VOC PASCAL 2007 Detection Challenge Oxford

Ondřej Chum
Andrew Zisserman
University of Oxford


Lambert's Coaches


HONDA


2628

Outline

- Representation
  - Exemplar model
- Learning
  - Detector
  - Classifier
- Detection
  - Discrimination
  - Non maxima suppression

O. Chum and A. Zisserman: An Exemplar Model for Learning Object Classes, *CVPR'07*

Representation

Spatial pyramid of histograms

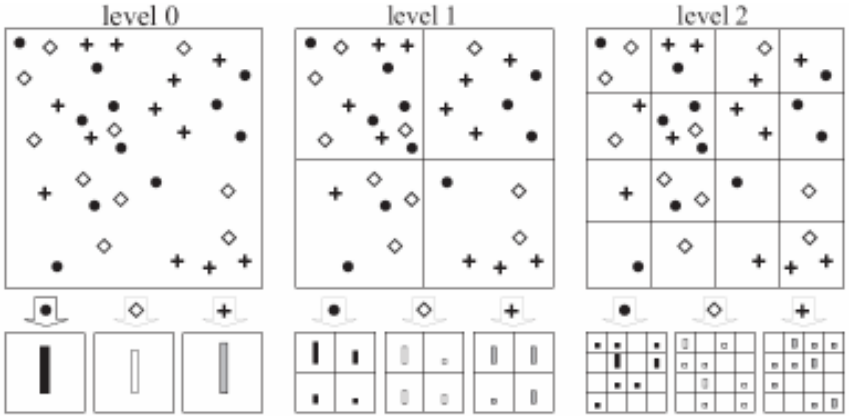
- Sparse features
  - Hessian Laplace operator
  - SIFT descriptor
  - Vector quantization: k-means
  - Weighted histograms
  - Weights of visual words are proportional to their discriminability for class
  - Spatial and scale pyramid
- Edge directions
  - Berkeley edge detector

Model

Sparse

Visual Words

Hessian Laplace + SIFT + k-means

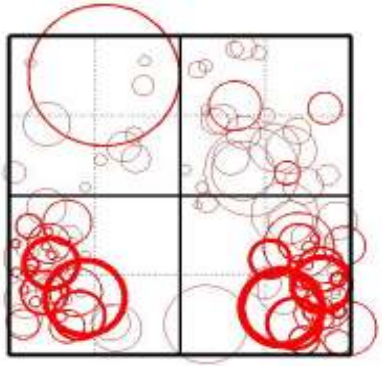


Lazebnik et al CVPR 2006

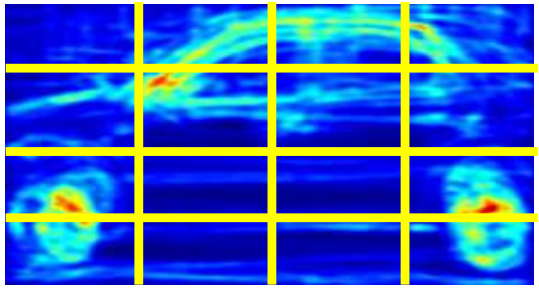
Dense

Edges

Berkeley edge detector



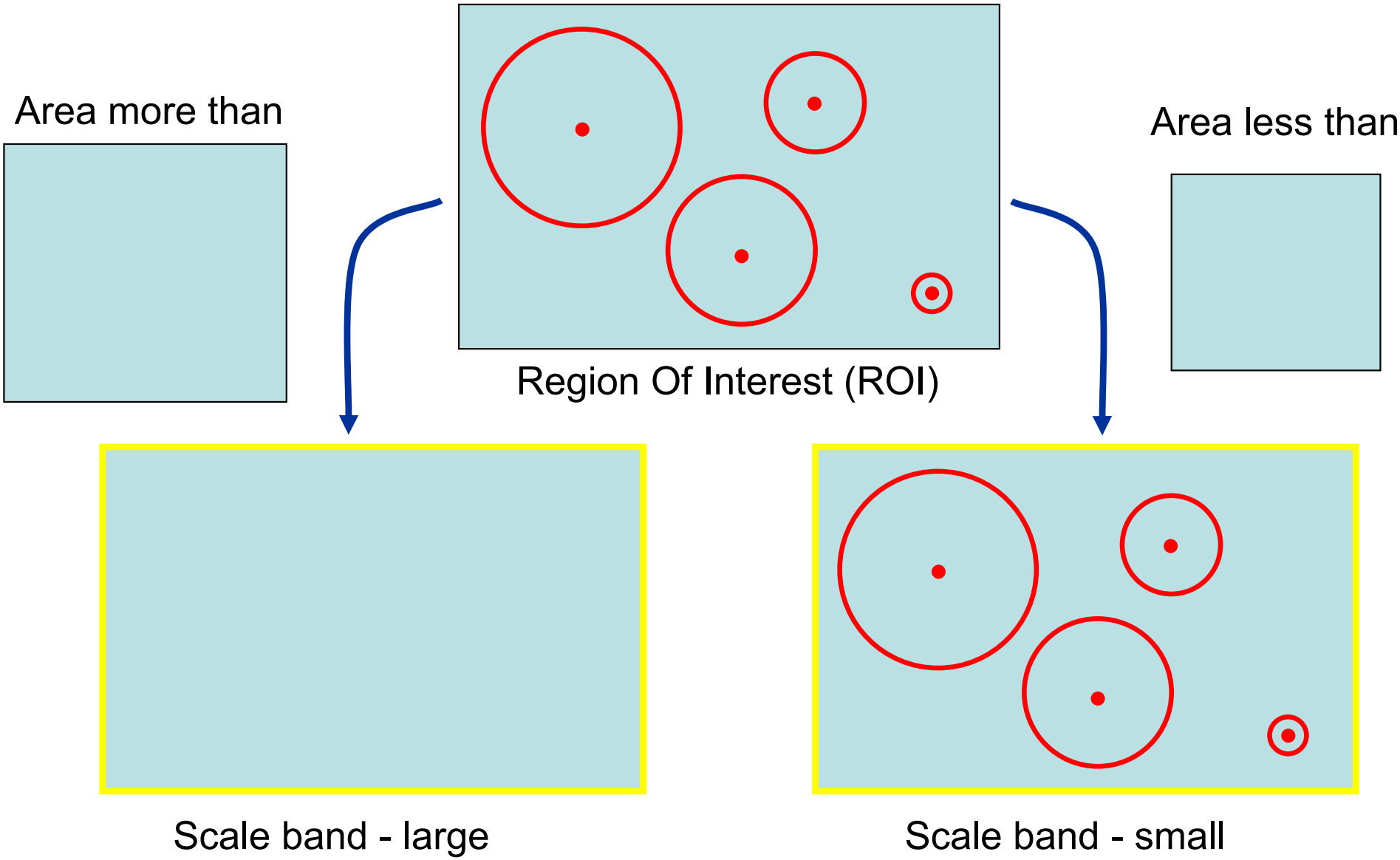
Represented by a sparse vector encoding spatial and scale layout of visual words



