04/03/12

Image and Region Categorization

Computer Vision

CS 543 / ECE 549

University of Illinois

Derek Hoiem

Ruiqi Guo

Last classes

- Object instance recognition: localizing an object instance in an image

- Face recognition: matching one face image to another

- Today: mapping images and regions to categories

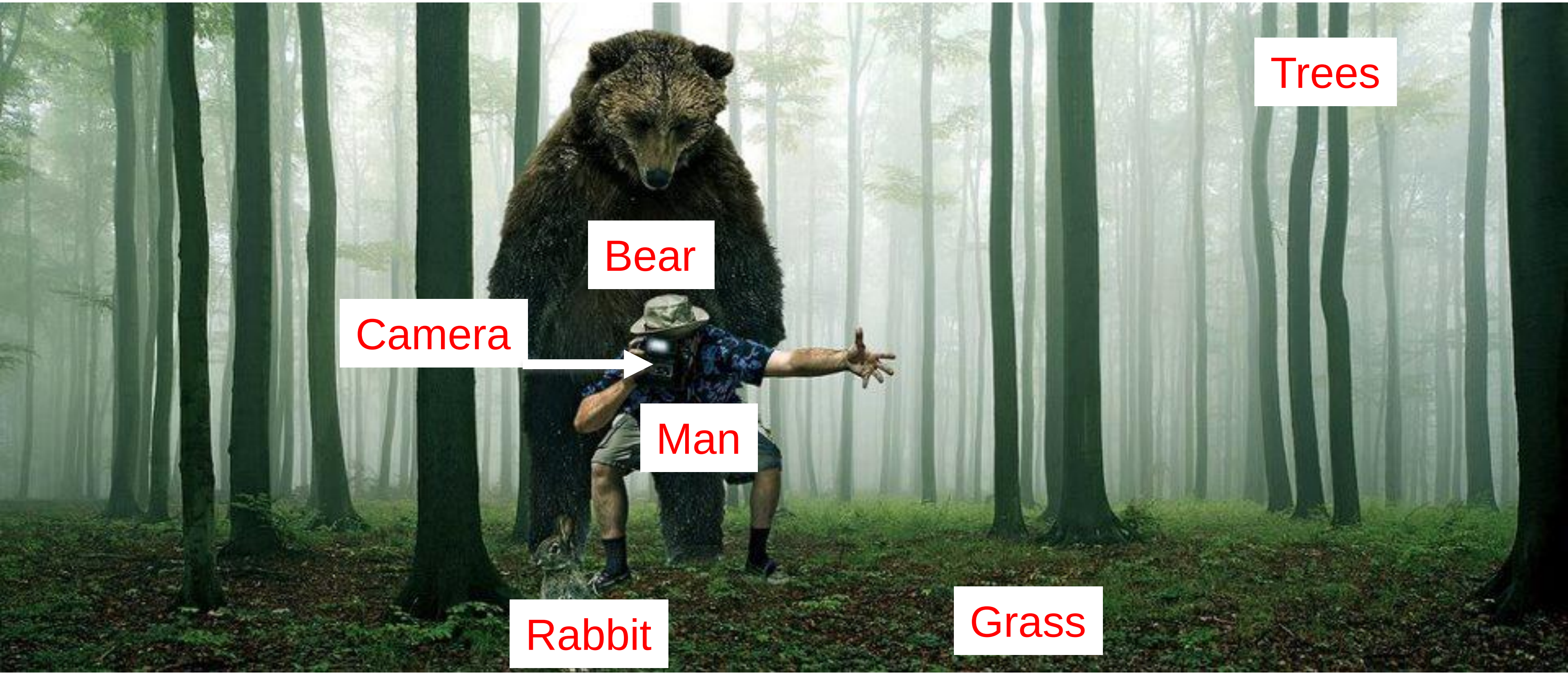
Today's class: image and region categorization

- Overview of image and region categorization
  - Task description
  - What is a category

- Representation
  - Image histograms
  - Bag of Word model
  - Region categorization

- Classifiers

What do you see in this image



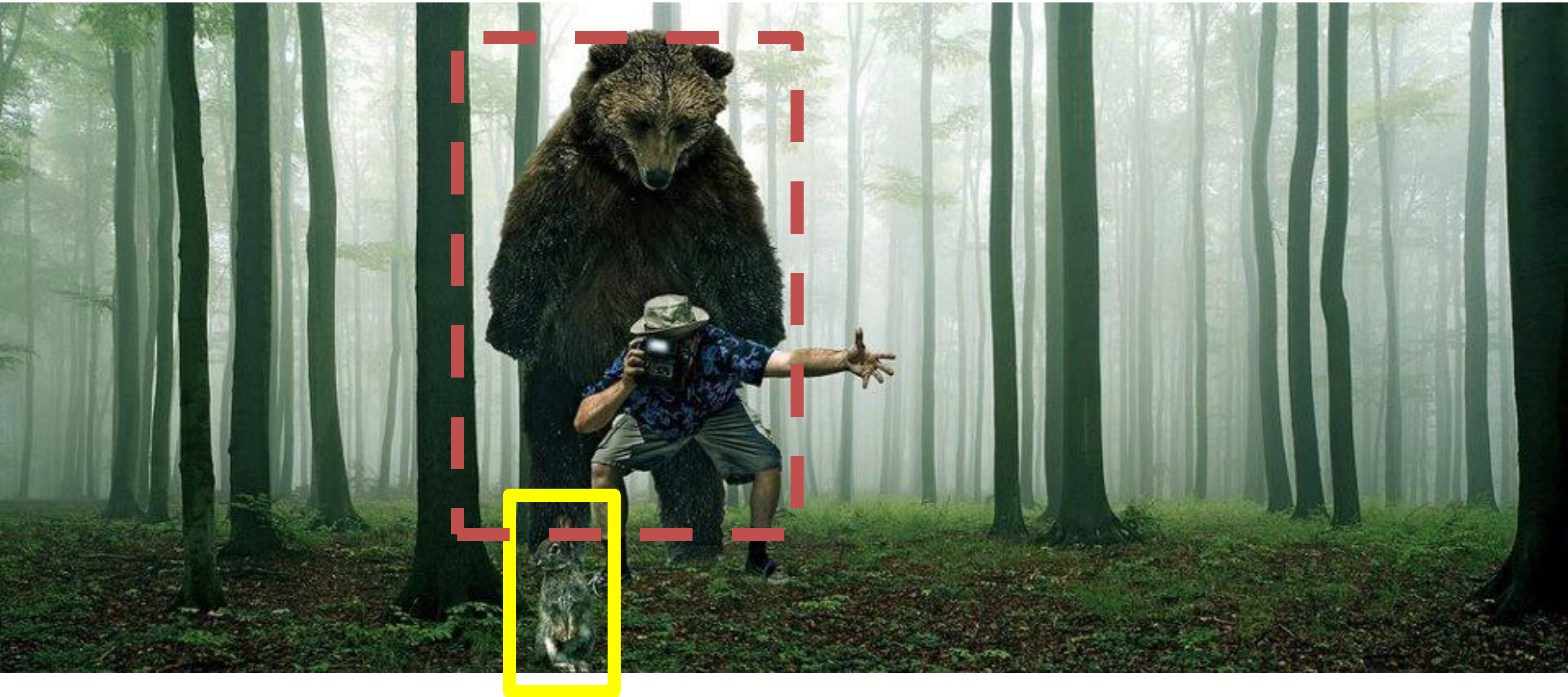
Forest

Why do we care about categories?

From an object's category, we can make predictions about its behavior in the future, beyond of what is immediately perceived.

Slide Credit: Torralba

describe, predict, or interact with the object based on visual cues



Is it **dangerous**?

Is it **alive**?

How **fast** does it run?

Is it **soft**?

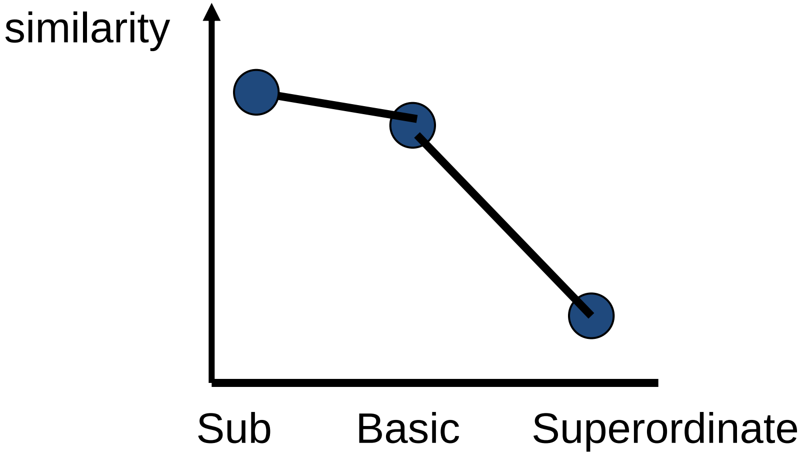
Does it have a **tail**?

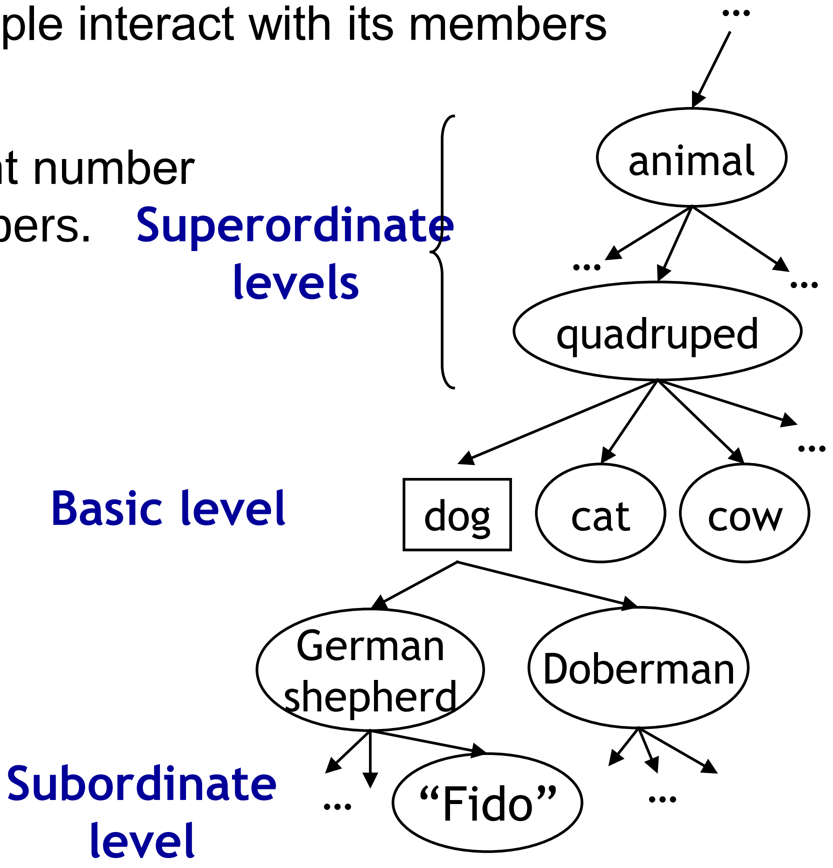
Can I **poke with it**?

Rosch’s (1976) Levels of Categorization

Definition of Basic Level:

- **Similar shape**: Basic level categories are the highest-level category for which their members have similar shapes.
- **Similar motor interactions**: … for which people interact with its members using similar motor sequences.
- **Common attributes**: … there are a significant number of attributes in common between pairs of members.




Superordinate levels

Basic level

Subordinate level

…

animal

…

…

quadruped

…

dog

cat

cow

German shepherd

Doberman

…

“Fido”

…

Levels of Categorization

- Rosch et al found that
  - People can tell whether an object belongs to a basic-level category faster

  - People tend to predict the basic category (e.g., “dog”) before superordinate (“animal”) or subordinate (“golden retriever”) categorie

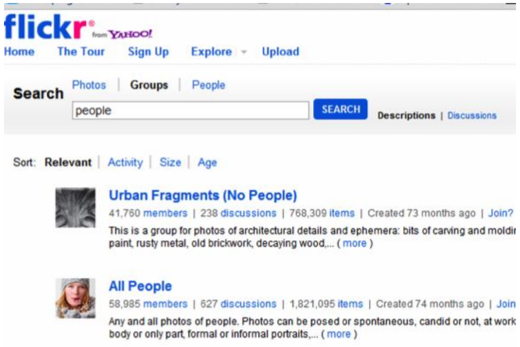
Visual categorization

- Given a small number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label.
- Many different ways to categorize

