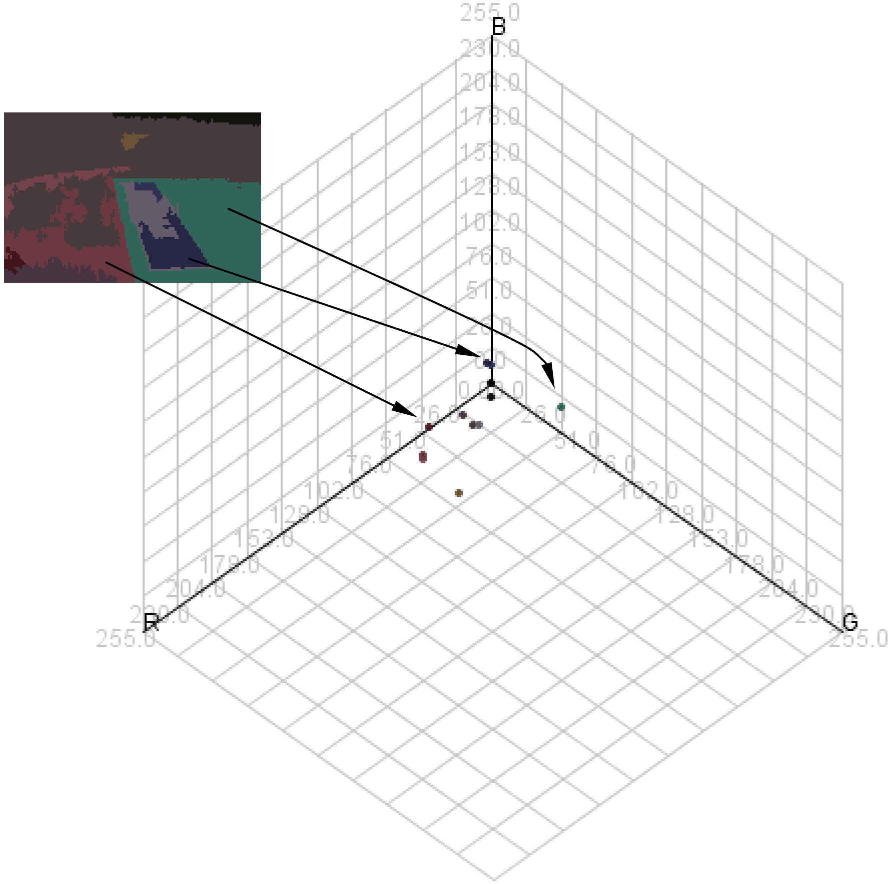
A Generative/Discriminative Learning Algorithm for Image Classification

Yi Li, Linda Shapiro and Jeff Bilmes

Department of Computer Science and Engineering
Department of Electrical Engineering
University of Washington

