

# Recognizing and Learning Object Categories

Based on work and slides by R. Fergus, P. Perona, A. Zisserman, A. Efros, J. Ponce, S. Lazebnik, C. Schmid, F. DiMaio, and others

## Traditional Problem: Single Object Recognition



## Most Objects Exhibit Considerable Intra-Class Variability



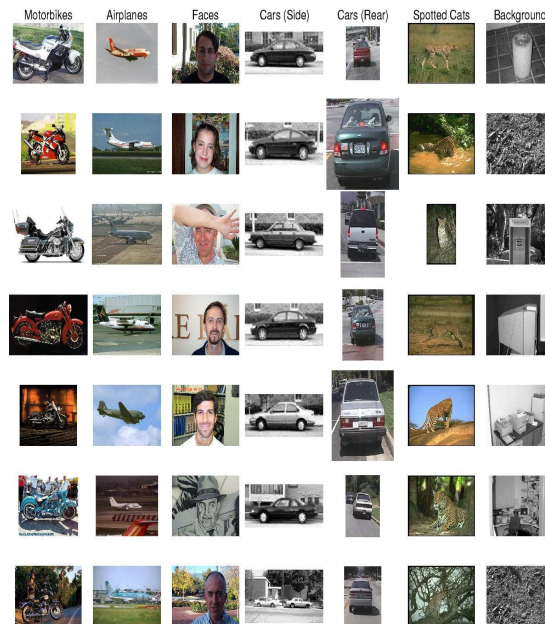
Task: Recognition of object **categories**

## Some object categories

Learn from just examples

Difficulties:

- ⌘ Size variation
- ⌘ Background clutter
- ⌘ Occlusion
- ⌘ Intra-class variation
- ⌘ Viewpoint variation
- ⌘ Illumination variation



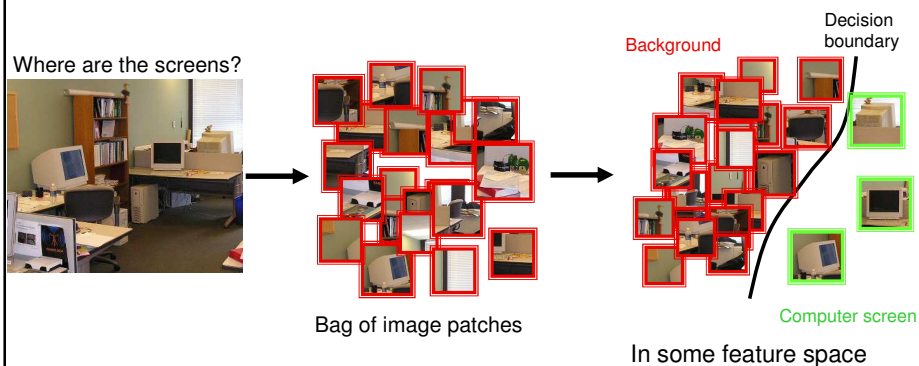


## Approach 1: Discriminative Methods

Object detection and recognition is formulated as a **classification problem**

The image is partitioned into a set of overlapping windows

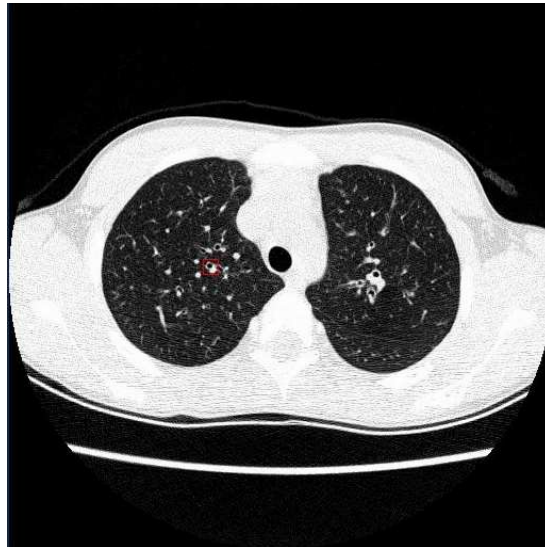
... and a decision is taken at each window about if it contains a target object or not



## HRCT Lung Image

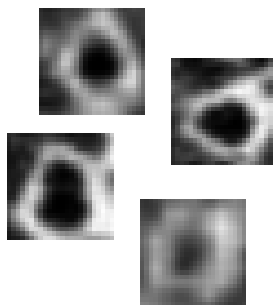


Dilated bronchus

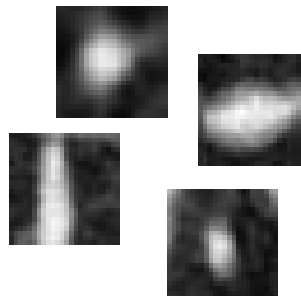


## Training Examples

Bronchiectasis  
(positive examples)



Non-Bronchiectasis  
(negative examples)

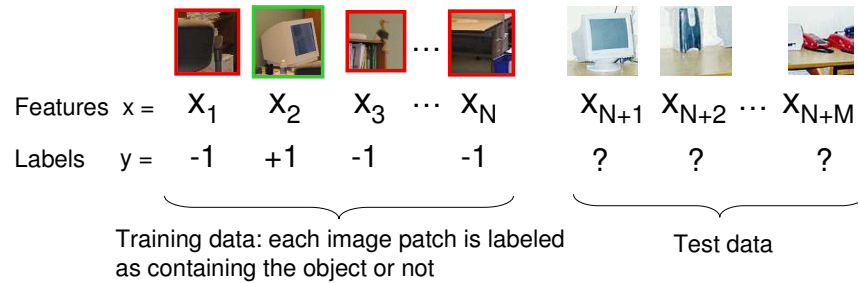


$24 \times 24$  images



# Formulation

## § Formulation: binary classification



- Classification function

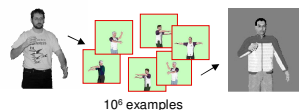
$$\hat{y} = F(x) \quad \text{Where } F(x) \text{ belongs to some family of functions}$$

- Minimize misclassification error

(Not that simple: we need some guarantees that there will be generalization)

## Discriminative Methods

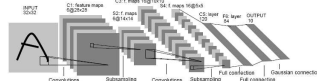
## Nearest Neighbor



Shakhnarovich, Viola, Darrell 2003  
Berg, Berg, Malik 2005

...

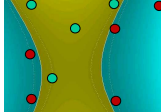
## Neural Networks



LeCun, Bottou, Bengio, Haffner 1998  
Rowley, Baluja, Kanade 1998

...

## Support Vector Machines and Kernels



Guyon, Vapnik  
Heisele, Serre, Poggio, 2001

...

## Conditional Random Fields



McCallum, Freitag, Pereira 2000  
Kumar, Hebert 2003

...

## Object categorization: the statistical viewpoint



$$p(\text{zebra} | \text{image})$$

vs.

$$p(\text{no zebra} | \text{image})$$

§ Bayes's rule:

$$\underbrace{\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})}}_{\text{posterior ratio}} = \underbrace{\frac{p(\text{image} | \text{zebra})}{p(\text{image} | \text{no zebra})}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(\text{zebra})}{p(\text{no zebra})}}_{\text{prior ratio}}$$

## Object categorization: the statistical viewpoint

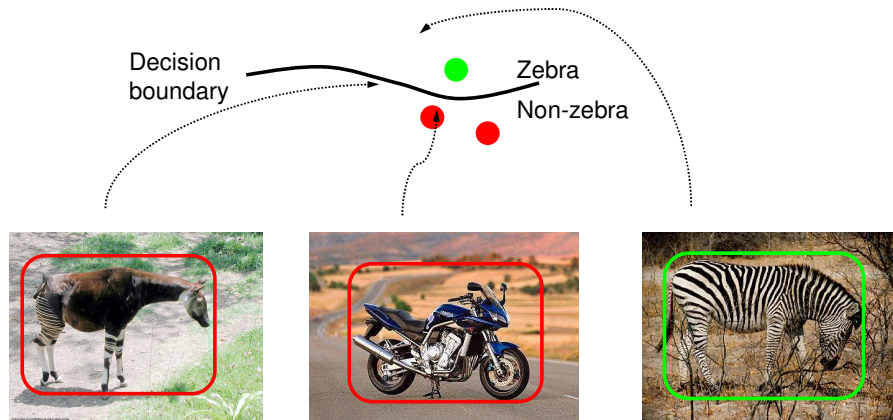
$$\underbrace{\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})}}_{\text{posterior ratio}} = \underbrace{\frac{p(\text{image} | \text{zebra})}{p(\text{image} | \text{no zebra})}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(\text{zebra})}{p(\text{no zebra})}}_{\text{prior ratio}}$$

§ Discriminative methods model the *posterior*

§ Generative methods model the *likelihood* and *prior*

## Discriminative



§ Direct modeling of  $\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})}$



## Generative

§ Model  $p(\text{image} | \text{zebra})$  and  $p(\text{image} | \text{no zebra})$



	$p(\text{image}   \text{zebra})$	$p(\text{image}   \text{no zebra})$
	Low	Middle
	High	Middle Low

## Three main issues

### § Representation

§ How to represent an object category

### § Learning

§ How to form the classifier, given training data

### § Recognition

§ How the classifier is to be used on novel data

## Constructing models of image content

Basic components: *local features* and *spatial relations*

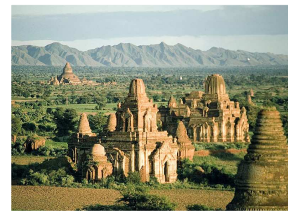
Textures



Objects

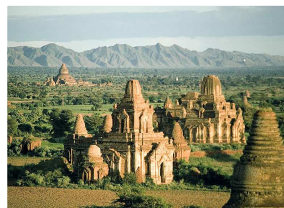


Scenes

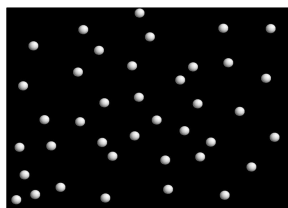


## Constructing models of image content

Basic components: *local features* and *spatial relations*

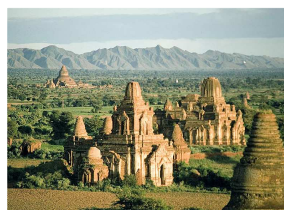


Local model

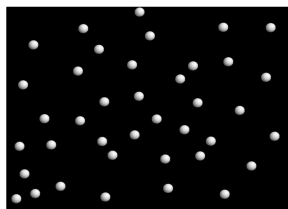


## Constructing models of image content

Basic components: *local features* and *spatial relations*



Local model



## Constructing models of image content

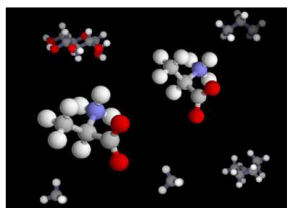
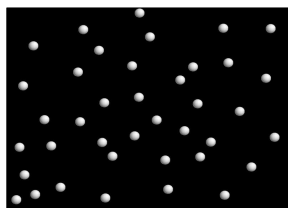
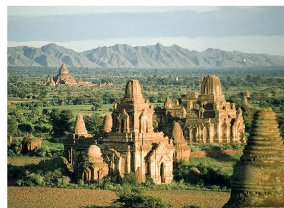
Basic components: *local features* and *spatial relations*



Local model



Semi-local model



## Constructing models of image content

Basic components: *local features* and *spatial relations*



Local model



Semi-local model





## Constructing models of image content

Basic components: *local features* and *spatial relations*  
(usually appearance)



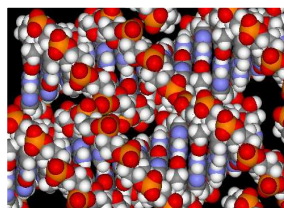
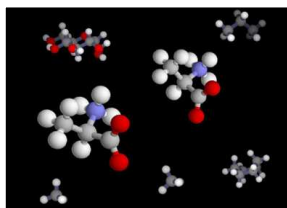
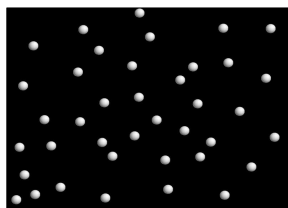
Local model



