# Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs

INRIA Rhône-Alpes

Grenoble, France

Funding: aceMedia, LAVA, Pascal Network

# Introduction

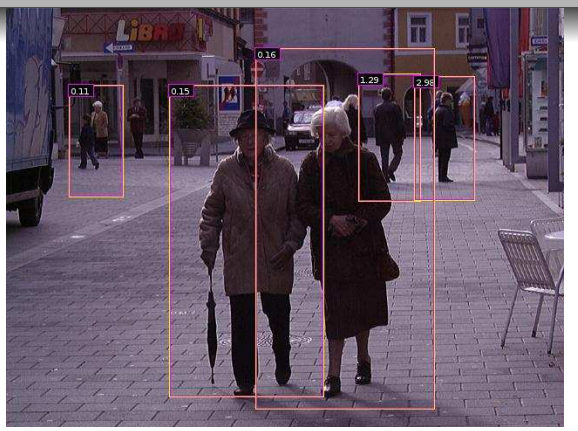
Detect & localize upright people
in static images

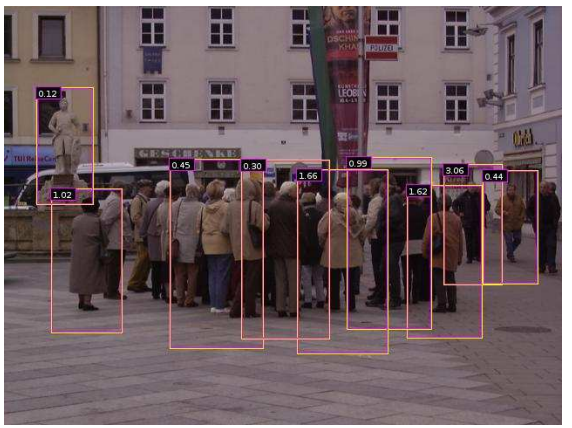
## Challenges

- Wide variety of articulated poses
- Variable appearance/clothing
- Complex backgrounds
- Unconstrained illumination
- Occlusions, different scales

## Applications

- Pedestrian detection for smart cars
- Film & media analysis
- Visual surveillance





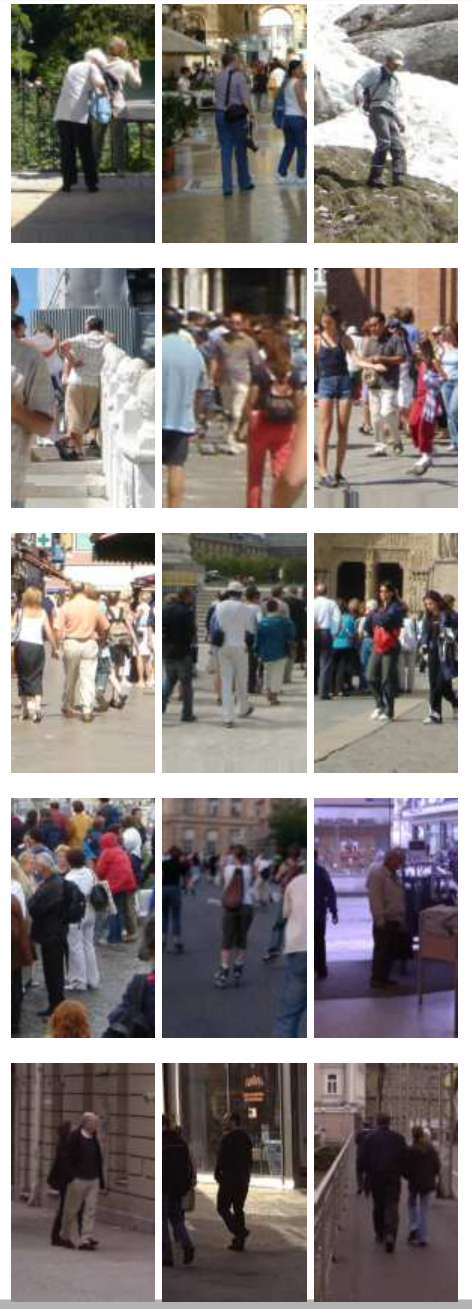
# Approach & Data Set

- We focus on building robust *feature sets*
- Classifier is just linear SVM on normalized image windows, is reliable & fast
- Moving window based detector with non-maximum suppression over scale-space

