Class-Specific Hough Forests for Object Detection

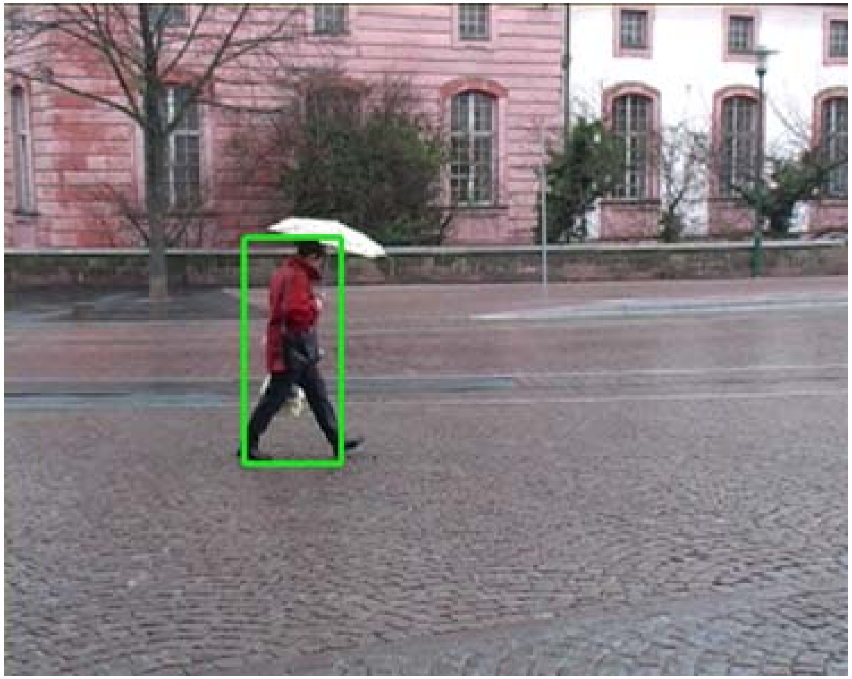
Juergen Gall[1] and Victor Lempitsky[2]

[1]BIWI, ETH Zurich
[1]Max-Planck-Institute for Informatics
[2]Microsoft Research Cambridge

Motivation

- Parts of an object provide useful spatial information
- Classification of object parts (foreground/background)
- Combine spatial information and class information during learning

Overview



Related Work

- Explicit model of object: Detect parts → Assemble parts together (e.g. Pictorial Structures)
- Implicit model of object: Learn relation of parts
  - Codebook based on appearance (e.g. Leibe et al. IJCV'08)
  - Codebook based on appearance and spatial information (Opelt et al. IJCV'08; Shotton et al. PAMI'08)
  - Grid-based classifier for object parts (Winn and Shotton CVPR'06)
  - Class-specific Hough forest: Generalized Hough transform within Random forest framework (Breiman ML'01)

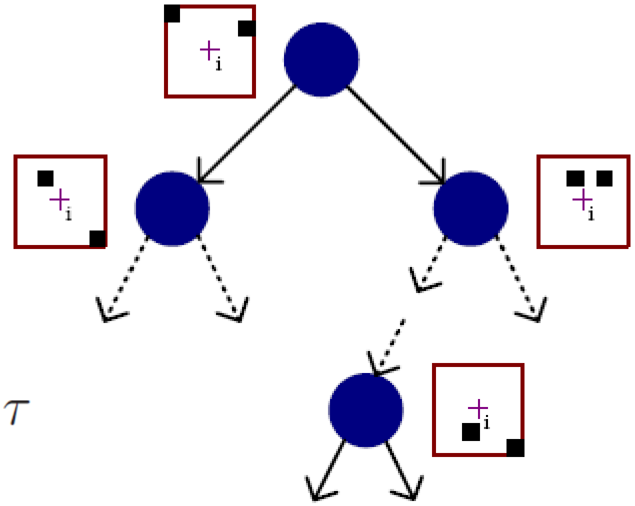
Random Forest

- Image patch:

$$\mathcal{I}_i = (I_i^1, I_i^2, \ldots I_i^C)$$

- Binary tests:

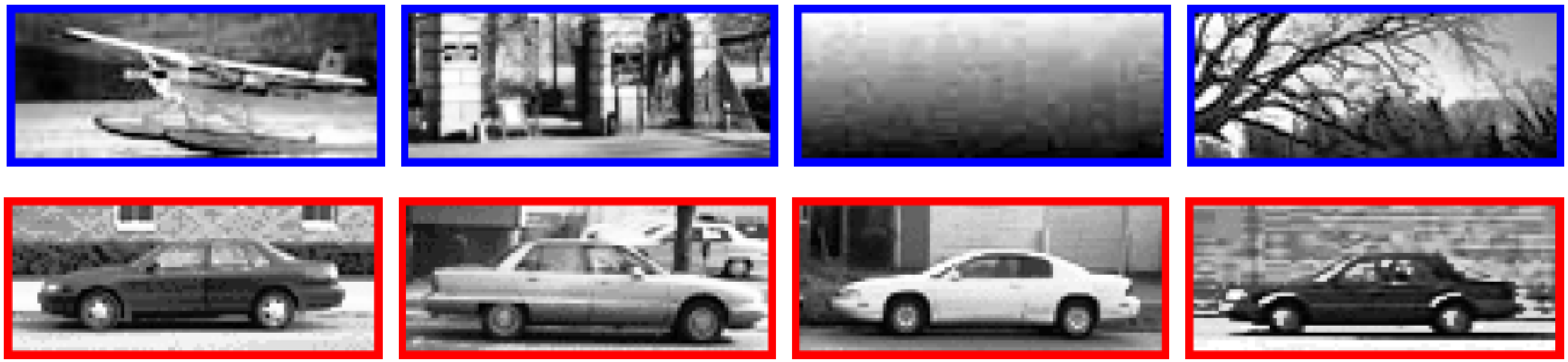
$$t_{a,p,q,r,s,\tau}(\mathcal{I}) = \begin{cases} 0, & \text{if } I^a(p,q) < I^a(r,s) + \tau \\ 1, & \text{otherwise.} \end{cases}$$



- Binary tests are selected during training from a random subset of all binary tests

Training

- Training set:



$$A = \{\mathcal{P}_i = (\mathcal{I}_i, c_i, \mathbf{d}_i)\}$$

- Class information: c_i (class label)
- Spatial information: $\mathbf{d}_i$ (relative position to object center)

Microsoft Research Cambridge
ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
BIWI Computer Vision
max planck institut informatik

Binary Tests Selection

- Test with optimal split:

$$\operatorname*{argmin}_k \Big(U_\star(\{p_i \,|\, t^k(\mathcal{I}_i)=0\}) + U_\star(\{p_i \,|\, t^k(\mathcal{I}_i)=1\}) \Big)$$

- Class-label uncertainty:

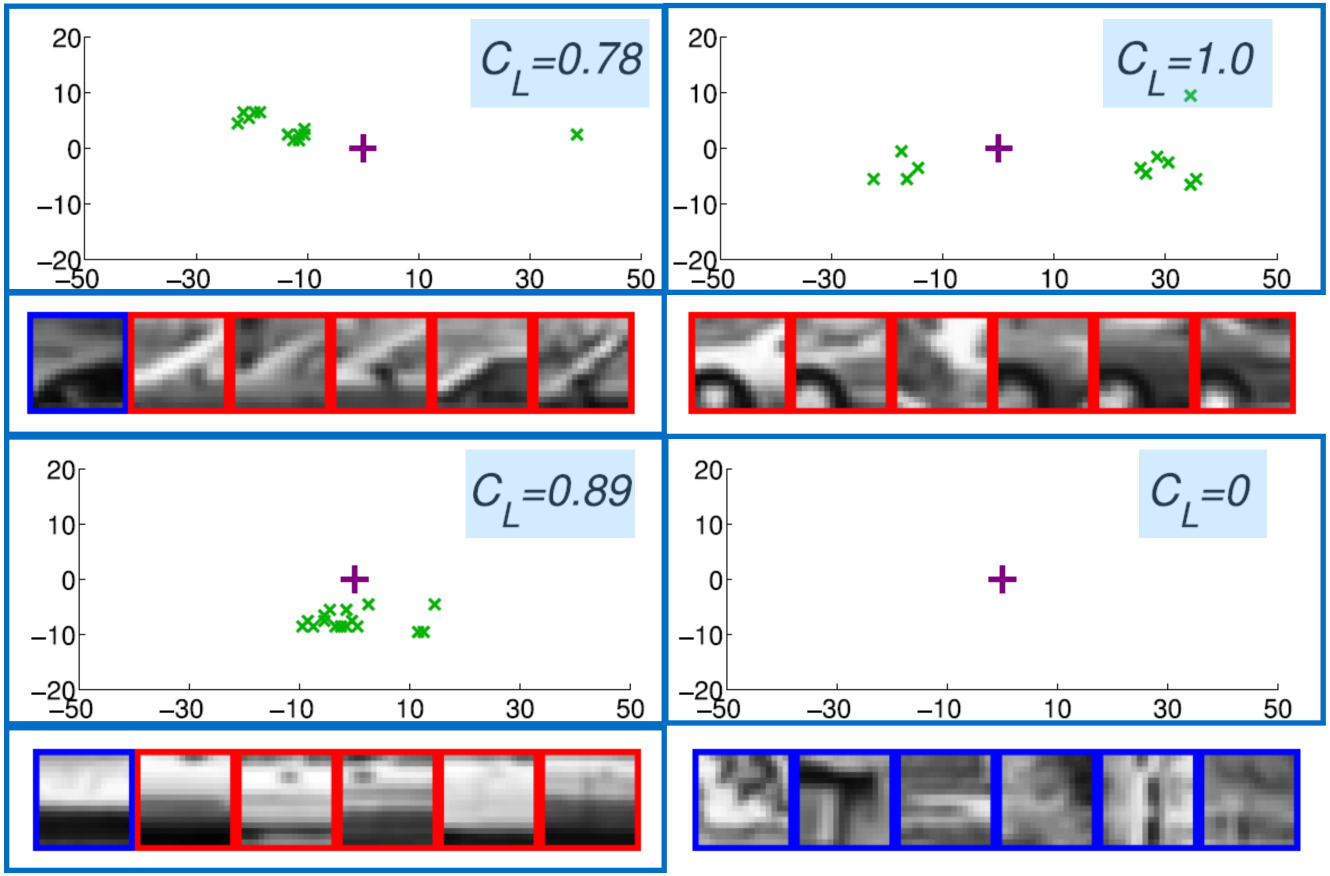
$$U_1(A) = |A| \cdot Entropy(\{c_i\})$$

- Offset uncertainty:

$$U_2(A) = \sum_{i:c_i=1} (\mathbf{d}_i - \mathbf{d}_A)^2$$

- Interleaved: Type of uncertainty is randomly selected for each node

Leaves



- Class probability: $C_L = \dfrac{\left|\{P_i \in L : c_i = 1\}\right|\left|\{P_i \in A : c_i = 0\}\right|}{\left|\{P_i \in L : c_i = 0\}\right|\left|\{P_i \in A : c_i = 1\}\right| + \left|\{P_i \in L : c_i = 1\}\right|\left|\{P_i \in A : c_i = 0\}\right|}$

Spatial probability

- For location x and given image patch $\mathcal{I}(y)$ and tree $\mathcal{T}$

$$p\big(E(\mathbf{x})|\,\mathcal{I}(\mathbf{y});\mathcal{T}\big) = \left[\frac{1}{|D_L|}\sum_{d\in D_L}\frac{1}{2\pi\sigma^2}\exp\left(-\frac{||(\mathbf{y}-\mathbf{x})-d||^2}{2\sigma^2}\right)\right]\cdot C_L$$

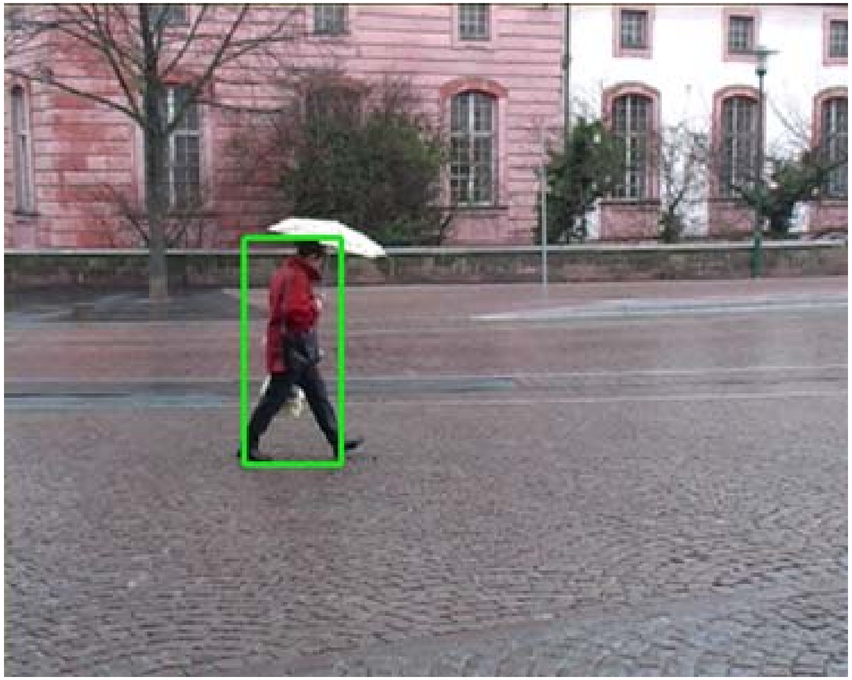
- Over all trees:

$$p\big(E(\mathbf{x})|\,\mathcal{I}(\mathbf{y});\{\mathcal{T}_t\}_{t=1}^T\big) = \frac{1}{T}\sum_{t=1}^{T} p\big(E(\mathbf{x})|\,\mathcal{I}(\mathbf{y});\mathcal{T}_t\big)$$

- Accumulation over all image patches:

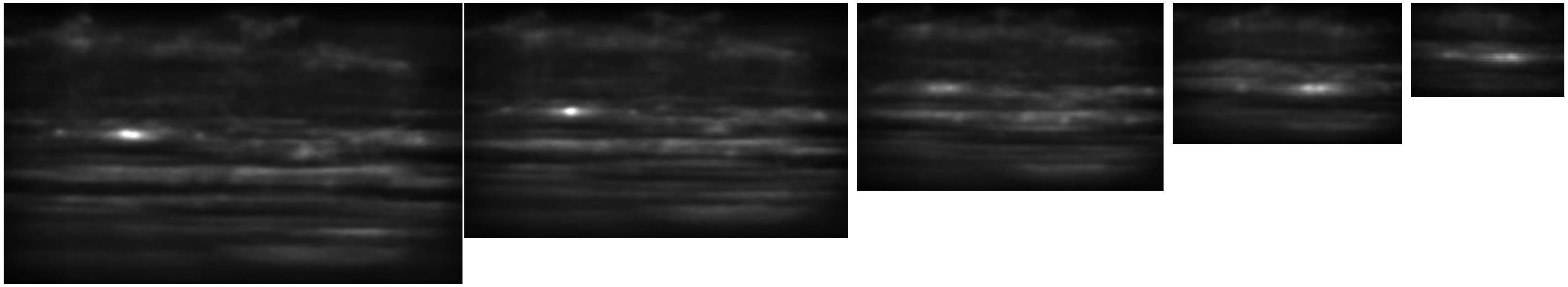
$$V(\mathbf{x}) = \sum_{y\in B(x)} p\big(E(\mathbf{x})|\,\mathcal{I}(\mathbf{y});\{\mathcal{T}_t\}_{t=1}^T\big)$$

Detection



Multi-Scale and Multi-Ratio

- Multi Scale: 3D Votes (x, y, scale)



- Multi-Ratio: 4D Votes (x, y, scale, ratio)

UIUC Cars - Multi Scale

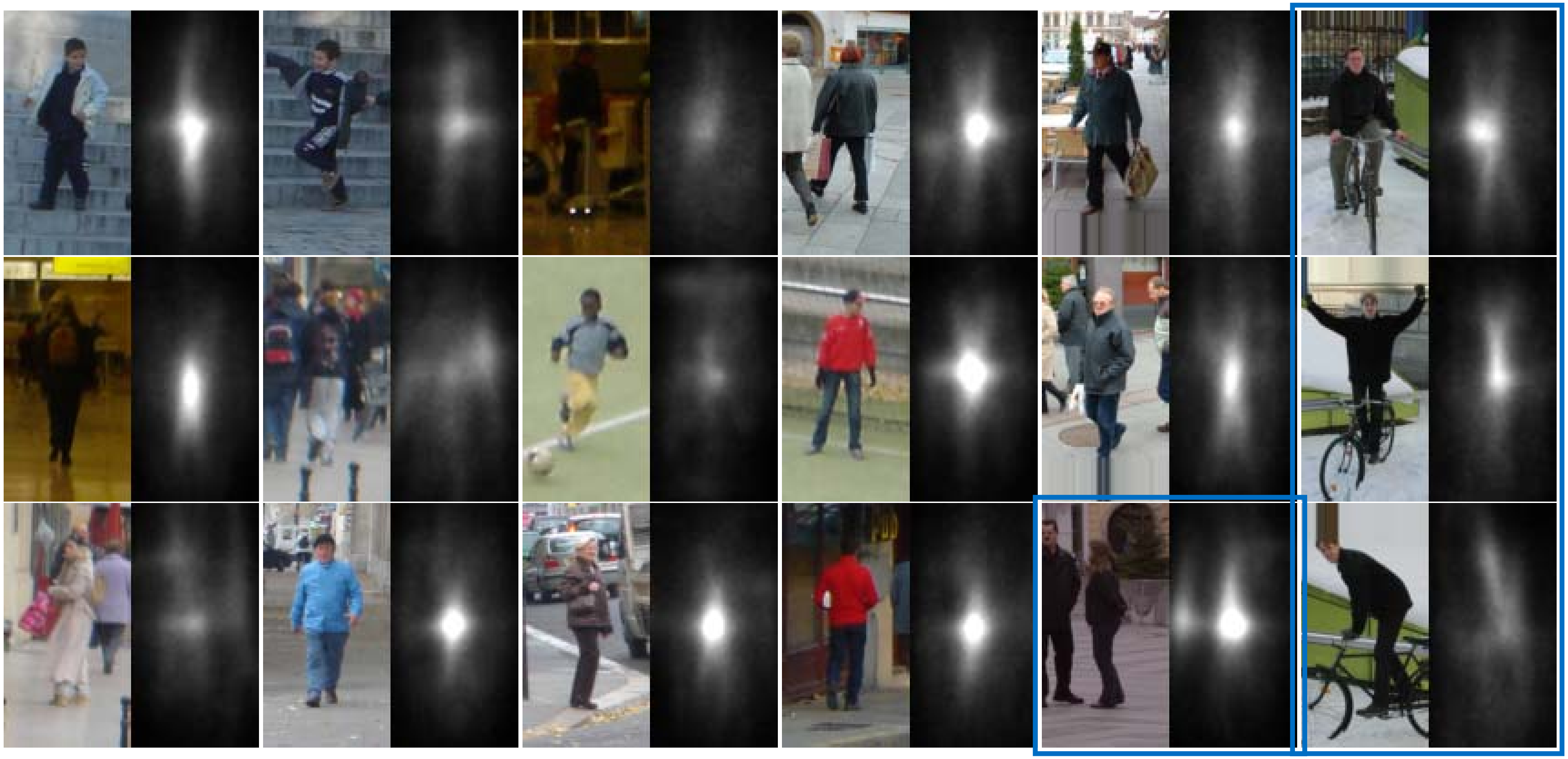
- Wrong (EER)