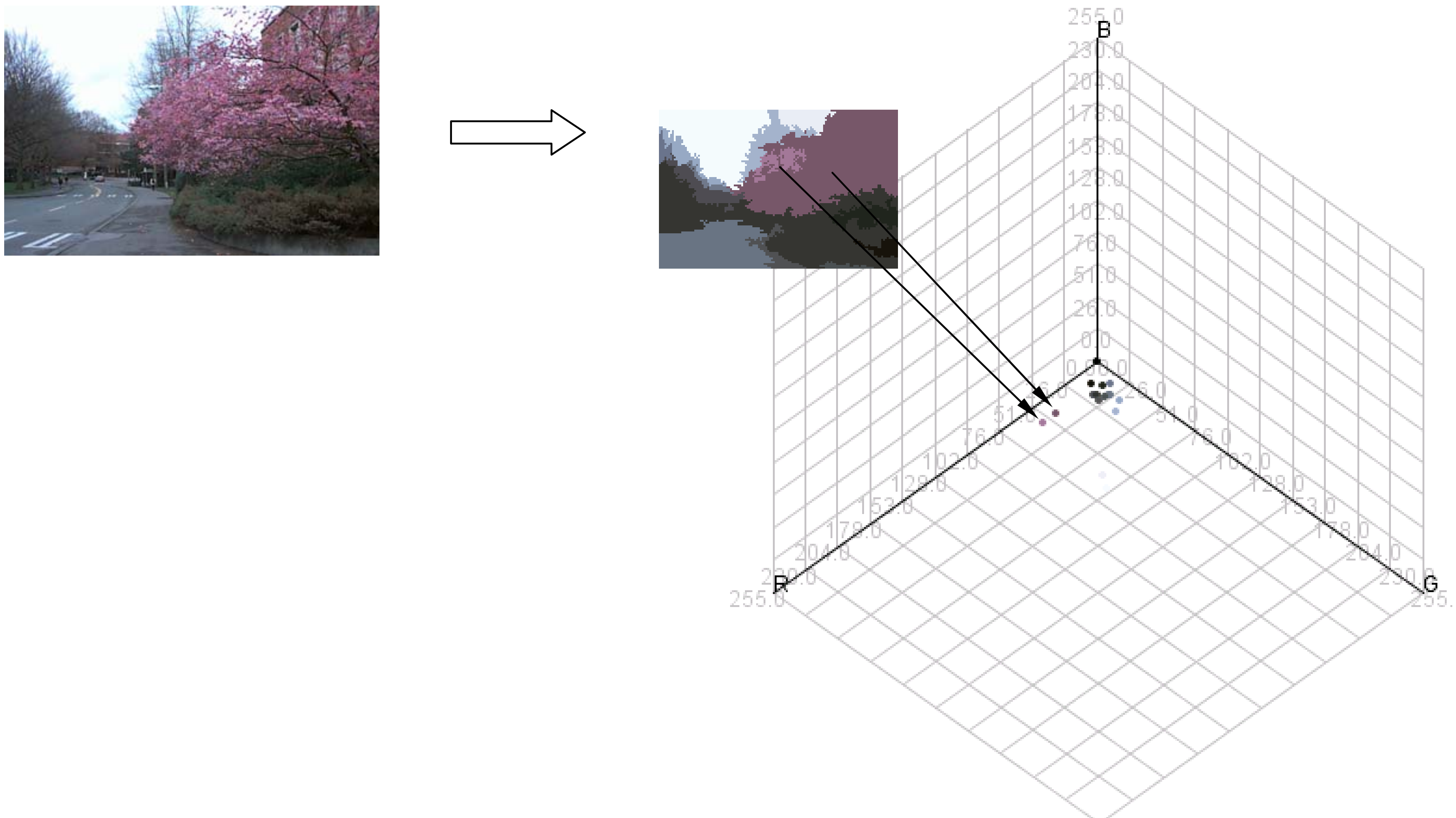
April 2005

Motivation
Scenario one: concept learning



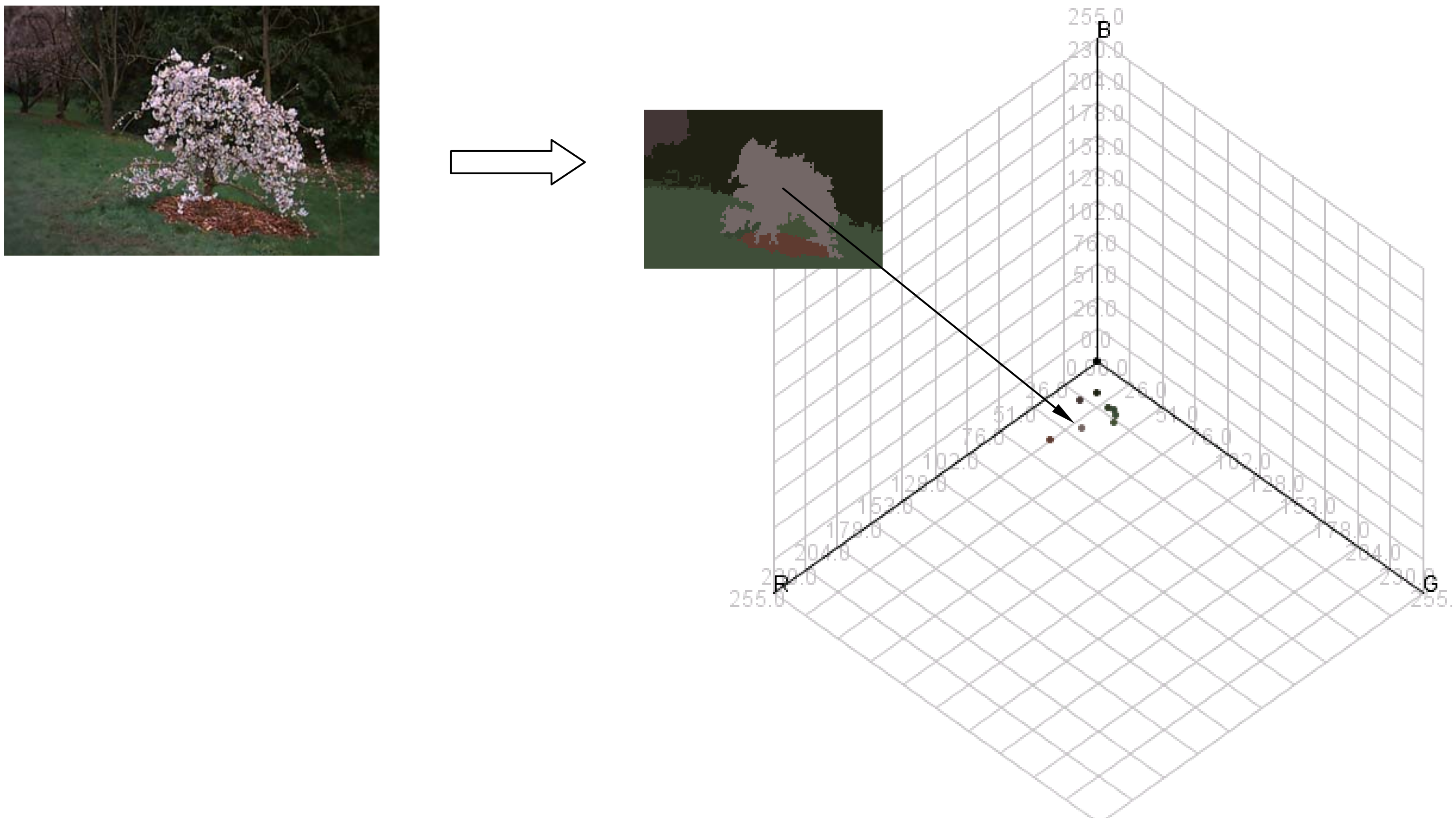
Motivation

Scenario two: Multiple Appearance

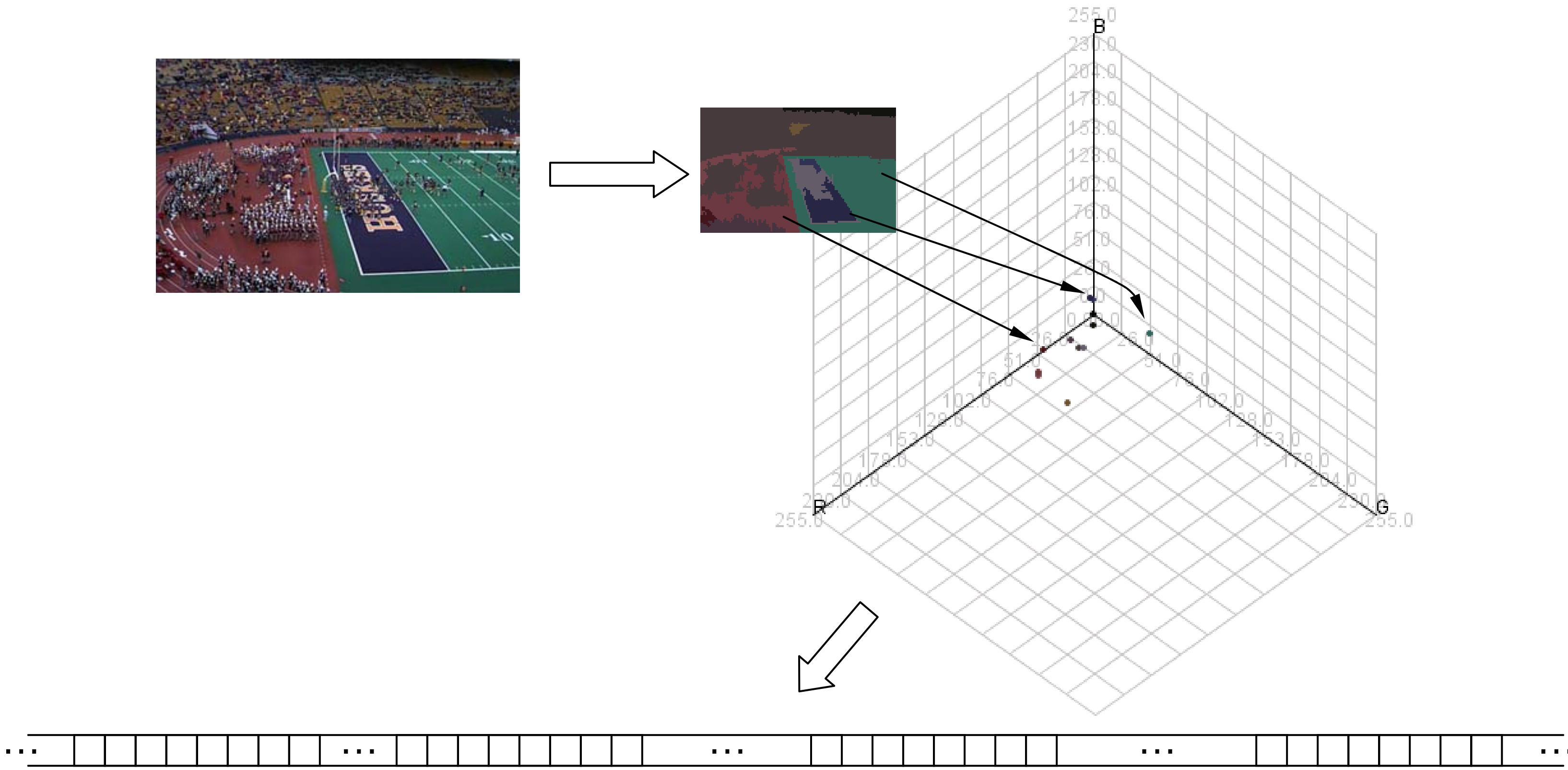


Motivation

Scenario two: Multiple Appearance

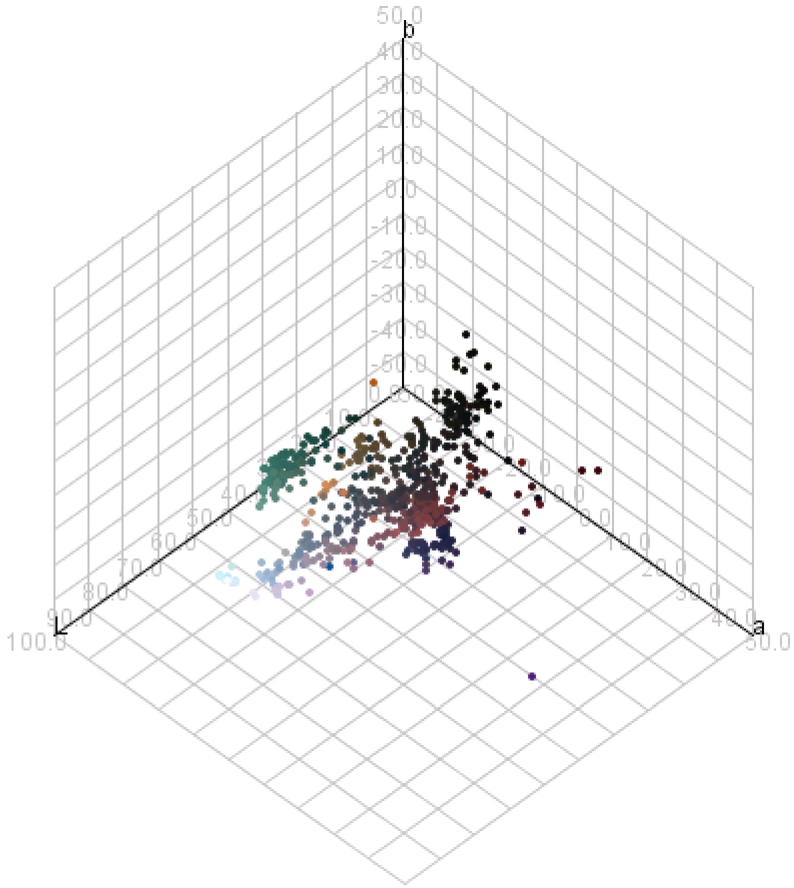


Motivation
Solution One: Histogram

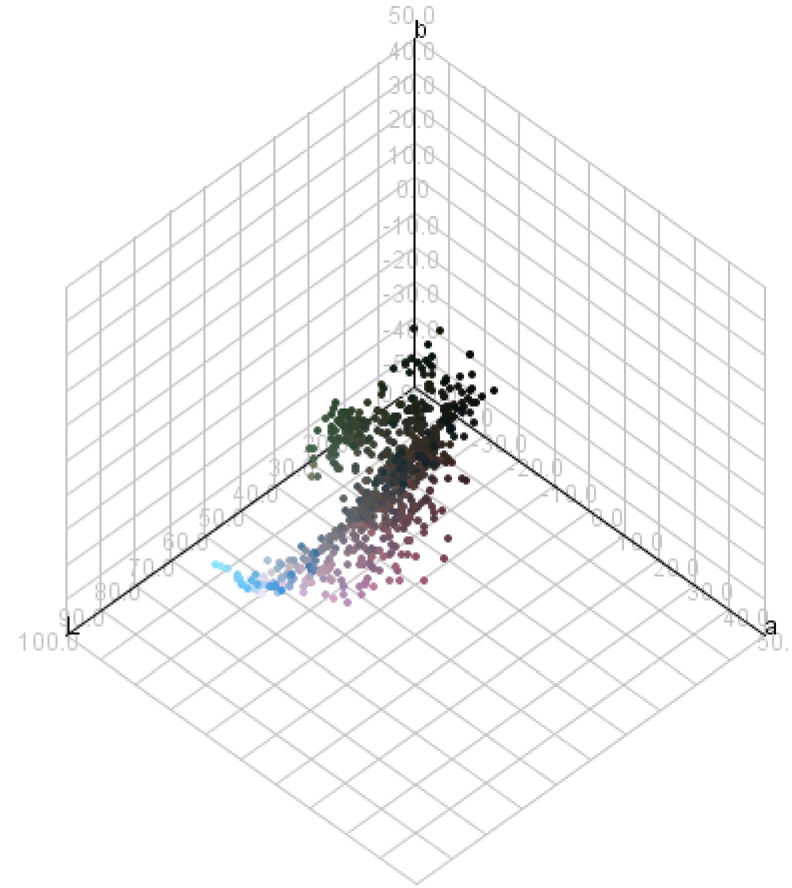


The problem: The Curse of Dimensionality

Our Solution:
Clustering Feature Points

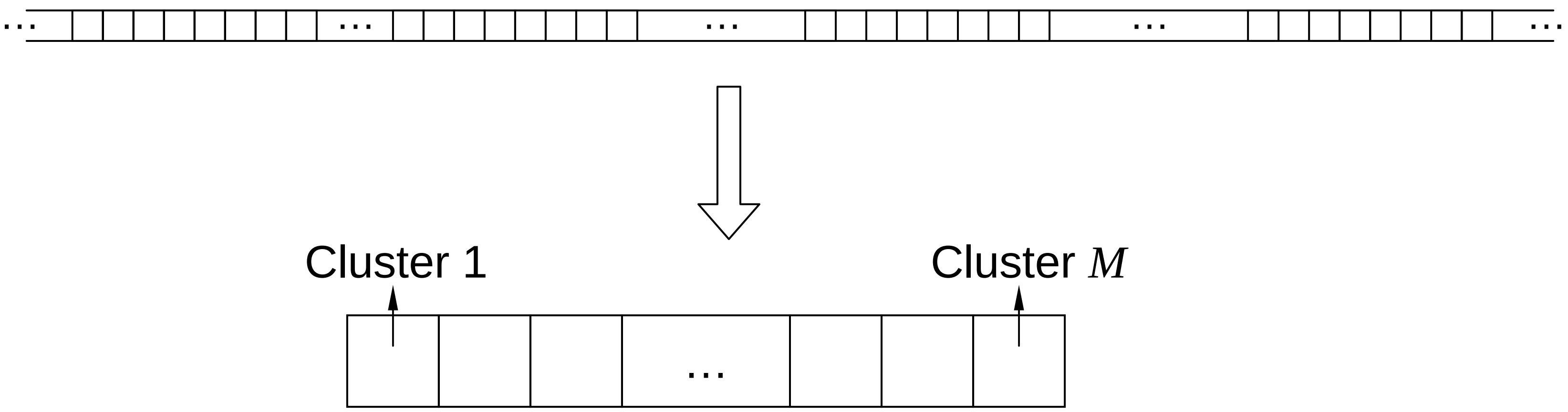


Football Field



Cherry Trees

Our Solution: Clustering Feature Points



- Gaussian Mixture Model is used to cluster the feature vectors

The Generative/Discriminative Approach Combines Different Feature Types

