TA Section: Problem Set 4

Professor Fei-Fei Li

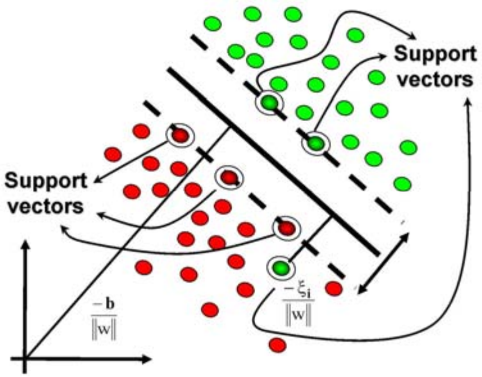
Stanford Vision Lab

outline

- Discriminative vs. Generative Classifiers
- Image representation and recognition models
  - Bag of Words Model
  - Part-based Model
    - Constellation Model
    - Pictorial Structures Model
  - Spatial Pyramid Matching (SPM)
  - ObjectBank

Discriminative vs Generative Classifiers

Training classifiers involves estimating f: X → Y, or P(Y|X)

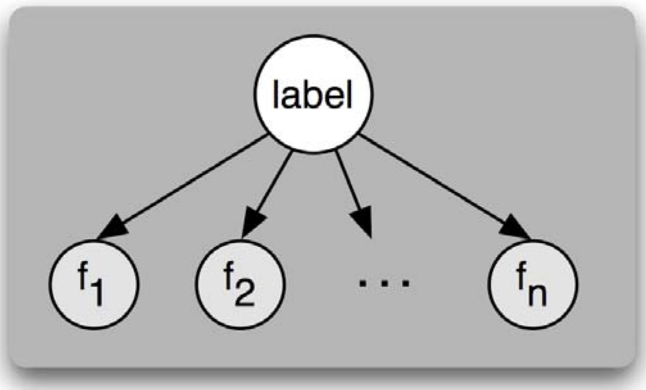


•Discriminative classifiers (e.g. logistic regression, SVM):

- We want to model P(Y|X)
- Assume some functional form for P(Y|X)
- Estimate parameters of P(Y|X) directly from training data

•Generative classifiers (e.g. naïve bayes):

- We want to model P(X, Y)
- Assume some functional form for P(X|Y), P(X)
- Estimate parameters of P(X|Y), P(X) directly from training data
- Use Bayes rule to calculate P(Y|X=x_i)



Discriminative vs Generative Classifiers

- Advantages of discriminative classifiers:
  - Typically faster at making predictions
  - Tend to have better performance
  - Direct modeling of what we want to optimize

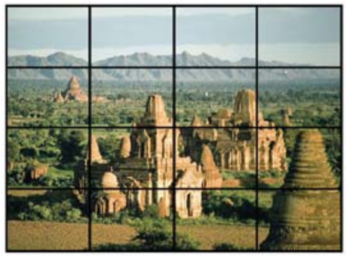
- Advantage of generative classifiers:
  - Can handle missing/partially labeled data
  - A new class (Y+1) can be added incrementally without training the complete model
  - Can generate samples from the training distribution

[Ulusoy & Bishop, 2005]

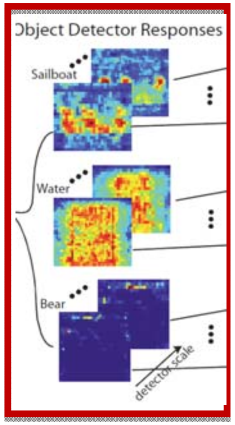
Image Representation

Bag of Words

Weakly Spatial Models

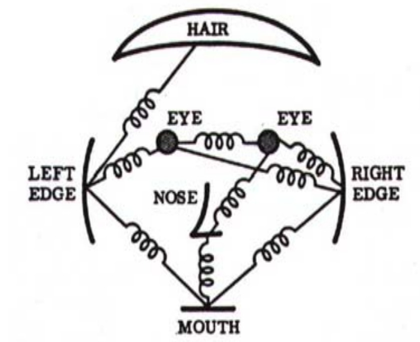


- Spatial Pyramid Matching



- Object-Bank

Part-based



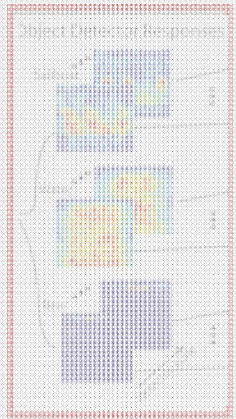
- Constellation Model
- Pictorial Structure

