Pset 4

TA Session
David Held
Feb 28, 2014

Overview

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

Overview

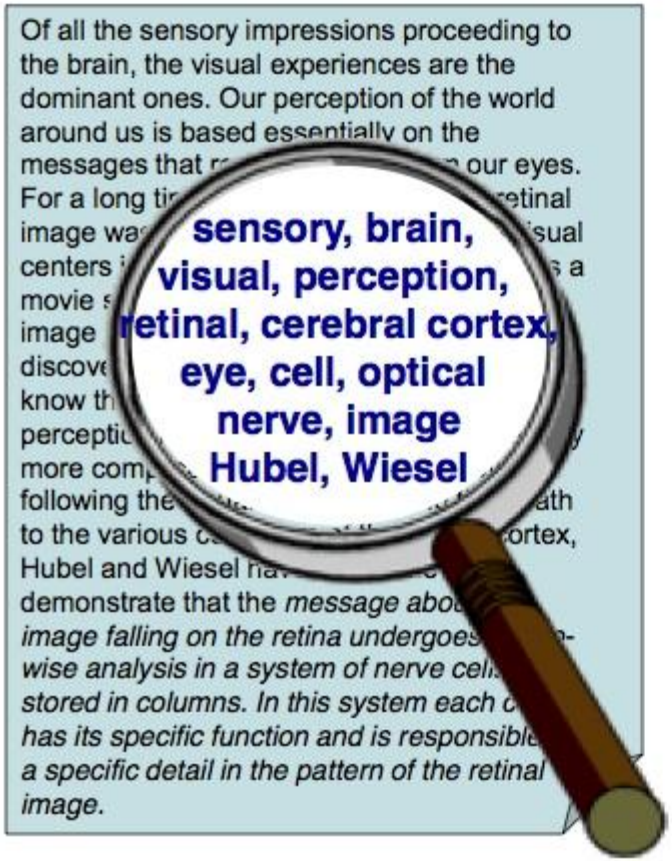
- Bag of Words
- Kernels
- Spatial Pyramid Matching
- Scene Classification
- Object Bank

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'

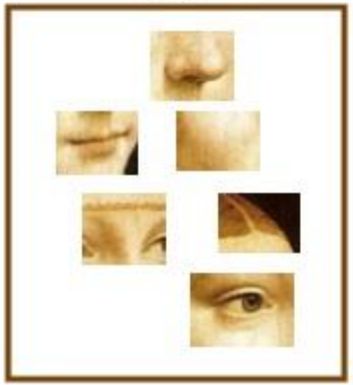
Analogy to documents



definition of “BoW”

– Independent features

face



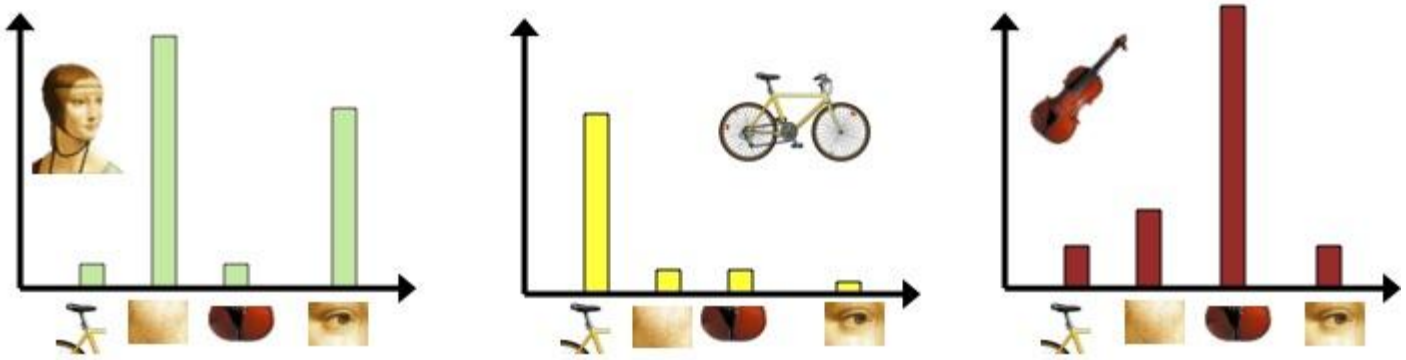
bike

violin

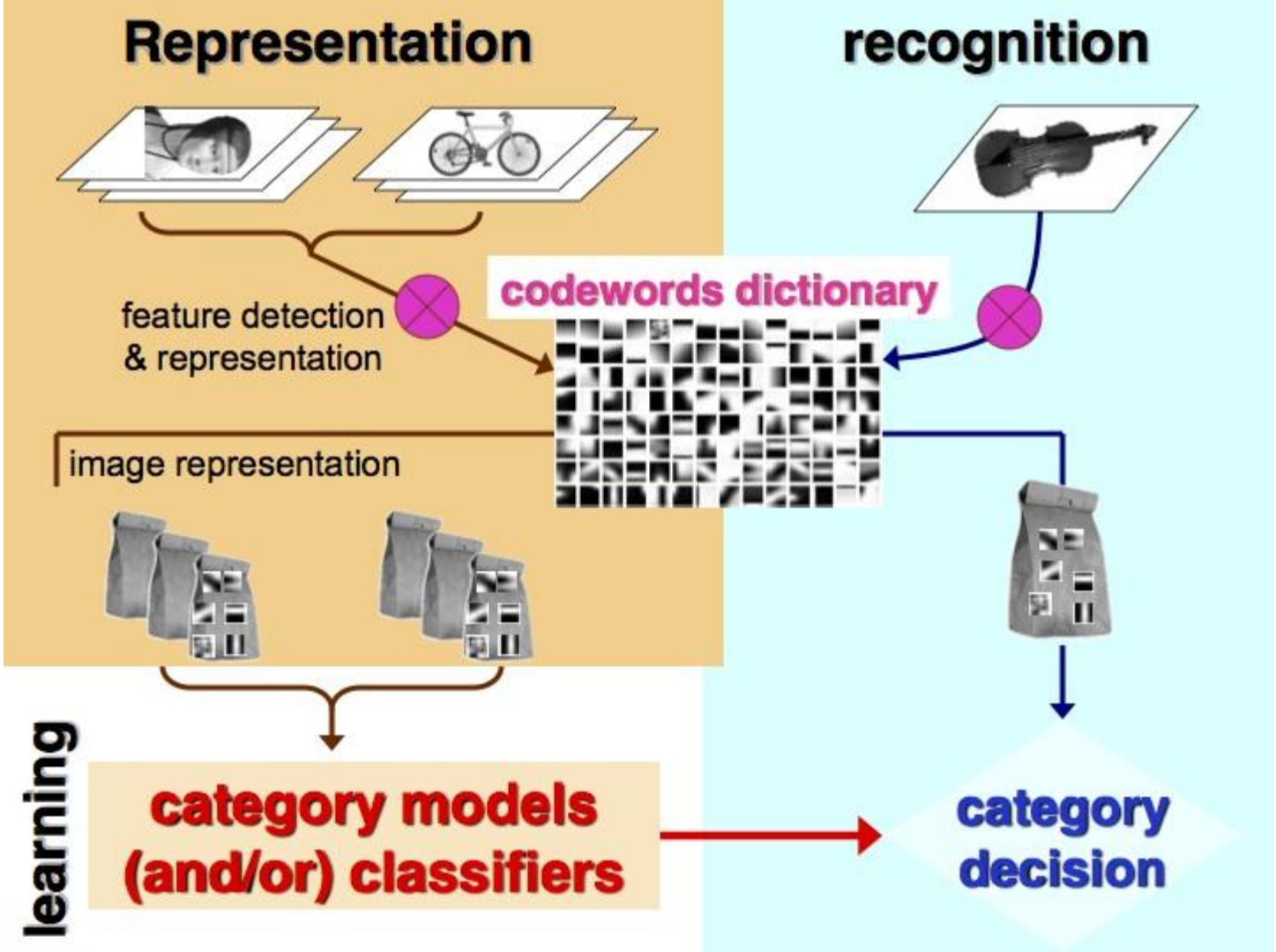
definition of “BoW”

