SAIR

Spatial AI & Robotics Lab

CSE 473/573-A

L21: DETECTION

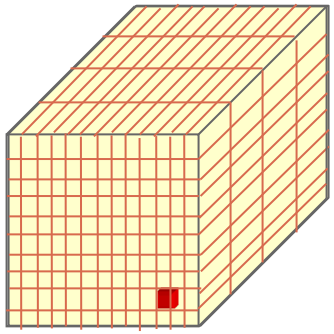
Chen Wang

Spatial AI & Robotics Lab

Department of Computer Science and Engineering

University at Buffalo The State University of New York

Why is detection hard?

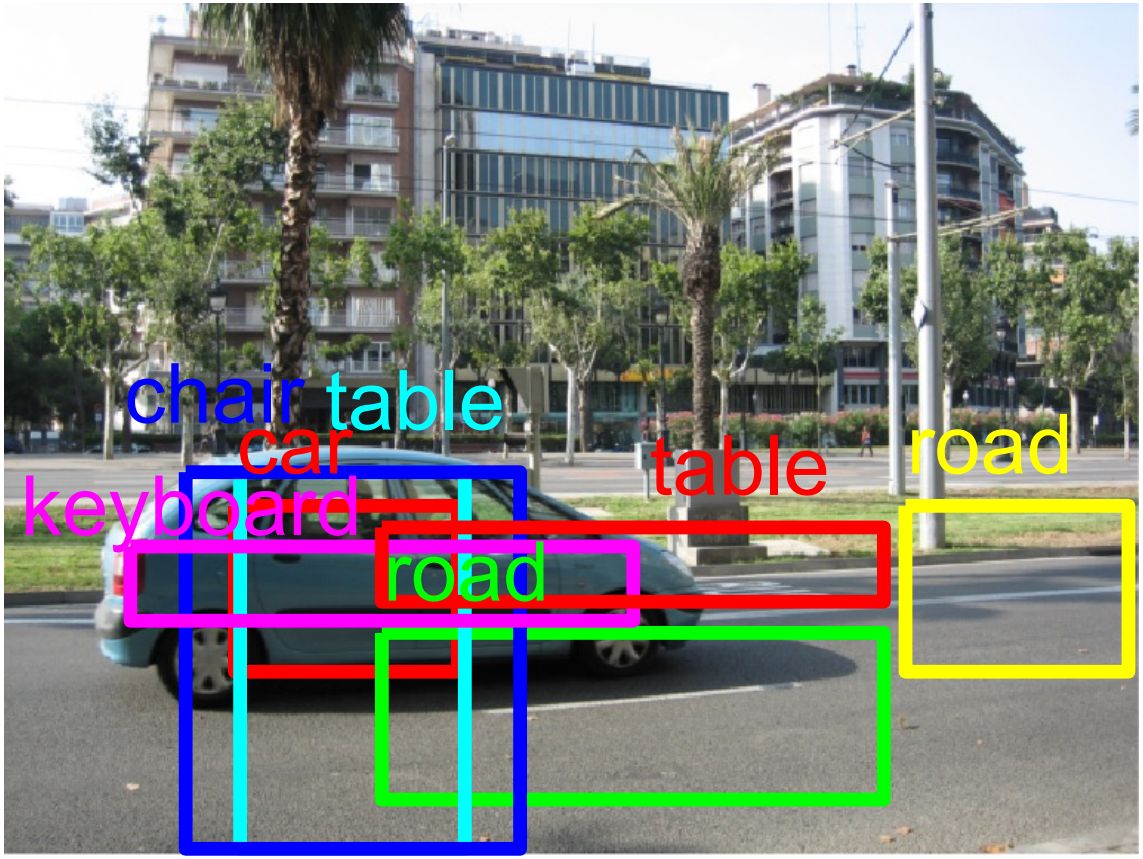


We want to do this for ~ 1000 objects

1,000,000 images/day

Why is detection hard?

If we have 1000 categories (detectors), and each detector produces 1 false every 10 images, we will have 100 false alarms per image… pretty much garbage…



Slide credit: A. Torralba

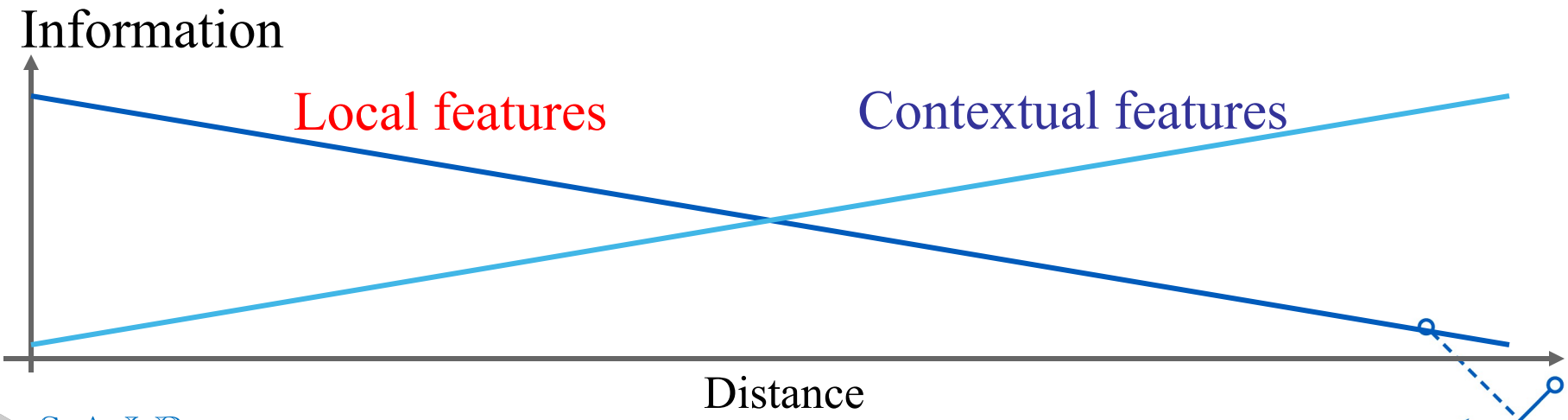
Local information is helpful

Slide credit: A. Torralba

Is local information enough?

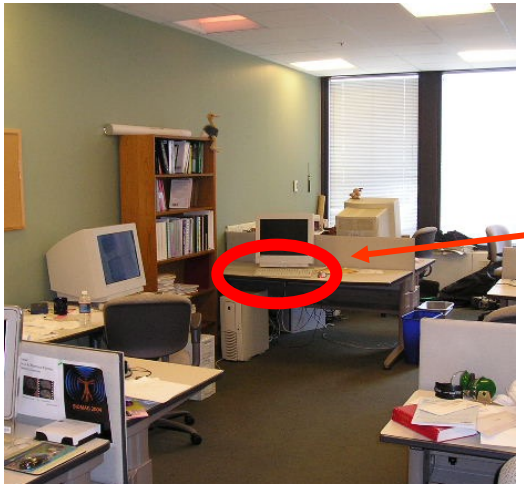




Is local information even enough?

We know there is a keyboard present in this scene even if we cannot see it clearly.

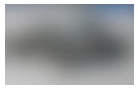


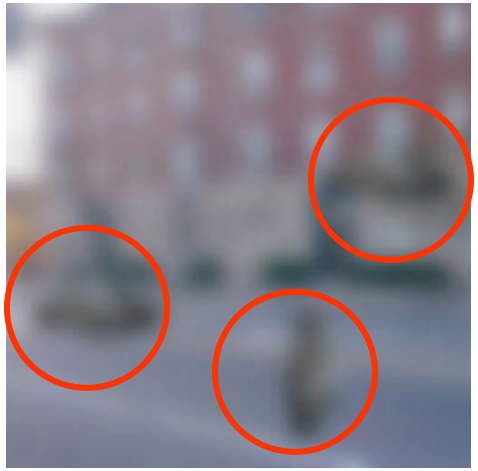
We know there is no keyboard present in this scene

… even if there is one indeed.

