Object Detection using Histograms of Oriented Gradients

Navneet Dalal, Bill Triggs

INRIA Rhône-Alpes

Grenoble, France

Thanks to Matthijs Douze for volunteering to help with the experiments

7 May, 2006

Pascal VOC 2006 Workshop

ECCV 2006, Graz, Austria

Talk Outline

- Current approaches
- Overall architecture
- Histogram of oriented gradient
  - Description of image encoding algorithm
- Multi-scale detection architecture
  - Fusion of detections at multiple scales and locations
- Key findings on Pascal VOC 2006
- Conclusions

Motivation

- Current Approaches
  - Dense feature sets based approaches
    - Papageorgiou & Poggio, 2000; Viola & Jones, 2001
  - Template or image fragments based approaches
    - Gavrila & Philomen, 1999; Vidal-Naquet & Ullman, 2003
  - Models based on key points
    - Leibe et al, 2005; Fergus et al, 2003
- Our Approach
  - Focus on creating robust encoding of images
  - Linear SVM as classifier on normalized image windows, is reliable & fast
  - Moving window based detector with non-maximum suppression over scale space

Overall Architecture

Learning Phase

Create normalised training data set

Encode images into feature vectors

Learn binary classifier

Object/Non-object decision

Detection Phase

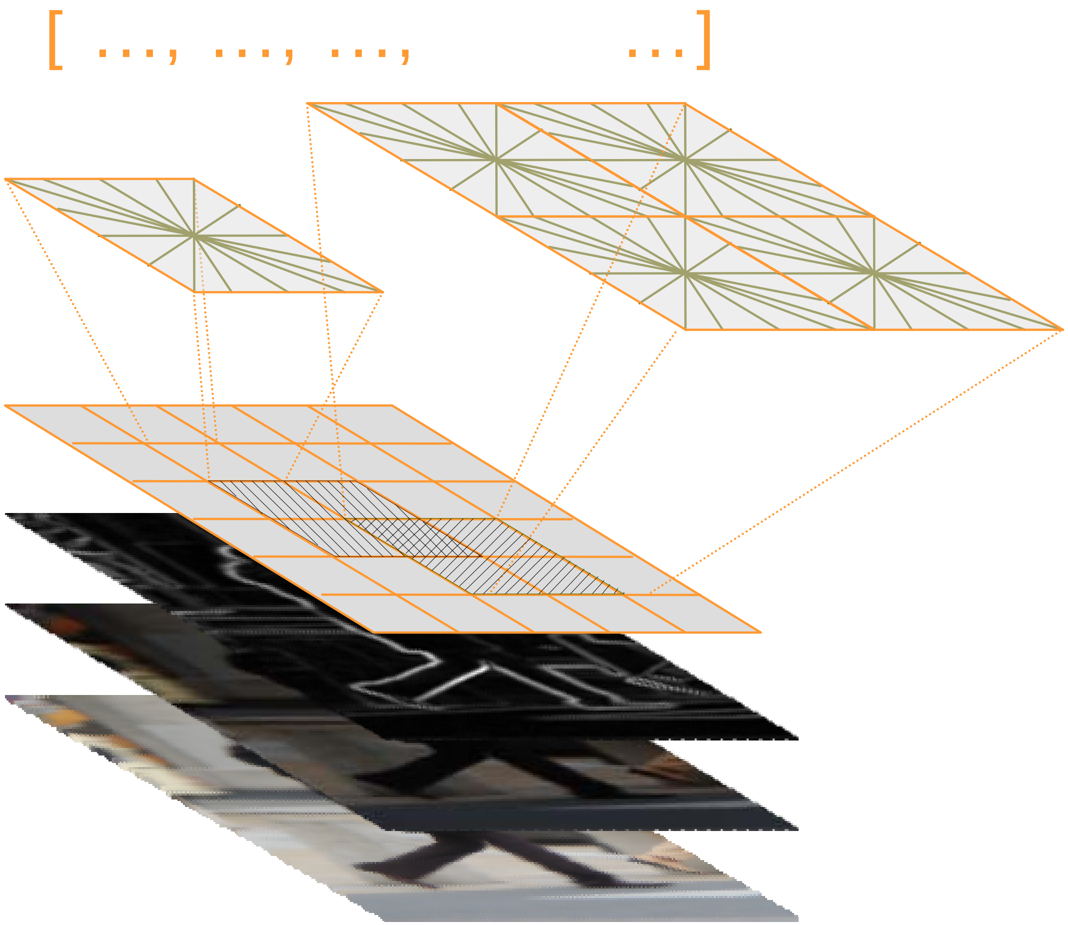
