Tutorial outline

- Overview (this)
- Image representation (60 mins, 9:15 - 10:30)
  - motivation, local features, global features, break
- Learning (90 mins, 10:30 - 12:30)
  - discriminative models, tea-break, generative models, break
- Object detection and recognition (90 mins, 12:30 - 2:00)
  - Dalal & Triggs, lunch-break, PASCAL challenge, *poselets* and their applications, tea-break
- Cross-modal search (60 mins, 2:30 - 3:30)

lunch-break 60 mins, break 15 mins, tea-break 20-30 mins

ICVGIP 2012, IIT Bombay

The eight Indian Conference on Computer Vision, Graphics and Image Processing

Tutorial

Histogram of Oriented Gradients

Subhransu Maji
Toyota Technological Institute at Chicago

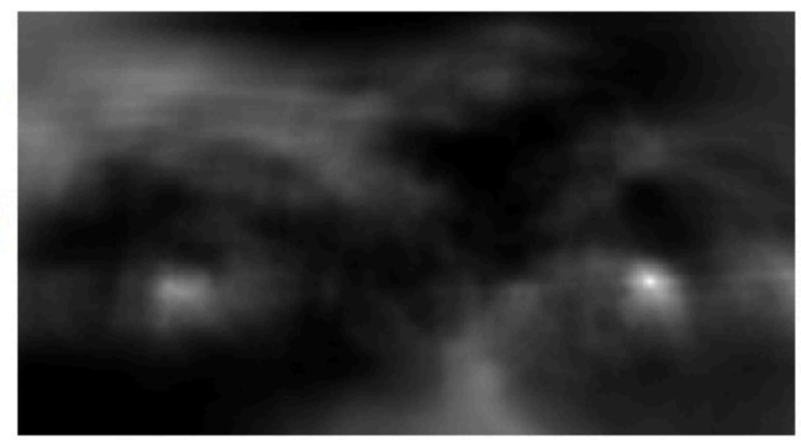
Slides courtesy Ross Girshick

(Review?) Template matching

- Consider matching with image patches
  - What could go wrong?