Slide credit: A. Torralba

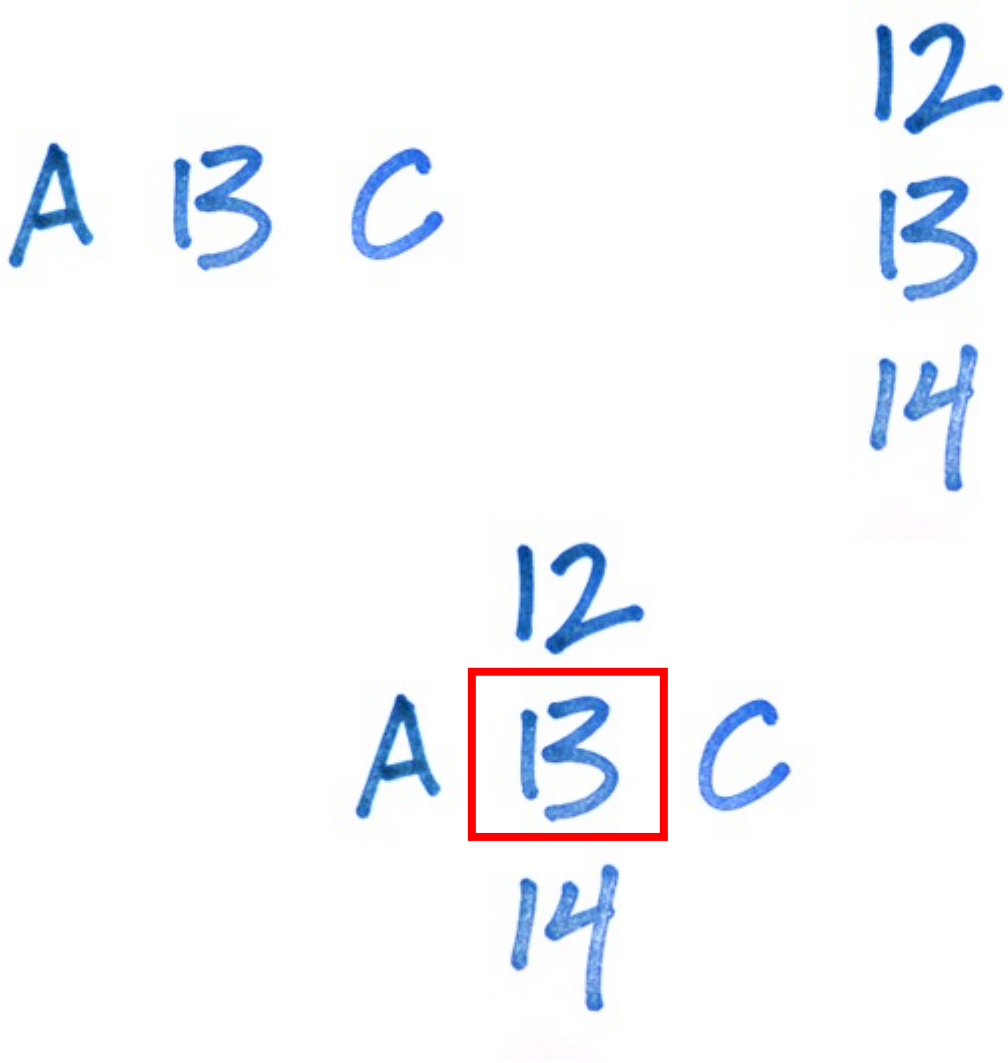
The multiple personalities of a blob



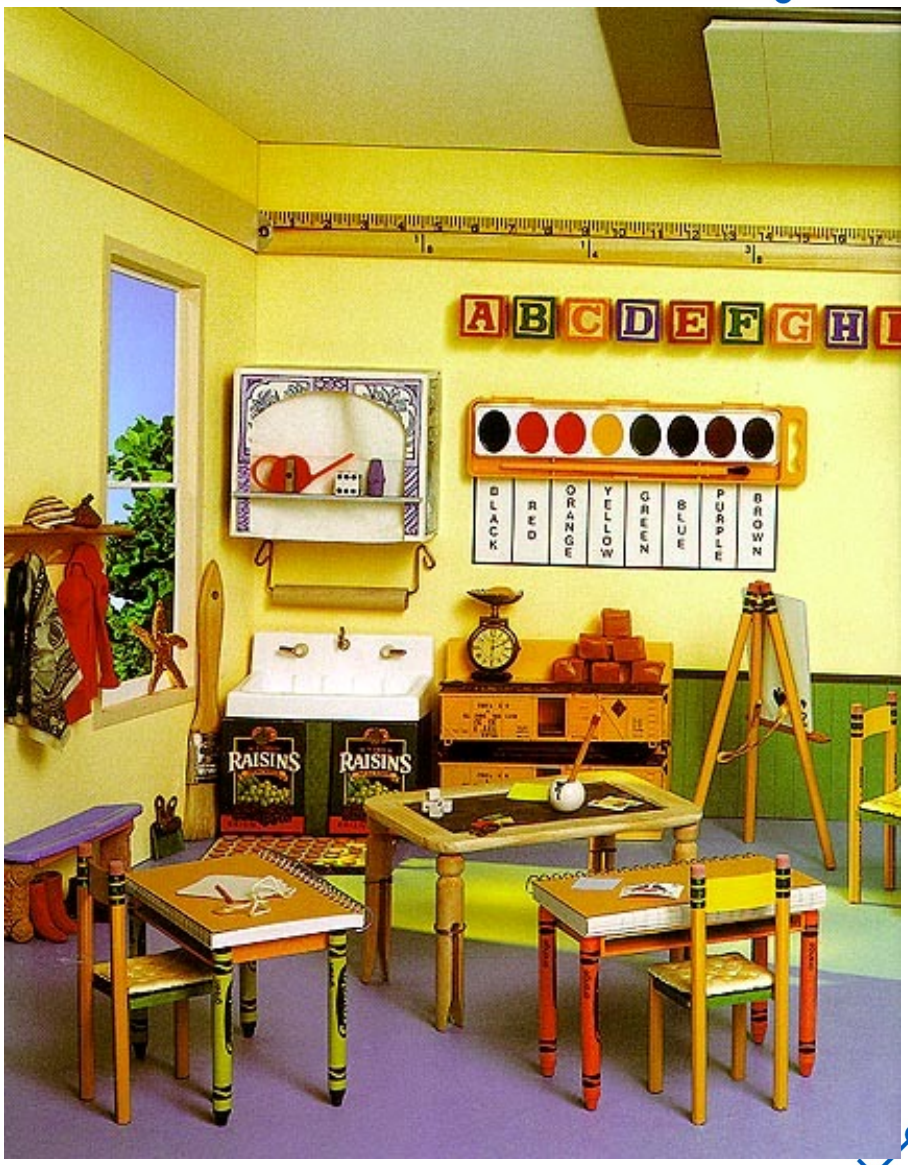


Slide credit: A. Torralba

The multiple personalities of a blob



The multiple personalities of a blob



Slide credit: A. Torralba

The context challenge

- What are the hidden objects?

Slide credit: A. Torralba

The context challenge

- What are the hidden objects?

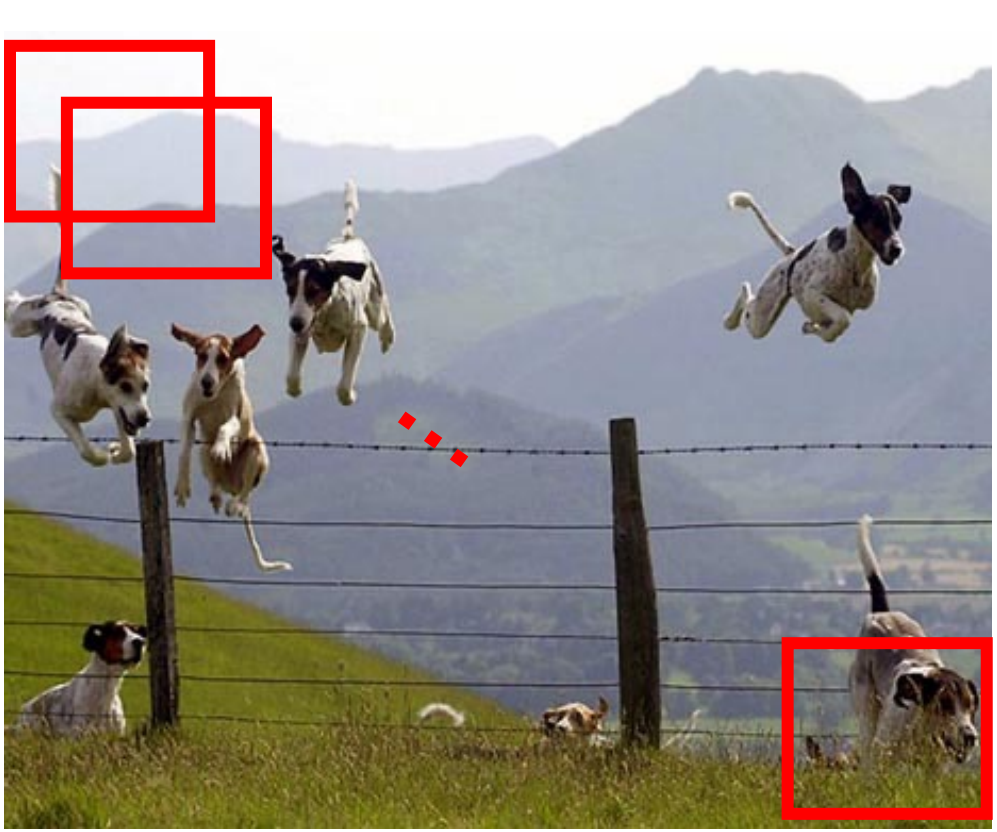
Slide credit: A. Torralba

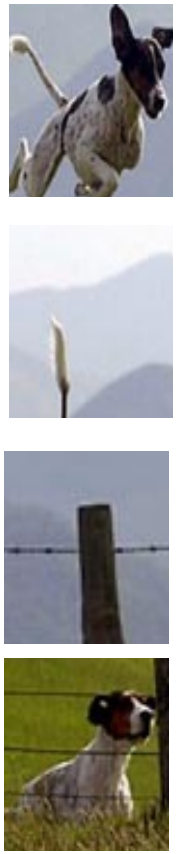
Object detection vs Scene Recognition

- Objects (even if deformable and articulated) probably have more **consistent shapes** than scenes.
- Scenes can be defined by distribution of “stuff” – materials and surfaces with **arbitrary shape**.
- Objects are **“things” that own their boundaries**
- Bag of words (BoW) were less popular for object detection because they **throw away shape** info.

Object Category Detection

- Focus on object search: “Where is it?”
- Build templates that quickly differentiate object patch from background patch





Dog Model

Object or
Non-Object?