Data set available
*http://pascal.inrialpes.fr/data/human/*

Data Set

| *Train* | *Test* |
|---|---|
| 614 positive images | 288 positive images |
| 1218 negative images | 453 negative images |
| 1208 positive windows | 566 positive windows |
| Overall 1774 human annotations + reflections | |



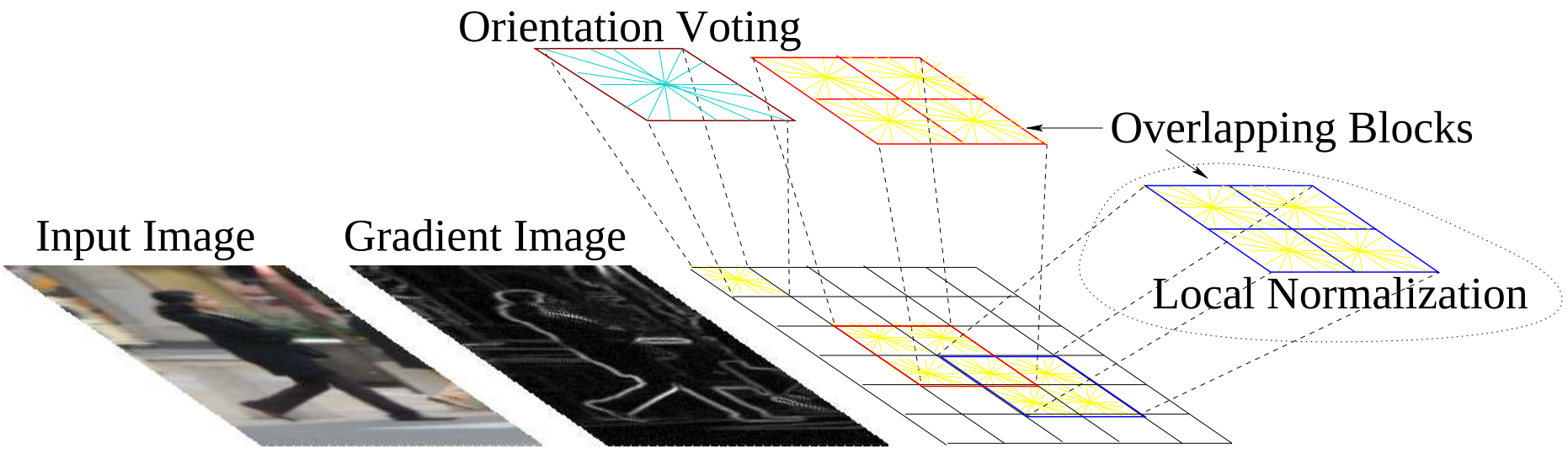
# Feature Sets

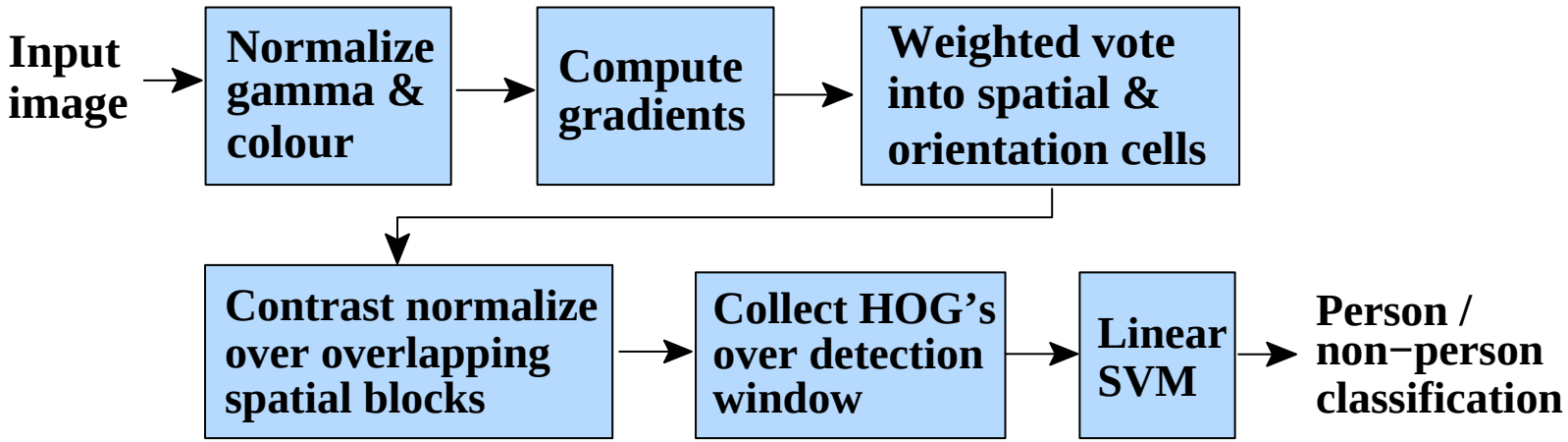
- Haar Wavelets + SVM: Papageorgiou & Poggio (2000), Mohan et al (2001), DePoortere et al (2002)
- Rectangular differential features + adaBoost: Viola & Jones (2001)
- Parts based binary orientation position histograms + adaBoost: Mikolajczyk et al (2004)
- Edge templates + nearest neighbor: Gavrila & Philomen (1999)
- Dynamic programming: Felzenszwalb & Huttenlocher (2000), Ioffe & Forsyth (1999)

