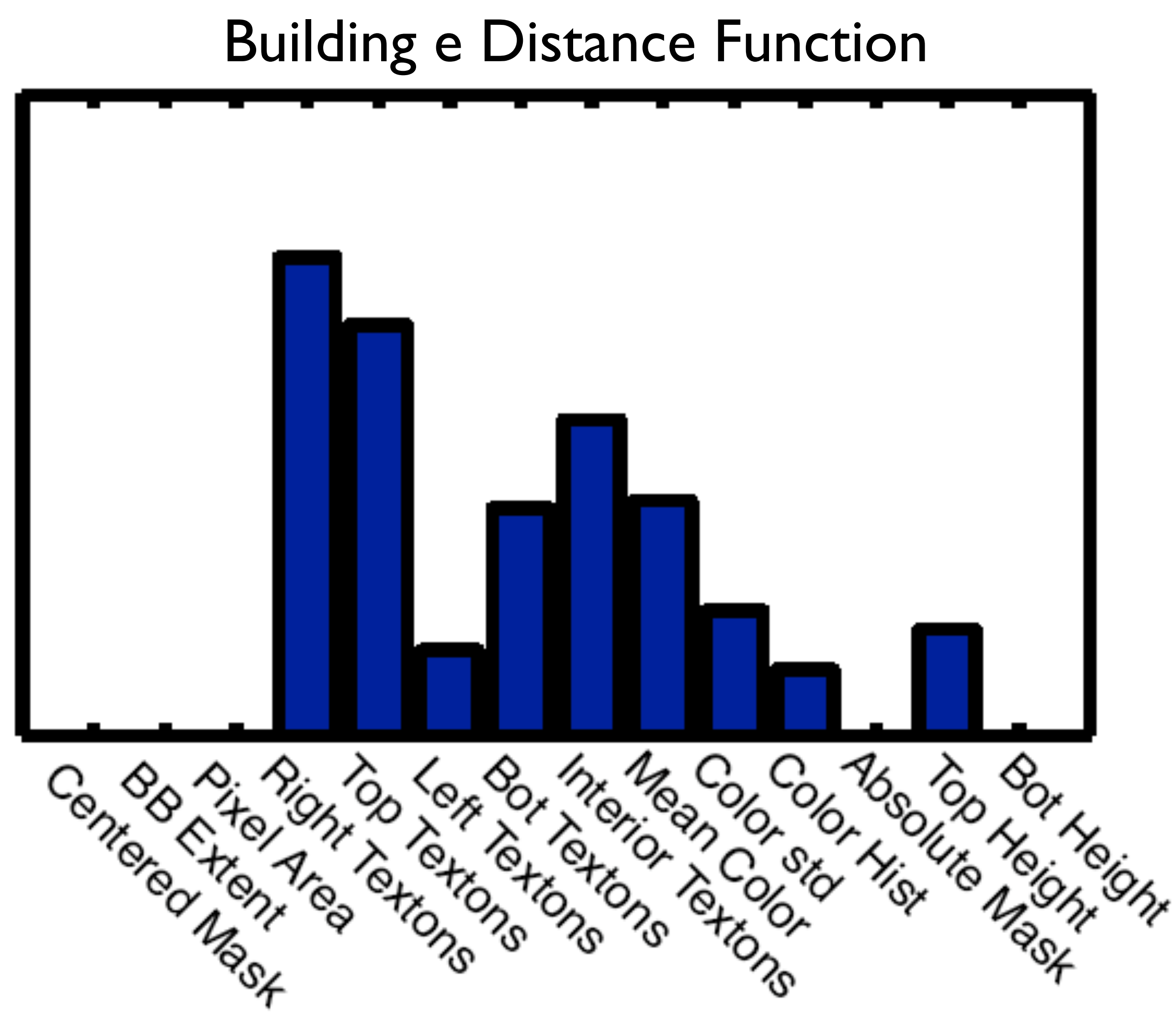


Distance "Similarity" Functions

- Positive Linear Combinations of Elementary Distances Computed Over 14 Features

$$D_e(z) = \mathbf{w}_e \cdot \mathbf{d}_{ez}$$

Building e



Learning Object Similarity

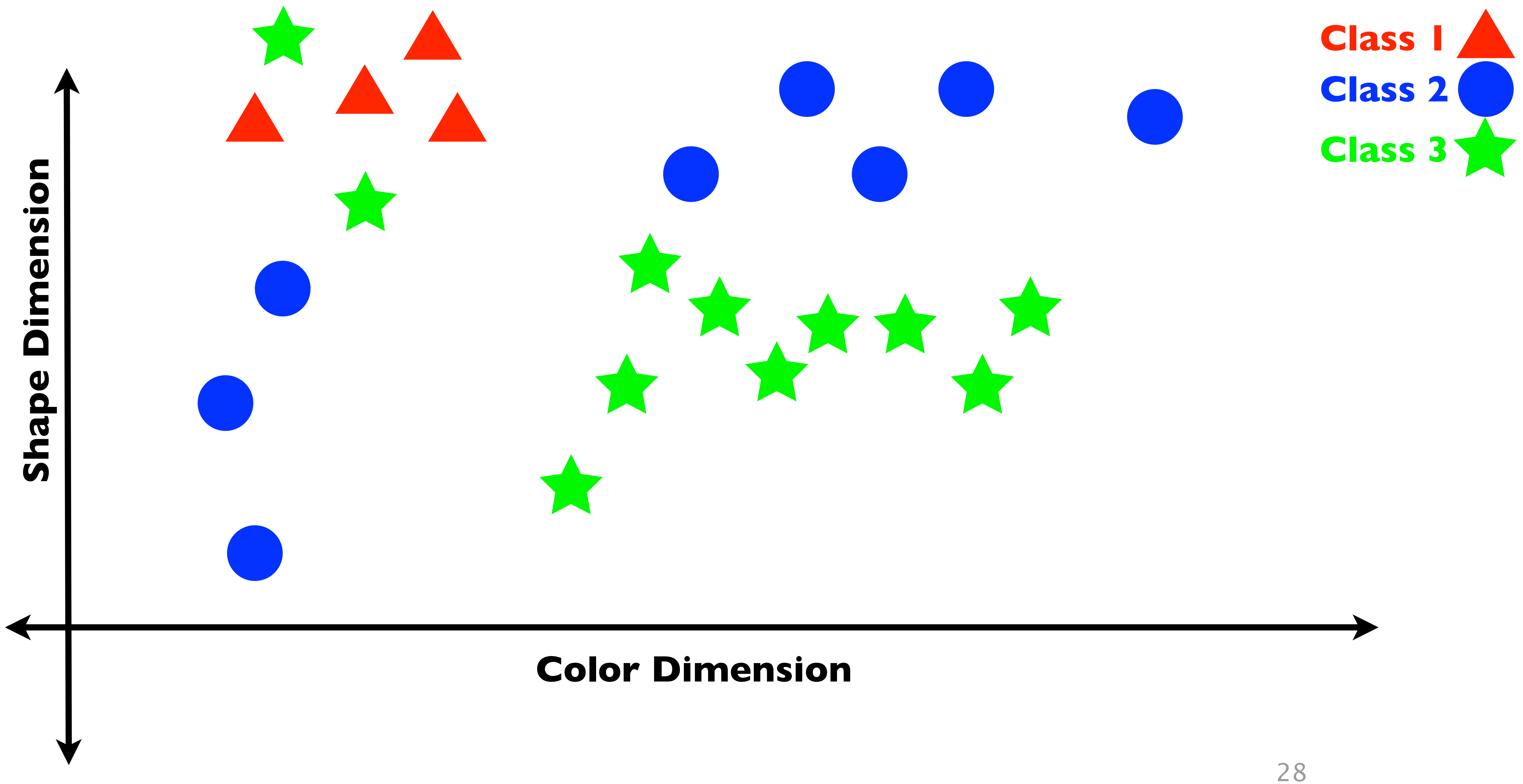
- Learn a different distance function for each exemplar in training set

- Formulation is similar to Frome et al [1,2]

