UNIVERSITY OF LEEDS

Implicit Color Segmentation Features for Pedestrian and Object Detection

Patrick Ott, Mark Everingham

Computer Vision Group, School of Computing

Presented at ICCV 2009

Pedestrian Detection?

UNIVERSITY OF LEEDS

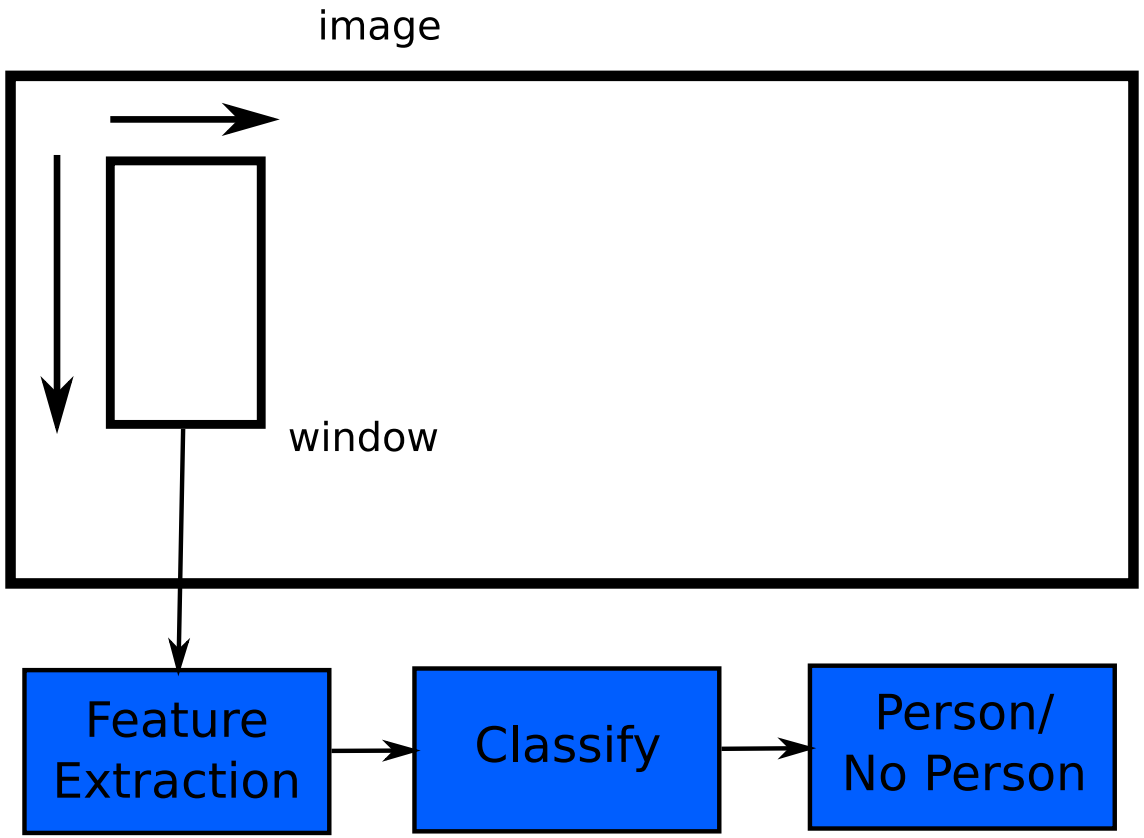
Overview

UNIVERSITY OF LEEDS

- Sliding window detector
- HOG features
- Segmentation as a detection cue
- CHOG
- Experiments