template

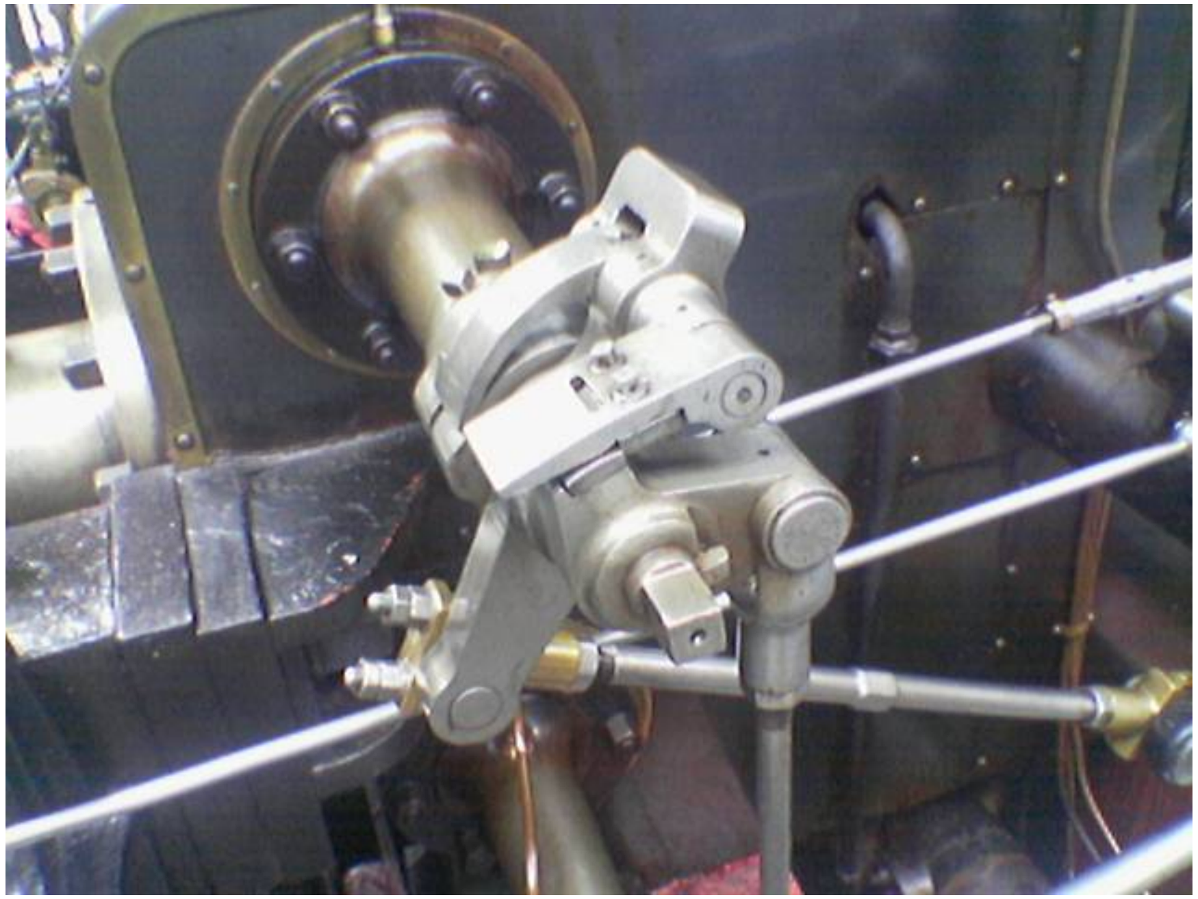
image



match quality
e.g., cross correlation

What is a feature map?

- Any transformation of an image into a new representation
- Example: transform an image into a binary edge map