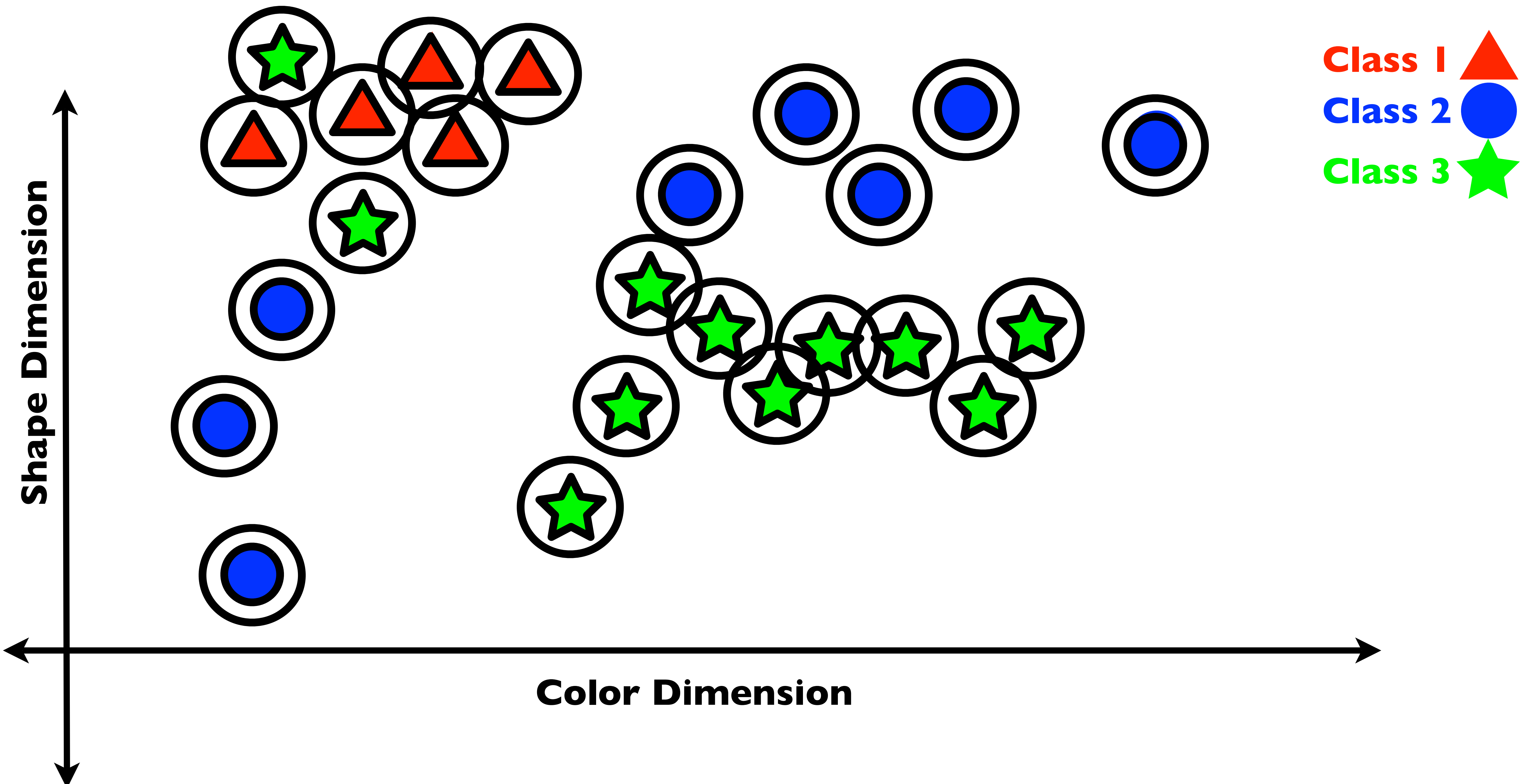
[1] Andrea Frome, Yoram Singer, Jitendra Malik. "Image Retrieval and Recognition Using Local Distance Functions." In NIPS, 2006.

[2] Andrea Frome, Yoram Singer, Fei Sha, Jitendra Malik. "Learning Globally-Consistent Local Distance Functions for Shape-Based Image Retrieval and Classification." In ICCV, 2007.