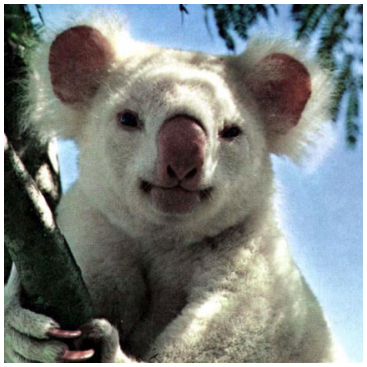
Challenges in modeling the object class

Illumination

Object pose

Clutter

Occlusions



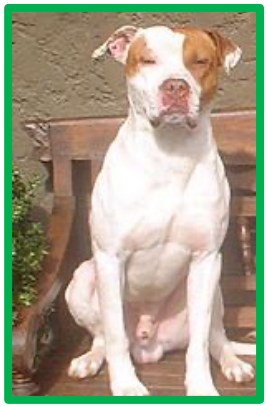
Intra-class appearance

Viewpoint

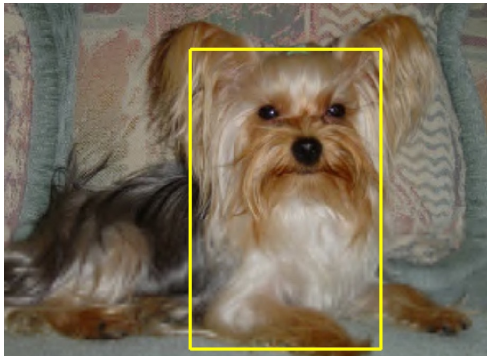
Slide from K. Grauman, B. Leibe

Challenges in modeling the non-object class

True Detections



Bad Localization

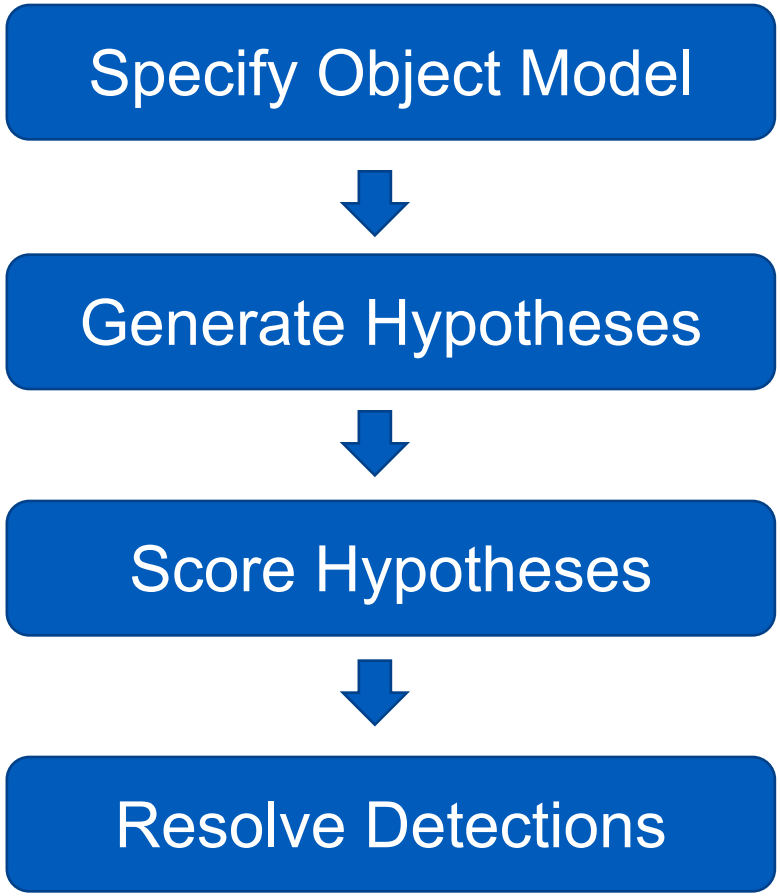


Confused with Similar Object

Misc. Background

Confused with Dissimilar Objects

General Process of Object Recognition

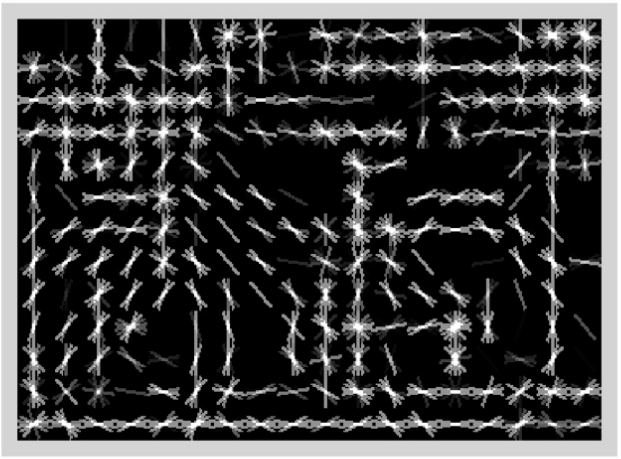


What are the object parameters?

Specifying an object model

1. Statistical Template in Bounding Box
   - Object is somewhere (x,y,w,h) in image
   - Features defined wrt bounding box coordinates

Image

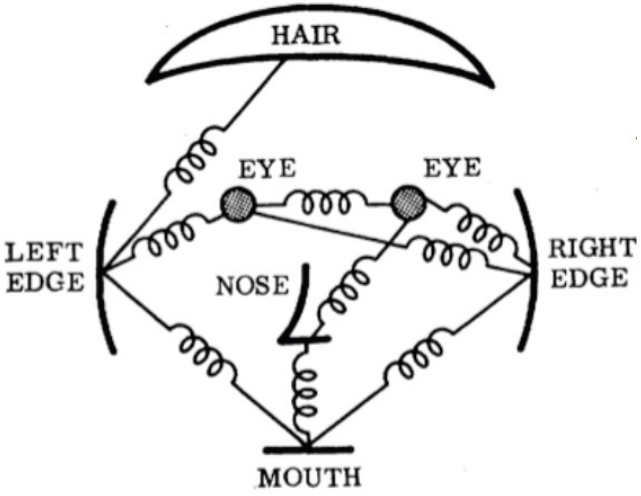


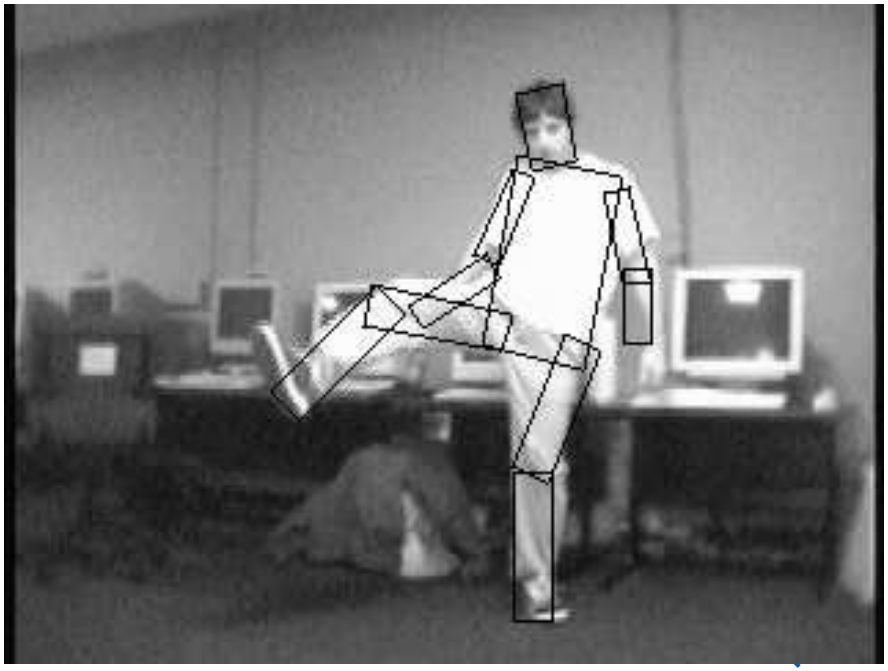
Template Visualization
(HoG features)