What type of place

What material

What species of animal



Search: people

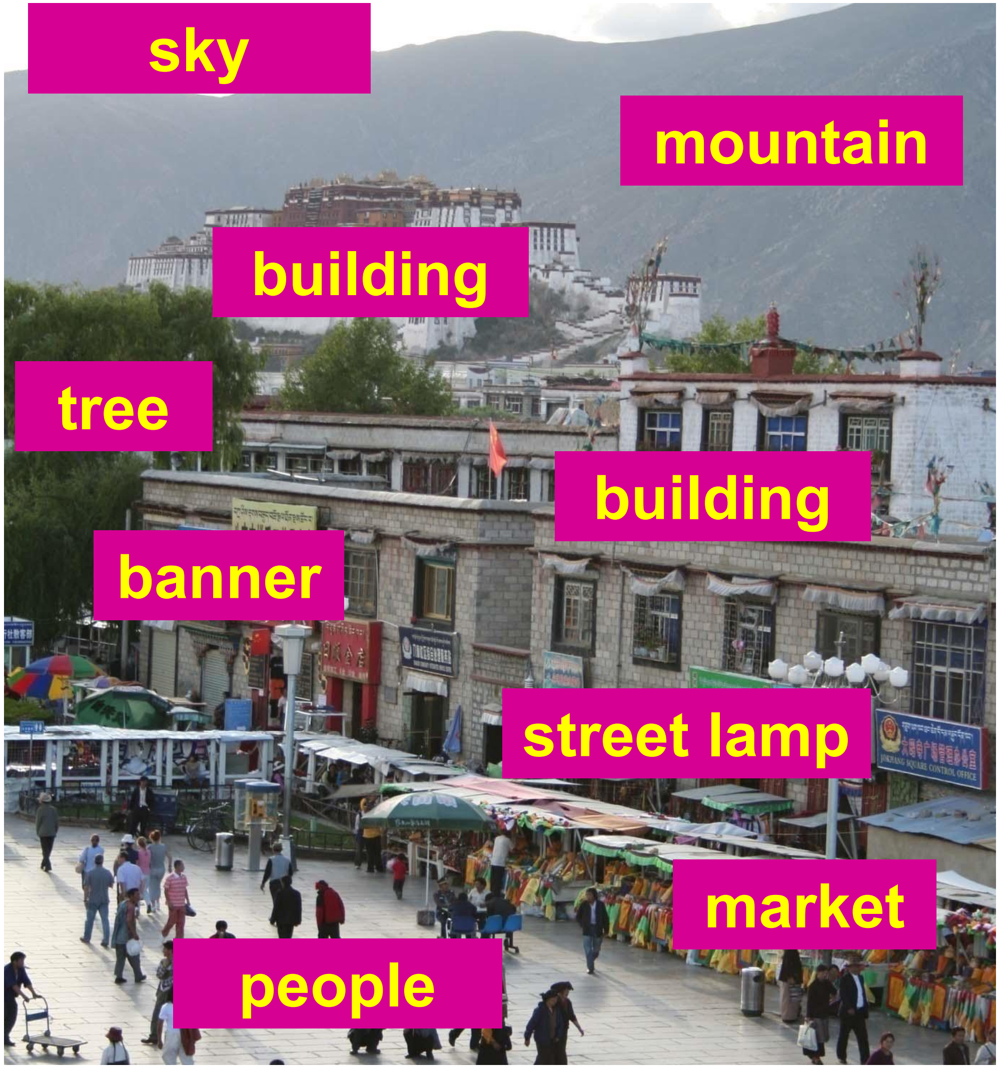
Urban Fragments (No People)
41,760 members | 238 discussions | 768,309 items | Created 73 months ago | Join?
This is a group for photos of architectural details and ephemera: bits of carving and moldi
paint, rusty metal, old brickwork, decaying wood,... (more)

All People
58,985 members | 627 discussions | 1,821,095 items | Created 74 months ago | Join
Any and all photos of people. Photos can be posed or spontaneous, candid or not, at work
body or only part, formal or informal portraits,... (more)

What textual tags

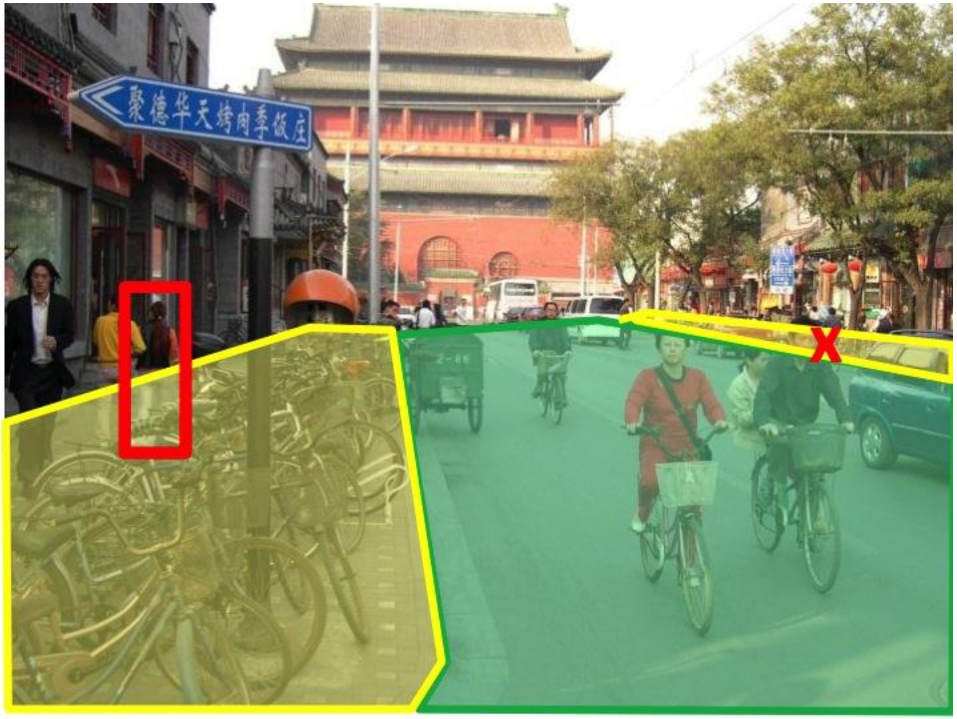
Region categorization

- Categorize each image regions
- Image parsing
- Semantic scene labeling



Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

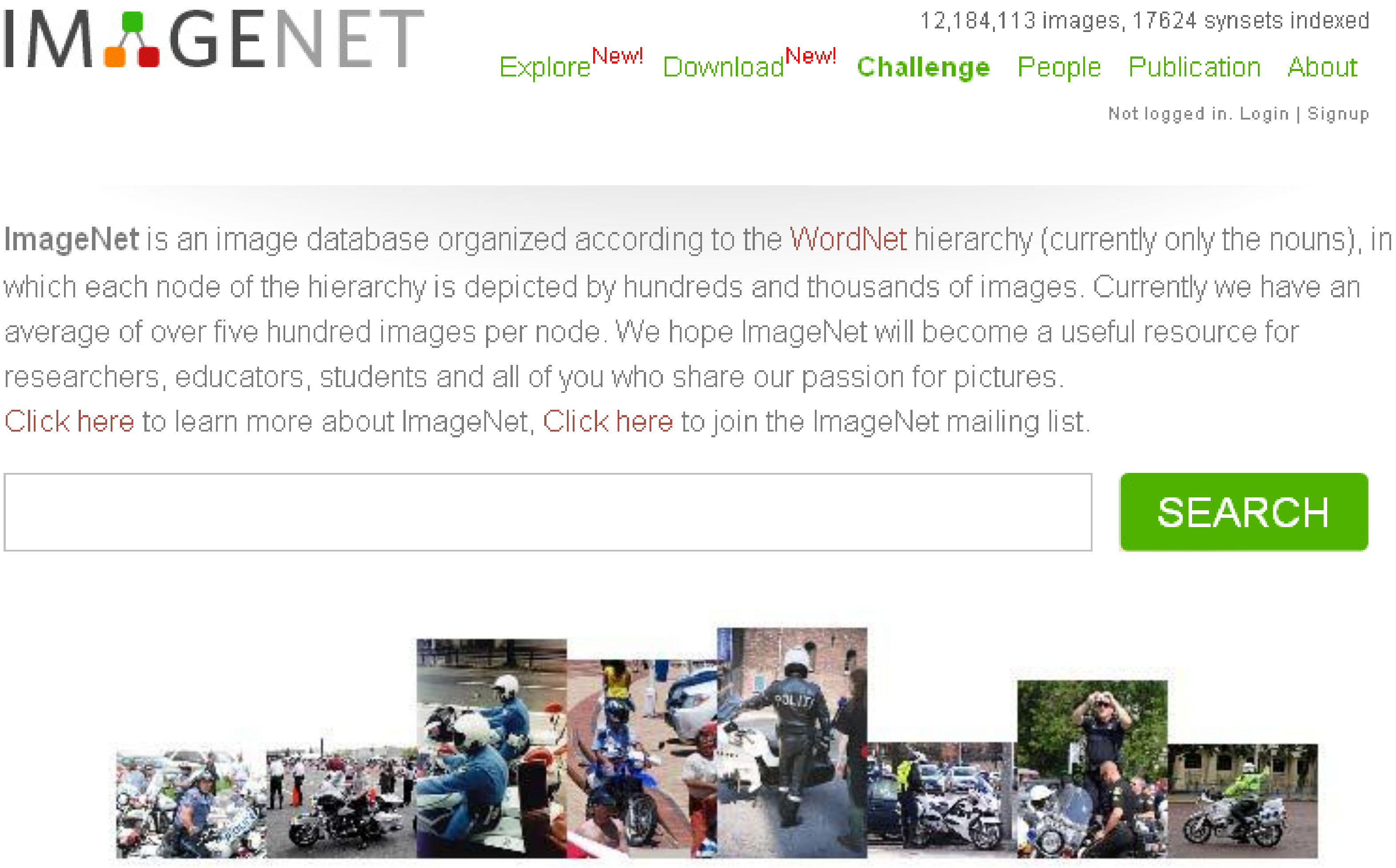
Inference about the scene



Is this person on the sidewalk?

How to get here?

Image categorization

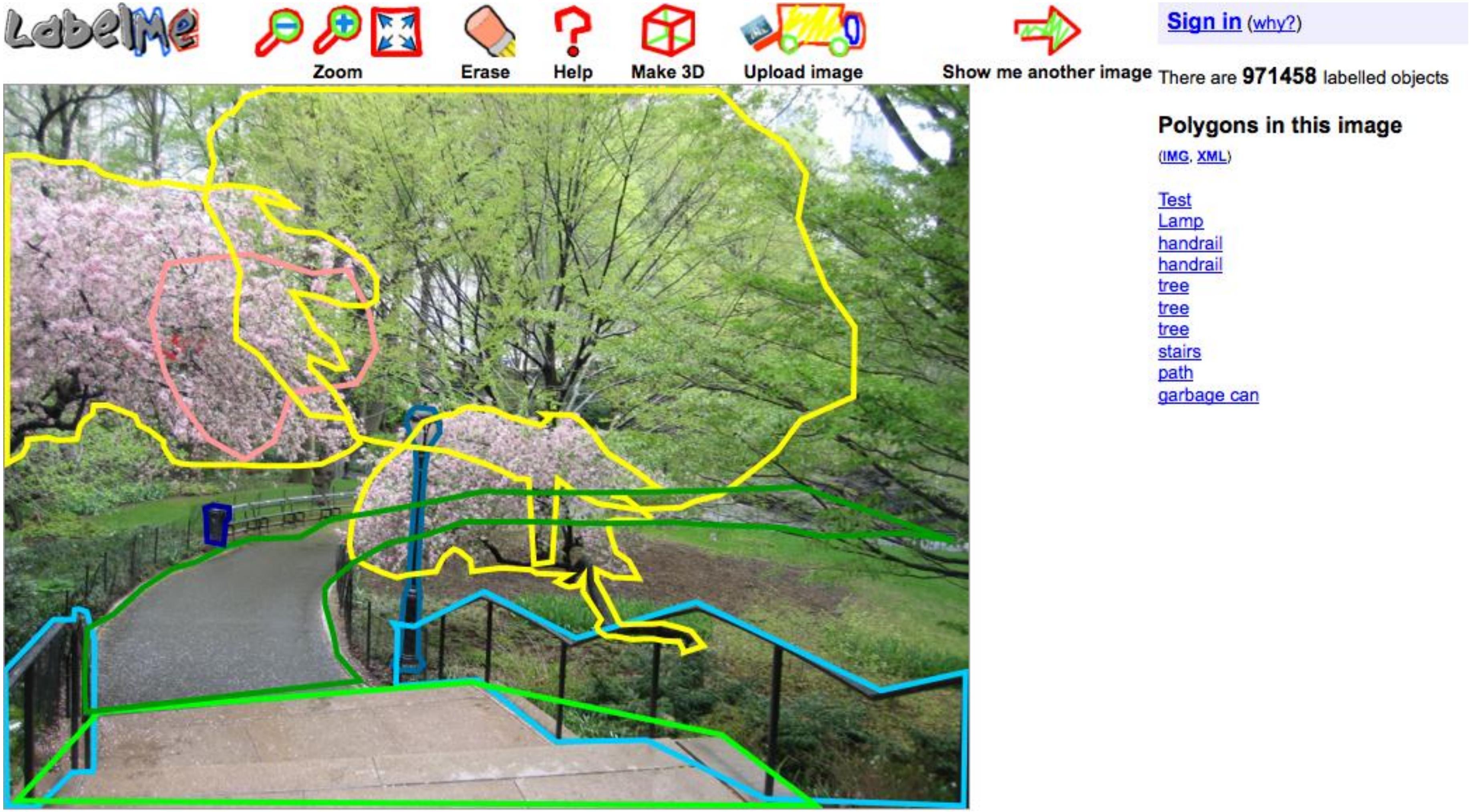


IMAGENET
12,184,113 images, 17624 synsets indexed
ExploreNew!
DownloadNew!
Challenge
People
Publication
About
Not logged in. Login | Signup
SEARCH

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.

