

Object Class Recognition

Readings: Yi Li's 2 Papers

- Abstract Regions (multiple different segmentations)
- Paper 1: EM as a Classifier (Generative Approach)
- Paper 2: EM Clustering, Feature Generation, Neural Net Classifier (Generative/Discriminative Approach)

Object Class Recognition using Images of Abstract Regions

Yi Li, Jeff A. Bilmes, and Linda G. Shapiro
Department of Computer Science and Engineering
Department of Electrical Engineering
University of Washington

Problem Statement

Given: Some images and their corresponding descriptions



{trees, grass, cherry trees}



{cheetah, trunk}



{mountains, sky}



{beach, sky, trees, water}

...

To solve: What object classes are present in new images



?



?



?

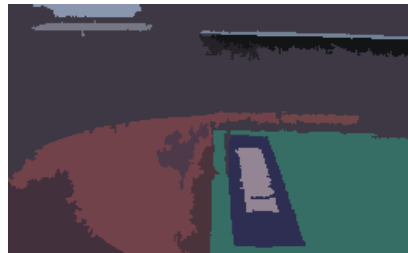


?

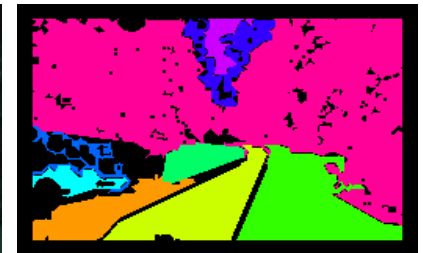
...

Image Features for Object Recognition

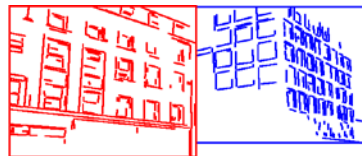
- Color



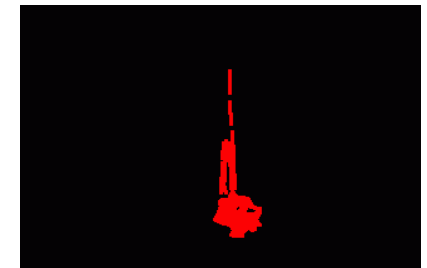
- Texture



- Structure



- Context



Abstract Regions

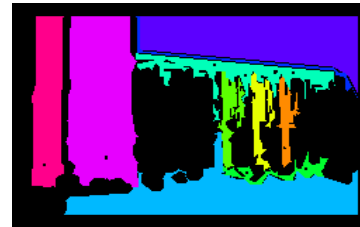
Original Images



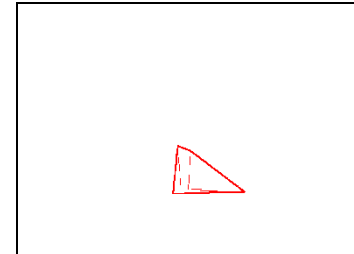
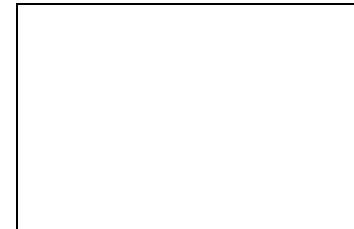
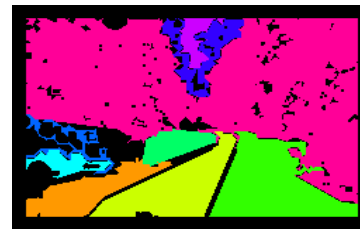
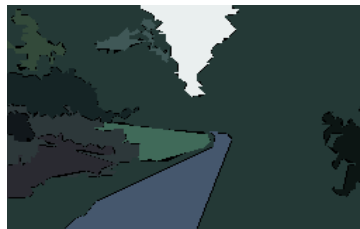
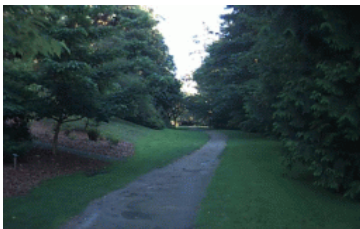
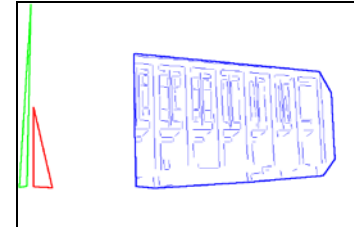
Color Regions