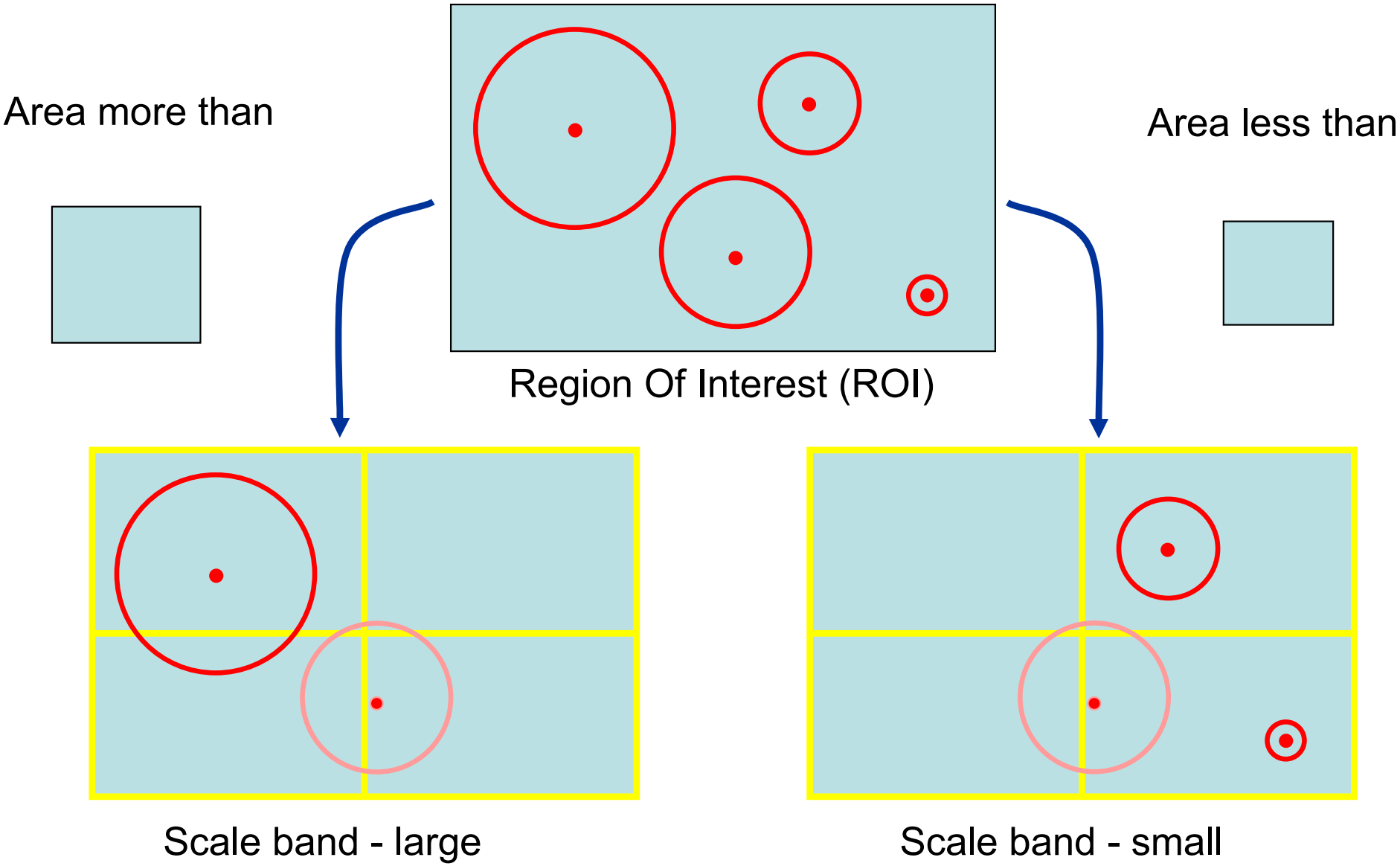
Represented by a histogram of edge directions encoding spatial layout