Region categorization



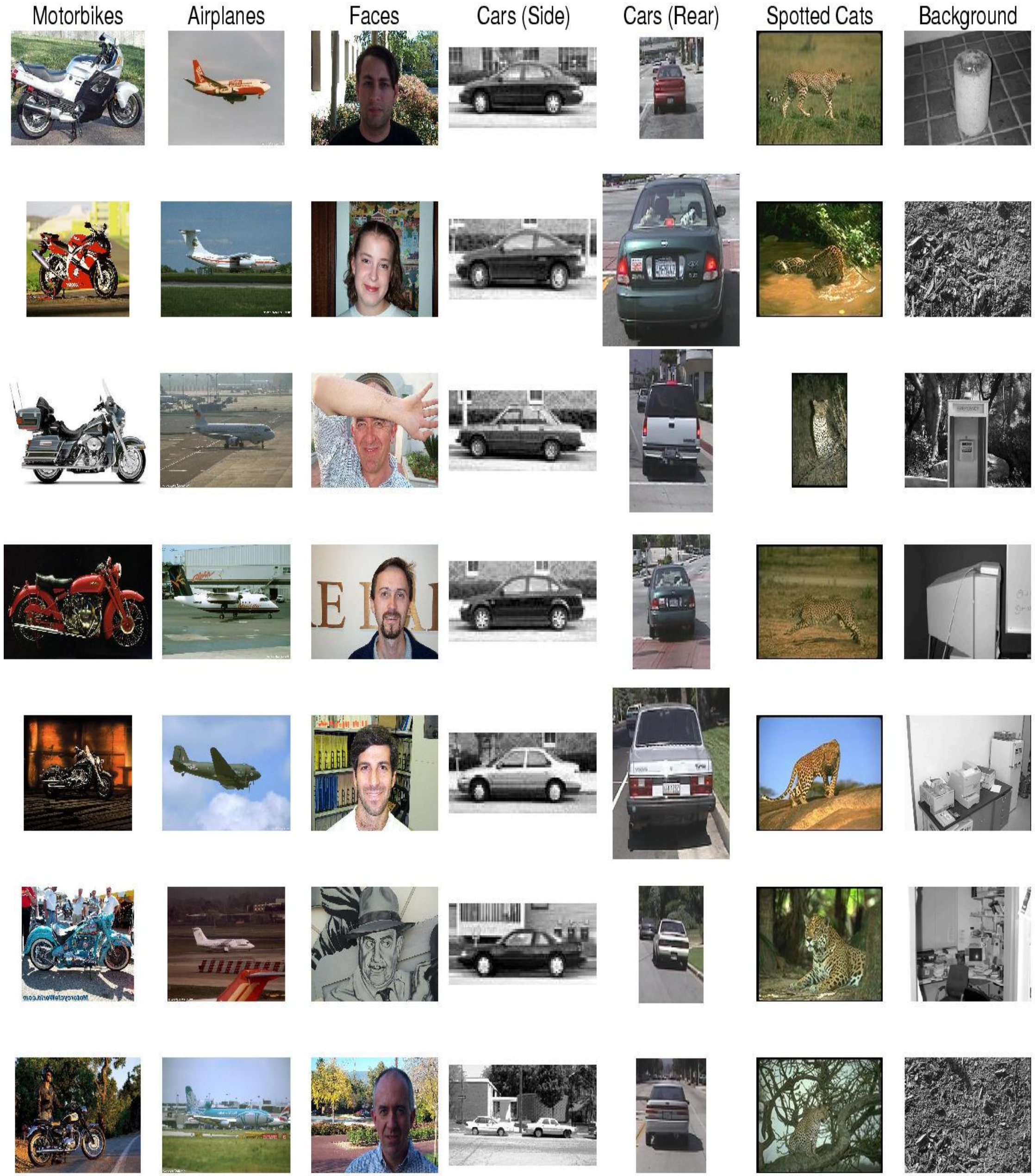
There are **971458** labelled objects

Polygons in this image

(IMG, XML)

Test
Lamp
handrail
handrail
tree
tree
tree
stairs
path
garbage can

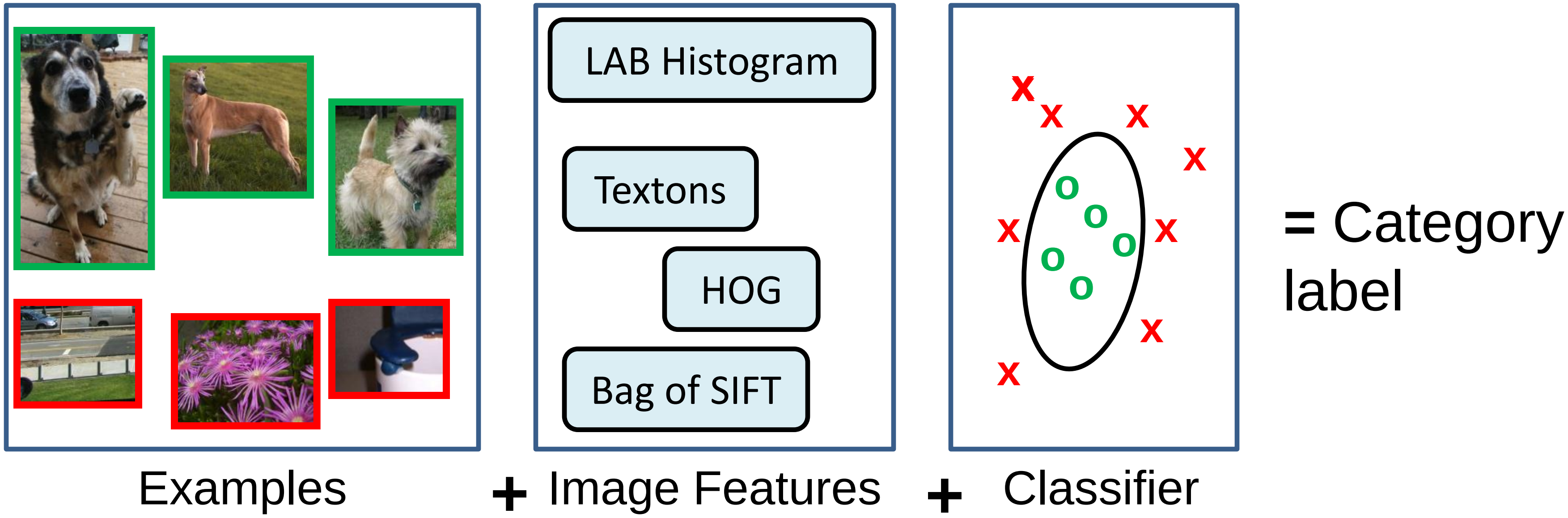
Categorization in computer vision



Difficulties:

- In class variation
- Similar looking categories
- Size variation
- Background clutter
- Occlusion

Supervised learning



LAB Histogram

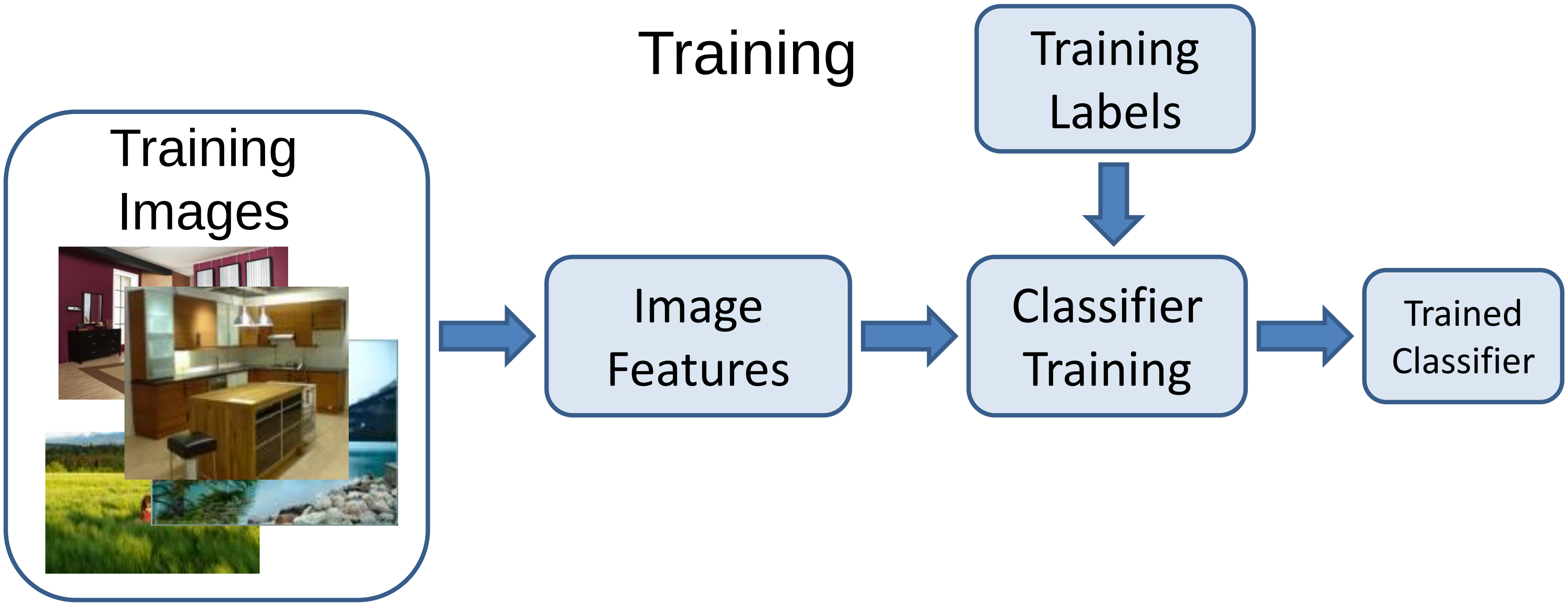
Textons

HOG

Bag of SIFT

Examples + Image Features + Classifier = Category label

Training phase



Training

Training Images

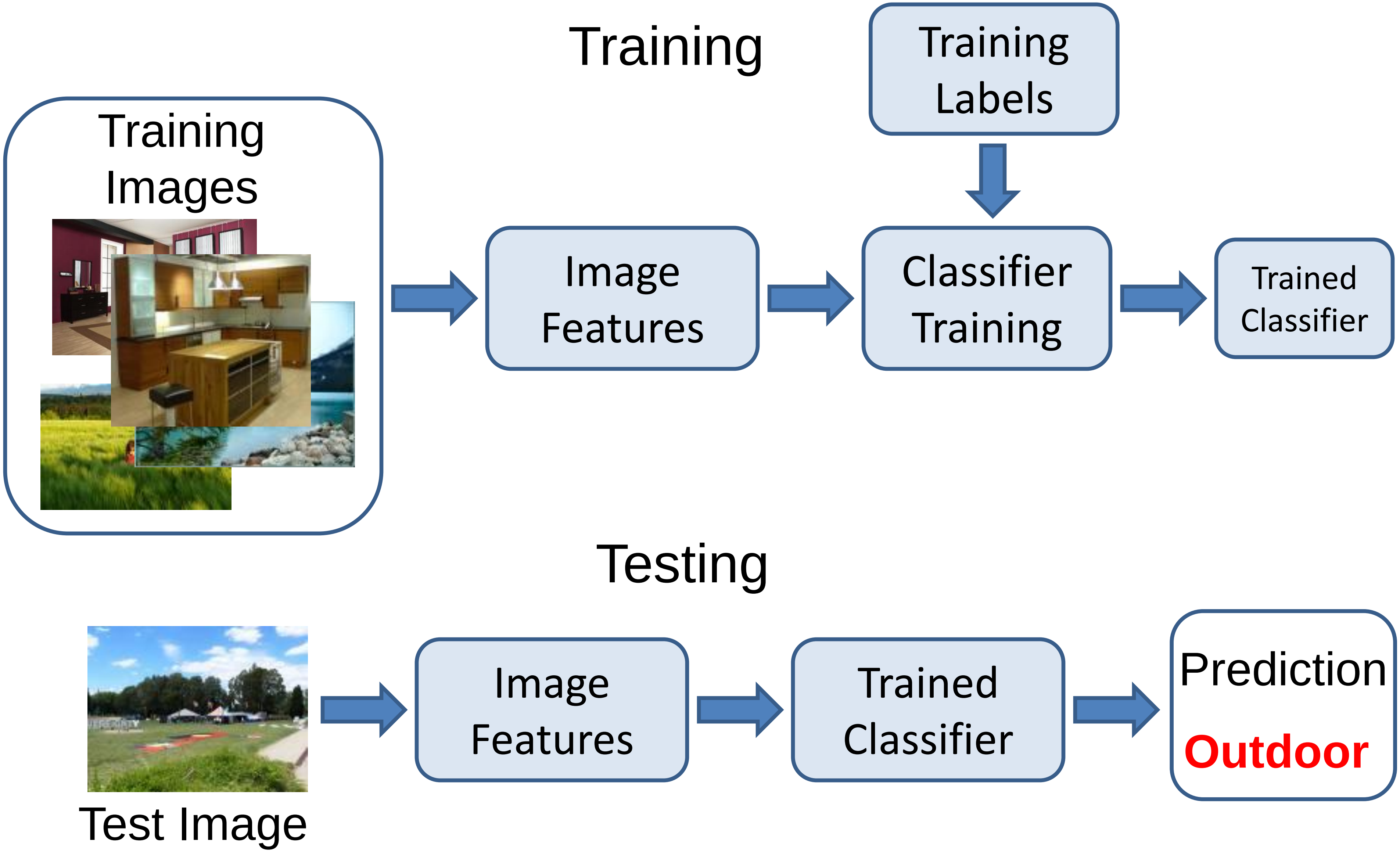
Image Features

Training Labels

Classifier Training

Trained Classifier

Testing phase



Training

Training Labels

Training Images

Image Features

Classifier Training

Trained Classifier

Testing

Test Image

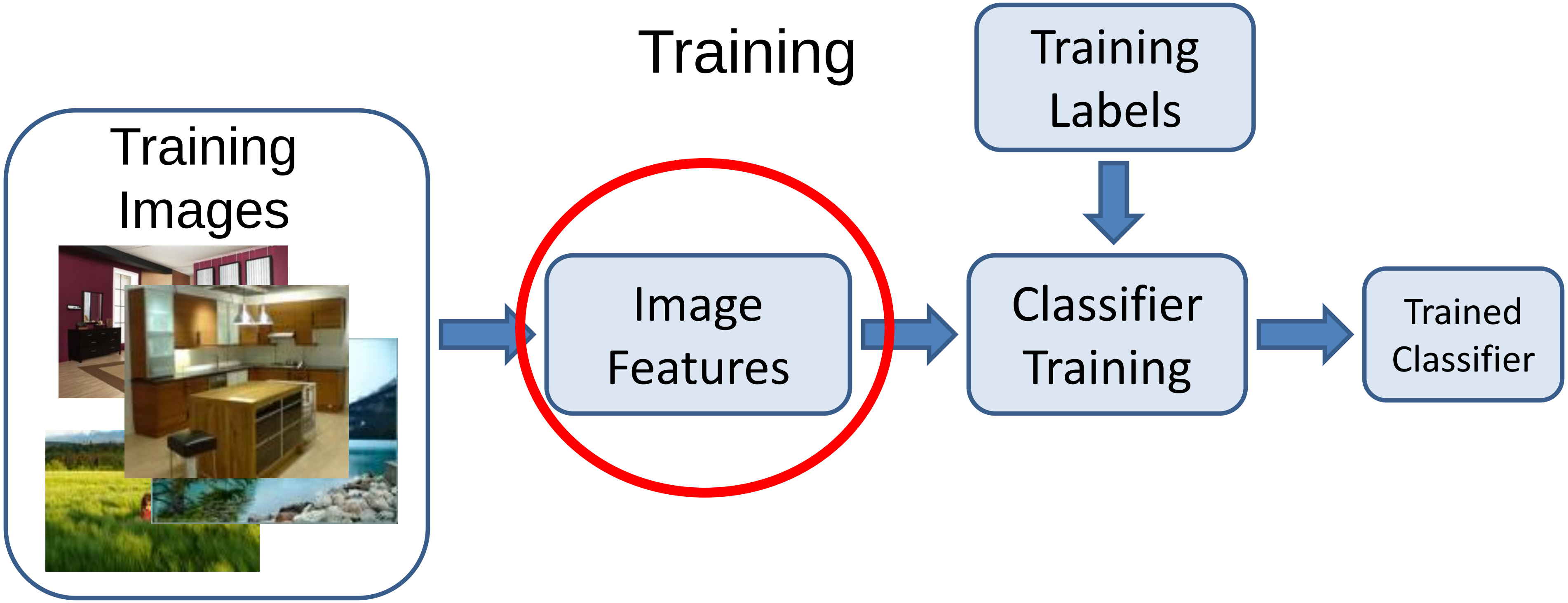
Image Features

Trained Classifier

Prediction

Outdoor

Part I: Image features



Training

Training Images

Image Features

Training Labels

Classifier Training

Trained Classifier

Right features depend on what you want to know

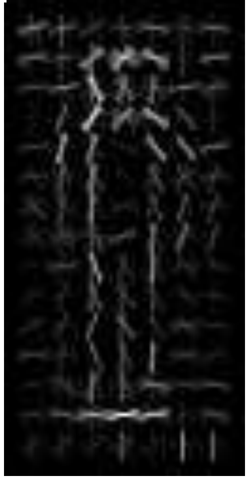
- Object: 2D shape
  - Local shape info, shading, shadows, texture
- Scene : overall layout
  - linear perspective, gradients
- Material properties: albedo, feel, hardness, …
  - Color, texture
- Motion
  - Optical flow, tracked points