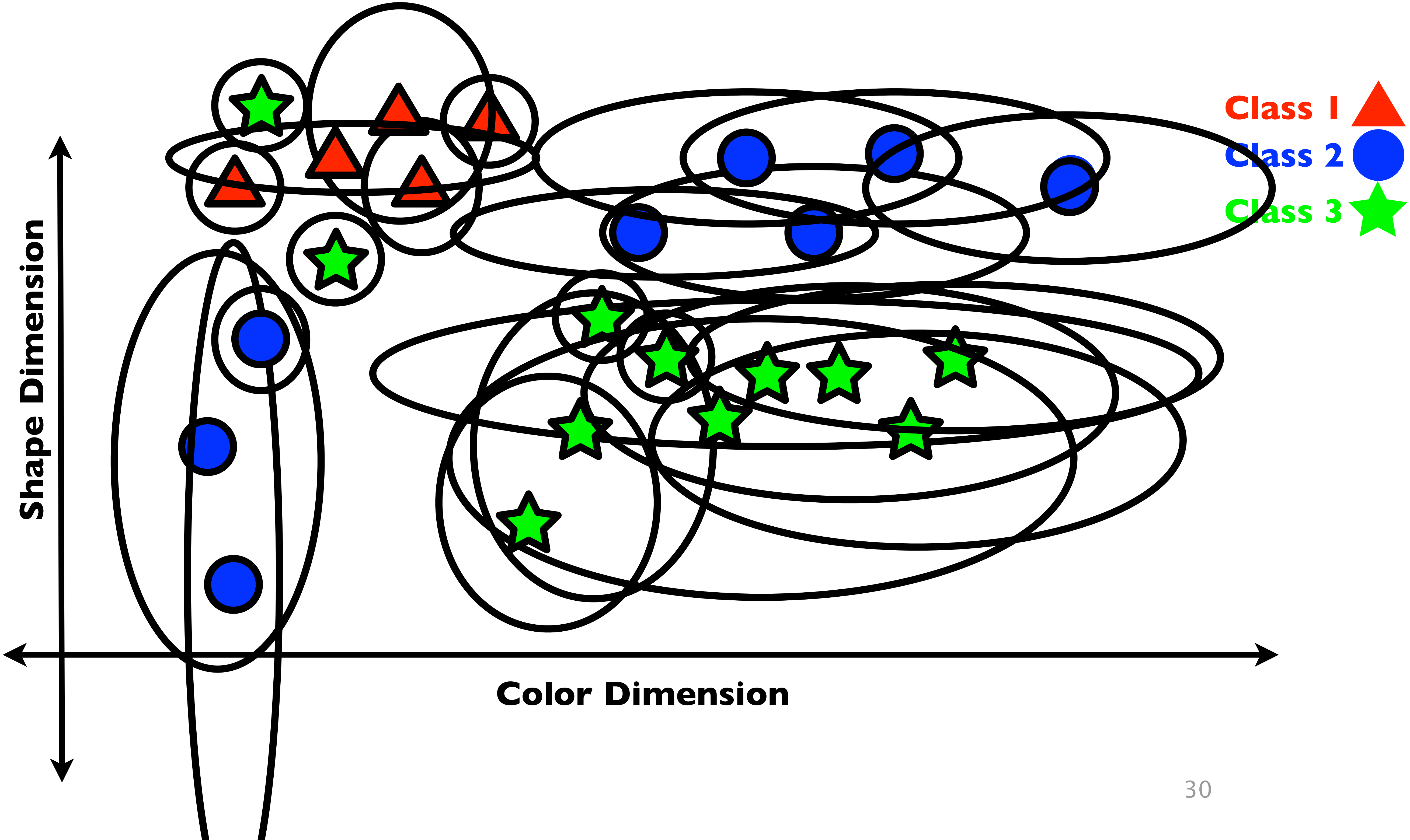
Non-parametric density estimation



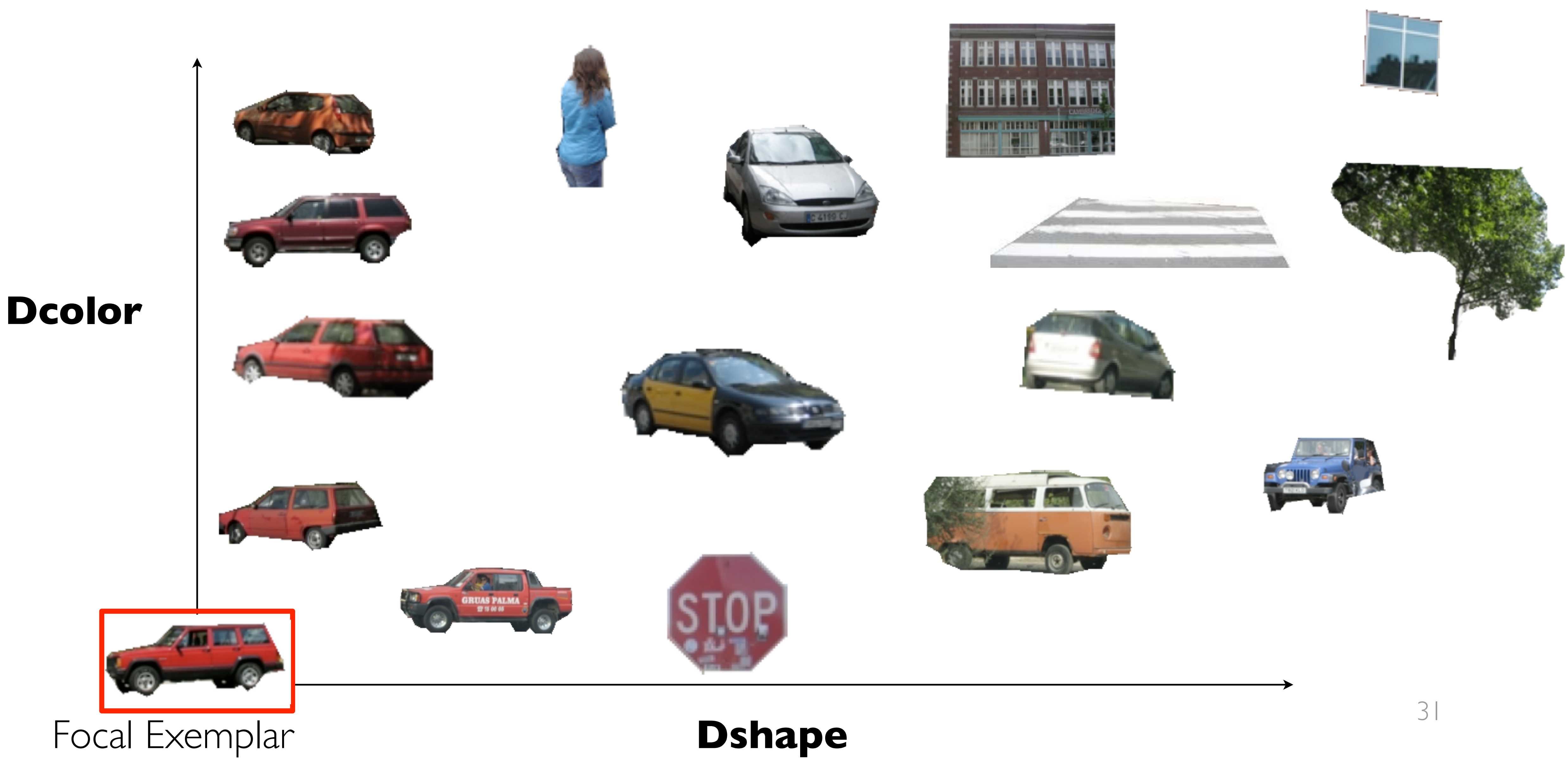
Non-parametric density estimation



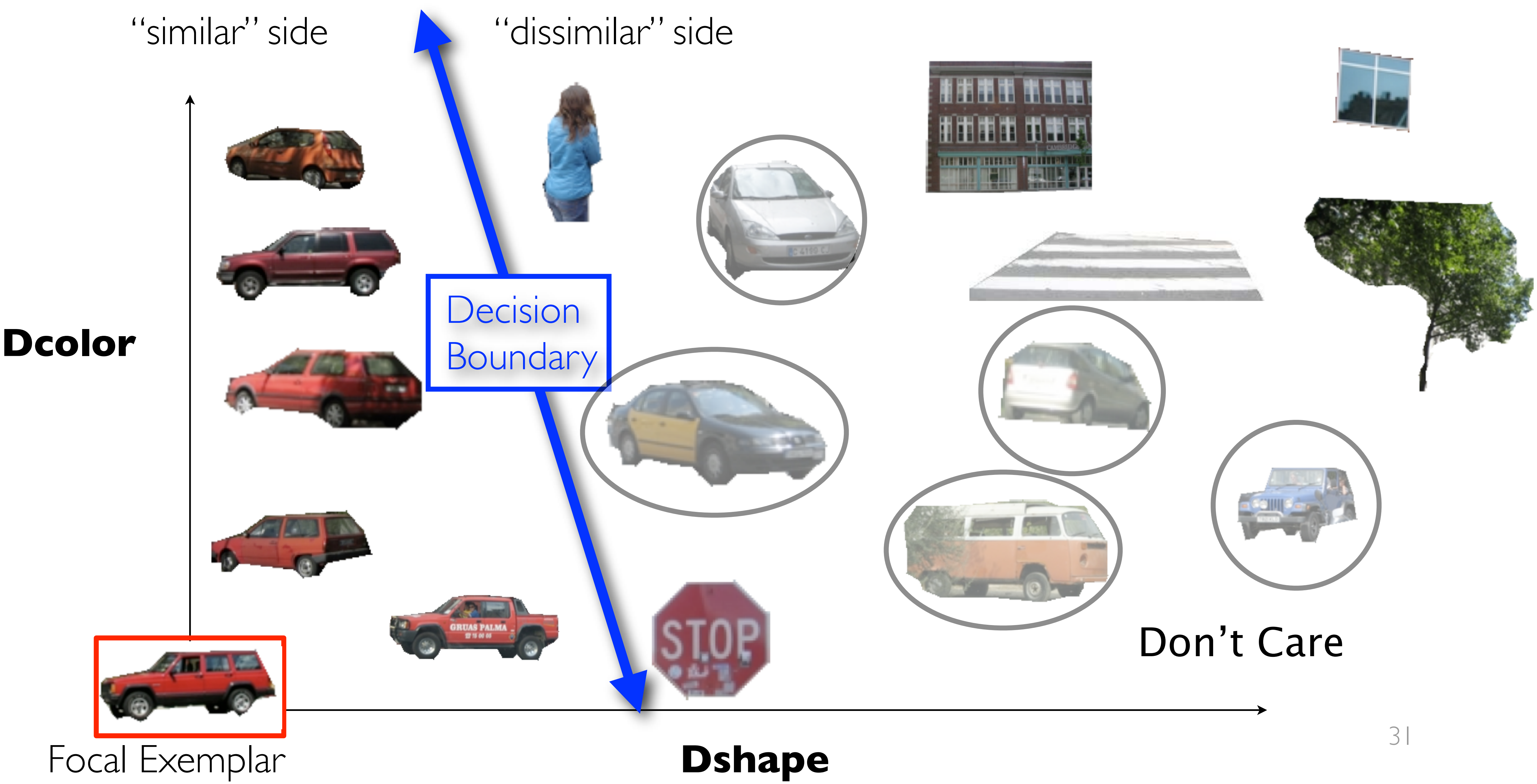
Non-parametric density estimation



Learning Distance Functions

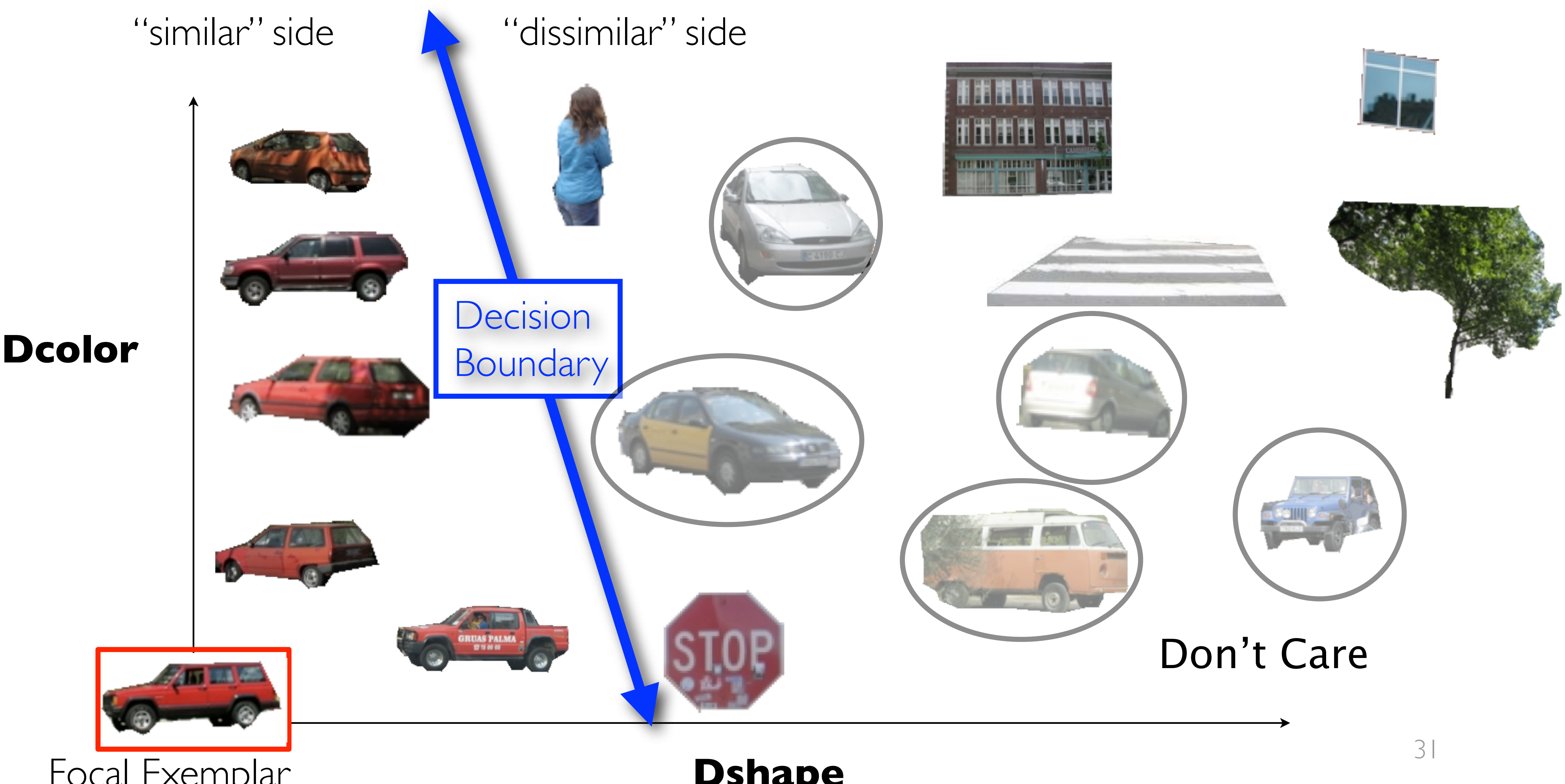


Learning Distance Functions



Learning Distance Functions

$$f(\mathbf{w}, \boldsymbol{\alpha}) = \sum_{i \in C} \alpha_i L(-\mathbf{w} \cdot \mathbf{d}_i) + \sum_{i \notin C} L(\mathbf{w} \cdot \mathbf{d}_i)$$



Learnin… …tions

w: positive weight vector

binary vector encodes which K exemplars are forced to be similar.

