

# **Pset 4**

TA Session

David Held

Feb 28, 2014

# Overview

- Bag of Words
- Kernels
- Spatial Pyramid Matching
- Scene Classification
- Object Bank

# Overview

- Bag of Words
- Kernels
- Spatial Pyramid Matching
- Scene Classification
- Object Bank



# Part 1: Bag-of-words models

This segment is based on the tutorial "*Recognizing and Learning Object Categories: Year 2007*", by Prof A. Torralba, R. Fergus and F. Li

**Object**



**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 reach our eyes. For a long time, the retinal image was considered as a movie screen. It is now discovered that the perception is more complex, following the path to the various centers of the brain. Hubel and Wiesel have demonstrated that the *message about the image falling on the retina undergoes a point-by-point analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.*

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

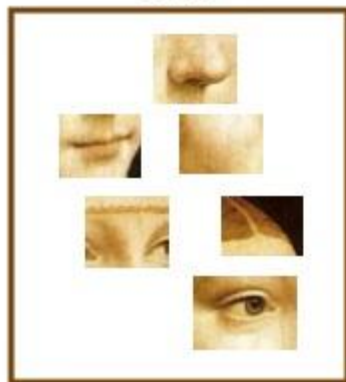
China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$575bn in 2004. The surplus is \$660bn. This surplus will also annoy the US because of China's deliberate policy to keep the yuan undervalued. The US government agrees that the yuan is undervalued and the government also needs to increase the demand so that the yuan can be used in the country. China has been permitted to trade within a narrow range but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

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

# definition of “BoW”

– Independent features

face



bike

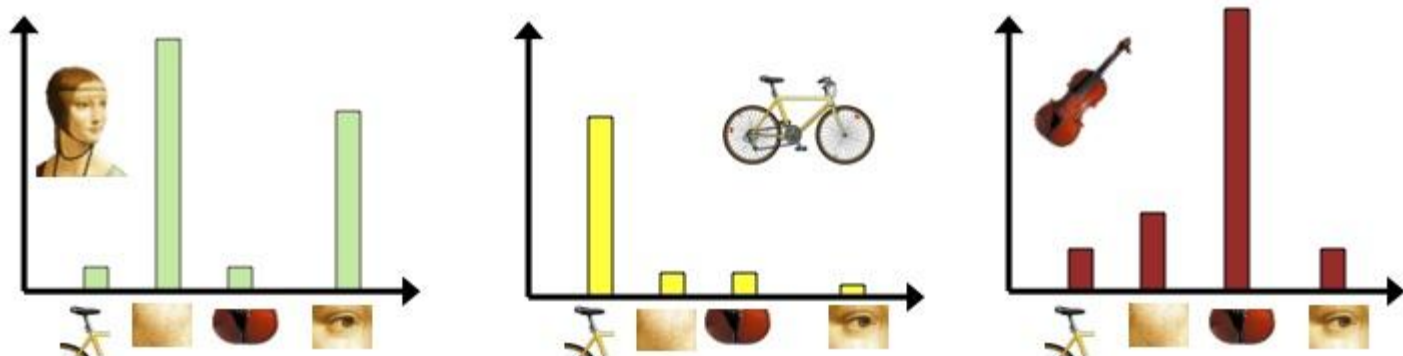


violin



# definition of “BoW”

- Independent features
- histogram representation



codewords dictionary



# Representation



feature detection  
& representation

**codewords dictionary**

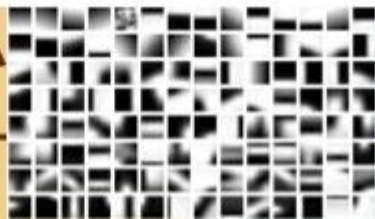


image representation



**learning**

**category models  
(and/or) classifiers**

# recognition

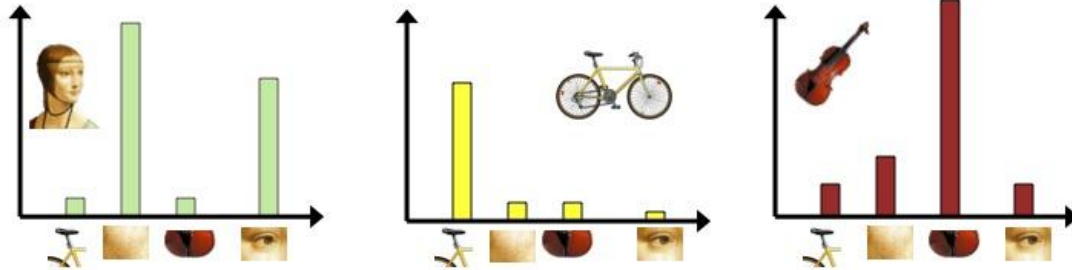


**category  
decision**

# Histogram Representation

## definition of “BoW”

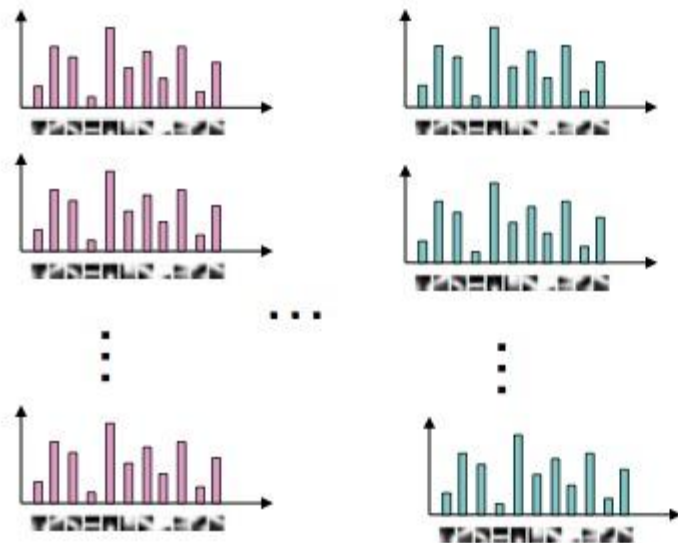
- Independent features
- histogram representation