No spatial info. **Spatial Specificity of Parts** Very strong but sparse

Image Representation

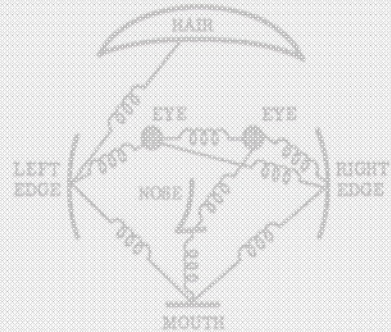
Bag of Words

Weakly Spatial Models



- Spatial Pyramid Matching
- Object-Bank

Part-based



- Constellation Model
- Pictorial Structure

No spatial info. **Spatial Specificity of Parts** Very strong but sparse

Bag-of-words Representation

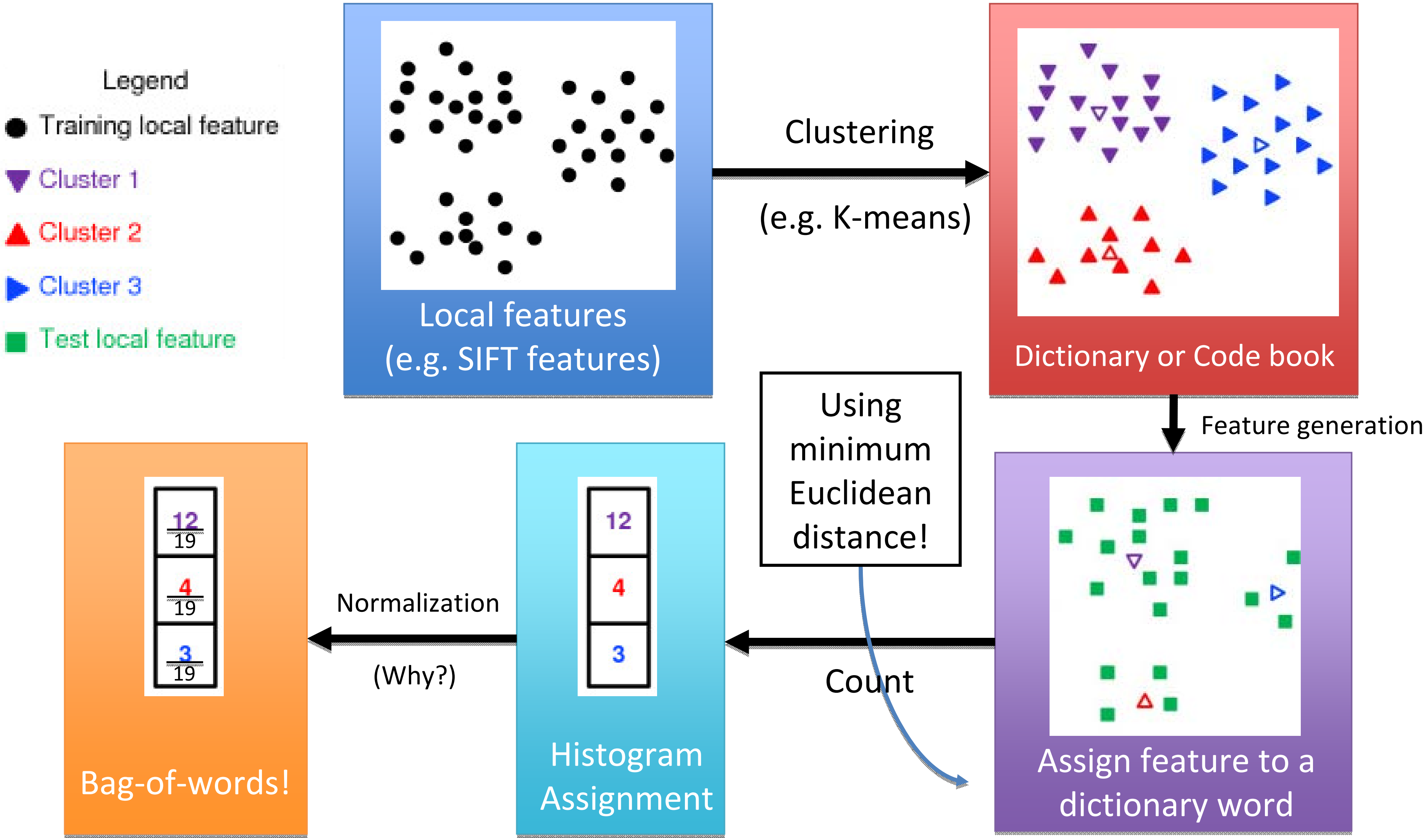
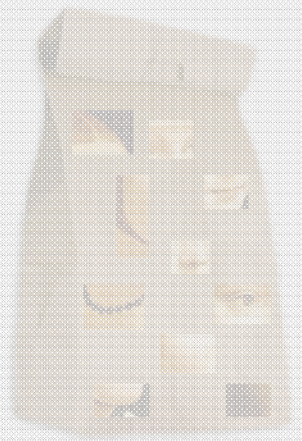
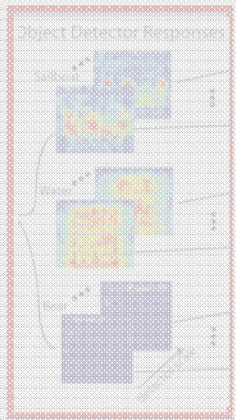


Image Representation

Bag of Words

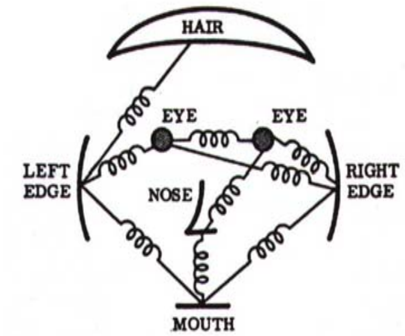


Weakly Spatial Models



- Spatial Pyramid Matching
- Object-Bank

Part-based

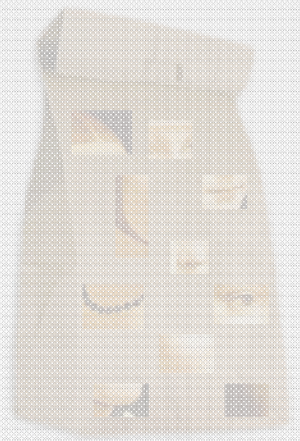


- Constellation Model
- Pictorial Structure

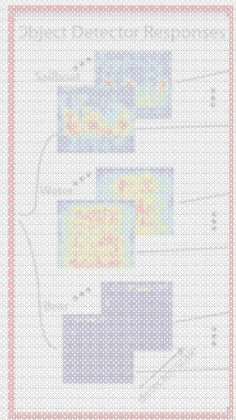
No spatial info. **Spatial Specificity of Parts** Very strong but sparse

Image Representation

Bag of Words



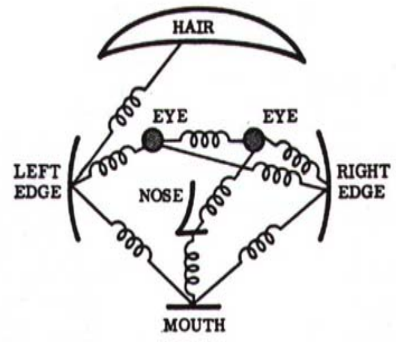
Weakly Spatial Models



•Spatial Pyramid Matching

•Object-Bank

Part-based



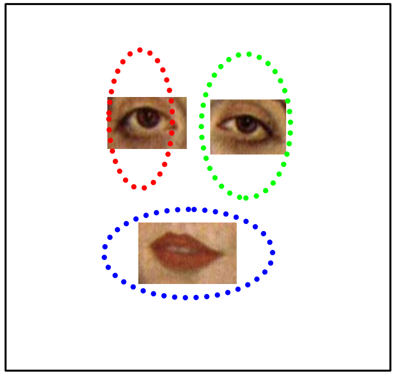
•Constellation Model
•Pictorial Structure

No spatial info. **Spatial Specificity of Parts** Very strong but sparse

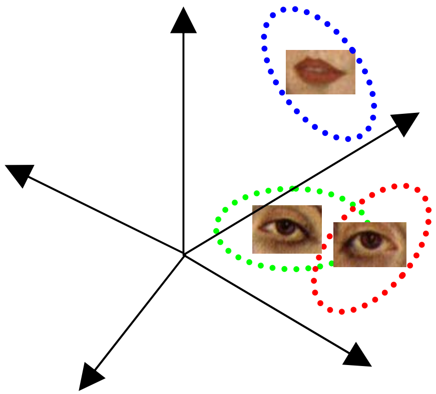
Generative probabilistic model (2)