Images from Felzenszwalb

Specifying an object model

2. Articulated parts model
   - Object is configuration of parts
   - Each part is detectable



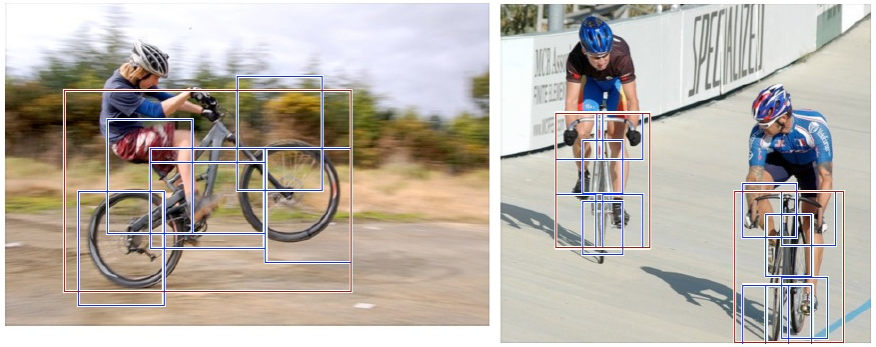


Images from Felzenszwalb

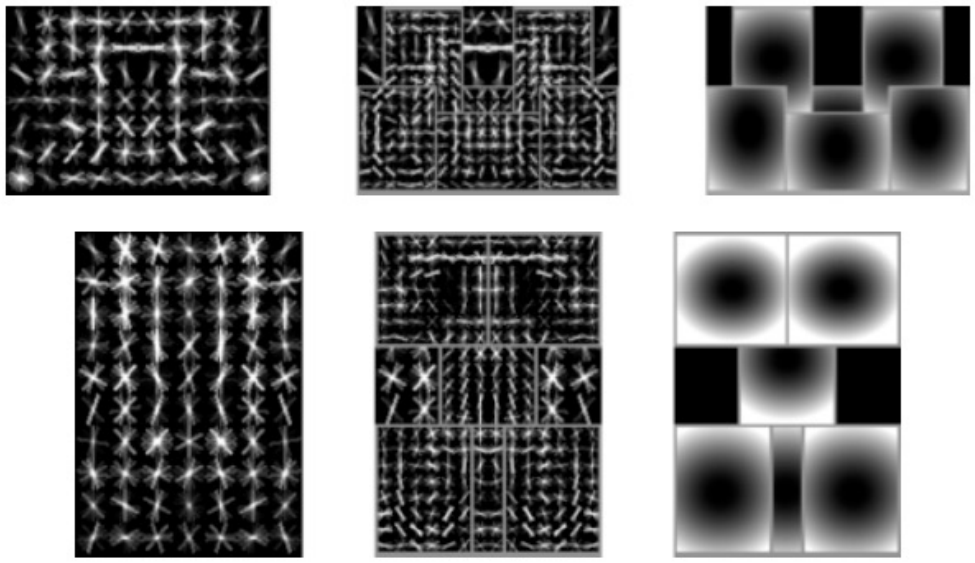
Specifying an object model

3. Hybrid template/parts model

Detections



Template Visualization
HoG features



root filters
coarse resolution

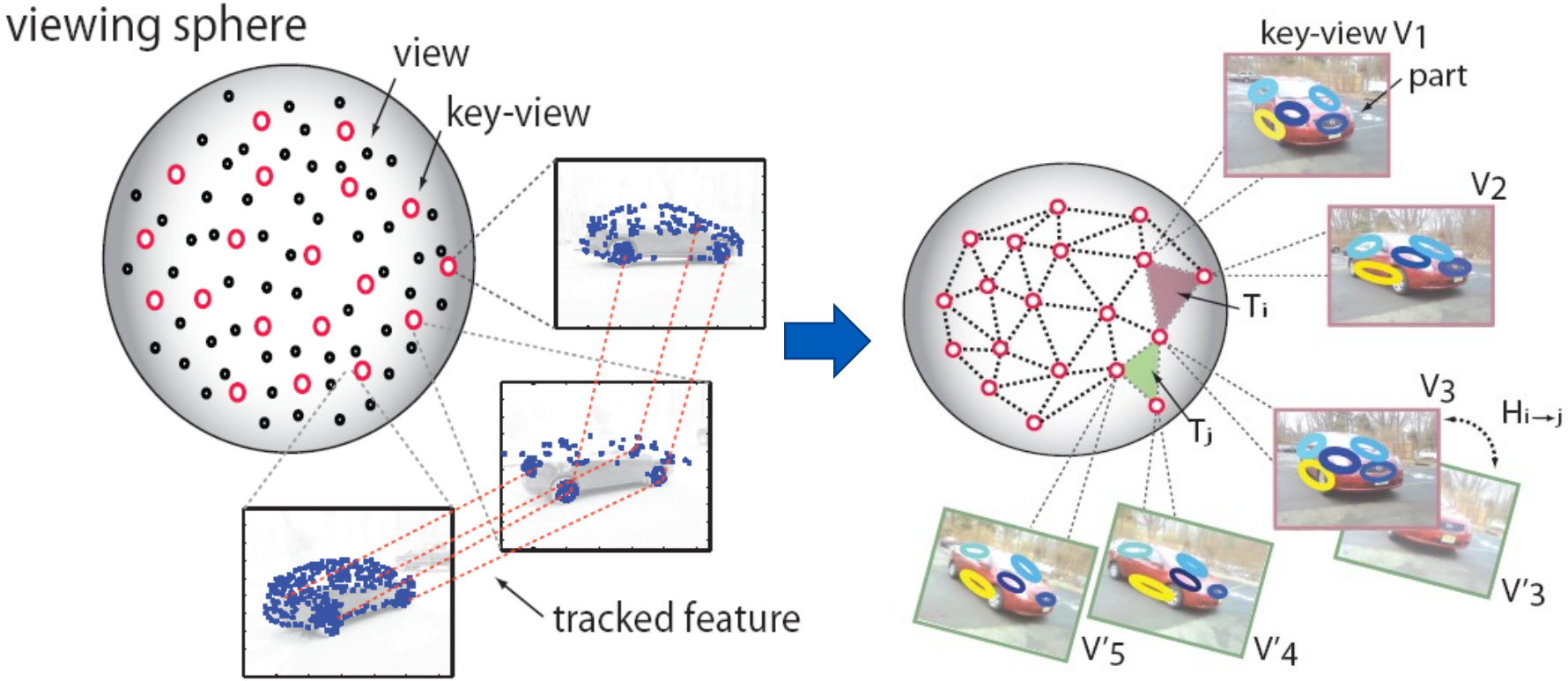
part filters
finer resolution

deformation
models

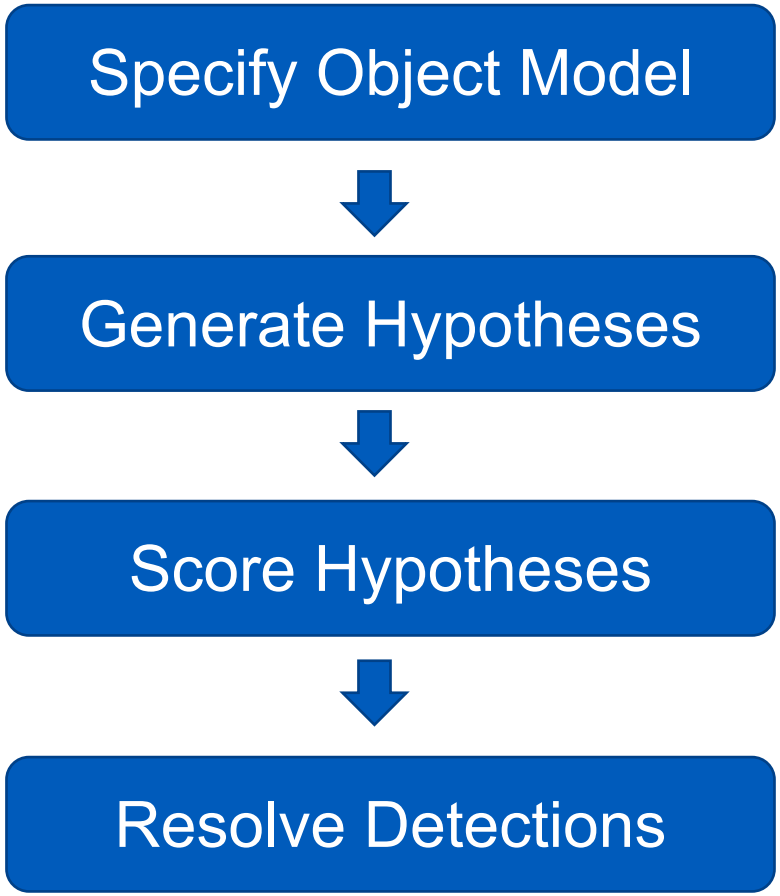
Felzenszwalb et al. 2008

Specifying an object model

4. 3D-ish model
   - Object is collection of 3D planar patches under affine transformation



General Process of Object Recognition



Propose an alignment of the model to the image

Generating hypotheses

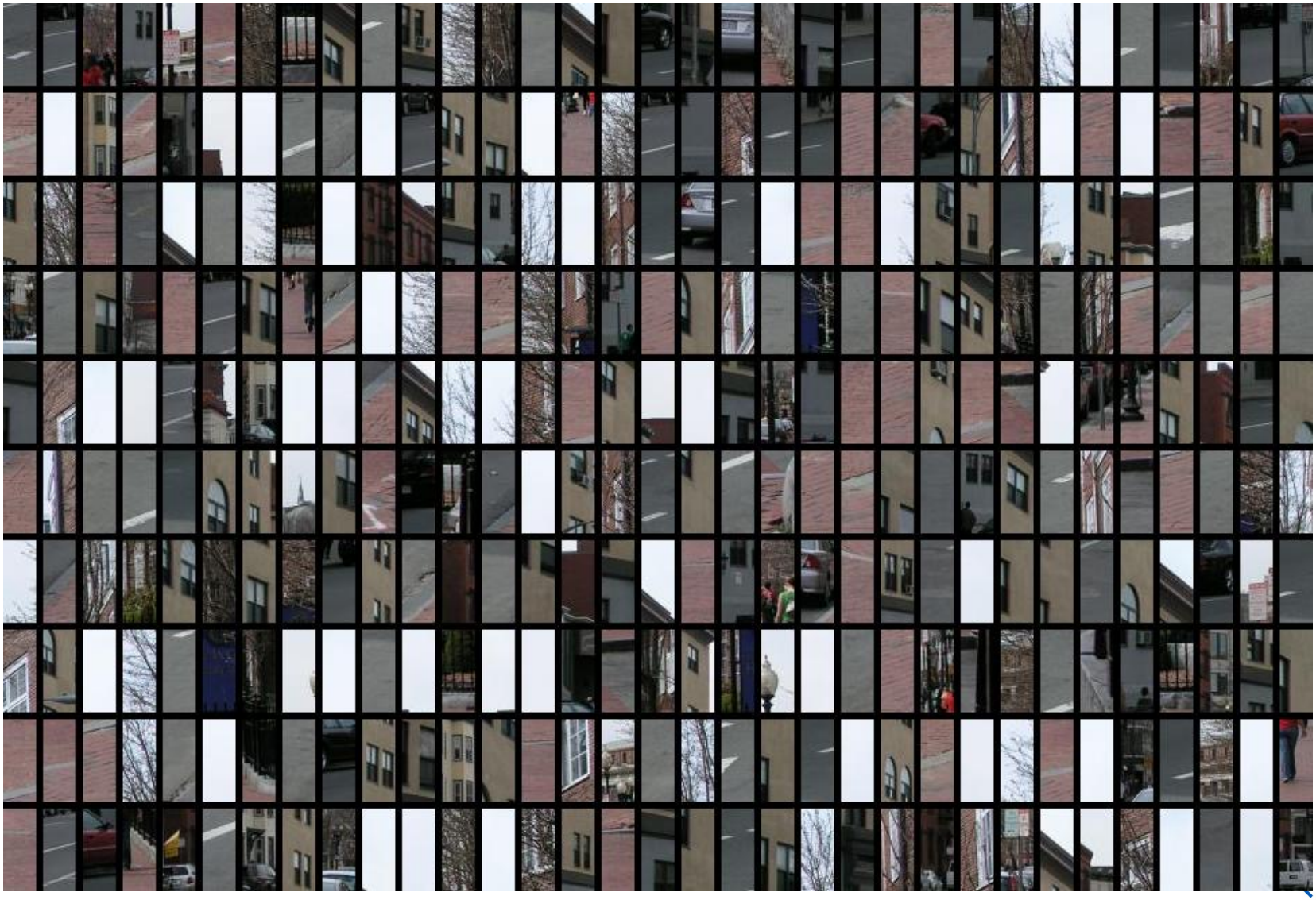
1. Sliding window
   - Test patch at each **location** and **scale**

Generating hypotheses

1. Sliding window
   - Test patch at each **location** and **scale (pyramids)**

Note – Template did not change size

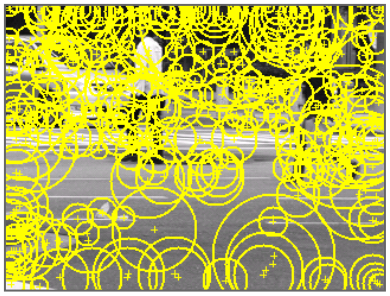
Each window is separately **classified**