codewords dictionary

# Discriminative classifiers

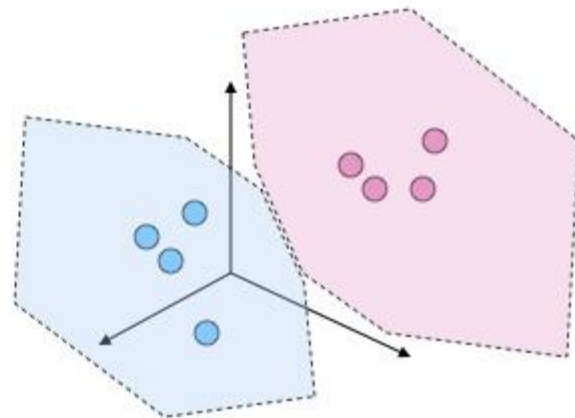
## category models



Class 1

Class N

## Model space

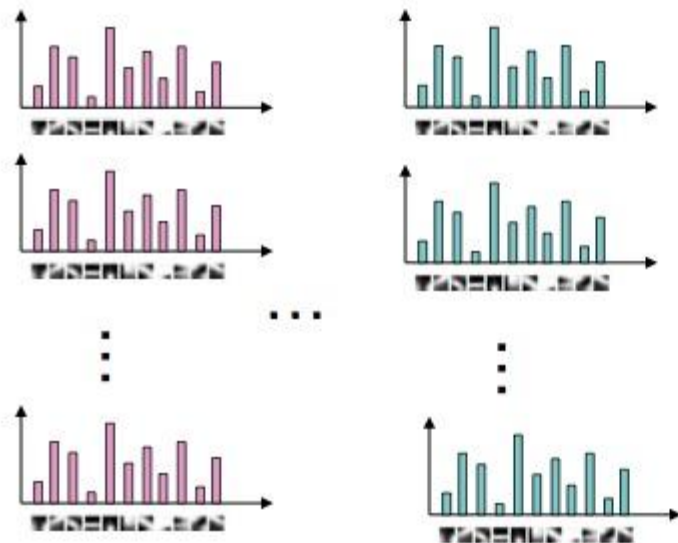


# Overview

- Bag of Words
- **Kernels**
- Spatial Pyramid Matching
- Scene Classification
- Object Bank

# Discriminative classifiers

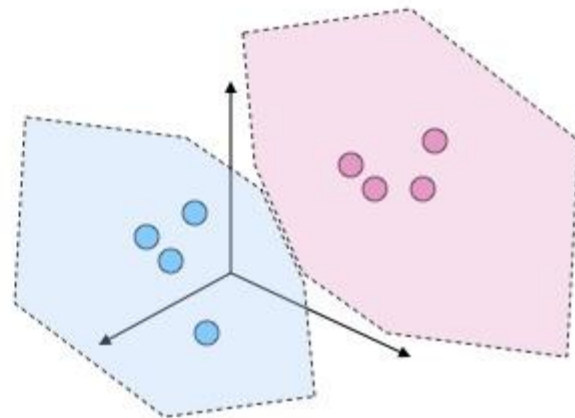
## category models



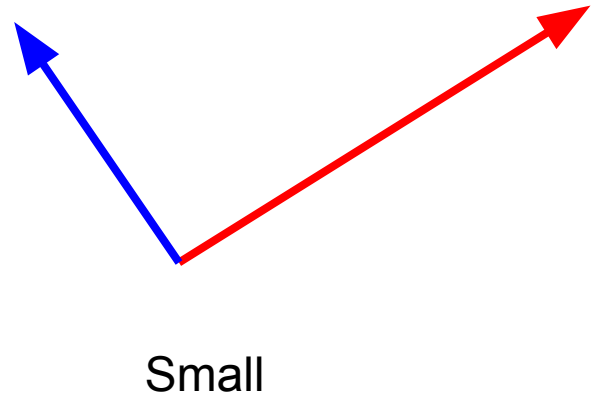
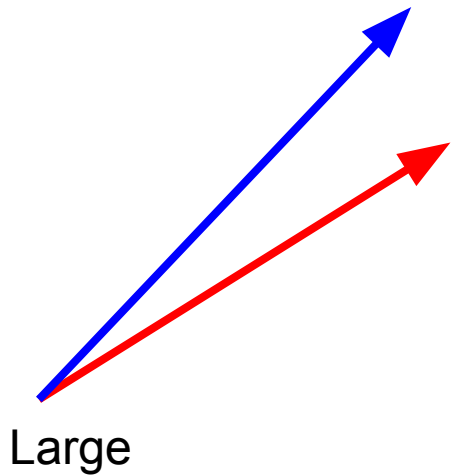
Class 1

Class N

## Model space



# How to compare vectors?



# Kernels

$$K(x, z) = x^T z \quad \text{Linear Kernel}$$

$$K(x, z) = (x^T z)^2$$

$$K(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right)$$

# Histogram Representation

## definition of “BoW”

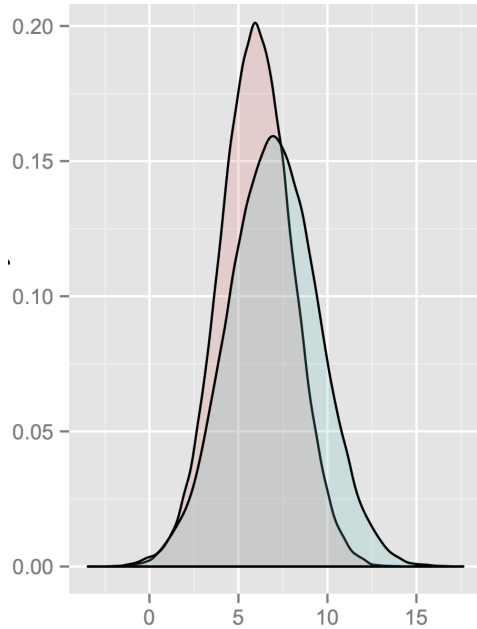
- Independent features
- histogram representation



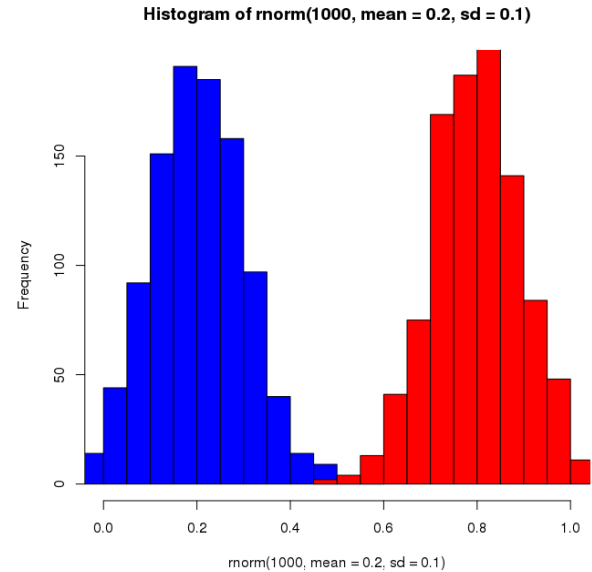
codewords dictionary



# How to Compare Histograms?



Large



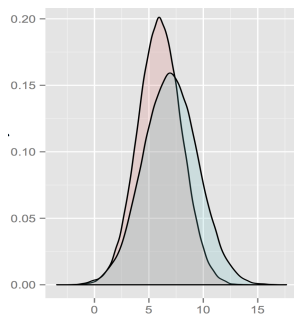
Small

# How to Compare Histograms?

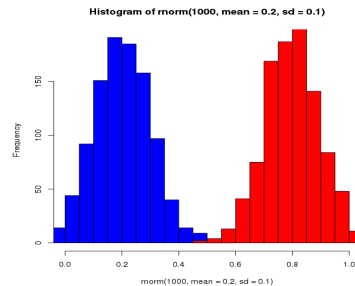
- Generalized Gaussian kernel:

$$K(h_1, h_2) = \exp\left(-\frac{1}{A} D(h_1, h_2)^2\right)$$

- $D$  can be Euclidean distance,  $\chi^2$  distance etc...