The distance between the histograms is measured by chi square

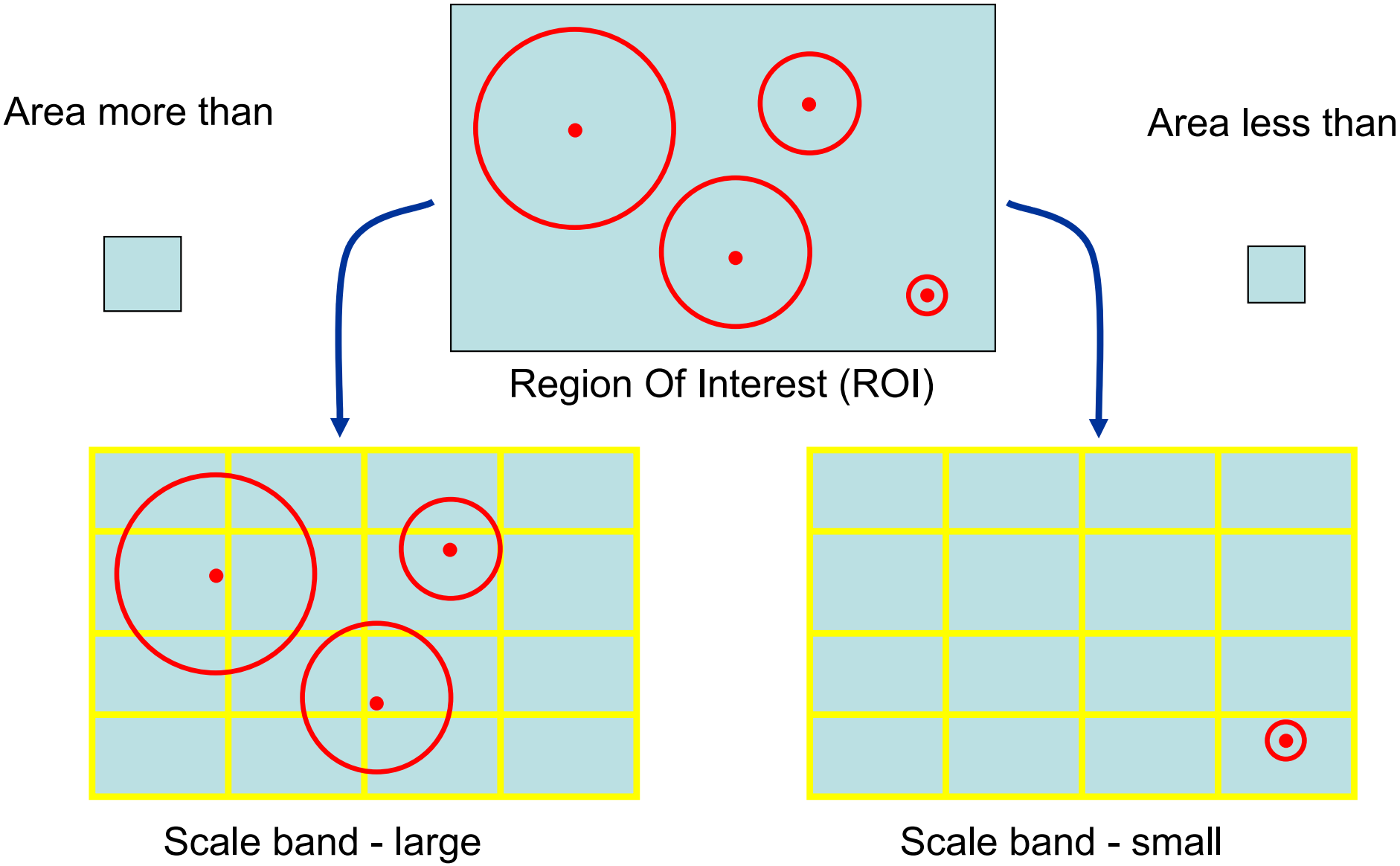
Spatial and Scale Pyramid



Spatial and Scale Pyramid



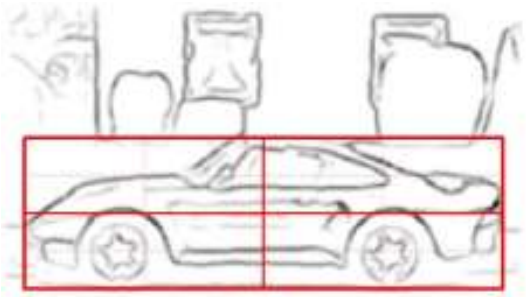
Spatial and Scale Pyramid

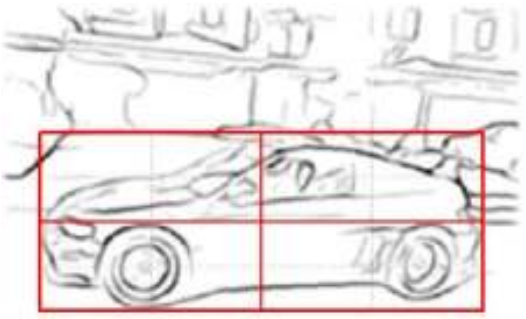