Scan image at all scales and locations

Run classifier to obtain object/non-object decisions

Fuse multiple detections in 3-D position & scale space

Object detections with bounding boxes

Descriptor Processing Chain



Object/Non-object

↑

Linear SVM

↑

Collect HOGs over detection window

↑

Contrast normalize over overlapping spatial cells

↑

Weighted vote in spatial & orientation cells

↑

Compute gradients

↑

Gamma compression

↑

Image Window

HOG Descriptors

HOG: Histogram of Oriented Gradients

Parameters

- Gradient scale
- Orientation bins
- Block overlap area

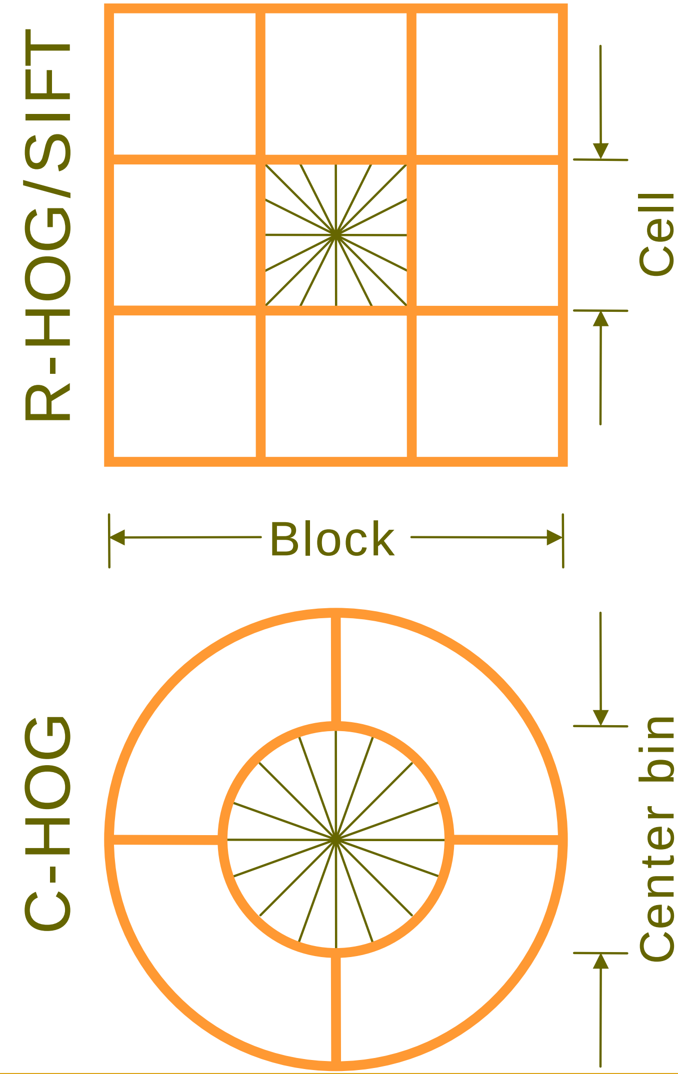
Schemes

- RGB or Lab, Color/gray-space
- Block normalization

L2-hys, $v \leftarrow v / \sqrt{\|v\|_2^2 + \varepsilon}$

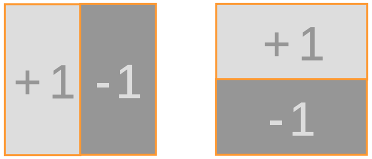
or

L1-sqrt, $v \leftarrow \sqrt{v / (\|v\|_1 + \varepsilon)}$



Lessons on HOGs

- No gradient smoothing, [1 0 -1] derivative mask
- Use gradient magnitude (no thresholding)
- Orientation voting into fine bins (20° wide bins)
- Spatial voting into coarser bins
- Strong local normalization
- Use overlapping blocks

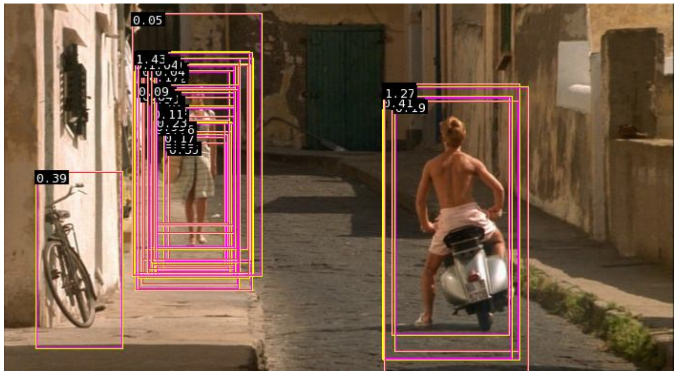


Fine grained features improve performance

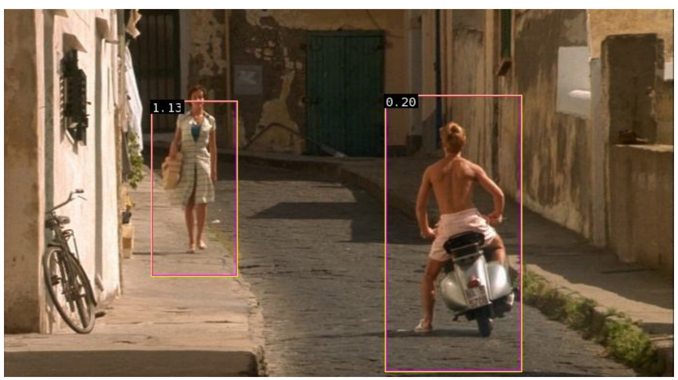
☺ Have 1-2 order lower false positives than other descriptors