General Principles of Representation

- Coverage
  - Ensure that all relevant info is captured
- Concision
  - Minimize number of features without sacrificing coverage
- Directness
  - Ideal features are independently useful for prediction

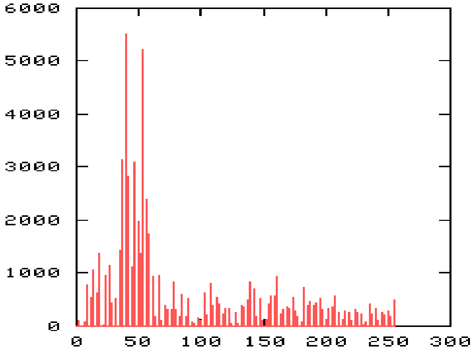
Image representations

Image Intensity



Gradient template

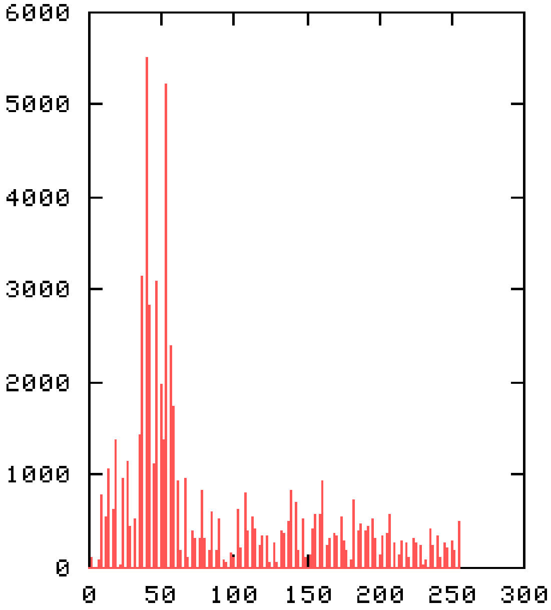
- Templates
  - Intensity, gradients, etc.

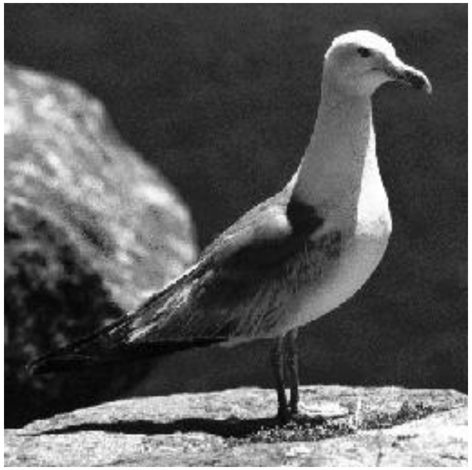


- Histograms
  - Color, texture, SIFT descriptors, etc.

- Average of features

Image Representations: Histograms





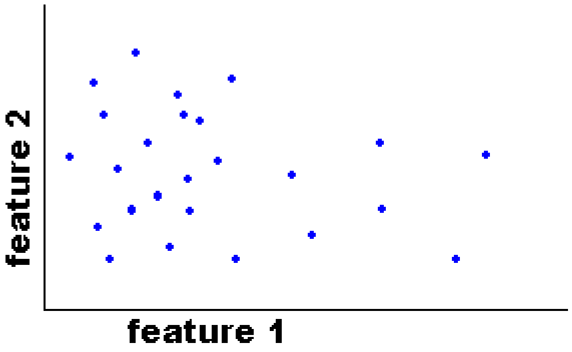
Global histogram

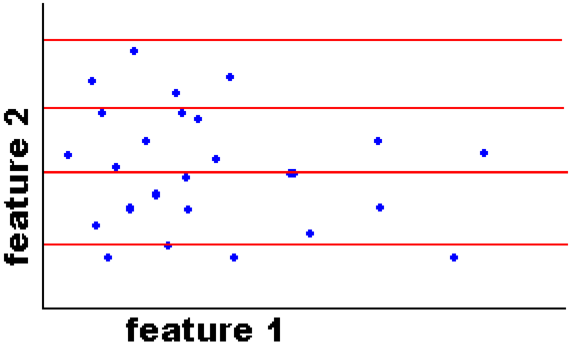
- Represent distribution of features
  - Color, texture, depth, …

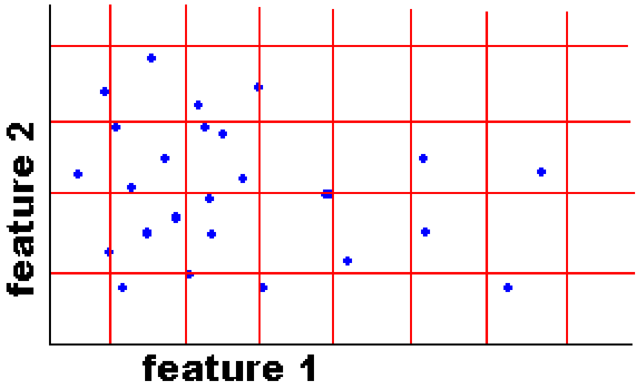
Images from Dave Kauchak

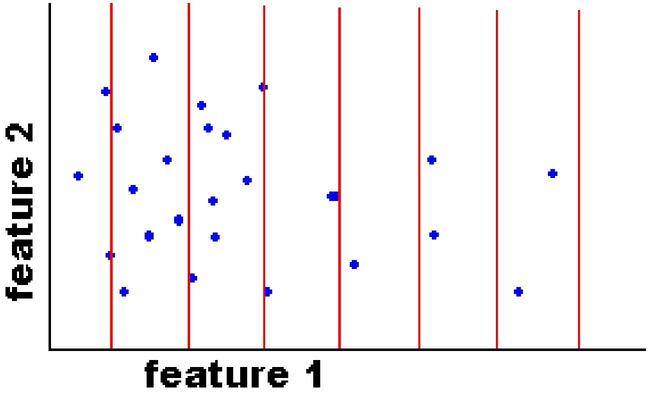
Image Representations: Histograms

Histogram: Probability or count of data in each bin









- Joint histogram
  - Requires lots of data
  - Loss of resolution to avoid empty bins

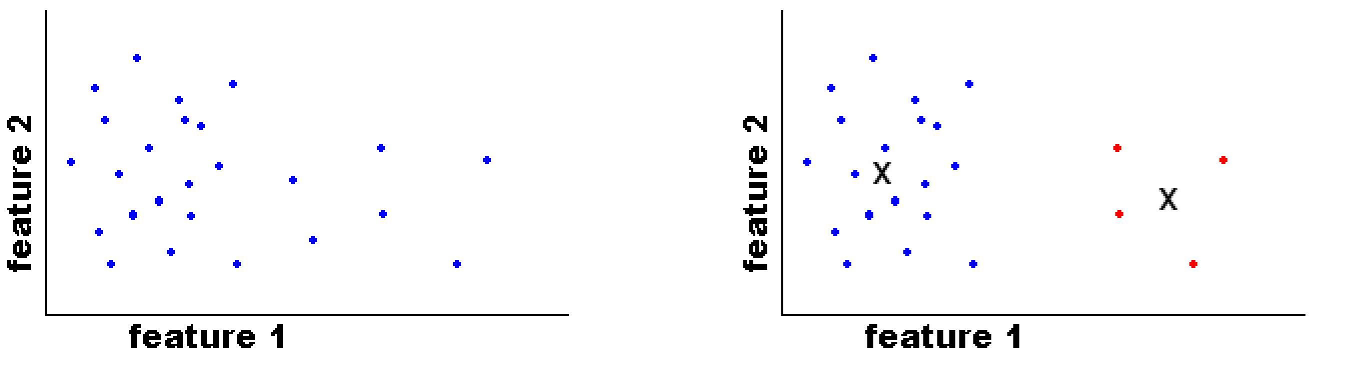
Marginal histogram

- Requires independent features
- More data/bin than joint histogram

Images from Dave Kauchak

Image Representations: Histograms

Clustering



Use the same cluster centers for all images

Images from Dave Kauchak

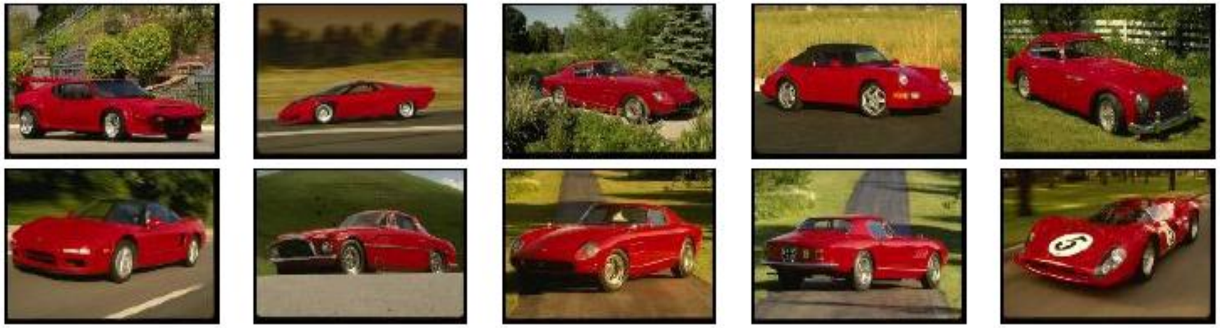
Computing histogram distance

$$\mathrm{histint}(h_i, h_j) = 1 - \sum_{m=1}^{K} \min\left(h_i(m), h_j(m)\right)$$

Histogram intersection (assuming normalized histograms)

$$\chi^2(h_i, h_j) = \frac{1}{2} \sum_{m=1}^{K} \frac{[h_i(m) - h_j(m)]^2}{h_i(m) + h_j(m)}$$

Chi-squared Histogram matching distance



Cars found by color histogram matching using chi-squared

Histograms: Implementation issues

- Quantization
  - Grids: fast but applicable only with few dimensions
  - Clustering: slower but can quantize data in higher dimensions

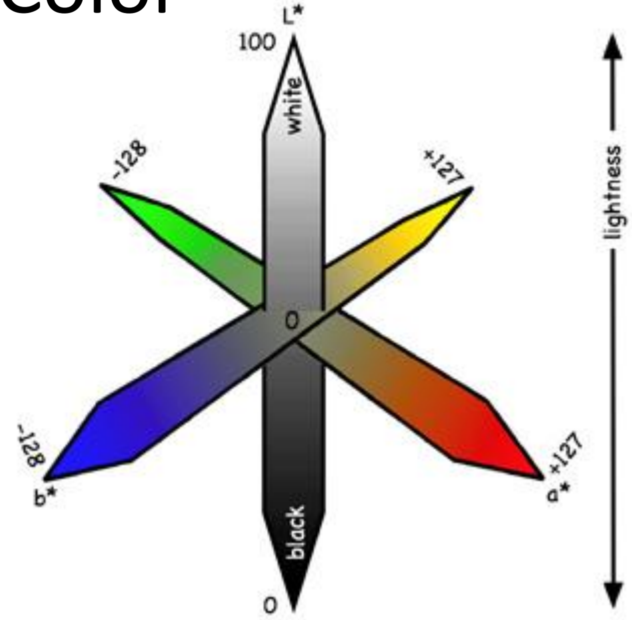
Few Bins
Need less data
Coarser representation

Many Bins
Need more data
Finer representation

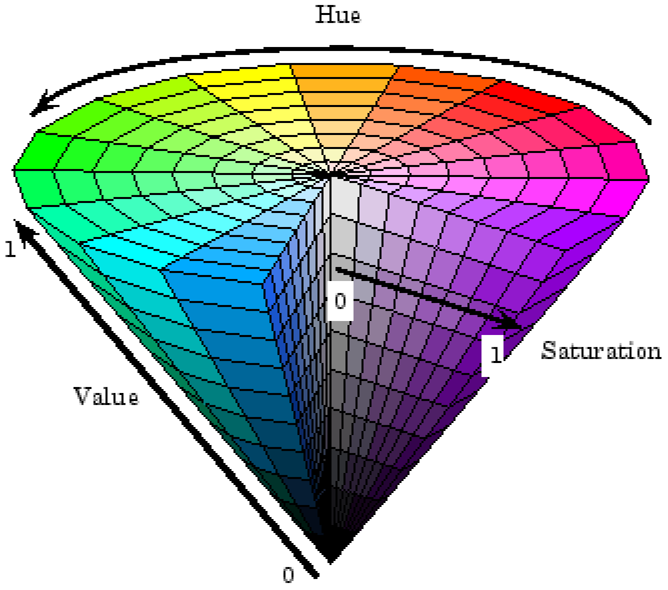
- Matching
  - Histogram intersection or Euclidean may be faster
  - Chi-squared often works better
  - Earth mover's distance is good for when nearby bins represent similar values

What kind of things do we compute histograms of?