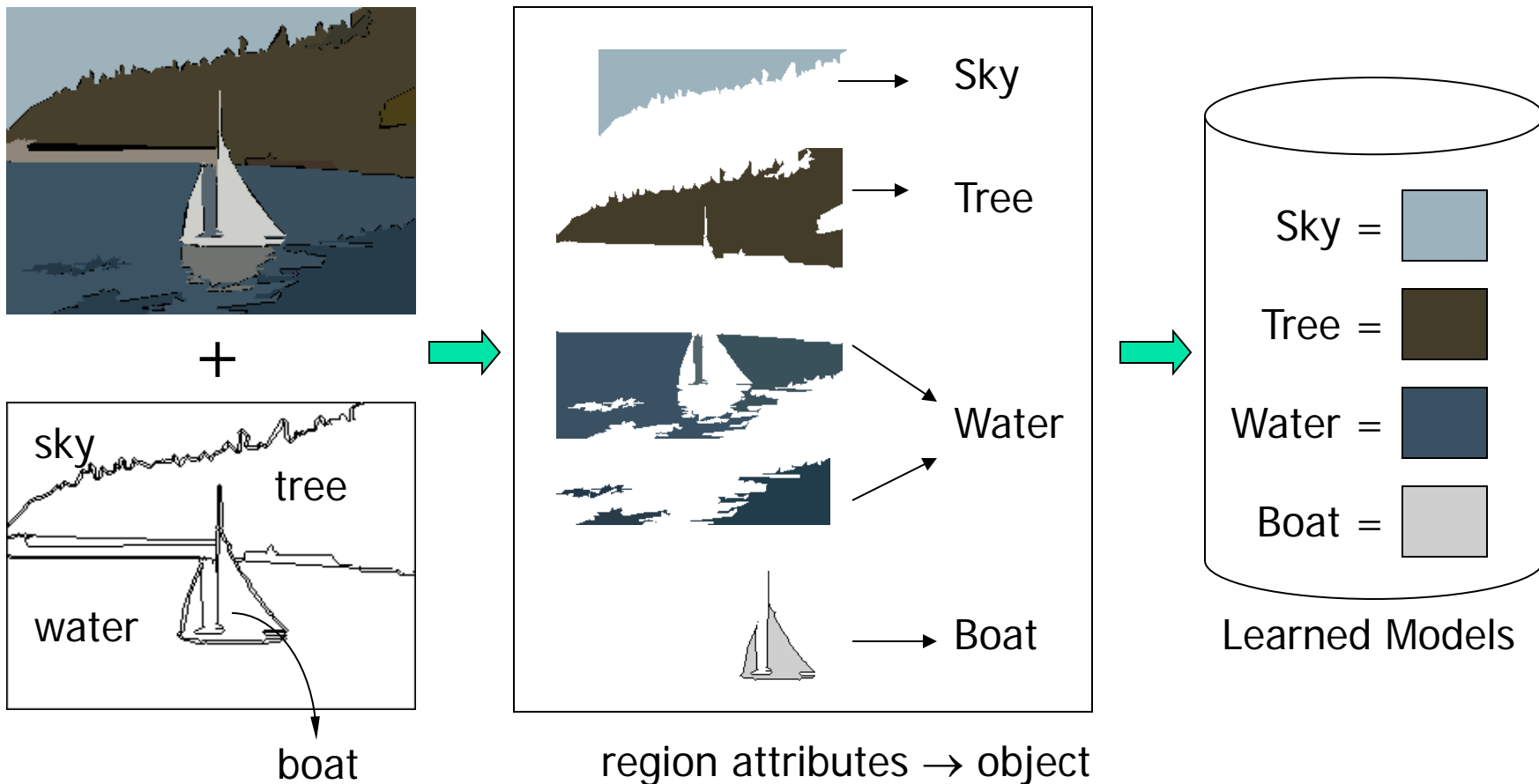
Texture Regions



Line Clusters



Object Model Learning (Ideal)



Our Scenario: **Abstract Regions**

Multiple segmentations whose regions are not labeled; a list of labels is provided for each training image.

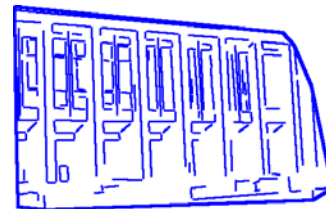
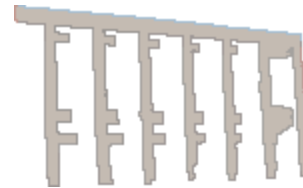
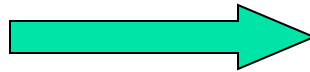
image



labels

{sky, building}

various different segmentations



region attributes from several different types of regions



Object Model Learning

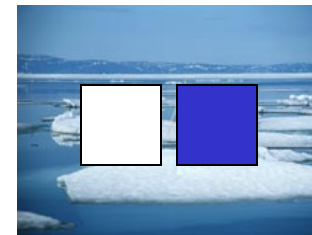
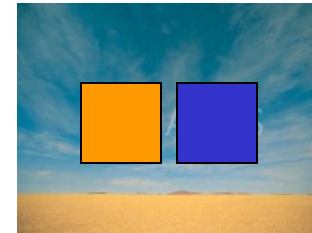
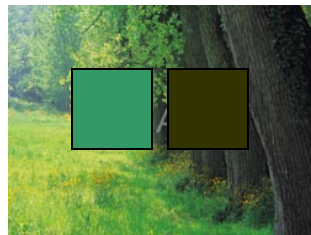
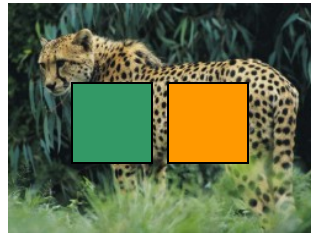
Assumptions:

1. The objects to be recognized can be modeled as multivariate Gaussian distributions.
2. The regions of an image can help us to recognize its objects.

Model Initial Estimation

- Estimate the initial model of an object using all the region features from all images that contain the object

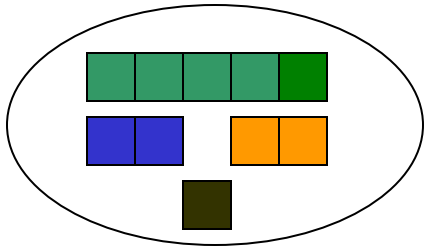
Tree



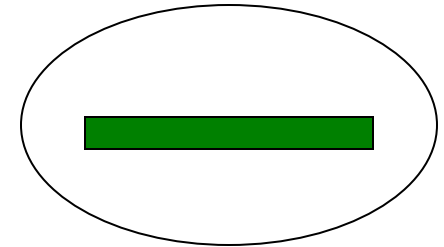
Sky

EM Variant

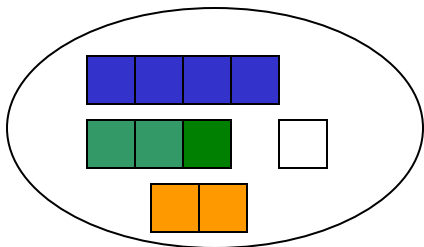
Initial Model for "trees"