---

- **Orientation histograms**: c.f. Freeman et al (1996), Lowe (1999)
- Other descriptors:
  - Shape contexts: Belongie et al (2002)
  - PCA-SIFT: Ke and Sukthankar (2004)

# Processing Chain



# HOG Descriptors

## Parameters

- Gradient scale
- Orientation bins
- Percentage of block overlap

## Schemes

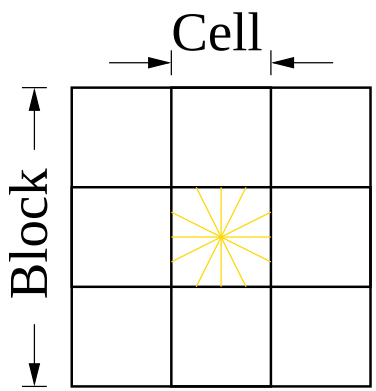
- RGB or Lab, color/gray-space
- Block normalization,

$L2$-norm, $\mathbf{v} \rightarrow \mathbf{v}/\sqrt{\|\mathbf{v}\|_2^2 + \epsilon^2}$

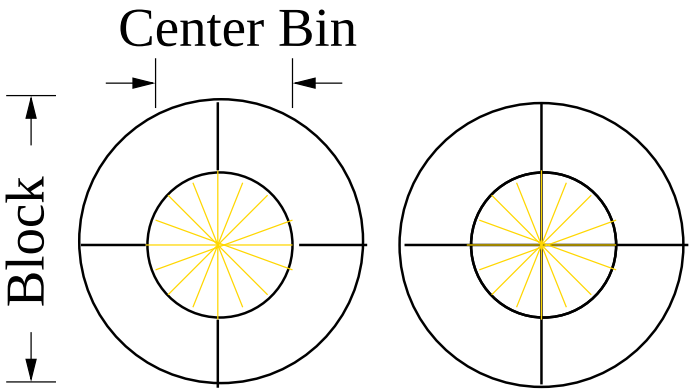
or

$L1$-norm, $\mathbf{v} \rightarrow \sqrt{\mathbf{v}/(\|\mathbf{v}\|_1 + \epsilon)}$