– Independent features

– histogram representation



codewords dictionary

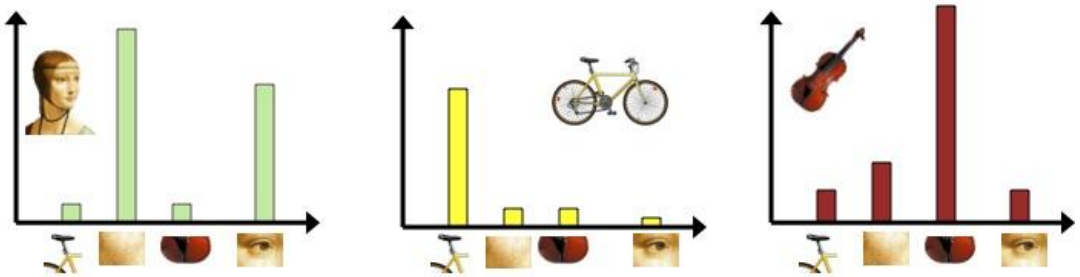

Representation
recognition
feature detection
& representation
codewords dictionary
image representation
learning
category models
(and/or) classifiers
category
decision

Histogram Representation

definition of “BoW”

– Independent features

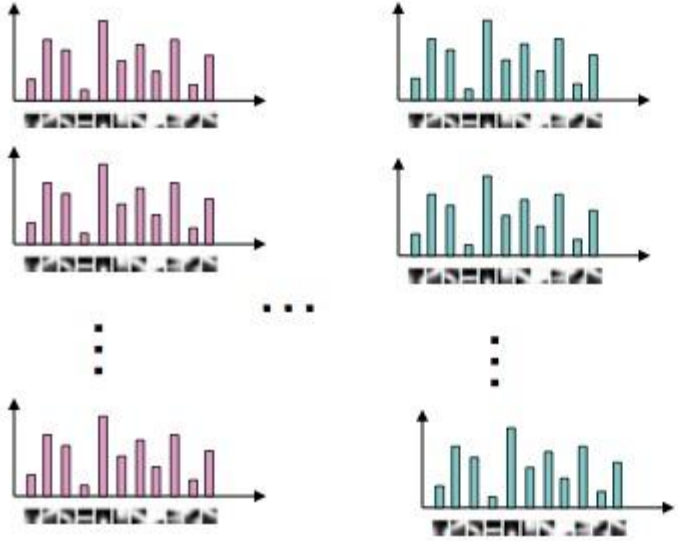
– histogram representation



codewords dictionary

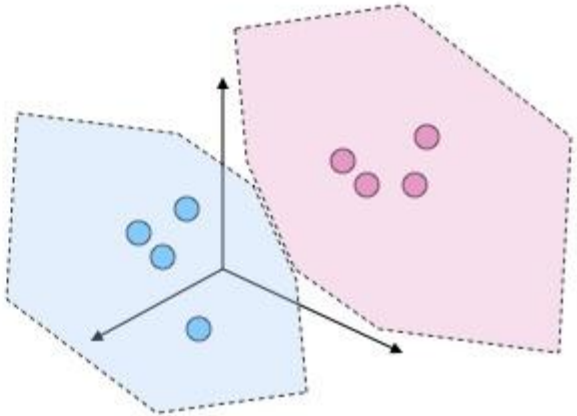
Discriminative classifiers

category models



Class 1 Class N

Model space

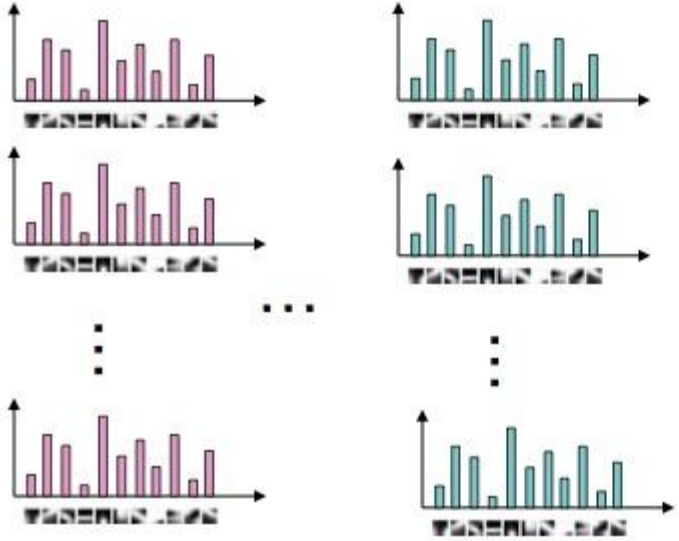


Overview

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

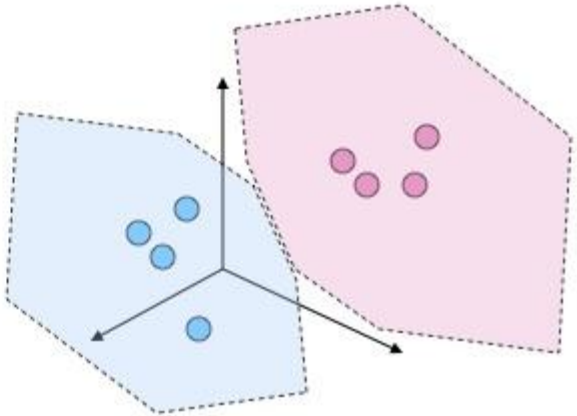
Discriminative classifiers

category models

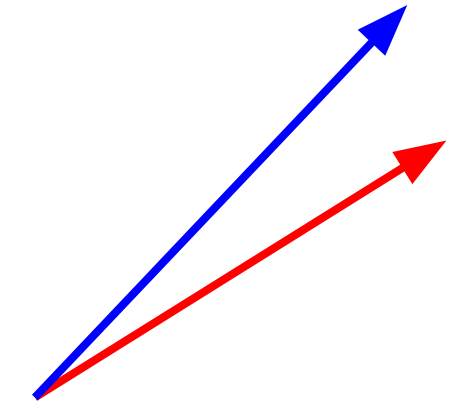


Class 1 Class N

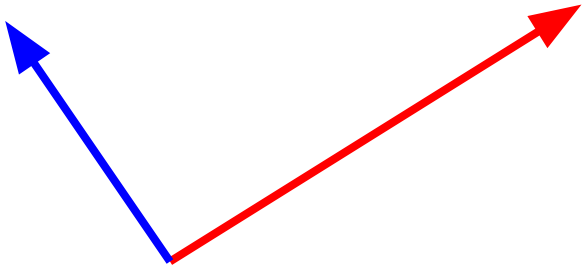
Model space



How to compare vectors?