Semi-local model



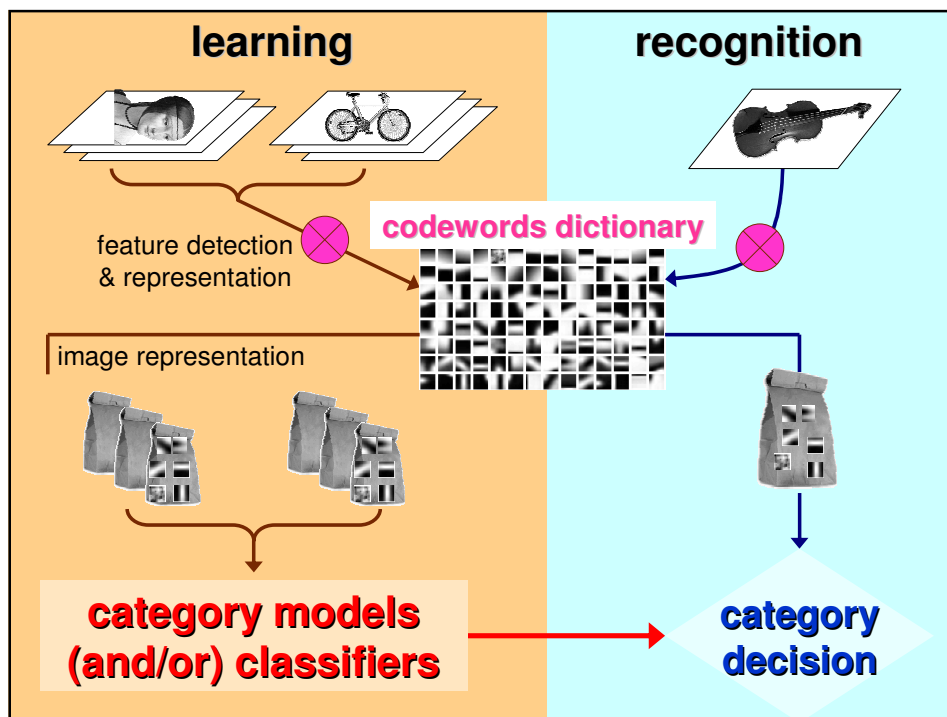
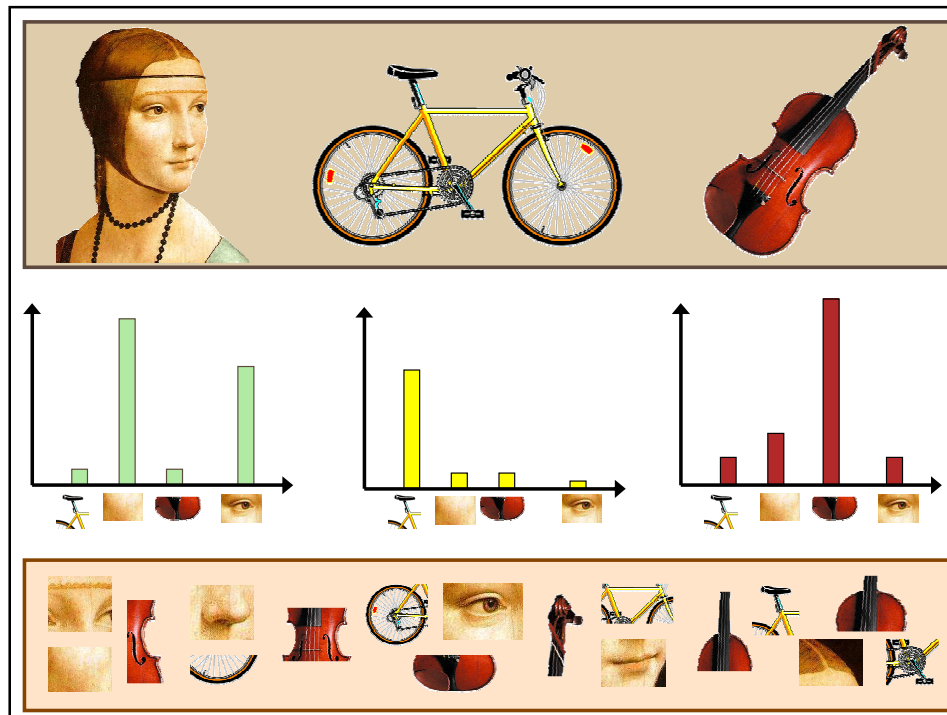
Global model



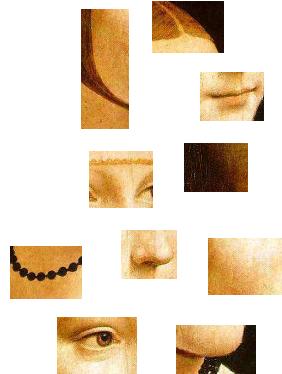
## Approach 2: Generative Methods using Bag of Words Models

- § An image is represented by a collection of “visual words” and their corresponding counts given a universal dictionary
- § Object categories are modeled by the distributions of these visual words
- § Although “bag of words” models can use both generative and discriminative approaches, here we will focus on generative models





## 1. Feature Detection and Representation



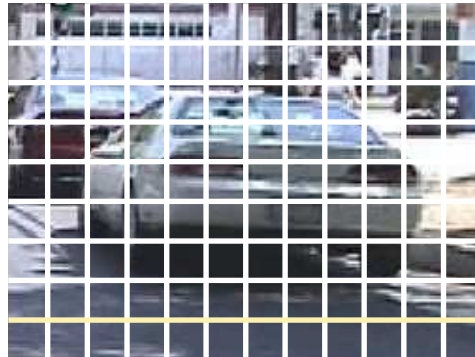
## Feature Detection

- § Sliding window
  - § Leung et al., 1999
  - § Viola et al., 1999
  - § Renninger et al. 2002



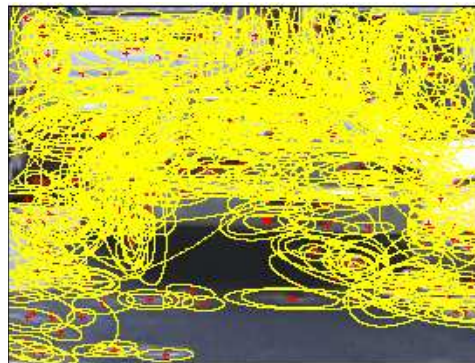
## Feature Detection

- § Sliding window
  - § Leung et al., 1999
  - § Viola et al., 1999
  - § Renninger et al., 2002
- § Regular grid
  - § Vogel et al., 2003
  - § Fei-Fei et al., 2005



## Feature Detection

- § Sliding window
  - § Leung et al., 1999
  - § Viola et al., 1999
  - § Renninger et al., 2002
- § Regular grid
  - § Vogel et al., 2003
  - § Fei-Fei et al., 2005
- § Interest point detector
  - § Csurka et al., 2004
  - § Fei-Fei et al., 2005
  - § Sivic et al., 2005



## Feature Detection

- § Sliding window
  - § Leung et al., 1999
  - § Viola et al., 1999
  - § Renninger et al., 2002
- § Regular grid
  - § Vogel et al., 2003
  - § Fei-Fei et al., 2005
- § Interest point detector
  - § Csurka et al., 2004
  - § Fei-Fei et al., 2005
  - § Sivic et al., 2005
- § Other methods
  - § Random sampling (Ullman et al., 2002)
  - § Segmentation based patches (Barnard et al., 2003)

## Feature Representation

Visual words, aka textons, aka keypoints:

K-means clustered pieces of the image

§ Various representations:

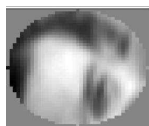
- § Filter bank responses
- § Image Patches
- § SIFT descriptors

All encode more-or-less the same thing ...

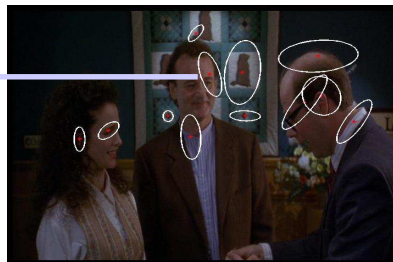


## Interest Point Features

  
**Compute  
SIFT  
descriptor**  
[Lowe'99]



**Normalize  
patch**



**Detect patches**

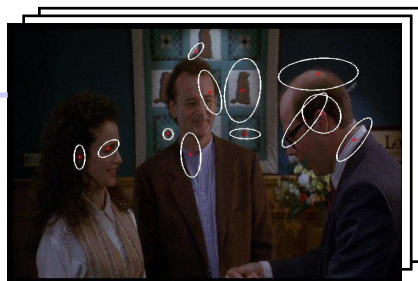
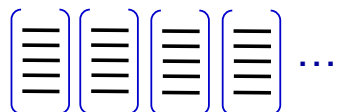
[Mikojaczyk and Schmid '02]

[Matas et al. '02]

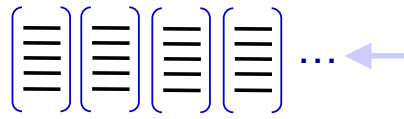
[Sivic et al. '03]

Slide credit: Josef Sivic

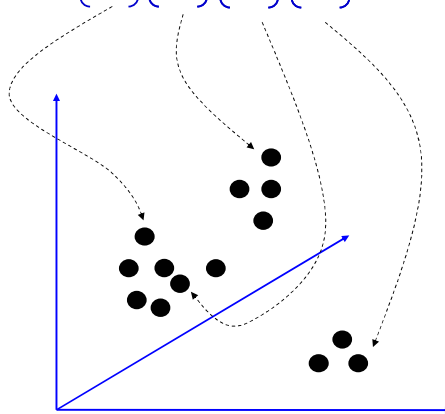
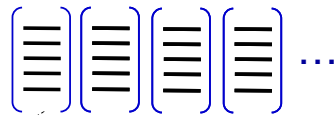
## Interest Point Features