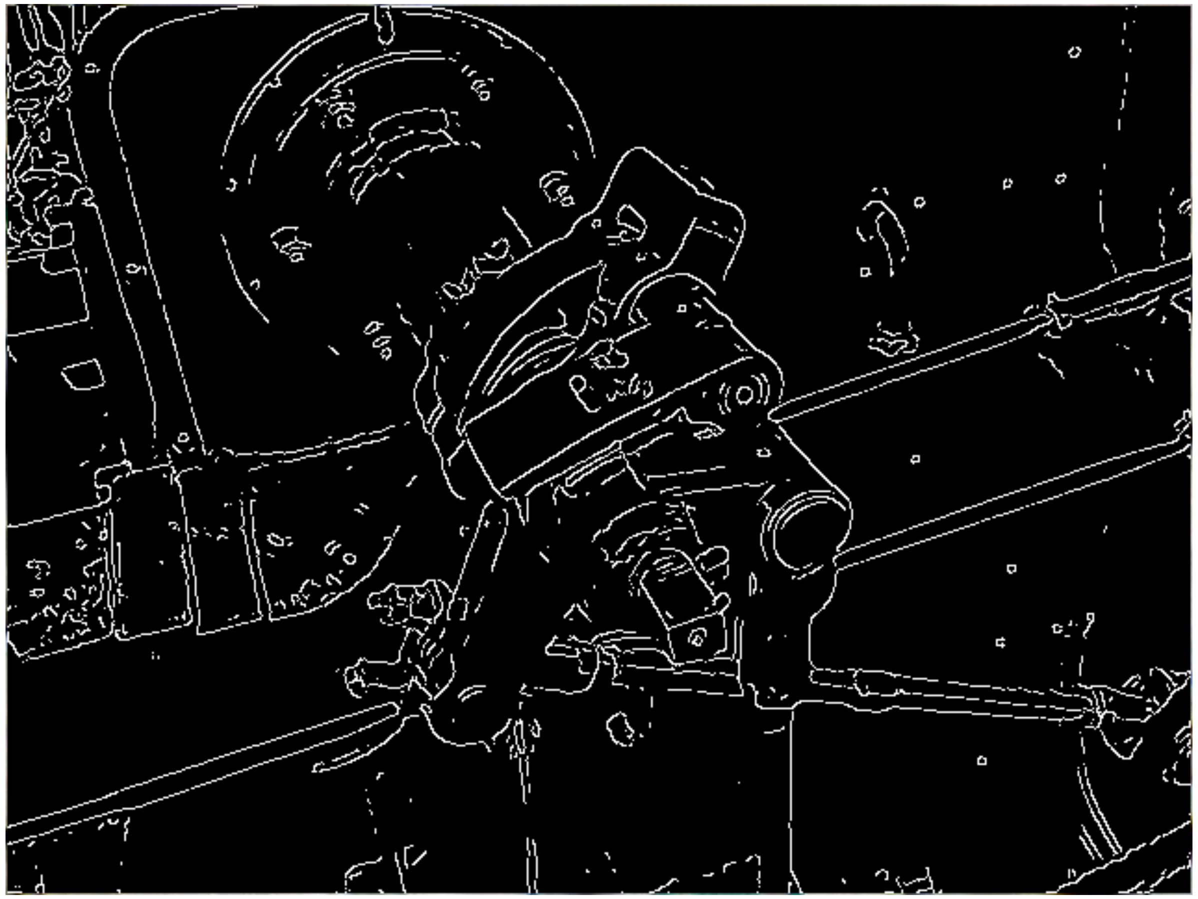


Image source: wikipedia

Feature map goals

- Introduce invariance
    - Bias, gain, nonlinear transformations
    - Small deformations

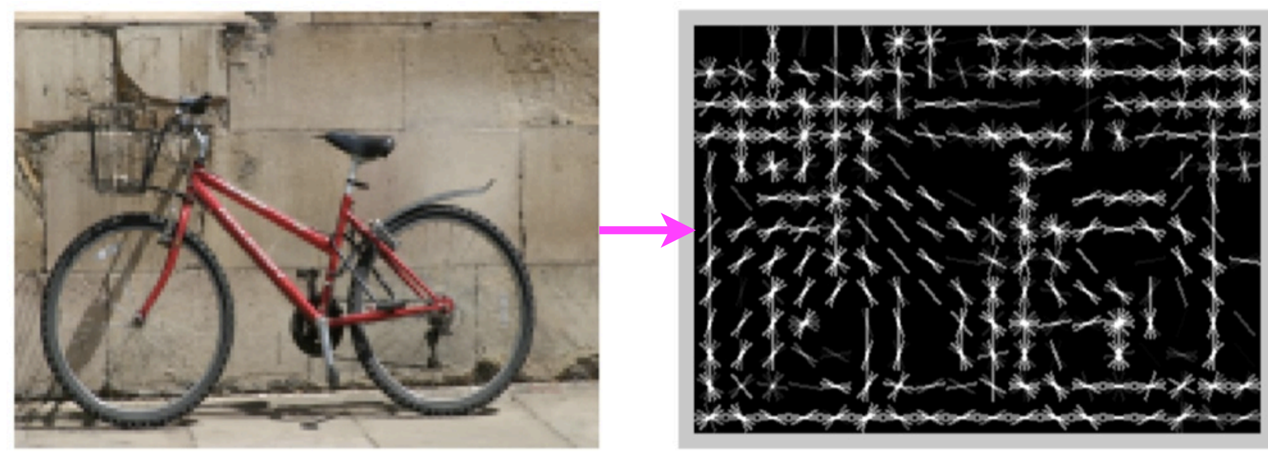
Figure 1.3: Variation in appearance due to a change in illumination

- Preserve larger scale spatial structure

Image: [Fergus05]

Histograms of Oriented Gradients (HOG)

- Introduce invariance
  - Bias / gain / nonlinear transformations
    - bias: gradients / gain: local normalization
    - nonlinearity: clamping magnitude, orientations
  - Small deformations
    - spatial subsampling
    - local "bag" models



- References
  - "Histograms of oriented gradients for human detection." N. Dalal and B. Triggs, CVPR 2005.
  - "Finding people in images and videos." N. Dalal, Ph.D. Thesis, Institut National Polytechnique de Grenoble, 2006.