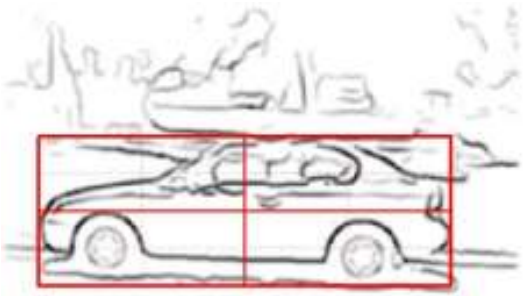
Histograms of Edge Directions

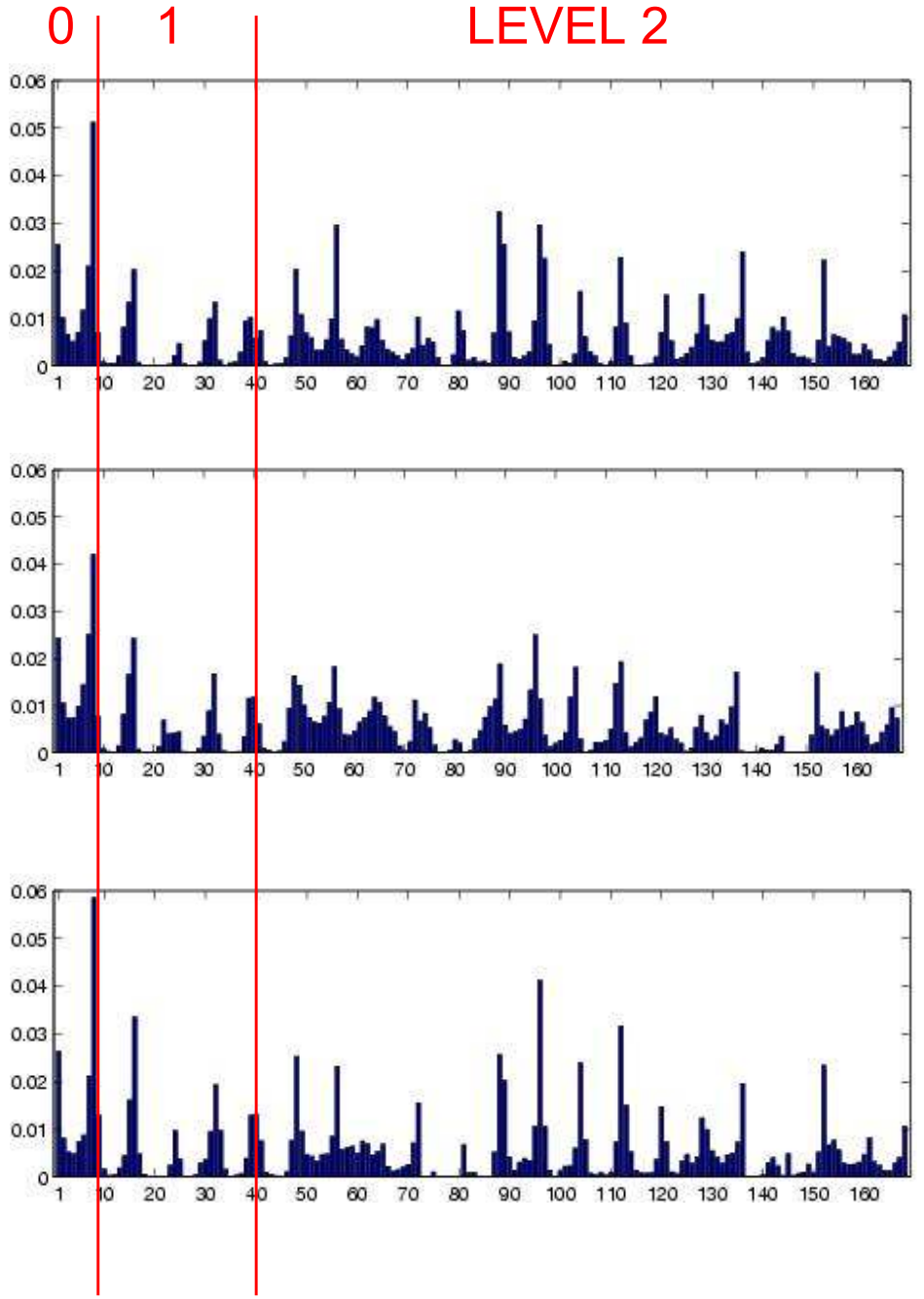
- Berkeley edge detector
- Spatial pyramid of edge directions
  - Using 8 directions
  - Gradient (contrast) flip invariant
  - Soft assignment

