Incremental Learning of Object Detectors Using a Visual Shape Alphabet

A. Opelt, A. Pinz & A. Zisserman

CVPR '06

Presented by Medha Bhargava*

* Several slides adapted from authors' presentation, CVPR '06

OUTLINE

- Motivation, Goals & Overview of the approach

- Learning the model

  Stage 1: the visual alphabet of shape

  Stage 2: jointly/incrementally learned detectors

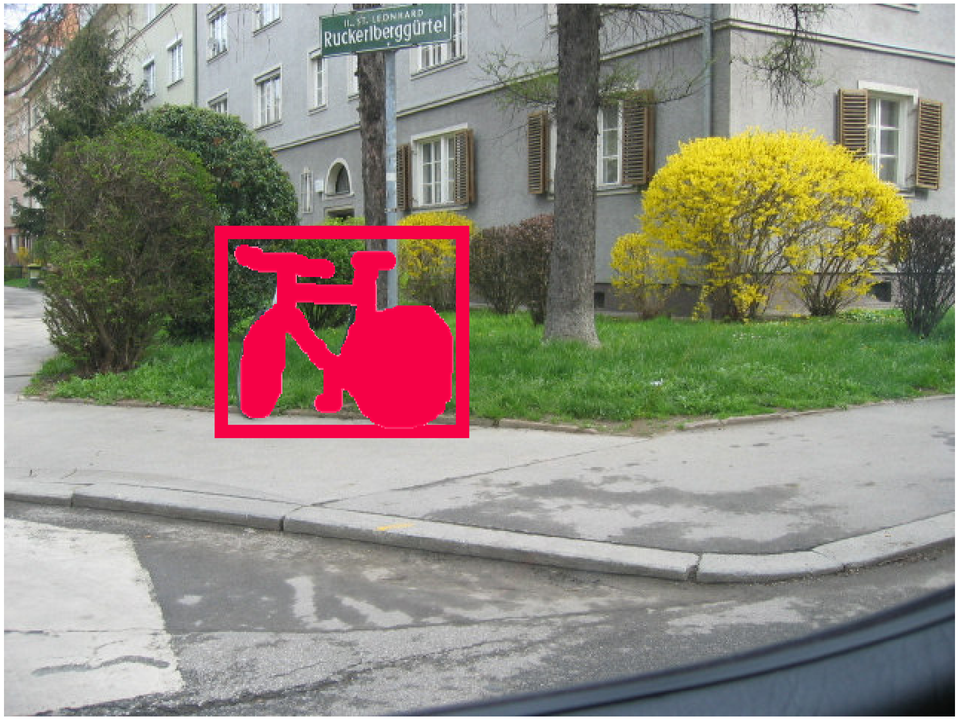
- Detection

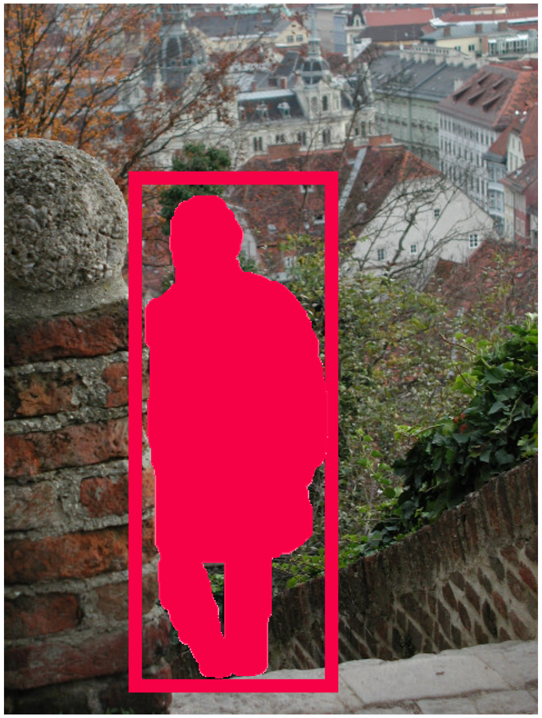
- Invariances (scale, rotation, viewpoint)

- Experiments and Results

- Summary

MOTIVATION





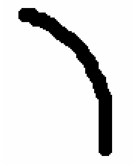
Class: **Bicycle** Class: **Person**

MOTIVATION

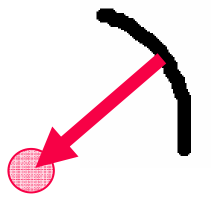
Classification + Localization + rough Segmentation

Proposed approach uses

Alphabet of Shape



+ Geometry



Goals

- Object *detection*
- Localization and crude segmentation
- Learning new models from previously trained detectors
  - Incremental learning
  - Sharing of model features ....underlying theme
- Sublinear learning complexity

New Concepts

- Boundary fragment based shape alphabet
- Incremental joint-AdaBoost algorithm

Boundary-Fragment-Model 1/2

- **Learning** the BFM

– *Training* set
  – Object delineated by bounding box
    – 20 images/class

– *Validation* set
  – Labeled as +ve/-ve image
  – Object centroid marked
    – 50 images/class -- 25 +ve, 25 –ve

– A candidate boundary fragment ***MUST***
  – match edge chains in +ve set
  – have good localization of the centroid in +ve set

Boundary-Fragment-Model 2/2

- **Scoring** a Boundary Fragment

$$\boxed{C(\gamma_i) = c_{match}(\gamma_i)\; c_{loc}(\gamma_i)}$$

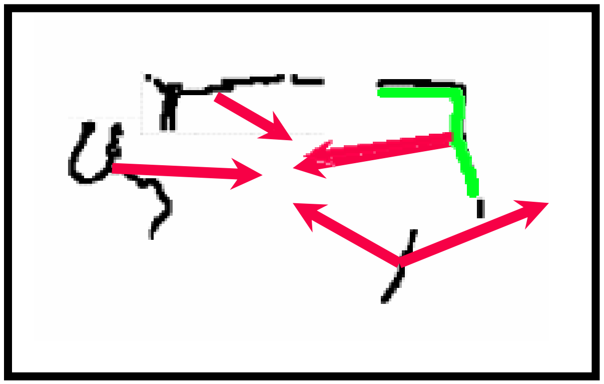
  - $c_{match}(\gamma_i)$: Ratio of cumulative Chamfer matching costs of fragment to edge chains in validation images

$$c_{match}(\gamma_i) = \frac{\sum_{i=1}^{L^+} distance(\gamma_i, V_i^+) / L^+}{\sum_{i=1}^{L^-} distance(\gamma_i, V_i^-) / L^-}$$

  - $c_{loc}(\gamma_i)$: pixel distance between true centroid and predicted centroid, averaged over +ve validation set