Large



Small

# Distance Between Histograms

## Functions for comparing histograms

Jan Puzicha, Yossi Rubner, Carlo Tomasi, Joachim M. Buhmann: [Empirical Evaluation of Dissimilarity Measures for Color and Texture](#). ICCV 1999

- L1 distance

$$D(h_1, h_2) = \sum_{i=1}^N |h_1(i) - h_2(i)|$$

- $\chi^2$  distance

$$D(h_1, h_2) = \sum_{i=1}^N \frac{(h_1(i) - h_2(i))^2}{h_1(i) + h_2(i)}$$

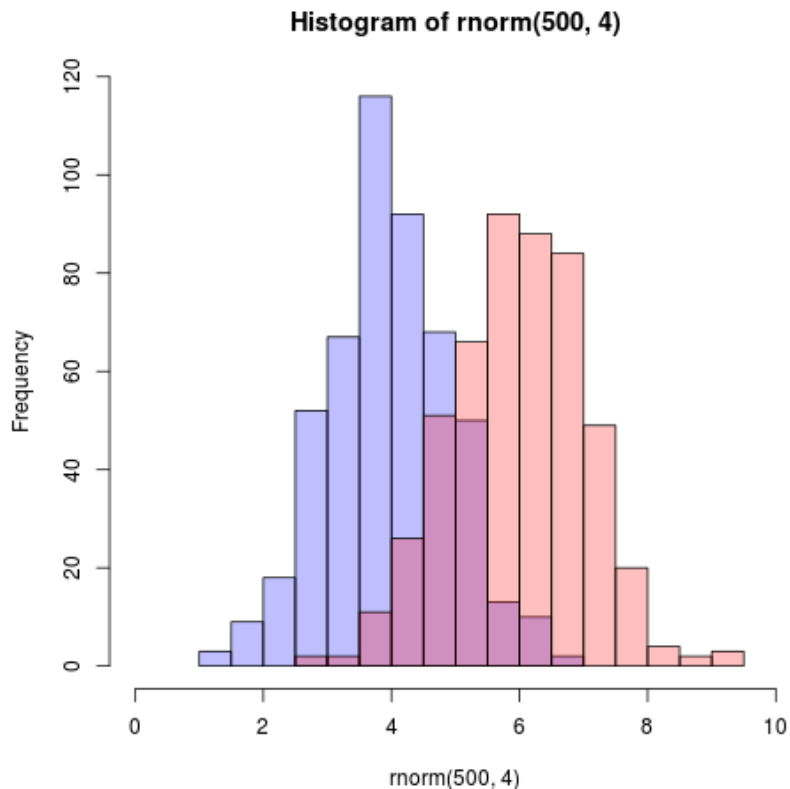
- Quadratic distance (*cross-bin*)

$$D(h_1, h_2) = \sum_{i,j} A_{ij} (h_1(i) - h_2(j))^2$$

# How to Compare Histograms

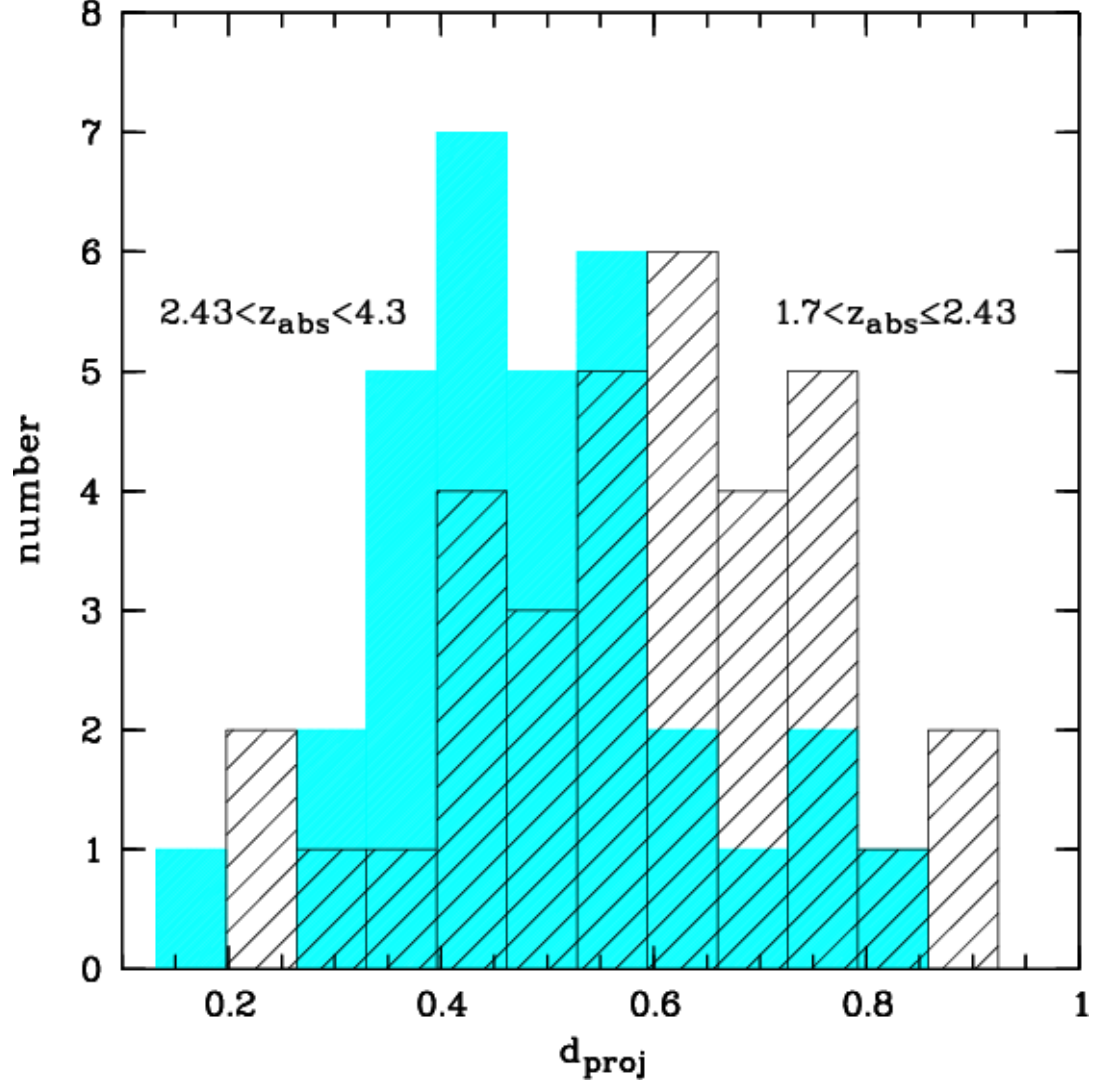
- Histogram intersection kernel:

$$I(h_1, h_2) = \sum_{i=1}^N \min(h_1(i), h_2(i))$$



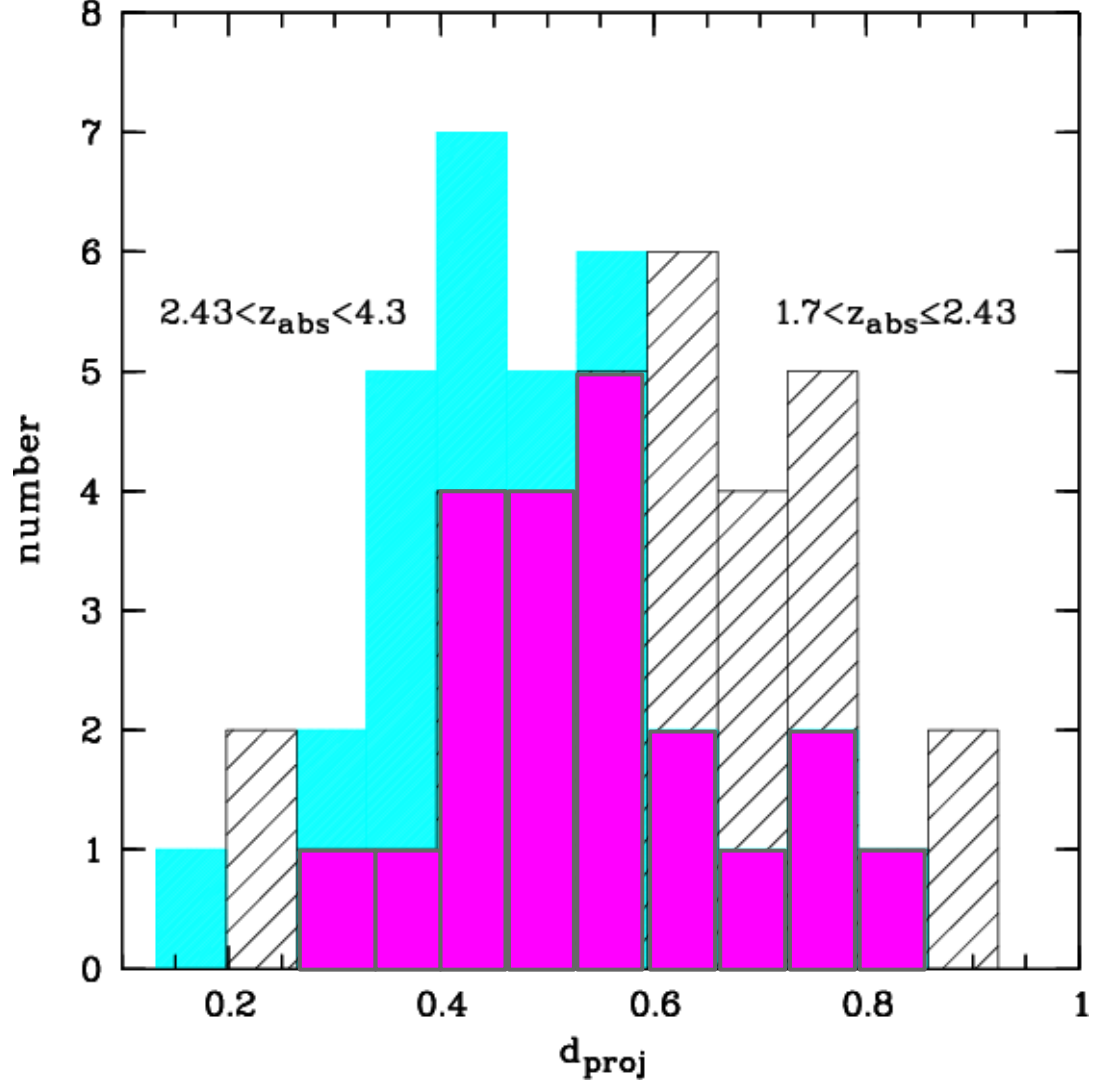
- Histogram intersection kernel:

$$I(h_1, h_2) = \sum_{i=1}^N \min(h_1(i), h_2(i))$$



- Histogram intersection kernel:

$$I(h_1, h_2) = \sum_{i=1}^N \min(h_1(i), h_2(i))$$



# Overview

- Bag of Words
- Kernels
- Spatial Pyramid Matching
- Scene Classification
- Object Bank

# Is this a good representation?

Bag of 'words'