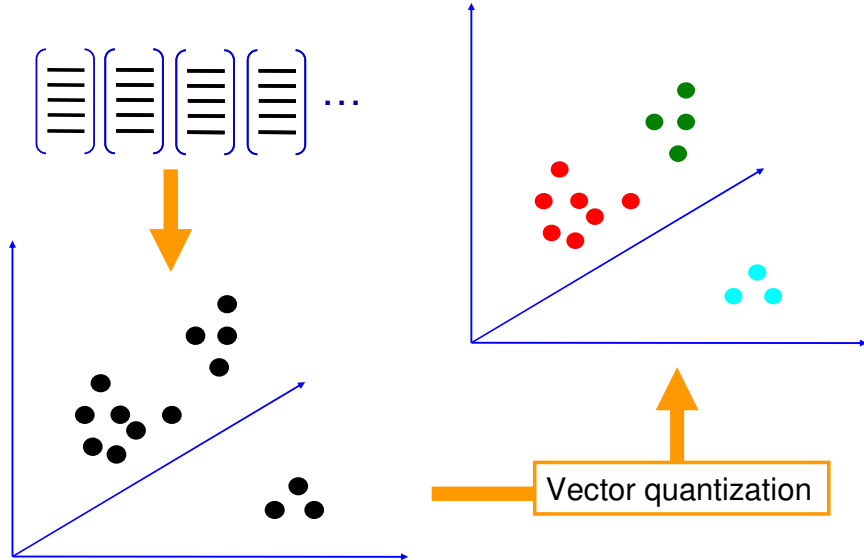
## Patch Features



## Dictionary Formation

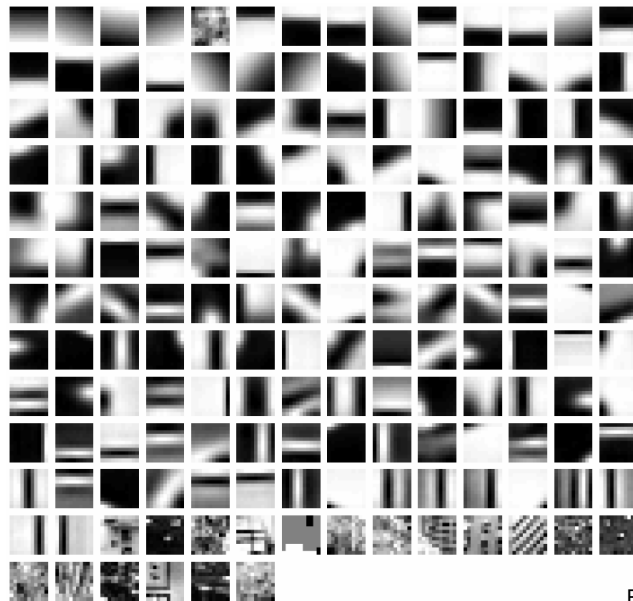


## Clustering (usually k-Means)



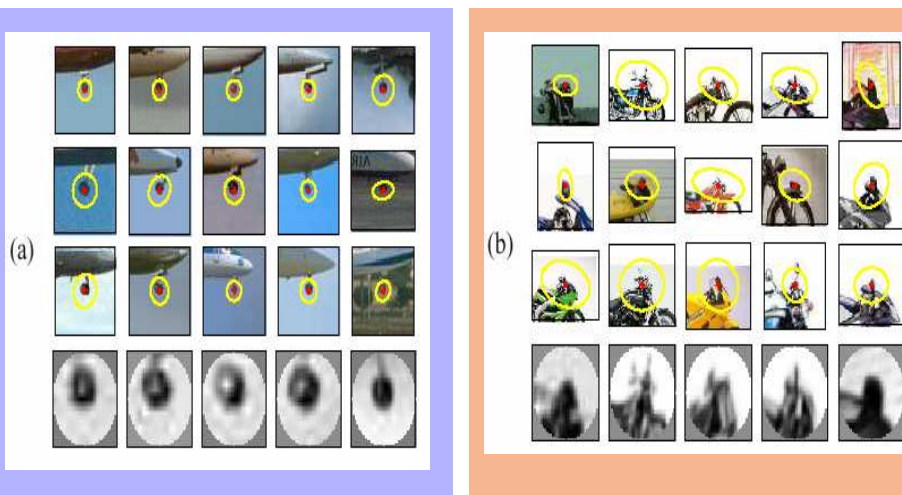
Slide credit: Josef Sivic

## Clustered Image Patches



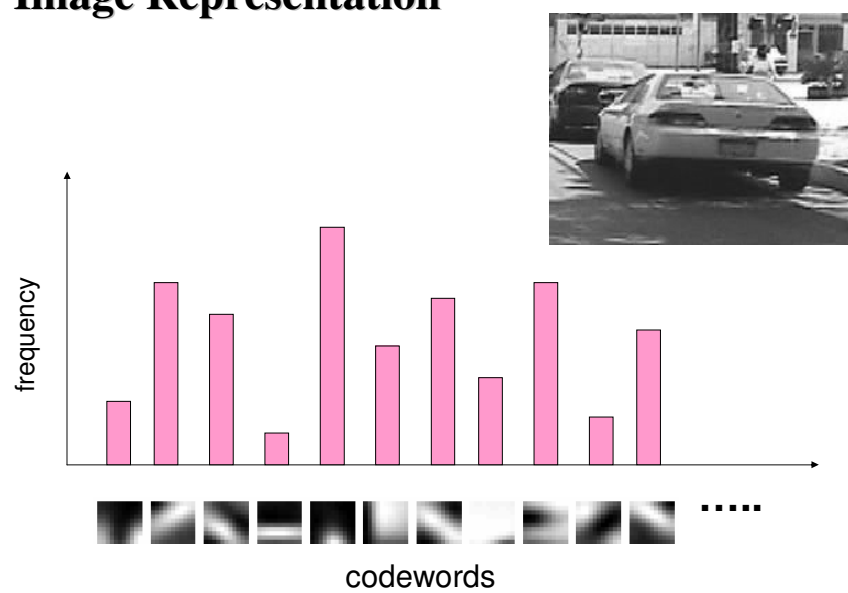
Fei-Fei et al. 2005

## Image Patch Examples of Codewords



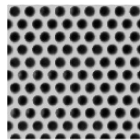
Sivic et al. 2005

## Image Representation



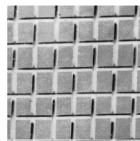
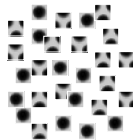
# 1. Local models for texture recognition

Training set



class 1

Feature  
extraction



class  $n$

Feature  
extraction

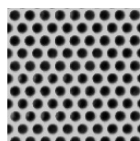


Bags of features

Lazebnik, Schmid & Ponce, CVPR 2003 and PAMI 2005

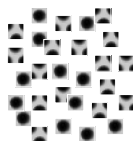
# 1. Local models for texture recognition

Training set



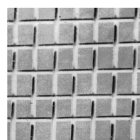
class 1

Feature  
extraction



bag of features"

Quantization



class  $n$

Feature  
extraction



Quantization

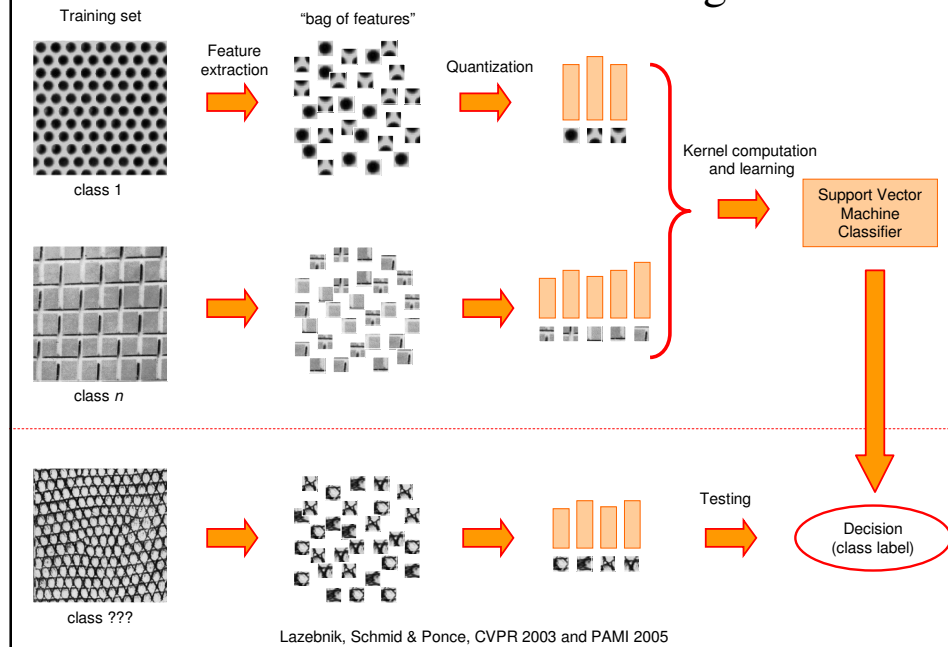


Kernel computation  
and learning

Support Vector  
Machine  
Classifier

Lazebnik, Schmid & Ponce, CVPR 2003 and PAMI 2005

# 1. Local models for texture recognition



## Local Models for Object Recognition

§ Serious limitations:

- § No spatial relations
- § No distinction between foreground and background
- § No localization capability





# Local Models for Object Recognition

## § Serious limitations:

- § No spatial relations
- § No distinction between foreground and background
- § No localization capability

## § And yet they work!

## Caltech6 dataset results

Object vs. background classification, ROC equal error rate



class	ours	other results		
	Zhang et al. (2005)	Willamowski et al. (2004)	Fergus et al. (2003)	
airplanes	<b>98.8</b>	97.1	90.2	
cars (rear)	98.3	<b>98.6</b>	90.3	
cars (side)	<b>95.0</b>	87.3	88.5	
faces	<b>100</b>	99.3	96.4	
motorbikes	<b>98.5</b>	98.0	92.5	
spotted cats	<b>97.0</b>	—	90.0	

bag of features

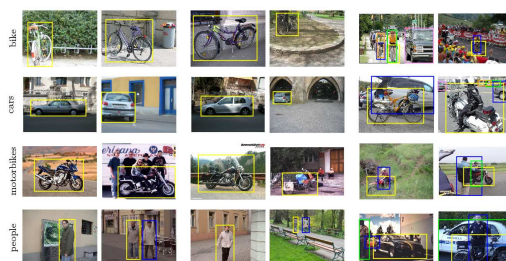
bag of features

constellation model

# Local Models for Object Recognition

## PASCAL 2005 challenge

<http://www.pascal-network.org/challenges/VOC>



Training: 684 images

Test set 1: 689 images

Test set 2: 956 images

class	test set 1	
	Zhang et al. (2005)	Larlus et al. (2006)
bikes	90.3	<b>93.0</b>
cars	93.0	<b>96.1</b>
motorbikes	96.2	<b>97.7</b>
people	91.6	<b>91.7</b>

class	test set 2	
	Zhang et al. (2005)	Deselaers et al. (2005)
bikes	<b>68.1</b>	66.7
cars	<b>74.1</b>	71.6
motorbikes	<b>79.7</b>	76.9
people	<b>75.3</b>	66.9

Object vs. background classification, ROC equal error rate

## § More comparisons: Xerox7, Graz, Caltech101, ...

## § The simplicity and effectiveness of the bag-of-features method make it a good baseline for evaluating novel approaches and datasets

# Object Recognition using Texture

## Object Categorization by Learned Universal Visual Dictionary

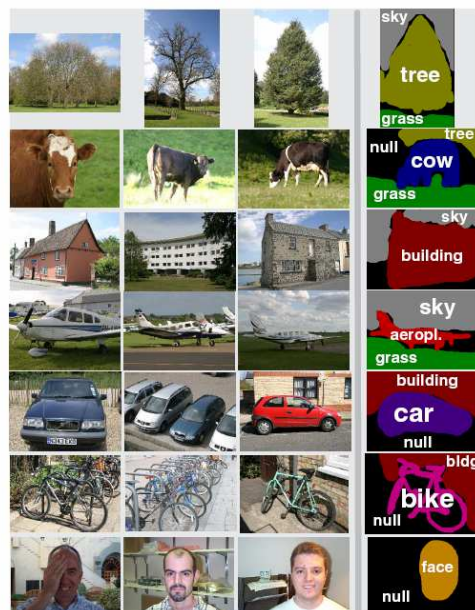
J. Winn, A. Criminisi and T. Minka

Microsoft Research, Cambridge, UK – <http://research.microsoft.com/vision/cambridge/recognition/>



## Learn Texture Model

- Representation:
  - Textons (rotation-variant)
- Clustering
  - K=2000
  - Then clever merging
  - Then fitting histogram with Gaussian
- Training
  - Labeled class data



## Results Movie



## Simple Works Well

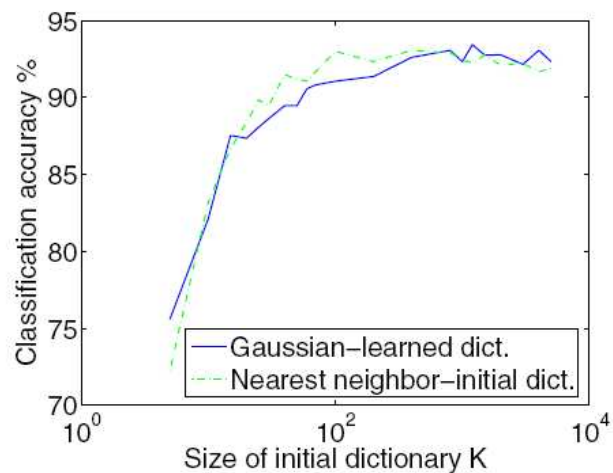
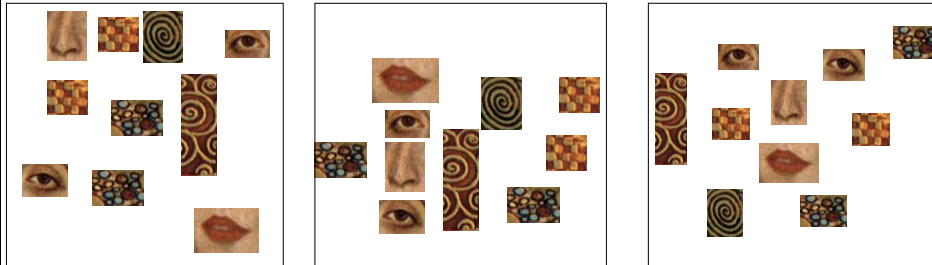


Figure 5: Comparing classification performance for Gaussian class models vs nearest neighbours classification.