☹ Slower than integral images of Viola & Jones, 2001

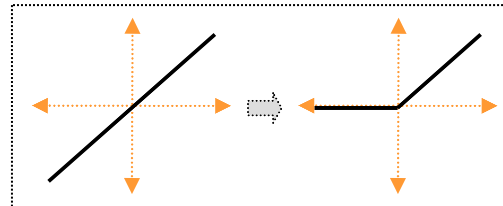
Multi-Scale Detection



After dense multi-scale scan of detection window

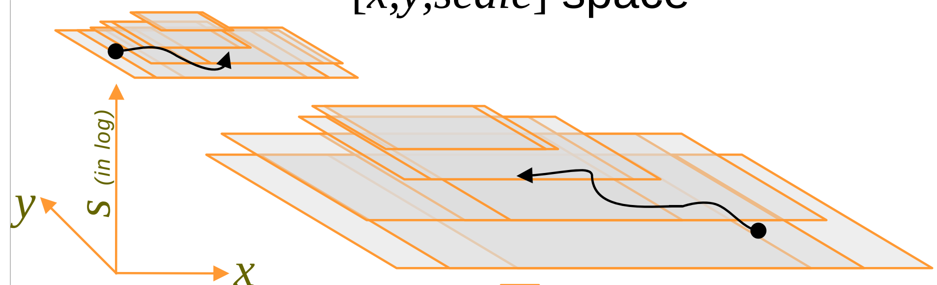


Final detections



Clip Detection Score

Map each detection to 3D [*x*,*y*,*scale*] space



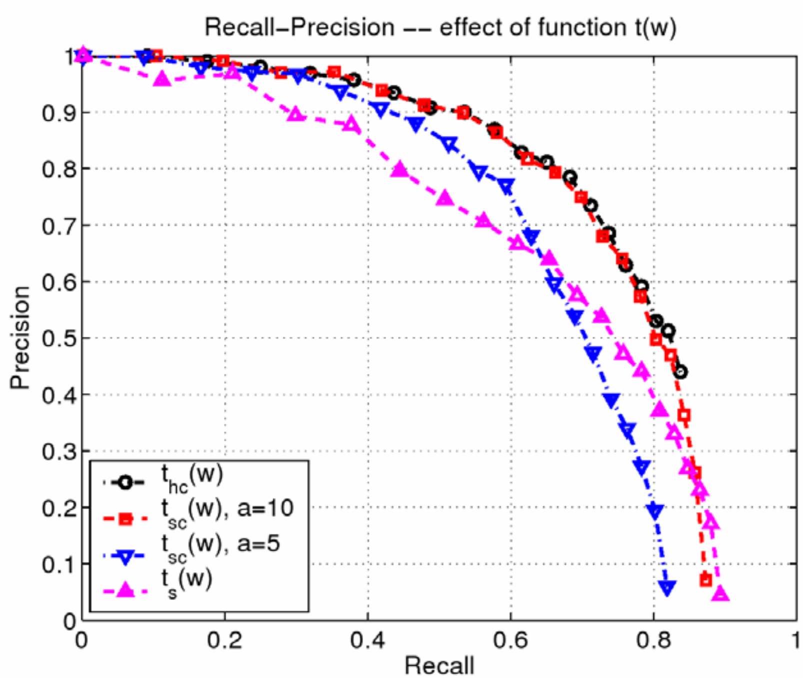
$$\mathrm{H}_i = [\exp(s_i)\sigma_x, \exp(s_i)\sigma_y, \sigma_s]$$

$$f(\mathbf{x}) = \sum_i^n w_i \exp\left(-\left\|(\mathbf{x}-\mathbf{x}_i)/\mathrm{H}_i^{-1}\right\|^2/2\right)$$

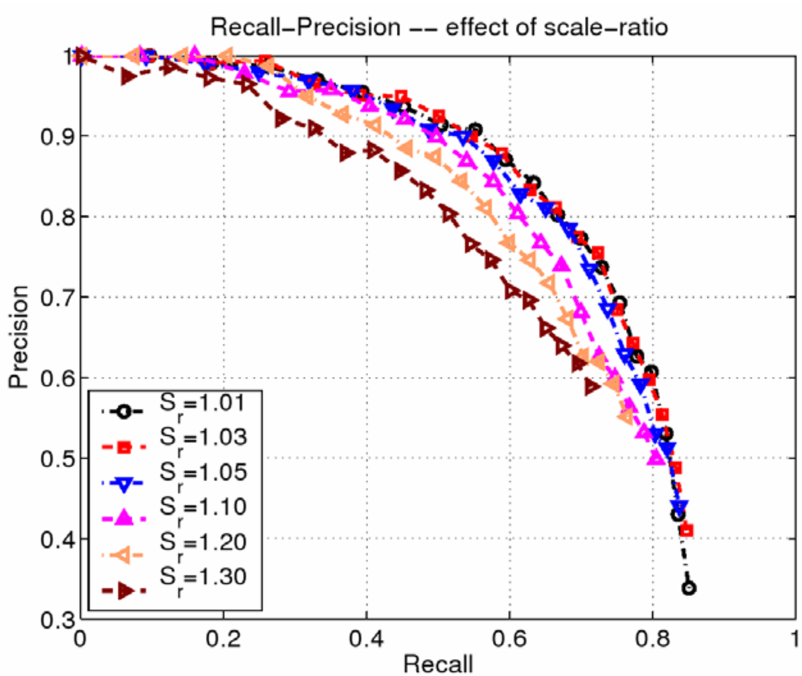
Apply robust mode detection, like mean shift