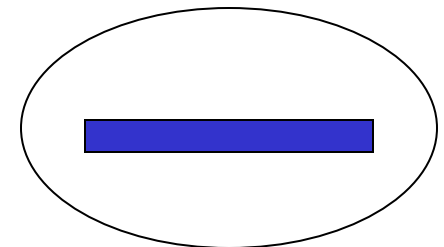
Final Model for "trees"



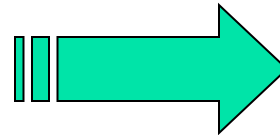
Initial Model for "sky"



Final Model for "sky"



EM





EM Variant

- **Fixed Gaussian components** (one Gaussian per object class) and **fixed weights** corresponding to the frequencies of the corresponding objects in the training data.
- **Customized initialization** uses only the training images that contain a particular object class to initialize its Gaussian.
- **Controlled expectation step** ensures that a feature vector only contributes to the Gaussian components representing objects present in its training image.
- **Extra background component** absorbs noise.

Gaussian for
trees

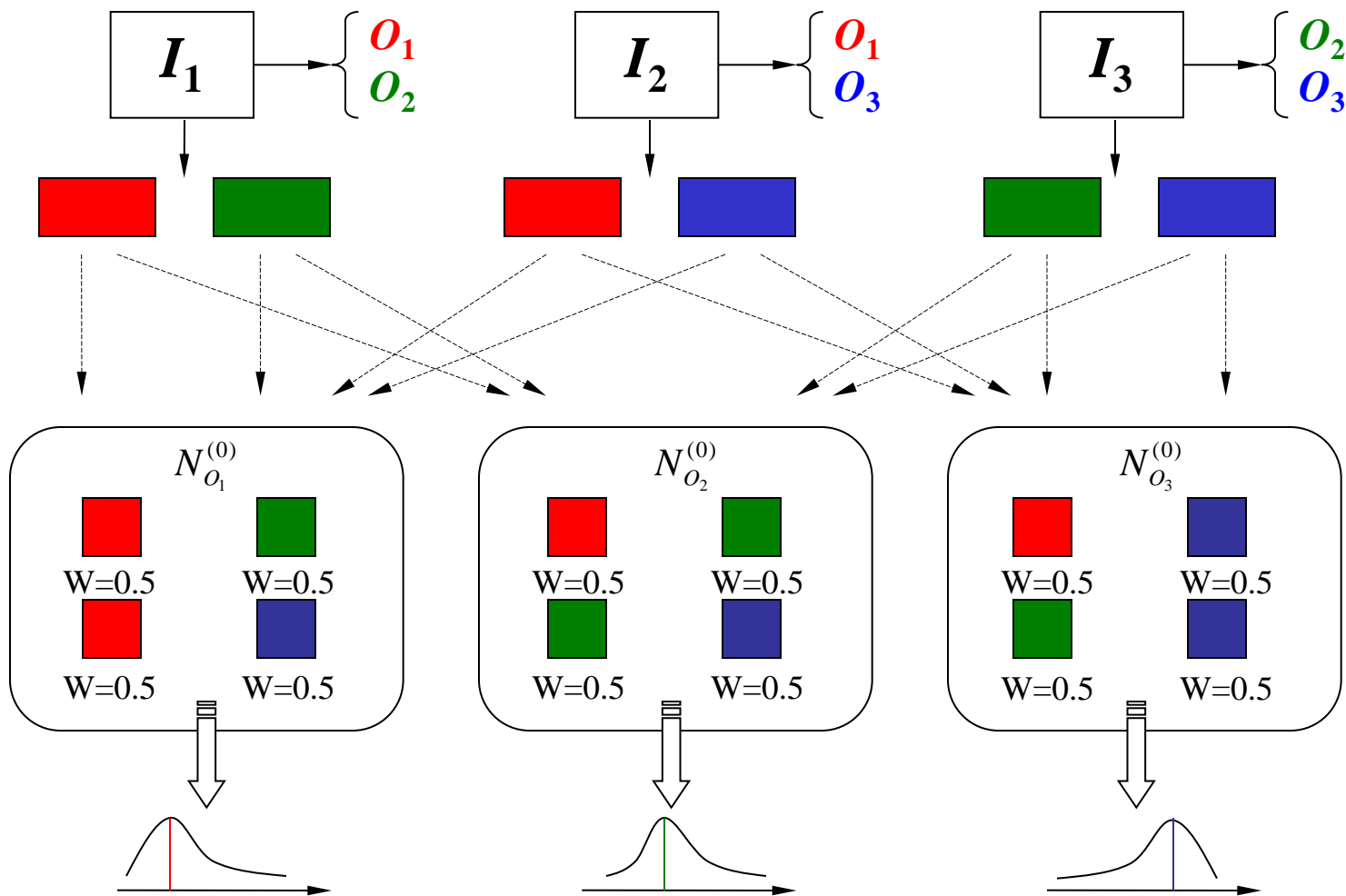
Gaussian for
buildings

Gaussian for
sky

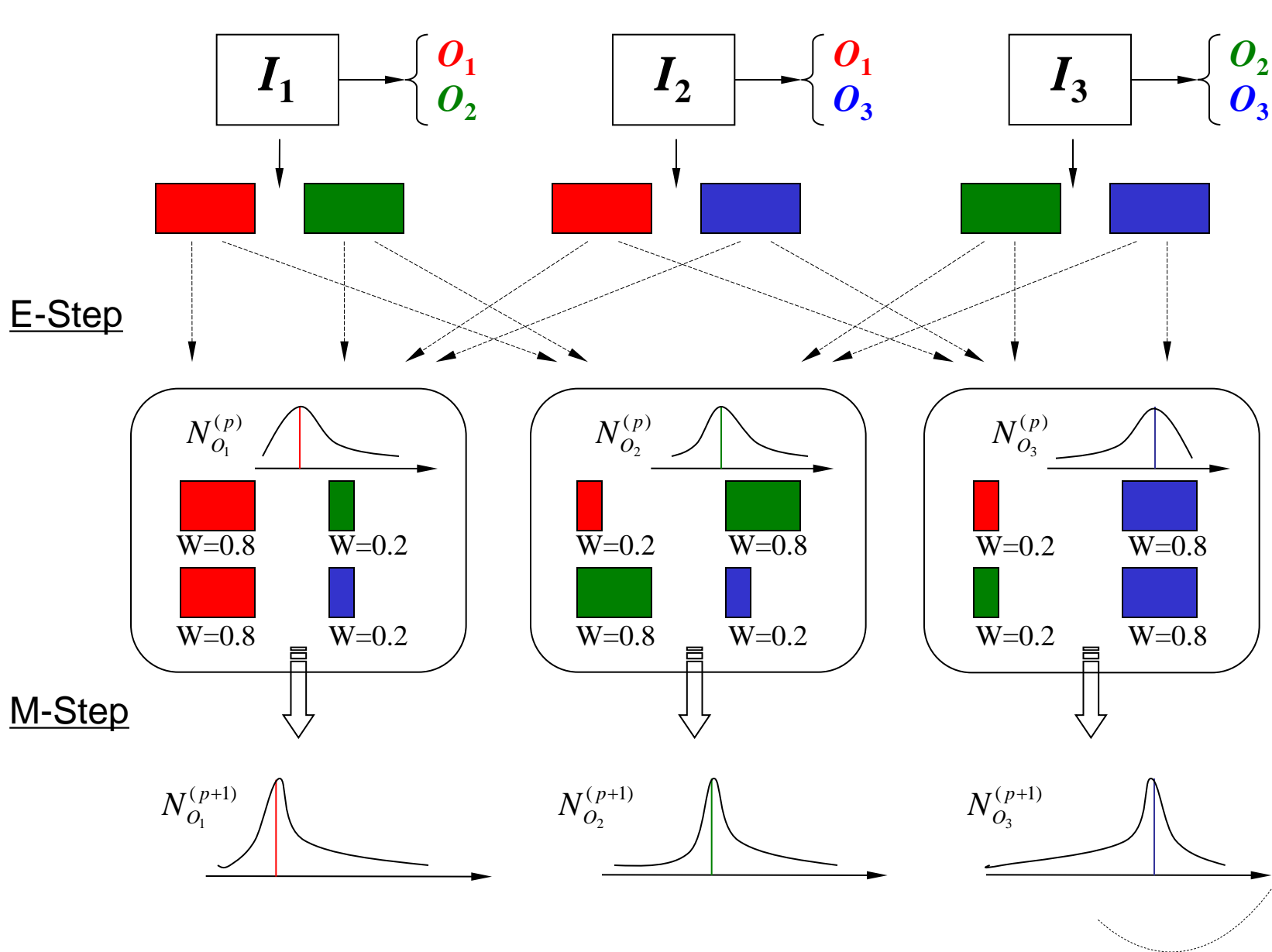
Gaussian for
background

1. Initialization Step (Example)

Image & description



2. Iteration Step (Example)