Large



Small

Kernels

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

Linear Kernel

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

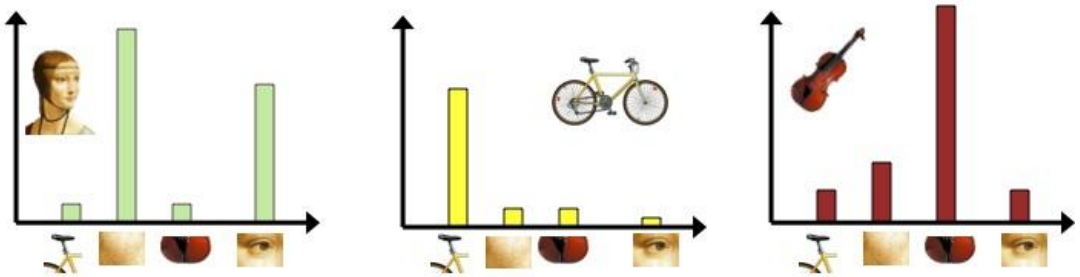
$$K(x,z) = \exp\left(-\frac{||x-z||^2}{2\sigma^2}\right)$$

Histogram Representation

definition of “BoW”

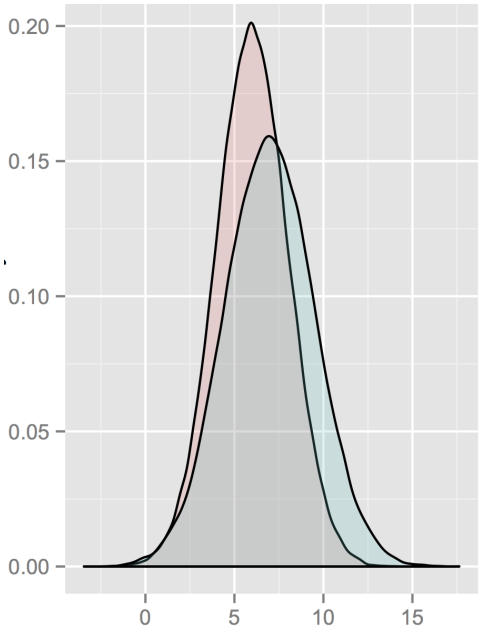
– Independent features

– histogram representation

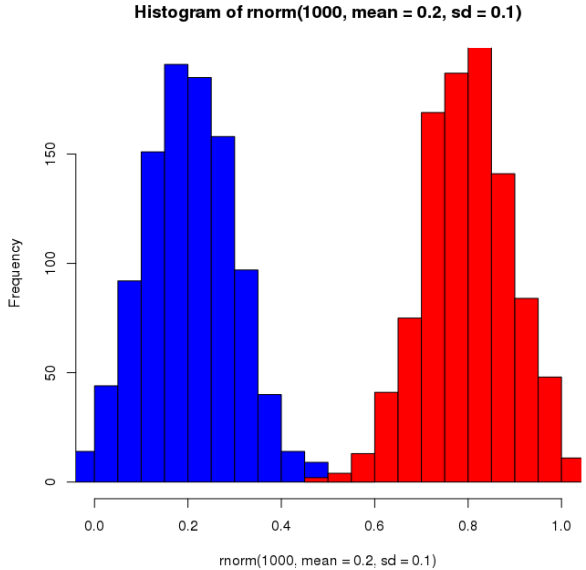


codewords dictionary

How to Compare Histograms?



Large



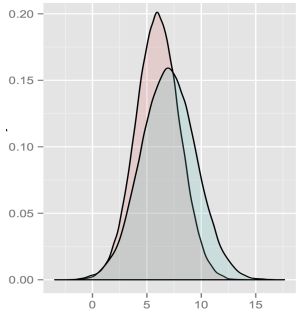
Small

How to Compare Histograms?

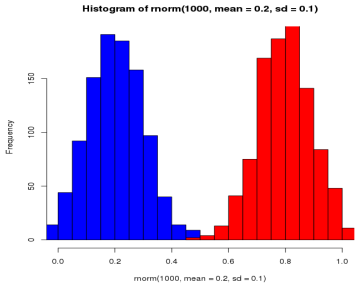
- Generalized Gaussian kernel:

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

- D can be Euclidean distance, χ^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

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

- χ^2 distance

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

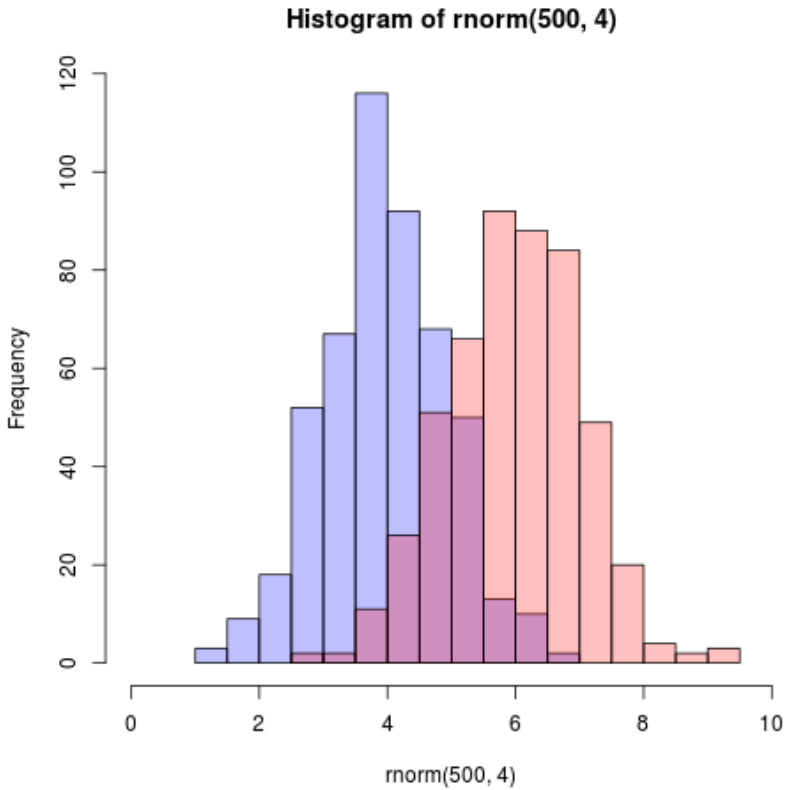
- Quadratic distance (*cross-bin*)