Performance Evaluation

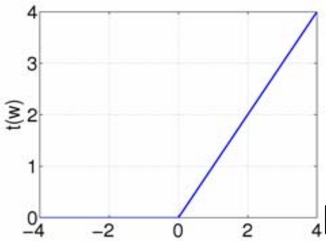
Transformation functions



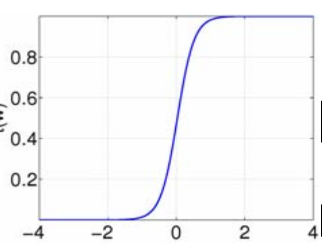
Scale-space pyramid steps



- p ... ter than sigmoid mapping
- e ... is very important



Hard clipping

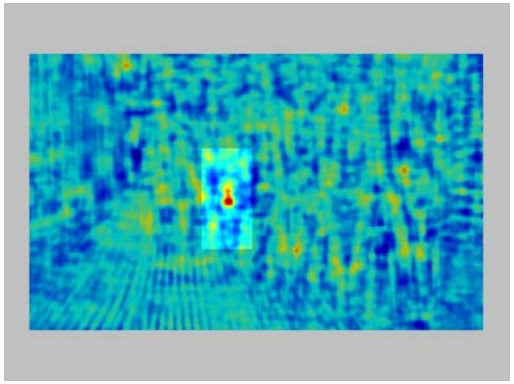


Sigmoid mapping

Effect of Smoothing

- Spatial smoothing proportional to window size performs best
- Relatively independent to smoothing across scales