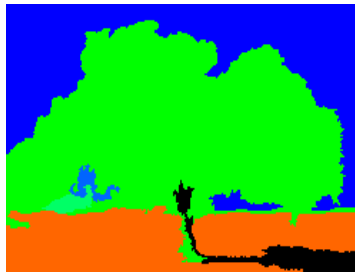


Recognition

Test Image



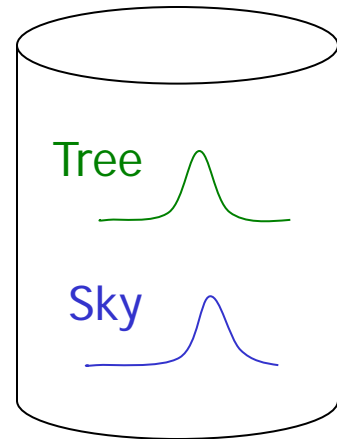
Color Regions



compare



Object Model Database



To calculate $p(\text{tree} / \text{image})$

$$p(\text{tree} / \text{image}) = f \left(\begin{array}{l} p(\text{tree} / \text{blue sky}) \\ p(\text{tree} / \text{green foliage}) \\ p(\text{tree} / \text{orange ground}) \\ p(\text{tree} / \text{black trunk}) \end{array} \right)$$

$$p(o | F_I^a) = f_{r^a \in F_I^a} (p(o | r^a))$$

f is a function that combines probabilities from all the color regions in the image.

What could it be?



Combining different abstract regions

- Treat the different types of regions **independently** and combine at the time of classification.

$$p(o | \{F_I^a\}) = \prod_a p(o | F_I^a)$$

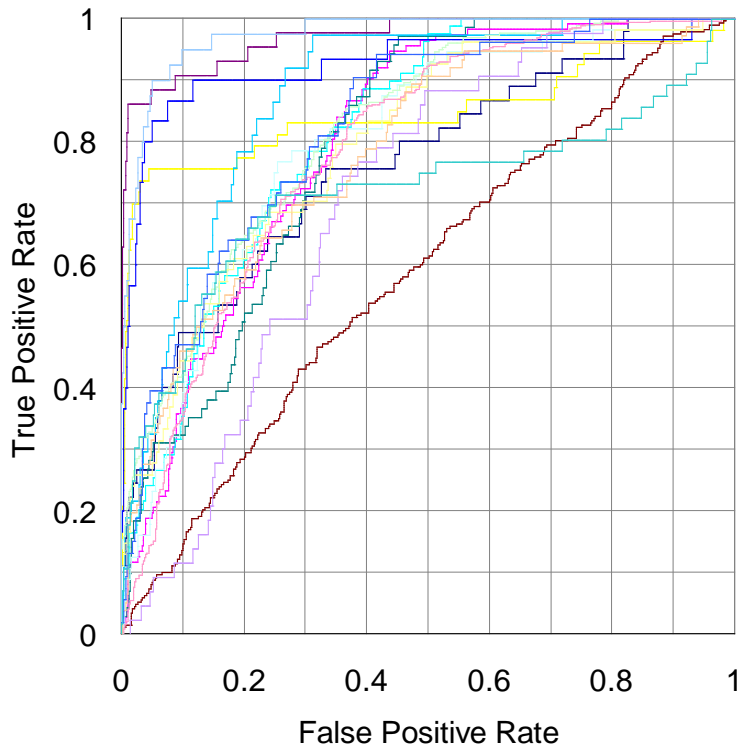
- Form **intersections** of the different types of regions, creating smaller regions that have both color and texture properties for classification.