- Color



L*a*b* color space

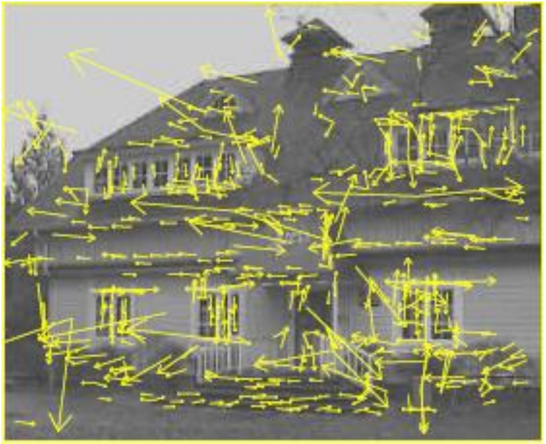


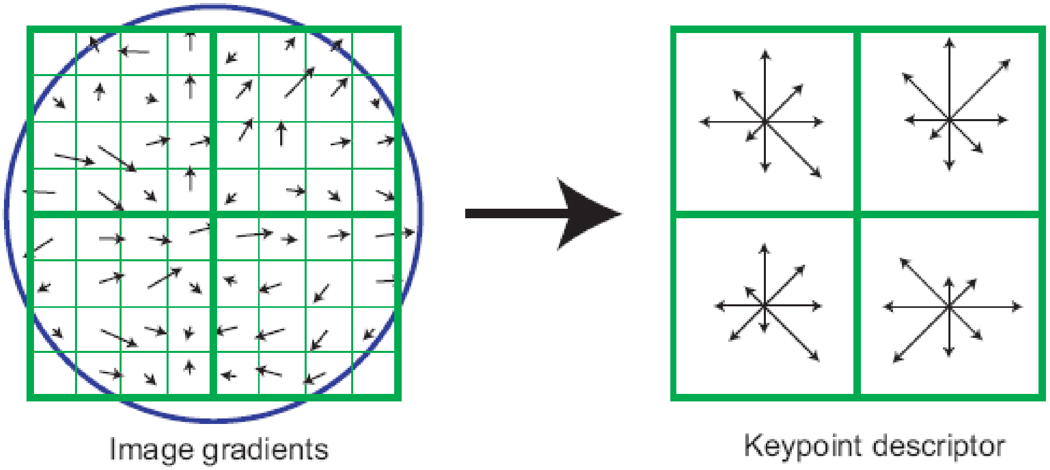
HSV color space

- Texture (filter banks or HOG over regions)

What kind of things do we compute histograms of?

- Histograms of descriptors

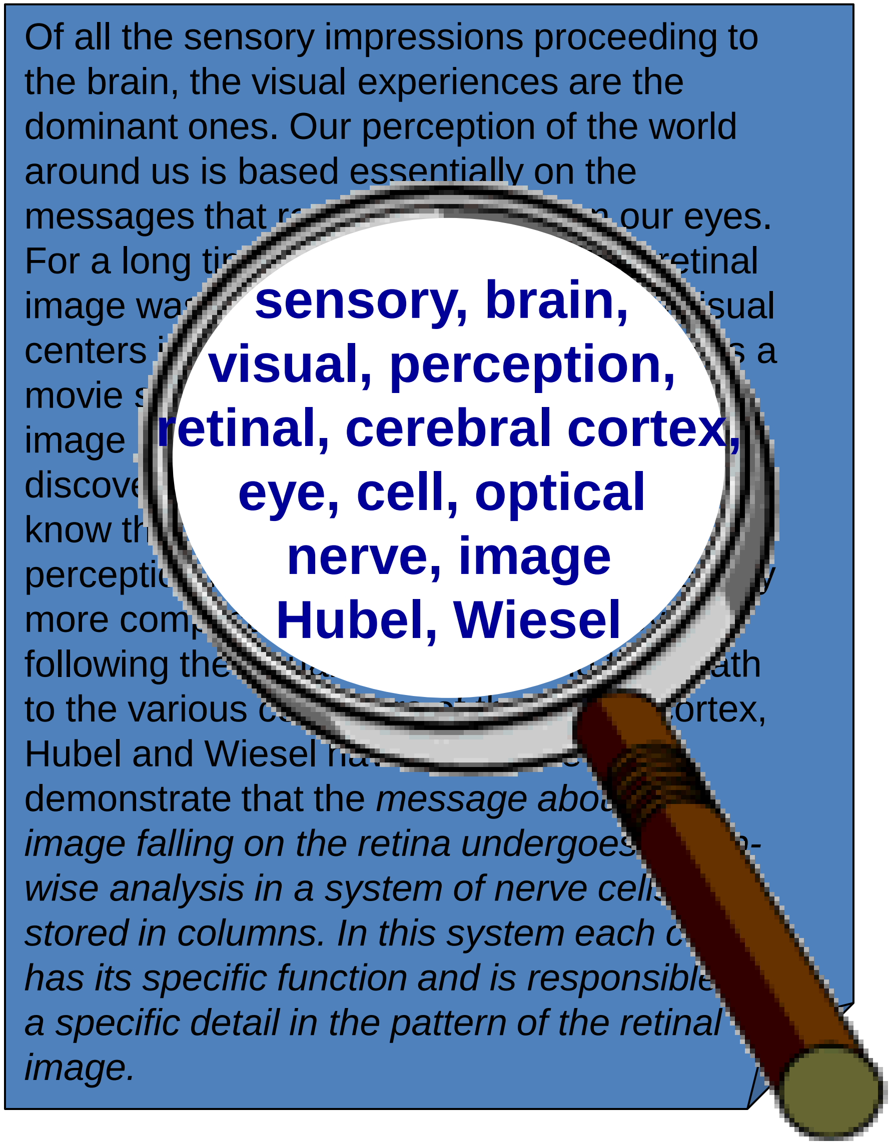




SIFT – Lowe IJCV 2004

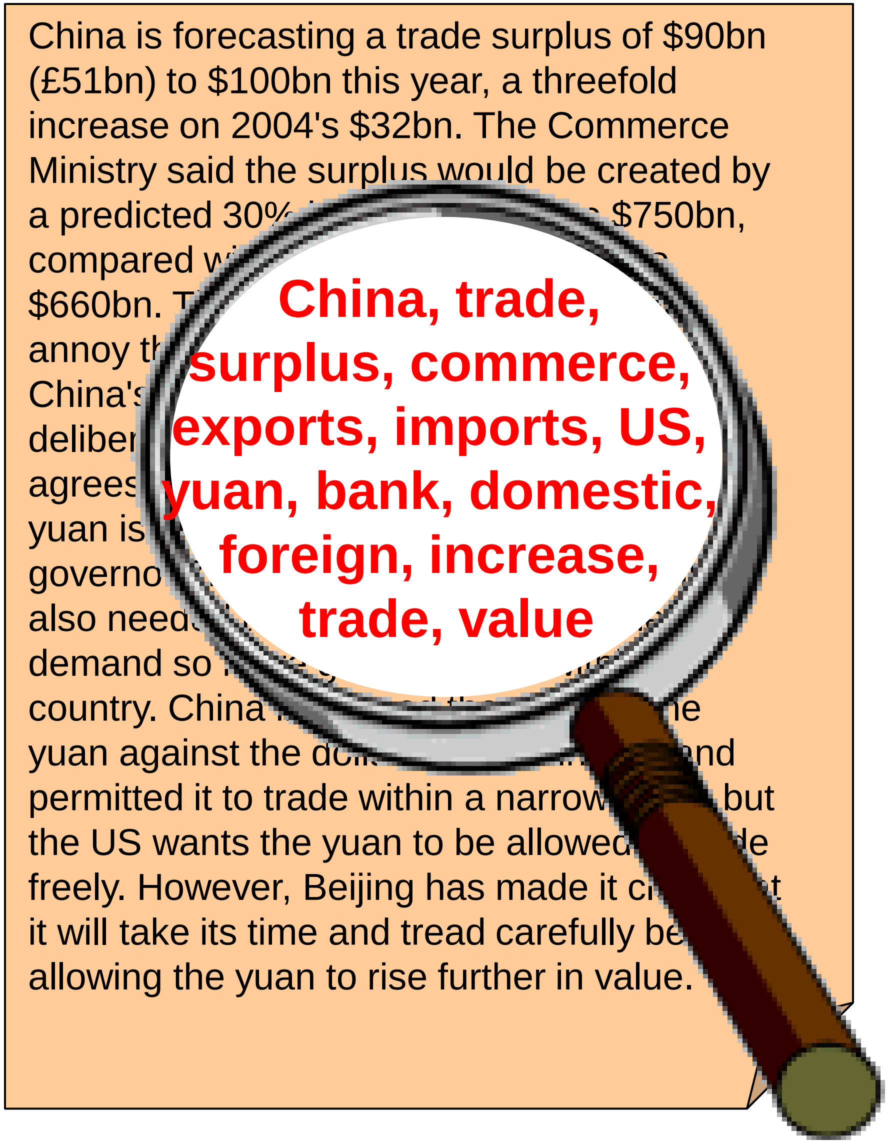
- “Bag of words”

Analogy to documents



Of all the sensory impressions proceeding to
the brain, the visual experiences are the
dominant ones. Our perception of the world
around us is based essentially on the
messages that r
our eyes.
For a long ti
retinal
image wa
visual
centers i
s a
movie s
image
discove
know th
perceptio
y
more comp
following the
ath
to the various c
ortex,
Hubel and Wiesel ha
demonstrate that the *message abo*
image falling on the retina undergoes
-
wise analysis in a system of nerve cell
stored in columns. In this system each c
has its specific function and is responsible
a specific detail in the pattern of the retinal
image.

sensory, brain,
visual, perception,
retinal, cerebral cortex,
eye, cell, optical
nerve, image
Hubel, Wiesel



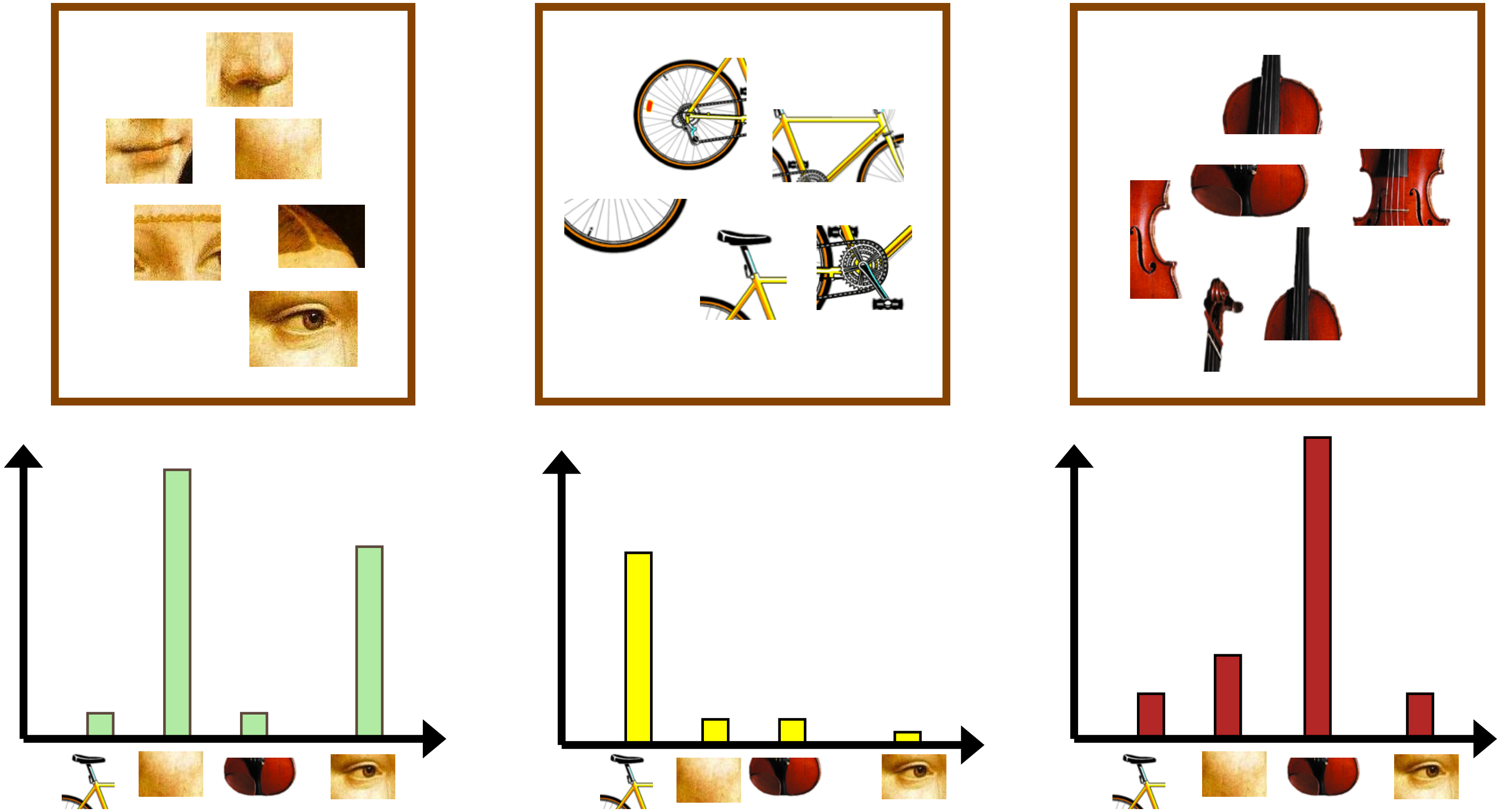
China is forecasting a trade surplus of $90bn
(£51bn) to $100bn this year, a threefold
increase on 2004's $32bn. The Commerce
Ministry said the surplus would be created by
a predicted 30%
$750bn,
compared w
$660bn. T
annoy th
China's
deliber
agrees
yuan is
governo
also need
demand so
country. China
e
yuan against the do
nd
permitted it to trade within a narrow
but
the US wants the yuan to be allowed
de
freely. However, Beijing has made it c
t
it will take its time and tread carefully be
allowing the yuan to rise further in value.