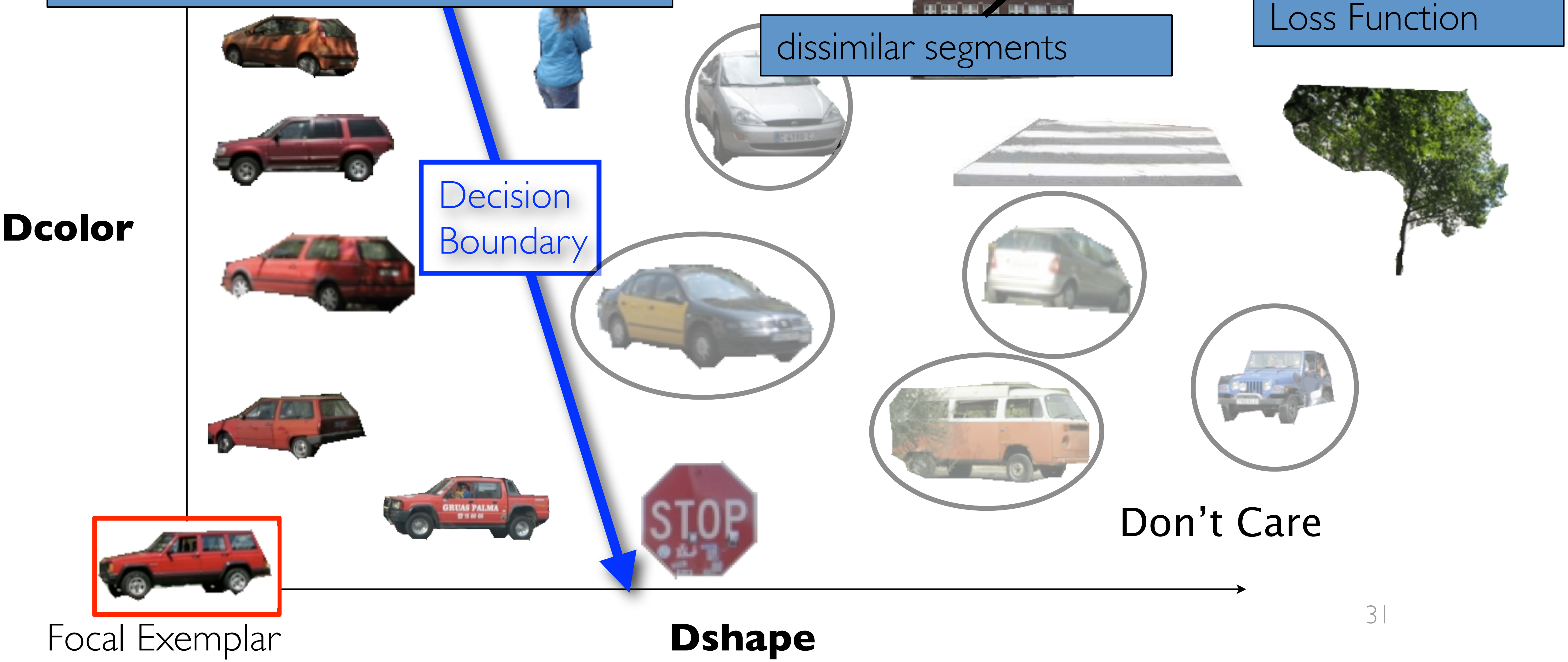
$$f(\mathbf{w},\boldsymbol{\alpha}) = \sum_{i\in C} \alpha_i L(-\mathbf{w}\cdot\mathbf{d}_i) + \sum_{i\notin C} L(\mathbf{w}\cdot\mathbf{d}_i)$$

C: candidate similar exemplars
exemplars with same label

ide

Loss Function

dissimilar segments



Learning Distance Functions

$$f(\mathbf{w}, \boldsymbol{\alpha}) \;=\; \sum_{i \in C} \alpha_i L(-\mathbf{w} \cdot \mathbf{d}_i) + \sum_{i \notin C} L(\mathbf{w} \cdot \mathbf{d}_i)$$

Iterative Optimization

$$\boldsymbol{\alpha}^k = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \sum_{i \in C} \alpha_i L(-\mathbf{w^k} \cdot \mathbf{d_i})$$

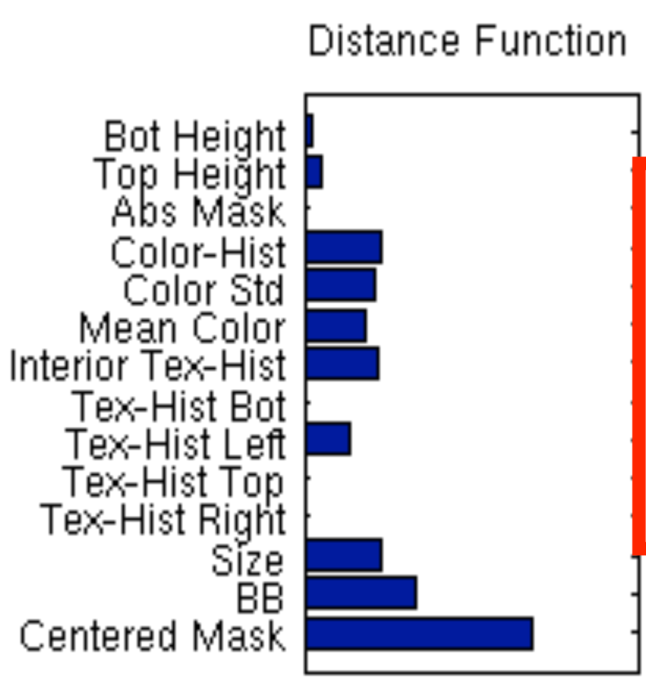
$$\mathbf{w}^{k+1} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i:\alpha_i^k = 1} L(-\mathbf{w} \cdot \mathbf{d}_i) + \sum_{i \notin C} L(\mathbf{w} \cdot \mathbf{d}_i)$$

alpha sums to K=10 (forced number of similar exemplars)
L: squared hinge-loss function (SVM optimization)
initialize with texton histogram distance (works well for a wide array of objects!)

Visualizing Distance Functions (Training Set)