## Problem with Bag of Words



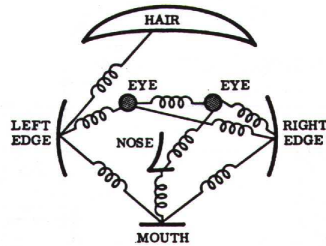
- § All have equal probability for bag-of-words methods
- § Location information is important

## Approach 3: Generative Methods using Part-Based Models

- § An object in an image is represented by a collection of parts, characterized by both their visual appearances and locations
- § Object categories are modeled by the appearance and spatial distributions of these characteristic parts
- § Issues for such models include efficient methods for finding correspondences between the object and the scene



## Model: Constellation of Parts



Fischler & Elschlager, 1973

Yuille, 1991

Brunelli & Poggio, 1993

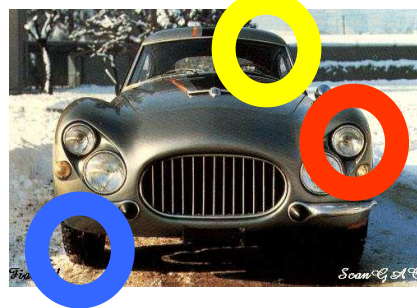
Lades, v.d. Malsburg et al. 1993

Cootes, Lanitis, Taylor et al. 1995

Amit & Geman, 1995, 1999

Perona et al. 1995, 1996, 1998, 2000

Felzenszwalb & Huttenlocher, 2000



## Representation

§ Object as set of parts

§ Generative representation

§ Model:

§ Relative locations between parts

§ Appearance of part

§ Issues:

§ How to model location

§ How to represent appearance

§ Sparse or dense (pixels or regions)

§ How to handle occlusion/clutter

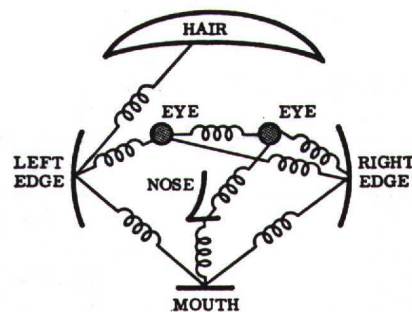
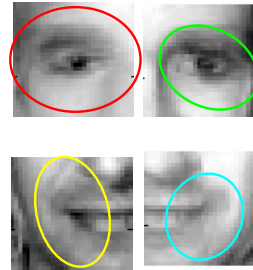


Figure from [Fischler73]

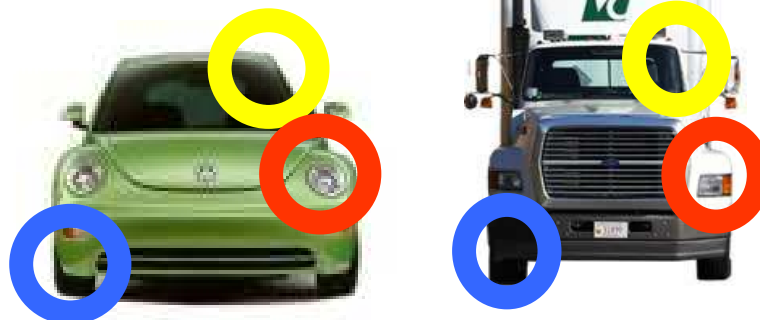
## Model Structure

- § Model **shape** using Gaussian distribution on image location between parts and scale of each part
- § Model **appearance** as patches of pixel intensities
- § Represent object class as graph of  $P$  image patches with parameters  $\theta$



## Sparse Representation

- § + Computationally tractable ( $10^5$  pixels     $10^1$  --  $10^2$  parts)
- § + Generative representation of class
- § + Avoid modeling global variability
- § + Success in specific object recognition

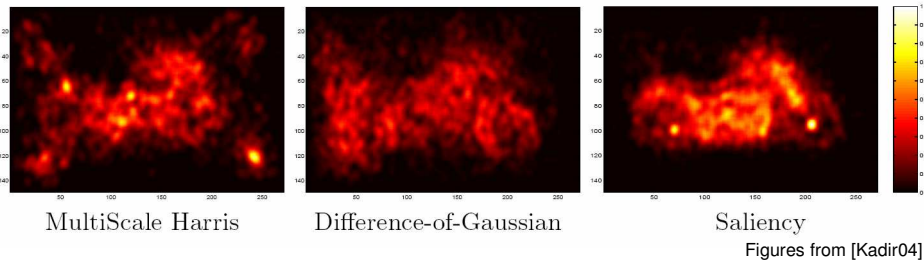
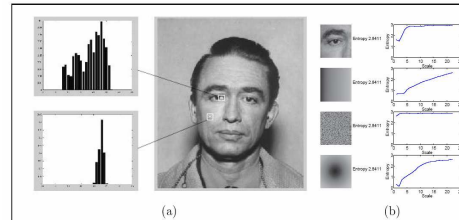


- § - Throws away most image information
- § - Parts need to be distinctive to separate from other classes



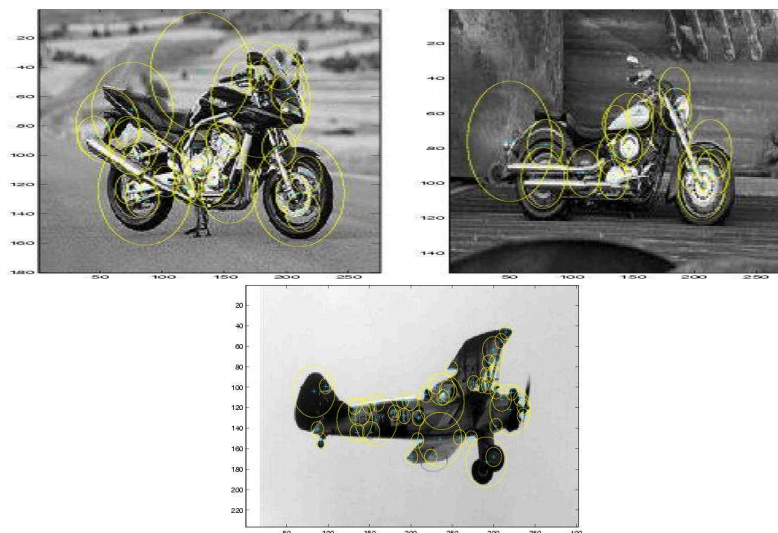
# Regions or Pixels?

- § # Regions << # Pixels
- § Regions increase tractability but lose information
- § Generally use regions:
  - § Local maxima of interest operators
  - § Can give scale/orientation invariance

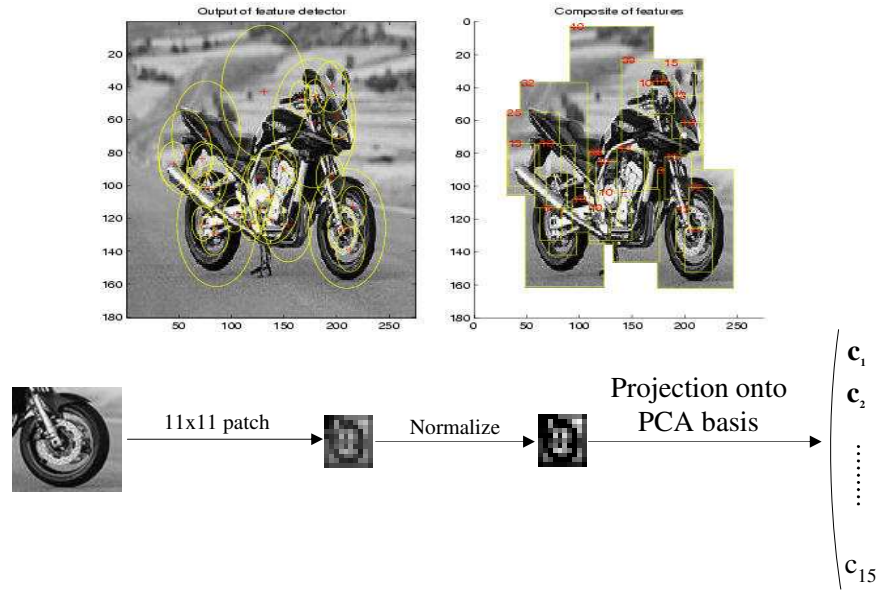


## Interest Operator

Kadir and Brady's interest operator  
Finds maxima in entropy over scale and location



## Representation of Appearance



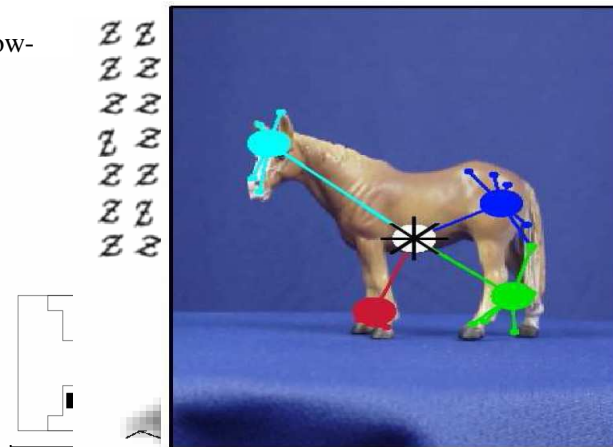
## Hierarchical Representations

§ Pixels   Pixel groupings   Parts   Object

§ Multi-scale approach increases number of low-level features

§ [Amit98]

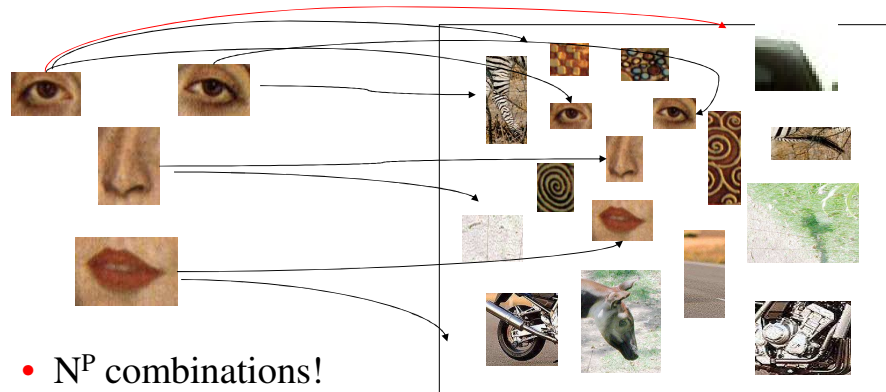
§ [Bouchard05]



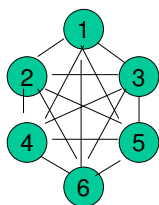
Images from [Amit98,Bouchard05]