## R-HOG/SIFT



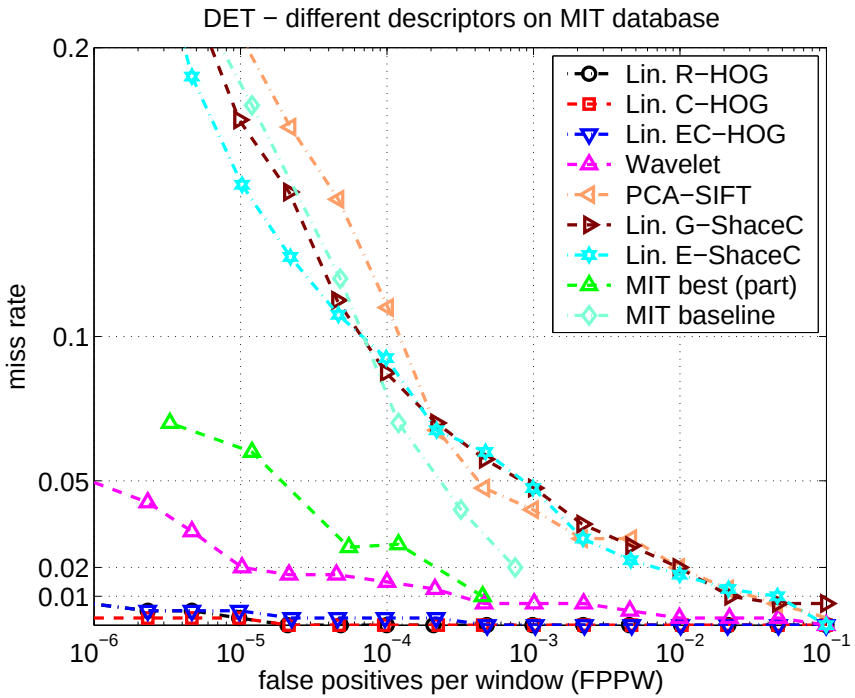

## C-HOG



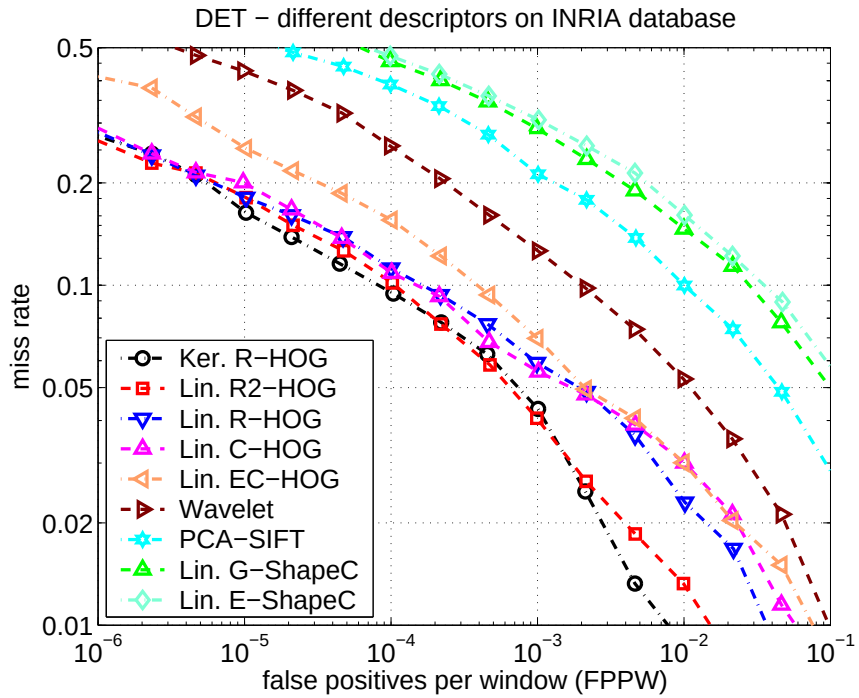

Radial Bins, Angular Bins

# Performance

## MIT pedestrian database



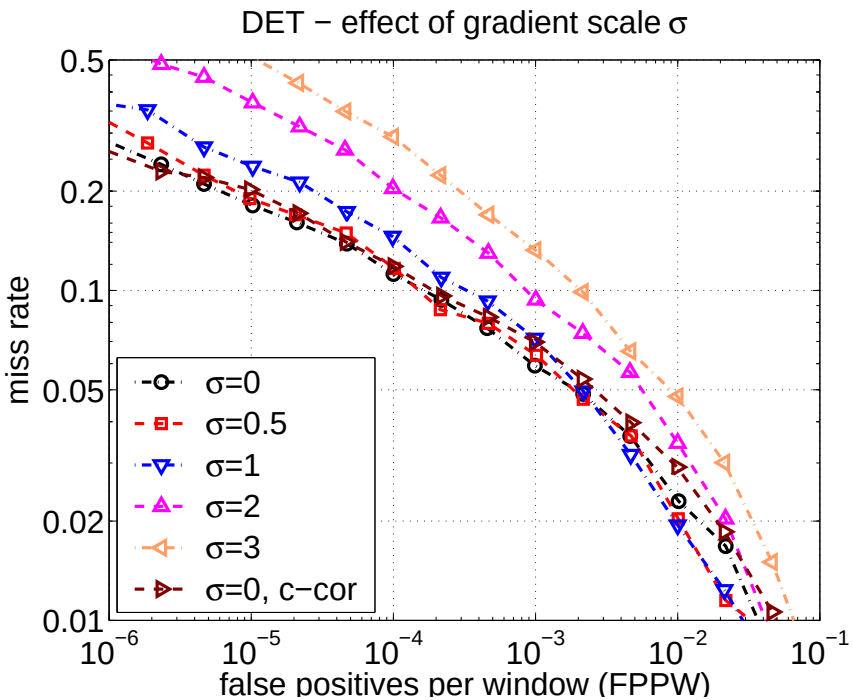

## INRIA person database



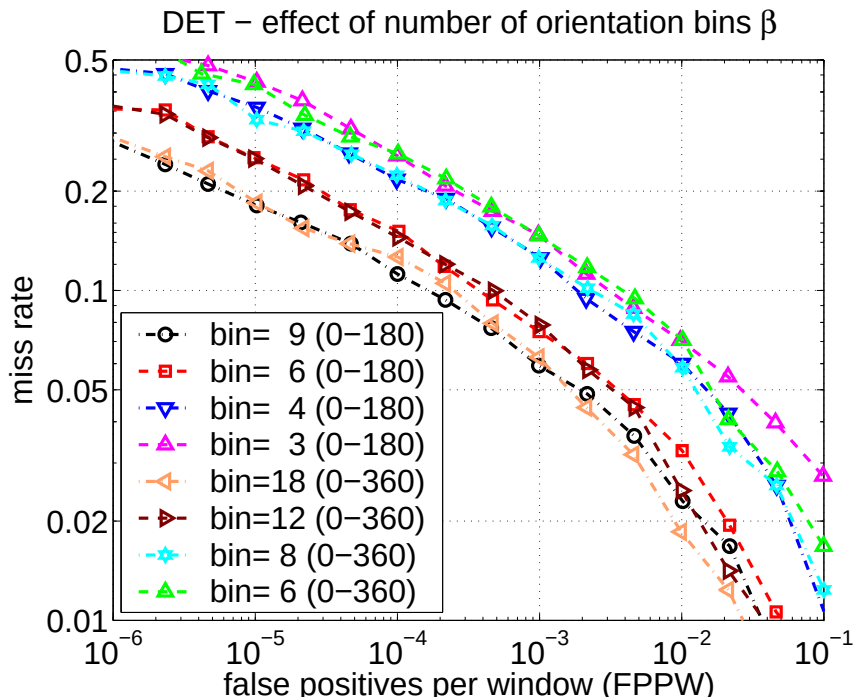