HOG feature computation

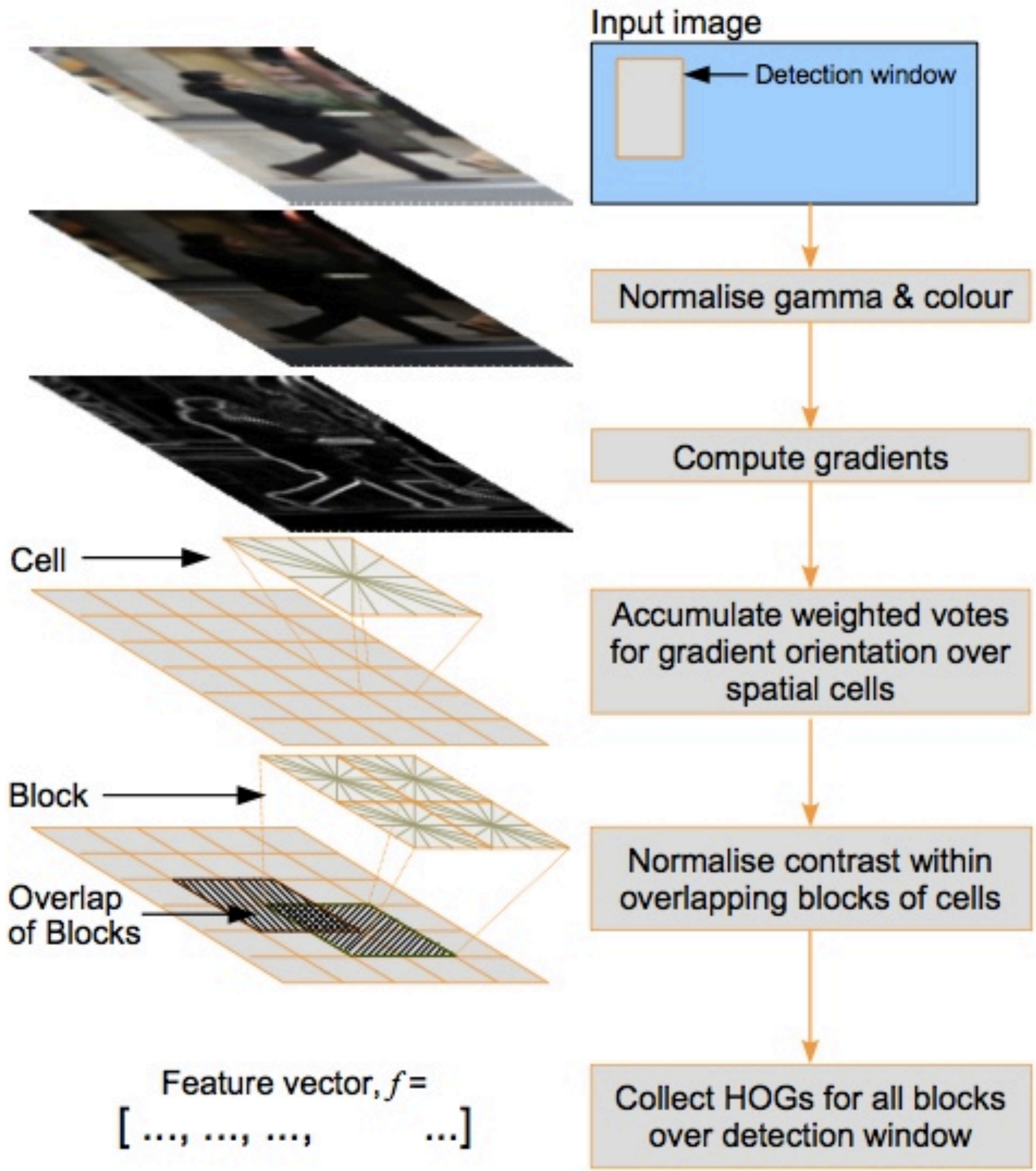