Phase 1: for learning object class o

- Treat each type of abstract region a (color, texture, structure) separately.

- Use the EM algorithm to construct a model that is a mixture of multivariate Gaussians over the features for type a regions.

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

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

- The probability that the feature vector from type-a region ***r*** of image ***I*$_i$** comes from component ***m*** is given by

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

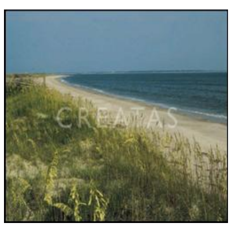
- Then the probability that image ***I*$_i$** has a region that comes from component ***m*** is

$$P(I_i, m^a) = f(\{P(X_{i,r}^a, m^a) | r = 1, 2, \ldots, n_i^a\})$$

where f is the aggregate function.

Aggregate Scores

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

Training the Classifier

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

$$\begin{array}{cc}
 & \begin{array}{cccc} \text{component 1} & \text{component 2} & & \text{component M} \end{array} \\
\begin{array}{c} \text{training} \\ \text{vectors} \end{array} &
\begin{bmatrix}
P(I_1^+, 1^a) & P(I_1^+, 2^a) & \cdots & P(I_1^+, M^a) \\
P(I_2^+, 1^a) & P(I_2^+, 2^a) & \cdots & P(I_2^+, M^a) \\
\vdots & & & \\
P(I_1^-, 1^a) & P(I_1^-, 2^a) & \cdots & P(I_1^-, M^a) \\
P(I_2^-, 1^a) & P(I_2^-, 2^a) & \cdots & P(I_2^-, M^a) \\
\vdots & & &
\end{bmatrix}
\end{array}$$

Phase 2 Learning

- Let $\mathbf{Y}_{I_i}^{1a:Ma}$ be row i of the matrix.