Overview of the Approach 1/2

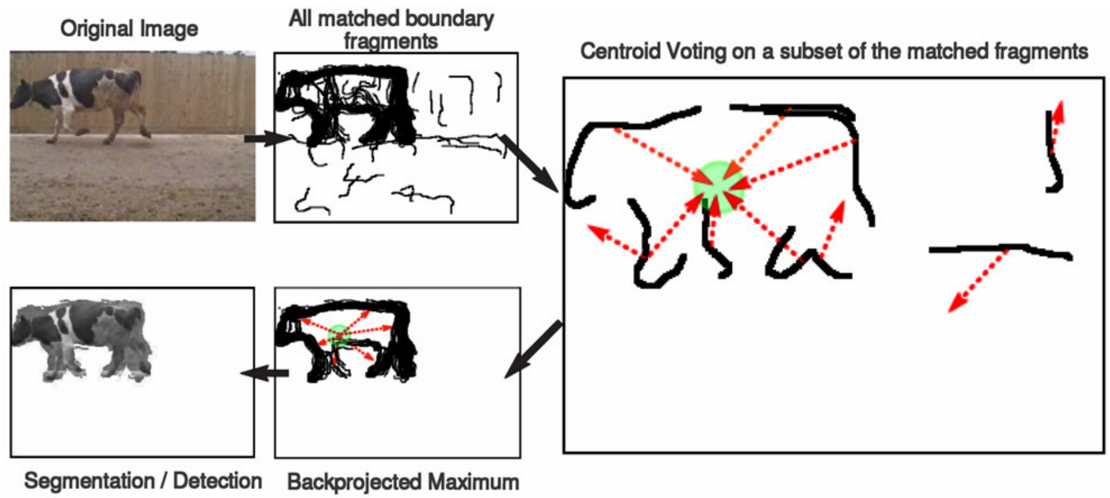
The Boundary-Fragment-Model



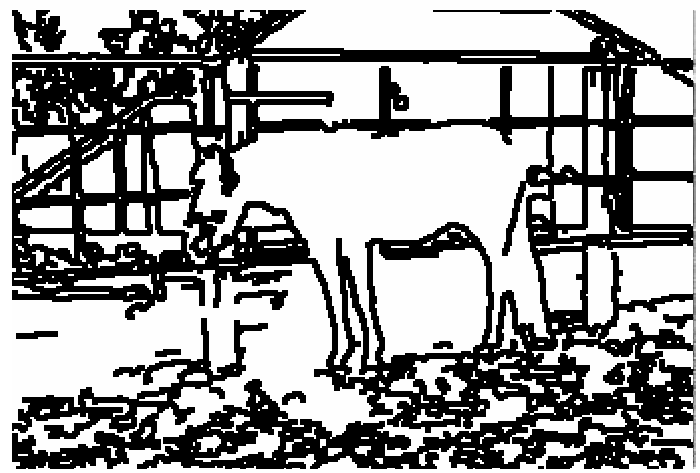
first proposed in
Opelt, Pinz and Zisserman ECCV 2006

Geometric model related to
Leibe, Leonardis and Schiele
(Workshop at ECCV 2004)

Similar model proposed by Shotton et al.
(ICCV 2005)

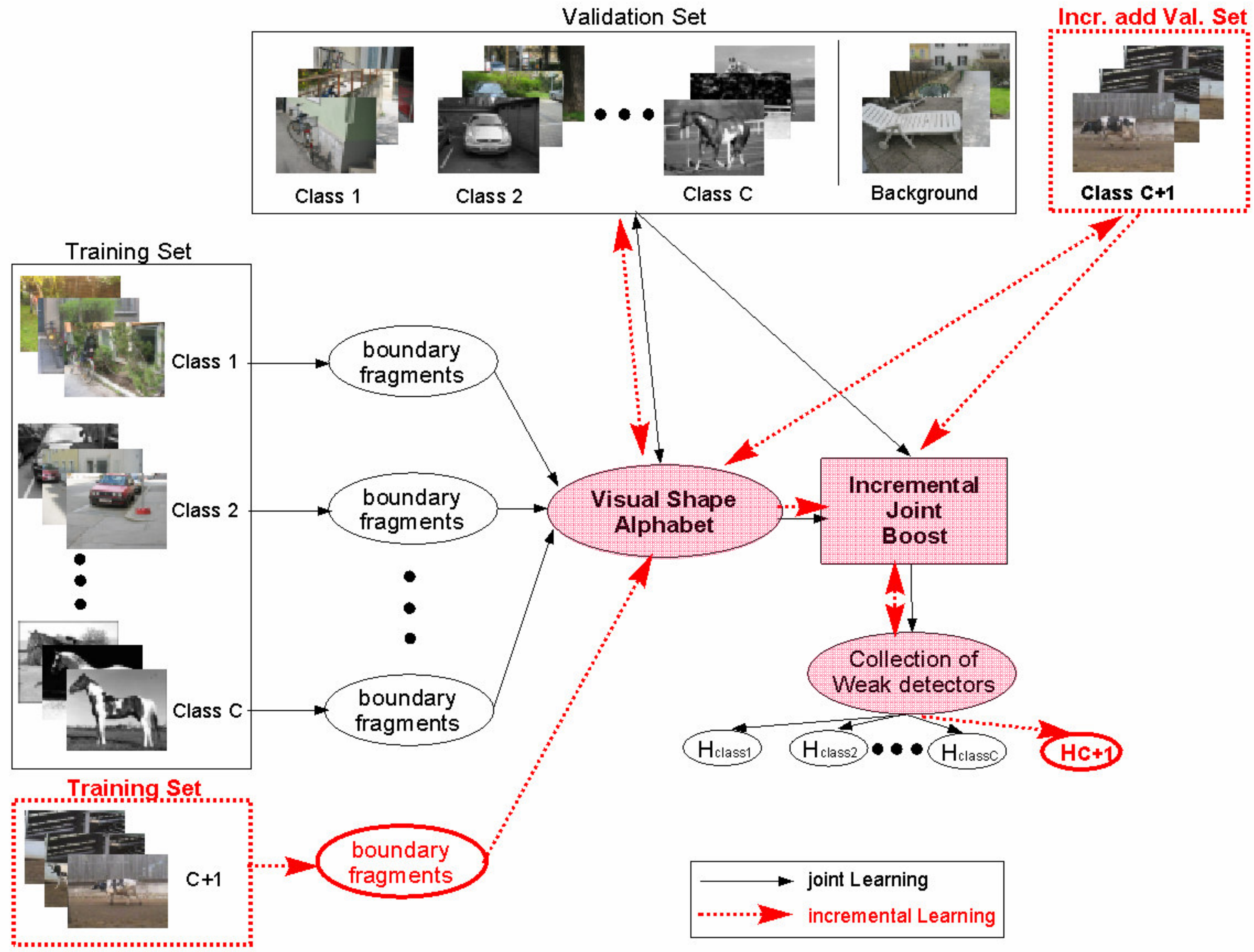


More categories →

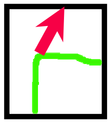


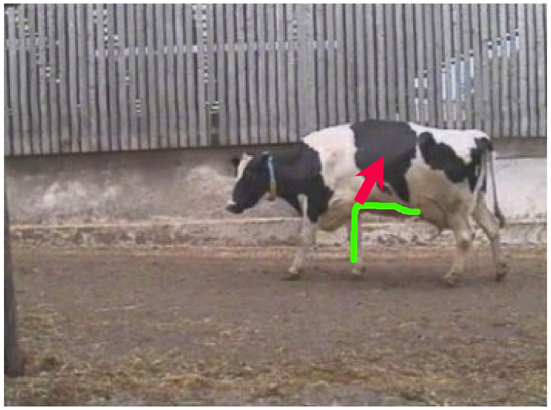
Two possibilities: Learning JOINTLY or INCREMENTALLY