- R/C-HOG give near perfect seperation on MIT database
- Have 1-2 orders of magnitude lower false positives than other descriptors

# Gradient Smoothening & Orientation Bins

## Gradient scale, $\sigma$




## Orientation bins, $\beta$

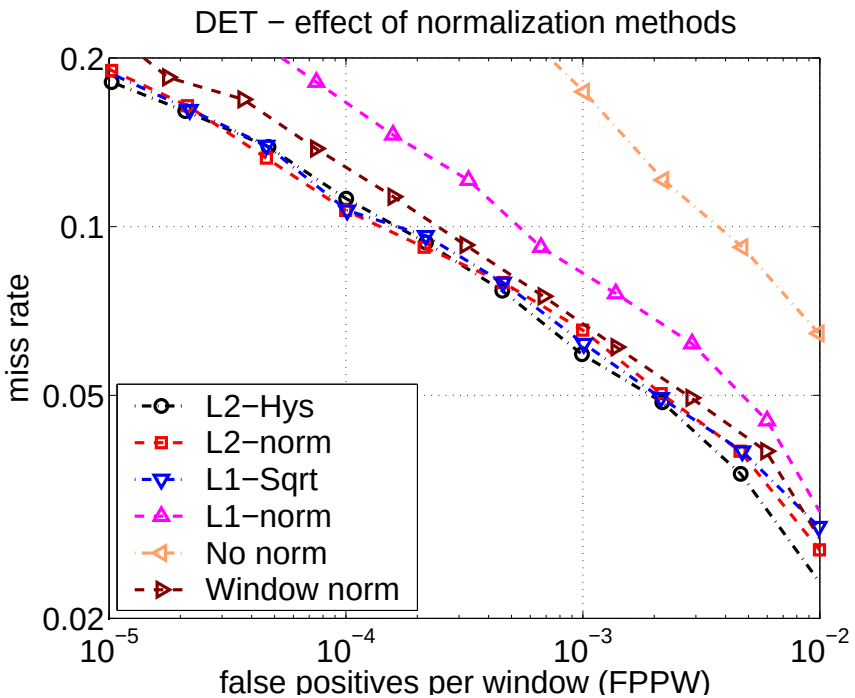



Using simple smoothed gradients & many orientations helps!