Detector's normalized image window size

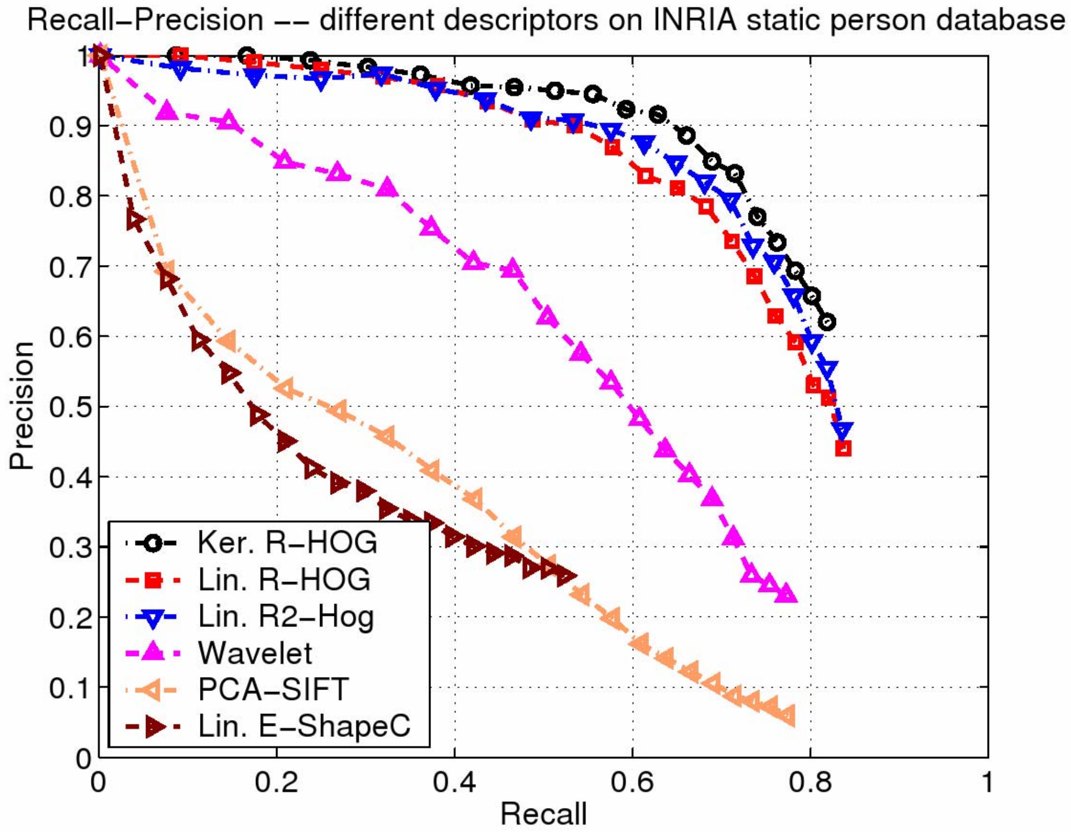


Detector's response at the given scale level

Overall robust non-maximum suppression is important

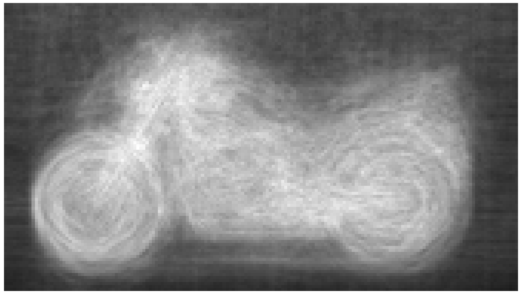
Overall Performance

Recall-precision on INRIA person database

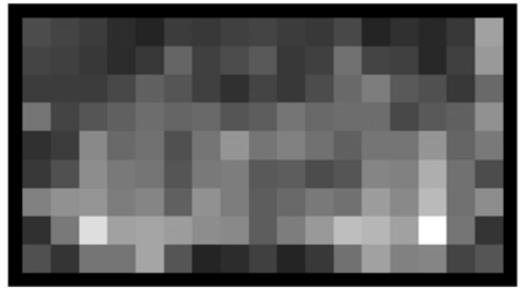


- R/C-HOG have 1-2 order lower false positives than other descriptors

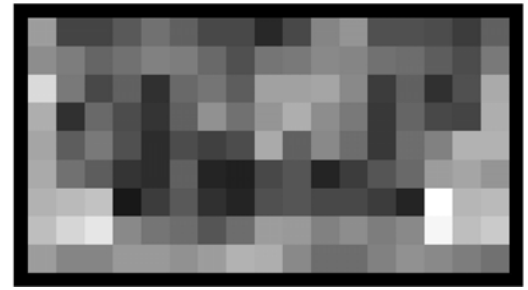
Descriptor Cues: Motorbikes



Average gradients



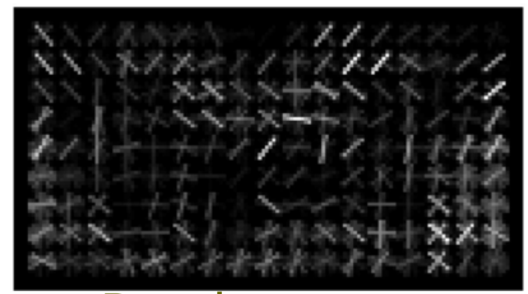
Weighted pos wts



Weighted neg wts

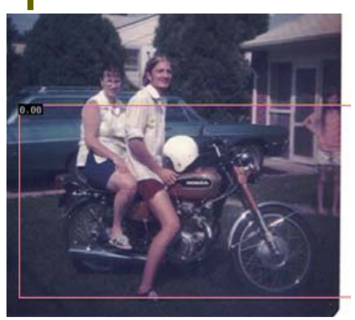
Input window

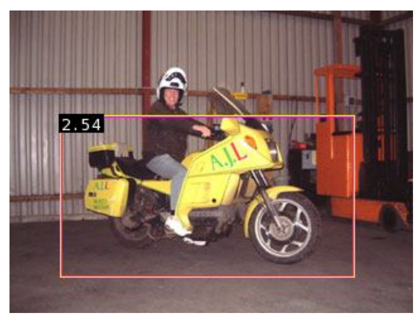
Dominant pos orientations



Dominant neg orientations

Detection Examples





Key Descriptor Parameters

Class	Window Size	Avg. Size	# of Orient-ation Bins	Orientat-ion Range	Gamma Compre-ssion	Normal-isation Method
Person	64×128	Height 96	9	0°–180°	$\sqrt{}$RGB	L2-Hys
Car	104×56	Height 48	18	0°–360°	$\sqrt{}$RGB	L1-Sqrt
Bus	120×80	Height 64	18	0°–360°	$\sqrt{}$RGB	L1-Sqrt
Motorbike	120×80	Width 112	18	0°–360°	$\sqrt{}$RGB	L1-Sqrt
Bicycle	104×64	Width 96	18	0°–360°	$\sqrt{}$RGB	L2-Hys
Cow	128×80	Width 56	18	0°–360°	$\sqrt{}$RGB	L2-Hys
Sheep	104×60	Height 56	18	0°–360°	$\sqrt{}$RGB	L2-Hys
Horse	128×80	Width 96	9	0°–180°	RGB	L1-Sqrt
Cat	96×56	Height 56	9	0°–180°	RGB	L1-Sqrt
Dog	96×56	Height 56	9	0°–180°	RGB	L1-Sqrt

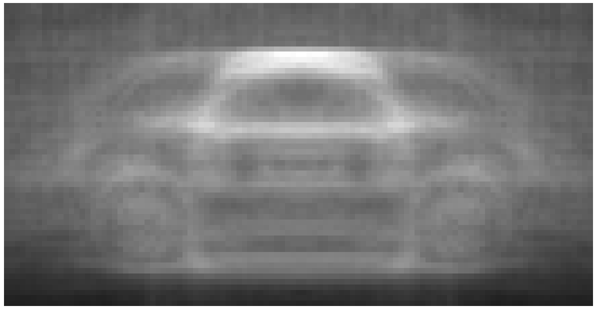
Conclusions

- Contributions
  - Robust feature encoding for object detection
  - Gives good performance for variety of object classes
  - Real time detection is possible

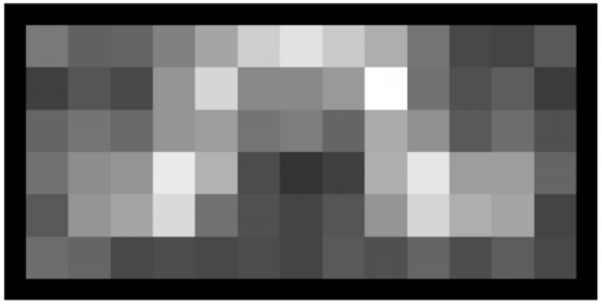
- Future Work
  - Part based detector for handling partial occlusions
  - Incorporate texture and color descriptors into the framework
  - One single optimization phase based on AdaBoost to learn most relevant descriptors