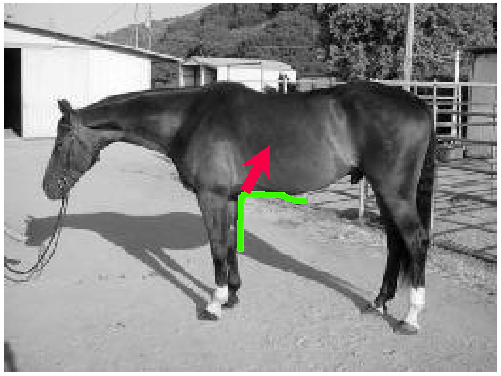
OVERVIEW OF THE APPROACH 2/2



Stage 1: The Visual Alphabet of Shape 1/3







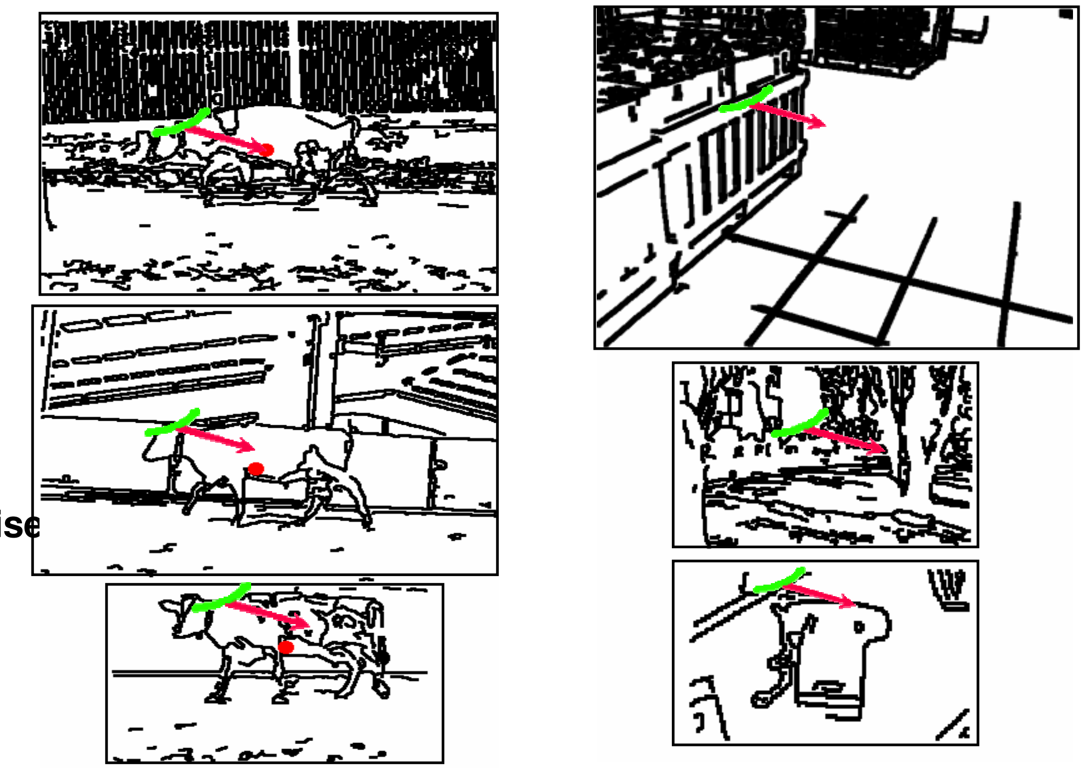
Stage 1: The Visual Alphabet of Shape 1/3

For each class Ci

For i=1:N trials

1. **Grow candidate fragment in training images around random starting point i**
2. **Evaluate the fragment at each step on the validation set of the category**
   → calculate costs
3. **If the fragments costs are above a certain threshold discard this fragment, otherwise go on with step 4.**

Validation set for the category: Cow



COSTS

STAGE 1: THE VISUAL ALPHABET OF SHAPE 2/3

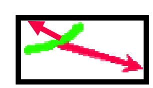
For each class Ci

For i=1:N trials

1. **…**
2. **…**
3. **…**
4. **Evaluate the boundary fragment on the validation sets of the other categories.**
5. **Add this fragment with costs on all categories and the geometric information to the alphabet**

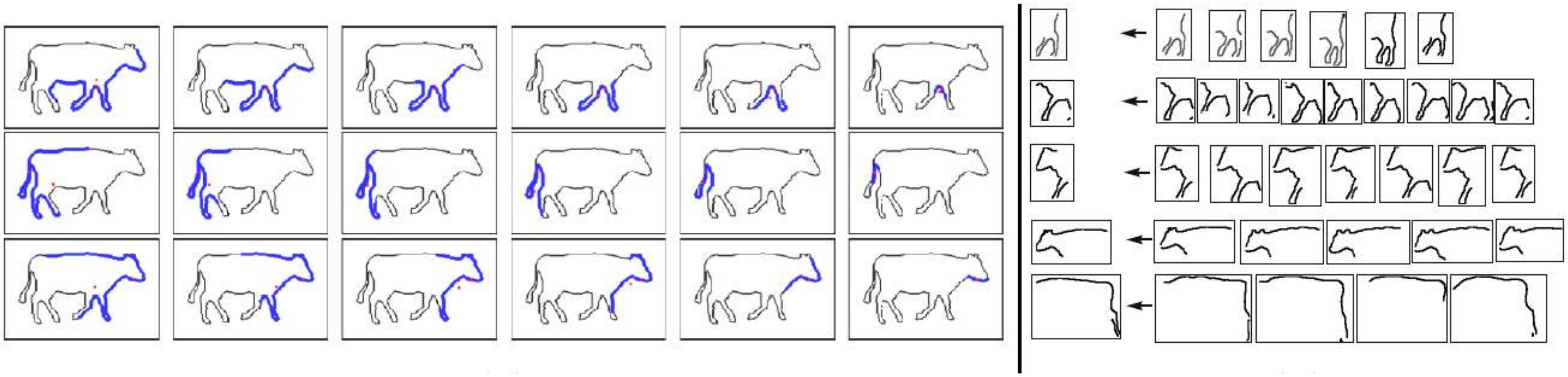
Horses **Faces**

Update centroid vectors

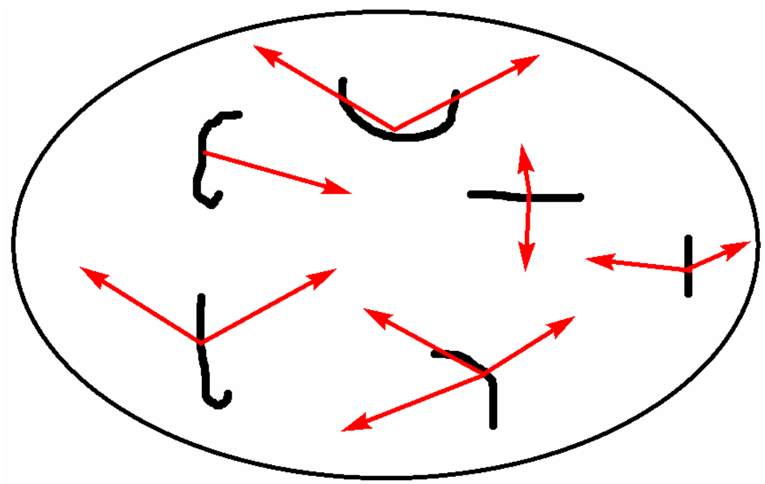


COSTS

Stage 1: The Visual Alphabet of Shape 3/3



Clustering shape



Visual Shape Alphabet

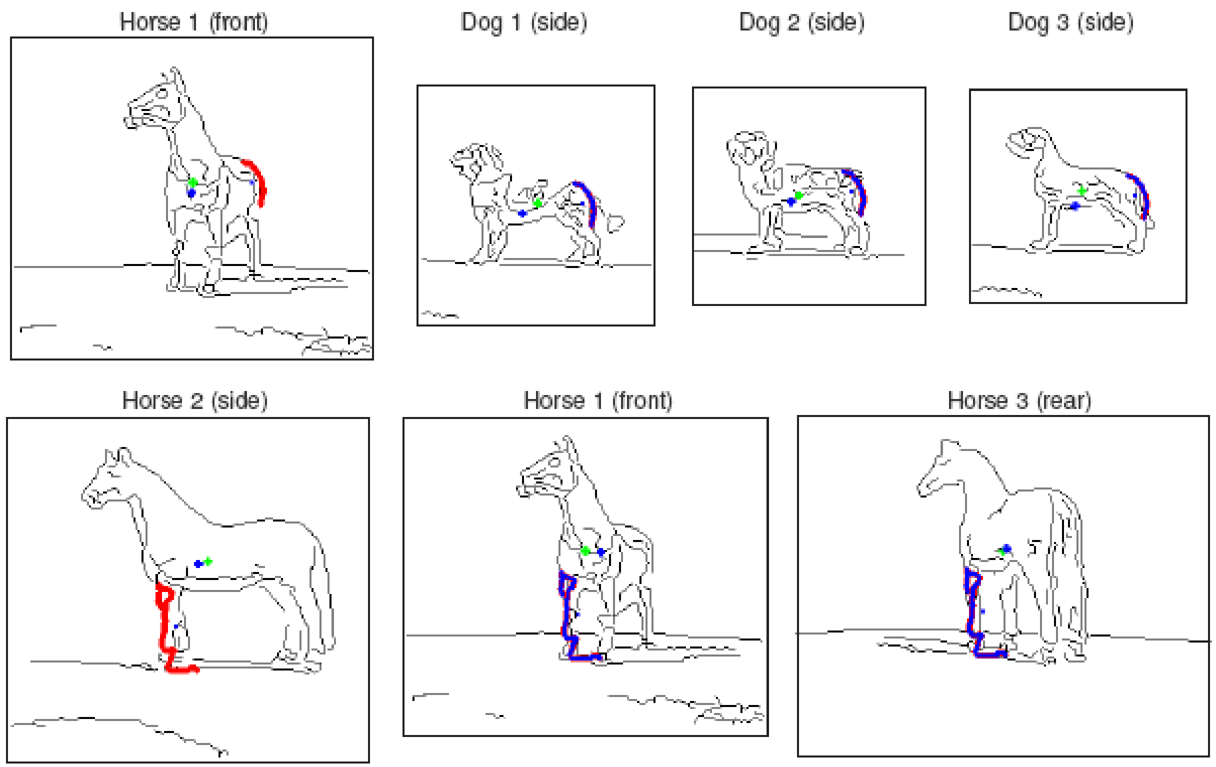
Incremental Learning

- **Enlarging the alphabet codebook**
  1. Add more boundary fragments
  2. Allow a single fragment to vote for additional object centroids
- **Sharing to build**
  1. If fragments from different categories match, update centroid info
  2. Evaluation of fragment on –ve validation set
  3. Granting additional voting privileges