Sliding Window Detection

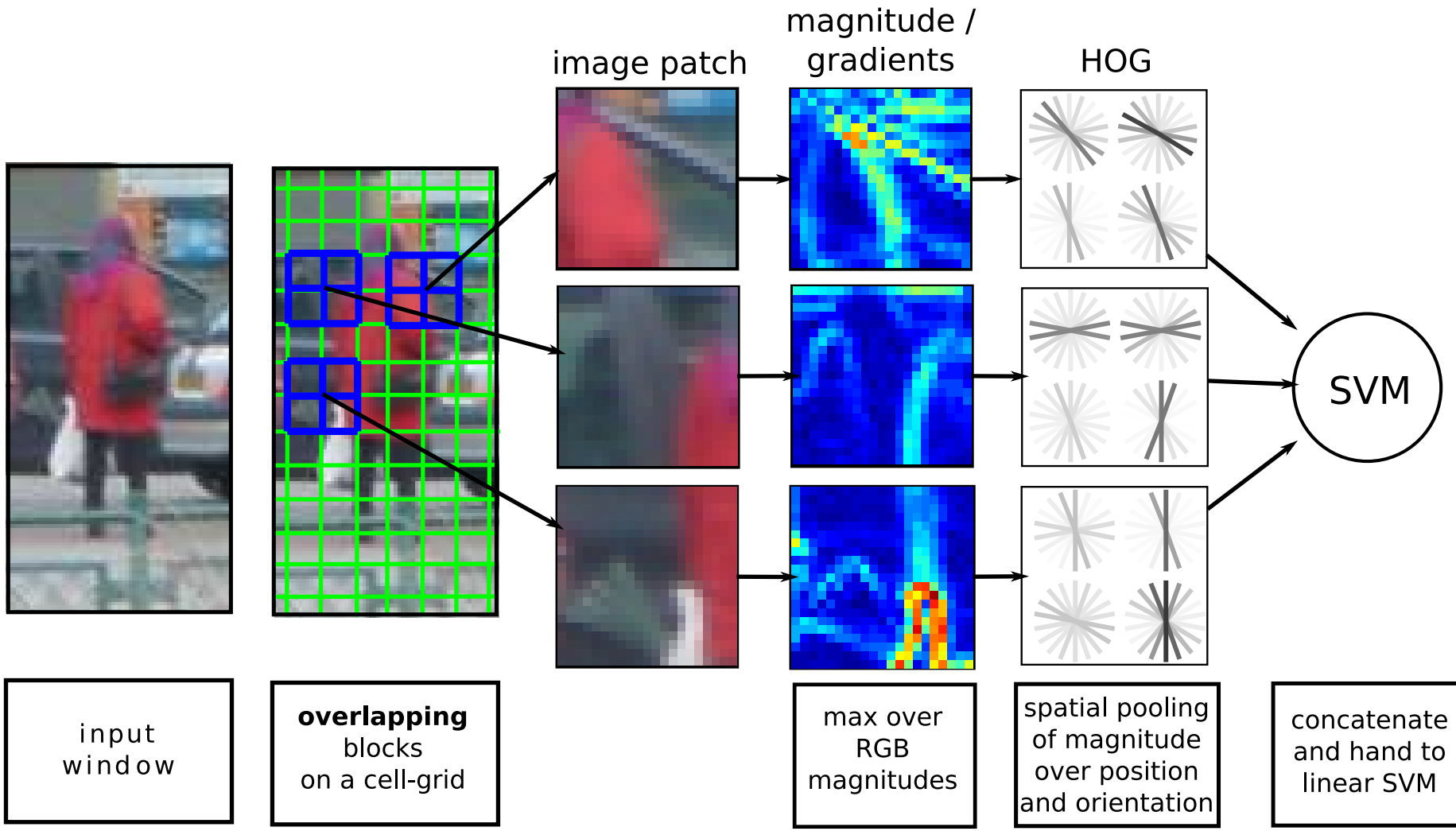
UNIVERSITY OF LEEDS



- Slide window over all positions in the image
- Rescale image to cover multiple scales

HOG, Dalal & Triggs 2005

UNIVERSITY OF LEEDS



HOG, Dalal & Triggs 2005

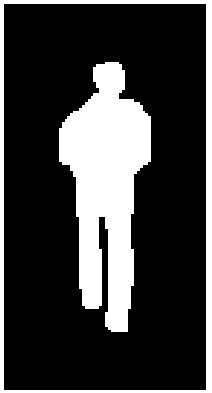
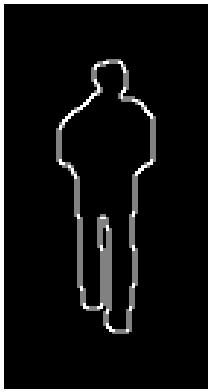
UNIVERSITY OF LEEDS

- Good
  - Local invariance due to spatial pooling
  - Invariance to brightness (gradients)
  - Invariance to contrast due to block normalization
- Not so good
  - Max-RGB gradients introduce background clutter
  - Object edges often not very strong
  - $\longrightarrow$ noisy HOG descriptors

Segmentation as a detection cue

UNIVERSITY OF LEEDS

- If we had a perfect segmentation...

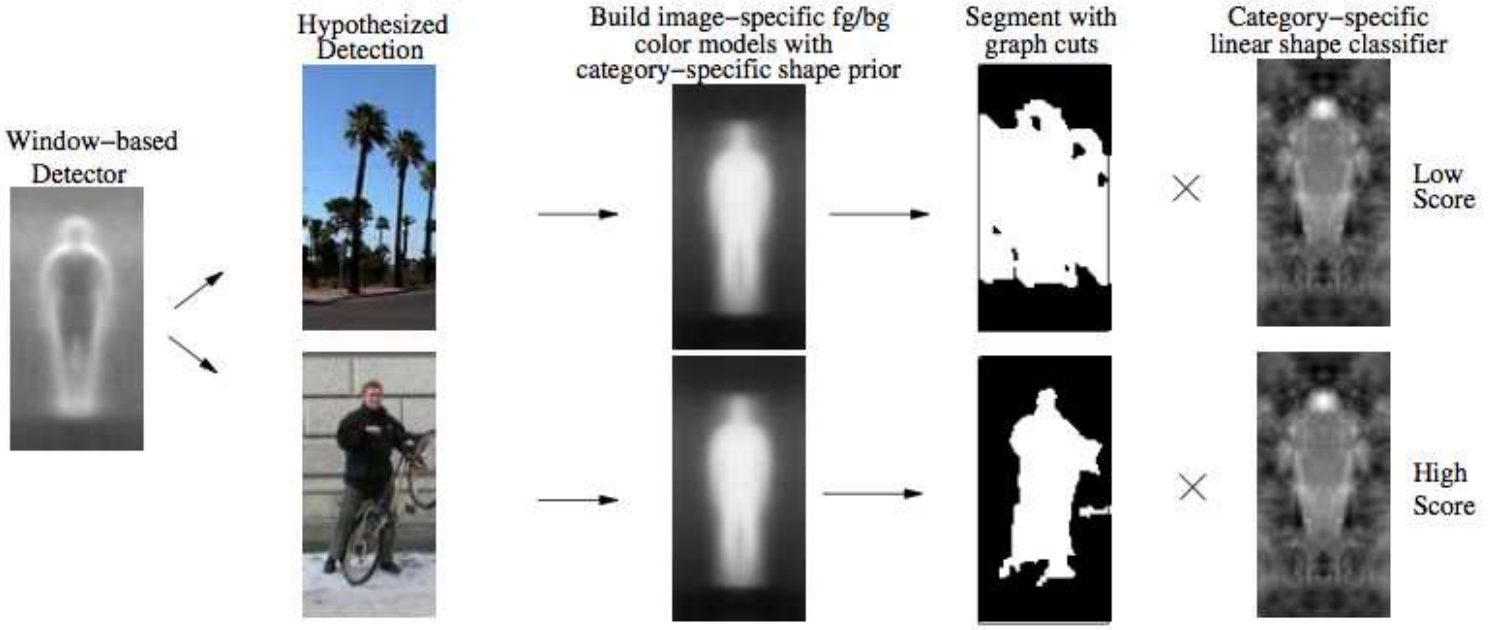




- Strong object edges
- No clutter in background or foreground
- Good starting point for a gradient-based detector