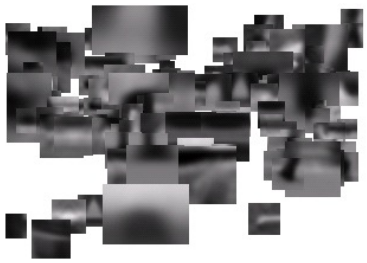
Generating hypotheses

2. Voting from patches/keypoints

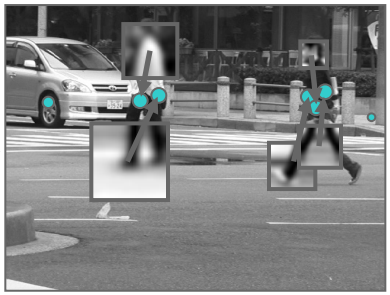
Interest Points



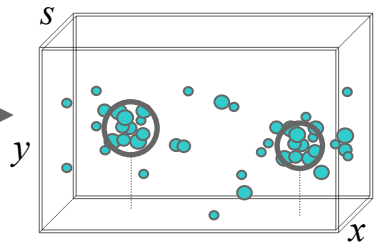
Matched Codebook Entries



Probabilistic Voting

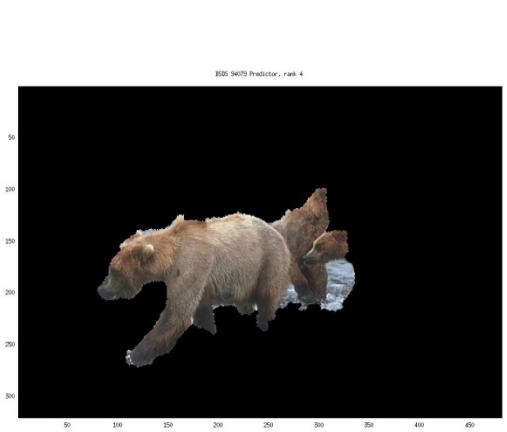


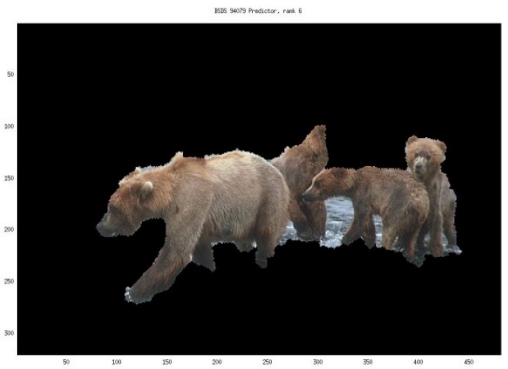
3D Voting Space (continuous)

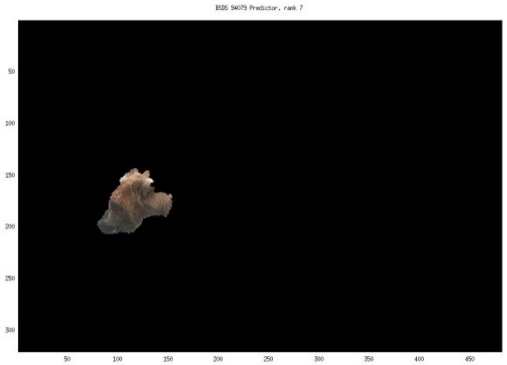


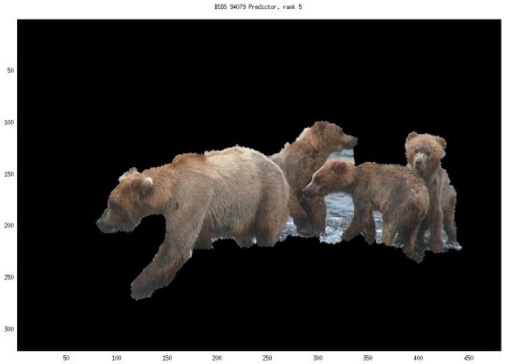
Generating hypotheses

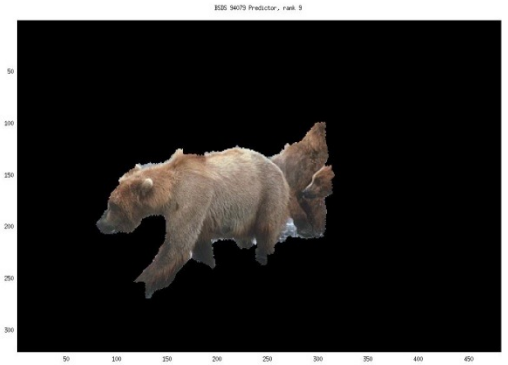
3. Region-based proposal



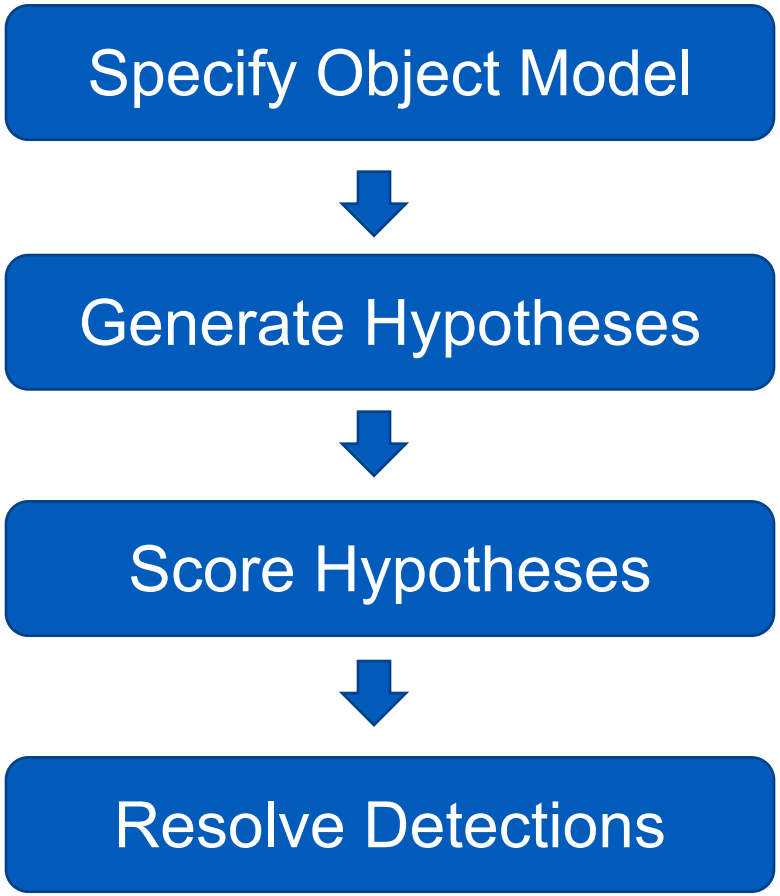






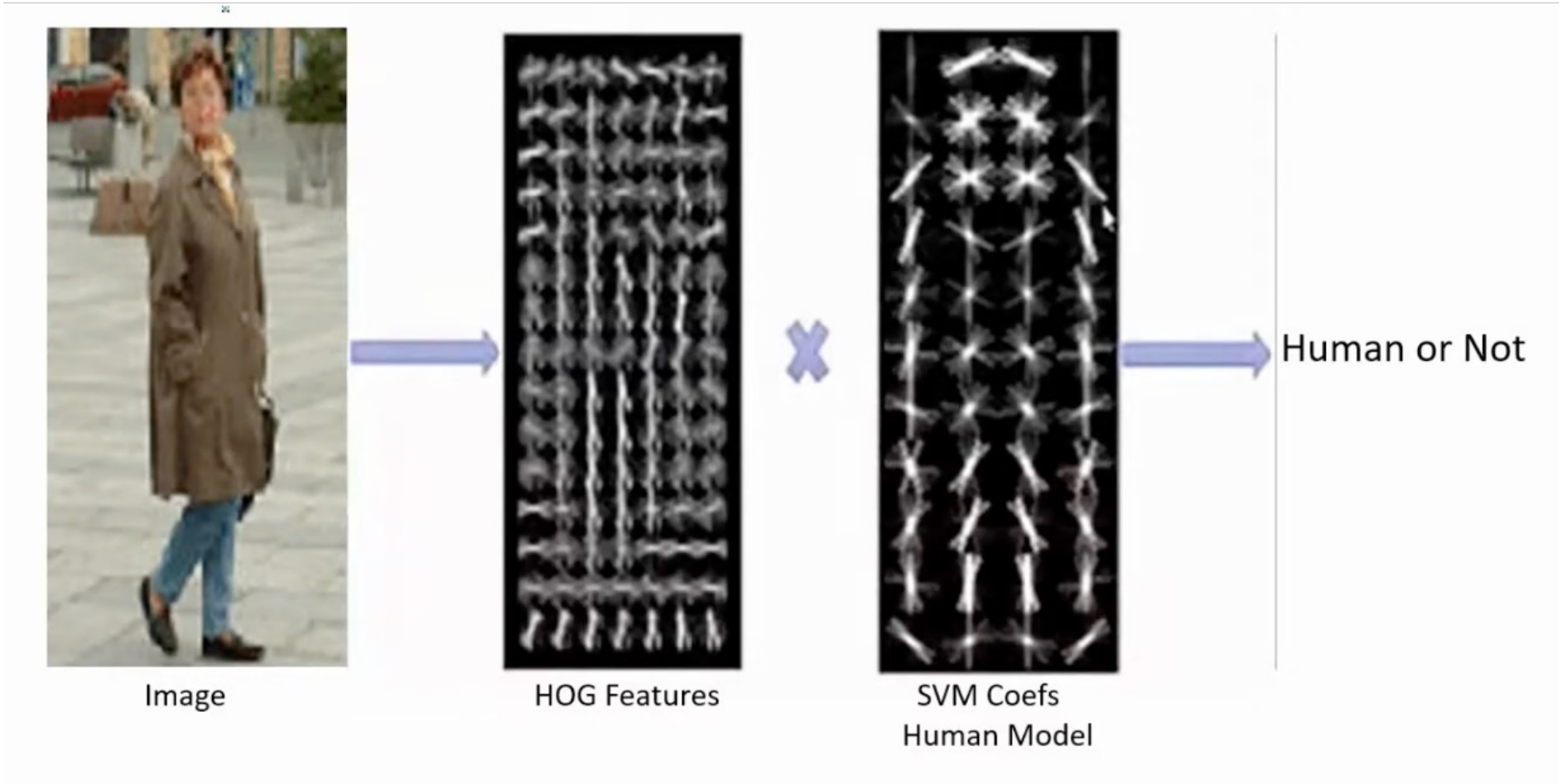


General Process of Object Recognition



Mainly-gradient based features, usually based on summary representation, many classifiers.