- Correct

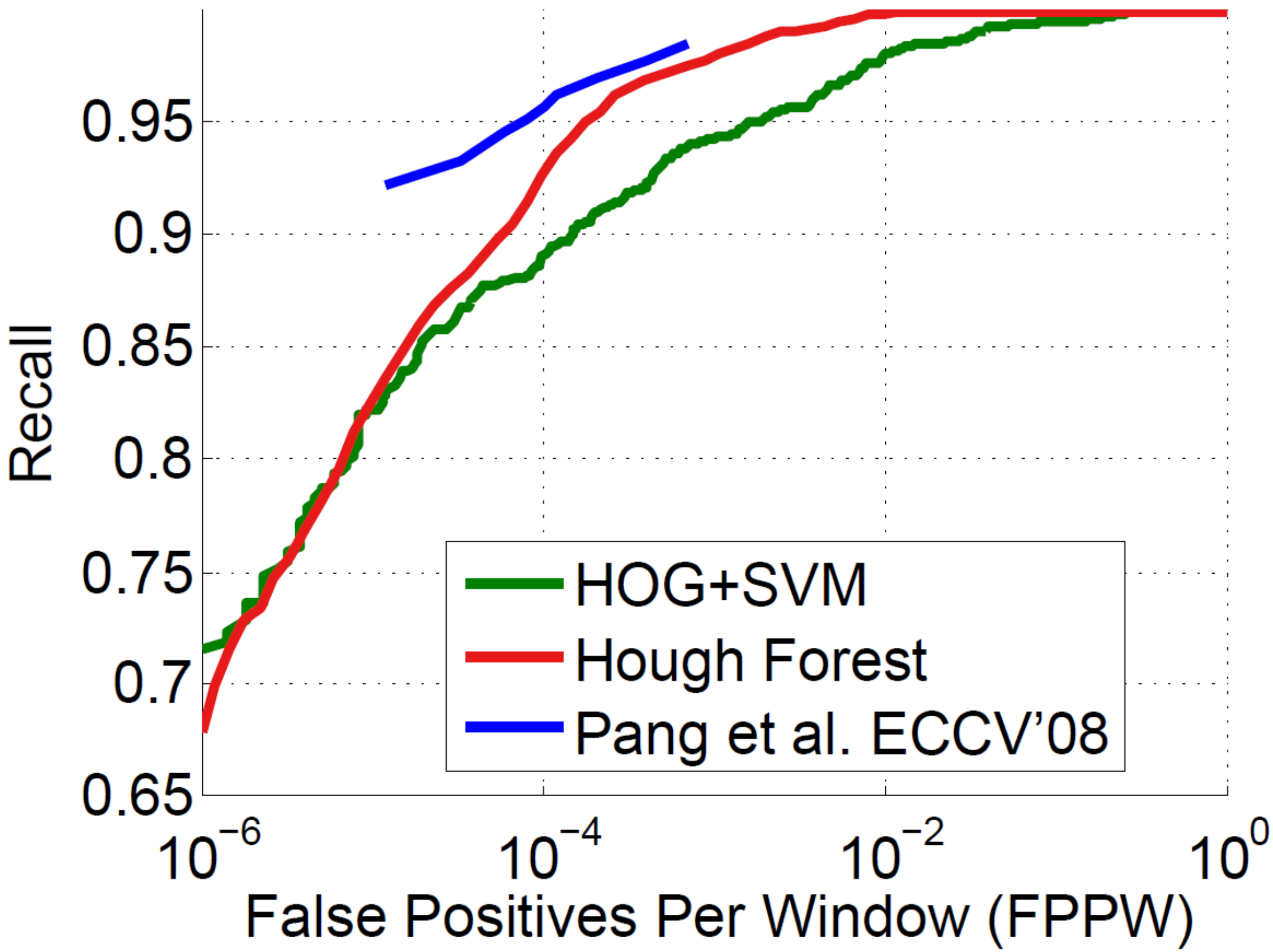
Comparison

Methods	UIUC-Single	UIUC-Multi
Hough-based methods		
Implicit Shape Model [10]	91%	–
ISM+verification [10]	97.5%	95%
Boundary Shape Model [17]	85%	–
Random forest based method		
LayoutCRF [27]	93%	–
State-of-the-art		
Mutch and Lowe CVPR'06 [15]	99.9%	90.6%
Lampert et al. CVPR'08 [9]	98.5%	98.6%
Our approach		
Hough Forest	98.5%	98.6%
HF - Weaker supervision	94.4%	–