Foreground model

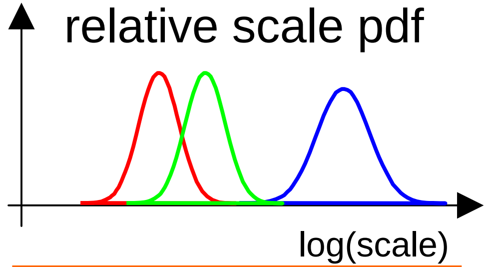
Gaussian shape pdf



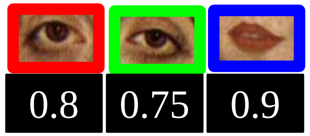
Gaussian part appearance pdf



Gaussian relative scale pdf

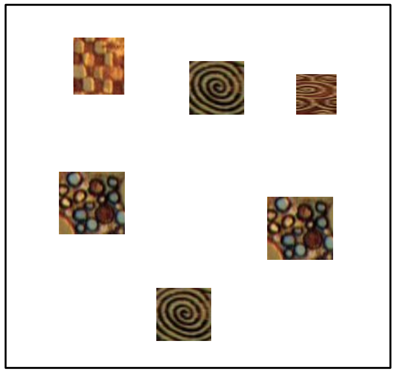


Prob. of detection

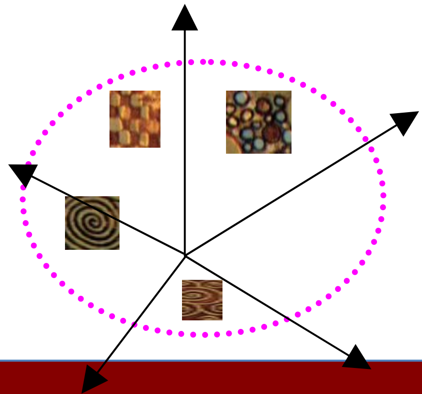


Clutter model

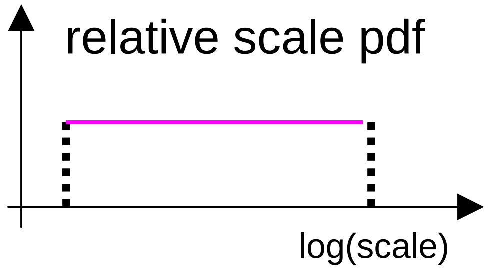
Uniform shape pdf



Gaussian background appearance pdf



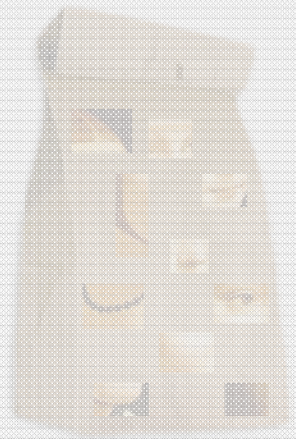
Uniform relative scale pdf



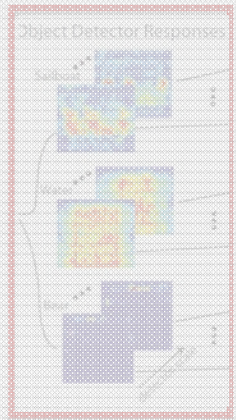
Poission pdf on # detections

Image Representation

Bag of Words

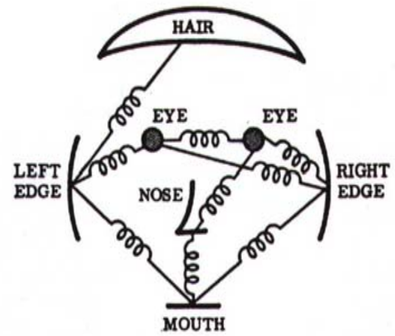


Weakly Spatial Models



- Spatial Pyramid Matching
- Object-Bank

Part-based



- Constellation Model
- Pictorial Structure

No spatial info. **Spatial Specificity of Parts** Very strong but sparse

Pictorial Structures

- Basic idea:
  We would like to represent an object by
  - a collection of parts
  - arranged in a deformable configuration

Examples of detected models

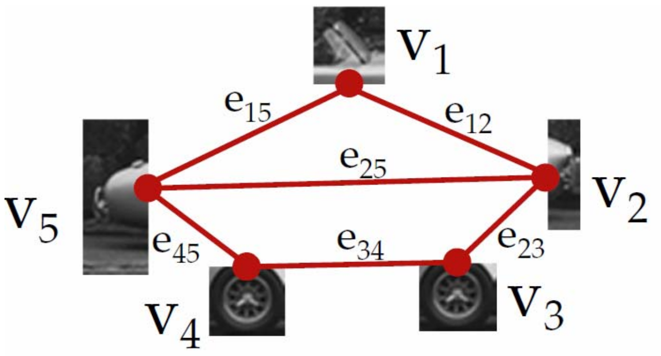
Parts sharing similar appearance

Pictorial Structures

- Local model of appearance with non-local geometric or spatial constraints
- Simultaneous use of appearance and spatial information
  - Simple part models alone are not discriminative
- The model needs to solve the tasks:
  - determine whether an object is visible in an image
  - determine where an object is in the image

Pictorial Structures

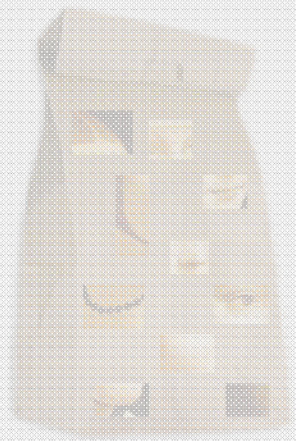
- Model is represented as an undirected graph structure G= (V, E), where V are the vertices and E are the edges



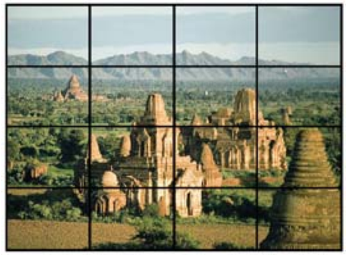
$$\underbrace{L^*}_{\text{Optimal parts configuration}} = \arg\min_L \left(\underbrace{\sum_{i=1}^{n}}_{\text{Sum over all parts}} \overbrace{m_i(l_i)}^{\text{Matching score of individual parts}} + \overbrace{\sum_{(v_i,v_j)\in E}}^{\text{Sum over all edges}} \underbrace{d_{ij}(l_i, l_j)}_{\text{Deformation score of connected parts}} \right)$$

Image Representation

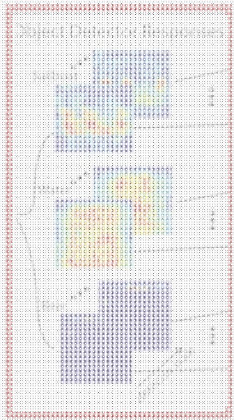
Bag of Words



Weakly Spatial Models

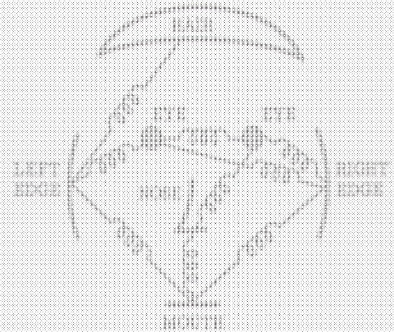


- Spatial Pyramid Matching



- Object-Bank

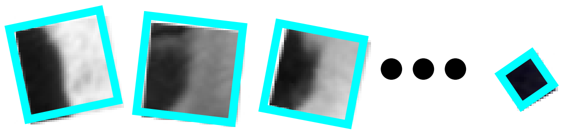
Part-based



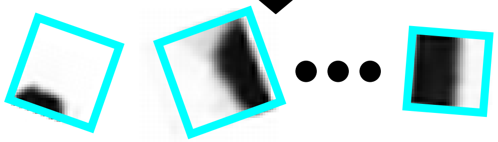
- Constellation Model
- Pictorial Structure