- Each such row is an aggregate feature vector for the type-**a** 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: $Y_{I_i}^{1^c:M^c}$
- Texture: $Y_{I_i}^{1^t:M^t}$
- Structure: $Y_{I_i}^{1^s:M^s}$

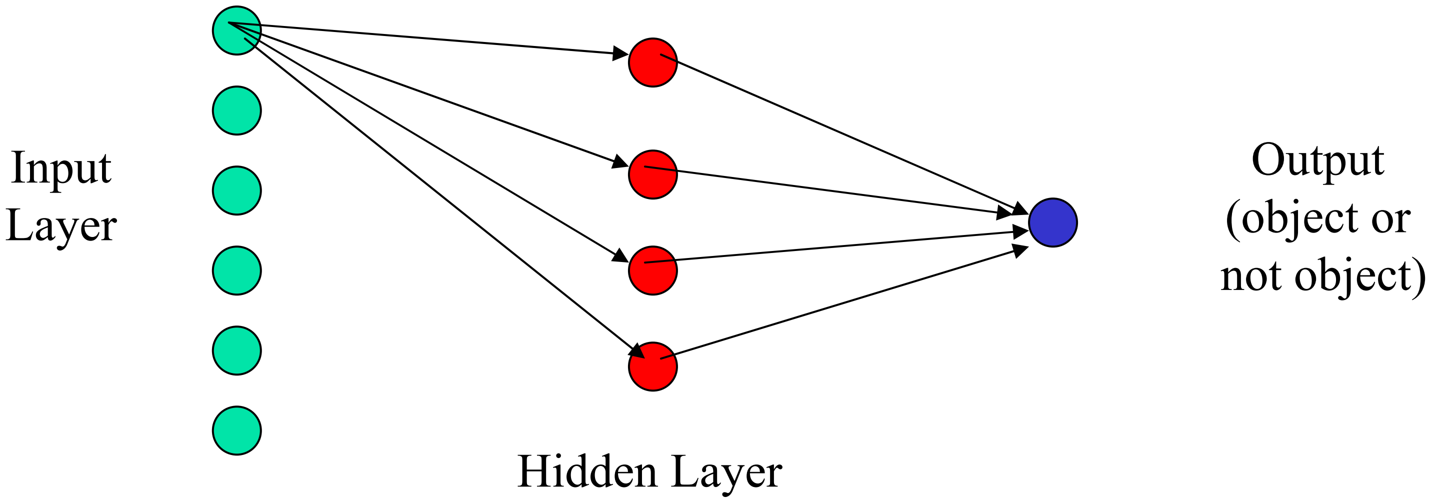
- and any more features we have (motion).

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

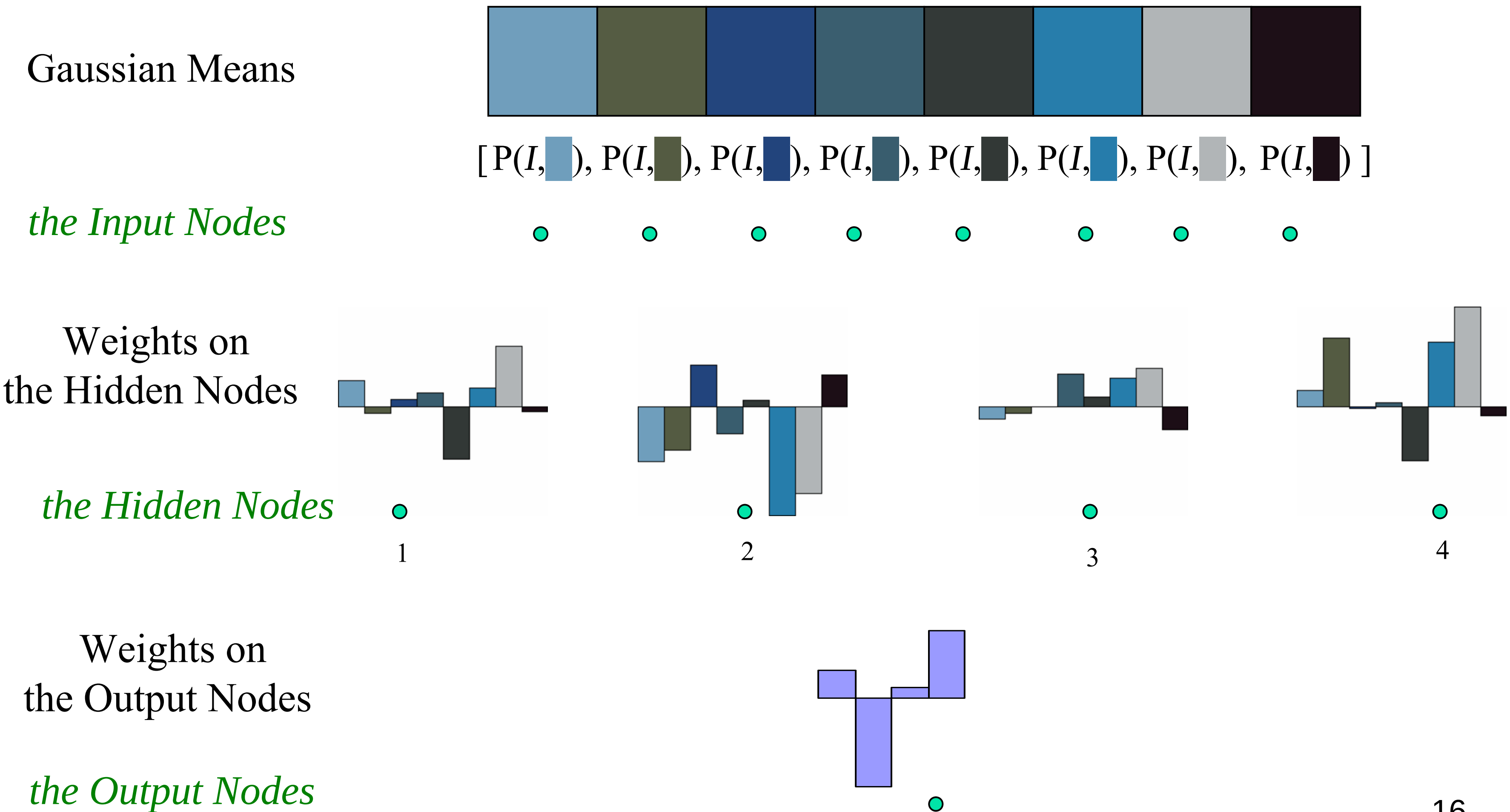
$$
\begin{array}{cccc}
 & color & texture & structure \\
\begin{array}{c} I_1^+ \\ I_2^+ \\ \vdots \\ I_1^- \\ I_2^- \\ \vdots \end{array} &
\begin{bmatrix} \cdots & Y_{I_1^+}^{m^c} & \cdots \\ \cdots & Y_{I_2^+}^{m^c} & \cdots \\ & \vdots & \\ \cdots & Y_{I_1^-}^{m^c} & \cdots \\ \cdots & Y_{I_1^-}^{m^c} & \cdots \\ & \vdots & \end{bmatrix} &
\begin{bmatrix} \cdots & Y_{I_1^+}^{m^t} & \cdots \\ \cdots & Y_{I_2^+}^{m^t} & \cdots \\ & \vdots & \\ \cdots & Y_{I_1^-}^{m^t} & \cdots \\ \cdots & Y_{I_1^-}^{m^t} & \cdots \\ & \vdots & \end{bmatrix} &
\begin{bmatrix} \cdots & Y_{I_1^+}^{m^s} & \cdots \\ \cdots & Y_{I_2^+}^{m^s} & \cdots \\ & \vdots & \\ \cdots & Y_{I_1^-}^{m^s} & \cdots \\ \cdots & Y_{I_1^-}^{m^s} & \cdots \\ & \vdots & \end{bmatrix}
\end{array}
\Longrightarrow
\begin{bmatrix} \cdots & Y_{I_1^+}^{m^c} & \cdots & Y_{I_1^+}^{m^t} & \cdots & Y_{I_1^+}^{m^s} & \cdots \\ \cdots & Y_{I_2^+}^{m^c} & \cdots & Y_{I_2^+}^{m^t} & \cdots & Y_{I_2^+}^{m^s} & \cdots \\ & & & \vdots & & & \\ \cdots & Y_{I_1^-}^{m^c} & \cdots & Y_{I_1^-}^{m^t} & \cdots & Y_{I_1^-}^{m^s} & \cdots \\ \cdots & Y_{I_1^-}^{m^c} & \cdots & Y_{I_1^-}^{m^t} & \cdots & Y_{I_1^-}^{m^s} & \cdots \\ & & & \vdots & & & \end{bmatrix}
$$