Pedestrians (INRIA)



Pedestrians (INRIA)



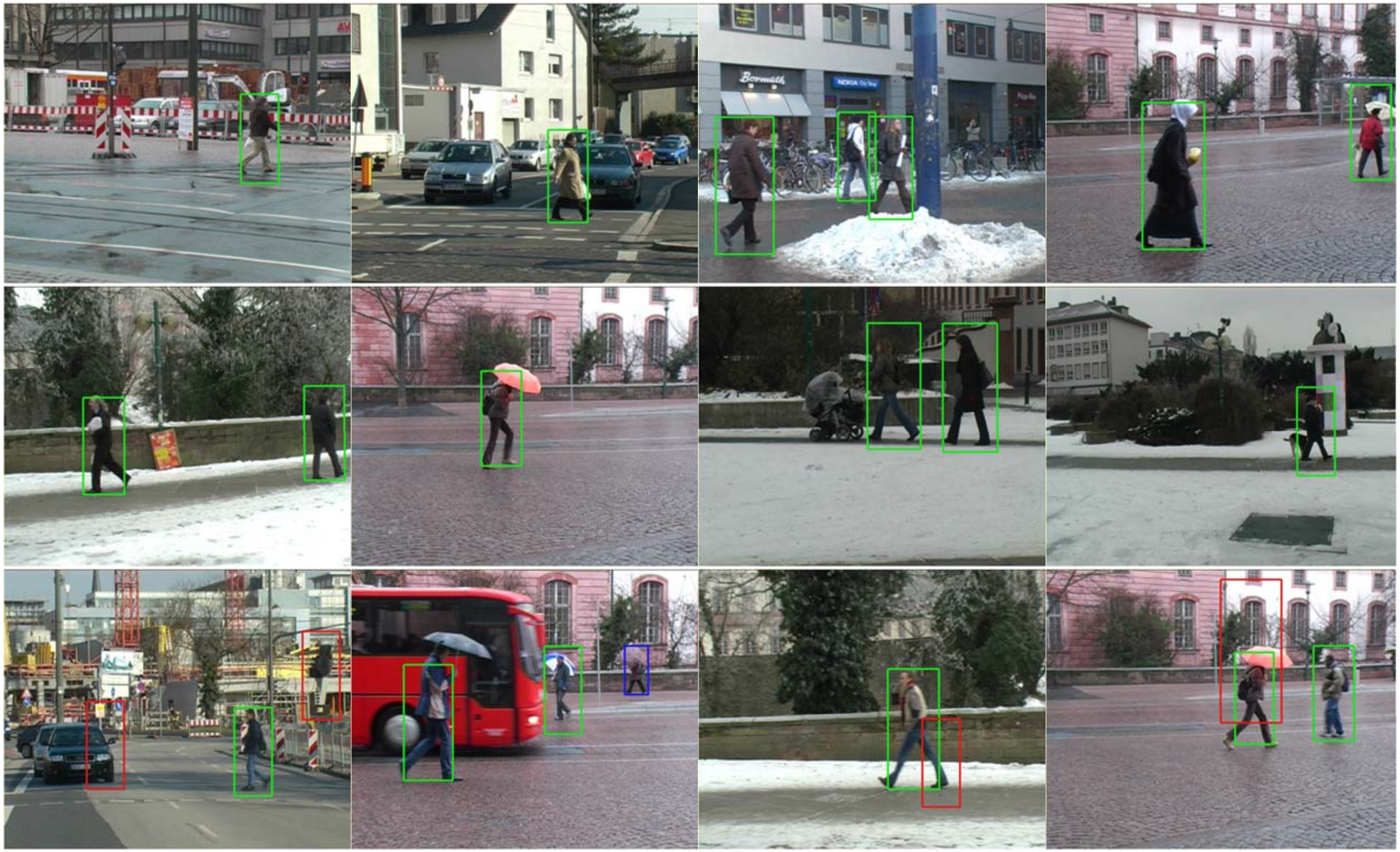
Microsoft Research Cambridge

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

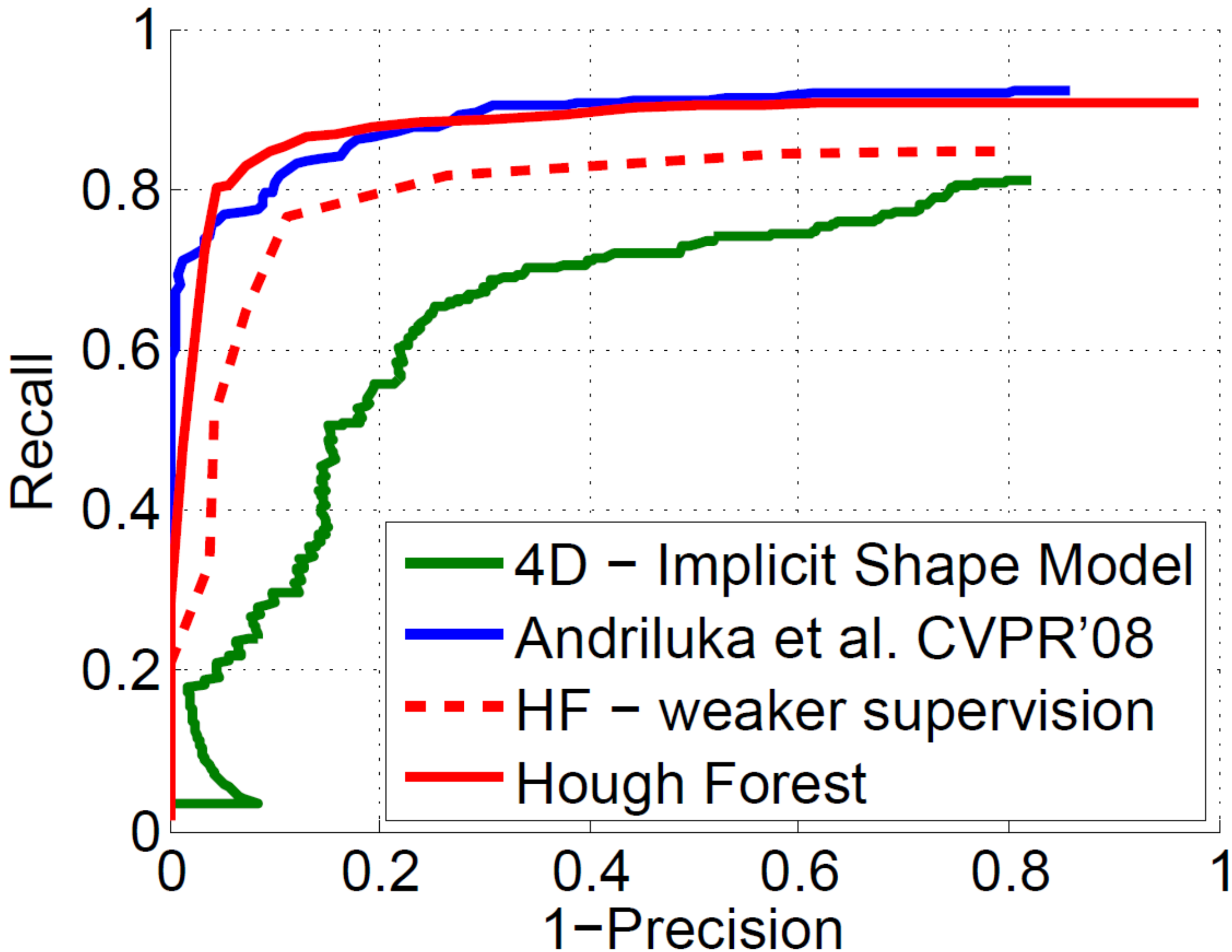