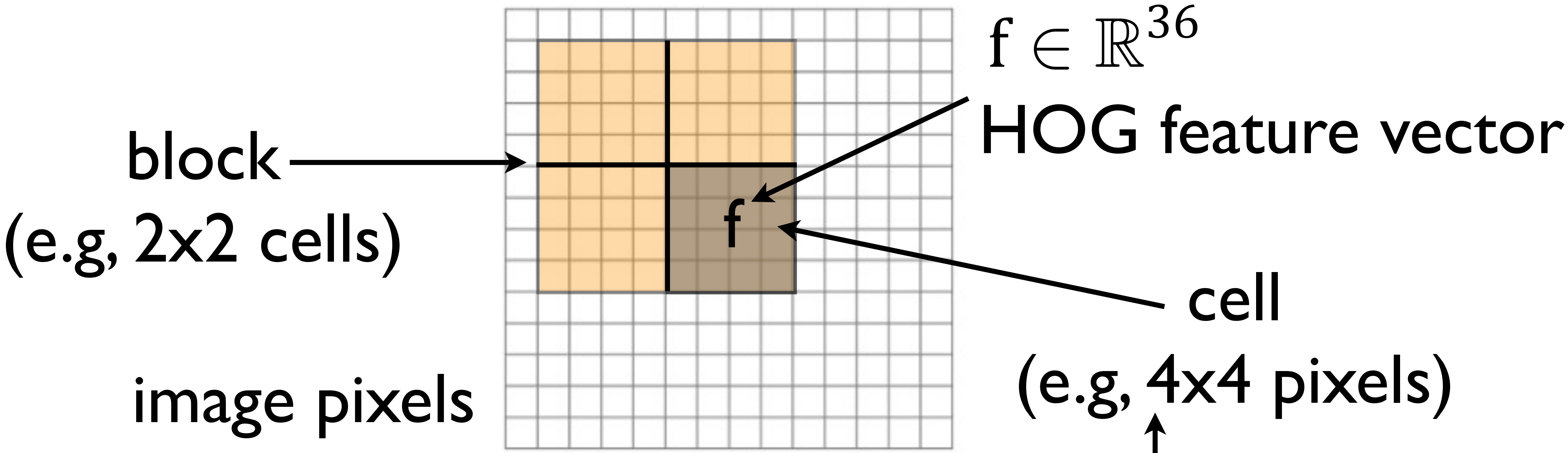
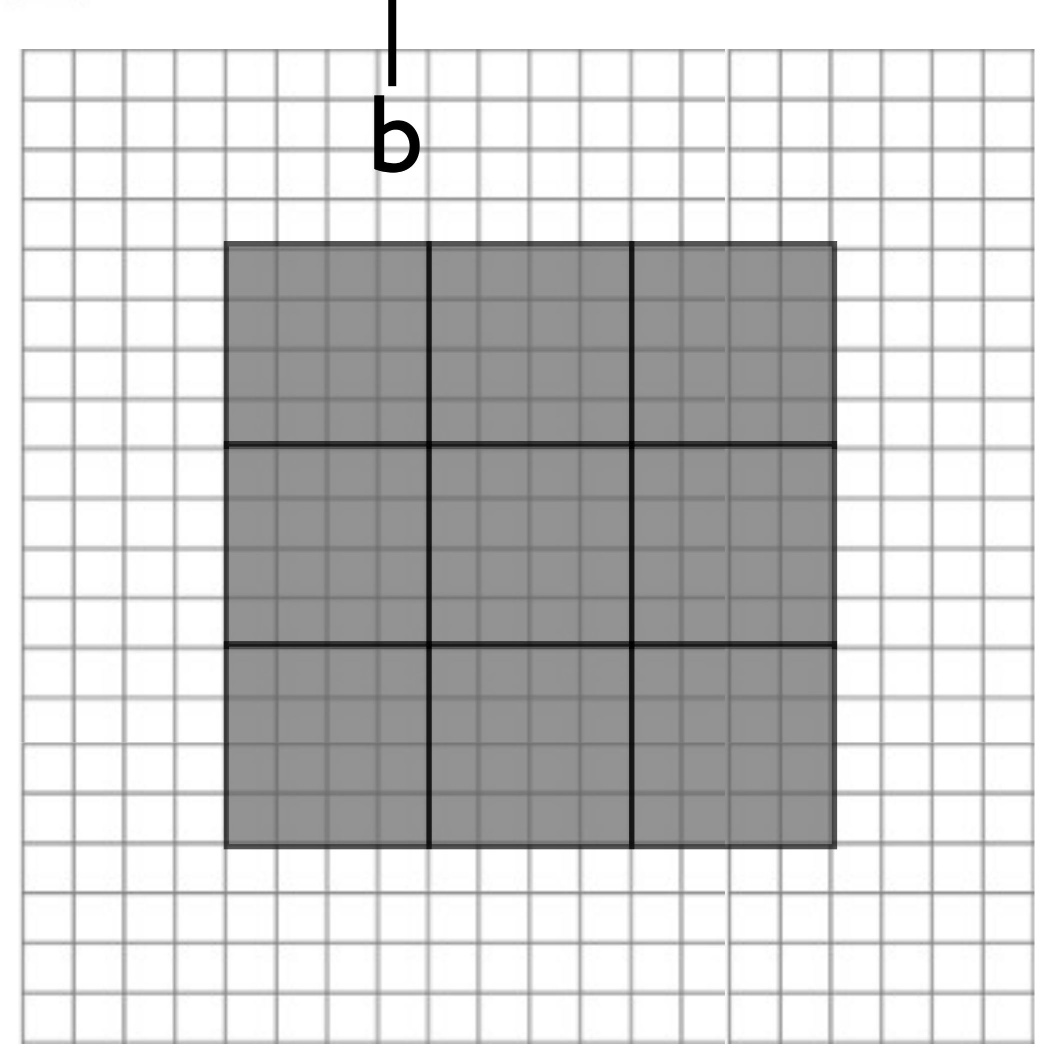