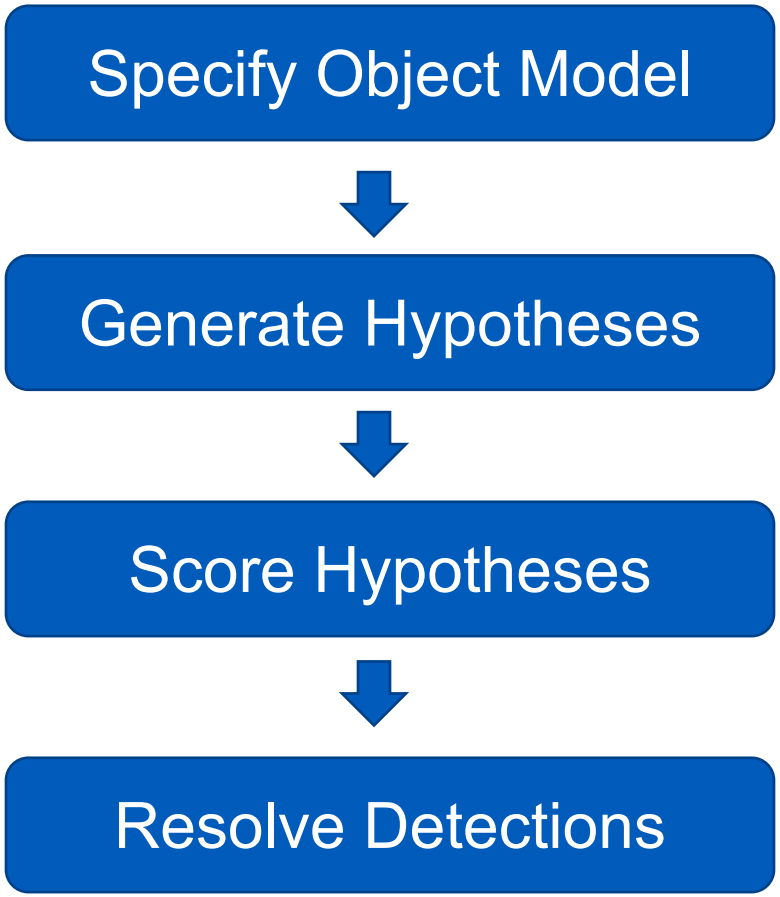
HOG + SVM for Object Detection



https://youtu.be/sDByI84n5mY

Histograms of Oriented Gradients for Human Detection, Navneet Dalal, Bill Triggs, International Conference on Computer Vision & Pattern Recognition, 2005

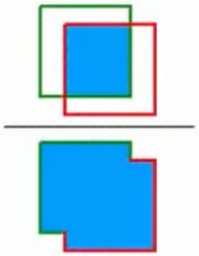
General Process of Object Recognition



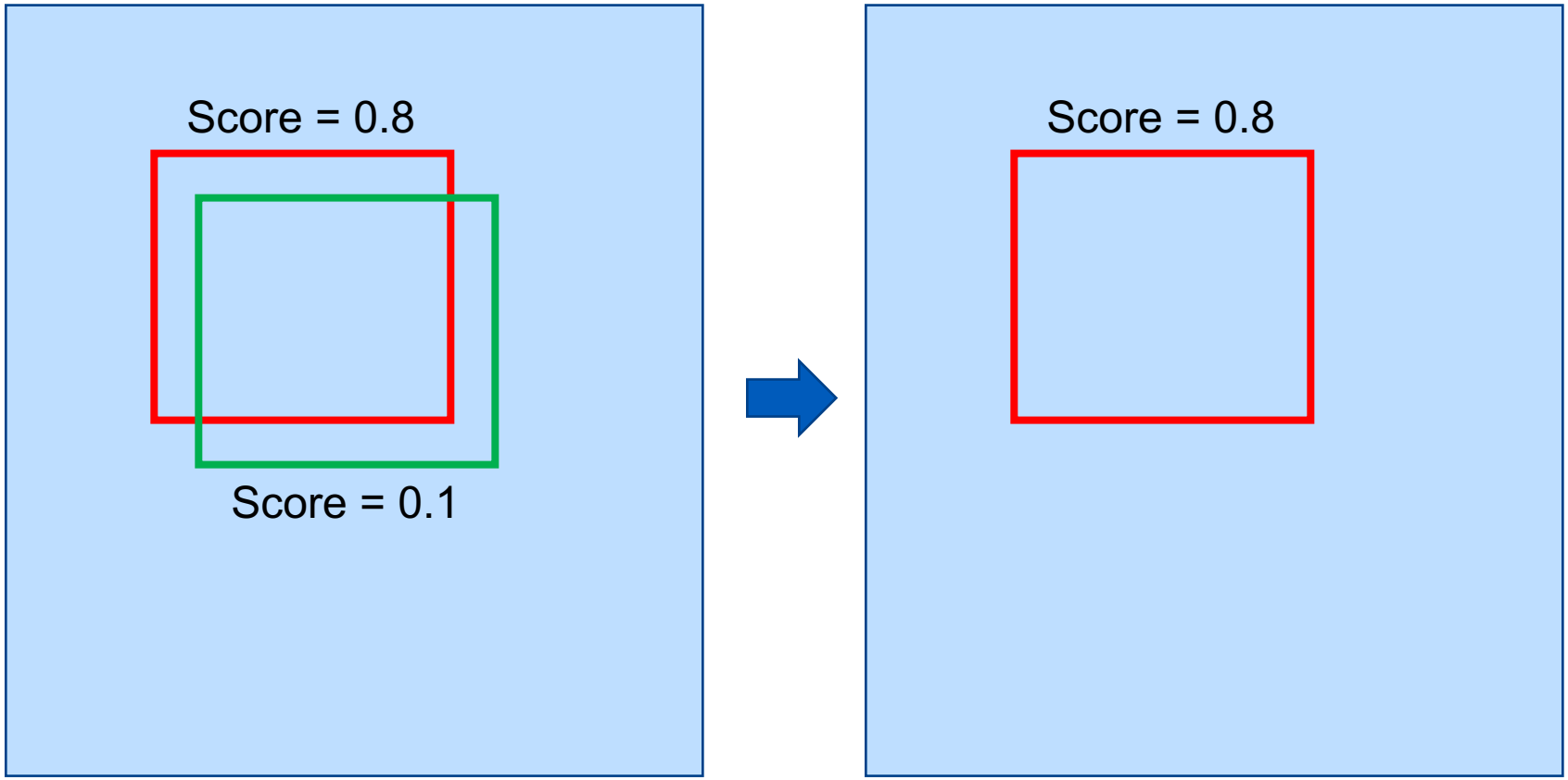
Rescore each proposed object based on whole set