Example of Sharing

Over Classes

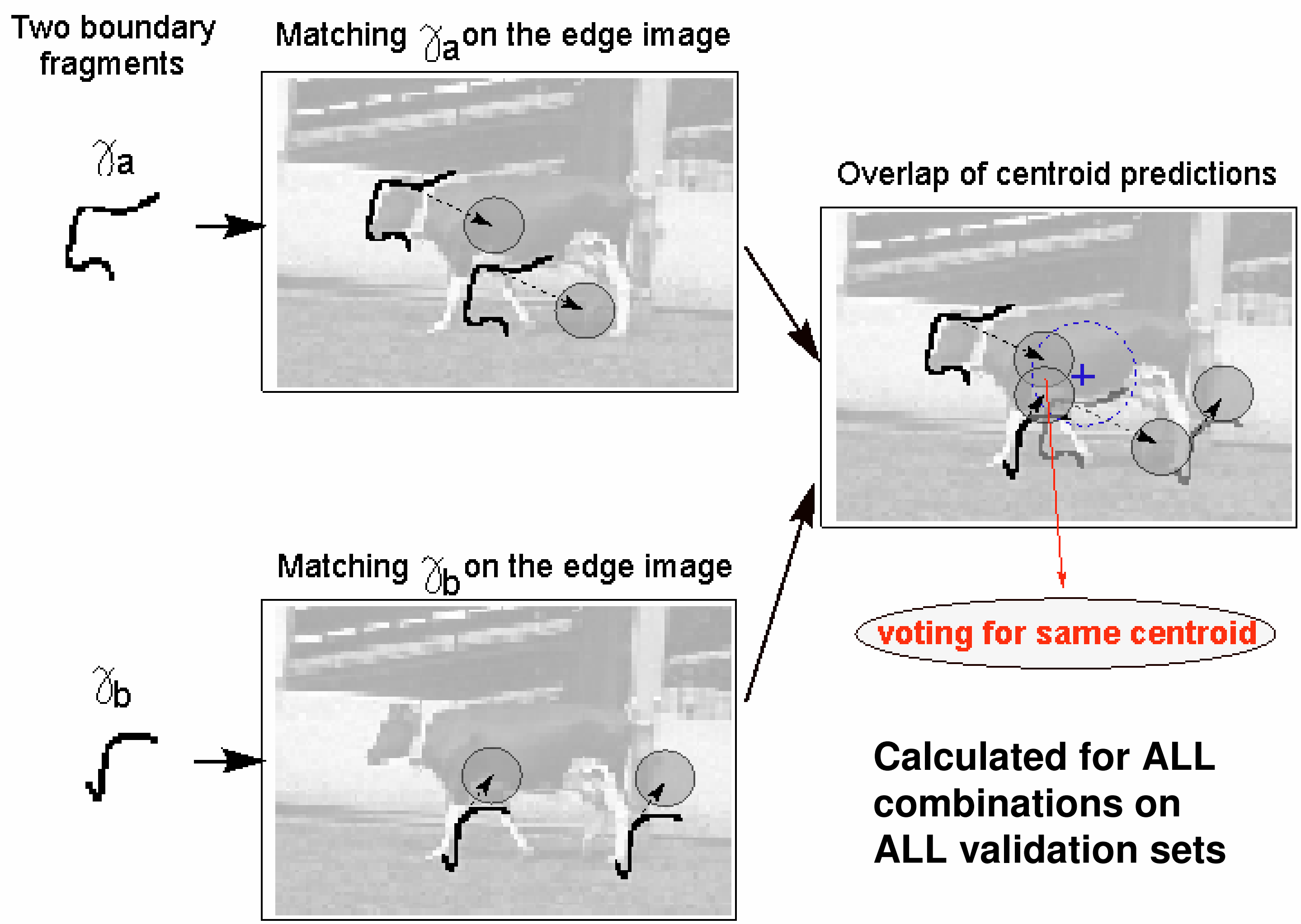
Over Aspects



Benefit: One class/aspect can build on what has been learnt from another

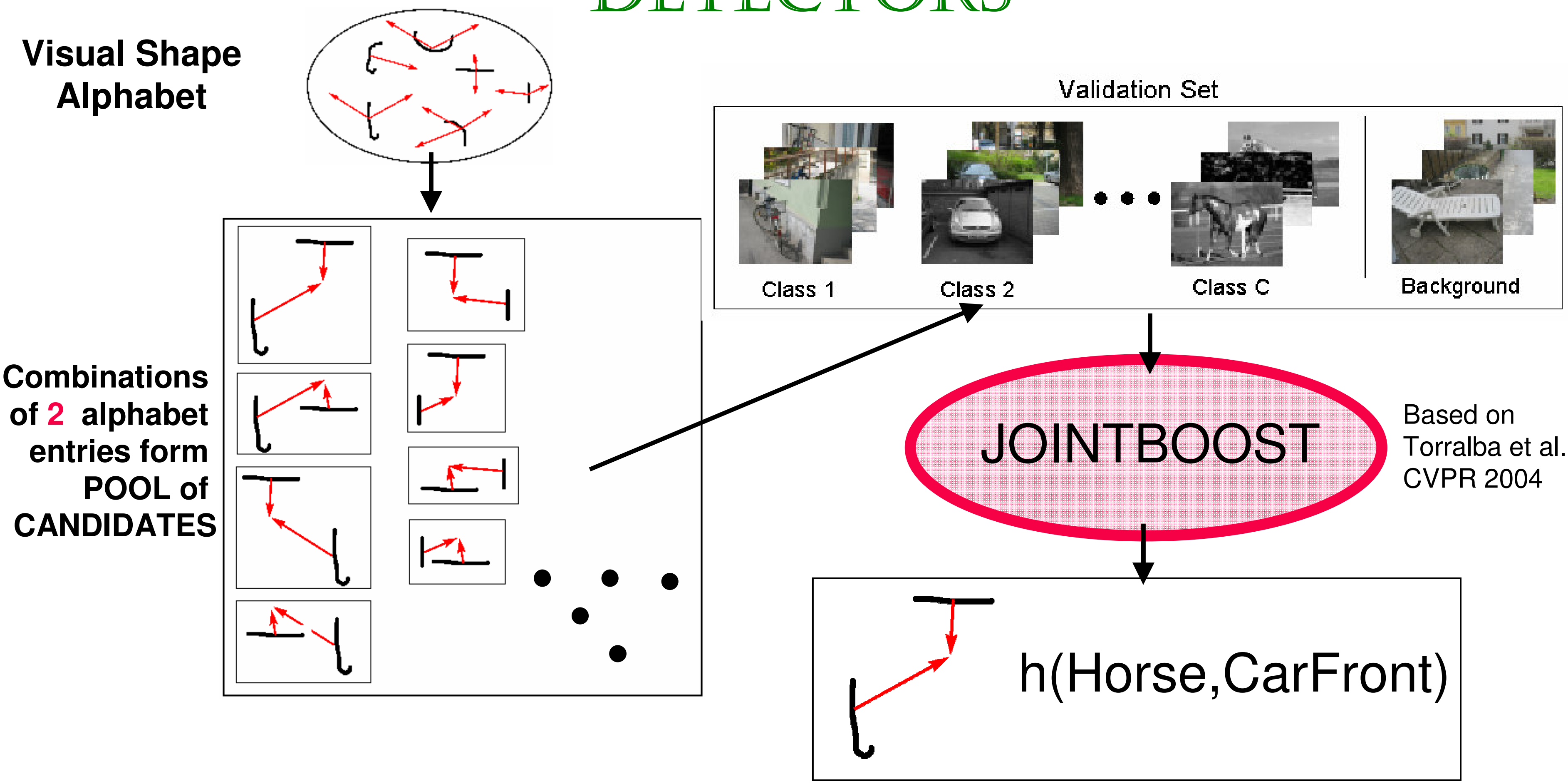
Stage 2: Weak Detector Candidates

Combinations of 2 boundary fragments as pool for learning



Calculated for ALL combinations on ALL validation sets

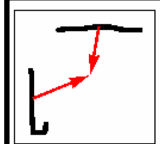
STAGE 2: JOINTLY LEARNED DETECTORS

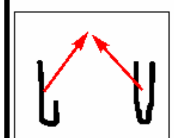


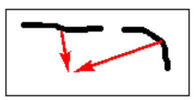
Each candidate that is valid (boosting) for at least one category is added to
Collection of weak detectors