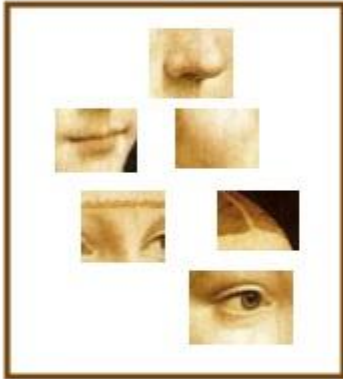


# Is this a good representation?

## definition of “BoW”

– Independent features

face

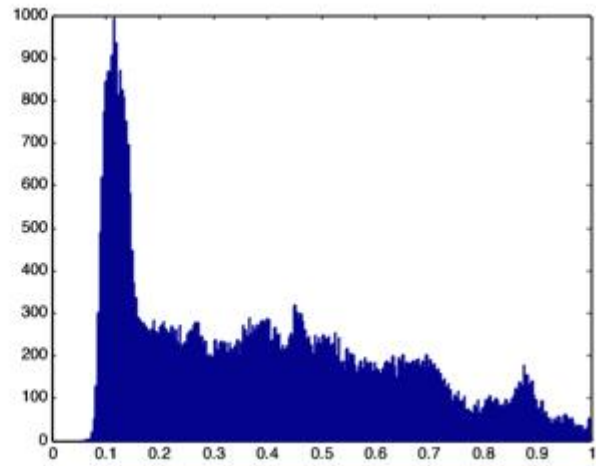


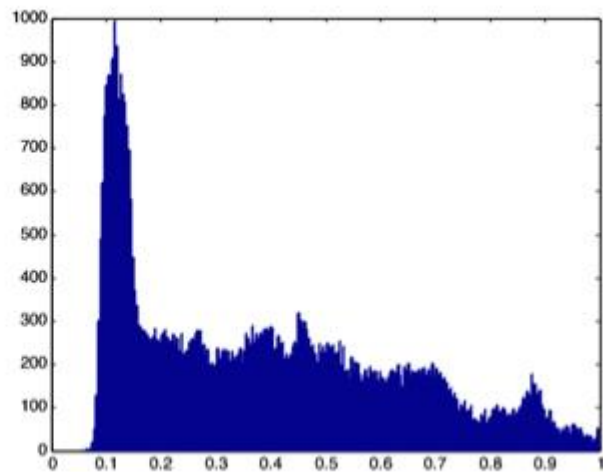
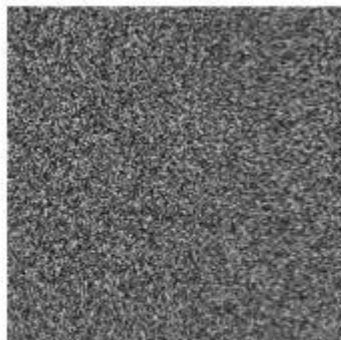
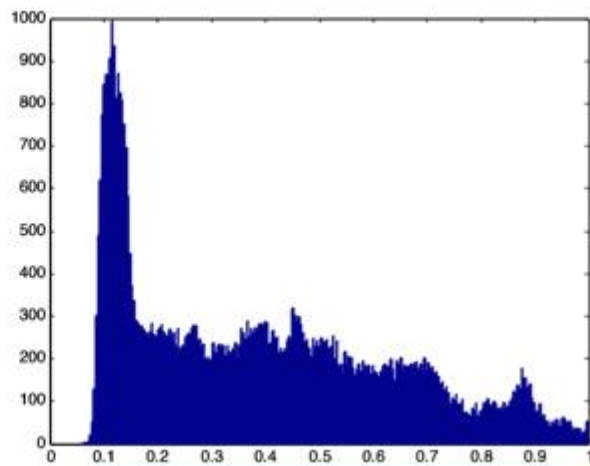
bike



violin



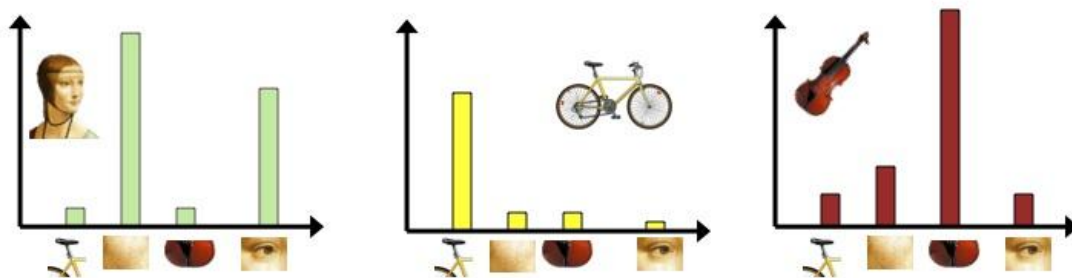




# Still Using a Histogram Representation

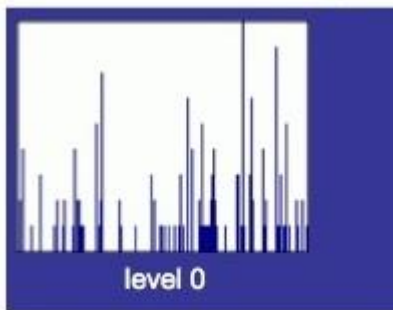
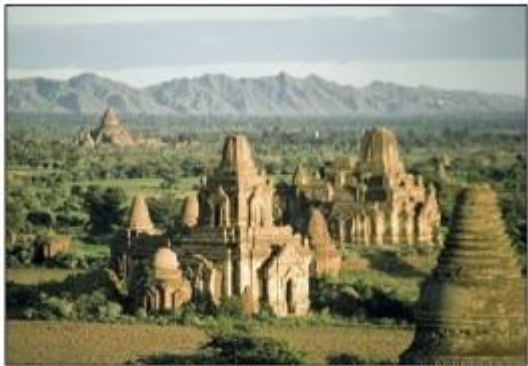
definition of “BoW”

- Independent features
- histogram representation



codewords dictionary

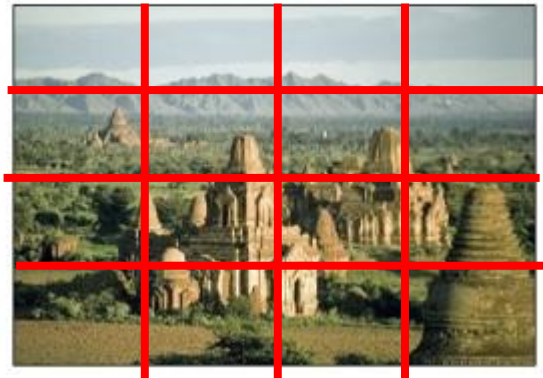
Feature extraction



Feature extraction



Feature extraction



# Pyramid matching

Indyk & Thaper (2003), Grauman & Darrell (2005)

Find maximum-weight matching (weight is inversely proportional to distance)

Original images



Feature histograms:

Level 3



Level 2



Level 1



Level 0



# Pyramid matching

Indyk & Thaper (2003), Grauman & Darrell (2005)

Find maximum-weight matching (weight is inversely proportional to distance)

Original images



Feature histograms:

Level 3



$\cap$



$= \mathcal{I}_3$

Level 2



$\cap$



$= \mathcal{I}_2$

Level 1



$\cap$



$= \mathcal{I}_1$

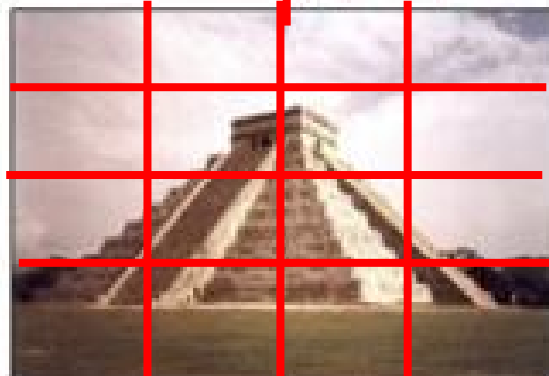
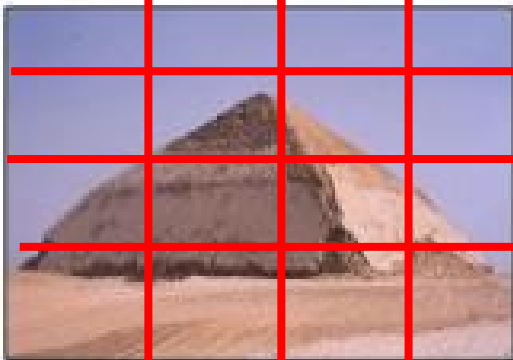
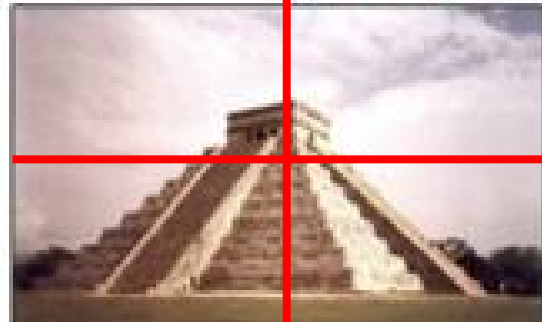
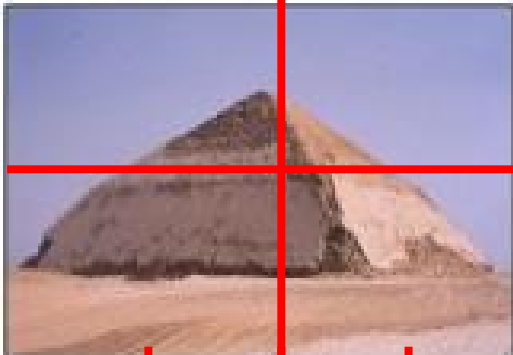
Level 0



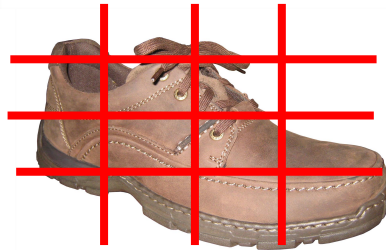
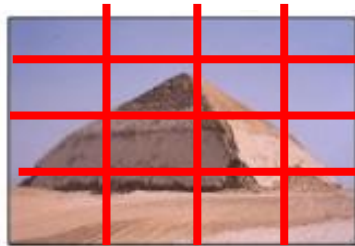
$\cap$



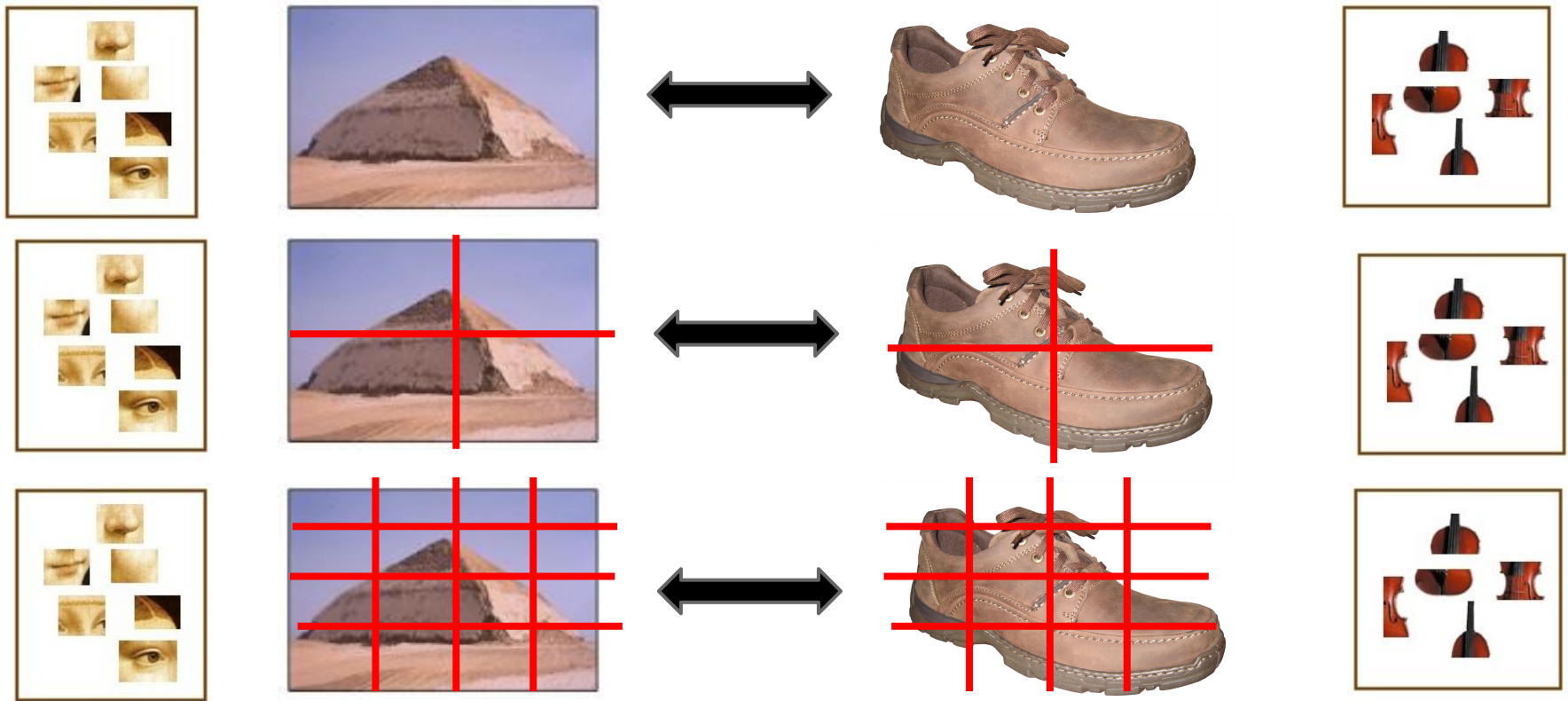
$= \mathcal{I}_0$







At which grid level will there be the most overlap due to random chance?



- Less of a chance of random overlap at finer grid levels
- Overlap at a larger grid size includes overlap at finer grid level

# Spatial Pyramid Matching Kernel

Total weight (value of *pyramid match kernel*):  $I_3 + \frac{1}{2}(I_2 - I_3) + \frac{1}{4}(I_1 - I_2) + \frac{1}{8}(I_0 - I_1)$

$$\begin{aligned}\kappa^L(X, Y) &= \mathcal{I}^L + \sum_{\ell=0}^{L-1} \frac{1}{2^{L-\ell}} (\mathcal{I}^\ell - \mathcal{I}^{\ell+1}) \\ &= \frac{1}{2^L} \mathcal{I}^0 + \sum_{\ell=1}^L \frac{1}{2^{L-\ell+1}} \mathcal{I}^\ell.\end{aligned}$$

- Less of a chance of random overlap at finer grid levels
- Overlap at a larger grid size includes overlap at finer grid level

# Overview

- Bag of Words
- Kernels
- Spatial Pyramid Matching
- Scene Classification
- Object Bank