Thank You

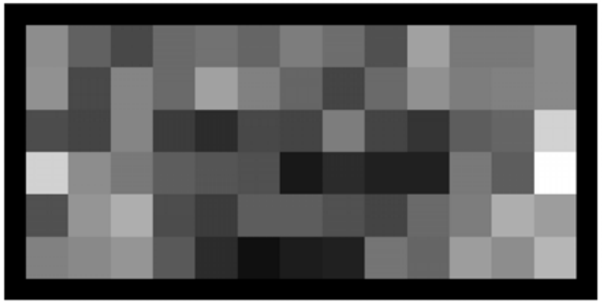
Descriptor Cues: Cars



Average gradients

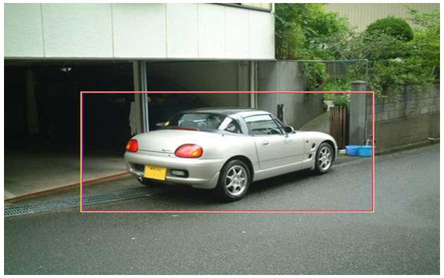


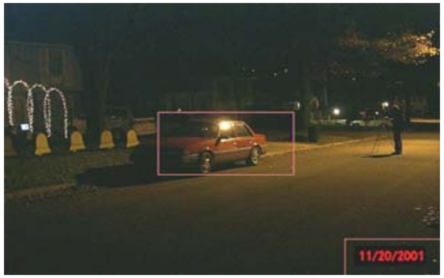
Weighted pos wts



Weighted neg wts

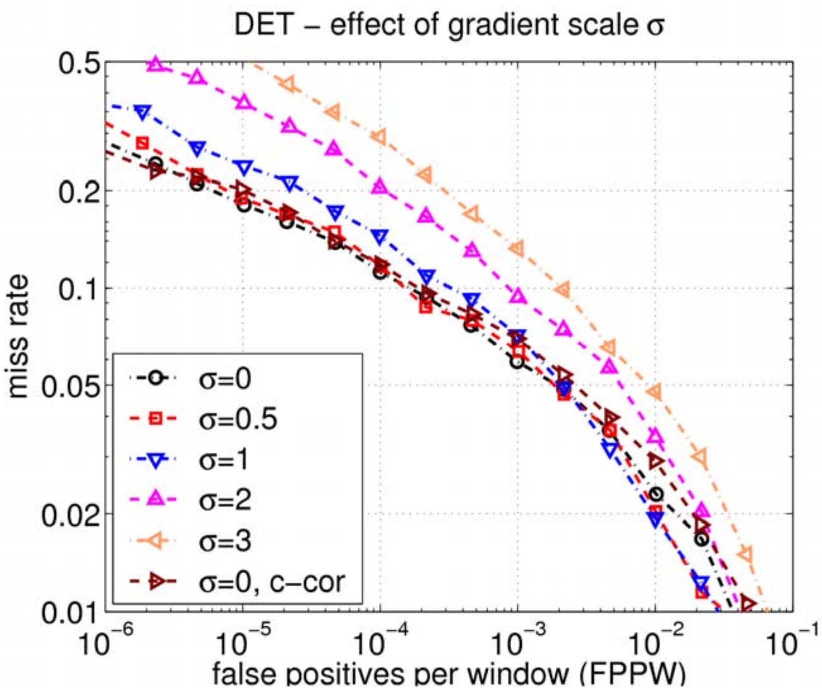
Detection Examples



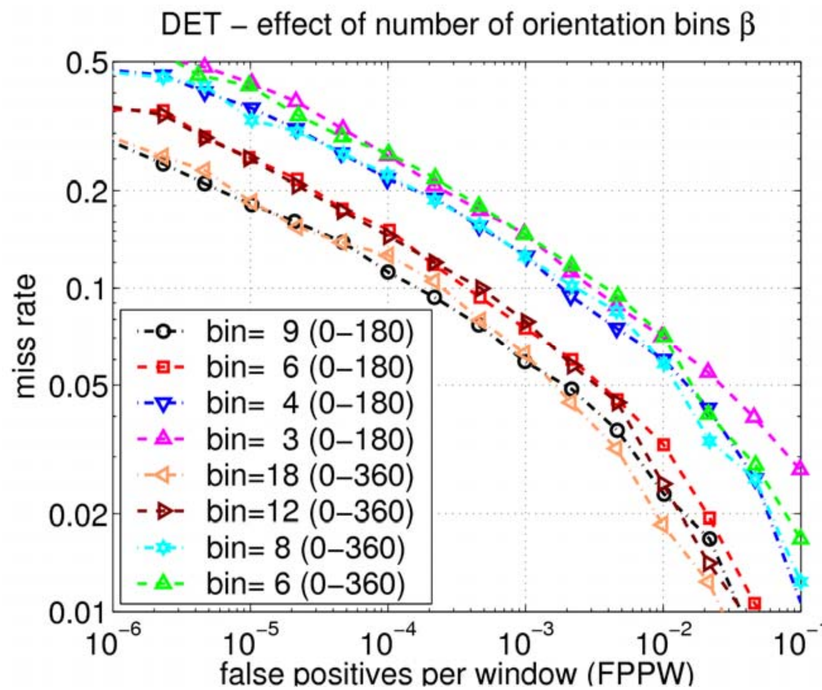


Effect of Parameters