Related Work

UNIVERSITY OF LEEDS

- Ramanan (CVPR 2007) used segmentation to verify detections



- Our goal: Incorporate segmentation into feature extraction stage

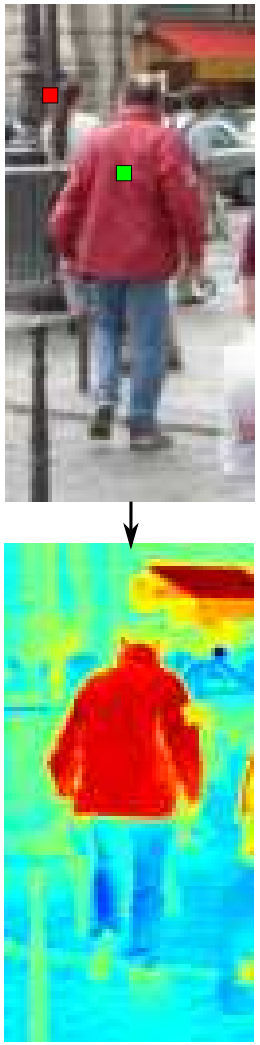
Obtaining soft segmentations

UNIVERSITY OF LEEDS

- Goal: Build a model for FG/BG, so we can ‘label’ pixels
- Local Gaussian assumption around $\mathbf{p}_{FG}$ (■) and $\mathbf{p}_{BG}$ (■) with means $\mathbf{m}_{FG}, \mathbf{m}_{BG}$
- Fast and simple projection:

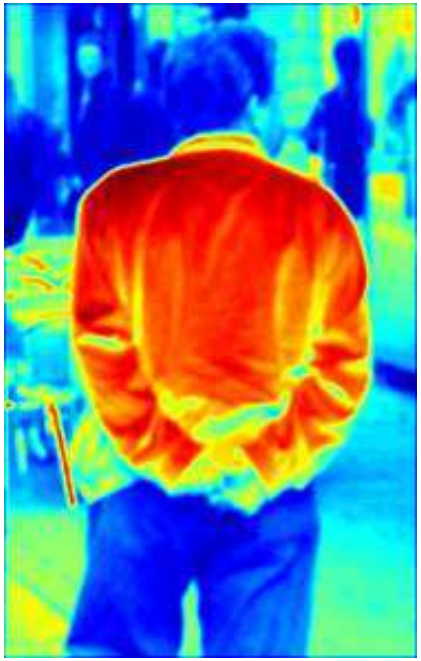
$$\hat{\mathbf{w}} \propto \left(\mathbf{m}_{FG} - \mathbf{m}_{BG}\right)$$

  - ♦ Simplified form of Fisher discriminant



Motivation

UNIVERSITY OF LEEDS

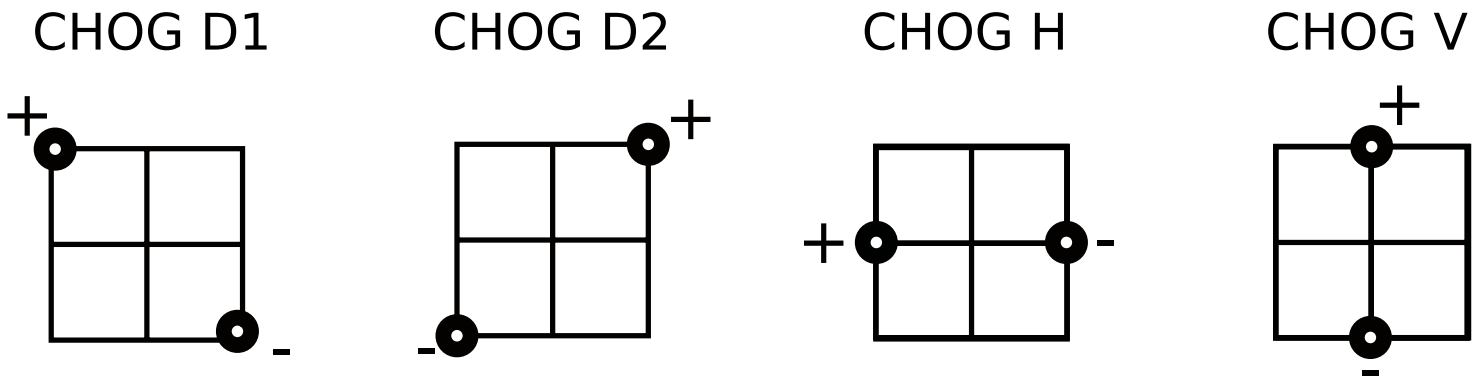


- Soft segmentation allows us to create an instance-specific cue about the object
- $\rightarrow$ ‘this jacket is white’ rather than ‘jackets are white’

Reference Points

UNIVERSITY OF LEEDS

- Our goal is to compute HOG features on the segmentation map S
- To compute $\hat{\mathbf{w}}$ locally we link the position of the reference points $\mathbf{p}_{FG}$ and $\mathbf{p}_{BG}$ to the position of the HOG block