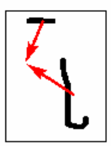
Stage 2: Incrementally learned detectors

Knowledge:
Collection of weak detectors
e.g. CarsSide, Horses, Bicycles

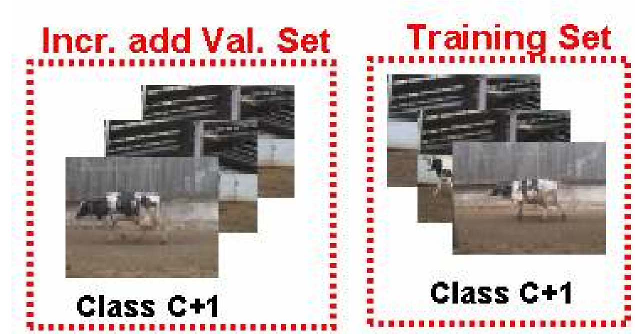

h(Horse,CarFront)


h(Horse)

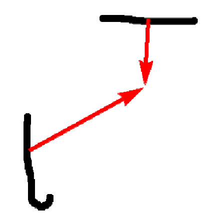

h(Horse)


h(Horse,Bicycle)

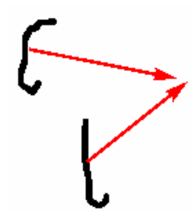
•
•
•



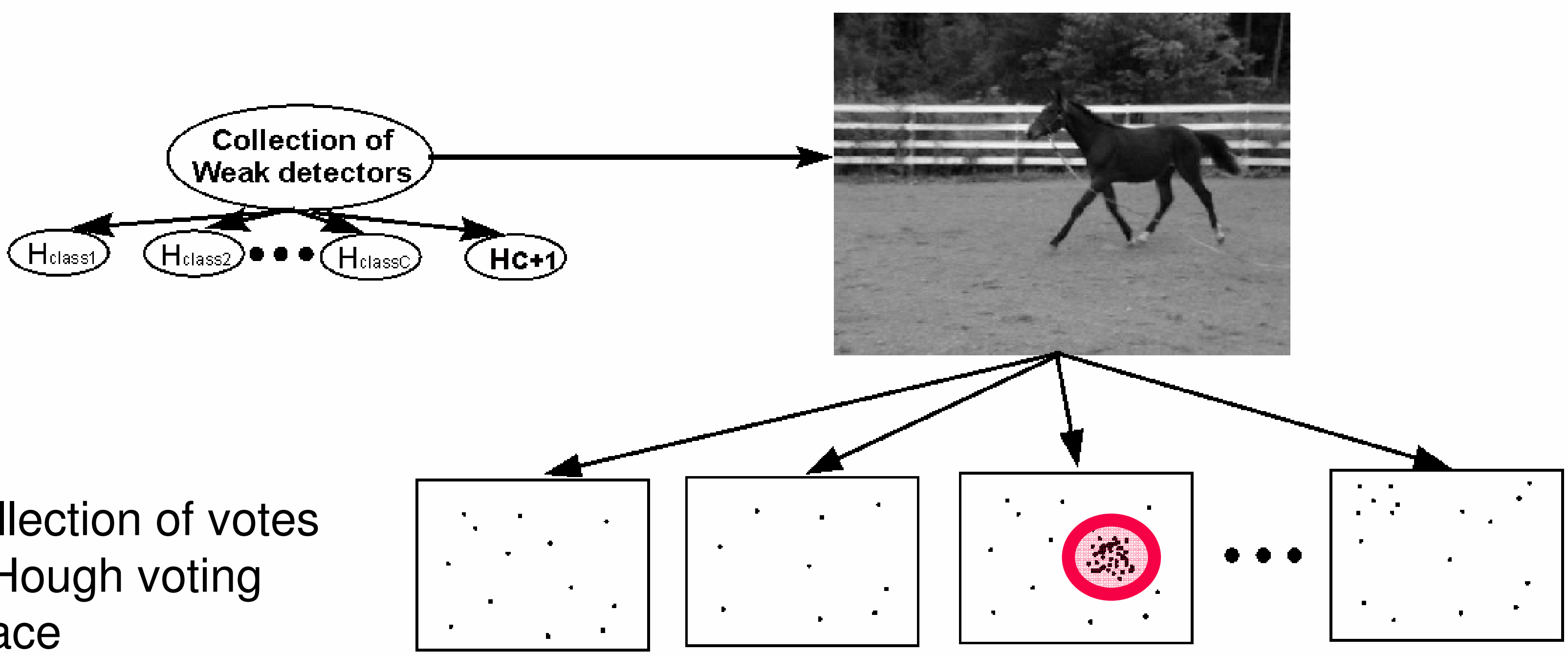
1. Update existing knowledge (share)


h(Horse,CarFront,**Cow**)

2. Add new weak detectors (discriminative)


h(**Cow**)

DETECTION FOR THE MULTICLASS CASE

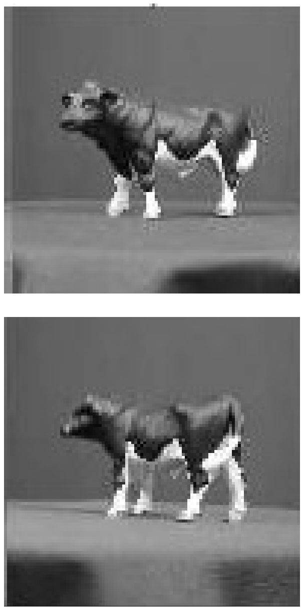


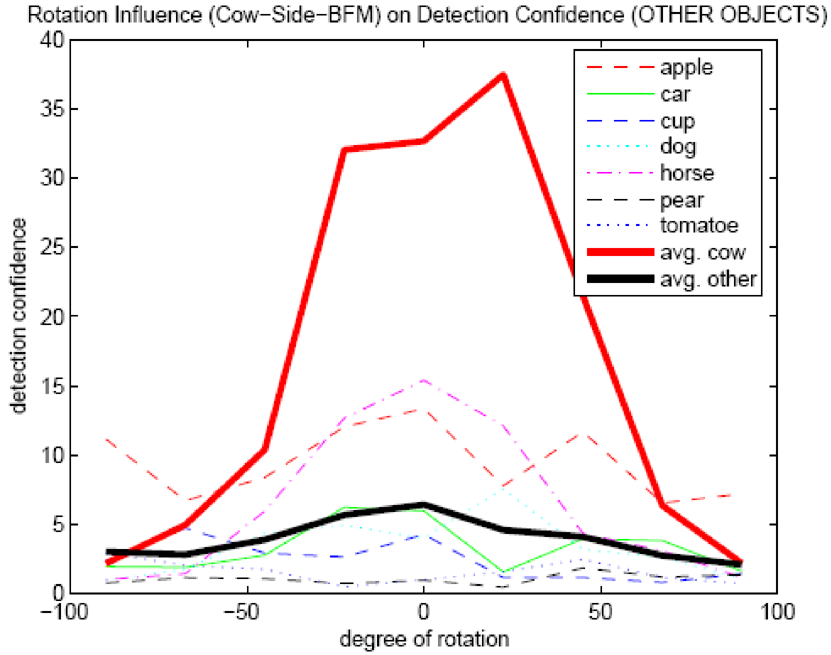
Collection of votes in Hough voting space