# Scene Classification

beach



city street



elevator



forest fire



fountain



highway



lightning storm



ocean



railway



rushing river



sky-clouds



snowing



waterfall



windmill farm



# Scene Classification



office



kitchen



living room



bedroom



store



industrial



tall building



inside city



street



highway



coast



open country



mountain

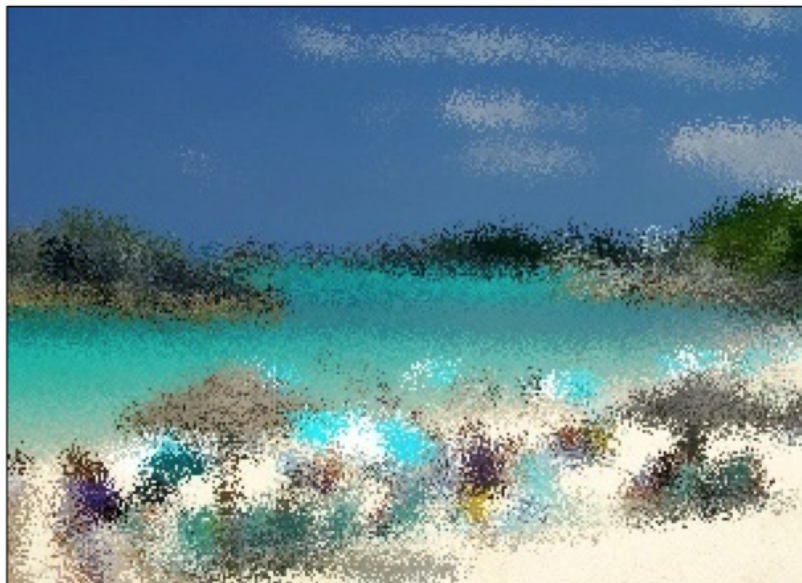


forest

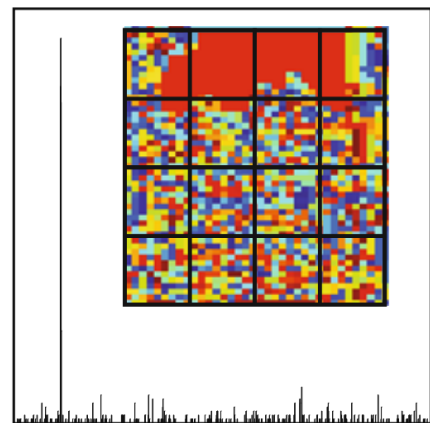
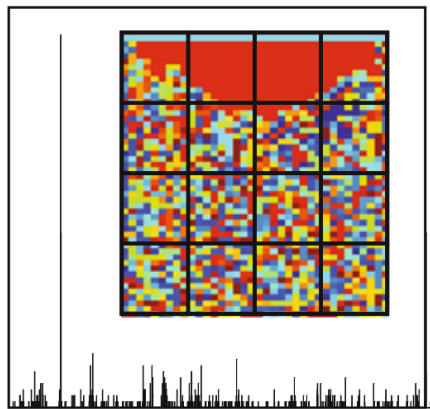


suburb

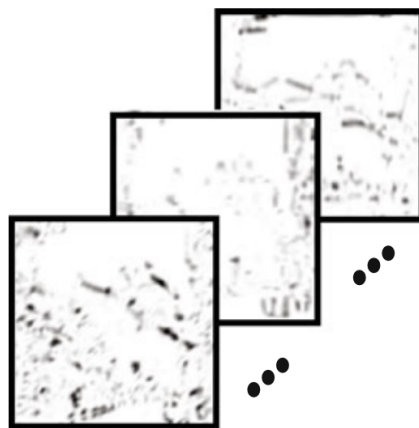
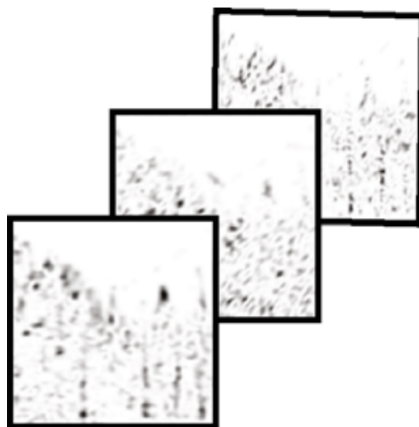
# Low-level Features?



SIFT-SPM (L=2)



GIST (filters)



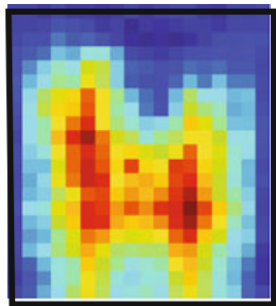
Original Image



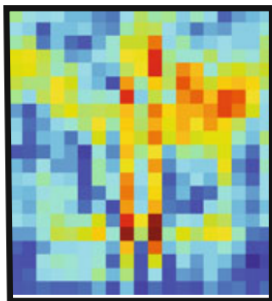


# Object Filters

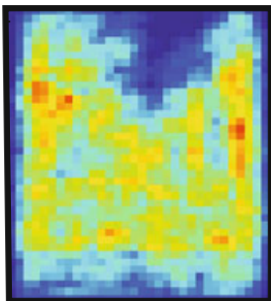
Tree



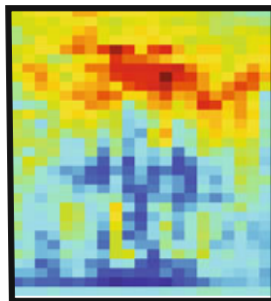
Mountain



Tower



Sky

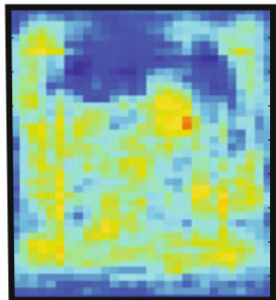


...