China, trade,
surplus, commerce,
exports, imports, US,
yuan, bank, domestic,
foreign, increase,
trade, value

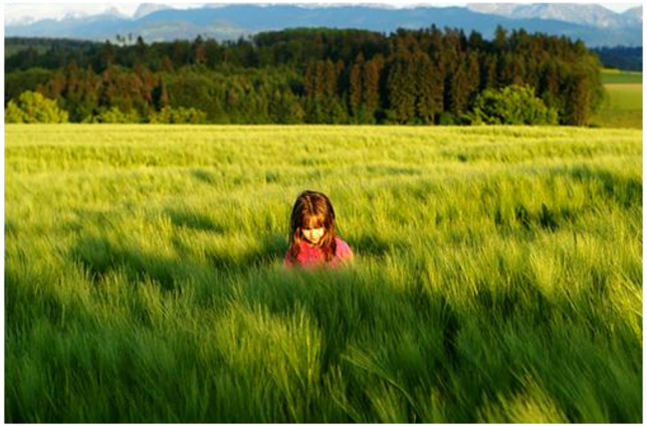
ICCV 2005 short course, L. Fei-Fei

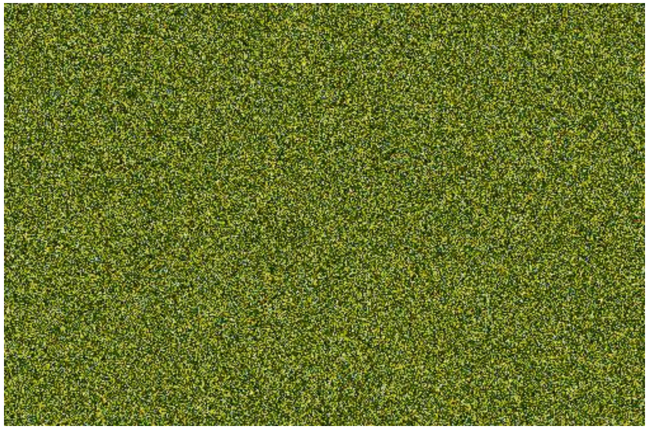
Bag of visual words

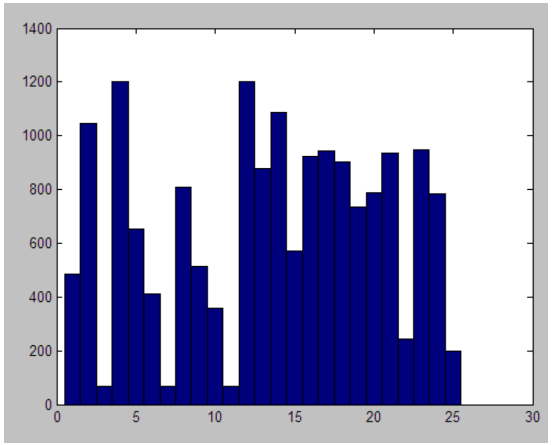
- Image patches
- Bow histogram
- Codewords

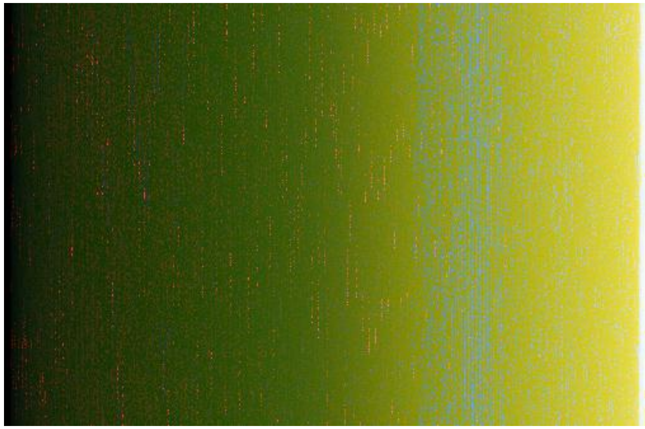


But what about layout?



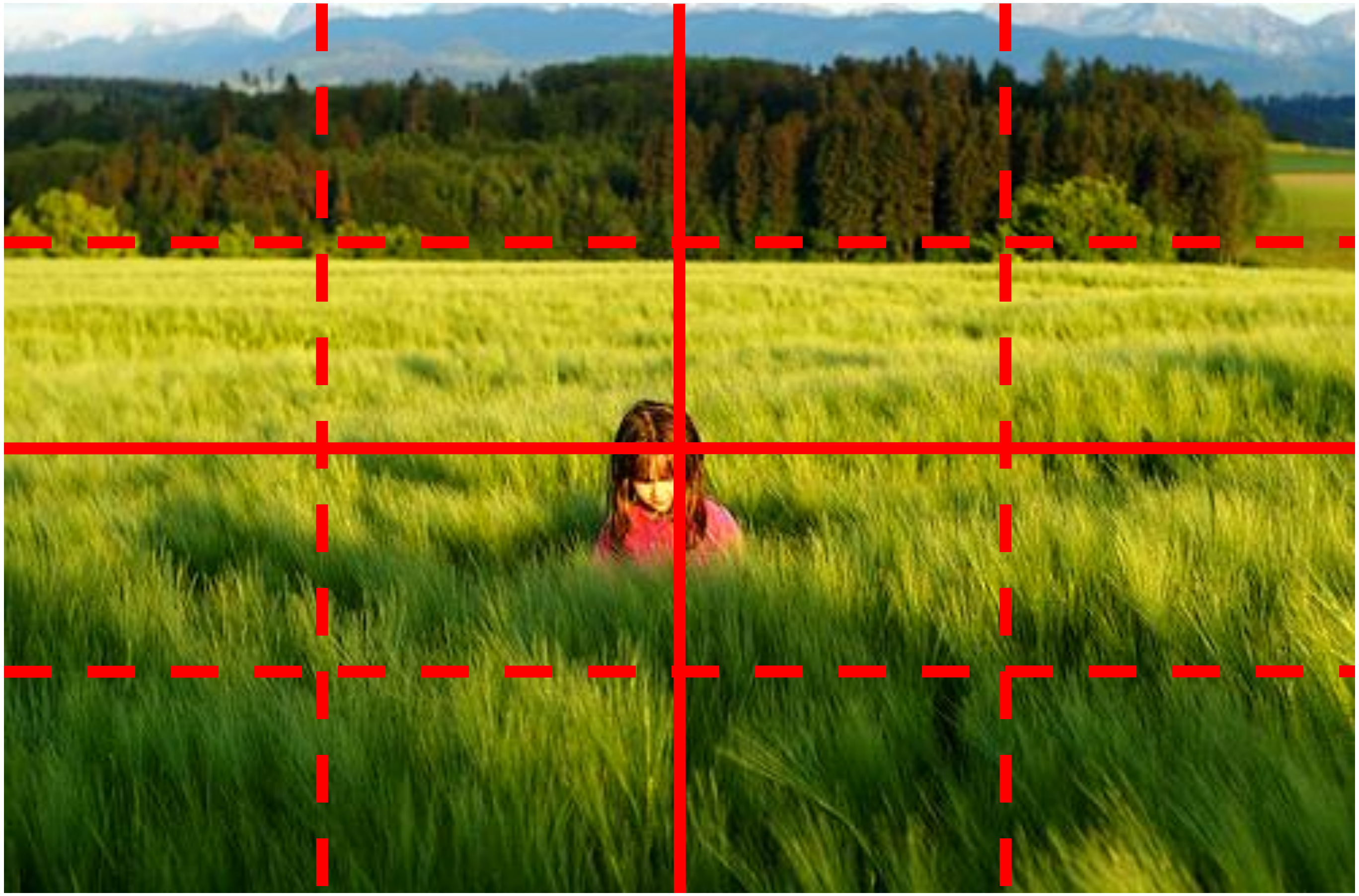






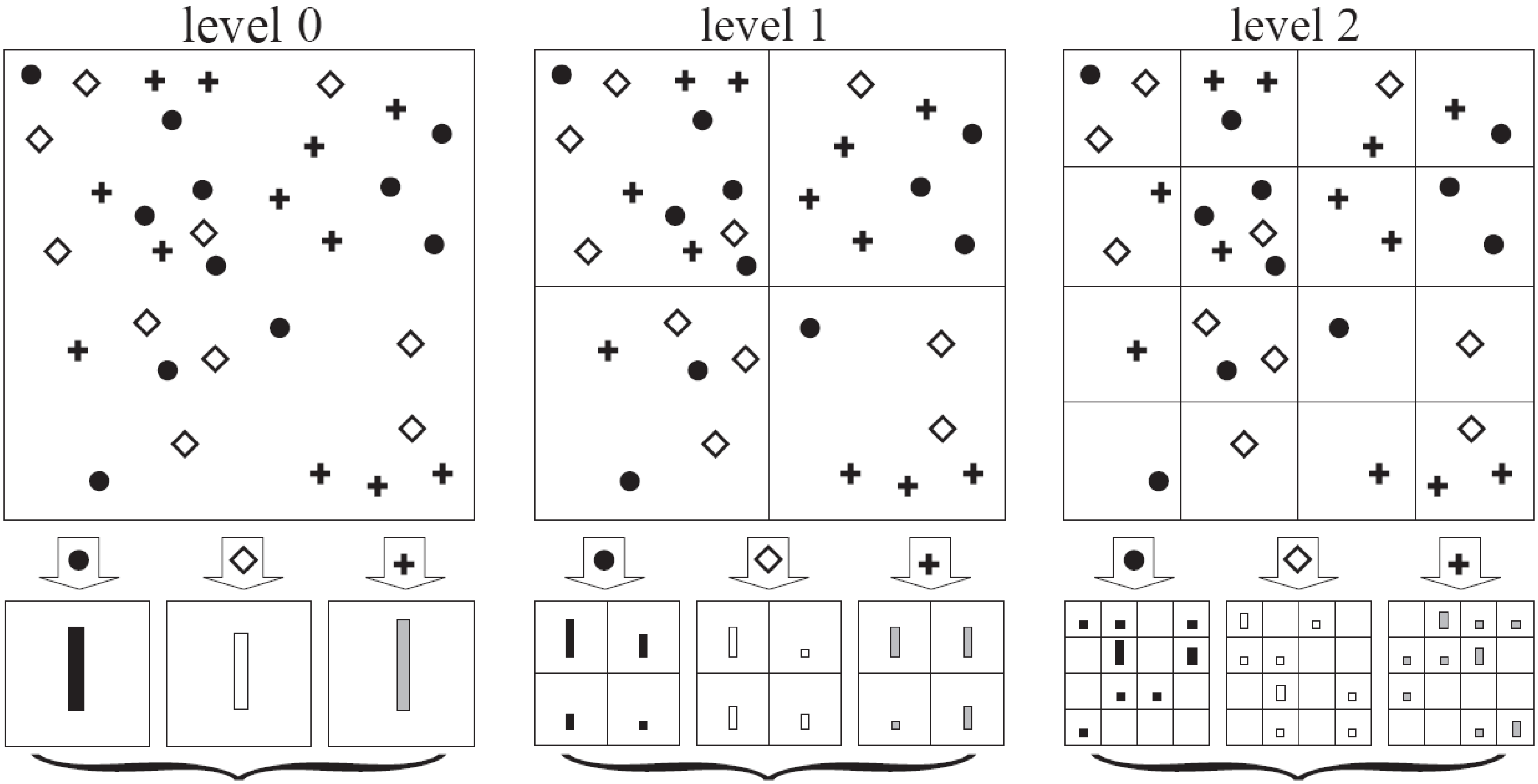
All of these images have the same color histogram

Spatial pyramid



Compute histogram in each spatial bin

Spatial pyramid



High number of features – PCA to reduce dimensionality

Image Categorization: Bag of Words

Training

1. Extract keypoints and descriptors for all training images
2. Cluster descriptors
3. Quantize descriptors using cluster centers to get “visual words”
4. Represent each image by normalized counts of “visual words”
5. Train classifier on labeled examples using histogram values as features

Testing

1. Extract keypoints/descriptors and quantize into visual words
2. Compute visual word histogram
3. Compute label or confidence using classifier

Often used features

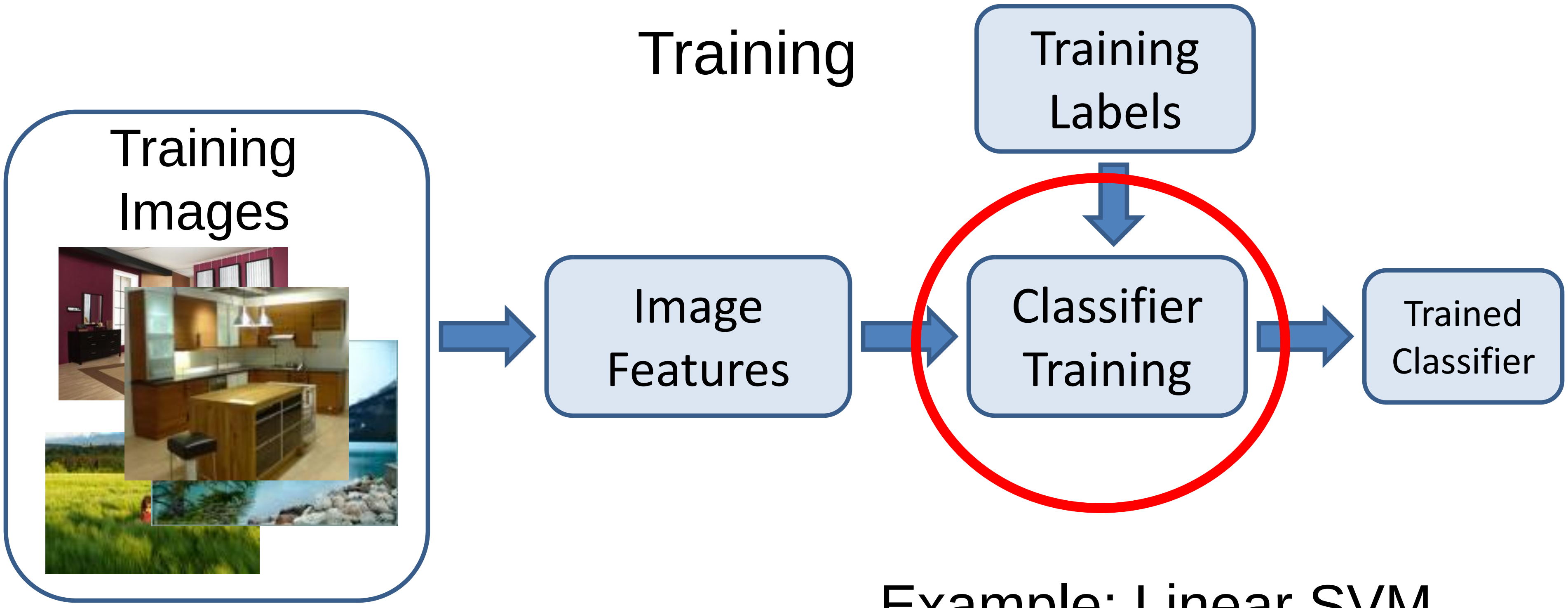
- Scene: GIST, Spatial pyramid BoW, color
- Object: Spatial pyramid BoW, HOG, color
- Material: texture, color