Mode above threshold
→ Detection Class 3

INVARIANCES

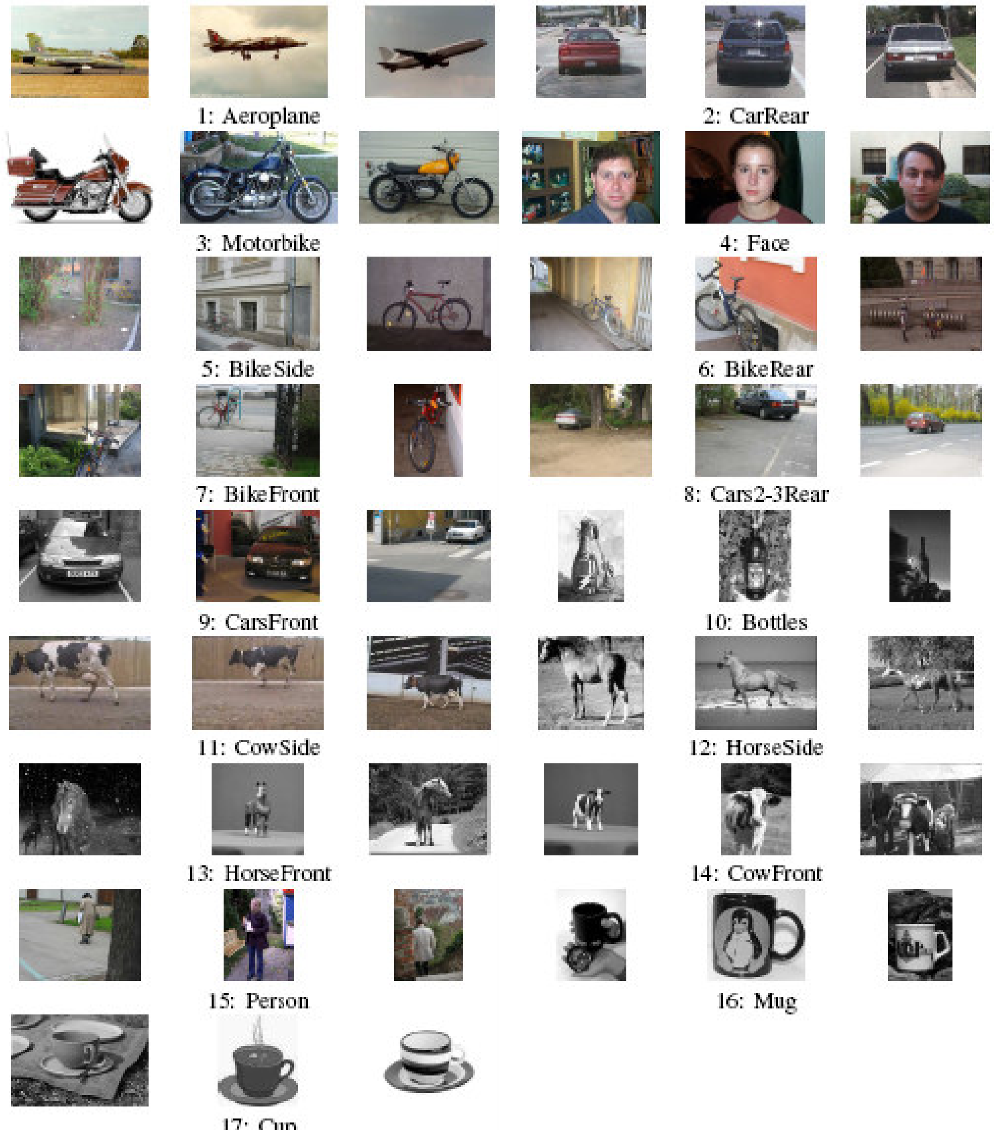
- **Translation →** Mode search in the Hough voting space
- **In-plane Rotation →** Hough voting with oriented model
- **Scale invariance →** 3D-Balloon-Meanshift-Mode-Est.
- **Viewpoint →**





EXPERIMENTS

MULTICLASS DATASET

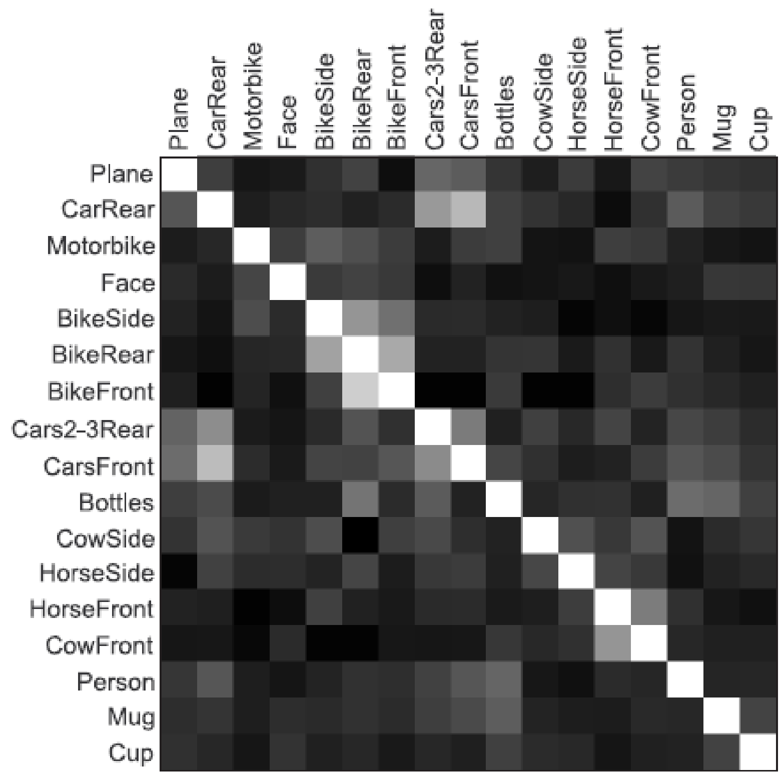


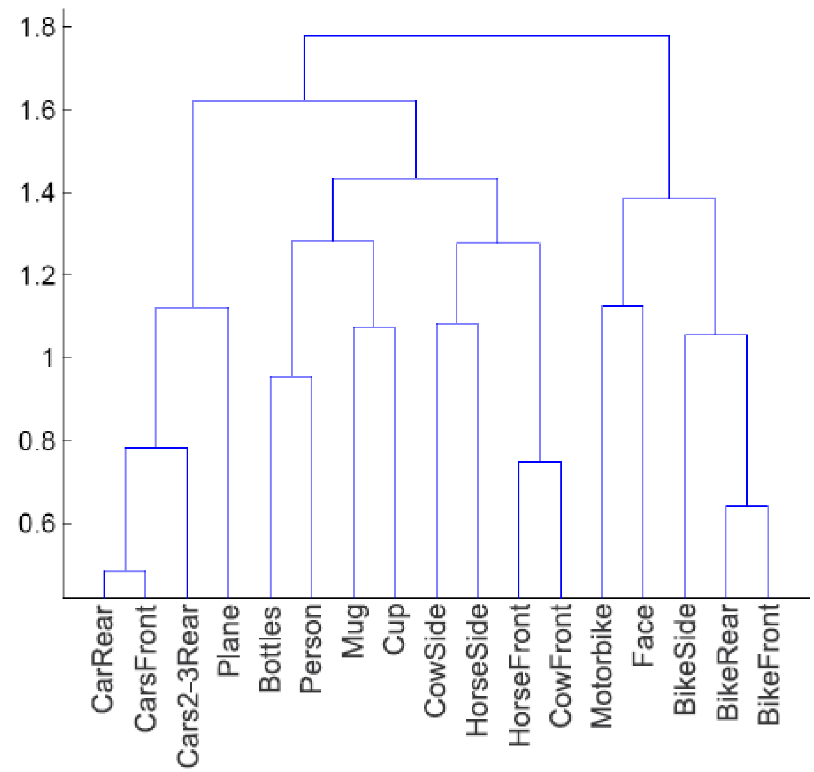
Collection of
17 categories,
From Caltech, Graz02,
Magee, ImageGoogle
(avalilable at:
http://emt.tugraz.at/~pinz/data)

Different numbers of
training images
per category (10-100)