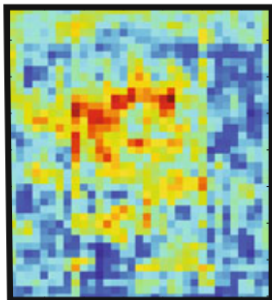
# Original Image



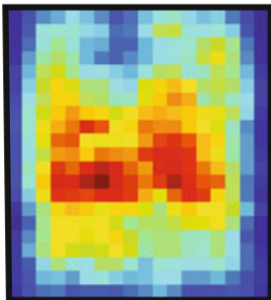
Tree



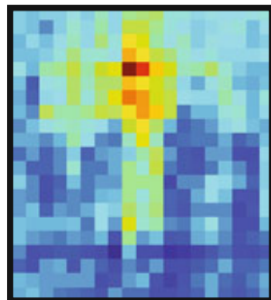
Mountain



Tower



Sky



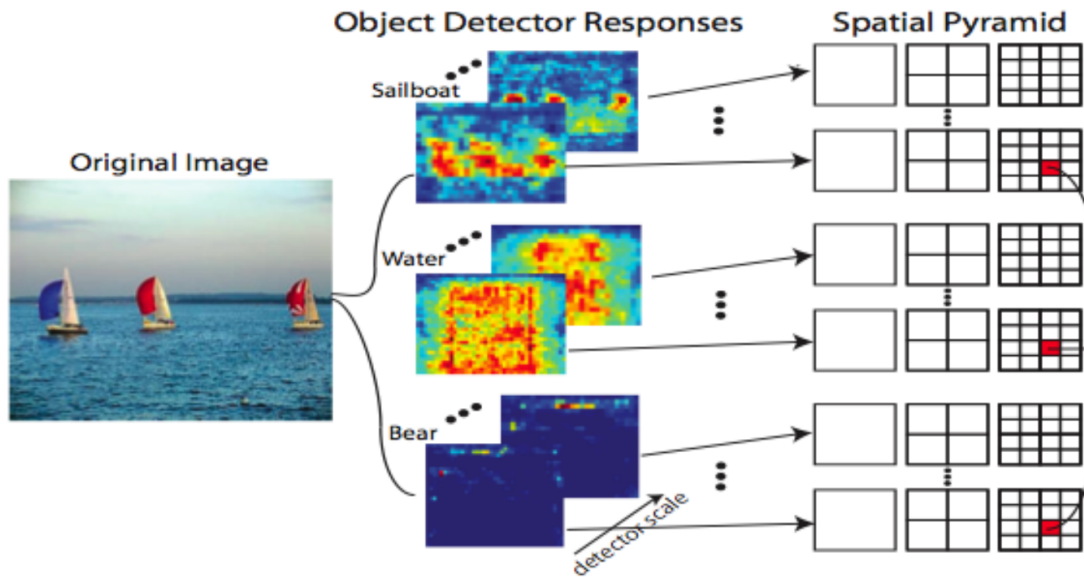
...



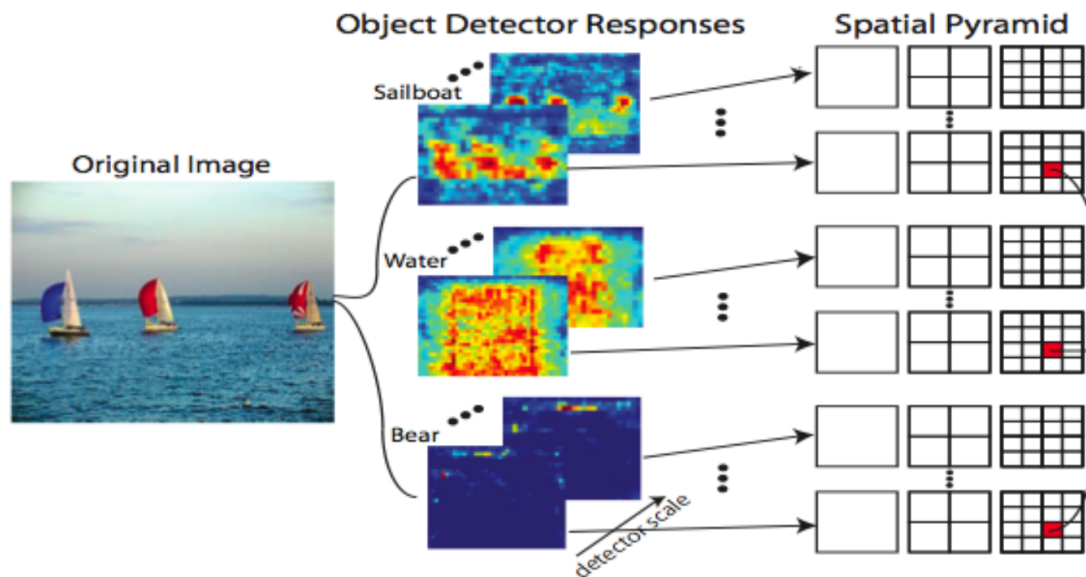
# Overview

- Bag of Words
- Kernels
- Spatial Pyramid Matching
- Scene Classification
- Object Bank

# Object Bank

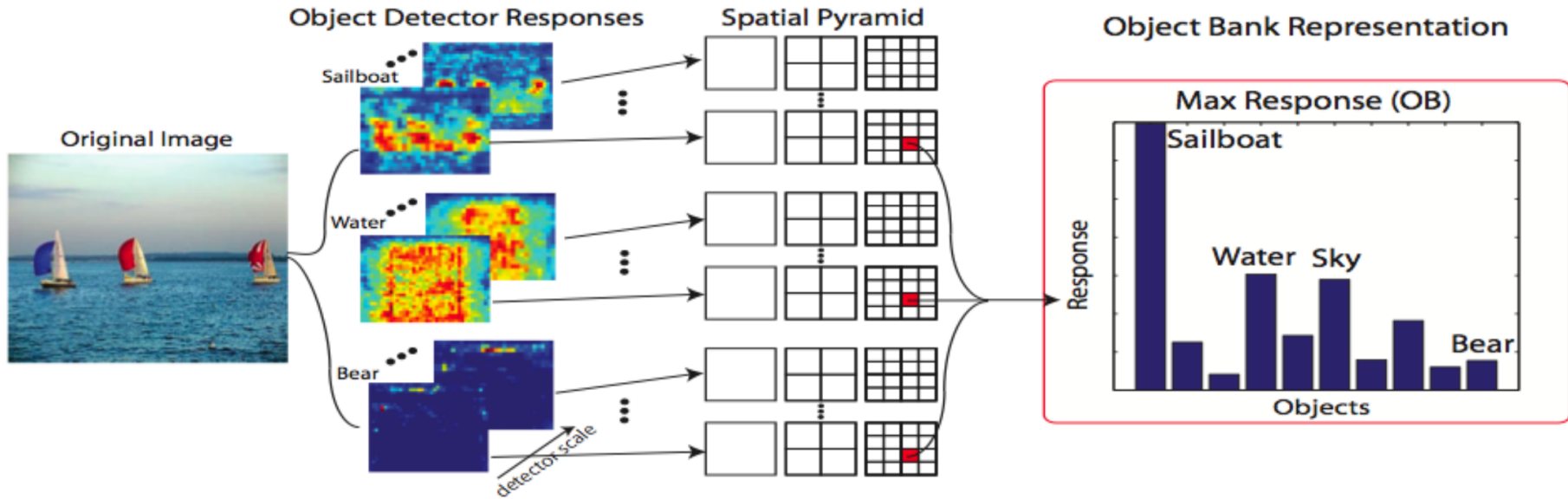


# Number of Grid Cells?



$$NumObjects \cdot NumScales \cdot (1^2 + 2^2 + 4^2)$$

# Object Bank



Questions?