Histograms of Edge Directions

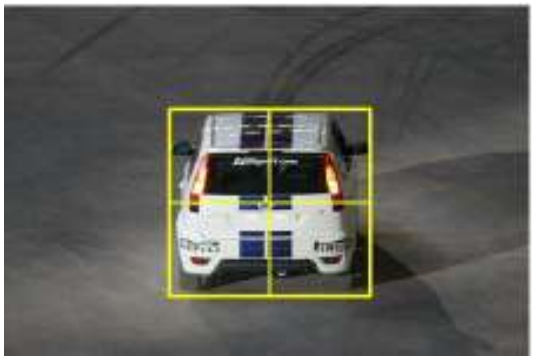


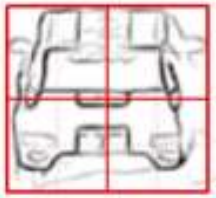


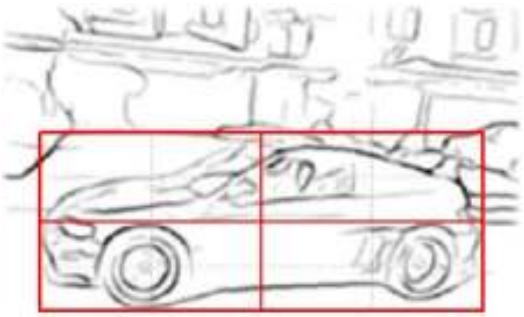


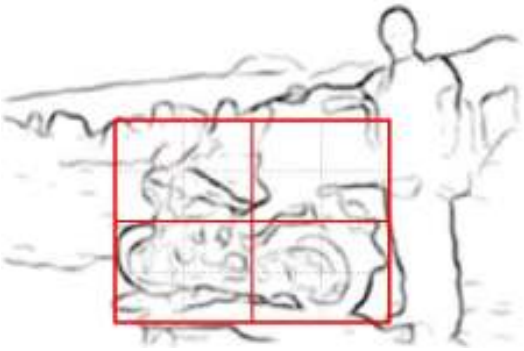


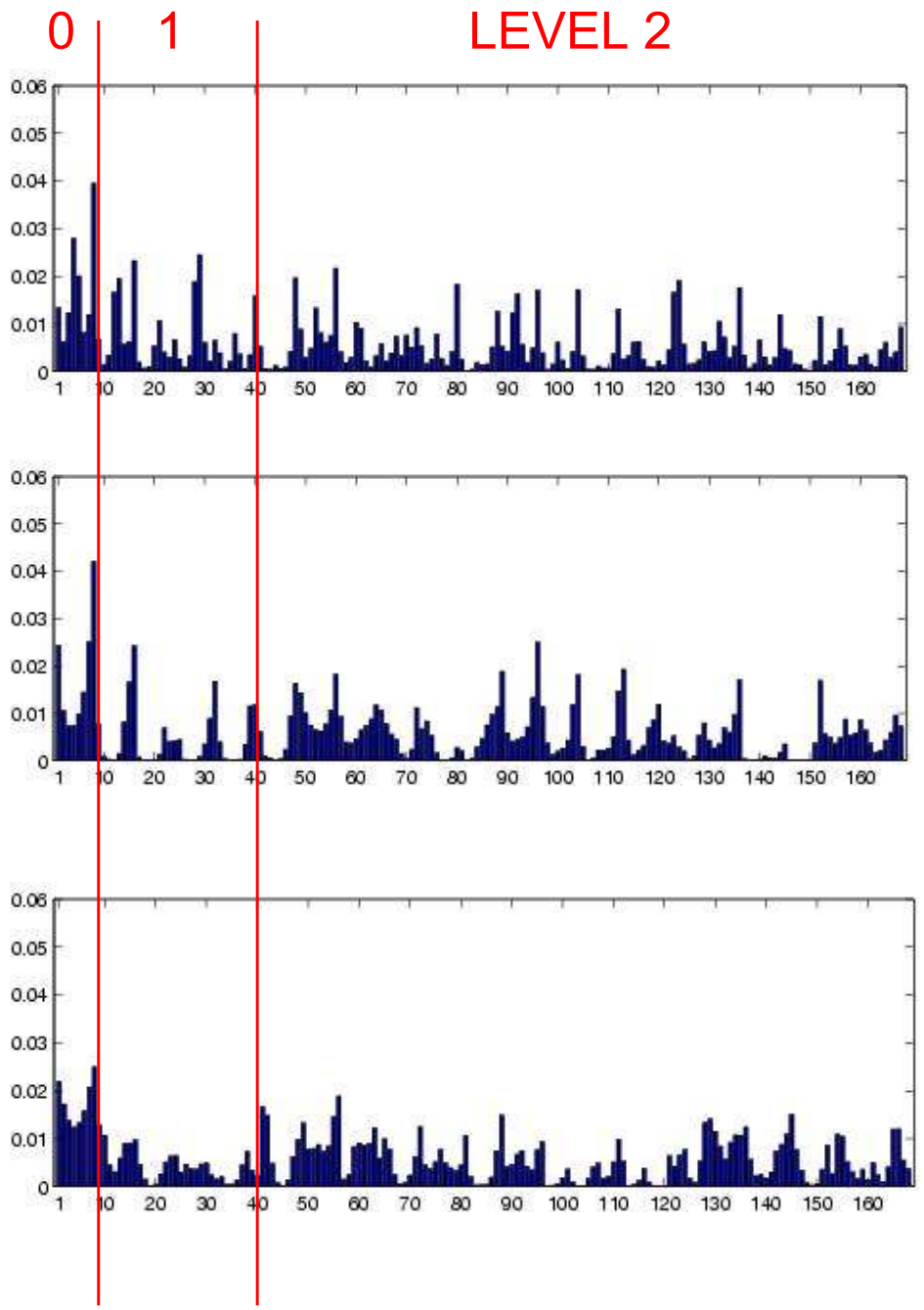
Histograms of Edge Directions











Learning

- A model for each aspect
- Visual words weights
  - Proportional to their relevance to the category (and aspect)
- Relation of features to object spatial layout
- SVM learning
  - Equal weights for all visual words

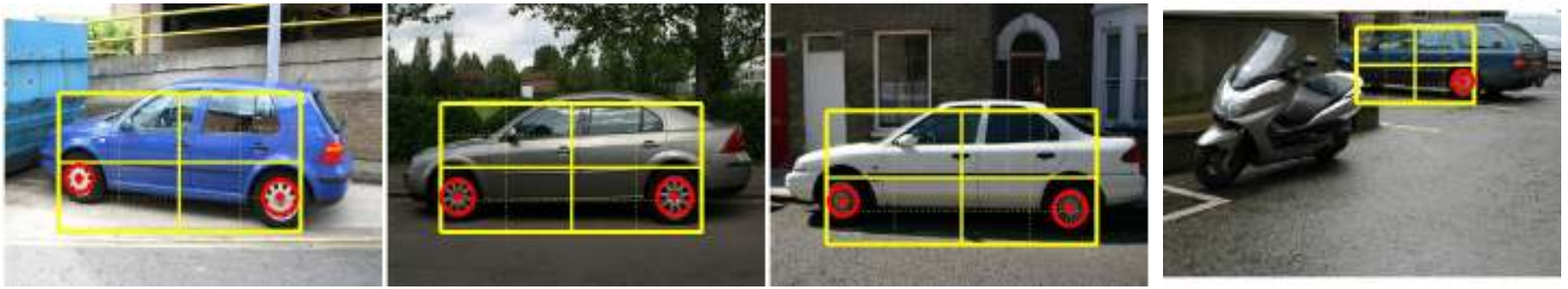
Learning Feature Weights

Vocabulary 3K

$$D(w) \sim \frac{\#\text{class labelled images containing } w}{\#\text{images in database containing } w}$$

G. Dorkó and C. Schmid: Selection of Scale-Invariant Parts for Object Class Recognition, ICCV 03

Learning feature – object relation



Position of visual word with respect to the object

We learn the position of the object with respect to the visual word

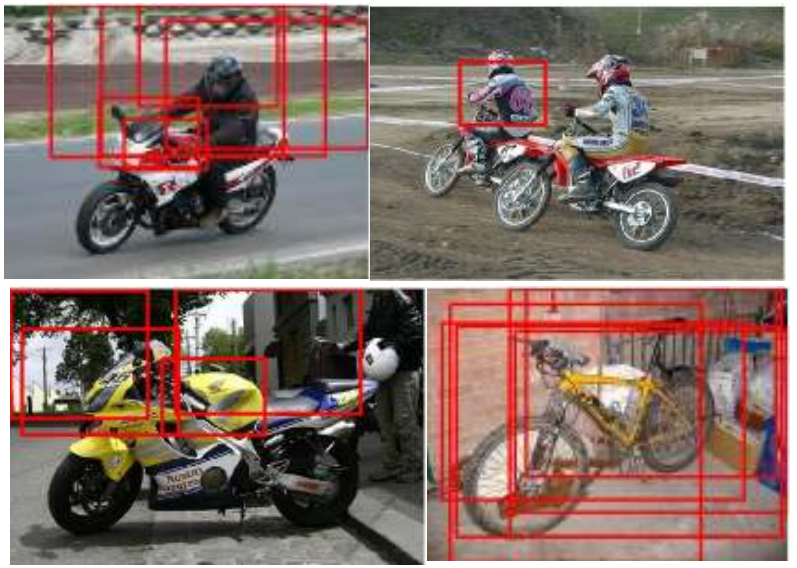
4K feature – object relations for different features and object positions
- features discriminative for the category
- similar relation in many exemplars (large clusters)