Different aspects and
similar categories

RESULTS 1/6

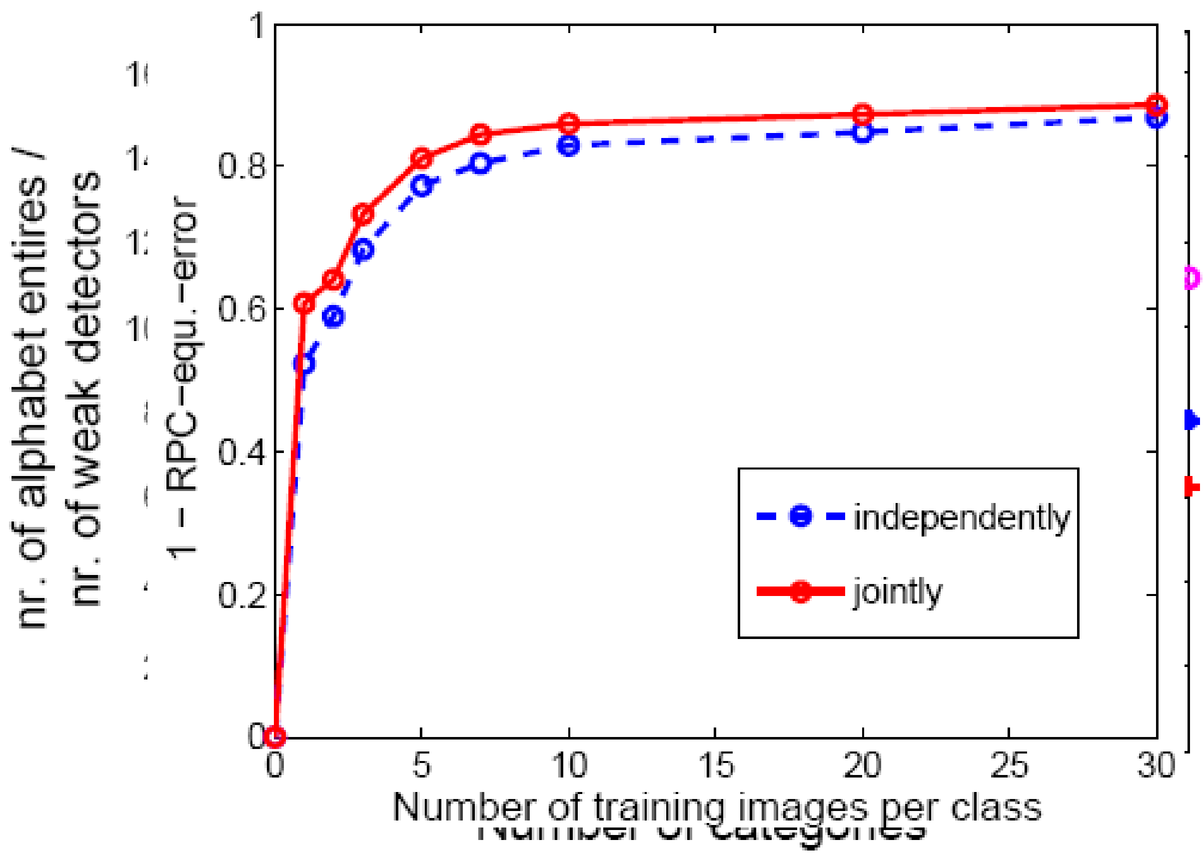




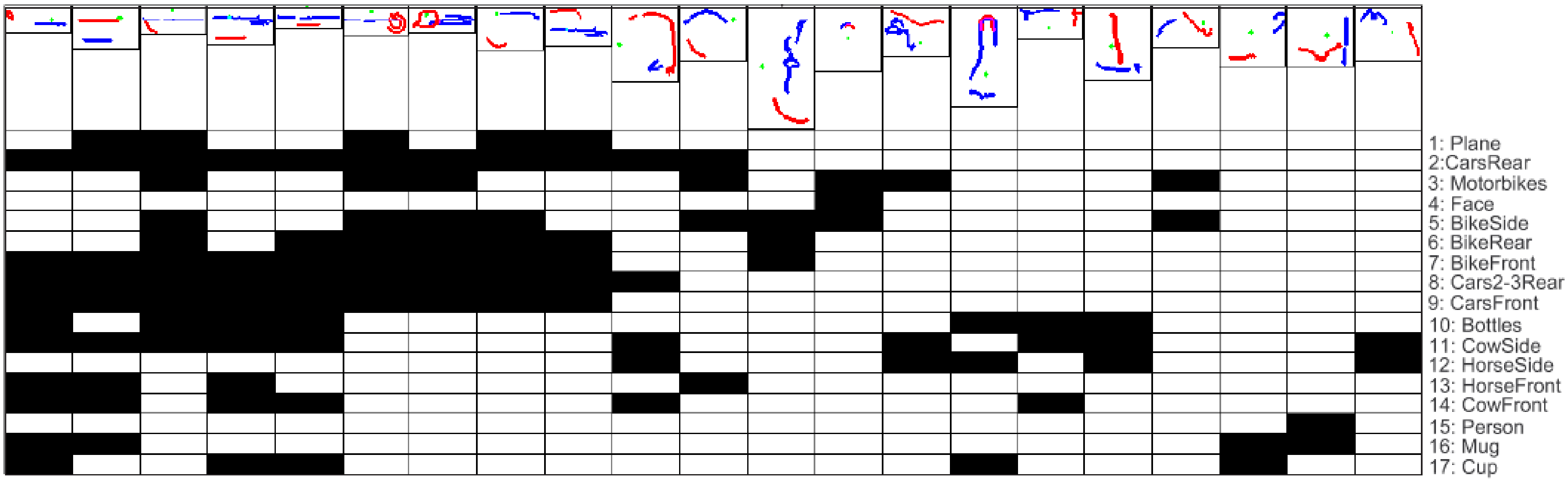
Similarities at the alphabet level

RESULTS 2/6

Incremental vs. Joint-Boosting



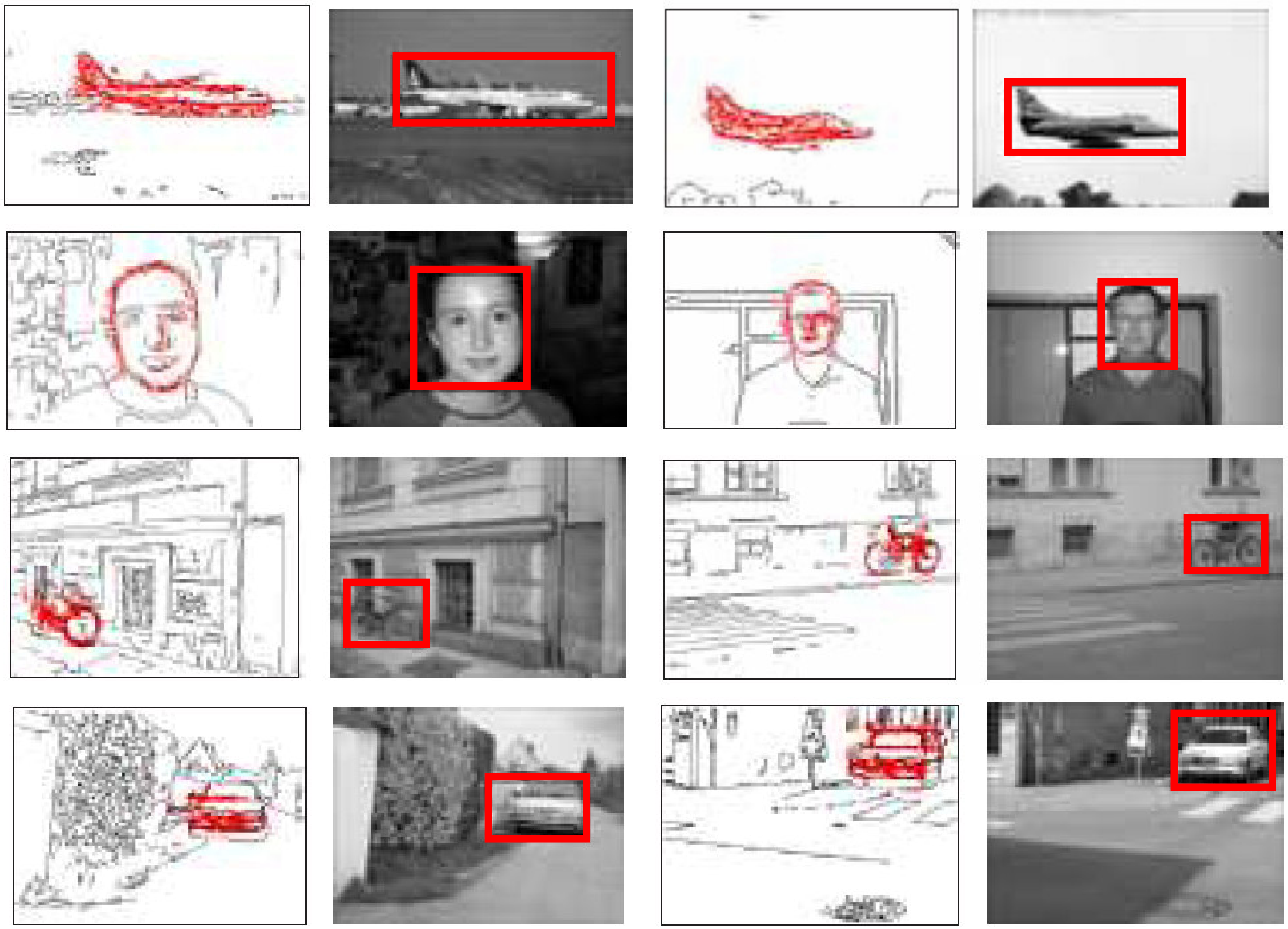
RESULTS 3/6



Sharing of weak detectors

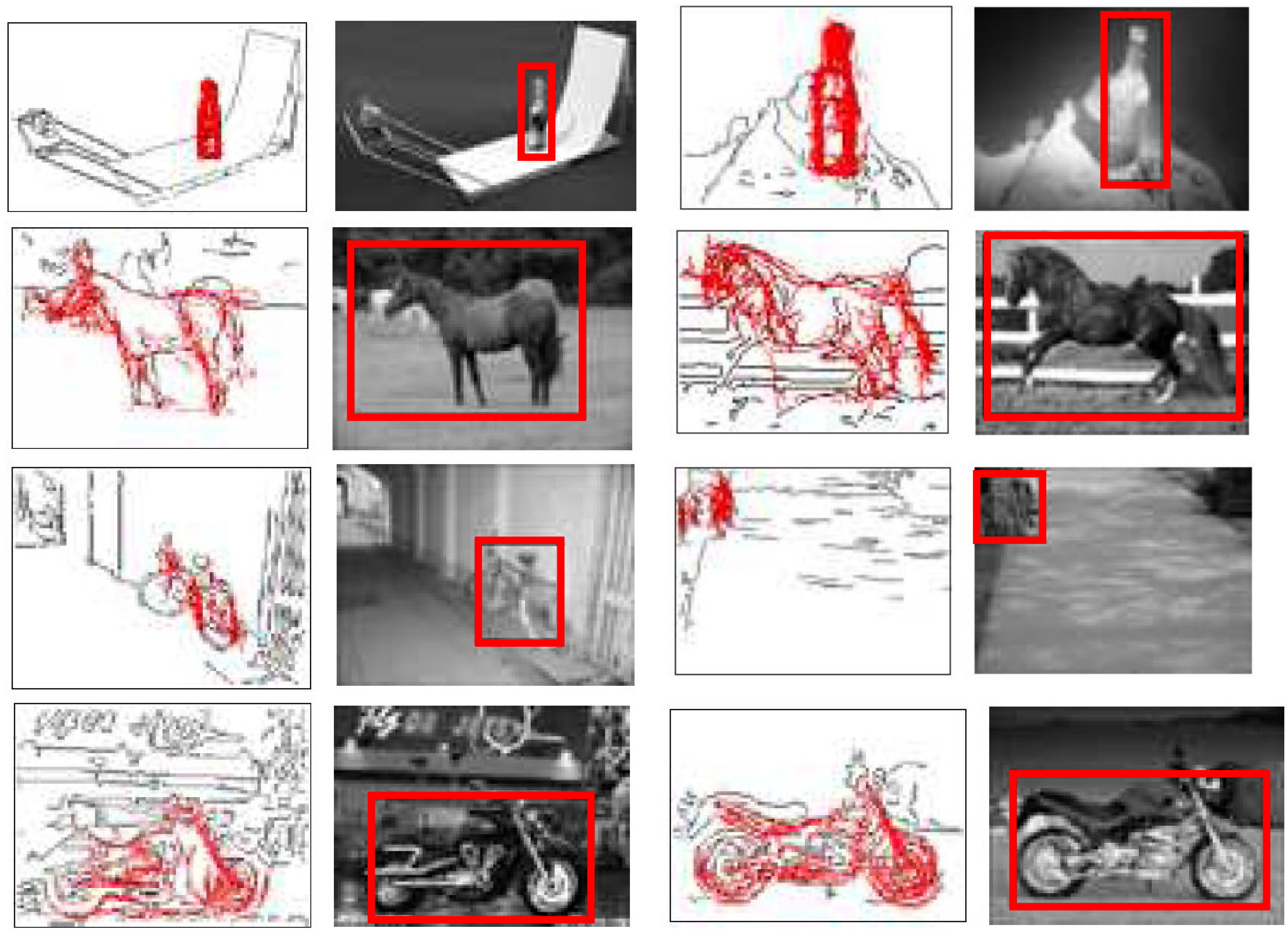
RESULTS 4/6

Examples of detection results



RESULTS 5/6

Examples of detection results



RESULTS 6/6

Detecion results: Independent learning, Joint learning, one-class, multi-class

Class	Plane	CarR	Mb	Face	B-S	B-R	B-F	Car23	CarF	Bottle	CowS	H-S	H-F	CowF	Pers.	Mug	Cup
Ref.	6.3 [5],C	6.1 [10],D	7.6 [15],D	6.0 [15],D							0.0 [10],D						
I,$\mathcal{T}$	7.4	2.3	4.4	3.6	28	25.0	41.7	12.5	10.0	9.0	0.0	8.2	13.8	18.0	47.4	6.7	18.8
J,$\mathcal{T}$	7.4	3.2	3.9	3.7	22.	20.8	31.3	12.5	7.6	10.7	0.0	7.8	11.5	12.0	42.0	6.7	12.5