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

How to Compare Histograms

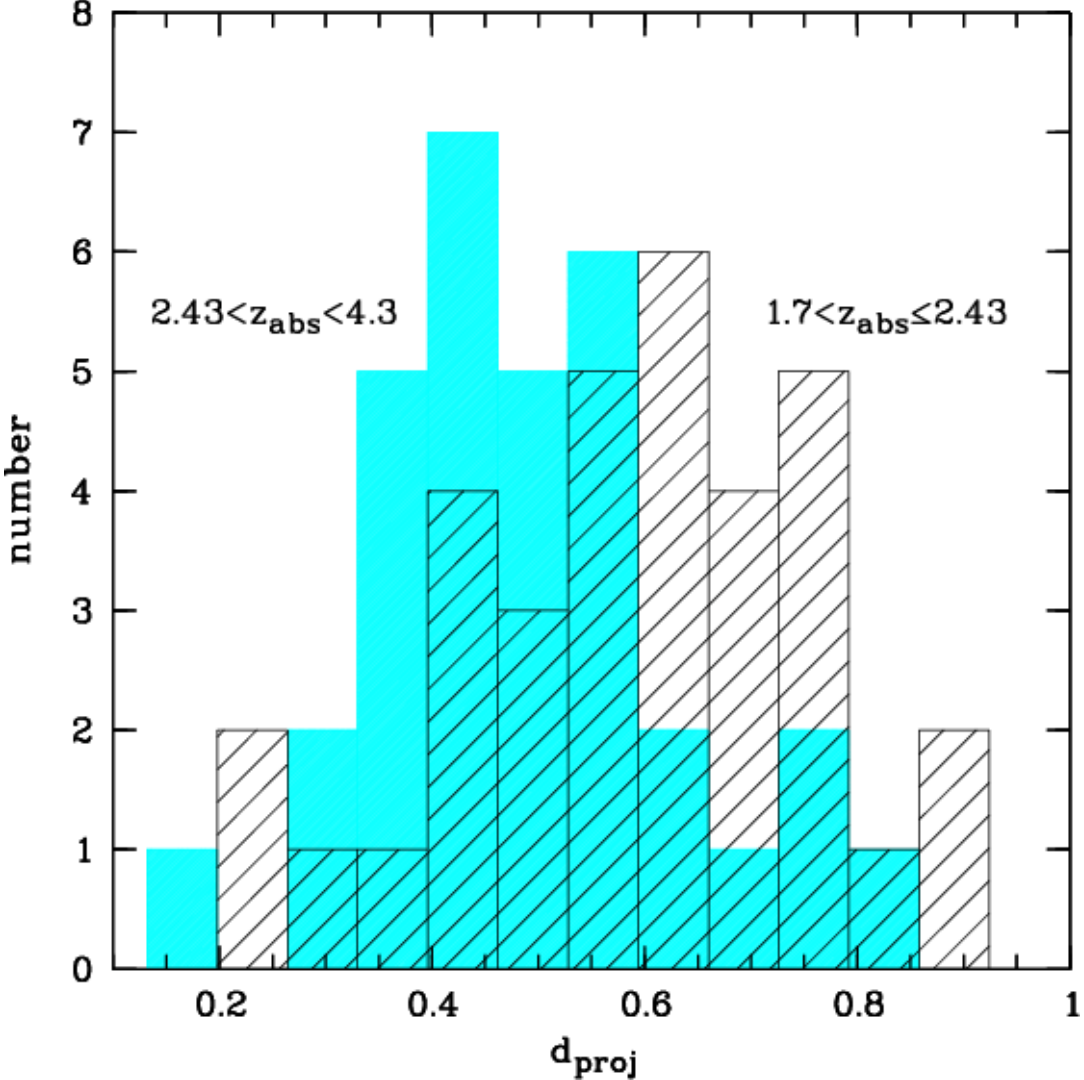
- Histogram intersection kernel:

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



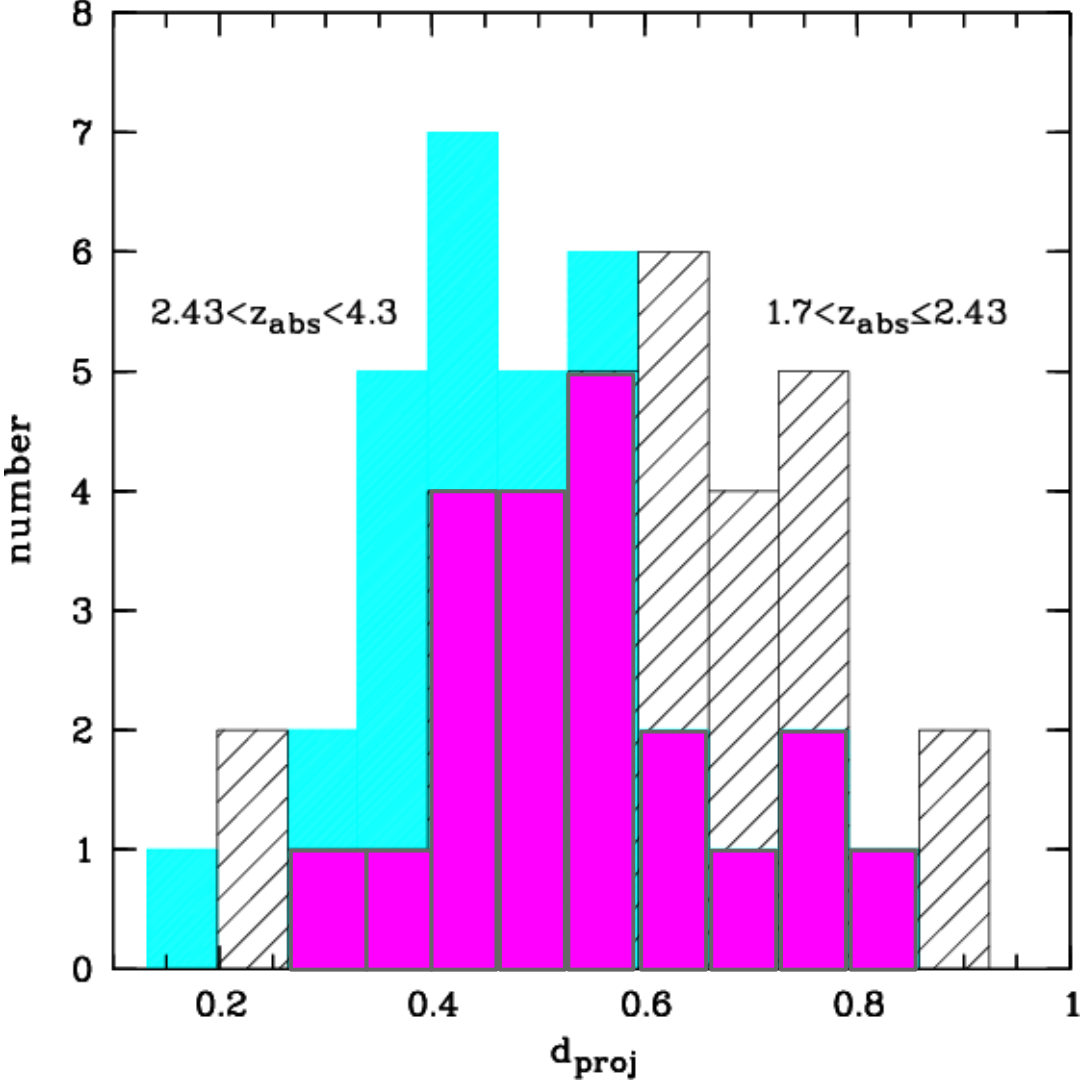
- Histogram intersection kernel:

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



- 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?