Query

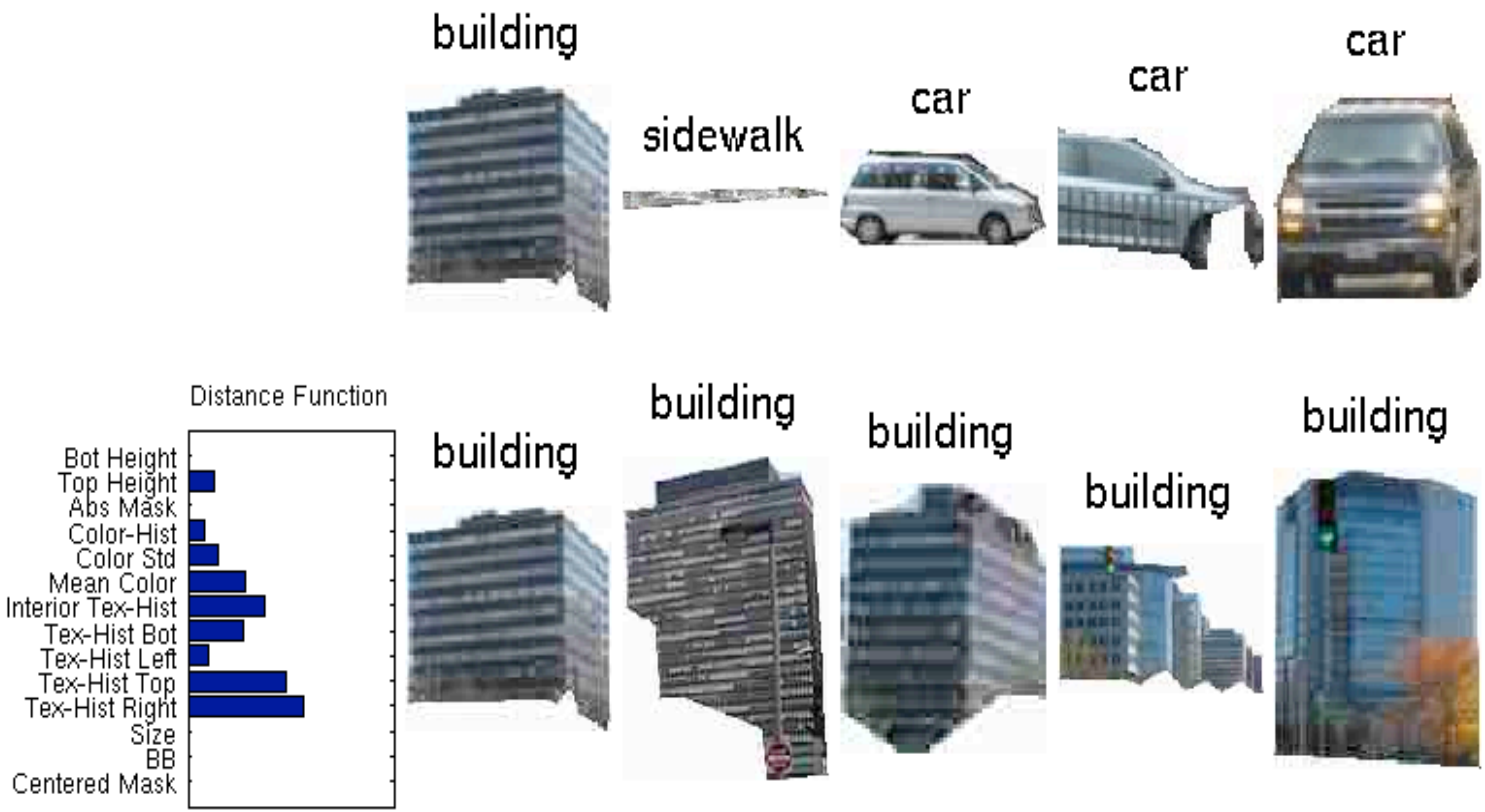
Top Neighbors with Tex-Hist Dist



Query

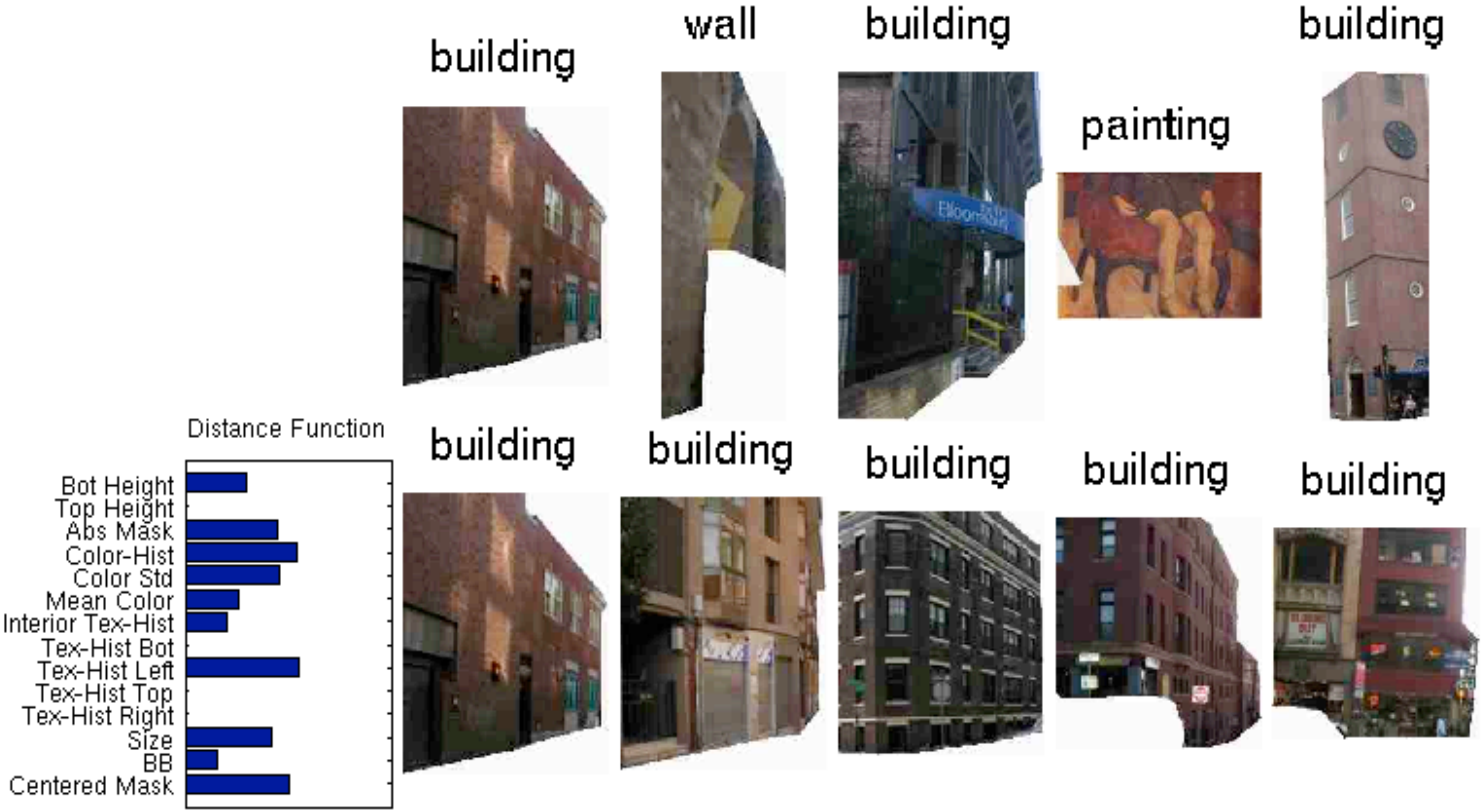
Top Neighbors with Learned Dist

Visualizing Distance Functions (Training Set)

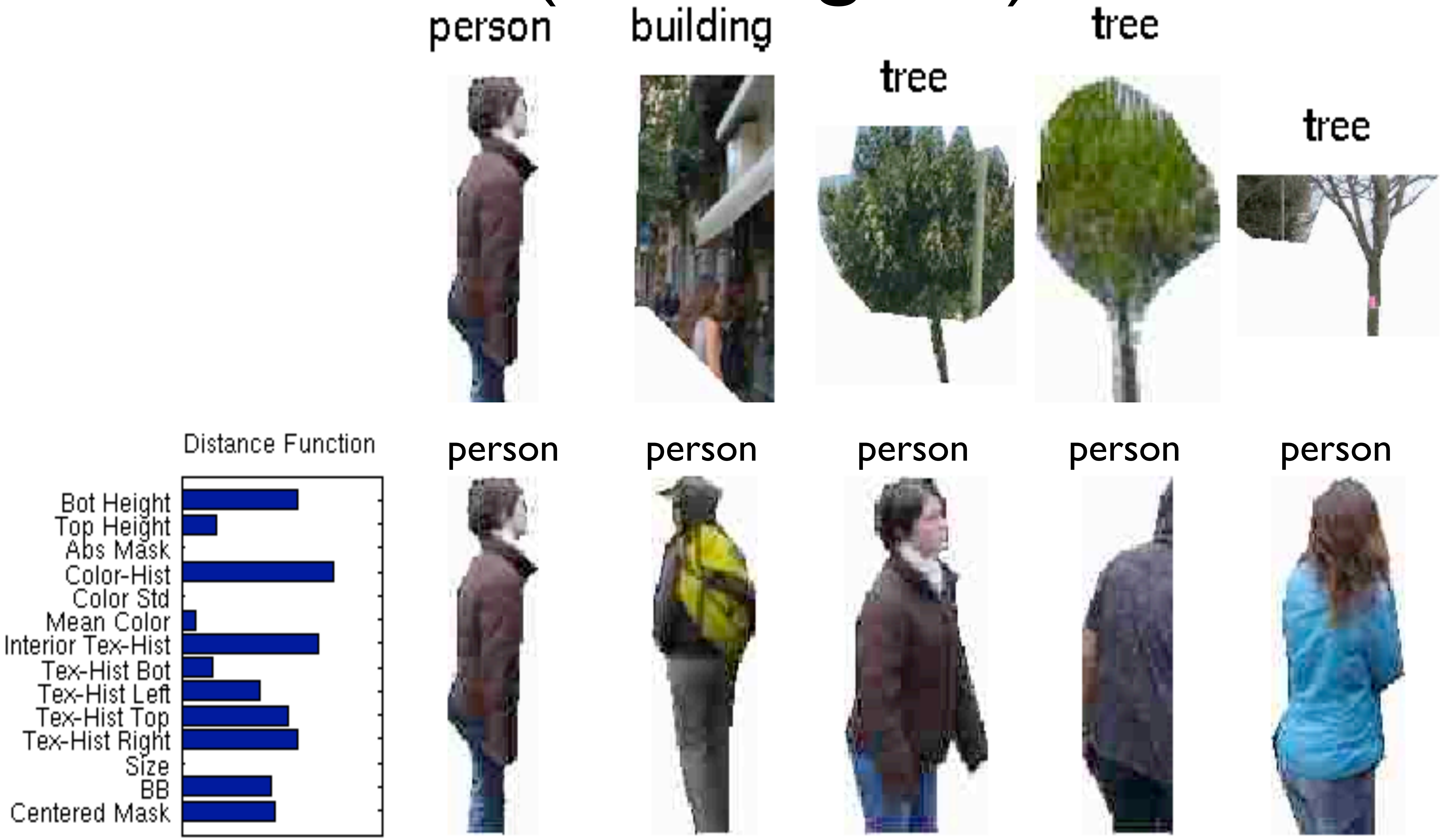


Visualizing Distance Functions (Training Set)