I,$\mathcal{M}$	1.1	7.0	6.2	1.4	10.	7.7	8.5	5.2	7.6	7.1	1.6	10.0	8.2	9.5	29.1	5.1	8.0
J,$\mathcal{M}$	1.5	4.3	4.5	1.6	8.9	5.9	7.7	3.8	8.5	6.1	1.3	11.0	4.7	6.8	27.7	5.8	8.3

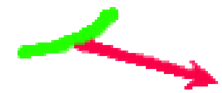
See paper for details!

Motorbikes: Shotton et al. 2005 → 7.6% Ours: 4.4 % (indep.), 3.9 % (joint)

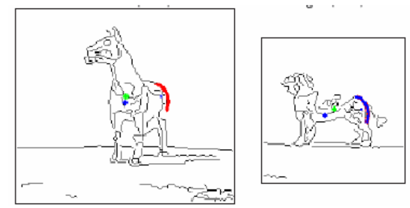
Bicycle (Rear): Ours: 25.0 % (indep.), 20.8 % (joint)

Cups: Ours: 18.8 % (indep.), 10.0 % (joint)

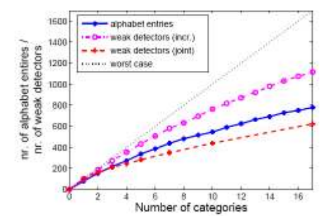
SUMMARY

- Shape and geometry for categorization and detection 

- Shared over categories (and aspects)

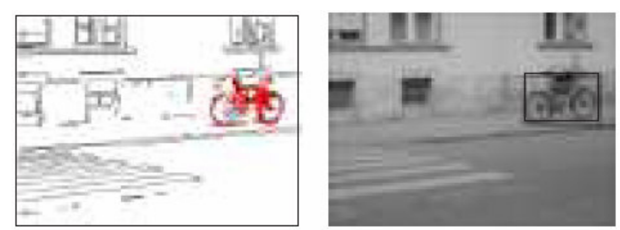


- Required number of weak detectors grows sublinearly with the number of categories



- Alphabet and the detector can be updated incrementally

- Joint learning gives better results with the same amount of training data



THANK YOU!