Gradient smoothing, σ



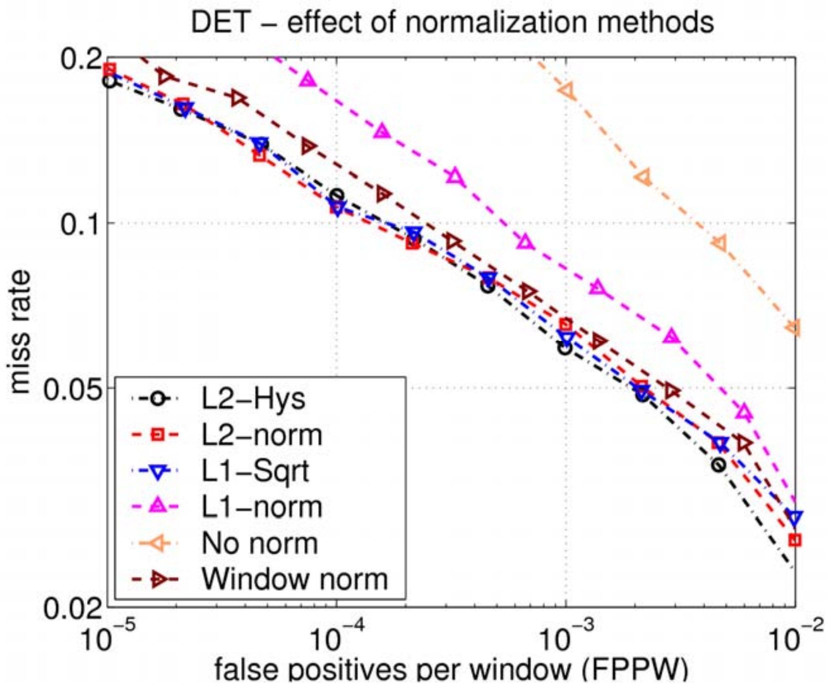
Orientation bins, β



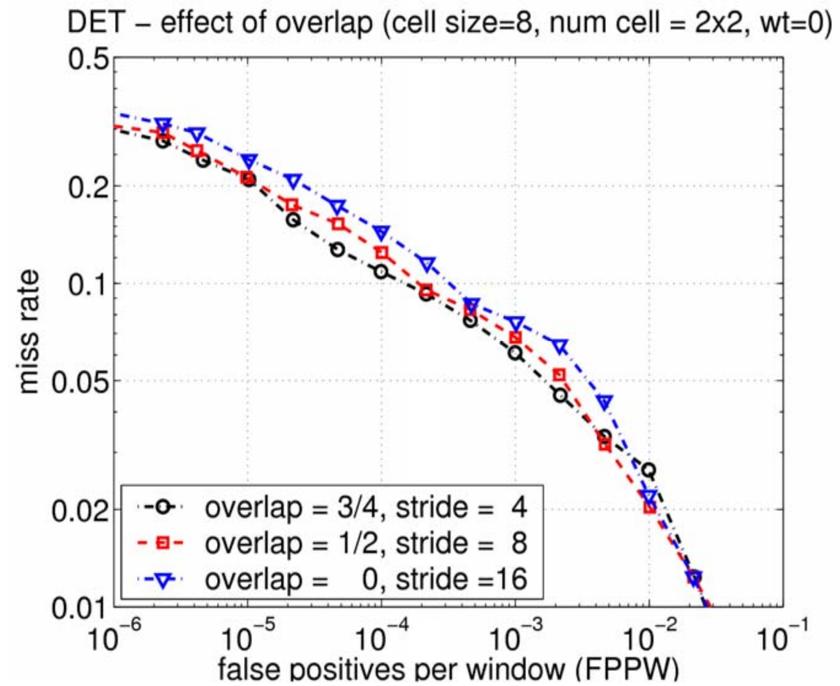
- Using simple smoothed gradients and many orientations helps!
- Gradient scale 3$\rightarrow$ 0 $\Rightarrow$ false positives drop by 10 times
- Orientation bins 45°$\rightarrow$ 20° $\Rightarrow$ false positives drop by 10 times