## The Correspondence Problem

- Model with  $P$  parts
- Image with  $N$  possible locations for each part

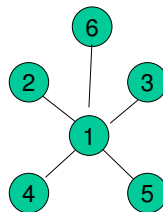


## Different Graph Structures



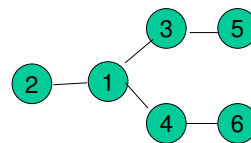
Fully connected

$$O(N^6)$$



Star structure

$$O(N^2)$$



Tree structure

$$O(N^2)$$

- Sparser graphs cannot capture all interactions between parts

## Some Class-Specific Graphs

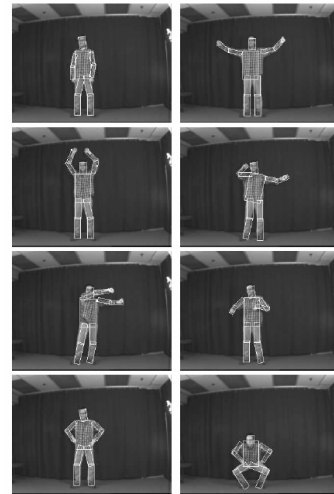
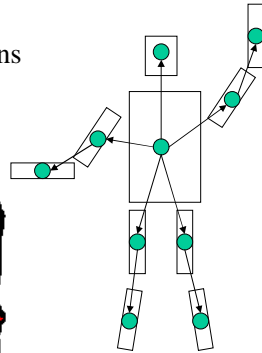
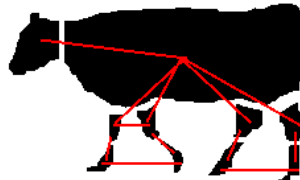
- § Articulated motion

- § People

- § Animals

- § Special parameterizations

- § Limb angles



Images from [Kumar05, Felzenszwalb05]

## Linear-Time Matching Algorithm

- § A *Dynamic Programming* implementation runs in **quadratic time**

- § Requires **tree configuration of parts**

- § Felzenszwalb & Huttenlocher (2000) developed **linear-time** matching algorithm

- § Additional constraint on part-to-part cost function  $d_{ij}$

- § **Basic “Trick”**: Parallelize minimization computation over entire image using a Generalized Distance Transform

# Distance Transforms

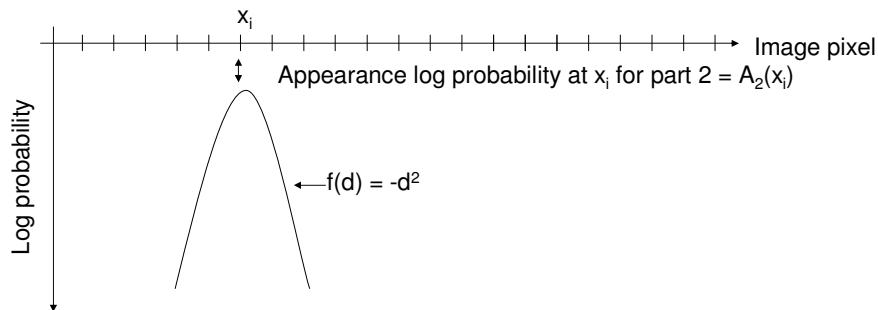
## Distance transforms

$O(N^2P)$   $O(NP)$  for tree structured models

## How it works

Assume location model is Gaussian (i.e.  $e^{-d^2}$ )

Consider a two part model with  $\mu=0$ ,  $\sigma=1$  on a 1-D image



Model



# Distance Transforms 2

For each position of landmark part, find best position for part 2

Finding most probable  $x_i$  is equivalent finding maximum over set of offset parabolas

Upper envelope computed in  $O(N)$  rather than obvious  $O(N^2)$  via distance transform [Felzenszwalb and Huttenlocher '05]

Add  $A_L(x)$  to upper envelope (offset by  $\mu$ ) to get overall probability map

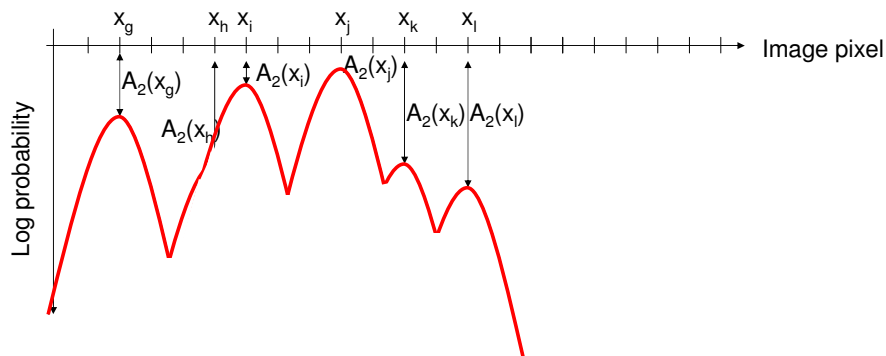


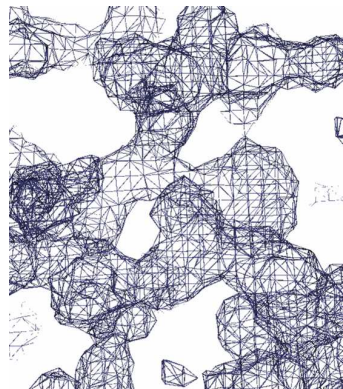


Figure from "Efficient Matching of Pictorial Structures," P. Felzenszwalb and D. Huttenlocher, *Proc. Computer Vision and Pattern Recognition Conf.*, 2000

## Using Pictorial Structures to Identify Proteins in X-ray Crystallographic Electron Density Maps

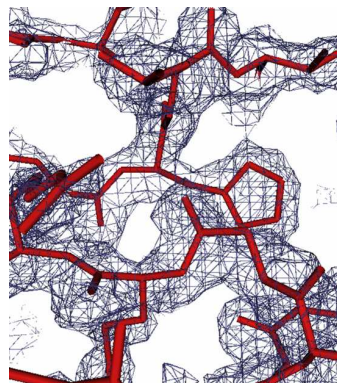
Frank DiMaio  
Jude Shavlik  
George N. Phillips, Jr.

## Task Overview



### Given

- Electron density for a region in a protein
- Protein's *topology*



### Find

- Atomic positions of individual atoms in the density map

## Pictorial Structures for Map Interpretation

**Basic Idea:** Build pictorial structure that is able to model *all configurations of a molecule*

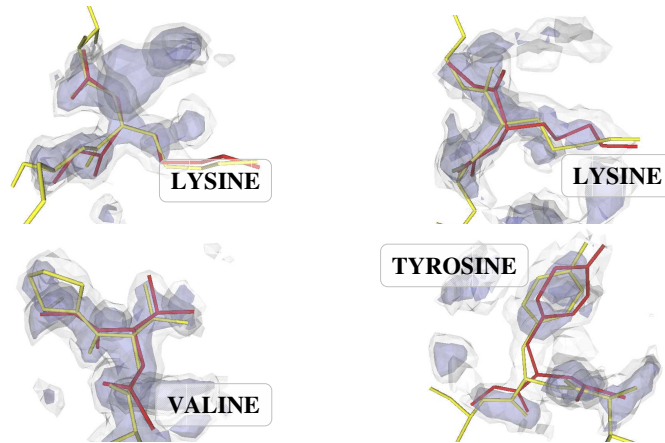
- § Each part in “collection of parts” corresponds to an **atom**
- § Model has **low-cost conformation** for **low-energy states** of the molecule





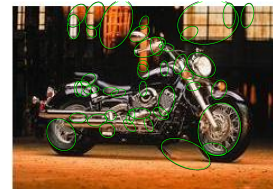
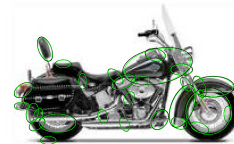
## Results

§ **PREDICTED** vs. **ACTUAL**



## Representation of Appearance

- § Invariance needs to match that of shape model
- § Insensitive to small shifts in translation/scale
  - § Compensate for jitter of features
  - § e.g. SIFT
- § Illumination invariance
  - § Normalize out
  - § Condition on illumination of landmark part



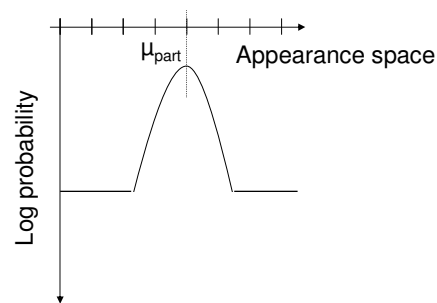
## Representation of Occlusion

### § Explicit

- § Additional match of each part to missing state

### § Implicit

- § Truncated minimum probability of appearance



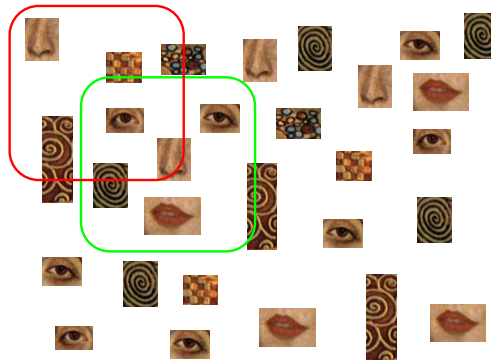
## Representation of Background Clutter

### § Explicit model

- § Generative model for clutter as well as foreground object

### § Use a sub-window

- § At correct position,  
no clutter is present



## Object Categorization: The Statistical Viewpoint



$$p(\text{zebra} | \text{image})$$

vs.

$$p(\text{no zebra} | \text{image})$$

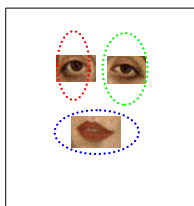
§ Bayes's rule:

$$\underbrace{\frac{p(\text{zebra} | \text{image})}{p(\text{no zebra} | \text{image})}}_{\text{posterior ratio}} = \underbrace{\frac{p(\text{image} | \text{zebra})}{p(\text{image} | \text{no zebra})}}_{\text{likelihood ratio}} \cdot \underbrace{\frac{p(\text{zebra})}{p(\text{no zebra})}}_{\text{prior ratio}}$$

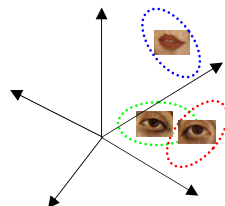
## Generative Probabilistic Model

Object model

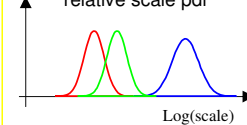
Gaussian shape pdf



Gaussian part appearance pdf



Gaussian relative scale pdf



Prob. of detection

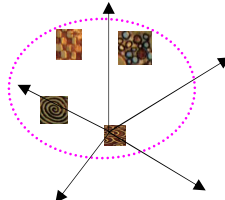


Background clutter model

Uniform shape pdf



Gaussian appearance pdf



Uniform relative scale pdf



Poisson pdf on # detections

## Model Structure

- Assume prior ratio is known or learned
- Find values for parameters  $\theta$  that maximizes the likelihood ratio

$$p(X, S, A | \theta) = \sum_{h \in H} p(X, S, A, h | \theta)$$

- $H$  is the set of all valid correspondences of image features to model parts, so  $|H| = O(N^P)$  in general
- Factor the likelihood to simplify computation (using Chain Rule)

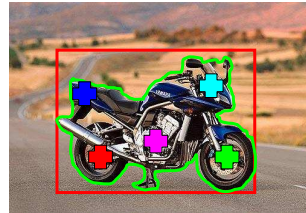
## Learning

## Learning Situations

### § Varying levels of supervision

- § Unsupervised
- § Image labels
- § Object centroid/bounding box
- § Segmented object
- § Manual correspondence (typically sub-optimal)

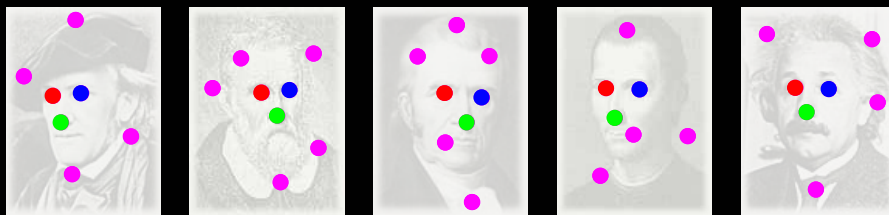
Contains a motorbike



- § Generative models naturally incorporate labelling information (or lack of it)
- § Discriminative schemes require labels for all data points