Used to generate hypotheses in the detection phase

SVM Classifier

Detection in the training images + ground truth annotation

Positive examples



Negative examples

Sufficient overlap (70%) with ground truth bounding box, multiple instances must originate from different exemplars

Hundreds of examples

SVM

with chi-square kernel

Overlap smaller than 20% with ground truth bounding box

Thousands of examples

- Sparse features (Vocabulary 10K)
- Edge orientation histograms

Four Aspects

Frontal

Unspecified

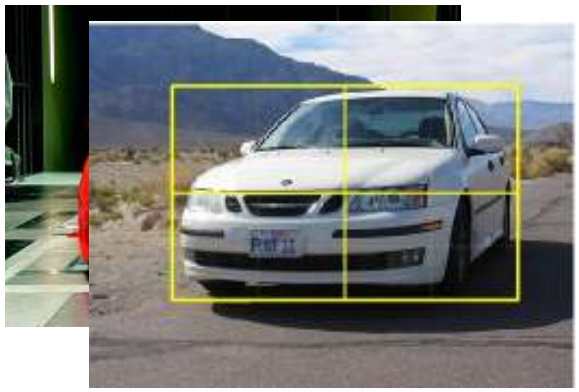
Right

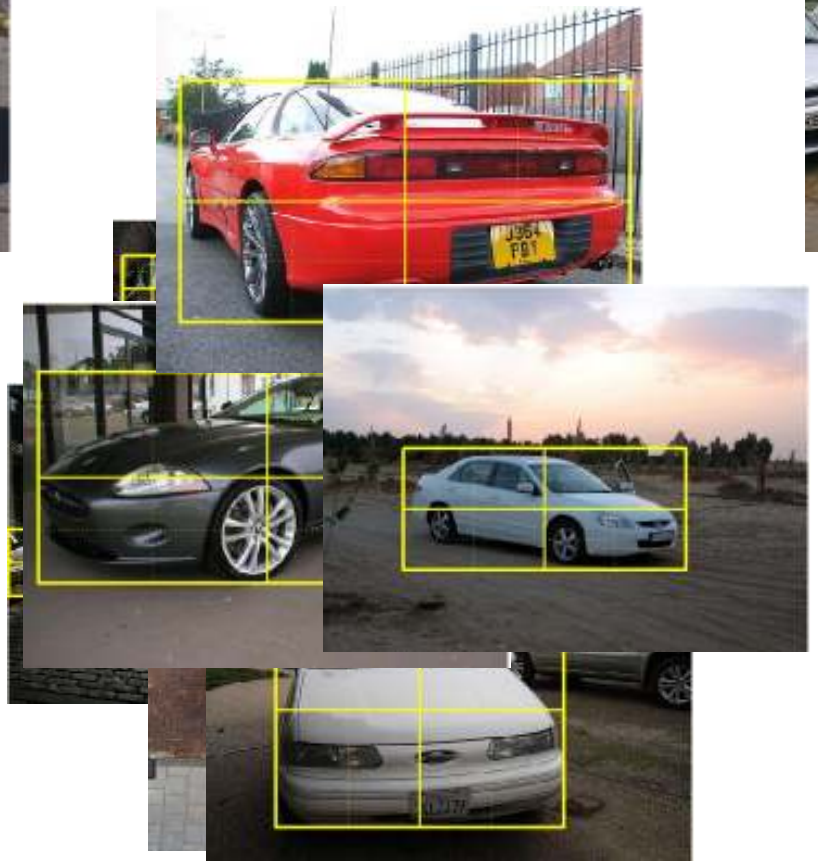
Left

Rear

Four Aspects

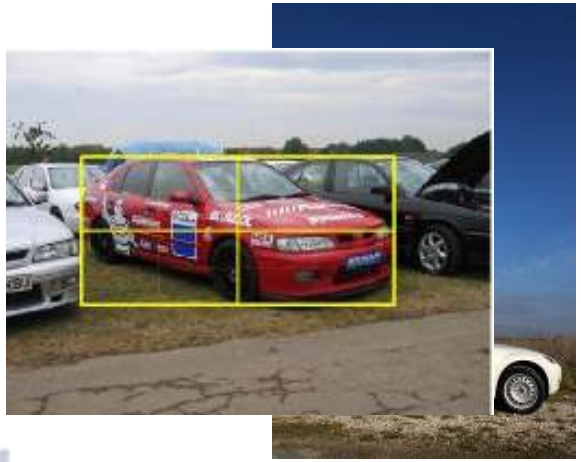
Frontal





Unspecified

Right



Left

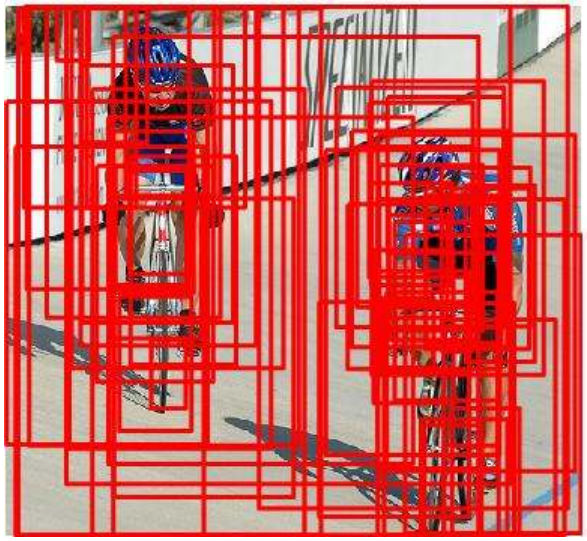
Rear

Detection

For every aspect of each category