… Continued

Normalization method

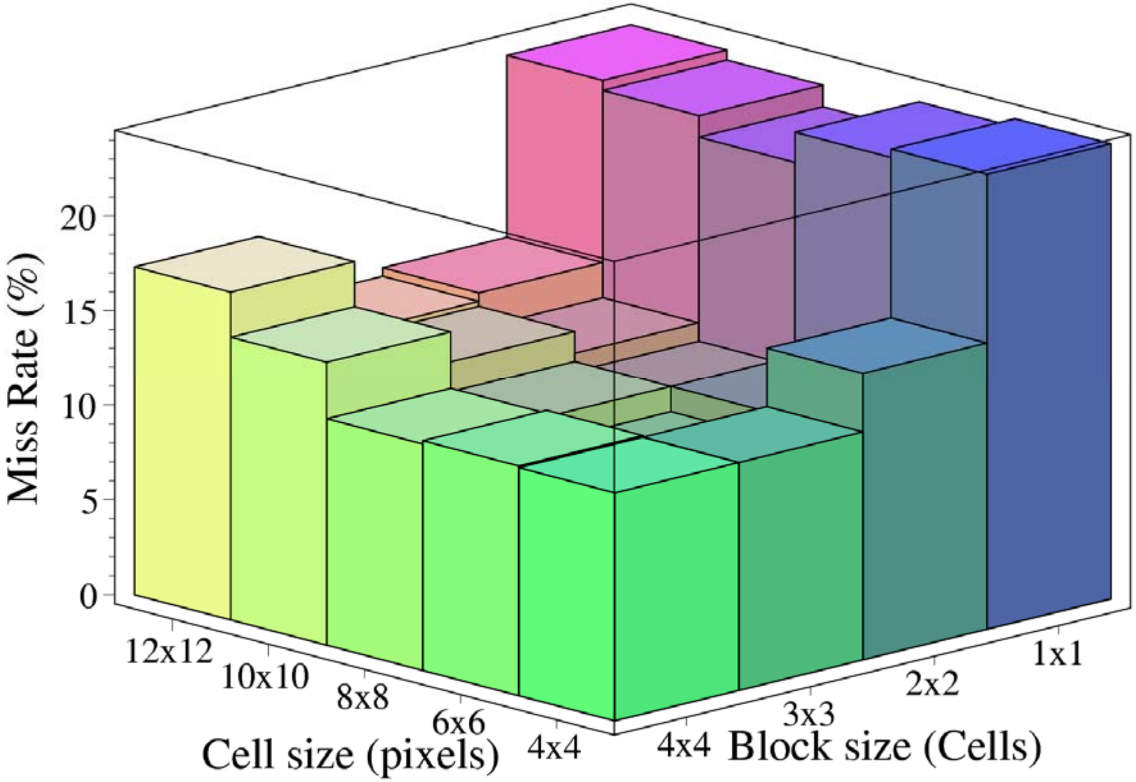


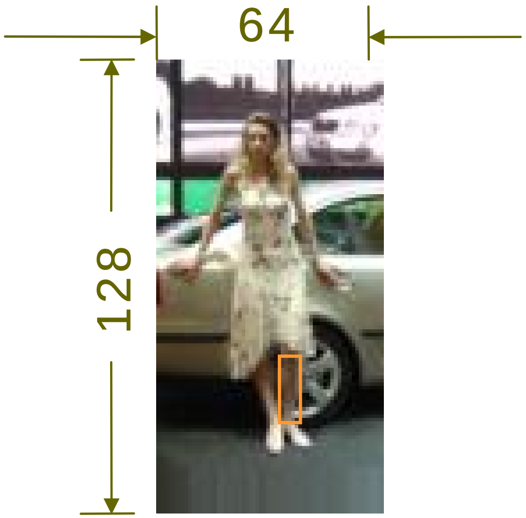
Block overlap



- Strong local normalization is essential
- Overlapping block increases performance, but descriptor size increases

Effect of Block and Cell Size

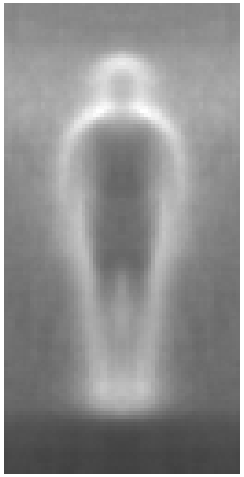


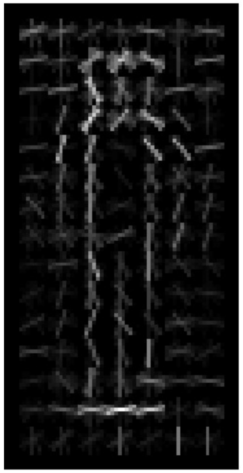


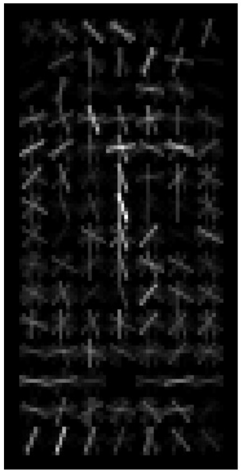
- Trade off between need for local spatial invariance and need for finer spatial resolution

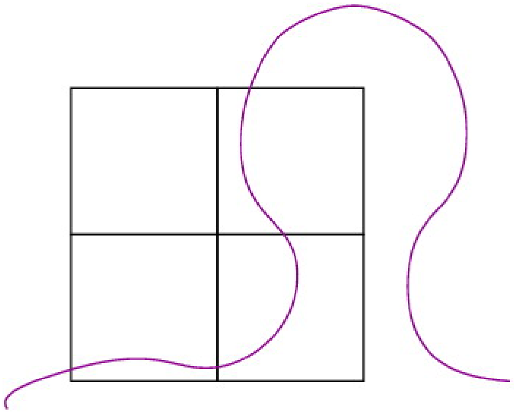
Descriptor Cues: Persons


Input example


Average gradients


Weighted pos wts


Weighted neg wts


Outside-in weights

- Most important cues are head, shoulder, leg silhouettes
- Vertical gradients inside a person are counted as negative
- Overlapping blocks just outside the contour are most important