BIWI Computer Vision

max planck institut informatik

Pedestrians (TUD)

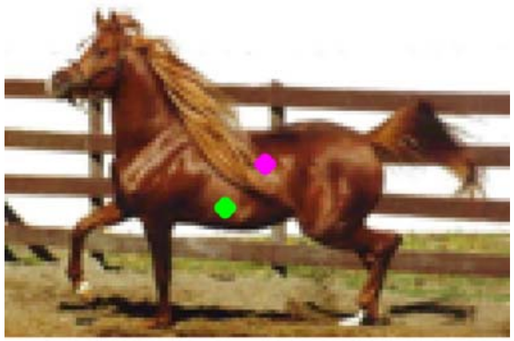
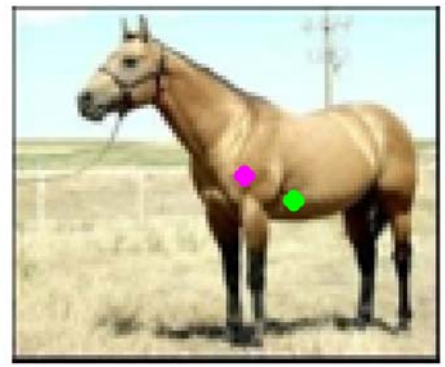


Pedestrians (TUD)