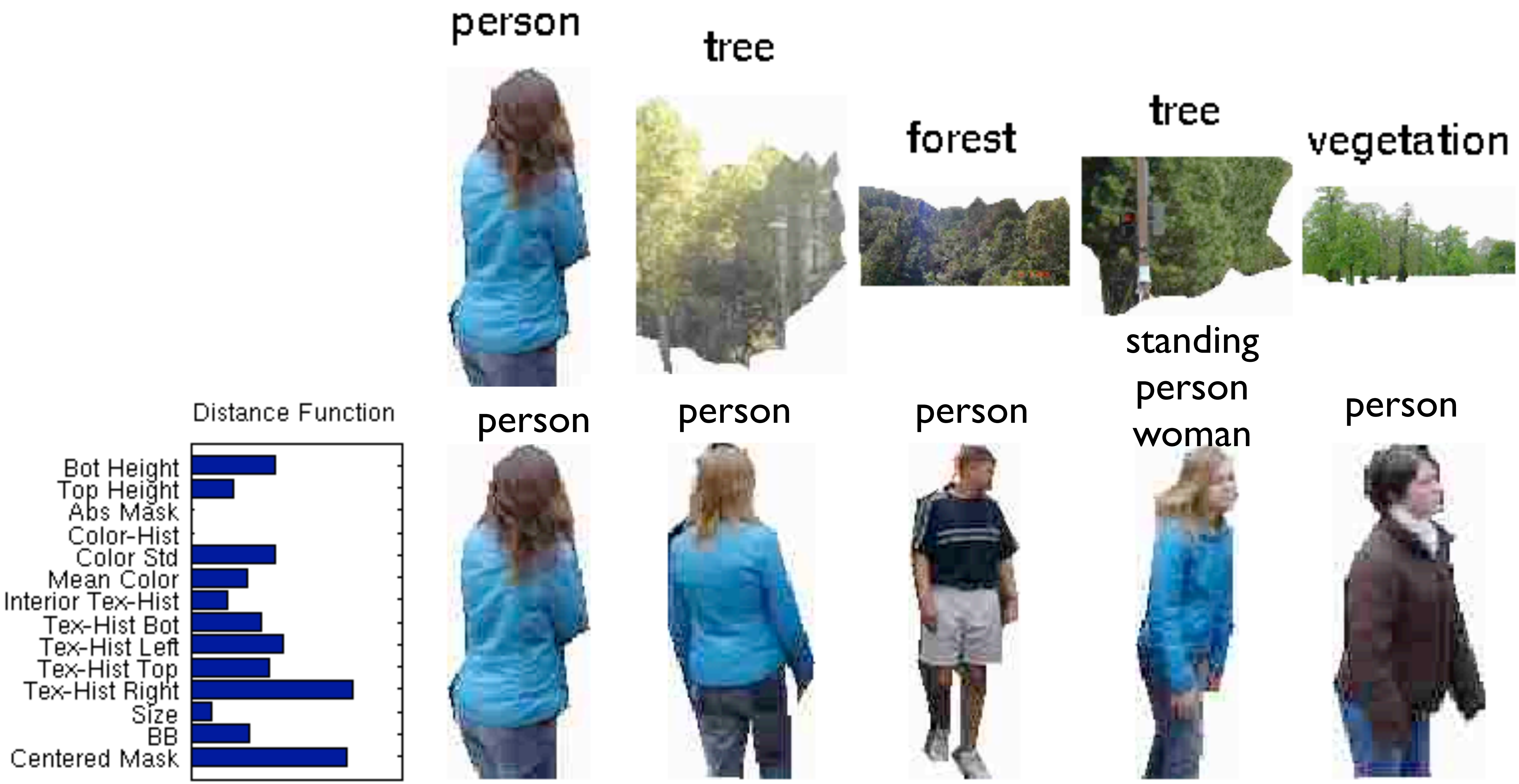
Visualizing Distance Functions (Training Set)



Visualizing Distance Functions (Training Set)

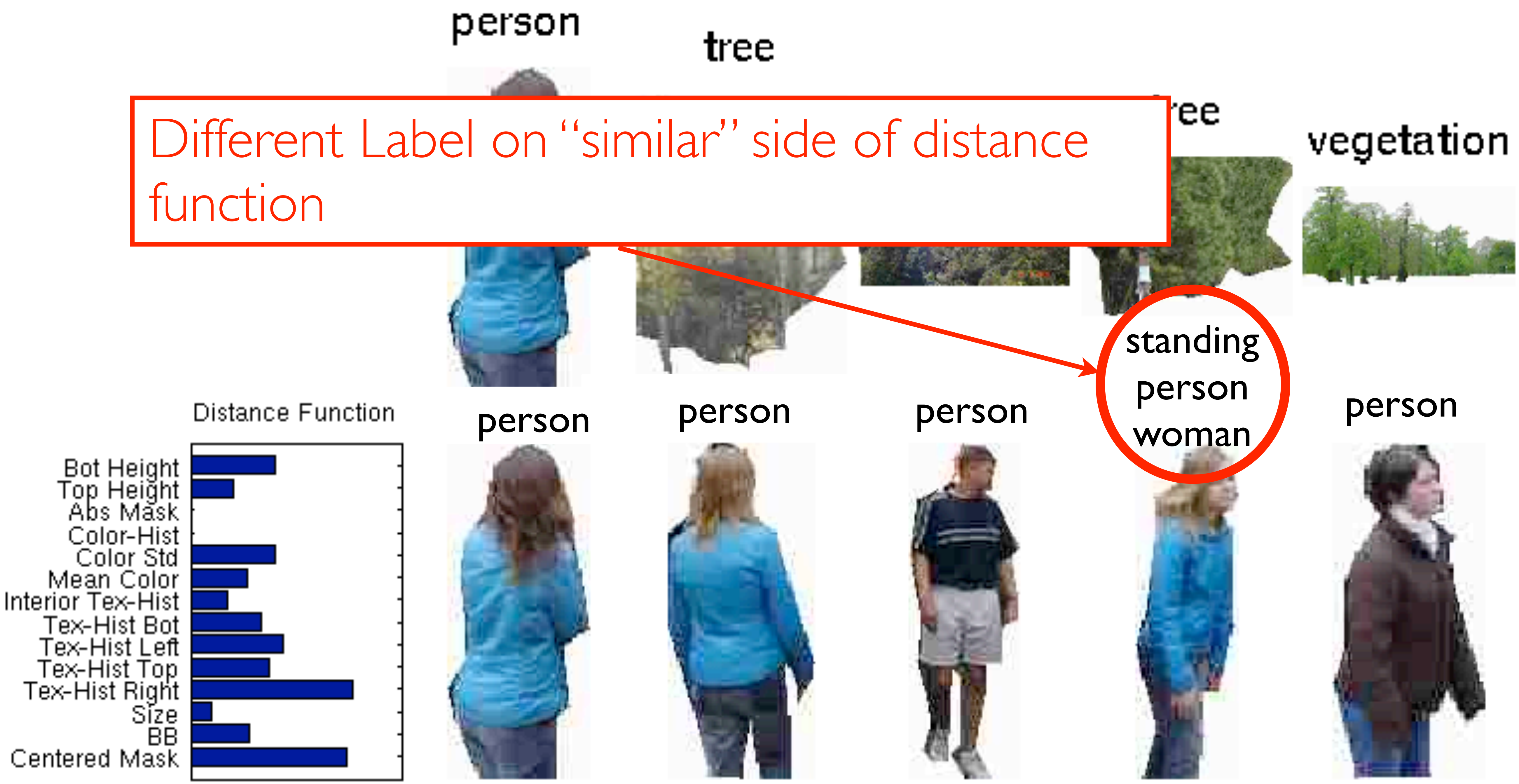


Visualizing Distance Functions (Training Set)



Visualizing Distance Functions (Training Set)