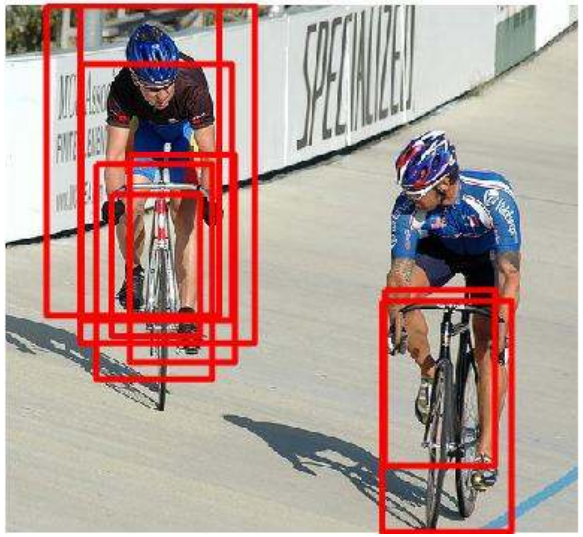
- Hypothesis generation
- Hypothesis scoring
  - Average distance to N closest exemplars of given model
  - Thresholded
- Hypothesis classification
  - SVM on features and edge orientations
- Non-maxima suppression
  - Based on bounding box overlap

Detection



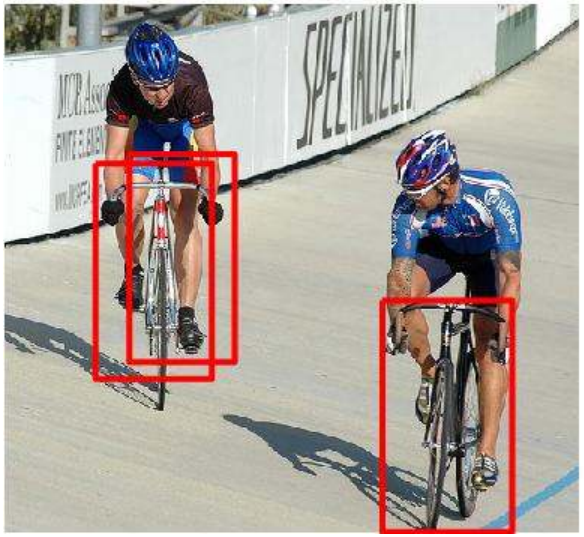
Hypotheses generation

using a single feature to hypothesize a bounding box

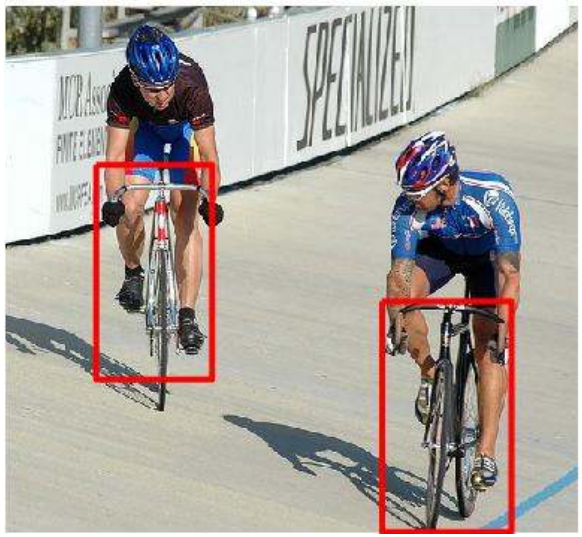


Hypotheses scoring

the exemplar model score is thresholded to prune the hypotheses



SVM classification



Non-maxima supression

Hypothesis Scoring

Test image

ROI

Sparse features

Edge features

Aspect ratio of the bounding box

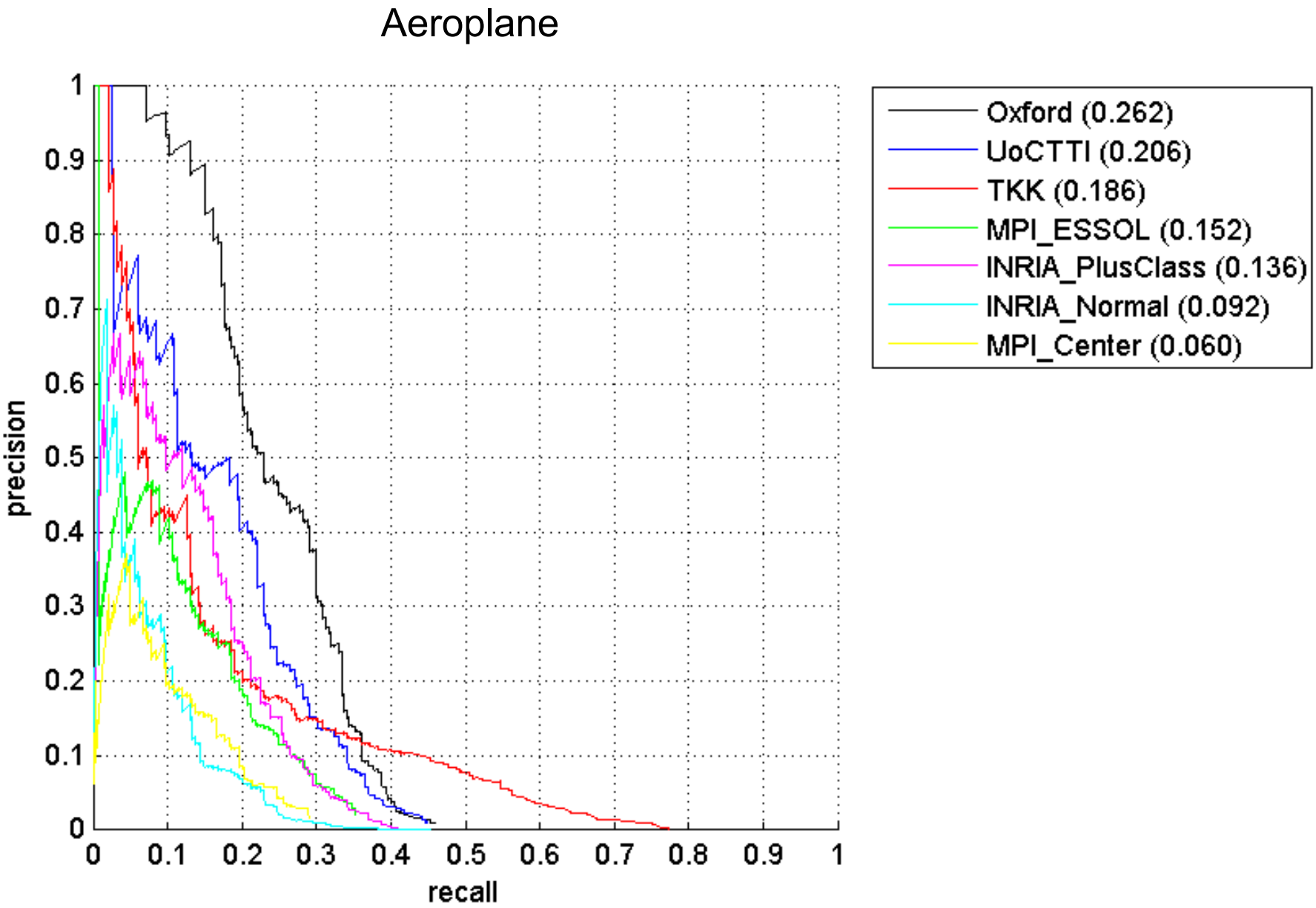
$$C_D = \sum_X (d(X^w, Y^w)) + \alpha\,(d(X^e, Y^e)) + \beta \frac{(A-\mu)^2}{\sigma^2}$$