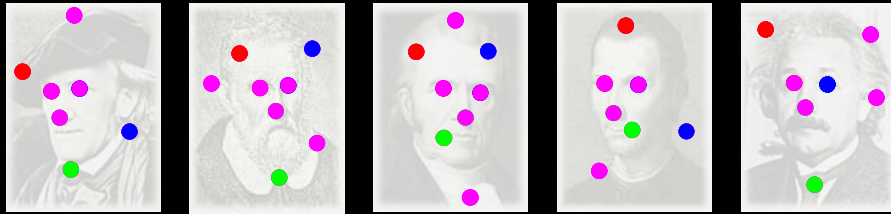
## Learning using EM

- Task: Estimation of model parameters
- Chicken and Egg type problem, since we initially know neither:
  - Model parameters
  - Assignment of regions to parts
- Let the assignments be a hidden variable and use EM algorithm to learn them and the model parameters



## Learning procedure

- Find regions & their location & appearance
- Initialize model parameters
- Use EM algorithm and iterate to convergence:
  - E-step: Compute assignments for which regions belong to which part (red, green and blue dots)
  - M-step: Update model parameters
- Try to maximize likelihood – consistency in shape & appearance



## Recognition

- § For each of  $P$  parts, run template over all locations in image
- § Detect local maxima, giving possible locations of each part
- § Given learned model, find maximum likelihood ratio of  $p(\mathbf{X}, \mathbf{S}, \mathbf{A} | \theta) / p(\mathbf{X}, \mathbf{S}, \mathbf{A} | \theta_{bg})$  for all possible correspondences –  $O(N^2P)$  where  $N$  = number of locations of each part in image
- § If greater than a threshold, signify object detected

# Experimental Procedure

Two series of experiments:

1. Scale variant (using pre-scaled images)
2. Scale invariant

**$P = 6-7$**

**$N = 20-30$**

**20-30 parameters/part**

**10-15 PCA features**

**Datasets:**

§ Motorbikes, Faces, Spotted cats, Airplanes, Cars from behind and side

§ **200 - 800 images**



**Training**

§ 50% images

§ No identification of object within image

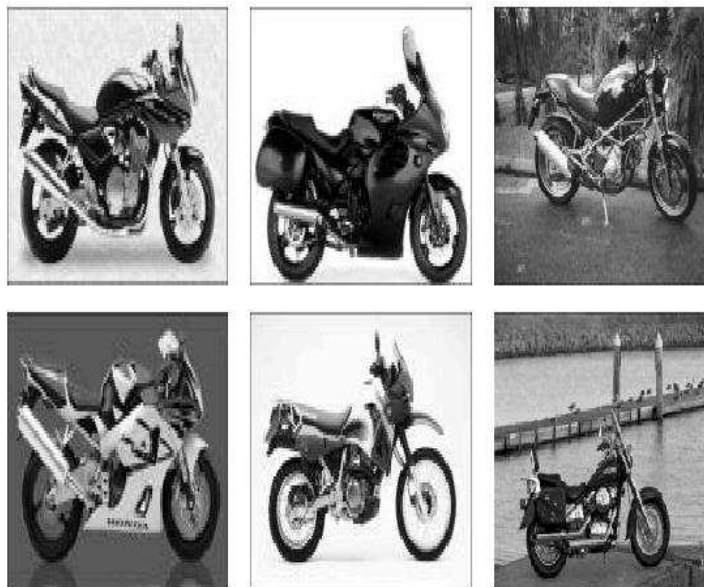
**Testing**

§ 50% images

§ Simple object present/absent test

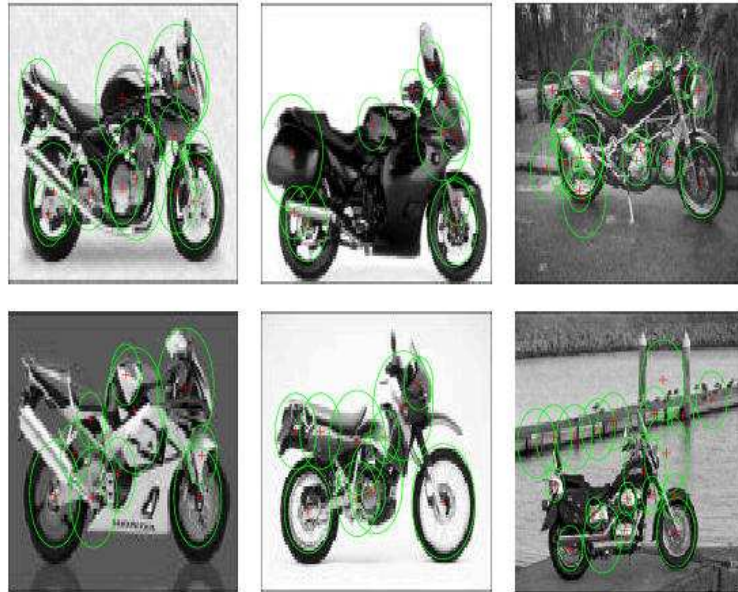
§ ROC equal error rate computed, using background set of images

## Motorbikes: Input Images

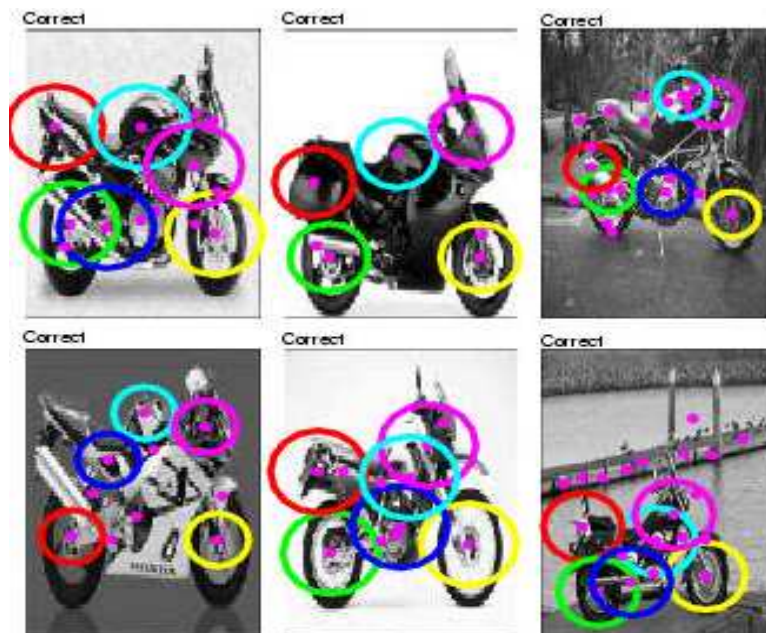


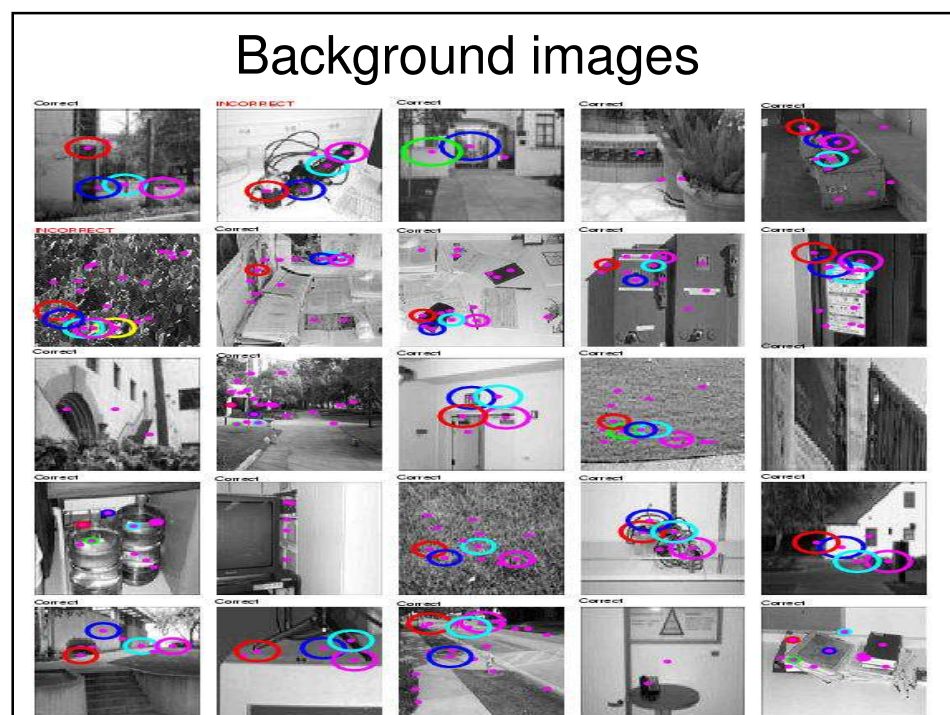
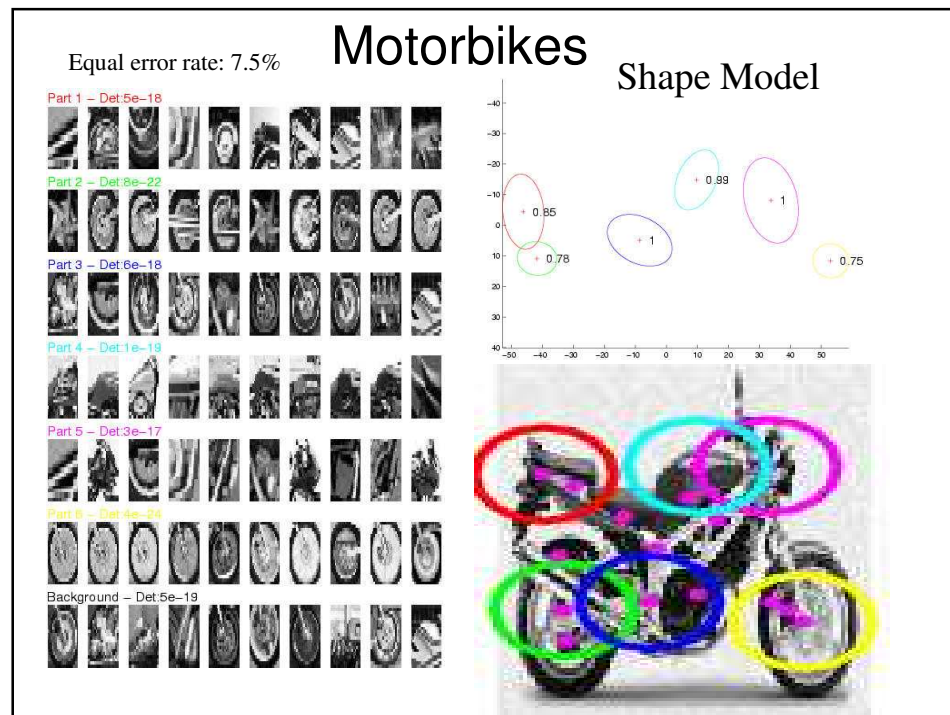