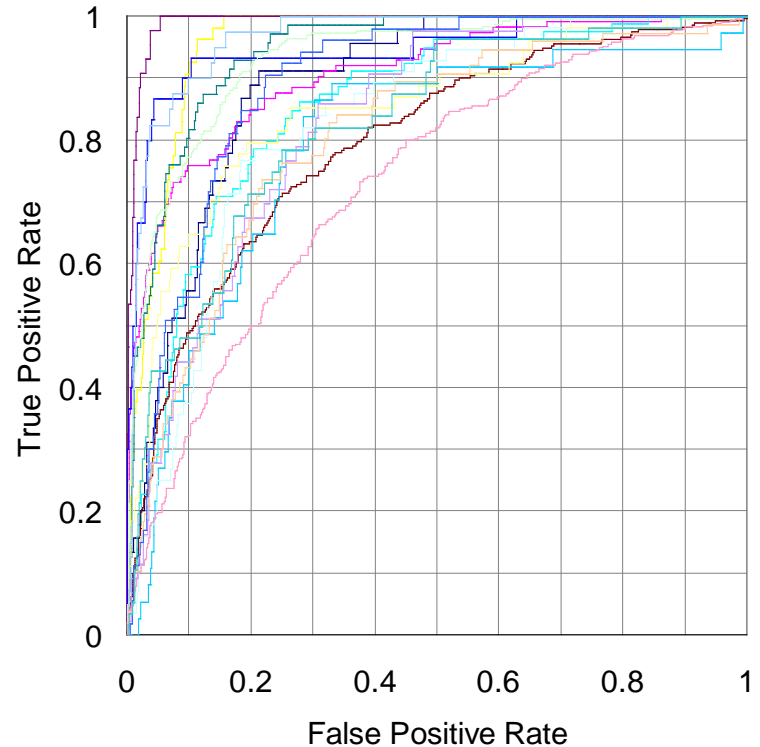
Experiments (on 860 images)

- 18 keywords: **mountains** (30), **orangutan** (37), **track** (40), **tree trunk** (43), **football field** (43), **beach** (45), **prairie grass** (53), **cherry tree** (53), **snow** (54), **zebra** (56), **polar bear** (56), **lion** (71), **water** (76), **chimpanzee** (79), **cheetah** (112), **sky** (259), **grass** (272), **tree** (361).
- A set of cross-validation experiments (80% as training set and the other 20% as test set)
- The poorest results are on object classes “**tree**,” “**grass**,” and “**water**,” each of which has a high variance; a single Gaussian model is insufficient.

ROC Charts



Independent Treatment of
Color and Texture



Using Intersections of
Color and Texture Regions

Sample Retrieval Results

cheetah



Sample Results (Cont.)

grass



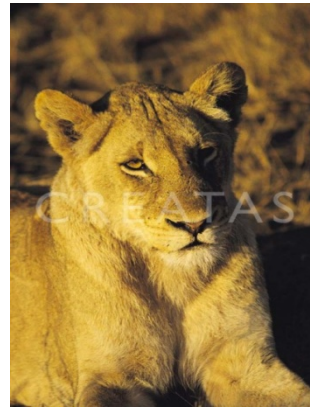
Sample Results (Cont.)

cherry tree



Sample Results (Cont.)

lion





Summary

- Designed a set of abstract region features: **color**, **texture**, **structure**, . . .
- Developed a new **semi-supervised EM-like algorithm** to recognize object classes in color photographic images of outdoor scenes; tested on 860 images.
- Compared **two different methods of combining** different types of abstract regions. The intersection method had a higher performance



A Generative/Discriminative Learning Algorithm for Image Classification

Y. Li, L. G. Shapiro, J. Bilmes
Department of Computer Science
Department of Electrical Engineering
University of Washington

Our New Approach to Combining Different Feature Types



Phase 1:

- Treat each type of abstract region separately
- For abstract region type a and for object class o , use the EM algorithm to construct a model that is a **mixture of multivariate Gaussians** over the features for type a regions.



This time Phase 1 is just EM clustering.

- For object class (tree) and abstract region type color we will have some preselected number M of clusters, each represented by a 3-dimensional Gaussian distribution in color space.

$$N_1(\mu_1, \Sigma_1) \quad N_2(\mu_2, \Sigma_2) \quad \dots \quad N_M(\mu_M, \Sigma_M)$$



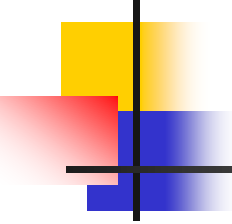
Consider only abstract region type color (c) and object class object (o)

- At the end of Phase 1, we can compute the distribution of **color feature vectors** X^c in an image containing object o .

$$P(X^c|o) = \sum_{m=1}^{M^c} w_m^c \cdot N(X^c; \mu_m^c, \Sigma_m^c)$$

This is called a
Gaussian
mixture model
for the object o

- M^c is the number of components for object o .
- The w 's are the weights of the components.
- The μ 's and Σ 's are the parameters of the components



Now we can determine which components are likely to be present in an image.