Color Normalization

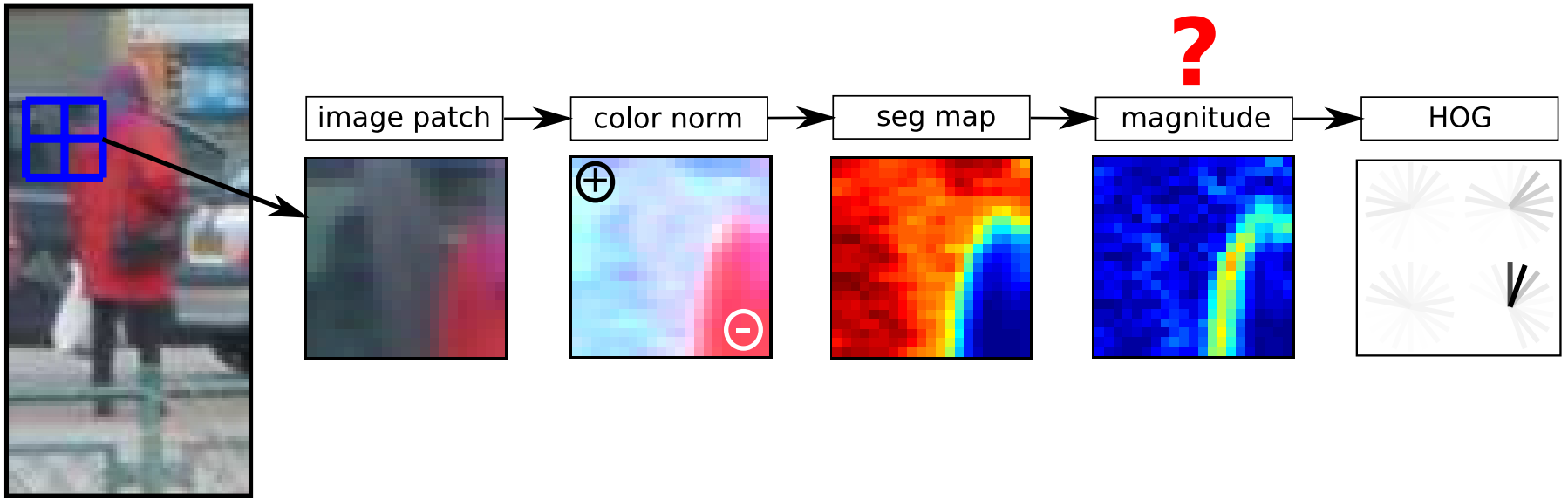
UNIVERSITY OF LEEDS

- Difference in means is dominated by intensity
  $\longrightarrow$ intensity invariant color normalization