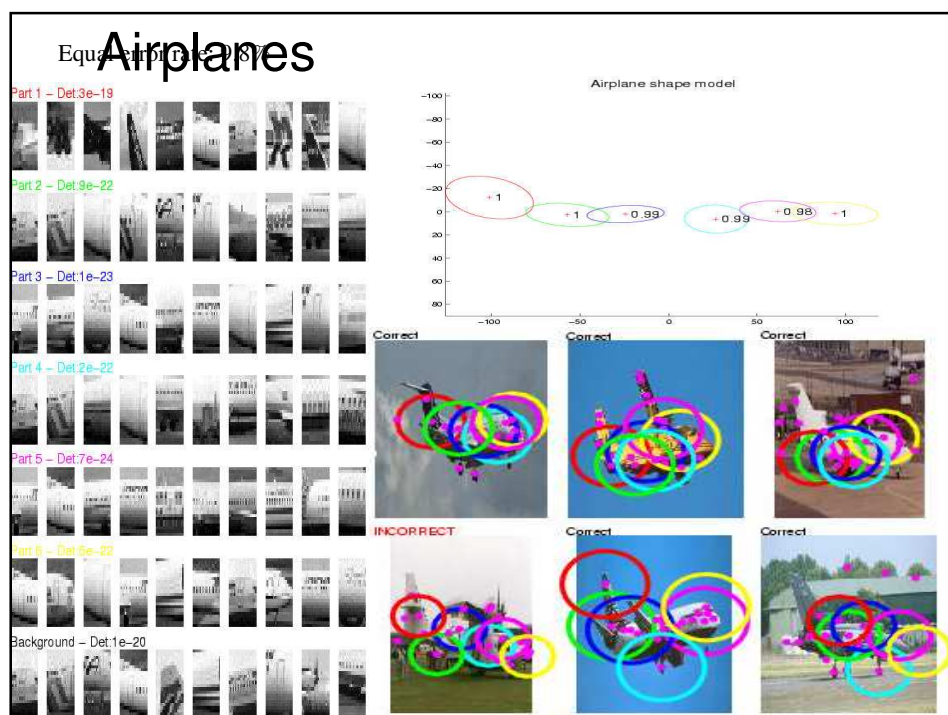
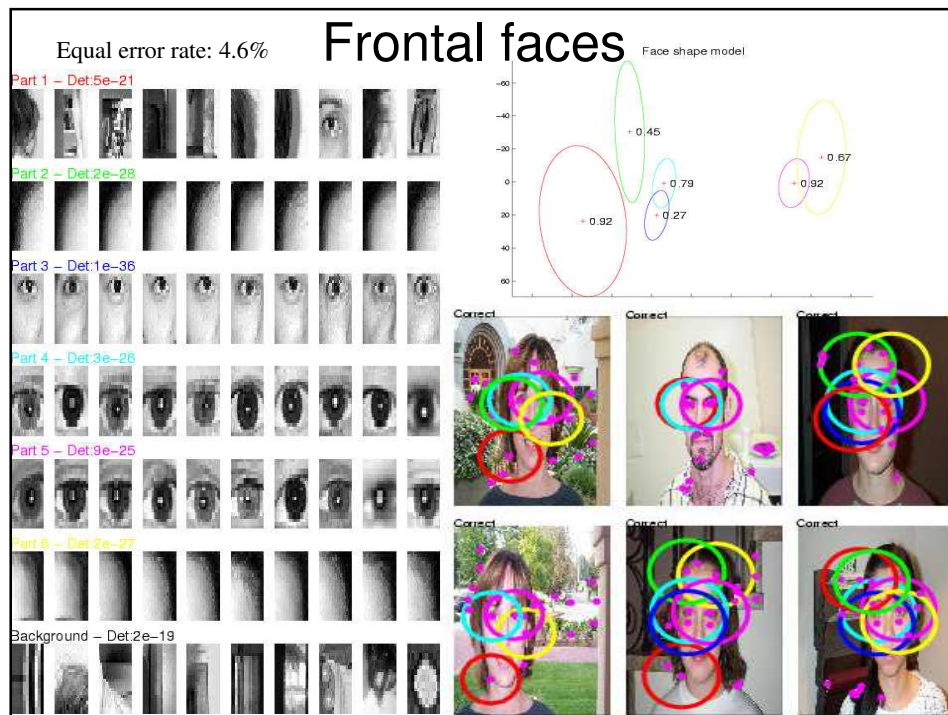
## Motorbikes: Features Detected



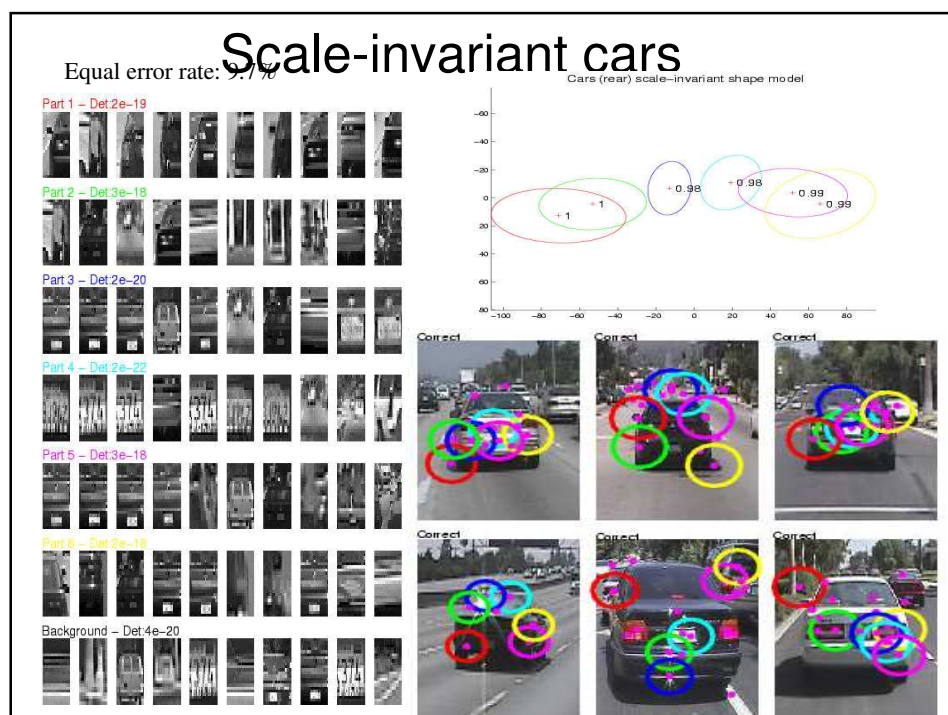
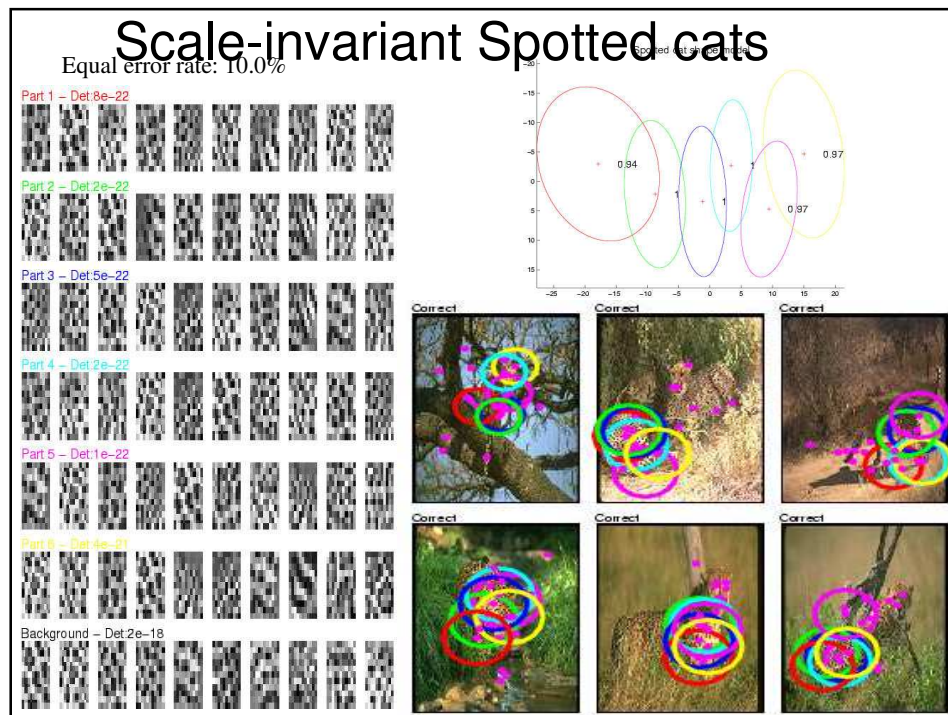
## Motorbikes: Max Likelihood Result



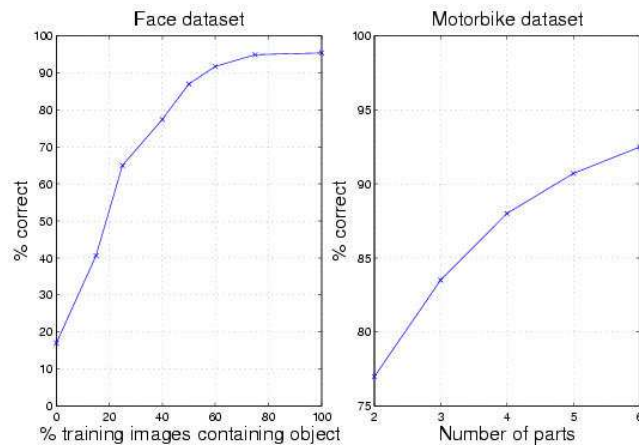








## Robustness of algorithm



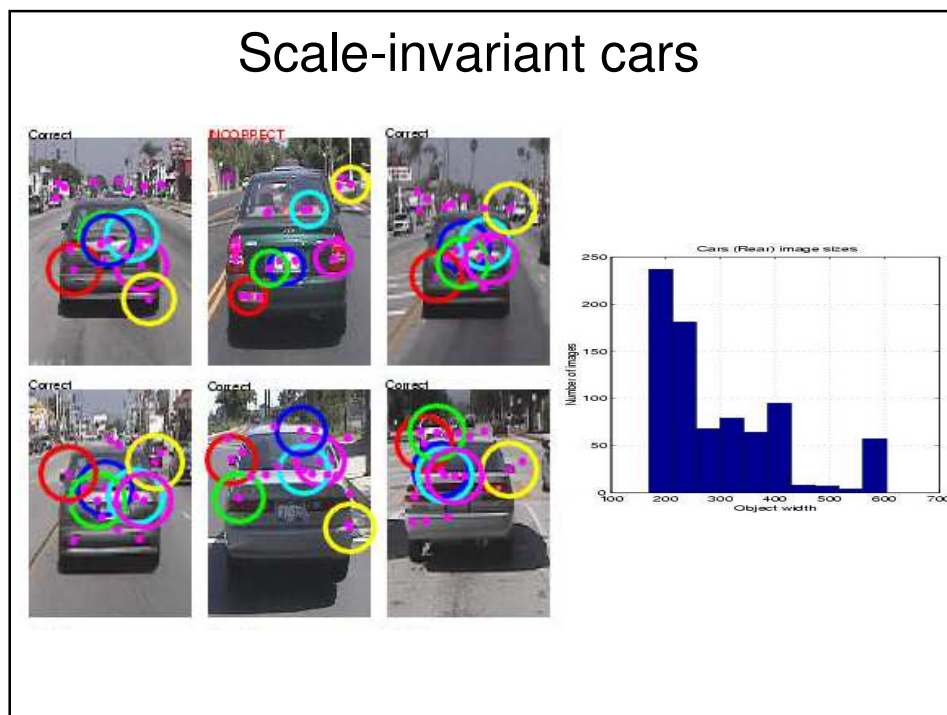
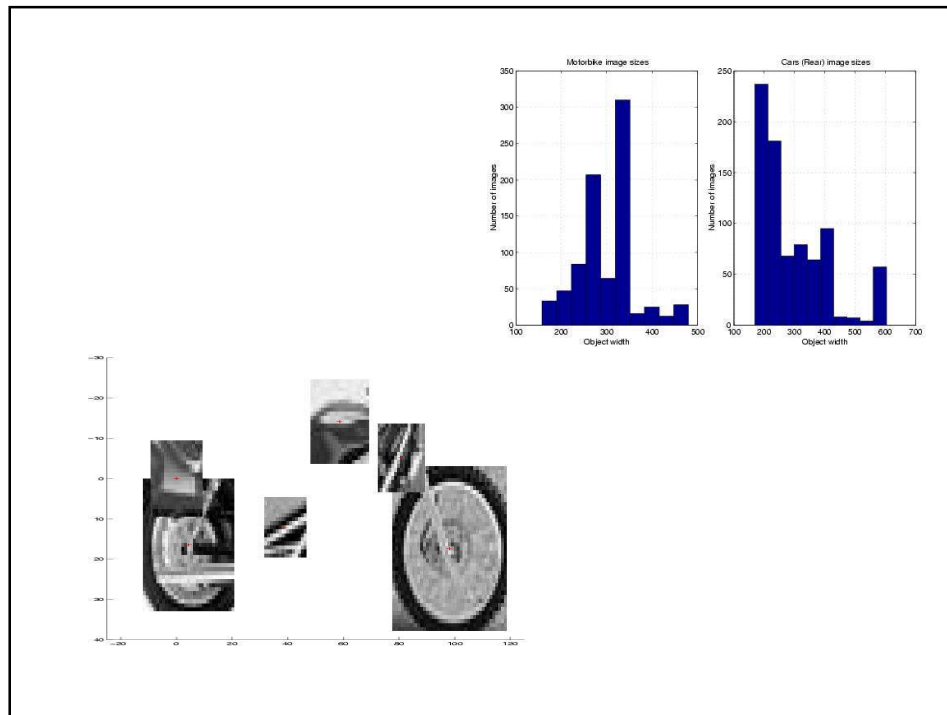
## ROC equal error rates