- The probability that the feature vector X from color region r of image I_i comes from component m is given by

$$P(X_{i,r}^c, m^c) = w_m^c \cdot N(X_{i,r}^c, \mu_m^c, \Sigma_m^c)$$

Just use the Gaussian function to decide on the probability of region.

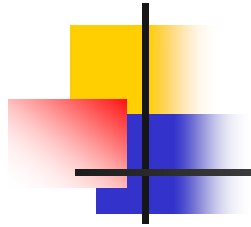
- Then the probability that image I_i has a region that comes from component m is

$$P(I_i, m^c) = f(\{P(X_{i,r}^c, m^c) | r = 1, 2, \dots\})$$

Probability of a component in image is max of all regions.

- where f is an aggregate function such as **mean** or **max**

Aggregate Scores for Color



Components

1 2 3 4 5 6 7 8



beach



.93	.16	.94	.24	.10	.99	.32	.00
-----	-----	-----	-----	-----	-----	-----	-----

beach



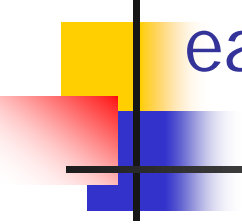
.66	.80	.00	.72	.19	.01	.22	.02
-----	-----	-----	-----	-----	-----	-----	-----

not
beach



.43	.03	.00	.00	.00	.00	.15	.00
-----	-----	-----	-----	-----	-----	-----	-----

We now use **positive** and **negative** training images, calculate for each the probabilities of regions of each component, and form a training matrix.


$$\begin{array}{l} I_1^+ \\ I_2^+ \\ \vdots \\ I_1^- \\ I_2^- \\ \vdots \end{array} \left[\begin{array}{cccc} P(I_1^+, 1^c) & P(I_1^+, 2^c) & \cdots & P(I_1^+, M^c) \\ P(I_2^+, 1^c) & P(I_2^+, 2^c) & \cdots & P(I_2^+, M^c) \\ \vdots & \vdots & & \\ P(I_1^-, 1^c) & P(I_1^-, 2^c) & \cdots & P(I_1^-, M^c) \\ P(I_2^-, 1^c) & P(I_2^-, 2^c) & \cdots & P(I_2^-, M^c) \\ \vdots & \vdots & & \end{array} \right]$$



Phase 2 Learning

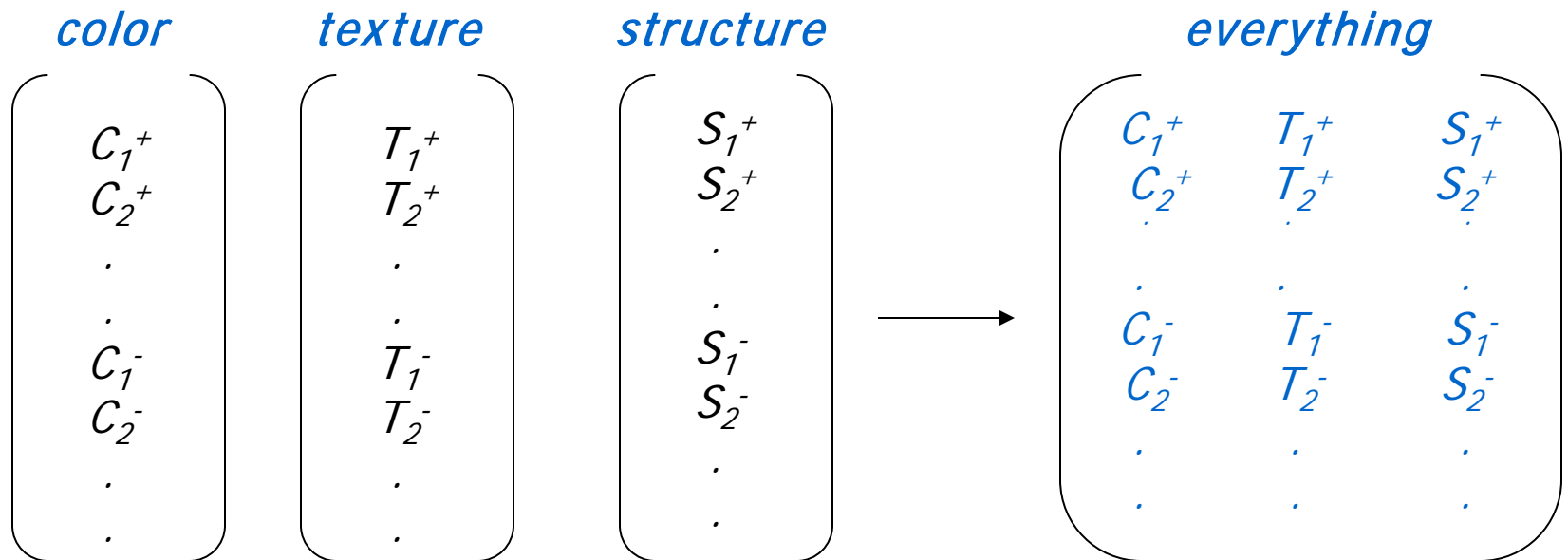
- Let C_i be row i of the training matrix.
- Each such row is a **feature vector** for the color features of regions of image I_i that relates them to the Phase 1 components.
- Now we can use a second-stage classifier to learn $P(o/I_i)$ for each object class o and image I_i .



Multiple Feature Case

- We calculate separate Gaussian mixture models for each different features type:
- Color: C_i
- Texture: T_i
- Structure: S_i
- and any more features we have (motion).

Now we concatenate the matrix rows from the different region types to obtain a **multi-feature-type training matrix**.





Discriminative Phase

- Given the training set for a **single object** (like tree)
- The classifier is given many samples of **positive and negative feature vectors** containing color, texture, and structure features all together.
- Yi trained a **neural net** classifier for object recognition.