Classification

- The training matrix is the input to a multi-layered perceptron that learns to classify new test images as either containing or not containing the object of interest.



Generative/Discriminative Approach: A Model for “beach”



Generative/Discriminative Approach: Experiments

- ICPR04 Data Set with General Labels
- Comparison to ALIP
  - the Benchmark Image Set
  - the 60K Image Set
- Comparison to MT
- Groundtruth Data Set
- Structure Feature Experiments
- VACE Test Image Set
- Comparison to Fergus and Dorko/Schmid

ICPR04 Data Set with General Labels

	EM-variant	EM-variant extension	Gen/Dis with Classical EM	Gen/Dis with EM-variant extension
African animal	71.8%	85.7%	89.2%	90.5%
arctic	80.0%	79.8%	90.0%	85.1%
beach	88.0%	90.8%	89.6%	91.1%
grass	76.9%	69.6%	75.4%	77.8%
mountain	94.0%	96.6%	97.5%	93.5%
primate	74.7%	86.9%	91.1%	90.9%
sky	91.9%	84.9%	93.0%	93.1%
stadium	95.2%	98.9%	99.9%	100.0%
tree	70.7%	79.0%	87.4%	88.2%
water	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
African	52	69	23	26	35	79	72	74
beach	32	44	38	39	51	48	59	64
buildings	64	43	40	41	67	70	70	78
buses	46	60	72	92	86	85	84	95
dinosaurs	100	88	70	37	86	89	94	93
elephants	40	53	8	27	38	64	64	69
flowers	90	85	52	33	78	87	86	91
food	68	63	49	41	66	77	84	85
horses	60	94	41	50	64	92	93	89
mountains	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

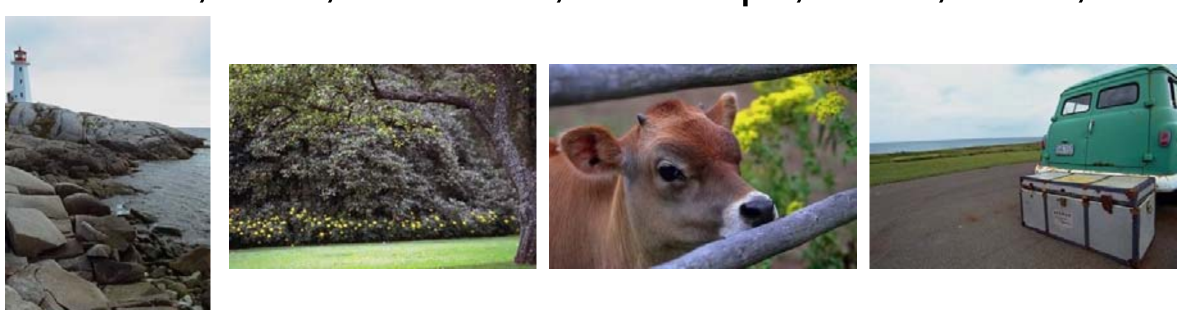
2. Bhutan, Asia, people, landscape, church