$$\mathbf{x}' = \frac{\mathbf{x}}{\max(x_r, x_g, x_b) + \epsilon}$$

Gradient Computation

UNIVERSITY OF LEEDS

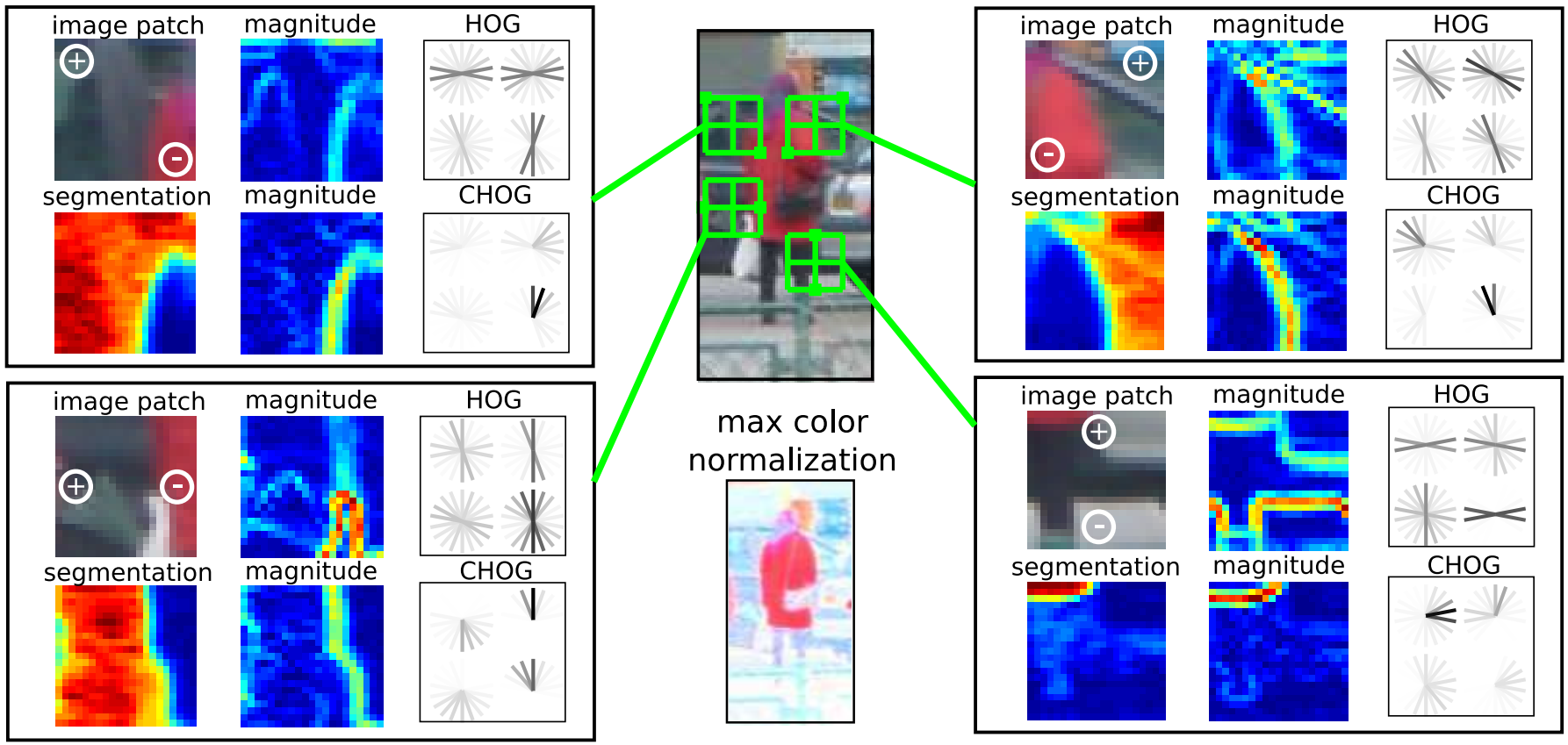


- Gradient of S as weighted sum of original image gradients

$$\frac{\partial}{\partial x}S = \hat{\mathbf{w}}^T \left\langle \frac{\partial}{\partial x}I_R, \frac{\partial}{\partial x}I_G, \frac{\partial}{\partial x}I_B \right\rangle^T$$

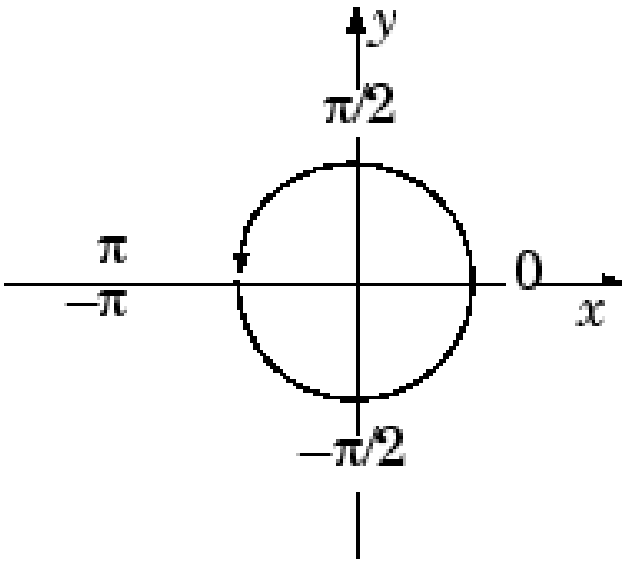
Discussion: CHOG

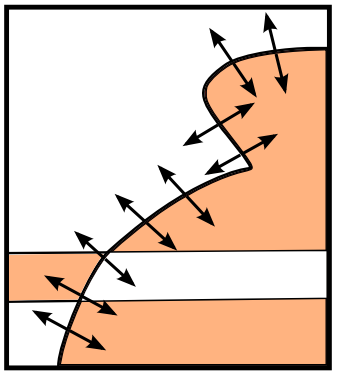
UNIVERSITY OF LEEDS

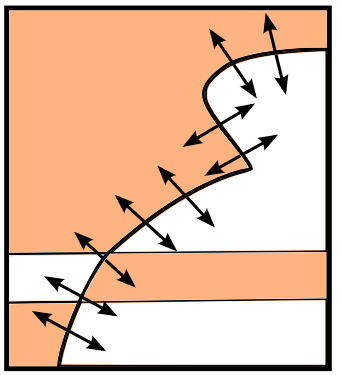


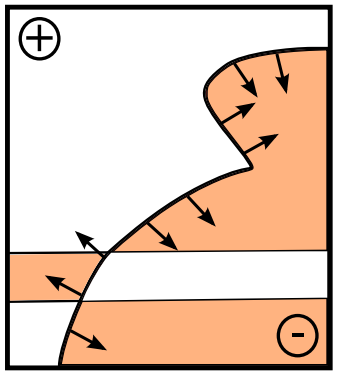
Discussion: Gradient Sign

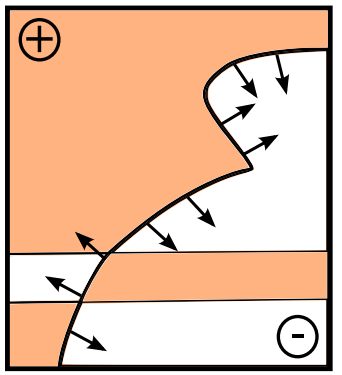
UNIVERSITY OF LEEDS











(a) HOG unsigned (b) CHOG (signed)

Classifier

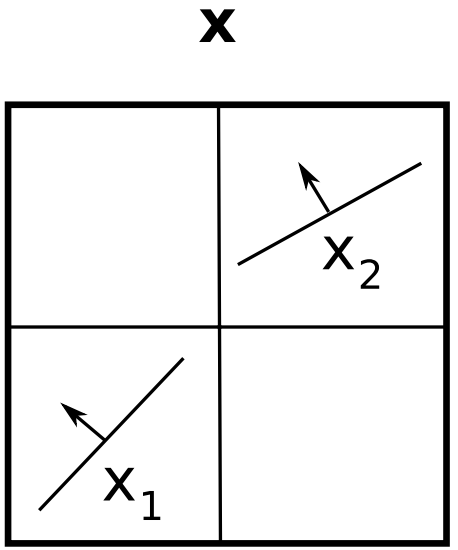
UNIVERSITY OF LEEDS

- Linear SVM & Kernel SVM with quadratic kernel

$$K(\mathbf{x}, \mathbf{y}) = \left(\mathbf{x}^T \mathbf{y} + 1\right)^2 = \phi(\mathbf{x}) \cdot \phi(\mathbf{y})$$

- The kernel is capable of modeling dependencies between any two features in the feature vector, i.e.