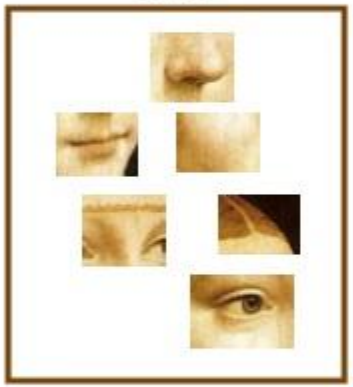
Bag of ‘words’

Is this a good representation?

definition of “BoW”

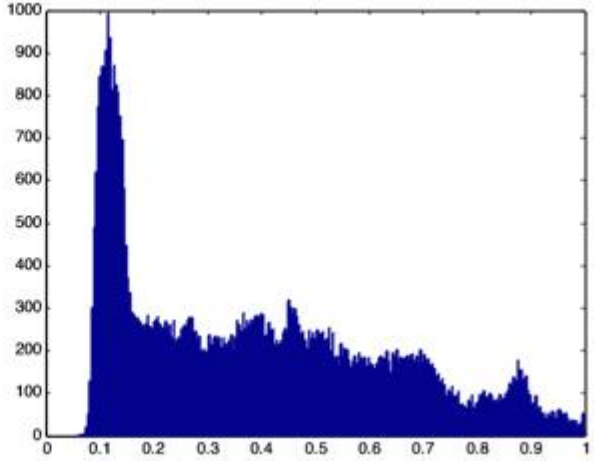
– Independent features

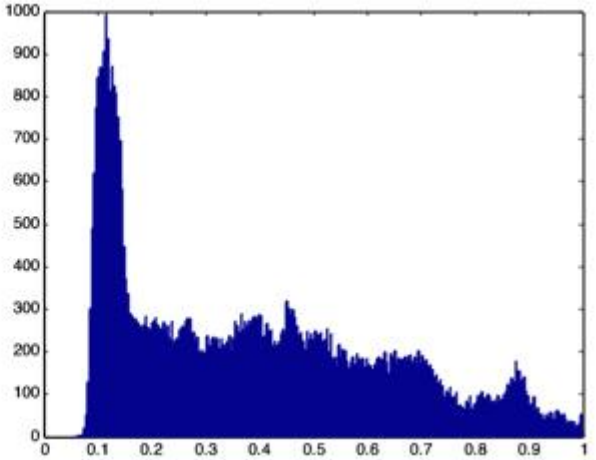
face

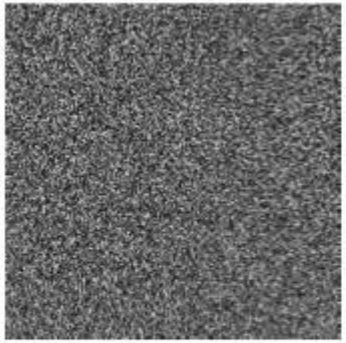


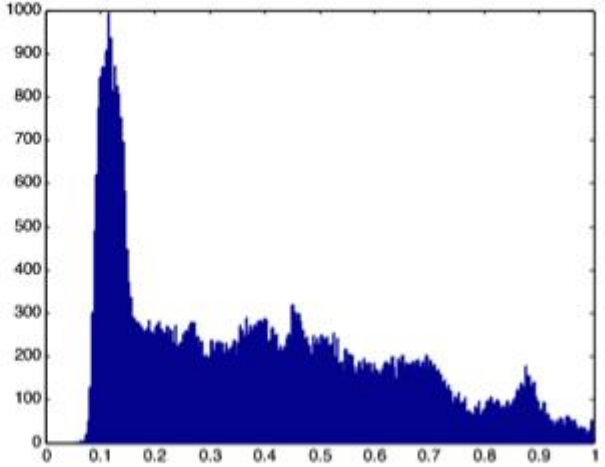
bike

violin


1000
900
800
700
600
500
400
300
200
100
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


1000
900
800
700
600
500
400
300
200
100
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1



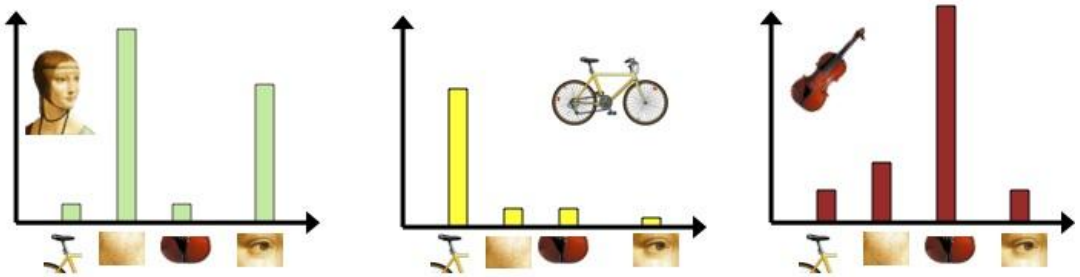

1000
900
800
700
600
500
400
300
200
100
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Still Using a Histogram Representation

definition of “BoW”

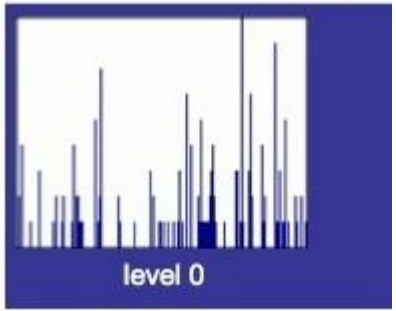
– Independent features

– histogram representation

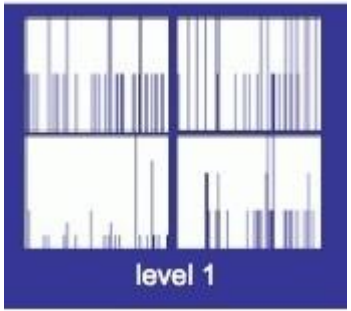


codewords dictionary

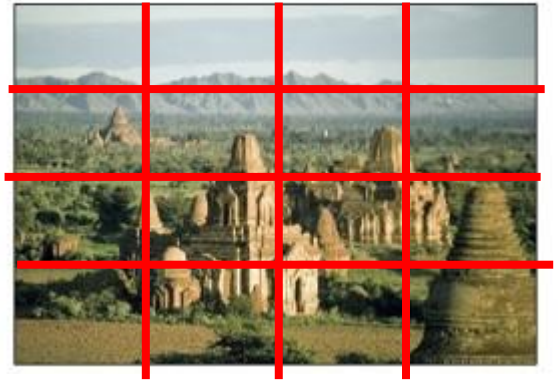
Feature extraction



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

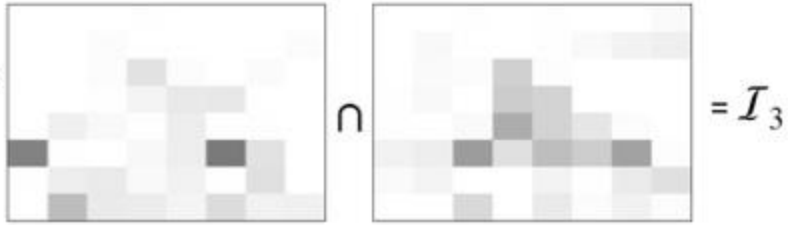
Indyk & Thaper (2003), Grauman & Darrell (2005)

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

Original images

Feature histograms:

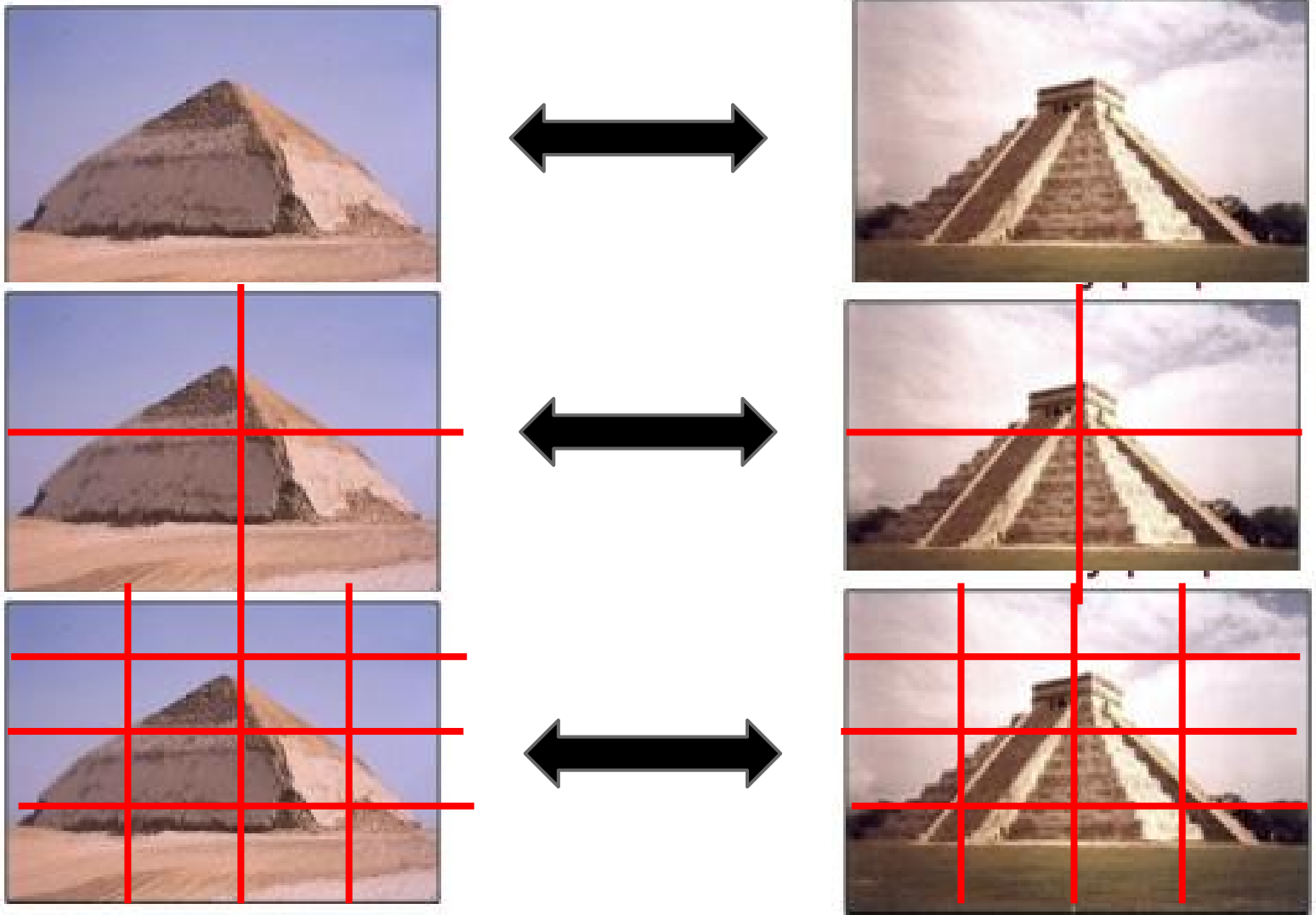
Level 3

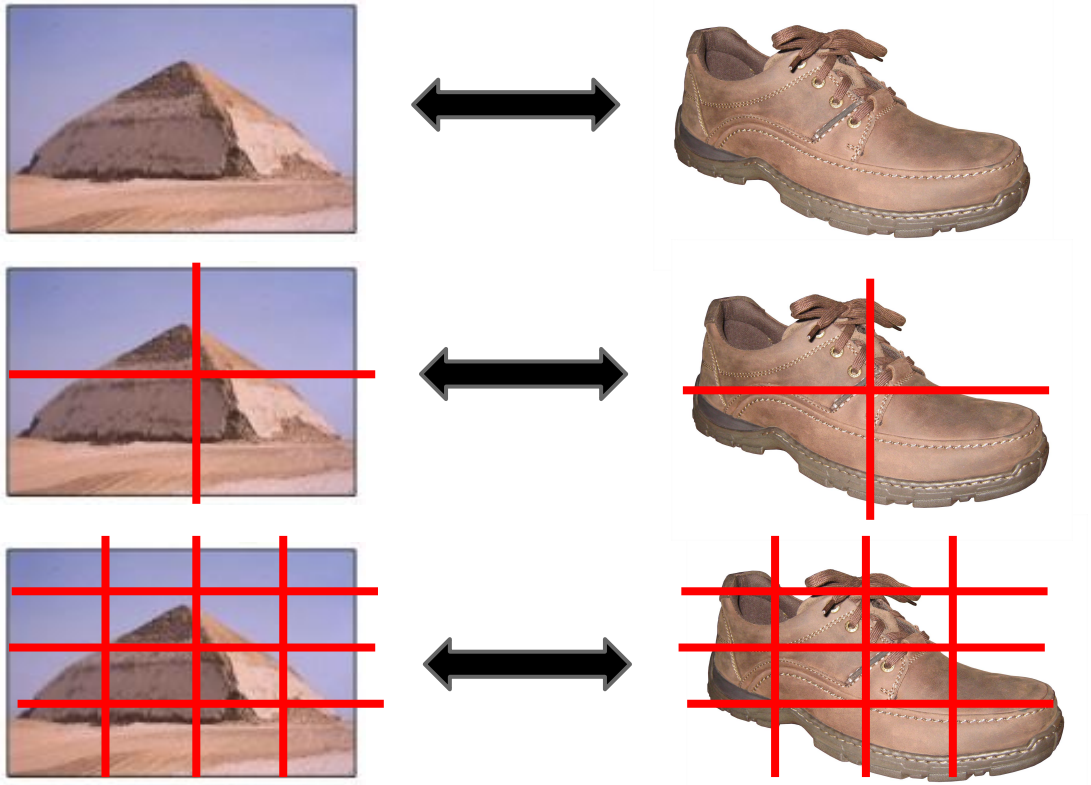


Level 2

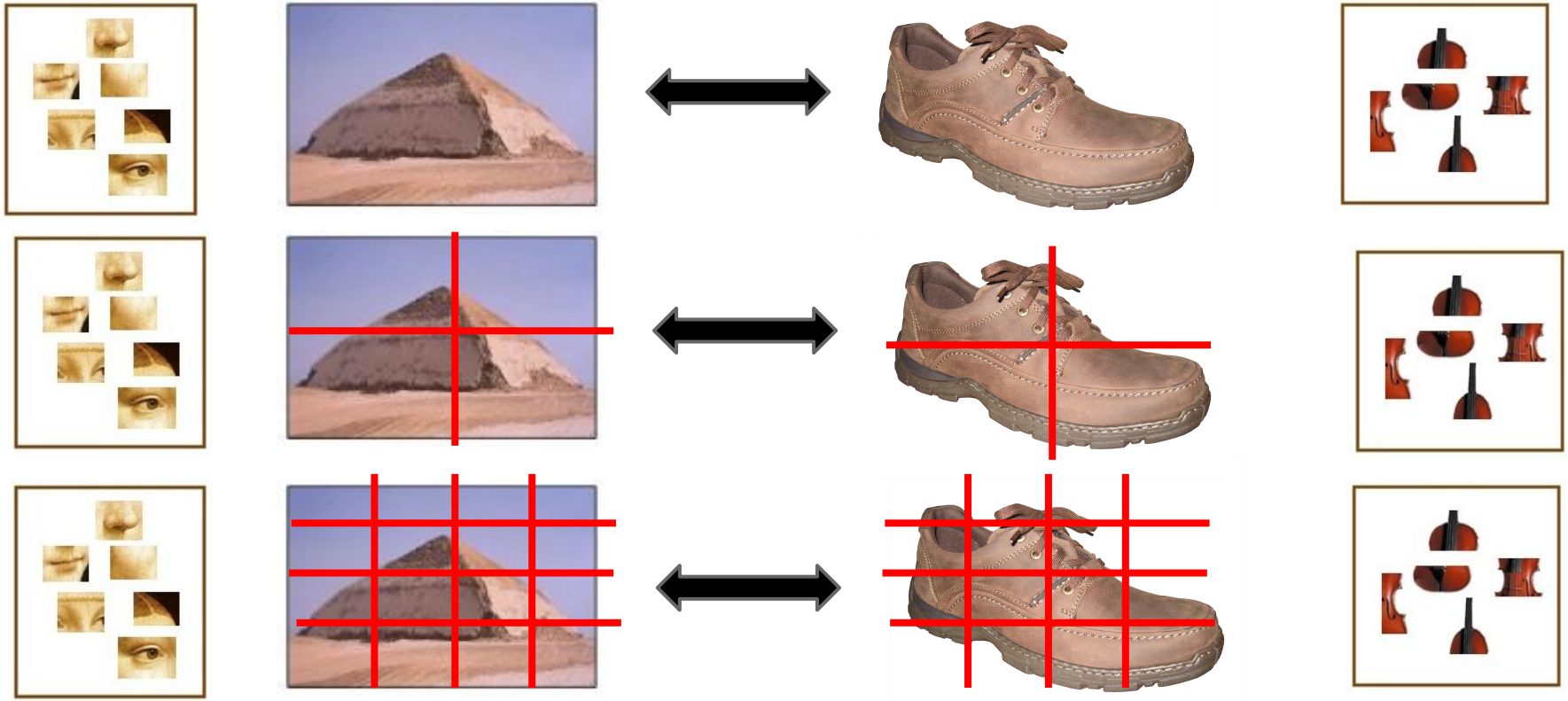
Level 1

Level 0





At which grid level will there be themost 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*): $I_3 + \frac{1}{2}(I_2 - I_3) + \frac{1}{4}(I_1 - I_2) + \frac{1}{8}(I_0 - I_1)$

$$\kappa^L(X,Y) = \mathcal{I}^L + \sum_{\ell=0}^{L-1} \frac{1}{2^{L-\ell}} \left(\mathcal{I}^\ell - \mathcal{I}^{\ell+1}\right)$$

$$= \frac{1}{2^L}\mathcal{I}^0 + \sum_{\ell=1}^{L} \frac{1}{2^{L-\ell+1}}\mathcal{I}^\ell .$$