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

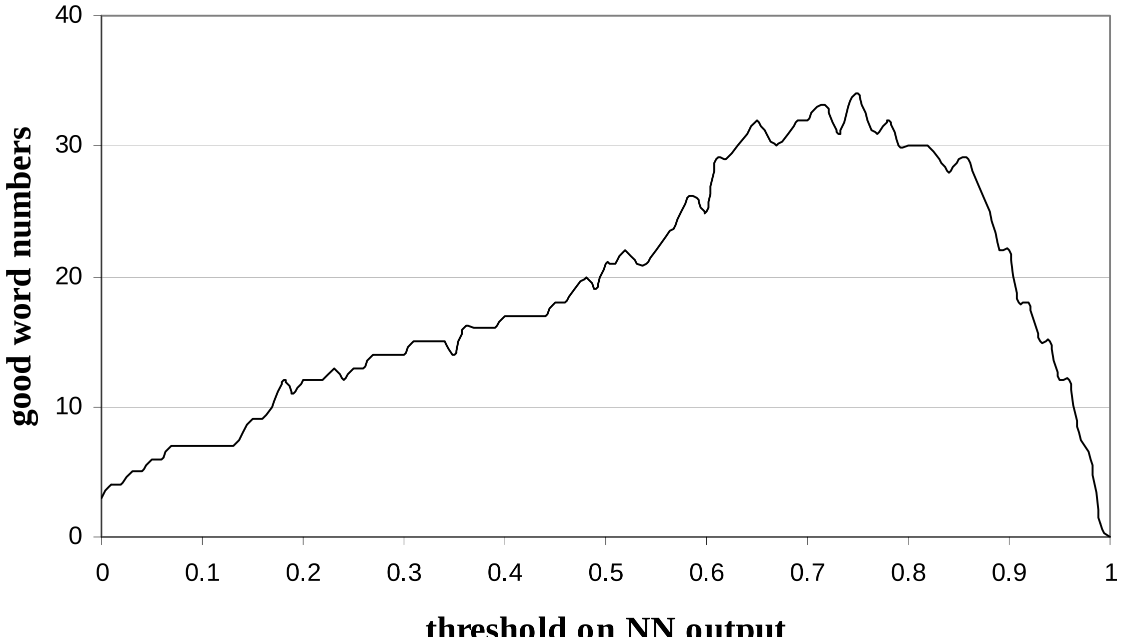
The table shows the percentage of test images whose true categories were included in the top-ranked categories.

Comparison to MT

- Machine Translation (MT) algorithm
  – 33 attributes for each region
- Generative / Discriminative approach
  – 3 color attributes and 12 texture attributes
- The feature vectors of 5000 Corel images were provided.
- 4500 training images and 500 test images.

Comparison to MT: The number of good words

- A word is “good” if its recall value is greater than 0.4 and its precision value is greater than 0.15.)
- MT approach learned 14 “good words”