No spatial info. **Spatial Specificity of Parts** Very strong but sparse

Start with Pyramid Matching Kernel for BoW Models



Sets of features



$$\mathbf{X} = \{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_m\}$$

$$\mathbf{Y} = \{\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_n\}$$

[Grauman & Darrell, 2005]

Pyramid Matching Kernel

- How do we build a **discriminative classifier** using the set representation?

- Kernel-based methods (e.g. SVM) are appealing for efficiency and generalization power.

- But what is an appropriate kernel?
  - Each instance is an unordered set of vectors
  - Varying number of vectors per instance

Pyramid Matching Kernel

- We can compare sets by computing a **partial matching** between their features

Number of newly matched pairs at level *i*

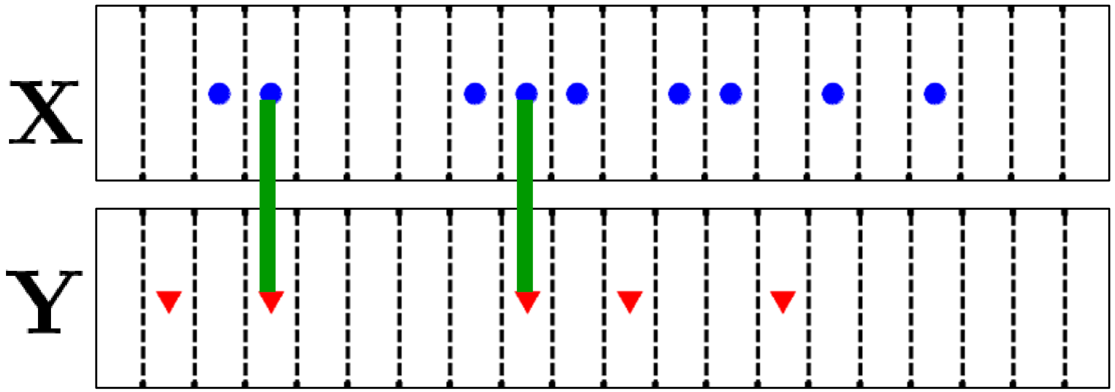
Approximate partial match similarity

$$K_{\Delta} = \sum_{i=0}^{L} w_i N_i$$

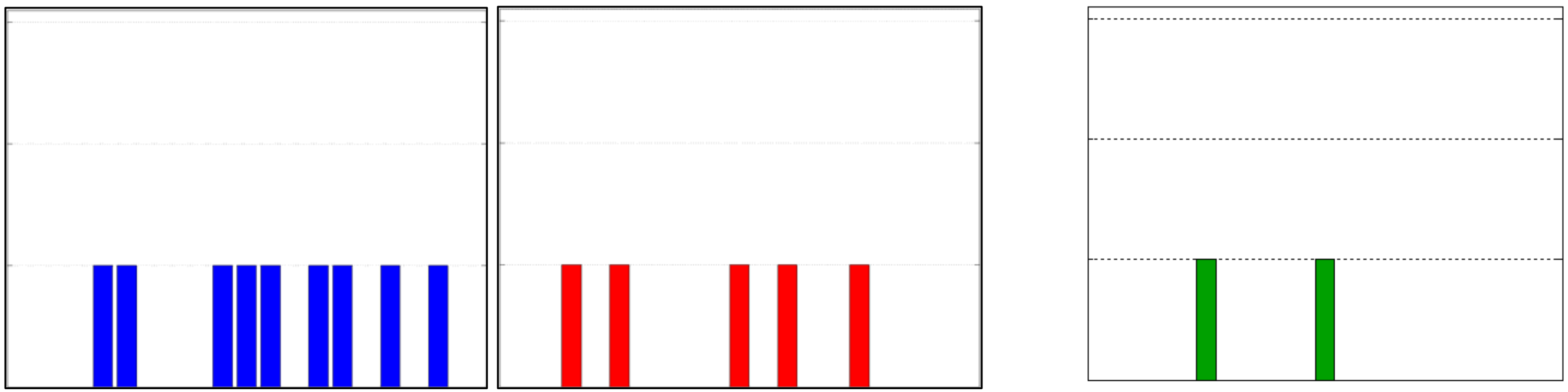
Measure of difficulty of a match at level *i*

Pyramid Matching Kernel (Example)