$$\phi(\mathbf{x}) = \left\langle \mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_1\mathbf{x}_2, \ldots \right.$$
$$\left. \sqrt{2}\mathbf{x}_1, \sqrt{2}\mathbf{x}_2, 1 \right\rangle$$



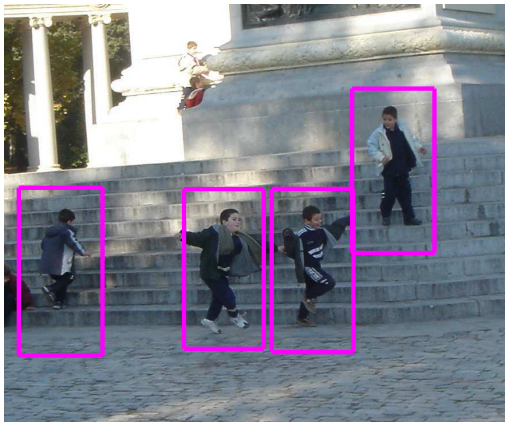
Experiments

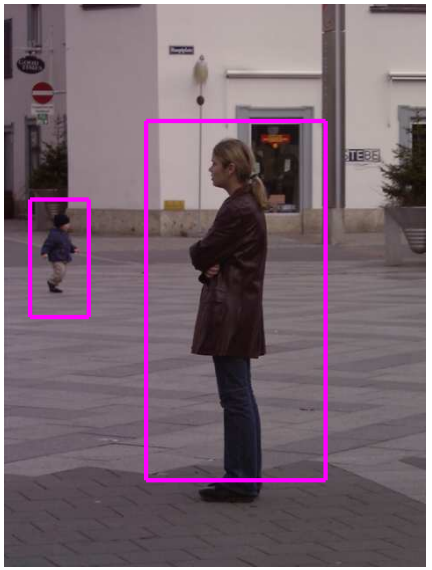
UNIVERSITY OF LEEDS

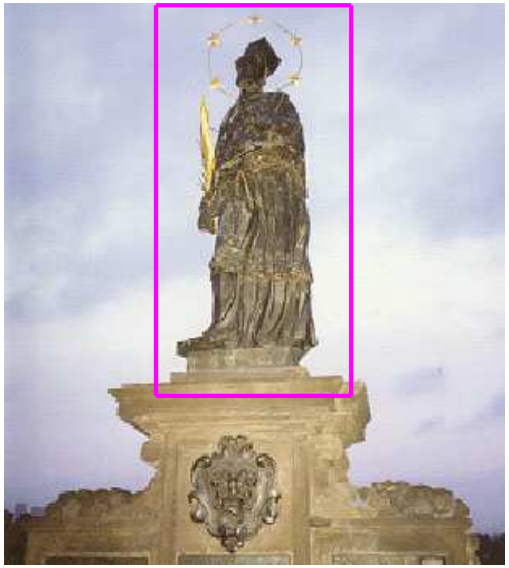
- INRIAPerson dataset $\rightarrow$ followed Dalal & Triggs evaluation protocol
  - Detection-Error-Tradeoff (DET) curves: FPPW vs miss-rate
- PASCAL VOC 2006 dataset $\rightarrow$ followed evaluation protocol
  - Precision-Recall curves