ICPR04 Data Set with General Labels

	EM-variant with single Gaussian per object	EM-variant extension to mixture models	Gen/Dis with Classical EM clustering	Gen/Dis with EM-variant extension
<i>African animal</i>	71.8%	85.7%	89.2%	90.5%
<i>arctic</i>	80.0%	79.8%	90.0%	85.1%
<i>beach</i>	88.0%	90.8%	89.6%	91.1%
<i>grass</i>	76.9%	69.6%	75.4%	77.8%
<i>mountain</i>	94.0%	96.6%	97.5%	93.5%
<i>primate</i>	74.7%	86.9%	91.1%	90.9%
<i>sky</i>	91.9%	84.9%	93.0%	93.1%
<i>stadium</i>	95.2%	98.9%	99.9%	100.0%
<i>tree</i>	70.7%	79.0%	87.4%	88.2%
<i>water</i>	82.9%	82.3%	83.1%	82.4%
MEAN	82.6%	85.4%	89.6%	89.3%



Comparison to ALIP: the Benchmark Image Set

- Test database used in SIMPLicity paper and ALIP paper.
- 10 classes (*African people, beach, buildings, buses, dinosaurs, elephants, flowers, food, horses, mountains*). 100 images each.



Comparison to ALIP: the Benchmark Image Set

	ALIP	cs	ts	st	ts+st	cs+st	cs+ts	cs+ts+st
<i>African</i>	52	69	23	26	35	79	72	74
<i>beach</i>	32	44	38	39	51	48	59	64
<i>buildings</i>	64	43	40	41	67	70	70	78
<i>buses</i>	46	60	72	92	86	85	84	95
<i>dinosaurs</i>	100	88	70	37	86	89	94	93
<i>elephants</i>	40	53	8	27	38	64	64	69
<i>flowers</i>	90	85	52	33	78	87	86	91
<i>food</i>	68	63	49	41	66	77	84	85
<i>horses</i>	60	94	41	50	64	92	93	89
<i>mountains</i>	84	43	33	26	43	63	55	65
MEAN	63.6	64.2	42.6	41.2	61.4	75.4	76.1	80.3

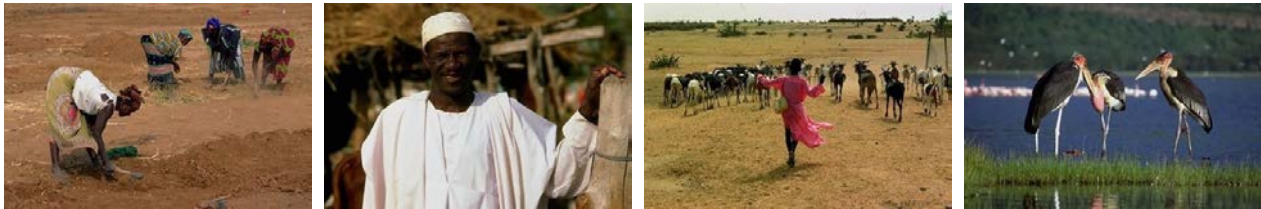


Comparison to ALIP: the 60K Image Set

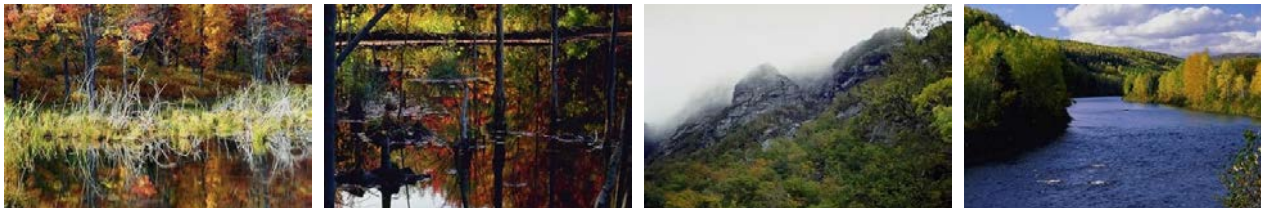
- 59,895 COREL images and 599 categories;
- Each category has about 100 images;
- 8 images per category were reserved for testing.
- To train on one category, all the available 92 positive images were used find the clusters. Those positive images, along with 1,000 randomly selected negative images were then used to train the MLPs.

Comparison to ALIP: the 60K Image Set

0. Africa, people, landscape, animal



1. autumn, tree, landscape, lake



2. Bhutan, Asia, people, landscape, church



Comparison to ALIP: the 60K Image Set

3. California, sea, beach, ocean, flower



4. Canada, sea, boat, house, flower, ocean



5. Canada, west, mountain, landscape, cloud, snow, lake





Comparison to ALIP: the 60K Image Set

Number of top-ranked categories required	1	2	3	4	5
ALIP	11.88	17.06	20.76	23.24	26.05
Gen/Dis	11.56	17.65	21.99	25.06	27.75

The table shows the percentage of test images whose true categories were included in the top-ranked categories.