Different Label on "similar" side of distance function



Labels Crossing Boundary

stop sign	sign	7.8%
pole	streetlight	6.7%
motorcycle	motorbike	6.2%
mountains	mountain	6.2%
ground grass	sidewalk	3.7%
grass	lawn	3.6%
road highway	road	3.4%
painting	picture	3.4%
sidewalk	road	3.2%
cloud	sky	3.1%
grass	ground grass	3.1%
mountain	mountains	2.7%

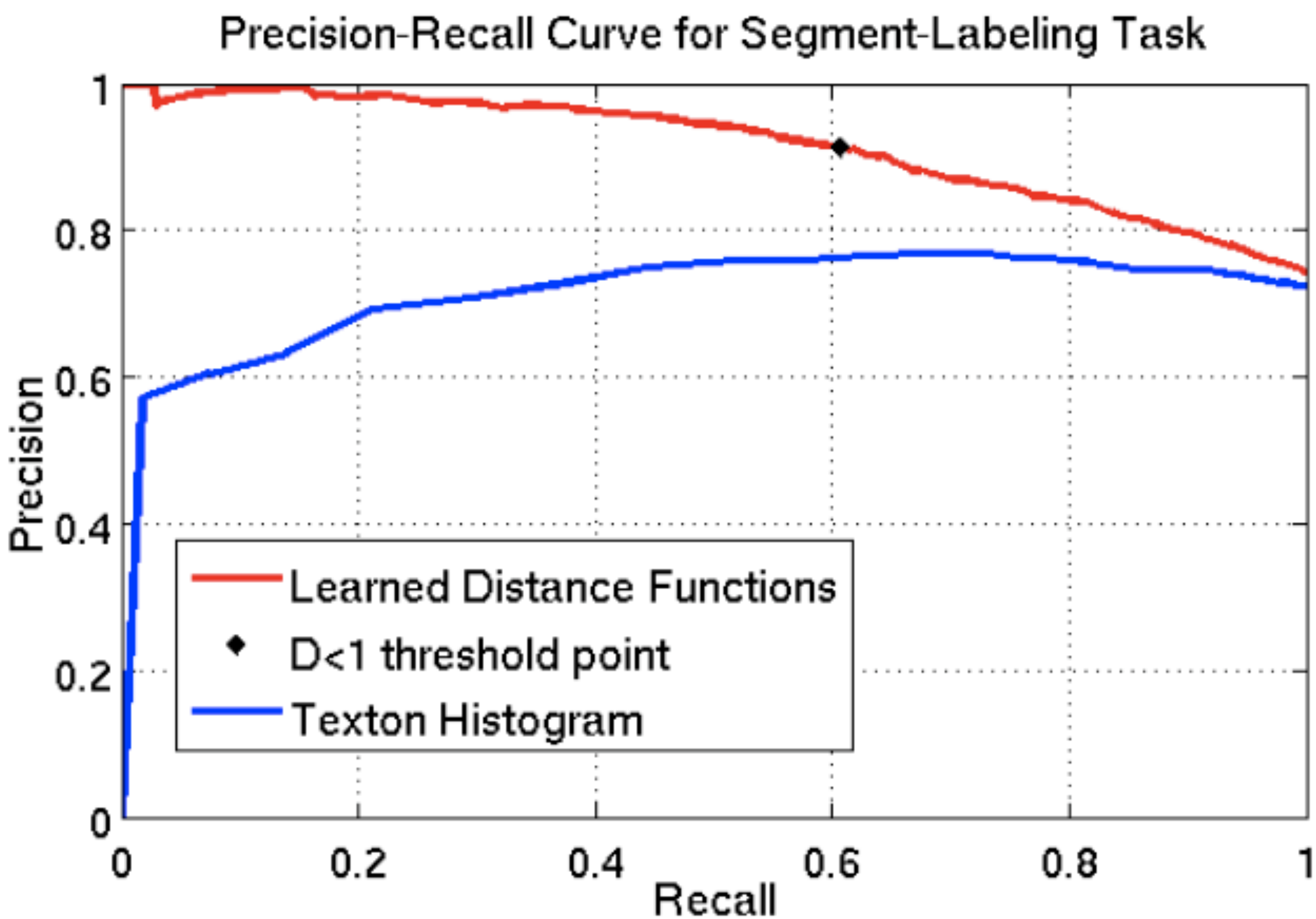
Table 2: Top dozen label confusions discovered after distance function learning.

Recognition in Test Set

- Compute the similarity between an input and all exemplars
- All exemplars with $D < 1$ are "associated" with the input
- Most occurring label from associations is propagated onto input
- Association confidence score favors more associations and smaller distances

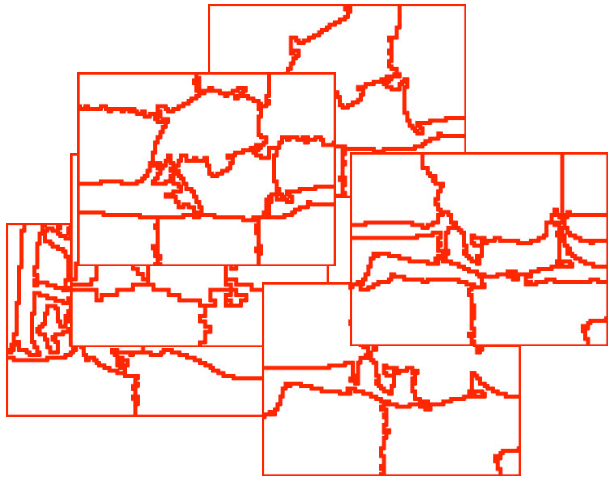
$$s(S,E) = 1 / \sum_{e \in E} \frac{1}{D_e(S)}$$

Performance on labeling perfect segments (test set)



Object Segmentation via Recognition

- Generate Multiple Segmentations (Hoiem 2005, Russell 2006, Malisiewicz 2007)
  - Mean-Shift and Normalized Cuts
  - Use pairs and triplets of adjacent segments
  - Generate about 10,000 segments per image