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

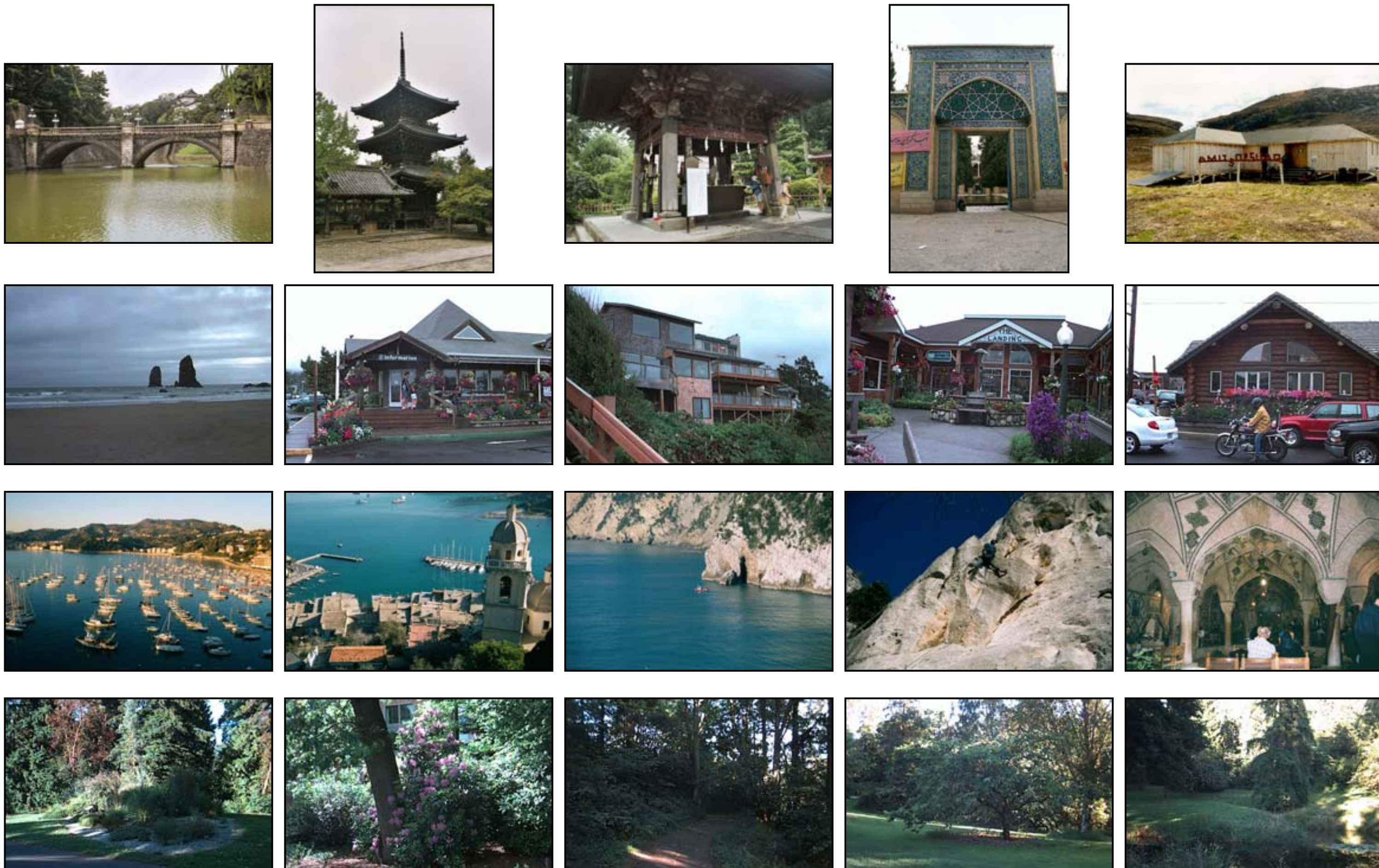
boat, house, water

building

Groundtruth Data Set: ROC Scores

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

Structure Feature Experiments

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

Structure Feature Experiments: ROC Scores

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

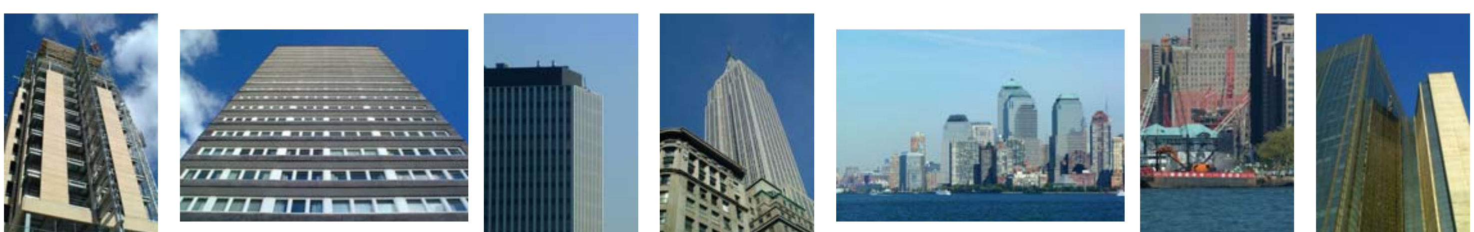
	bus	*house/ building*	*skyscraper*
Structure only	90.0	78.7	88.7
Structure + Color Seg	92.4	85.3	92.6
Structure^2 + Color Seg	94.0	86.0	91.9

Structure Feature Experiments: Top ranked result samples

bus

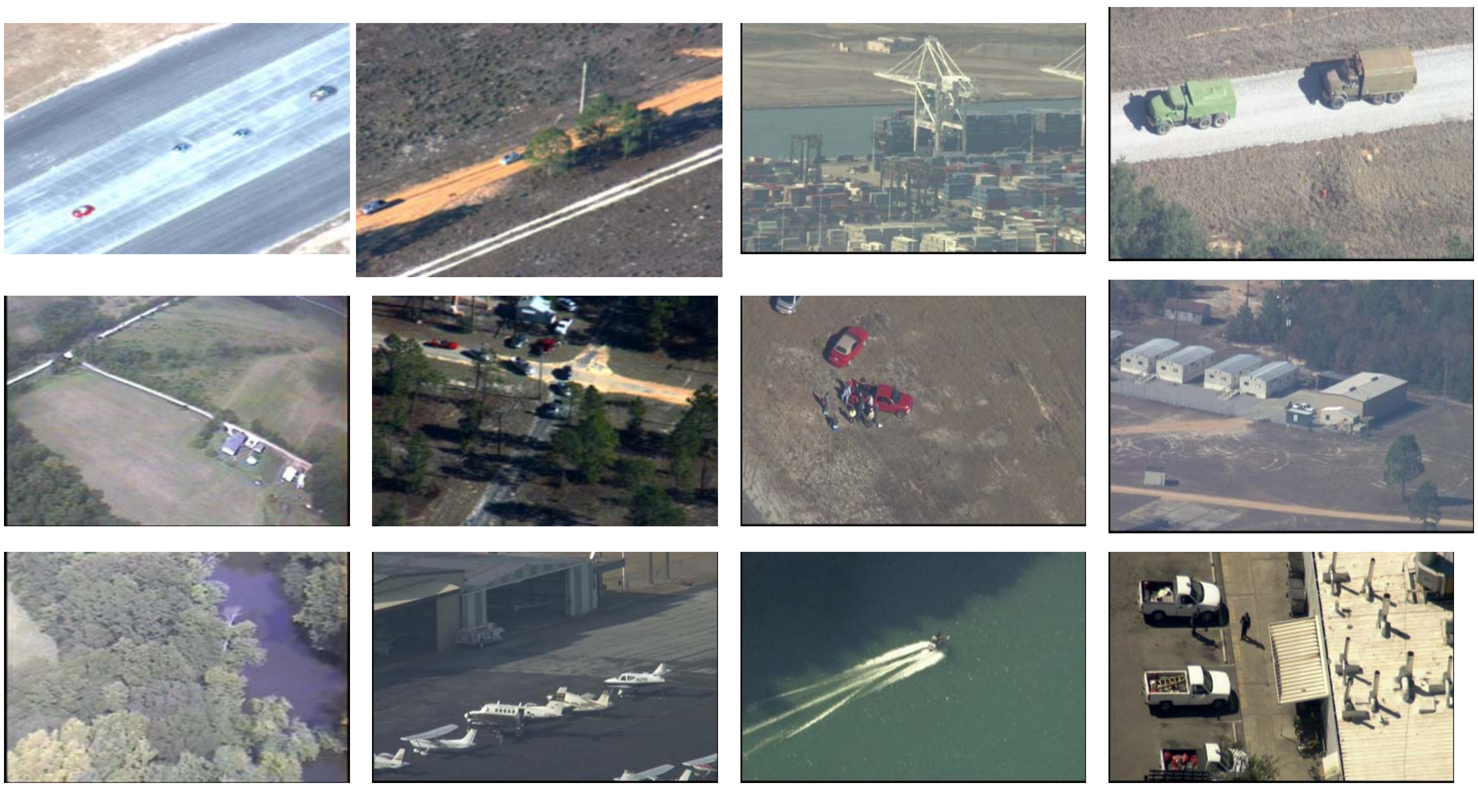
houses and buildings

skyscrapers



VACE Test Image Set

- 828 images and 10 object classes
- from Boeing, VIVID, and NGA videos



VACE Test Image Set: ROC Scores

	airplane	*car*	*dirt road*	*field*	*forest*	*house*	*paved road*	*people*	*runway*	*tree*	**MEAN**
cs	81.2	81.6	86.8	77.2	83.3	82.4	79.9	**83.9**	92.9	77.5	82.7
st	83.5	68.8	70.1	68.2	71.3	78.2	66.9	49.7	80.3	61.0	69.8
cs+st	90.1	78.9	86.4	**77.5**	86.4	83.7	81.5	83.9	93.9	77.5	84.0
cs+ts	78.4	81.1	**89.5**	74.2	86.7	80.8	79.8	83.8	**94.4**	**80.6**	82.9
cs+ts+st	**91.1**	**82.3**	88.1	74.1	**87.6**	**84.9**	**87.5**	79.7	93.6	77.1	84.6

*cs: color seg.   ts: texture seg.   st: structure

Top Results for *airplane*, *dirt road*, *field*, *runway* , and *tree*

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%