Resolving detection scores

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$

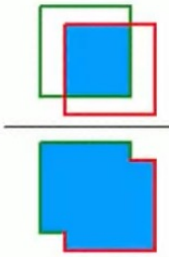


1. Non-max suppression

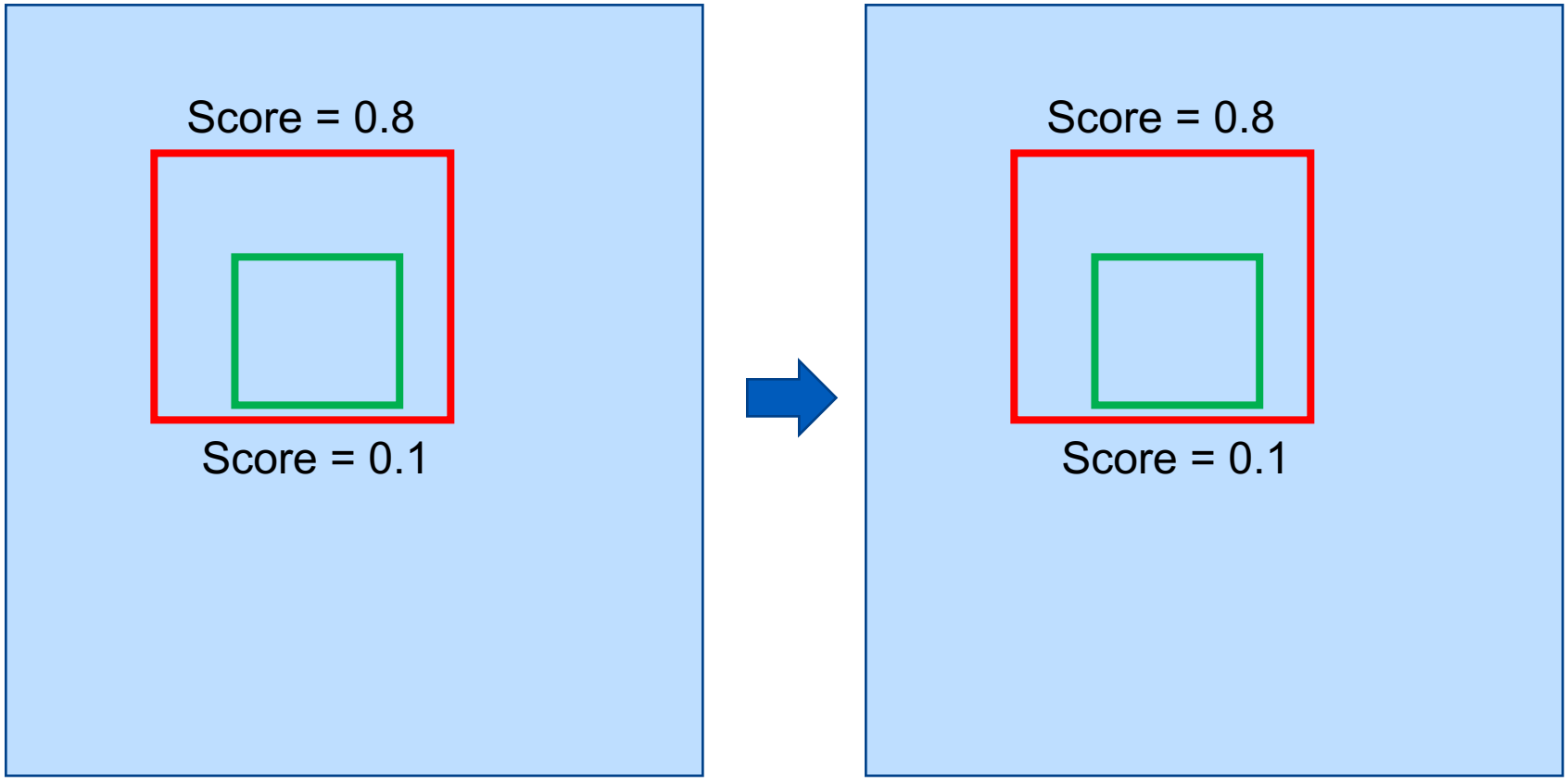


Resolving detection scores

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



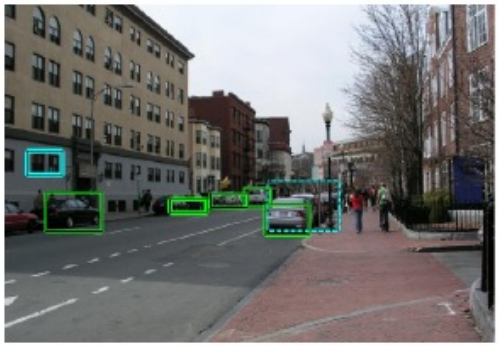
1. Non-max suppression



"Overlap" score is below some threshold

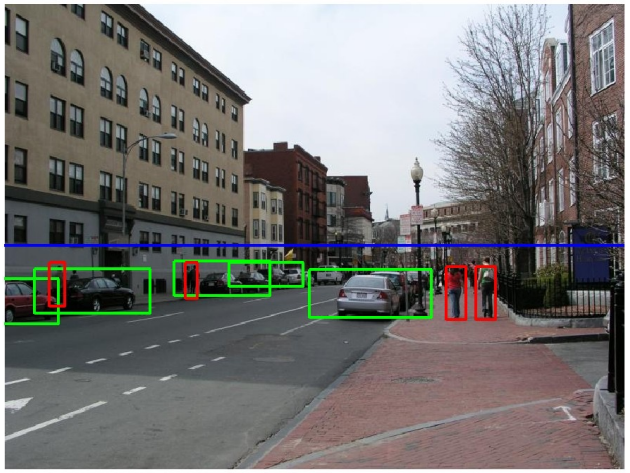
Resolving Detection Scores

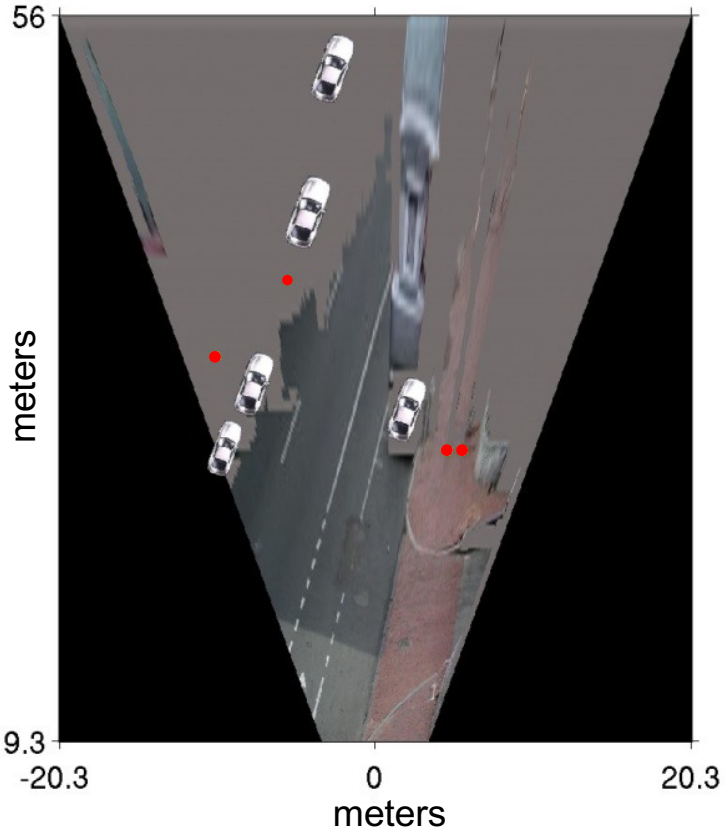
2. Context/reasoning



(g) Car Detections: Local

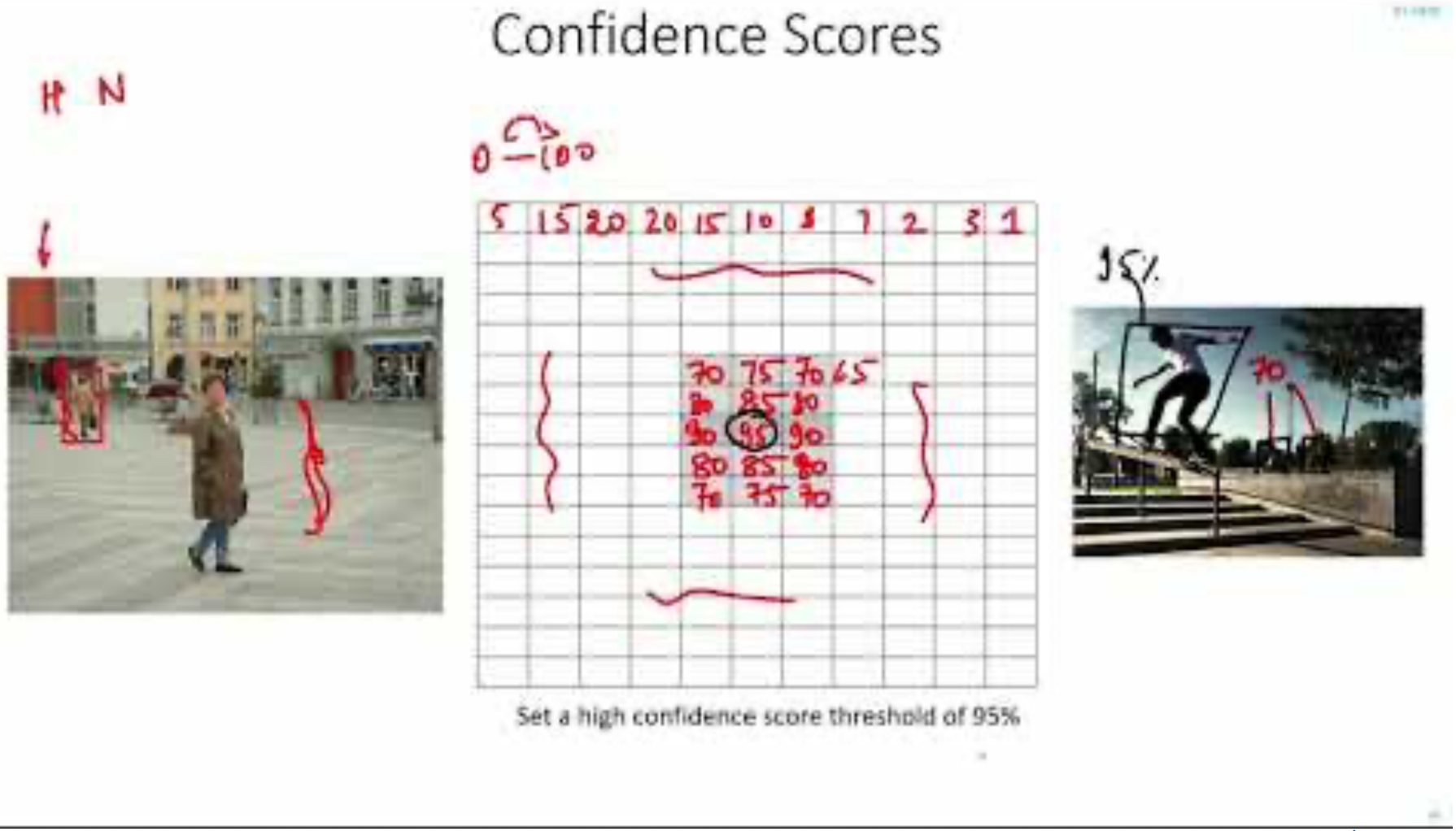
(h) Ped Detections: Local





Hoiem et al. 2006

Non-Max Suppression



https://youtu.be/07jFApuhh4I

How to measure performance?

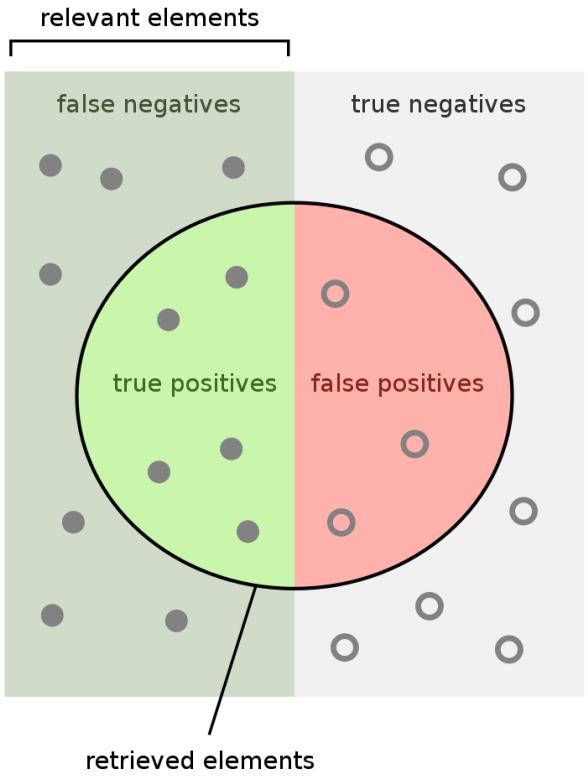
- Precision – How precise are the answers you gave?
  - (# Relevant)/(# Total Returned)
- Recall – How many did you find of the ones that could be found?
  - (# Relevant)/(# Total Relevant)
- F Measure – Harmonic Mean of Precision and Recall
  - (2*Precision*Recall)/(Precison+Recall)

Precision, Recall, F1 (Recap)

$$\text{Precision} = \frac{\text{True Positive}}{\text{Predicted Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Actual Positive}}$$

$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$



How many retrieved items are relevant?

Precision =

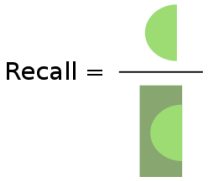
How many relevant items are retrieved?