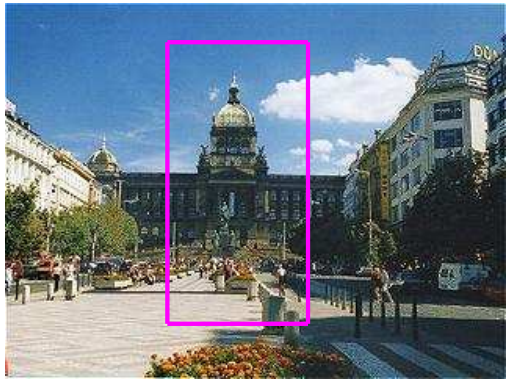
Detection Examples

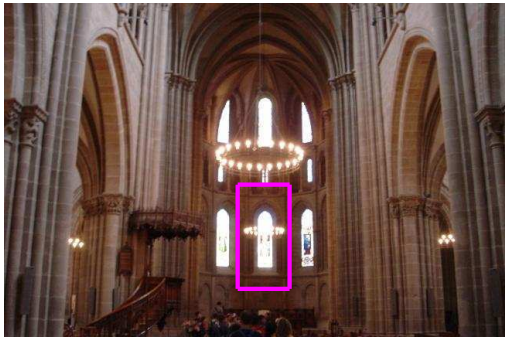
UNIVERSITY OF LEEDS





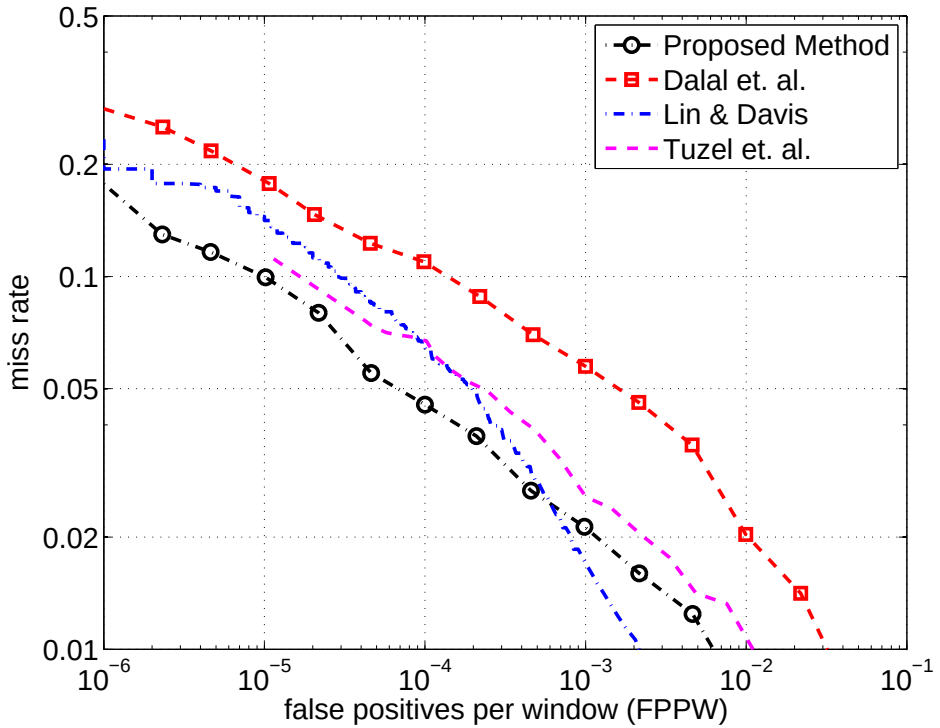






Quantitative Results

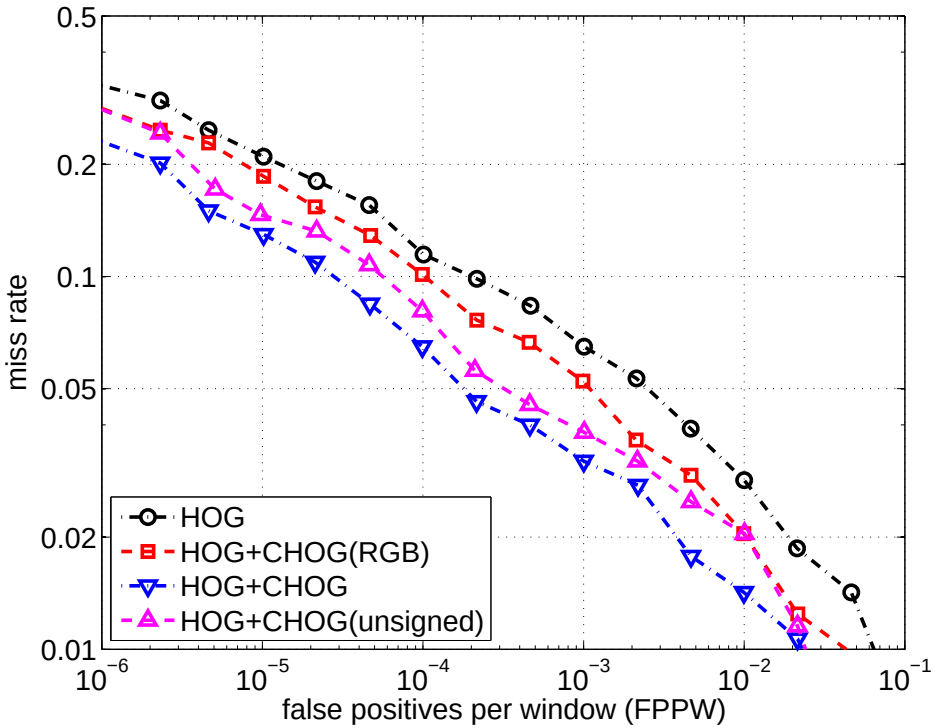
UNIVERSITY OF LEEDS



- For FPPW rate $\leqslant 10^{-4}$ we outperform all other methods
- $\sim 60\%$ improvement compared to Dalal&Triggs at 10^{-4} FPPW.

Descriptor Scheme

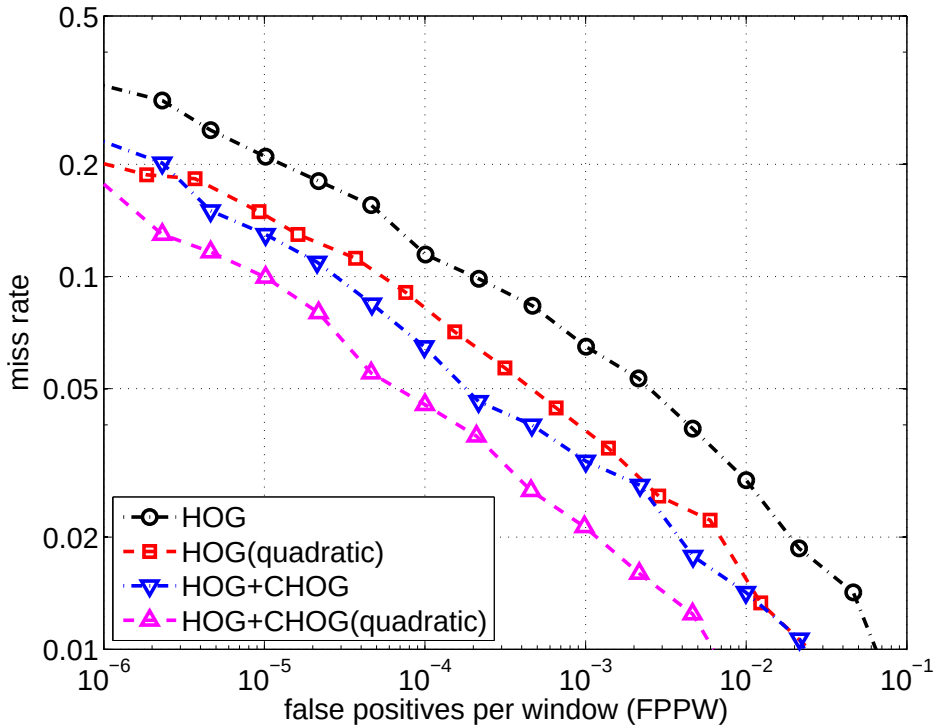
UNIVERSITY OF LEEDS



- 43% relative improvement at 10^{-4} FPPW when using HOG+CHOG
- Signed gradients for CHOG account for 30% of this improvement