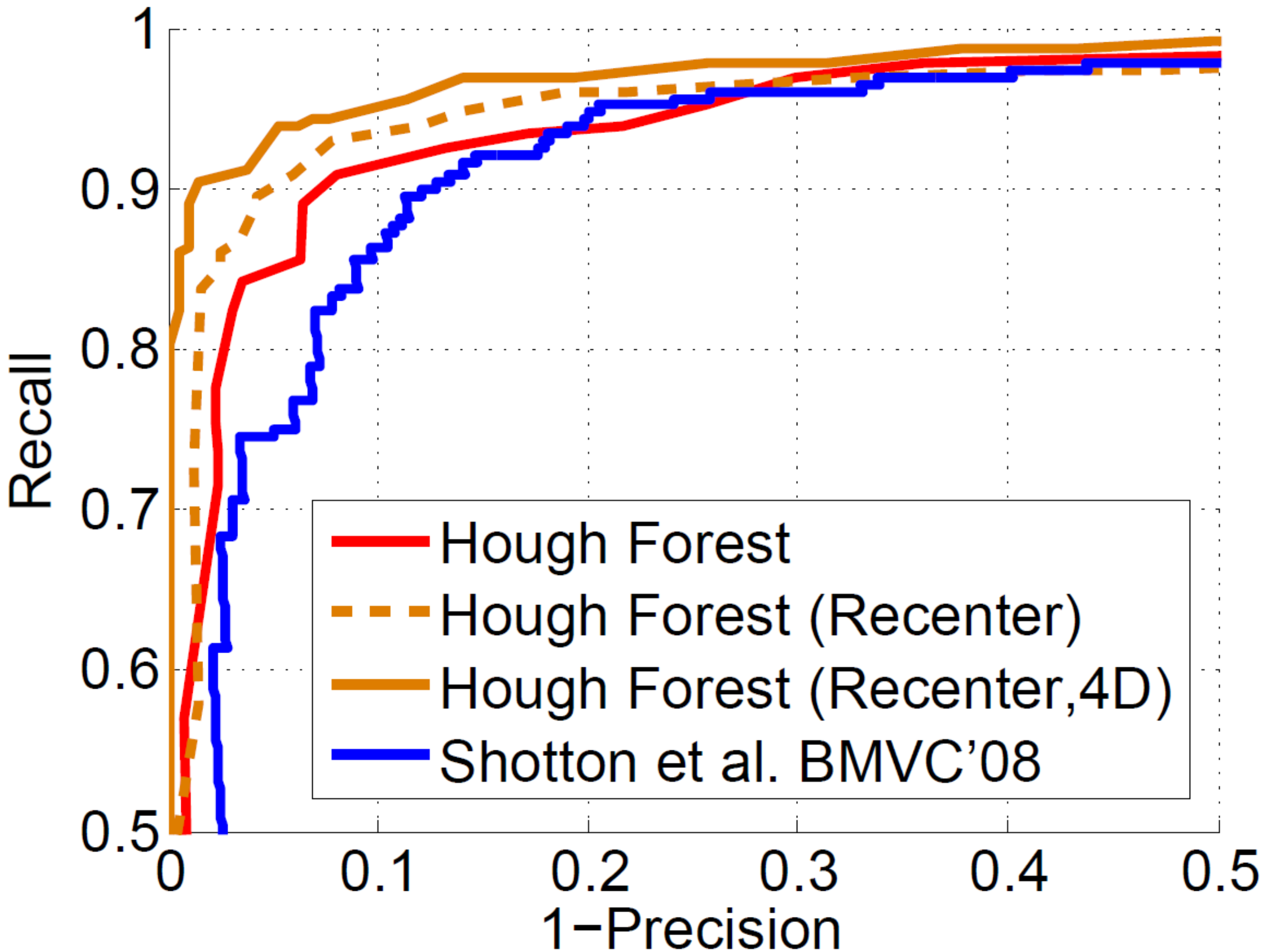


Recenter

- Object’s center ≠ Centre of bounding box
- Split training data → Estimate centers iteratively

Weizmann Horses



Summary

- Superior to previous methods using related techniques
- State-of-the-art for several datasets
- Advantages over related Hough-based methods:
  - Combine spatial information and class information
  - No sparse features like SIFT
  - GPU → real-time performance is feasible
  - Large and high-dimensional datasets
  - Bounding box-annotated training data is sufficient
- Focus: Get strong signal → Improve Detection
- 2-class problem → Multi-class problem

Microsoft Research Cambridge

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

BIWI Computer Vision

max planck institut informatik

Thank you for your attention.

The major part of the research project was undertaken when Juergen Gall was an intern with Microsoft Research Cambridge. The advice from Toby Sharp, Jamie Shotton, and other members of the Computer Vision Group at MSRC is gratefully acknowledged. We would also like to thank all the researchers, who have collected and published the datasets we used for the evaluation.