Things to remember about representation

- Most features can be thought of as templates, histograms (counts), or combinations

- Think about the right features for the problem
  - Coverage
  - Concision
  - Directness

Part 2: Classifiers



Training

Training Images

Image Features

Training Labels

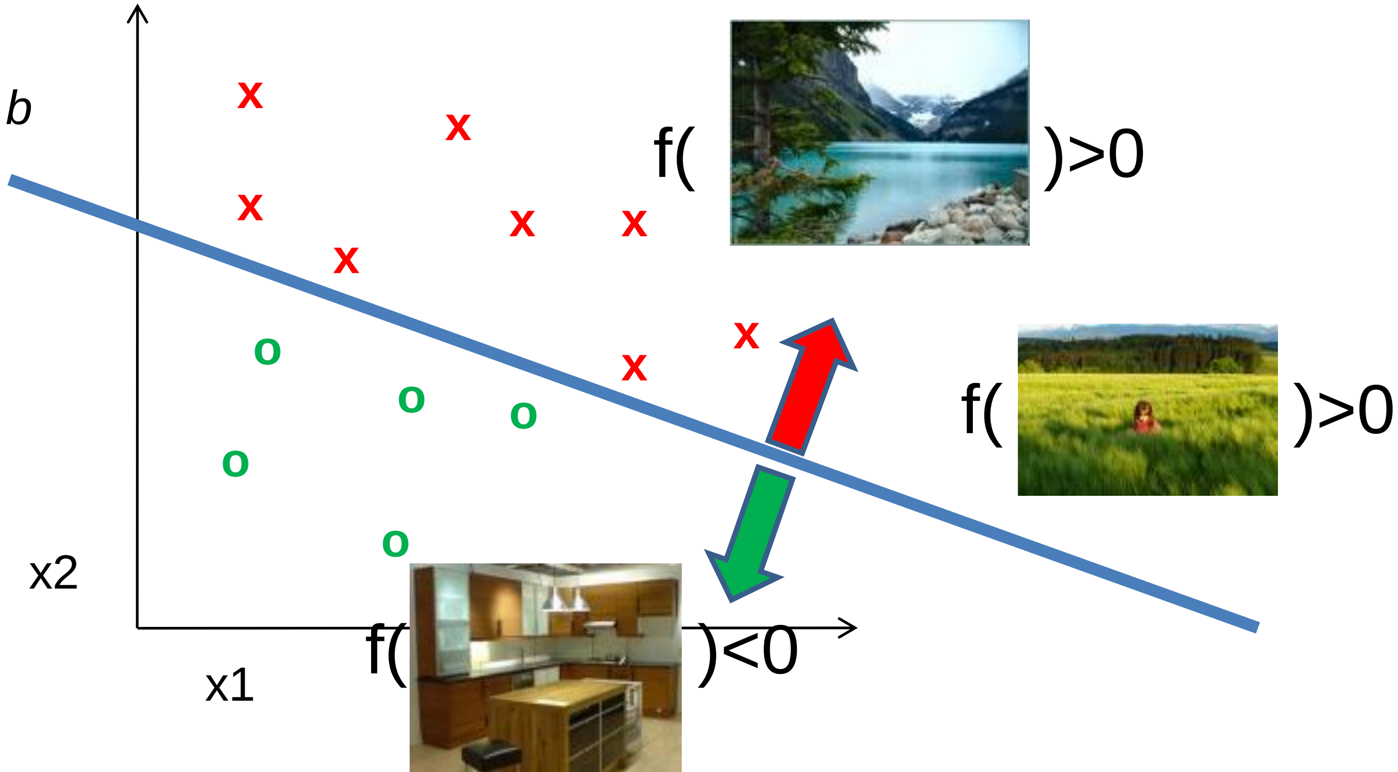
Classifier Training

Trained Classifier

Example: Linear SVM

Linear classifier

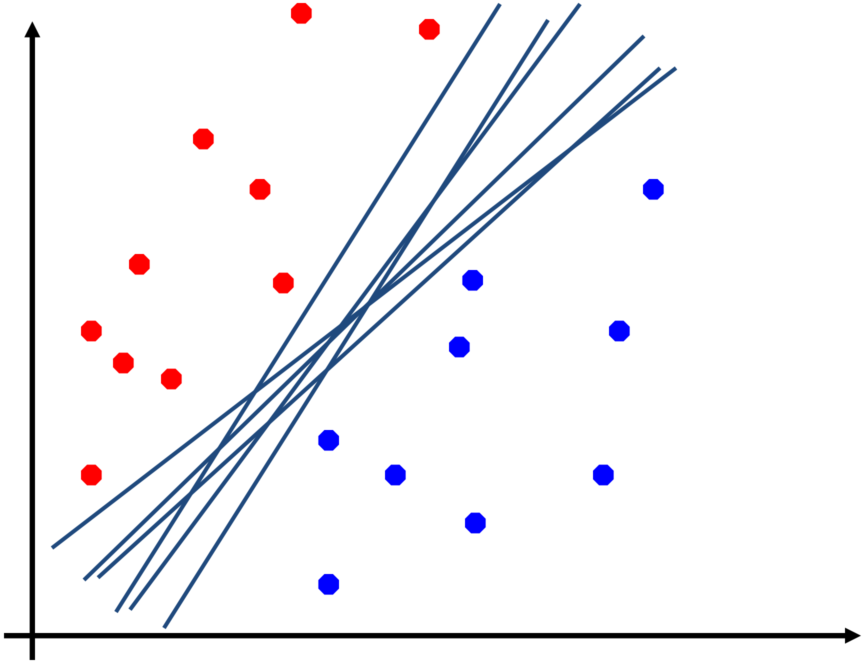
Finding the linear hyperplane that separate examples of different categories



$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$

Linear Separators

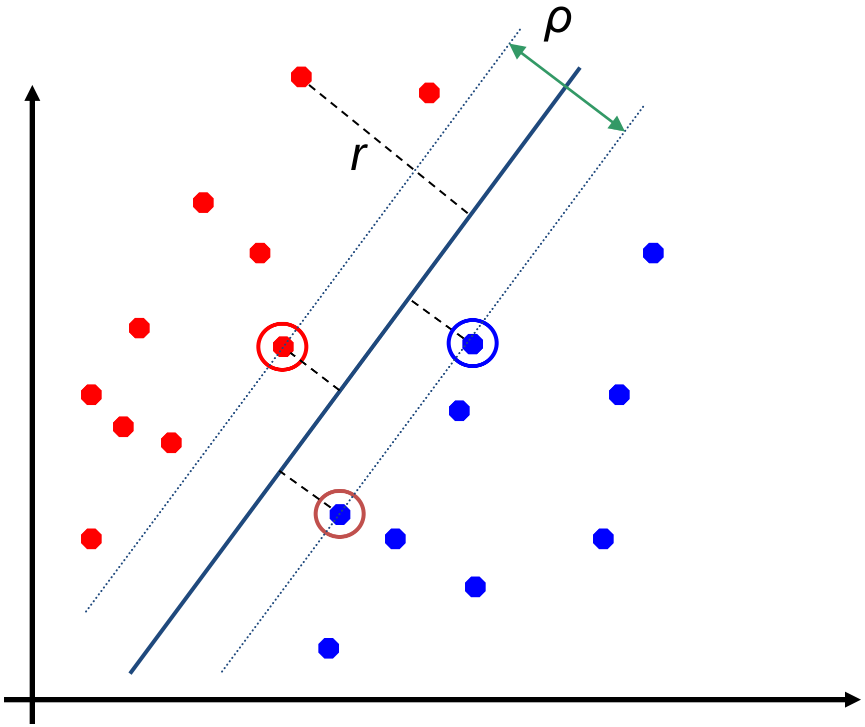
- Which of the linear separators is optimal?



*http://www.cs.utexas.edu/~mooney/cs391L/slides/**svm.ppt***
http://www1.cs.columbia.edu/~belhumeur/courses/biometrics/2010/svm.ppt

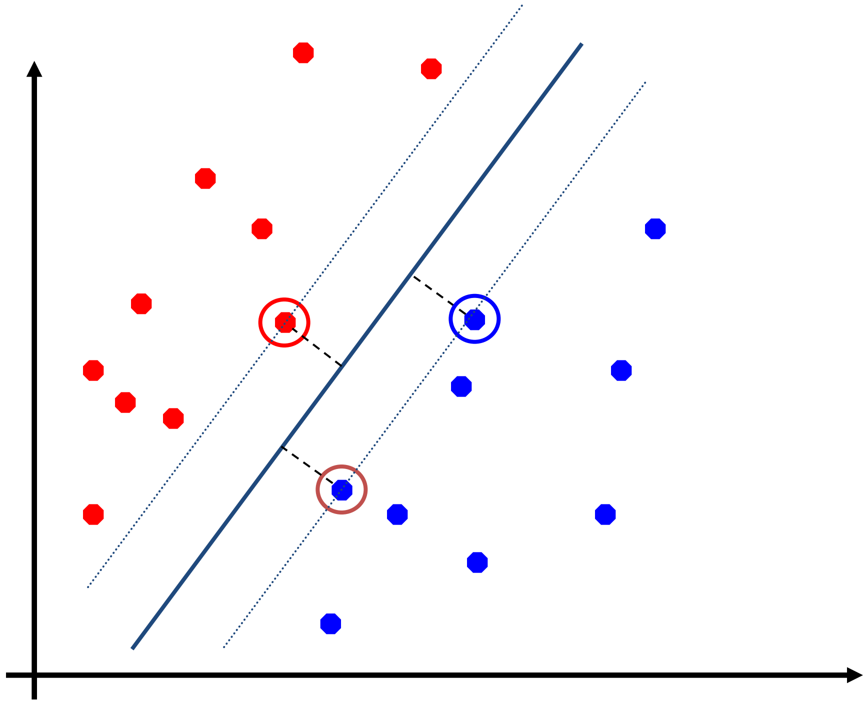
Classification Margin

- Distance from example $\mathbf{x}_i$ to the separator is $r = \dfrac{\mathbf{w}^T\mathbf{x}_i + b}{\|\mathbf{w}\|}$
- Examples closest to the hyperplane are ***support vectors***.
- ***Margin*** *ρ* of the separator is the distance between support vectors.



Maximum Margin Classification

- Implies that only support vectors matter; other training examples are ignorable.



Linear SVM Mathematically

- Let training set $\{(\mathbf{x}_i, y_i)\}_{i=1..n}$, $\mathbf{x}_i \in \mathbf{R}^d$, $y_i \in \{-1, 1\}$ be separated by a hyperplane with margin ρ. Then for each training example $(\mathbf{x}_i, y_i)$:

$$\begin{array}{ll} \mathbf{w}^\mathbf{T}\mathbf{x}_i + b \leq -\rho/2 & \text{if } y_i = -1 \\ \mathbf{w}^\mathbf{T}\mathbf{x}_i + b \geq \rho/2 & \text{if } y_i = 1 \end{array} \quad \Leftrightarrow \quad y_i(\mathbf{w}^\mathbf{T}\mathbf{x}_i + b) \geq \rho/2$$

- For every support vector $\mathbf{x}_s$ the above inequality is an equality. After rescaling **w** and *b* by $\rho/2$ in the equality, we obtain that distance between each $\mathbf{x}_s$ and the hyperplane is $r = \frac{\mathrm{y}_s(\mathbf{w}^T\mathbf{x}_s + b)}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}$

- Then the margin can be expressed through (rescaled) **w** and b as:

$$\rho = 2r = \frac{2}{\|\mathbf{w}\|}$$

Solving the Optimization Problem

Quadratic programming with linear constraints

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2$$

$$\text{s.t.} \quad y_i(\mathbf{w}^T\mathbf{x}_i + b)\ {}^{3}\ 1$$

$$\Updownarrow$$

Lagrangian Function

$$\text{minimize} \quad L_p(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \right)$$

$$\text{s.t.} \quad \alpha_i \ge 0$$

Solving the Optimization Problem

$$\boxed{\begin{array}{c} \text{minimize} \;\; L_p(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \right) \\[1em] \text{s.t.} \qquad \alpha_i \geq 0 \end{array}}$$

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \quad \Longrightarrow \quad \mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L_p}{\partial b} = 0 \quad \Longrightarrow \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

Solving the Optimization Problem

$$\text{minimize}\;\; L_p(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \right)$$

$$\text{s.t.} \qquad \alpha_i \geq 0$$

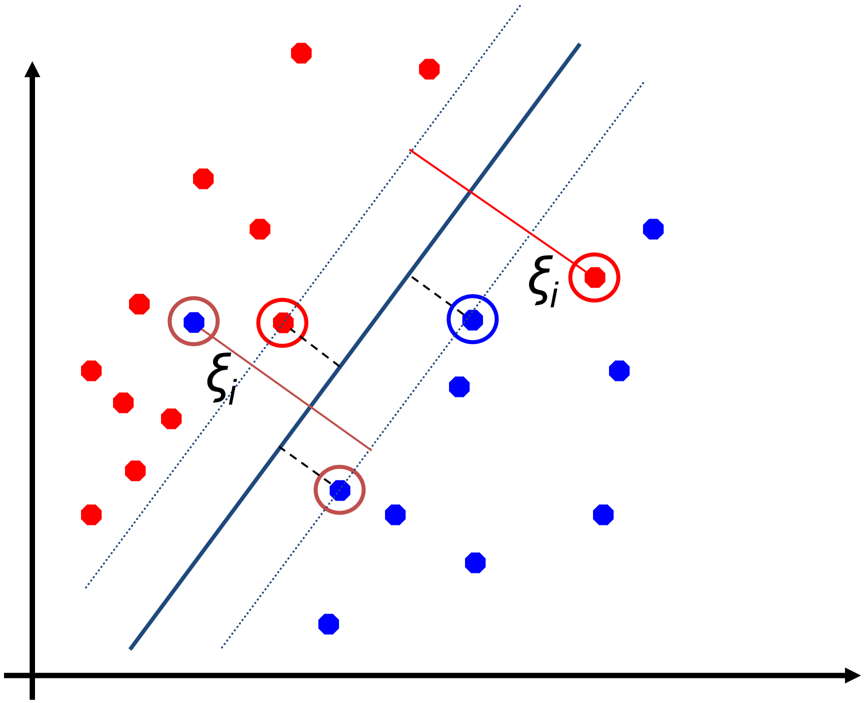
Lagrangian Dual Problem

$$\text{maximize}\;\; \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$\text{s.t.} \qquad \alpha_i \geq 0 \text{ , and } \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

Soft Margin Classification

- What if the training set is not linearly separable?
- *Slack variables* ξ_i can be added to allow misclassification of difficult or noisy examples, resulting margin called *soft*.