Choice of Kernel

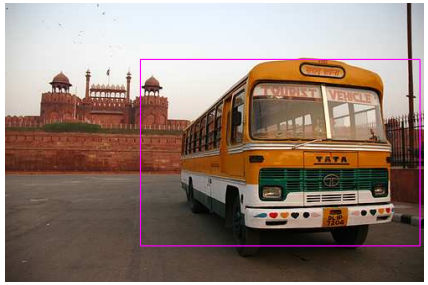
UNIVERSITY OF LEEDS

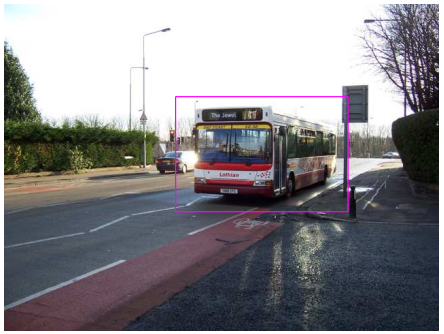


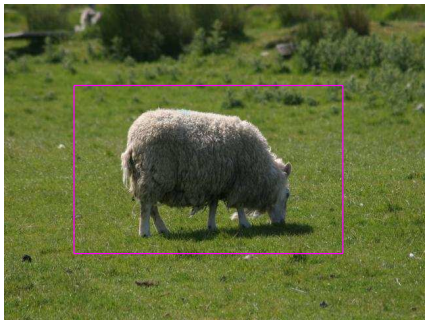
- 31% relative improvement for both, HOG and HOG+CHOG, with quadratic kernel

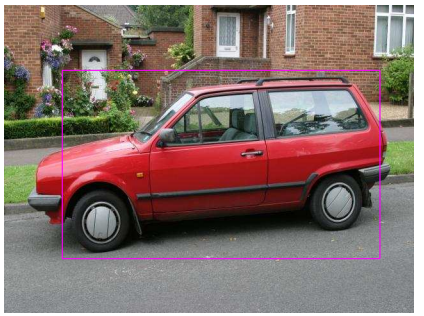
VOC Detection Examples

UNIVERSITY OF LEEDS





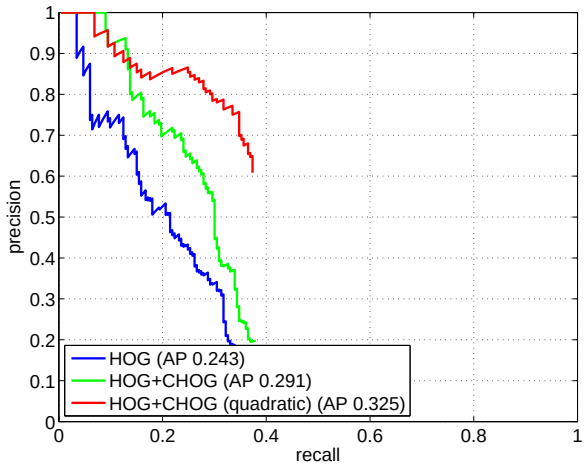




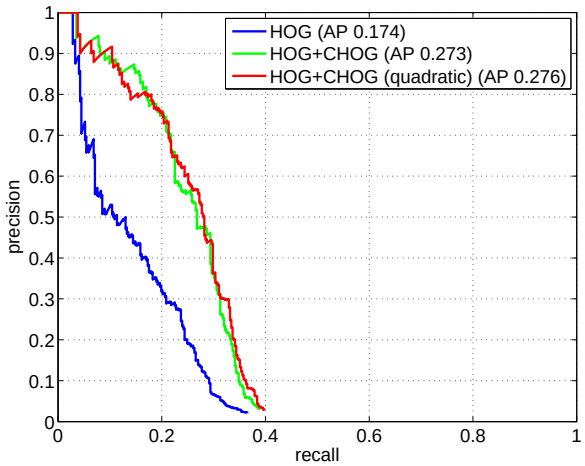
Experiments

UNIVERSITY OF LEEDS

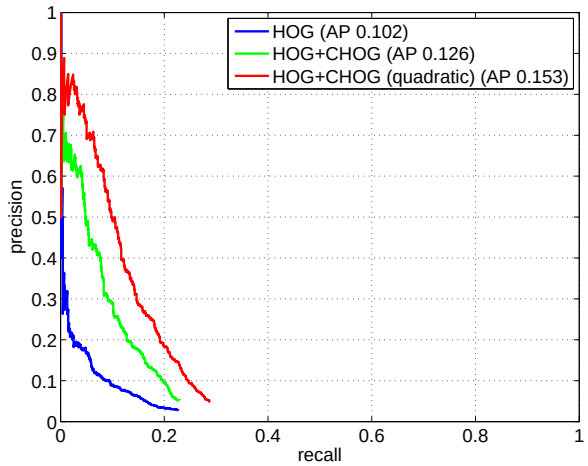
- PASCAL VOC 2006



(a) Bus



(b) Sheep



(c) Person

- Improvements in Average Precision up to 50% on various classes.

Conclusions

UNIVERSITY OF LEEDS

- Incorporated segmentation cues directly in the feature extraction process
- CHOG is an extension of HOG to segmentation and color information
- $\rightarrow$ captures stronger and coherent edges
- Substantial performance improvement over HOG for pedestrian and object detection
- State-of-the-art results on INRIAPerson

Thank you

UNIVERSITY OF LEEDS

Any questions?