Level 0

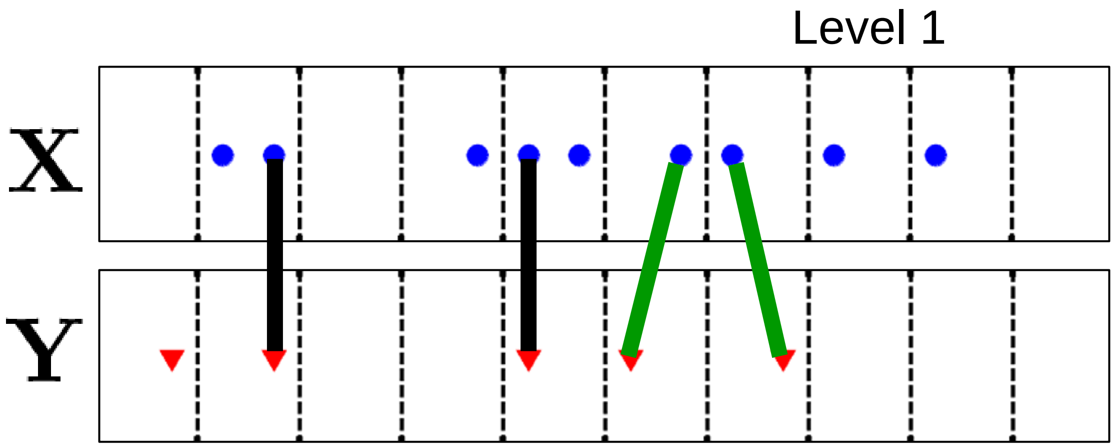


$$\rightarrow \quad N_0 = 2$$
$$w_0 = 1$$

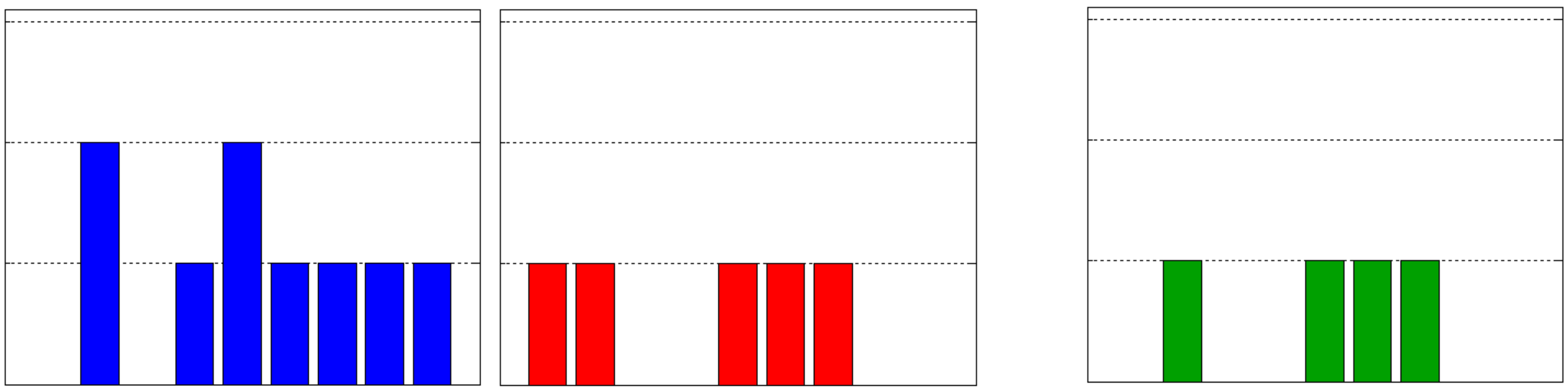


$H_0(\mathbf{X})$ $H_0(\mathbf{Y})$ $\mathcal{I}_0 = 2$

Pyramid Matching Kernel (Example)

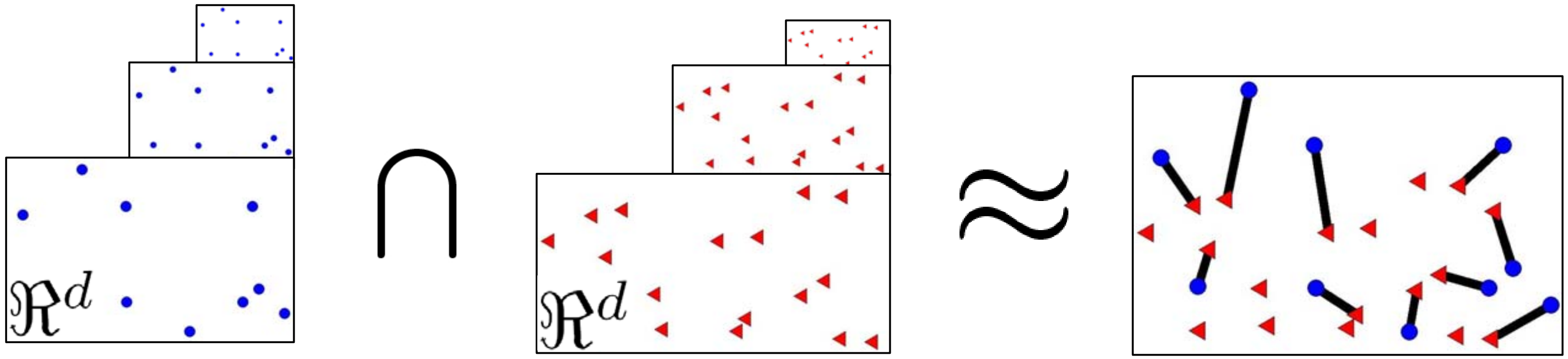


$$N_1 = 4 - 2 = 2$$
$$w_1 = \frac{1}{2}$$



$H_1(\mathbf{X})$ $H_1(\mathbf{Y})$ $\mathcal{I}_1 = 4$

Pyramid Matching Kernel



optimal partial matching between sets of features

$$K_{\Delta}\left(\Psi(\mathbf{X}), \Psi(\mathbf{Y})\right) = \sum_{i=0}^{L} \underbrace{\frac{1}{2^i}}_{\text{difficulty of a match at level i}} \left(\underbrace{\mathcal{I}\left(H_i(\mathbf{X}), H_i(\mathbf{Y})\right) - \mathcal{I}\left(H_{i-1}(\mathbf{X}), H_{i-1}(\mathbf{Y})\right)}_{\text{number of new matches at level i}} \right)$$

Spatial Pyramid Matching

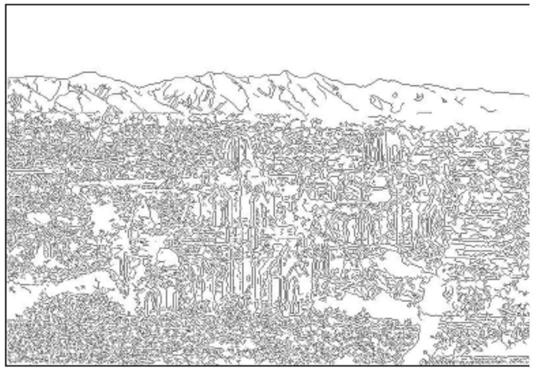
- Pyramid Match Kernel (Grauman & Darrell)

  *Pyramid in **feature** space, ignore location*

- Spatial Pyramid (Lazebnik et al)

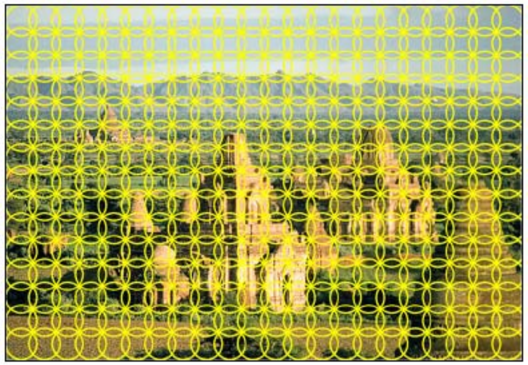
  *Pyramid in **image** space, quantize features*

Features:



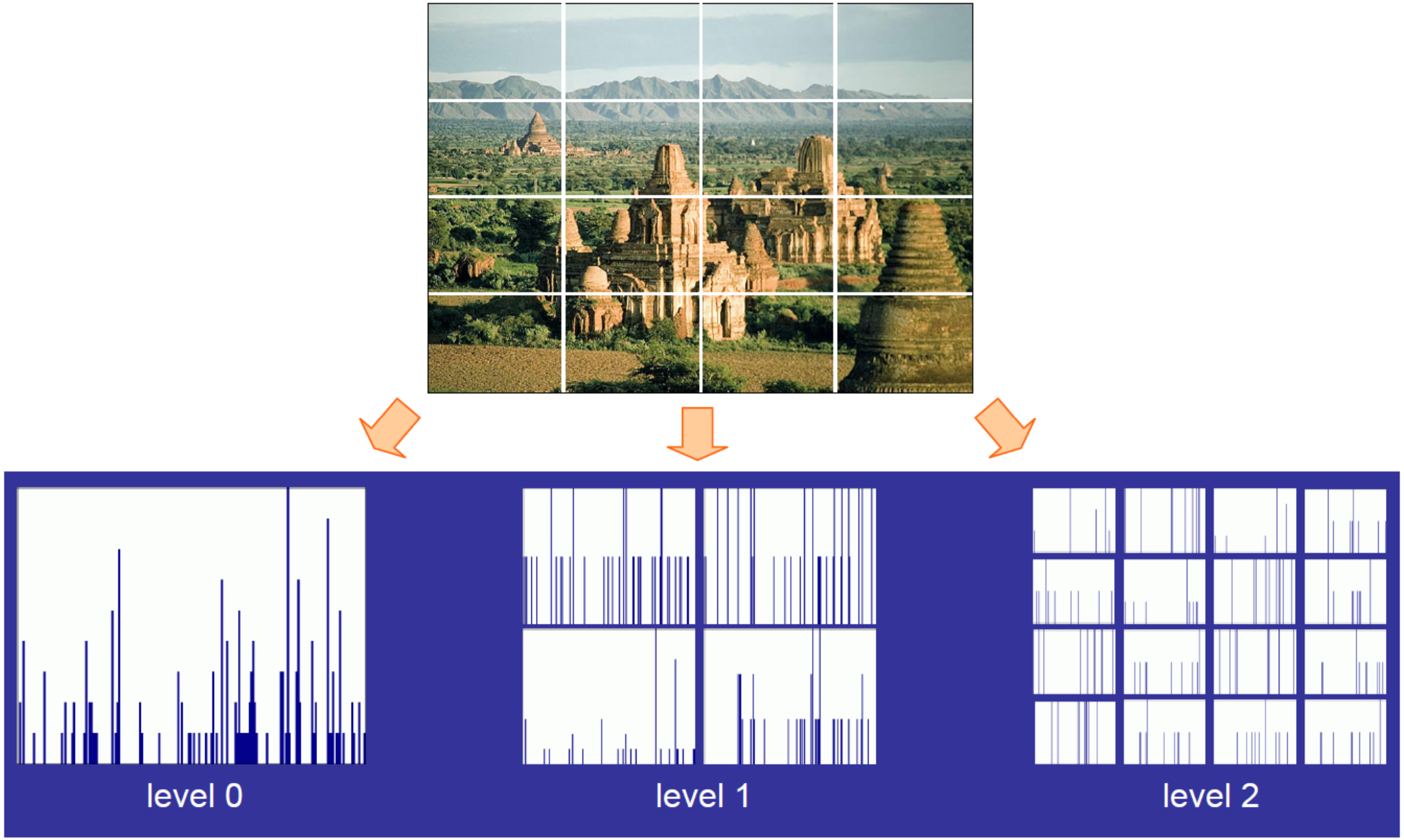
Weak (edge orientations)

OR

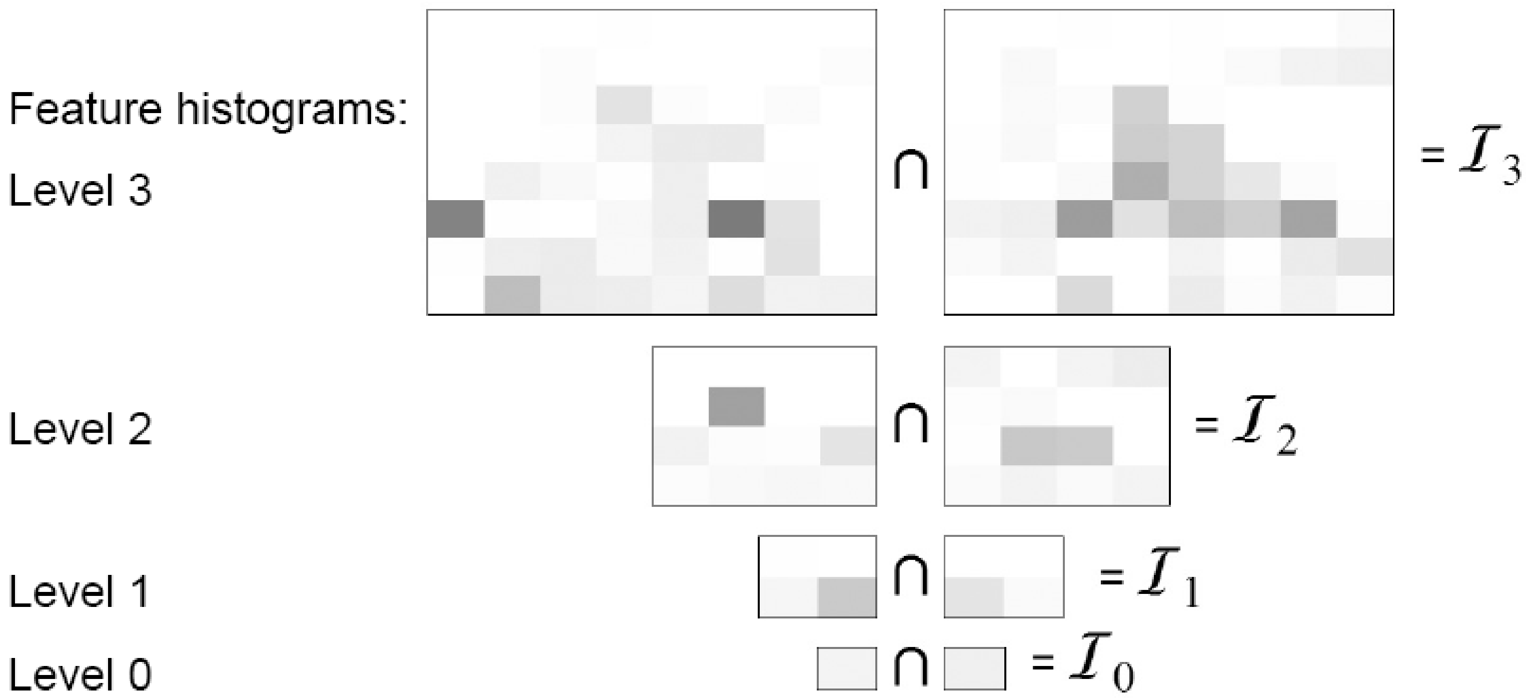


Strong (SIFT)