Large Margin Linear Classifier

- Formulation:

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

such that

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

- Parameter *C* is a trade off factor

Large Margin Linear Classifier

- Formulation: (Lagrangian Dual Problem)

$$\text{maximize} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

such that

$$0 \le \alpha_i \le C$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

Linear SVMs: Recap

- The classifier is a *separating hyperplane.*
- Most “important” training points are support vectors; they define the hyperplane.
- Quadratic optimization algorithms can identify which training points $\mathbf{x}_i$ are support vectors with non-zero Lagrangian multipliers α_i**.**

Multiclass classification (one vs all)

- Learning a function for each category: $f_i(x)$
  - y=1: for examples in this category
  - y=-1: for examples not in this category

- Finding the class with the largest function value

$$\bar{c} = \arg\max_c f_c(x)$$

Measuring classification performance

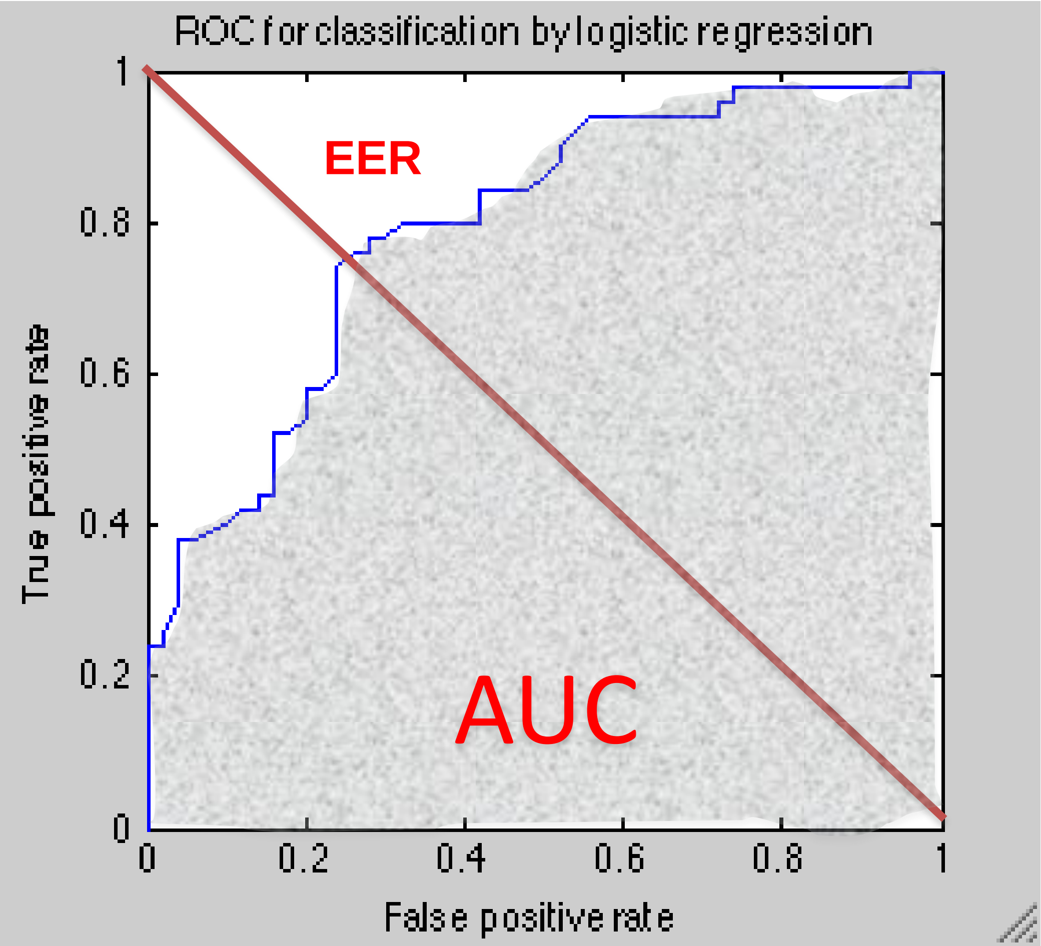
- Confusion matrix
- Accuracy
  - (TP+TN)/(TP+TN+FP+FN)
- True Positive Rate=Recall
  - TP/(TP+FN)
- False Positive Rate
  - FP/(FP+TN)
- Precision
  - TP/(TP+FP)
- F1 Score
  - 2*Recall*Precision/(Recall+Precision)

		Predicted	
		Positive	Negative
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

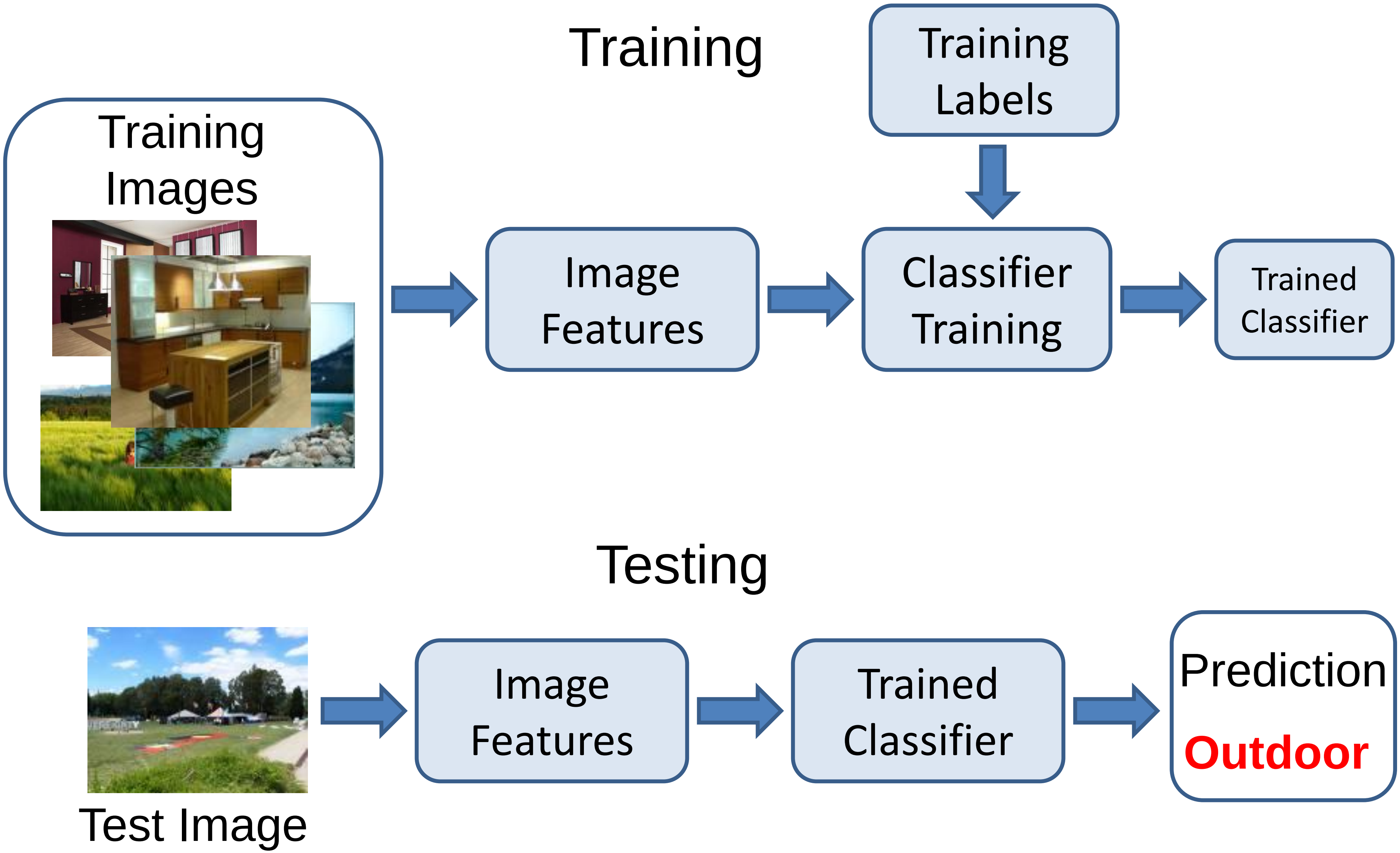
		Predicted class		
		Class1	Class2	Class3
Actual class	Class1	40	1	6
	Class2	3	25	7
	Class3	4	9	10

ROC curve

- Receiver_operating_characteristic
  - Area under the curve (AUC)
  - Equal Error Rate (EER)



Pipeline



Training

Training Images

Image Features

Training Labels

Classifier Training

Trained Classifier

Testing

Test Image

Image Features

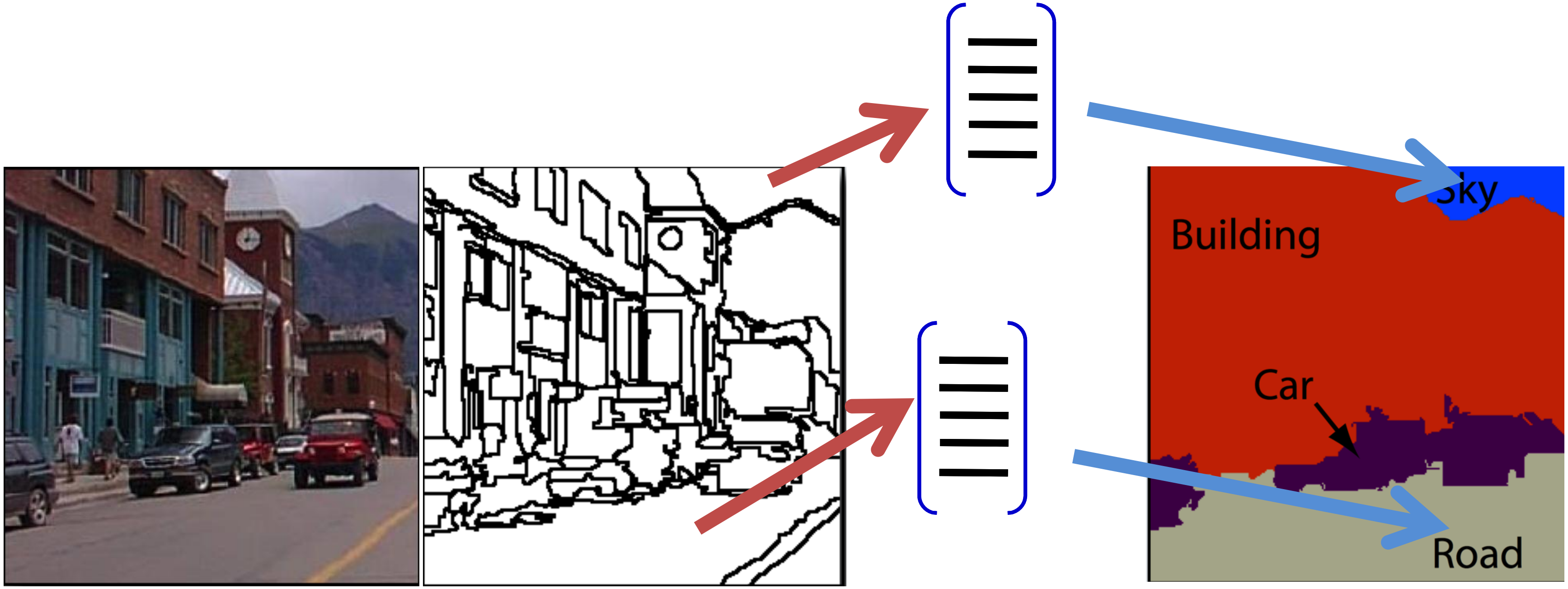
Trained Classifier

Prediction

Outdoor

Region Representation

- Segment the image
- Use features to represent each image segment



Joseph Tighe and Svetlana Lazebnik

Region Representation

- Color, texture, BoW
  - Only computed within the local region

- Shape of regions
- Position in the image

Working with regions

- Spatial support is important – multiple segmentation

- Spatial consistency – MRF smoothing

HW5, Prob2

- Training and testing images for 8 categories

- Implement representation: color histogram

- BoW model: preprocessed descriptors
  - Learning dictionary using K-Means
  - Learning classifier (NN, SVM)

- Report final result (confusion matrix)