Pre-scaled data (identical settings):

Dataset	Total size of dataset	~ Object width (pixels)	Model			
			Motorbikes	Faces	Airplanes	Spotted Cats
Motorbikes	800	200	92.5	50	51	56
Faces	435	300	33	96.4	32	32
Airplanes	800	300	64	63	90.2	53
Spotted Cats	200	80	48	44	51	90.0

Scale-invariant learning and recognition:

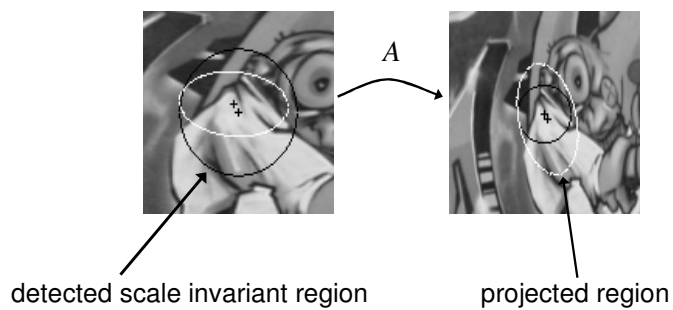
Dataset	Total size of dataset	Object size range (pixels)	Pre-scaled performance	Unscaled performance
Motorbikes	800	200-480	95.0	93.3
Airplanes	800	200-500	94.0	93.0
Cars (Rear)	800	100-550	84.8	90.3





## Adding Viewpoint Invariance

§ Locally approximated by an affine transformation





## Affine-Invariant Patches

Lindeberg & Garding (1997); Mikolajczyk & Schmid (2002);  
Tell & Carlsson (2000); Tuytelaars & Van Gool (2002)



Idea:

3D objects are never planar  
in the large, but they are  
always planar in the small



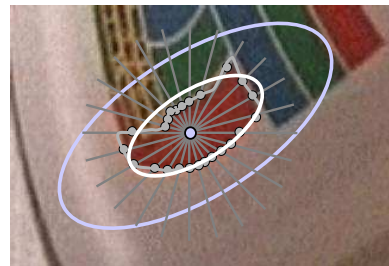
Representation: Local  
invariants and their  
spatial layout

## Intensity-based Method for Detecting Affine-Invariant Interest Points

Tuytelaars et al., 2000

1. Search for intensity extrema
2. Observe intensity profile along rays
3. Search for maximum of invariant  
function  $f(t)$  along each ray
4. Connect local maxima
5. Fit ellipse
6. Double ellipse size

$$f(t) = \frac{\text{abs}(I_0 - I)}{\max\left(\frac{\int \text{abs}(I_0 - I) dt}{t}, d\right)}$$



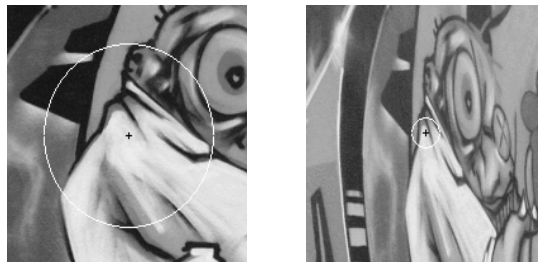
## Affine Invariant Harris Interest Points

- § Localization & scale influence affine neighborhood
  - § => affine invariant Harris points (Mikolajczyk & Schmid'02)
- § Iterative estimation of these parameters
  - § **localization** – local maximum of the Harris measure
  - § **scale** – automatic scale selection with the Laplacian
  - § **affine neighborhood** – normalization with second moment matrix
  - § Repeat estimation until convergence
- § Initialization with multi-scale interest points

## Affine invariant Harris points

- § Iterative estimation of localization, scale, neighborhood

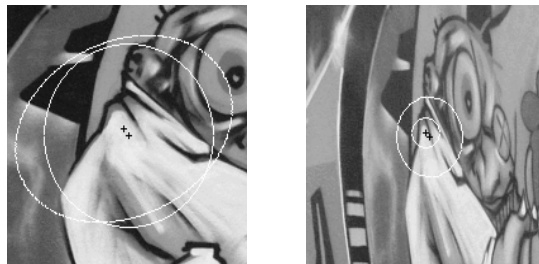
Initial points



## Affine invariant Harris points

- § Iterative estimation of localization, scale, neighborhood

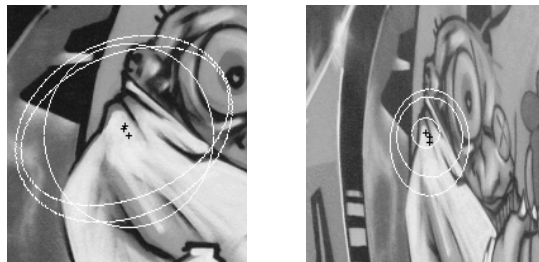
Iteration #1



## Affine invariant Harris points

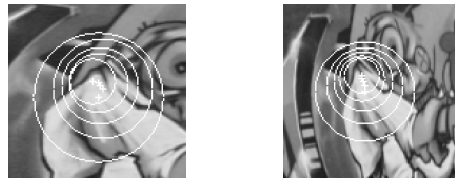
- § Iterative estimation of localization, scale, neighborhood

Iteration #2



## Affine invariant Harris points

§ Initialization with multi-scale interest points



§ Iterative modification of location, scale and neighborhood



## Affine Invariant Interest Point Detection



## Application: Image Retrieval

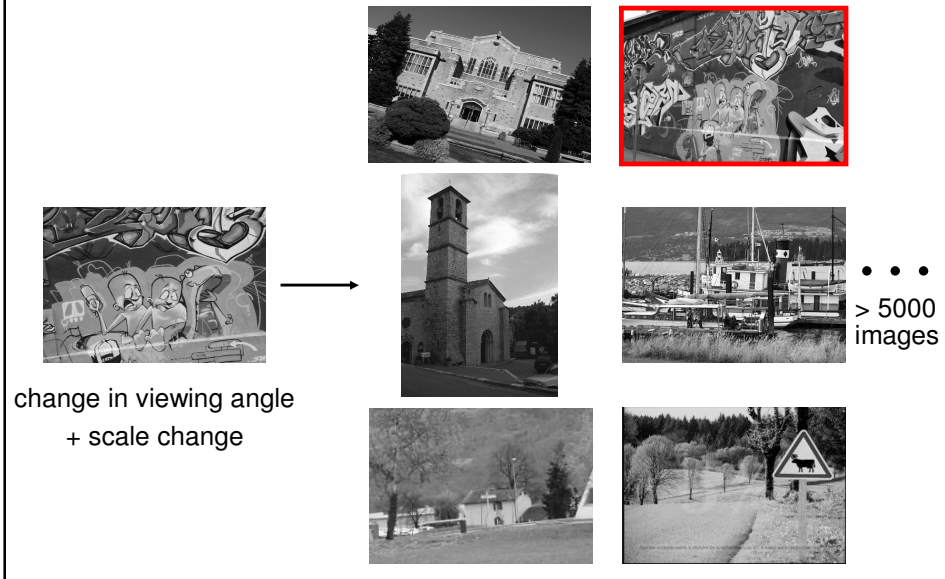


## Matches



22 correct matches

## Application: Image Retrieval



## Matches



33 correct matches





Figure 2: Model gallery: sample input images and renderings of the corresponding models.

Jean Ponce<sup>1</sup>, Svetlana Lazebnik<sup>1</sup>, Fredrick Rothganger<sup>1</sup>, Cordelia Schmid<sup>2</sup>

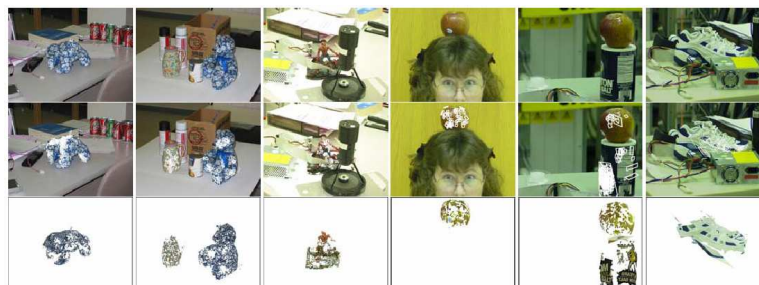


Figure 3: Object recognition experiments. The three rows of this figure show (respectively) input images, model patches matched to these images, and recognized models rendered in their estimated pose. Note that the teddy bear in the leftmost column is in a pose quite different from those used to acquire its model. Also note the significant amount of clutter and occlusion in each image.

Jean Ponce<sup>1</sup>, Svetlana Lazebnik<sup>1</sup>, Fredrick Rothganger<sup>1</sup>, Cordelia Schmid<sup>2</sup>



## Application: Photo Tourism

§ <http://phototour.cs.washington.edu/>

§ Detect and match local patch features across images of a scene taken by many different people and found via shared image databases such as Flickr

# Photo Tourism

Exploring photo collections in 3D

Noah Snavely   Steven M. Seitz   Richard Szeliski  
*University of Washington*   *Microsoft Research*

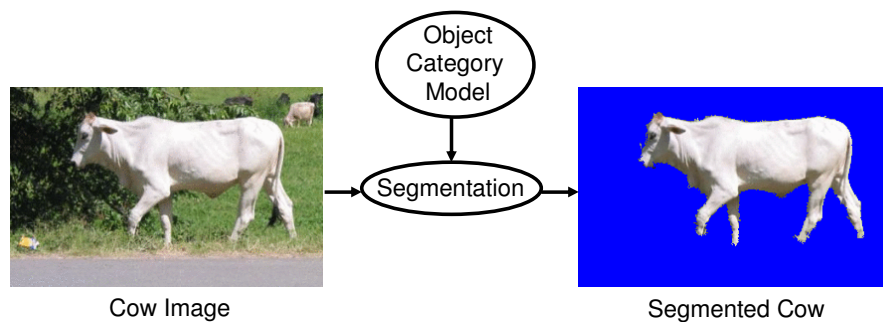
SIGGRAPH 2006

## Probabilistic Parts and Structure Models Summary

- § Correspondence problem
- § Efficient methods for large # parts and # positions in image
- § Challenge to get representation with desired invariance
- § Minimal supervision
  
- § Future directions:
  - § Multiple views
  - § Approaches to learning
  - § Multiple category training

## Combining Segmentation and Recognition

- § Example: Given an image and object category, segment the object



Segmentation should (ideally) be

- shaped like the object, e.g., cow-like
- obtained efficiently in an unsupervised manner
- able to handle self-occlusion