Spatial Pyramid Matching



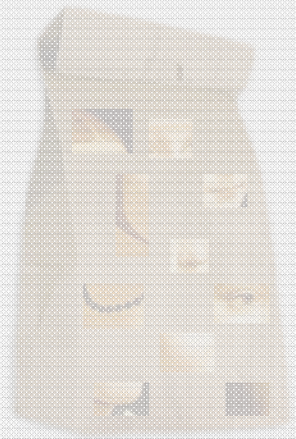
Spatial Pyramid Matching



Total weight (value of *pyramid match kernel*): $\mathcal{I}_3 + \frac{1}{2}(\mathcal{I}_2 - \mathcal{I}_3) + \frac{1}{4}(\mathcal{I}_1 - \mathcal{I}_2) + \frac{1}{8}(\mathcal{I}_0 - \mathcal{I}_1)$

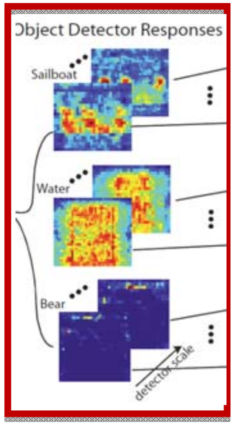
Image Representation

Bag of Words



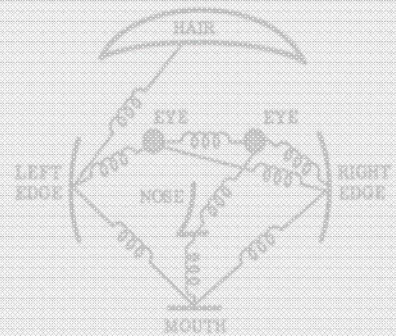
Weakly Spatial Models

•Spatial Pyramid Matching



•Object-Bank

Part-based



•Constellation Model
•Pictorial Structure

No spatial info. **Spatial Specificity of Parts** Very strong but sparse

Image classification

Event: "Sailing"

High level tasks

Object Bank

High Level Objects based

Sailboat, water, sky, tree, …

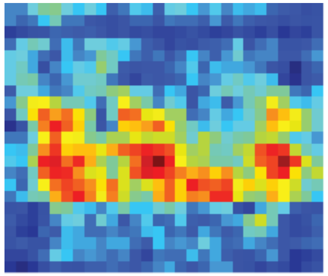


Image representation

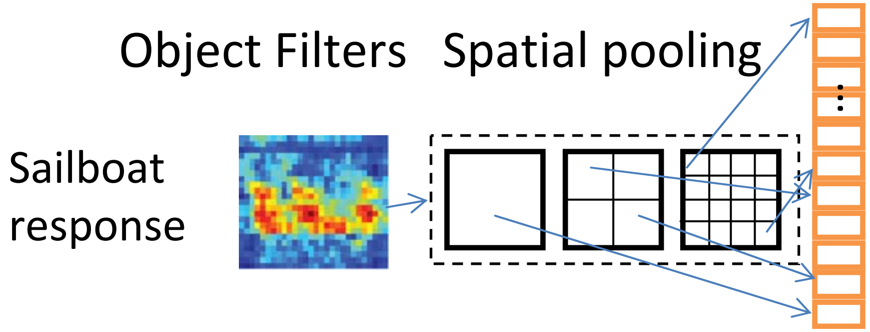
Semantic Gap

Low level feature

HoG, Gist , SIFT, Color, Texture, Bag of Words (BoW),Spatial Pyramid (SPM)