- 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

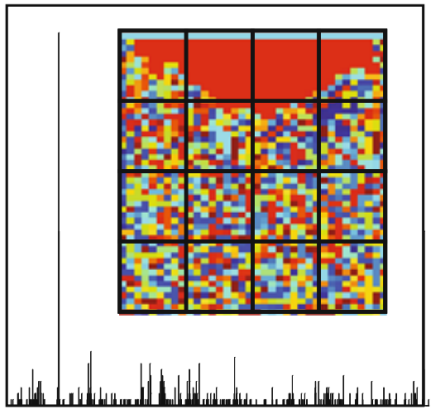
Scene Classification

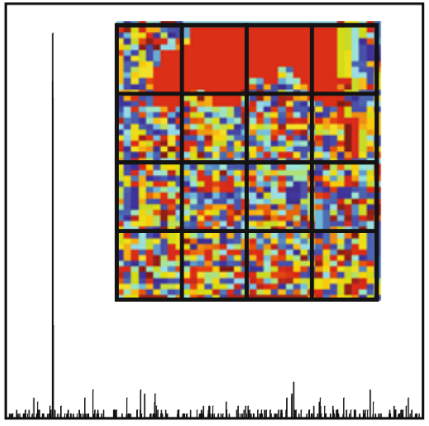
Low-level Features?

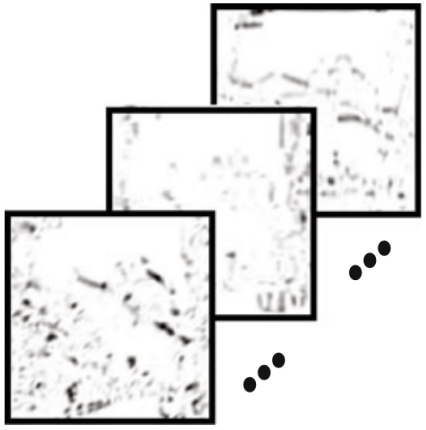
SIFT-SPM (L=2)

GIST (filters)

Original Image



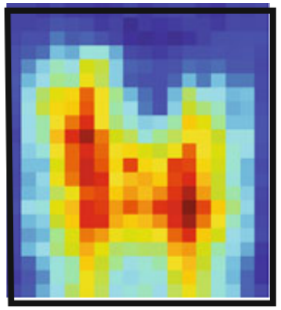




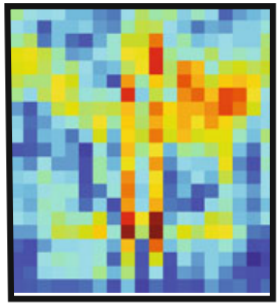
Object Filters

Original Image

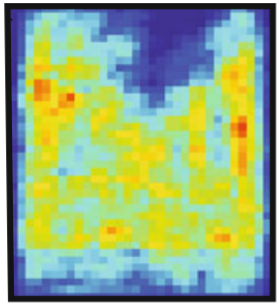
Tree



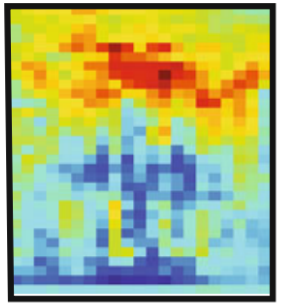
Mountain



Tower

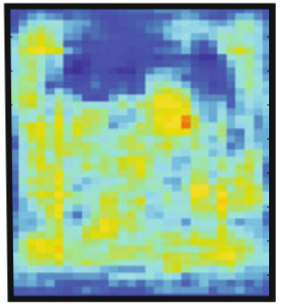


Sky

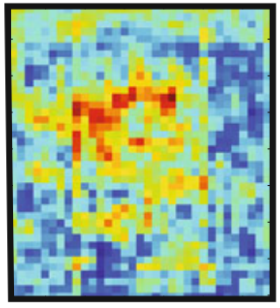


...