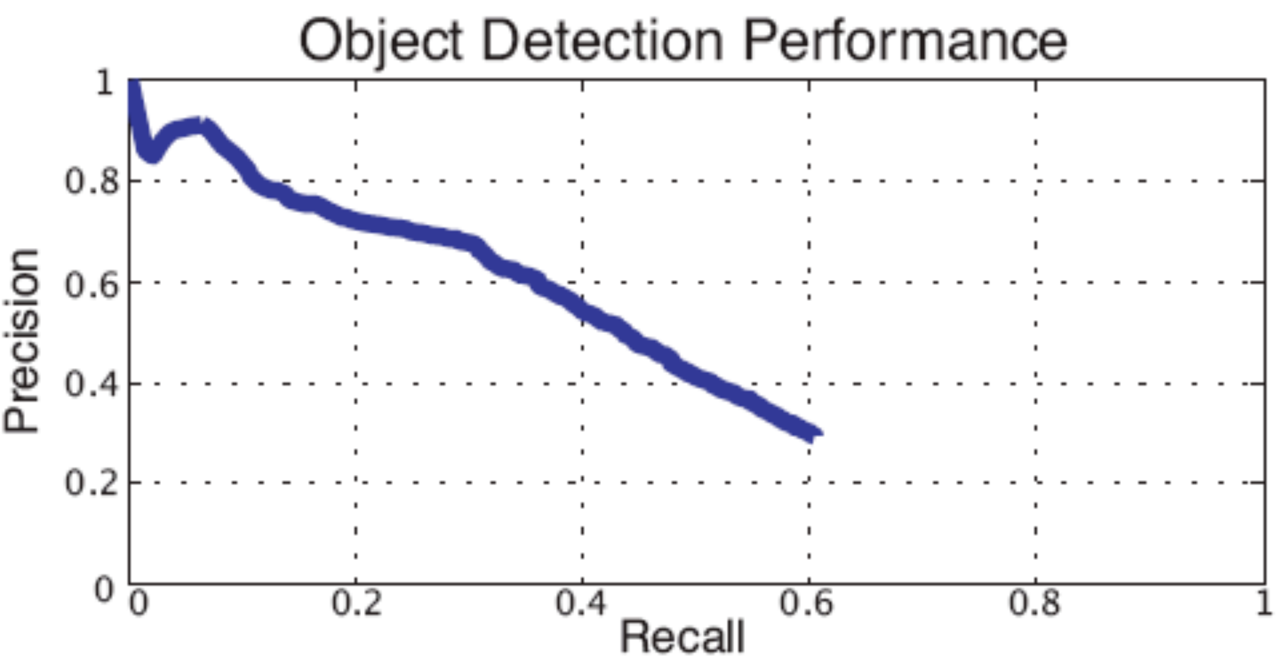


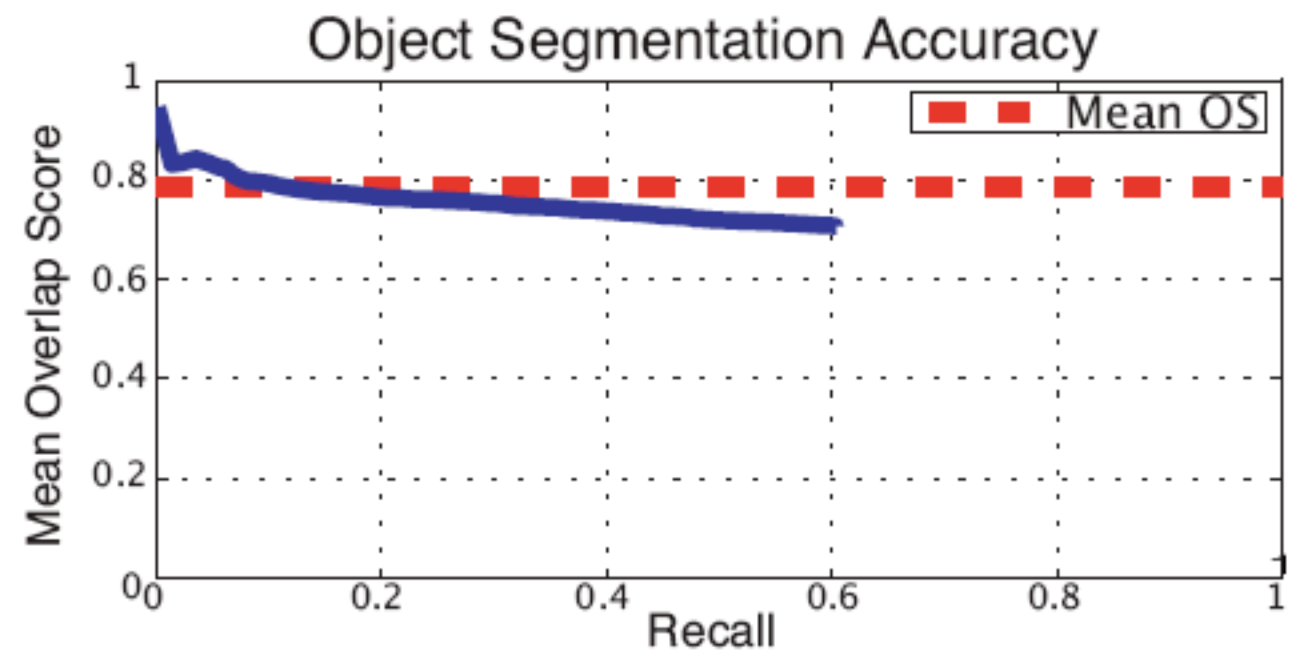
- Enhance training with bad segments
- Apply learned distance functions to bottom-up segments

Example Associations

Bottom-Up
Segments

Quantitative Evaluation





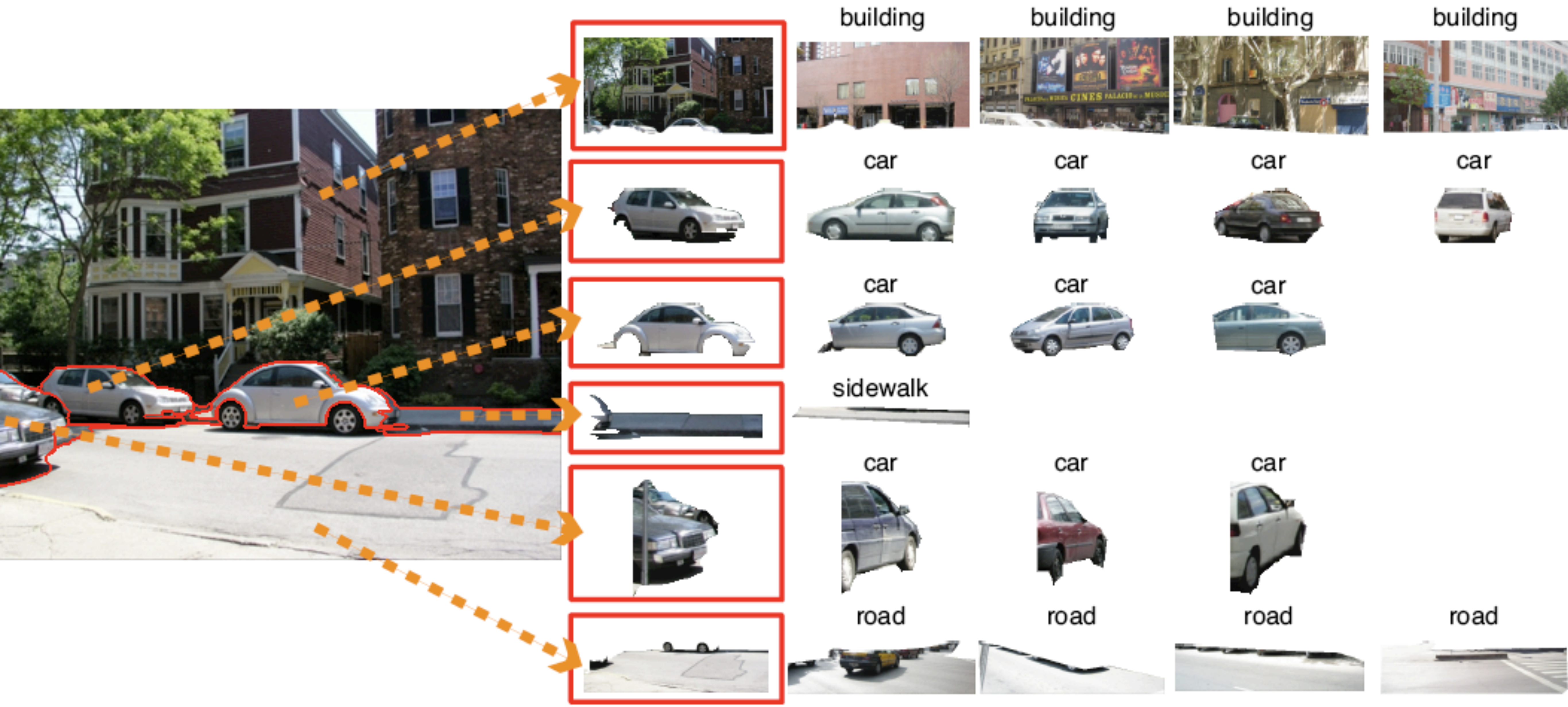
OS(A,B) = Overlap Score = intersection(A,B) / union(A,B)

Object hypothesis is correct if labels match and OS > .5

*We do not penalize for multiple correct overlapping associations

Toward Image Parsing

Toward Image Parsing

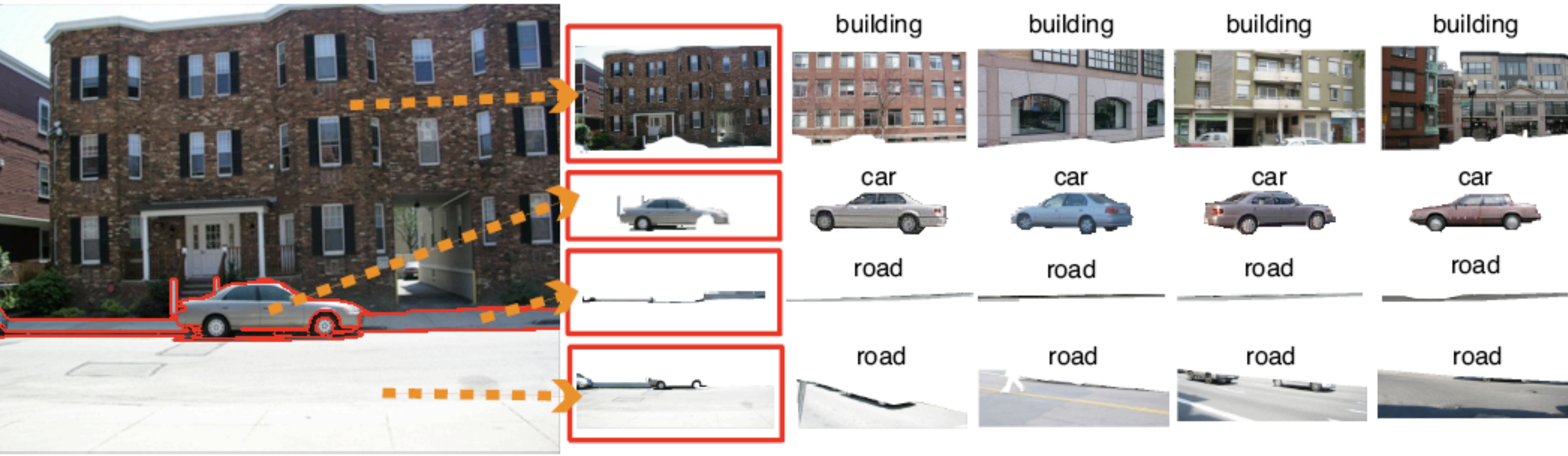


Conclusion and Future Work

- A multi-class exemplar-based object recognition system
- Segment and Recognize objects in LabelMe images

- Address scalability of the proposed approach
- Cleverly integrate object associations to parse images

Thank You



Questions?