- Reducing gradient scale from 3 to 0 decreases false positives by 10 times
- Increasing orientation bins from 4 to 9 decreases false positives by 10 times

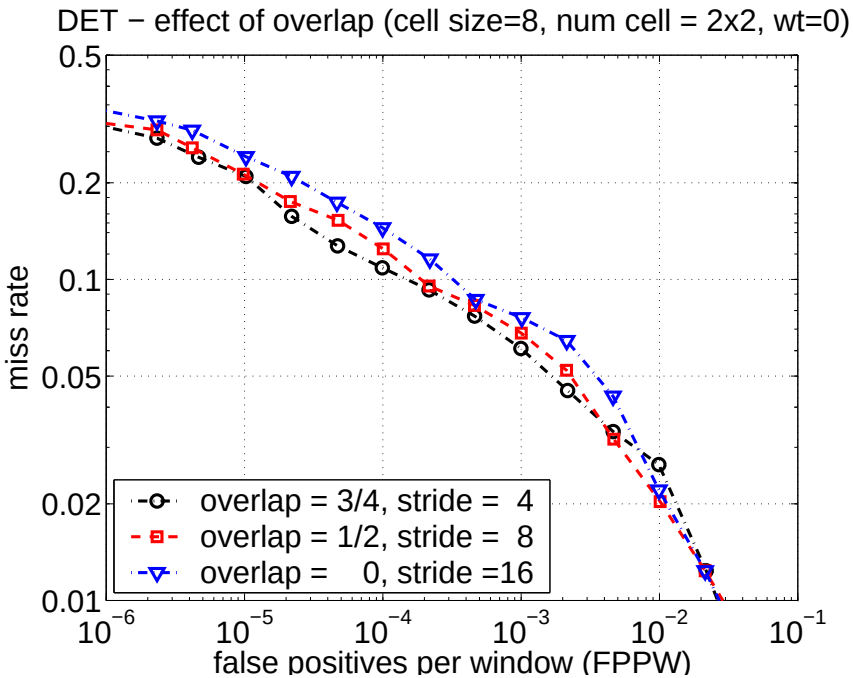
# Normalization Method & Block Overlap

## Normalization method



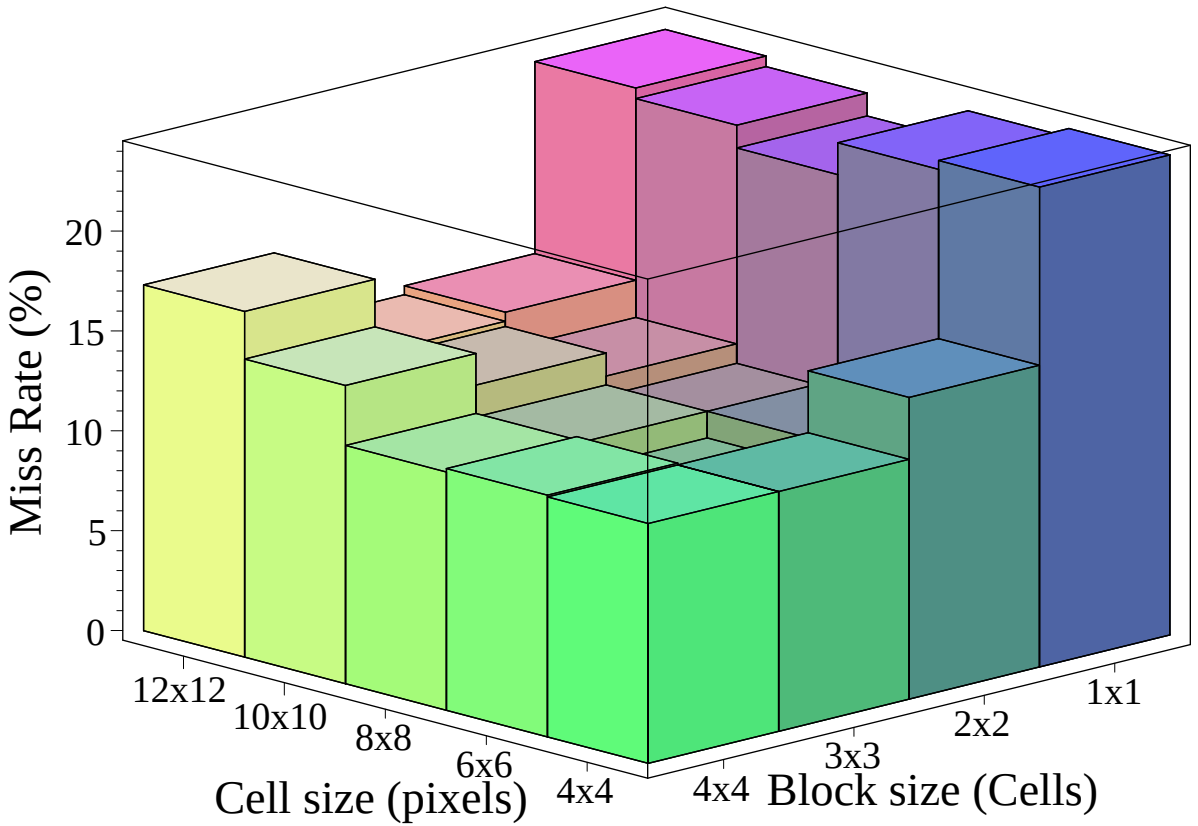

Strong local normalization is essential

## Block overlap




Overlapping blocks improves performance, but descriptor size increases

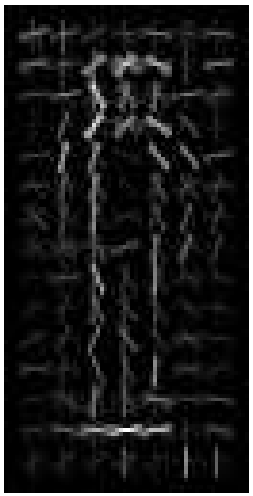
# Effect of Block & Cell Size



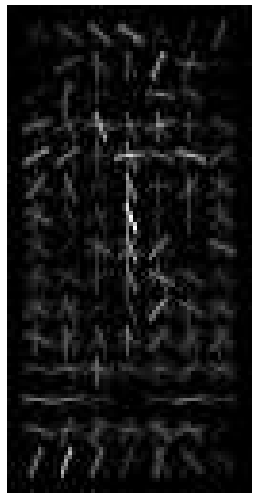

Trade off between need for local spatial invariance and need for finer spatial resolution

# Descriptor Cues

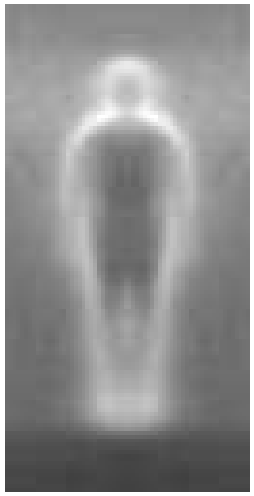
input image



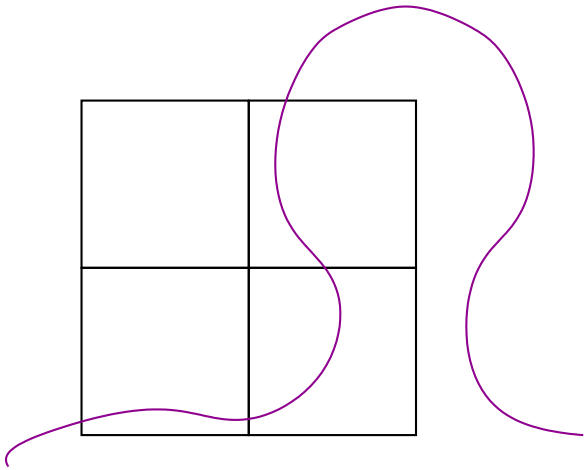
weighted pos wts



weighted neg wts



avg. grad



outside in block

- The most important cues are head, shoulder, leg silhouettes
- Vertical gradients inside the person count as negative
- Overlapping blocks those just outside the contour are the most important

# Conclusions

Fine grained features improve performance

- No gradient smoothning, $[-1, 0, 1]$ derivative mask
- Use gradient magnitude (no thresholding)
- Orientation voting into fine bins
- Spatial voting into coarser bins
- Strong local normalization
- Overlapping normalization blocks

- A general object classifier
- Also works well for other classes
- Linear SVM is reliable & fast, but not optimal
- Human detection: 90% at $10^{-4}$ false positives per window

# Demo

No temporal smoothning