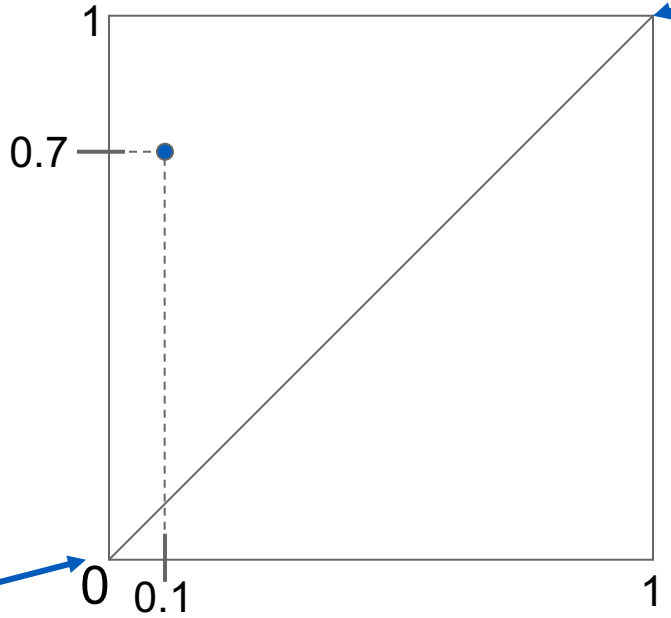
Recall =

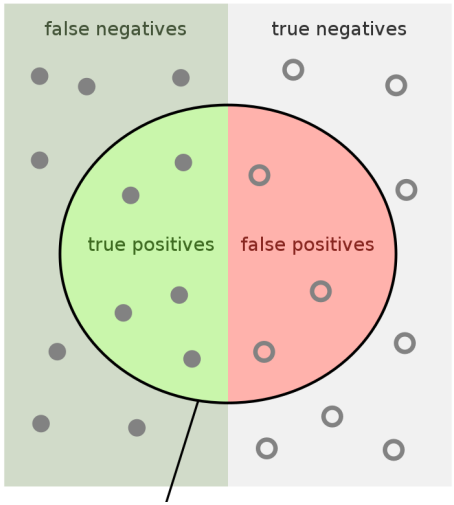
ROC curve (Recap)

- How can we measure the performance of a feature matcher?

Lowest threshold
all pairs as match







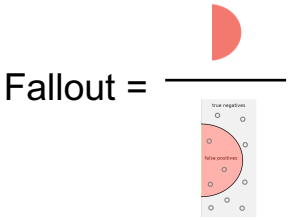
True Positive Rate

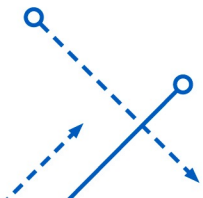
$$\frac{\#\,TP}{\#\,TP + FN}$$

Highest threshold
no match

False Positive Rate

$$\frac{\#\,FP}{\#\,FP + TN}$$

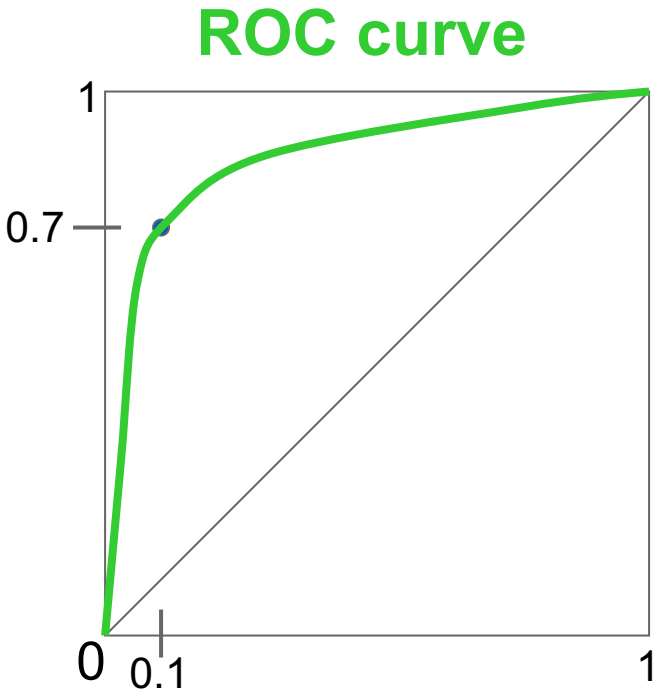




ROC curve (Recap)

ROC Curves (Receiver Operator Characteristic)

- Generated by counting # correct/incorrect matches, for different thresholds.
- Want to maximize area under the curve (AUC)
- Useful for comparing different feature matching methods.



true positive rate

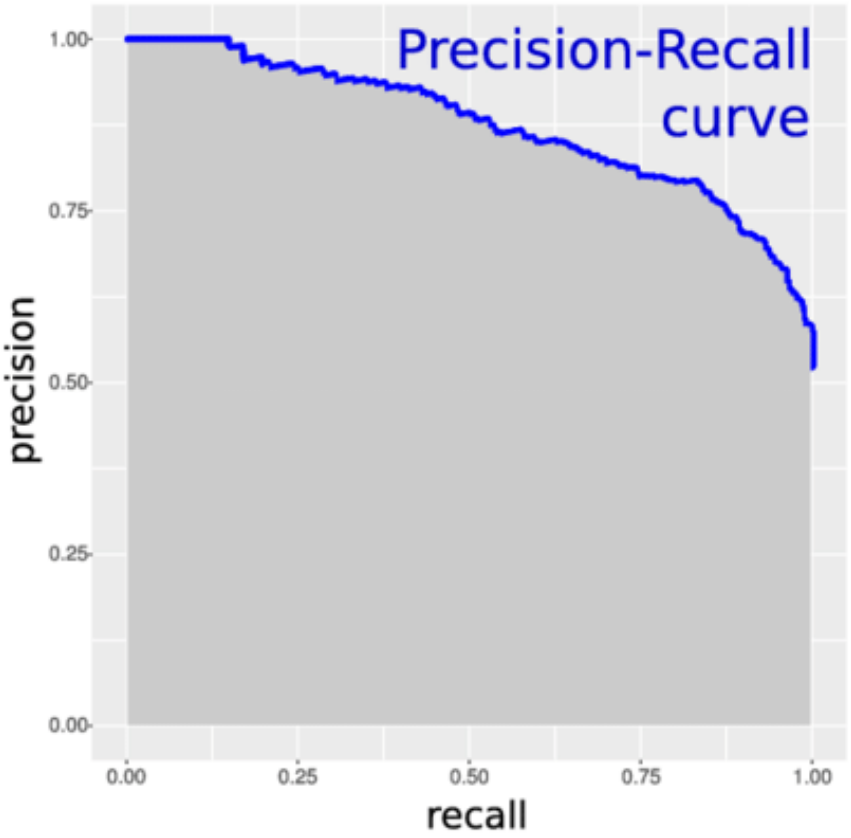
$$\frac{\#\ TP}{\#\ TP + FN}$$

false positive rate

$$\frac{\#\ FP}{\#\ FP + TN}$$

Precision and Recall Curve (PR curve)

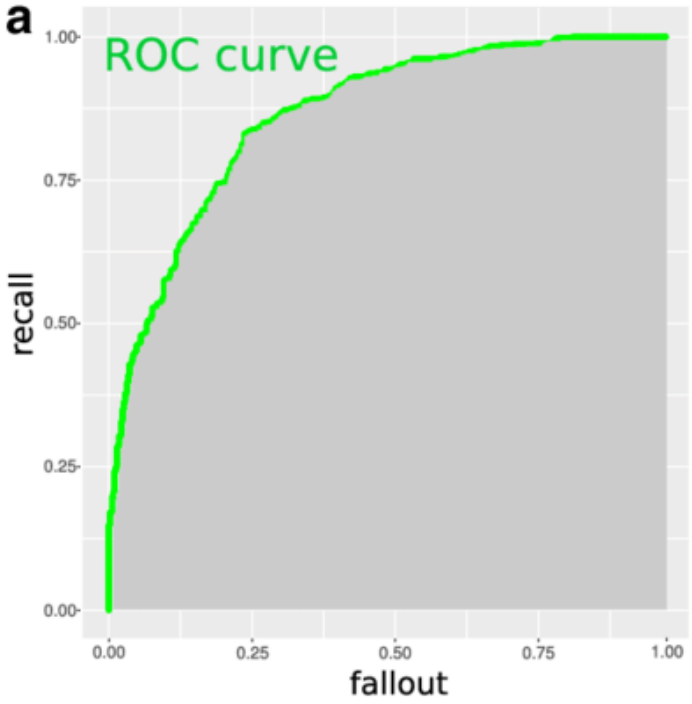
precision = #relevant / #returned
recall = #relevant / #total relevant

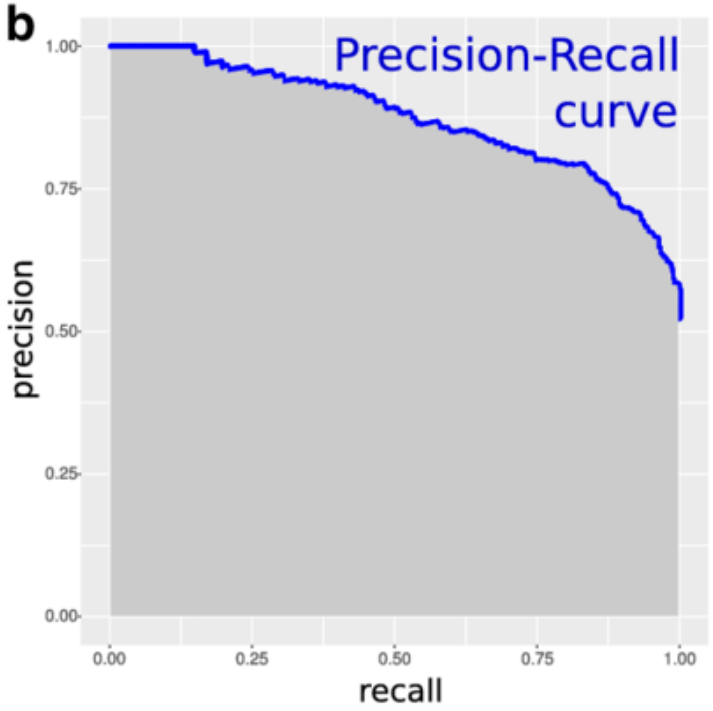


Slide credit: Ondrej Chum

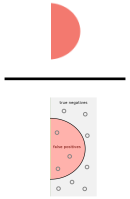
ROC curve VS. PR curve

Recall = 





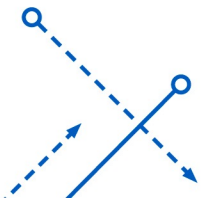
Precision = 

Fallout = 

Recall = 

ROC curve VS. PR curve

- ROC Curves (Lecture 4)
  - summarize the trade-off between the **true positive rate** and **false positive rate** for a predictive model using different probability thresholds.
- PR curves
  - summarize the trade-off between the **true positive rate** and the **positive predictive value** for a predictive model using different probability thresholds.



ROC curve VS. PR curve

- ROC: good for balanced datasets/classes.
- PR curves: appropriate for imbalanced datasets.
- If the proportion of positive to negative instances changes in a test set, the ROC curves will not change, so do not depend on class distributions.
- ROC “weights” equally true positives and true negatives (not expected sometimes, e.g., place recognition).

More explanation