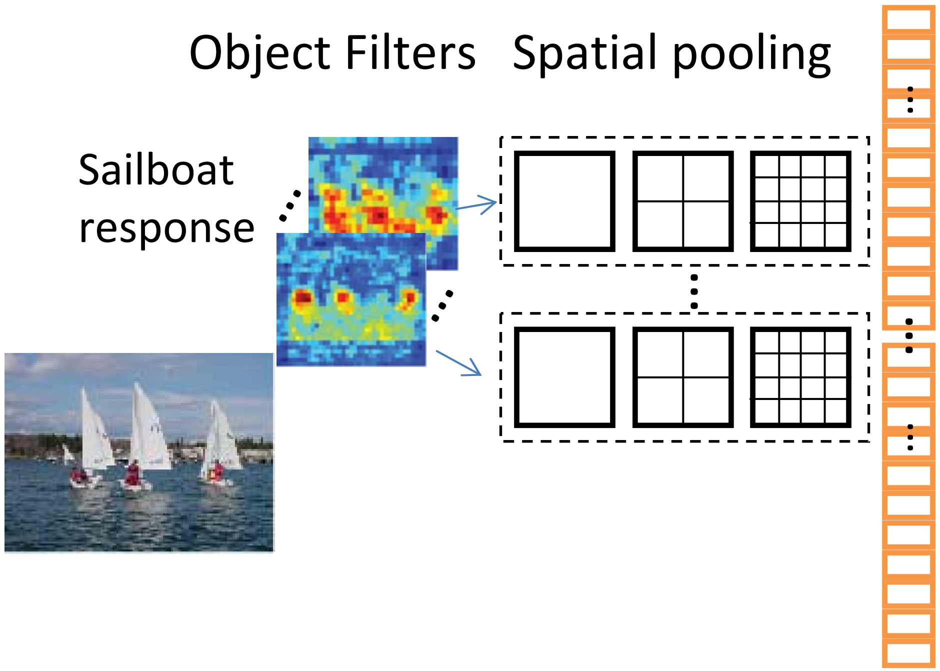
Li et al. 2010

Object Bank representation



Li et al. 2010

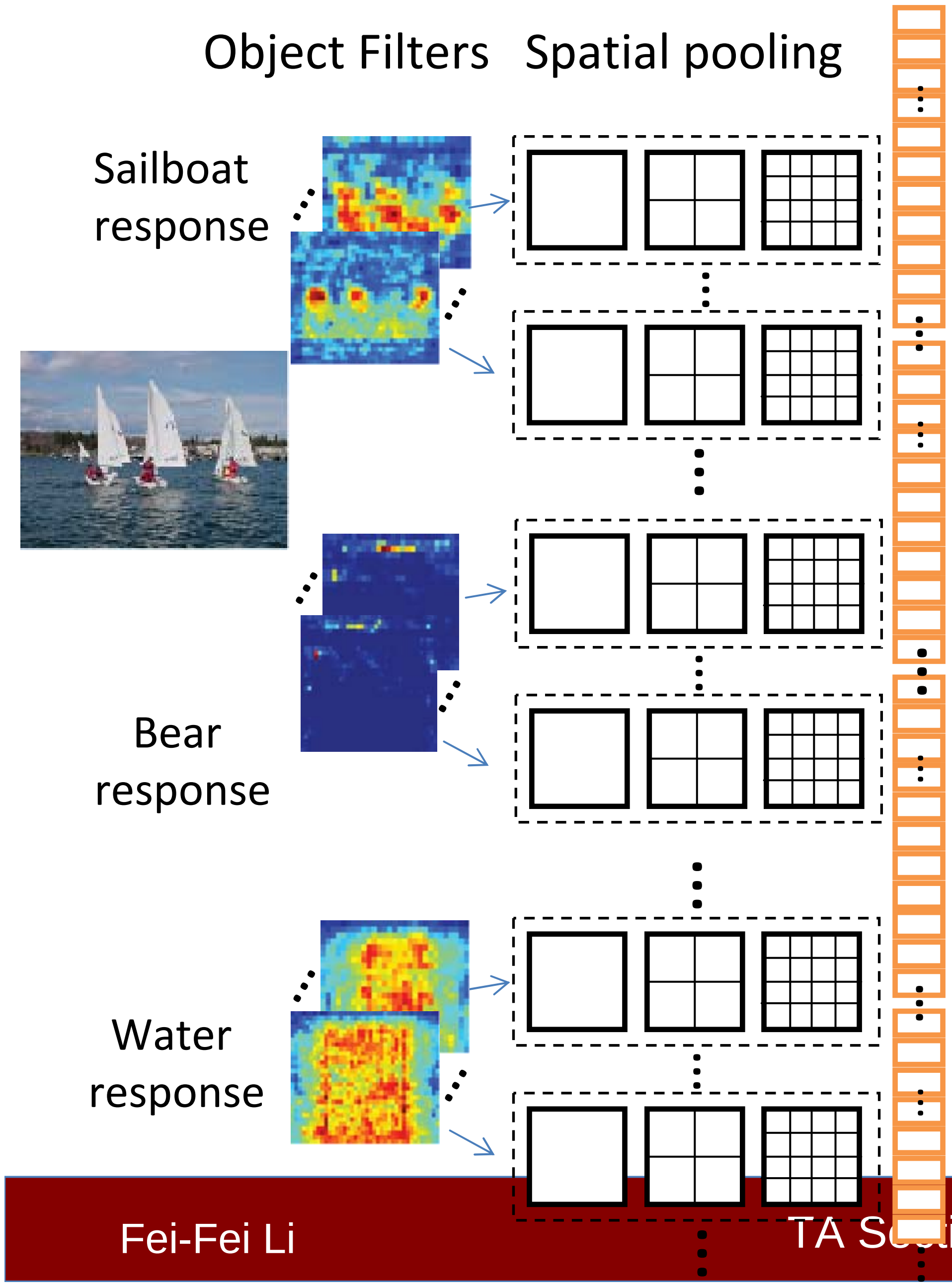
Object Bank representation



Object size variance

Li et al. 2010

Object Bank representation

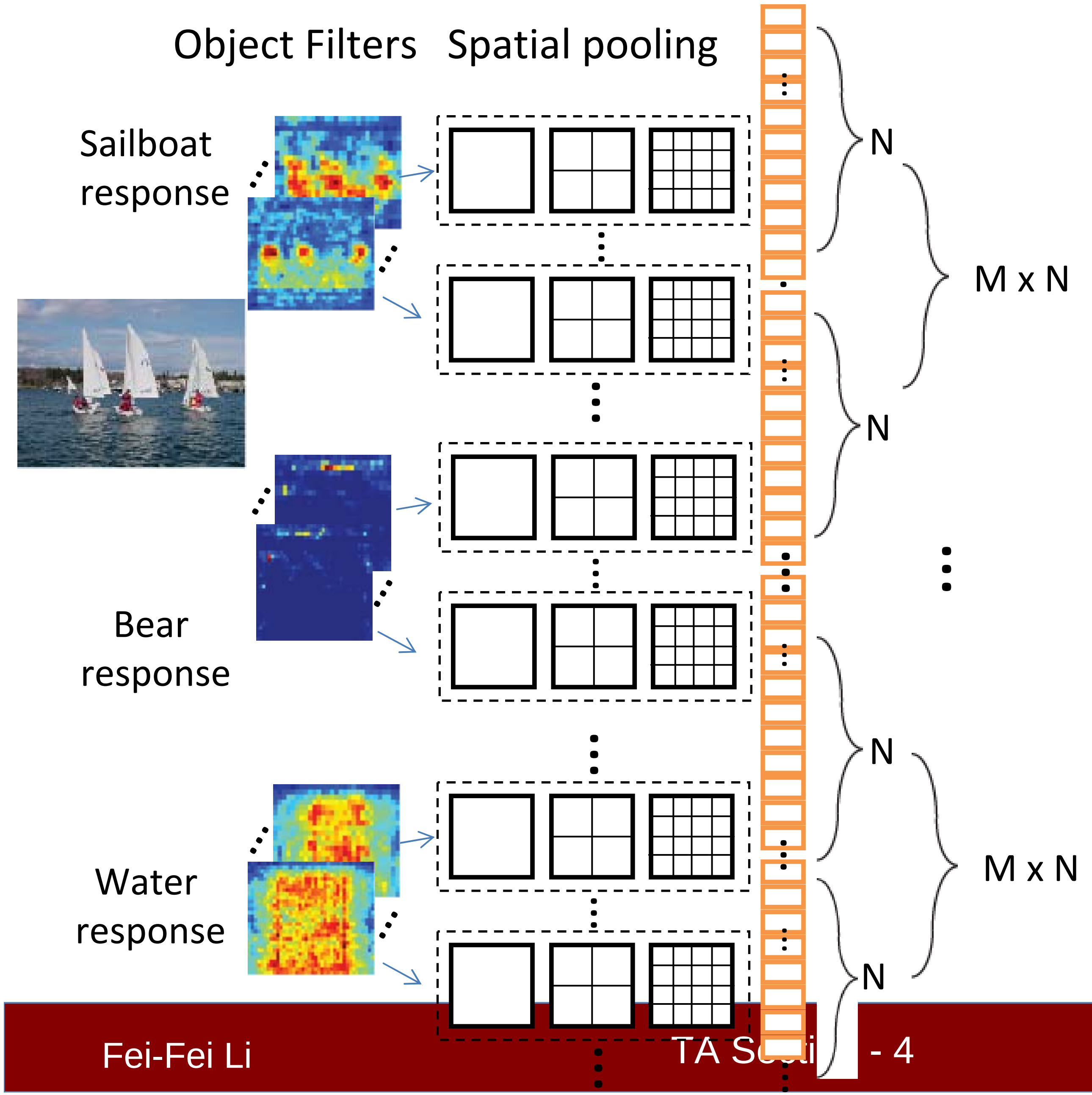


Implementation details

- ~ 200 object detectors
- Felzenswalb et al. 2008
- Hoeim et al. 2005
- 3-level spatial pyramid
- for each grid: max of each object

Li et al. 2010

Object Bank representation



N: nGridsperScale,
M: nScale,
O: nObject

M x N x O= thousands

E.g., In our setting:
12 x 21x 177=44604

Li et al. 2010

A word about Q2 in PS4

- We'd like you to understand the differences between BoW, SPM, and ObjectBank
- We'd like you to use what you've learned so far, be creative, and come up with interesting ways of encoding image information for an image recognition task
- Extra credits are given especially to innovation and good performances