Summing over 5 nearest exemplars

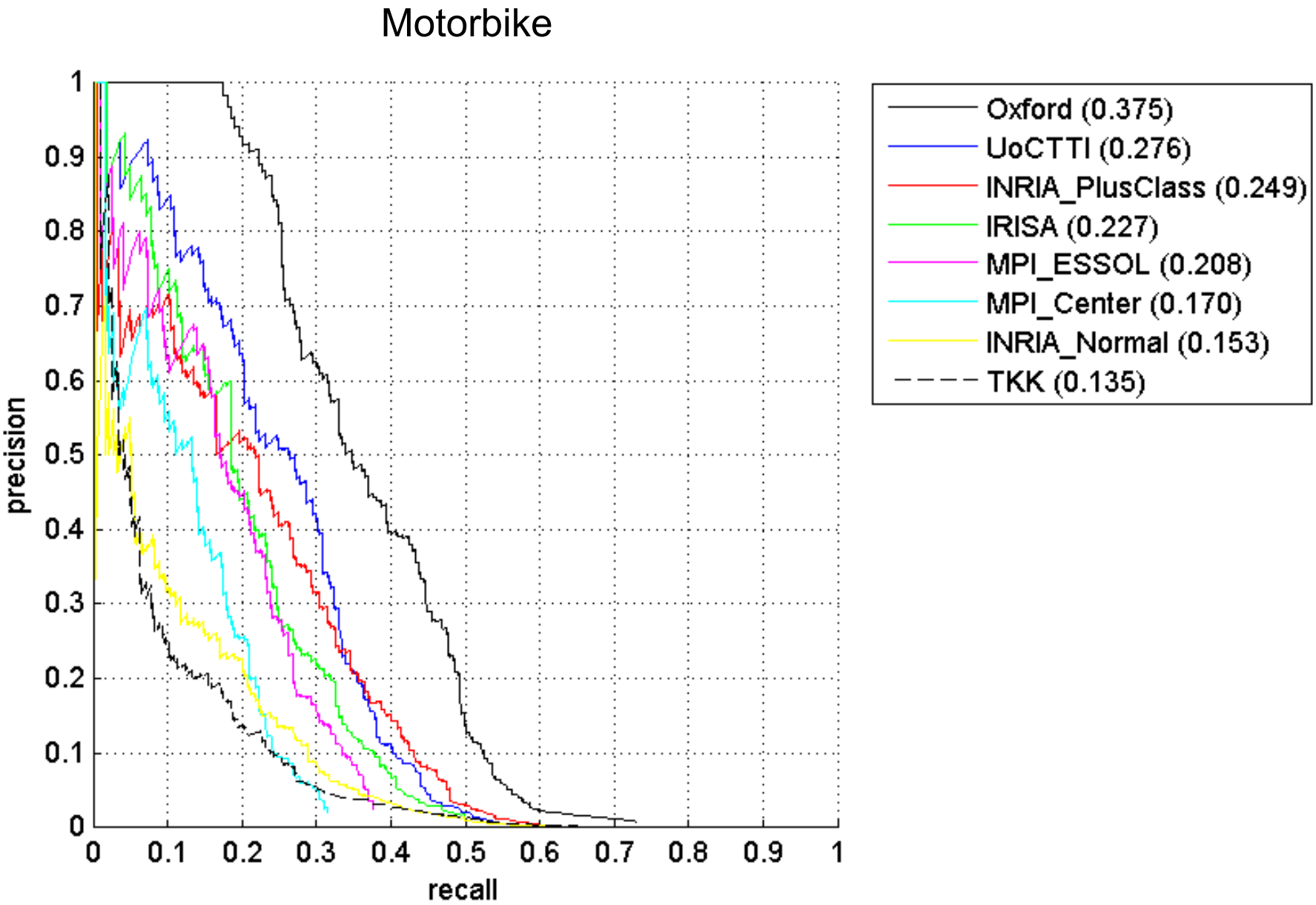
1 2 3 4 5

Exemplars from the training set

Results


Aeroplane
precision
recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Oxford (0.262)
UoCTTI (0.206)
TKK (0.186)
MPI_ESSOL (0.152)
INRIA_PlusClass (0.136)
INRIA_Normal (0.092)
MPI_Center (0.060)

Results


Motorbike
precision
recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Oxford (0.375)
UoCTTI (0.276)
INRIA_PlusClass (0.249)
IRISA (0.227)
MPI_ESSOL (0.208)
MPI_Center (0.170)
INRIA_Normal (0.153)
TKK (0.135)