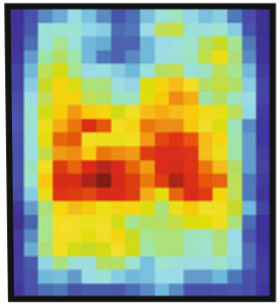
Tree



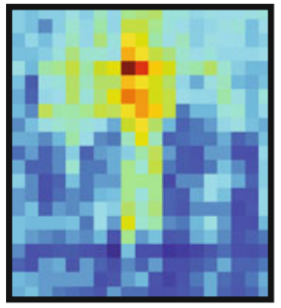
Mountain



Tower



Sky

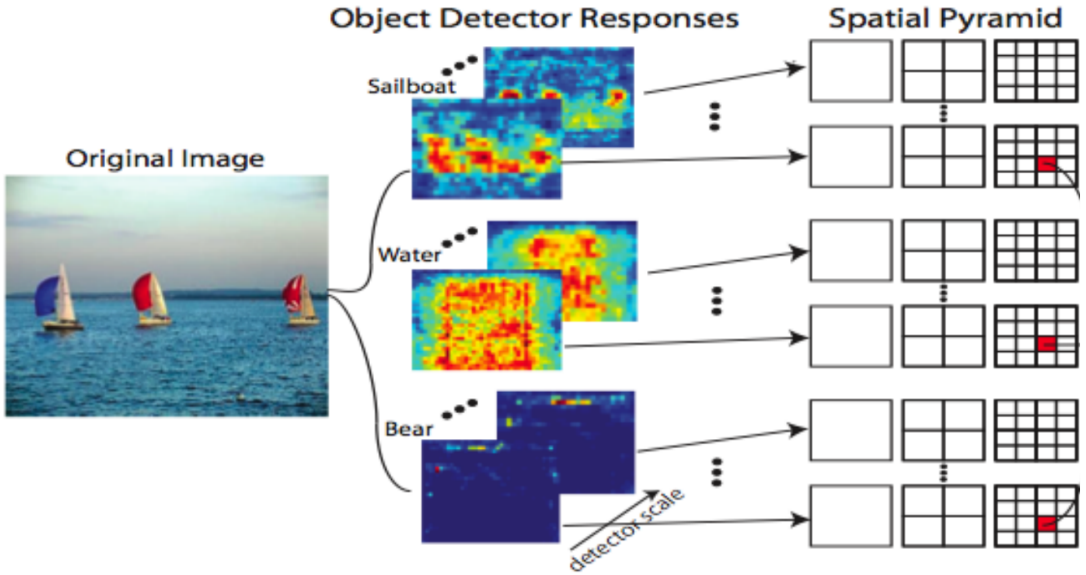


...

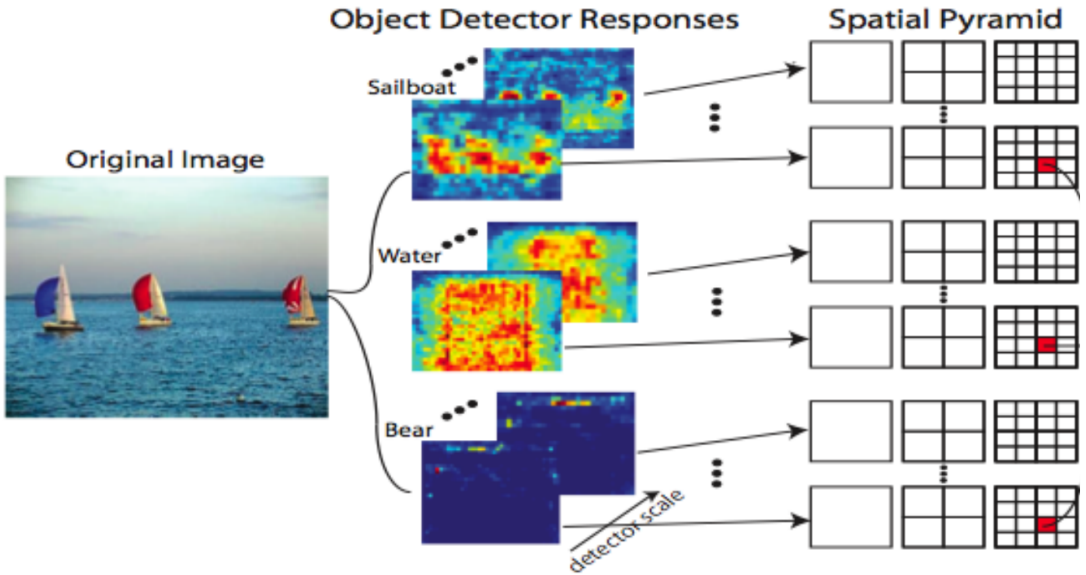
Overview

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

Object Bank

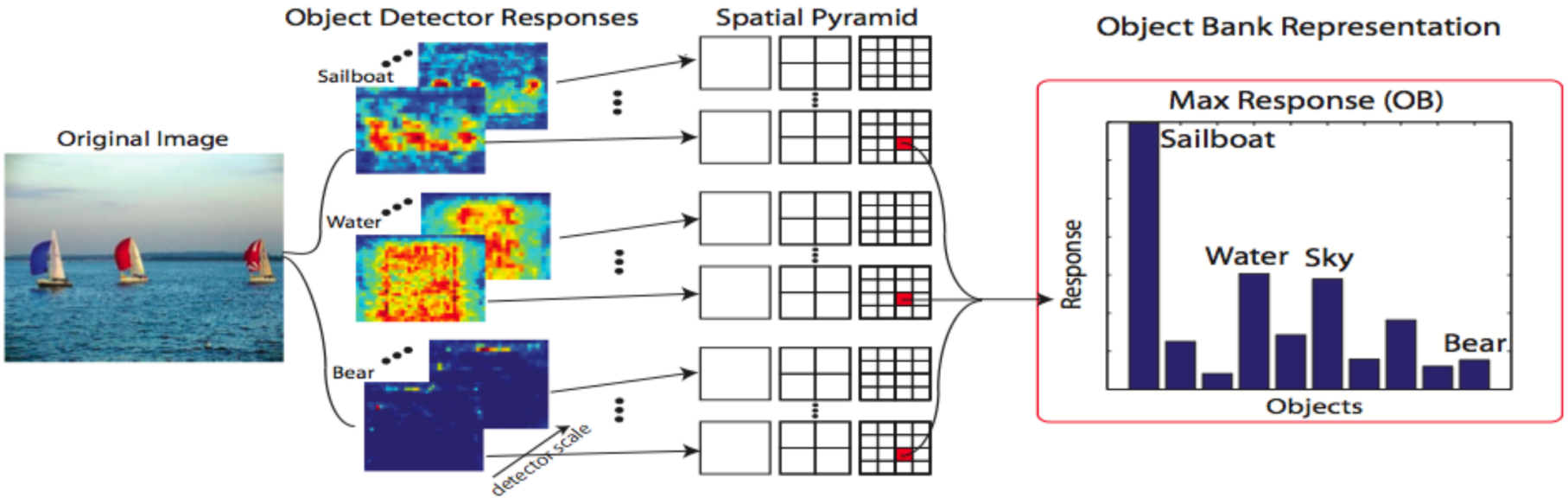

Object Detector Responses
Spatial Pyramid
Original Image
Sailboat
Water
Bear
detector scale

Number of Grid Cells?



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

Object Bank



Questions?