Groundtruth Data Set

- UW Ground truth database (1224 images)
- 31 elementary object categories: *river* (30), *beach* (31), *bridge* (33), *track* (35), *pole* (38), *football field* (41), *frozen lake* (42), *lantern* (42), *husky stadium* (44), *hill* (49), *cherry tree* (54), *car* (60), *boat* (67), *stone* (70), *ground* (81), *flower* (85), *lake* (86), *sidewalk* (88), *street* (96), *snow* (98), *cloud* (119), *rock* (122), *house* (175), *bush* (178), *mountain* (231), *water* (290), *building* (316), *grass* (322), *people* (344), *tree* (589), *sky* (659)
- 20 high-level concepts: *Asian city*, *Australia*, *Barcelona*, *campus*, *Cannon Beach*, *Columbia Gorge*, *European city*, *Geneva*, *Green Lake*, *Greenland*, *Indonesia*, *indoor*, *Iran*, *Italy*, *Japan*, *park*, *San Juans*, *spring flowers*, *Swiss mountains*, and *Yellowstone*.



beach, sky, tree, water



people, street, tree



*building, grass, people,
sidewalk, sky, tree*



*building, bush, sky,
tree, water*



*flower, house, people,
pole, sidewalk, sky*



*flower, grass, house,
pole, sky, street, tree*



*building, flower, sky,
tree, water*



*boat, rock, sky,
tree, water*



building, car, people, tree



car, people, sky

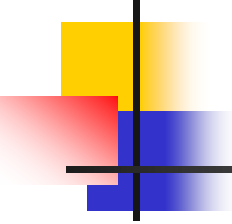


boat, house, water



building

Groundtruth Data Set: ROC Scores



<i>street</i>	60.4	<i>tree</i>	80.8	<i>stone</i>	87.1	<i>columbia gorge</i>	94.5
<i>people</i>	68.0	<i>bush</i>	81.0	<i>hill</i>	87.4	<i>green lake</i>	94.9
<i>rock</i>	73.5	<i>flower</i>	81.1	<i>mountain</i>	88.3	<i>italy</i>	95.1
<i>sky</i>	74.1	<i>iran</i>	82.2	<i>beach</i>	89.0	<i>swiss moutains</i>	95.7
<i>ground</i>	74.3	<i>bridge</i>	82.7	<i>snow</i>	92.0	<i>sanjuans</i>	96.5
<i>river</i>	74.7	<i>car</i>	82.9	<i>lake</i>	92.8	<i>cherry tree</i>	96.9
<i>grass</i>	74.9	<i>pole</i>	83.3	<i>frozen lake</i>	92.8	<i>indoor</i>	97.0
<i>building</i>	75.4	<i>yellowstone</i>	83.7	<i>japan</i>	92.9	<i>greenland</i>	98.7
<i>cloud</i>	75.4	<i>water</i>	83.9	<i>campus</i>	92.9	<i>cannon beach</i>	99.2
<i>boat</i>	76.8	<i>indonesia</i>	84.3	<i>barcelona</i>	92.9	<i>track</i>	99.6
<i>lantern</i>	78.1	<i>sidewalk</i>	85.7	<i>geneva</i>	93.3	<i>football field</i>	99.8
<i>australia</i>	79.7	<i>asian city</i>	86.7	<i>park</i>	94.0	<i>husky stadium</i>	100.0
<i>house</i>	80.1	<i>european city</i>	87.0	<i>spring flowers</i>	94.4		

Groundtruth Data Set: Top Results

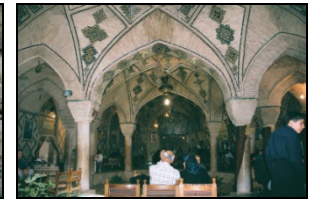
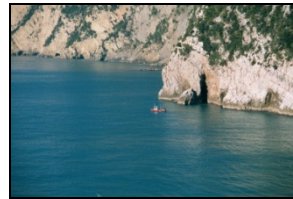
Asian city



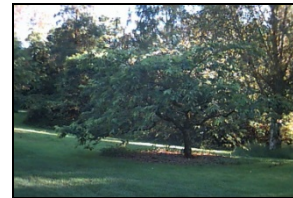
Cannon beach



Italy



park



Groundtruth Data Set: Top Results

sky



spring flowers



tree



water



Groundtruth Data Set: Annotation Samples



tree(97.3),
bush(91.6),
spring flowers(90.3),
flower(84.4),
park(84.3),
sidewalk(67.5),
grass(52.5),
pole(34.1)



sky(99.8),
Columbia gorge(98.8),
lantern(94.2), **street**(89.2),
house(85.8), **bridge**(80.8),
car(80.5), **hill**(78.3),
boat(73.1), **pole**(72.3),
water(64.3), **mountain**(63.8),
building(9.5)



sky(95.1), **Iran**(89.3),
house(88.6),
building(80.1),
boat(71.7), **bridge**(67.0),
water(13.5), **tree**(7.7)



Italy(99.9), **grass**(98.5),
sky(93.8), **rock**(88.8),
boat(80.1), **water**(77.1),
Iran(64.2), **stone**(63.9),
bridge(59.6), **European**(56.3),
sidewalk(51.1), **house**(5.3)



Comparison to Fergus and to Dorko/Schmid using their Features

Using their features and image sets, we compared our generative / discriminative approach to those of Fergus and Dorko/Schmid.

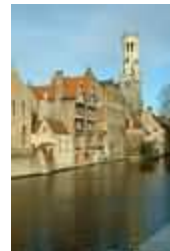
The image set contained 1074 airplane images, 826 motor bike images, 450 face images, and 900 background. Half were used to train and half to test. We added half the background images to the training set for our negative examples.

	Fergus	Dorko/Schmid	Ours
airplanes	90.2%	96.0%	96.6%
faces	96.4%	96.8%	96.5%
motorbikes	92.5%	98.0%	99.2%

Structure Feature Experiments

(from other data sets with more manmade structures)

- 1,951 total from freefoto.com
- **bus** (1,013) **house/building** (609) **skyscraper** (329)



Structure Feature Experiments: Area Under the ROC Curves

1. Structure (with color pairs)

- Attributes (10)
 - Color pair
 - Number of lines
 - Orientation of lines
 - Line overlap
 - Line intersection

2. Structure (with color pairs) + Color Segmentation

3. Structure (without color pairs) + Color Segmentation

	bus	house/ building	skyscraper
Structure only	0.900	0.787	0.887
Structure + Color Seg	0.924	0.853	0.926
Structure ² + Color Seg	0.940	0.860	0.919