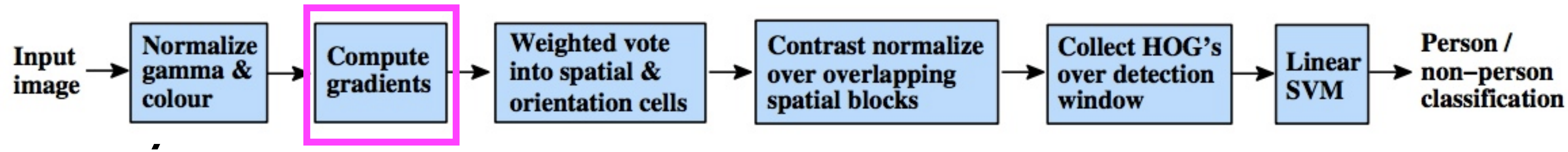
Image: [Dalal06]

HOG terminology



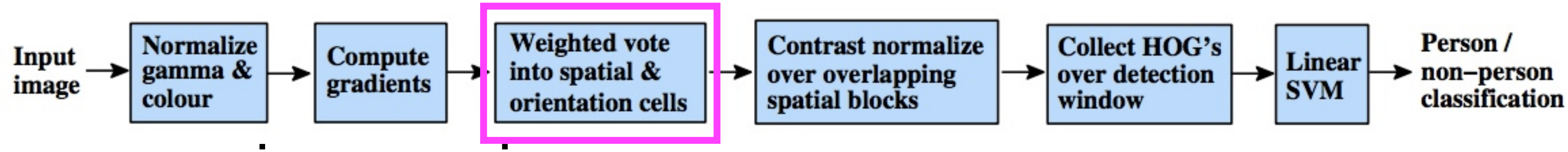


- Original image: H x W x 3
- Feature map: H’ x W’ x D
  - For example
    - H’ = floor(H/b) - 2
    - W’ = floor(W/b) - 2
    - D = 36

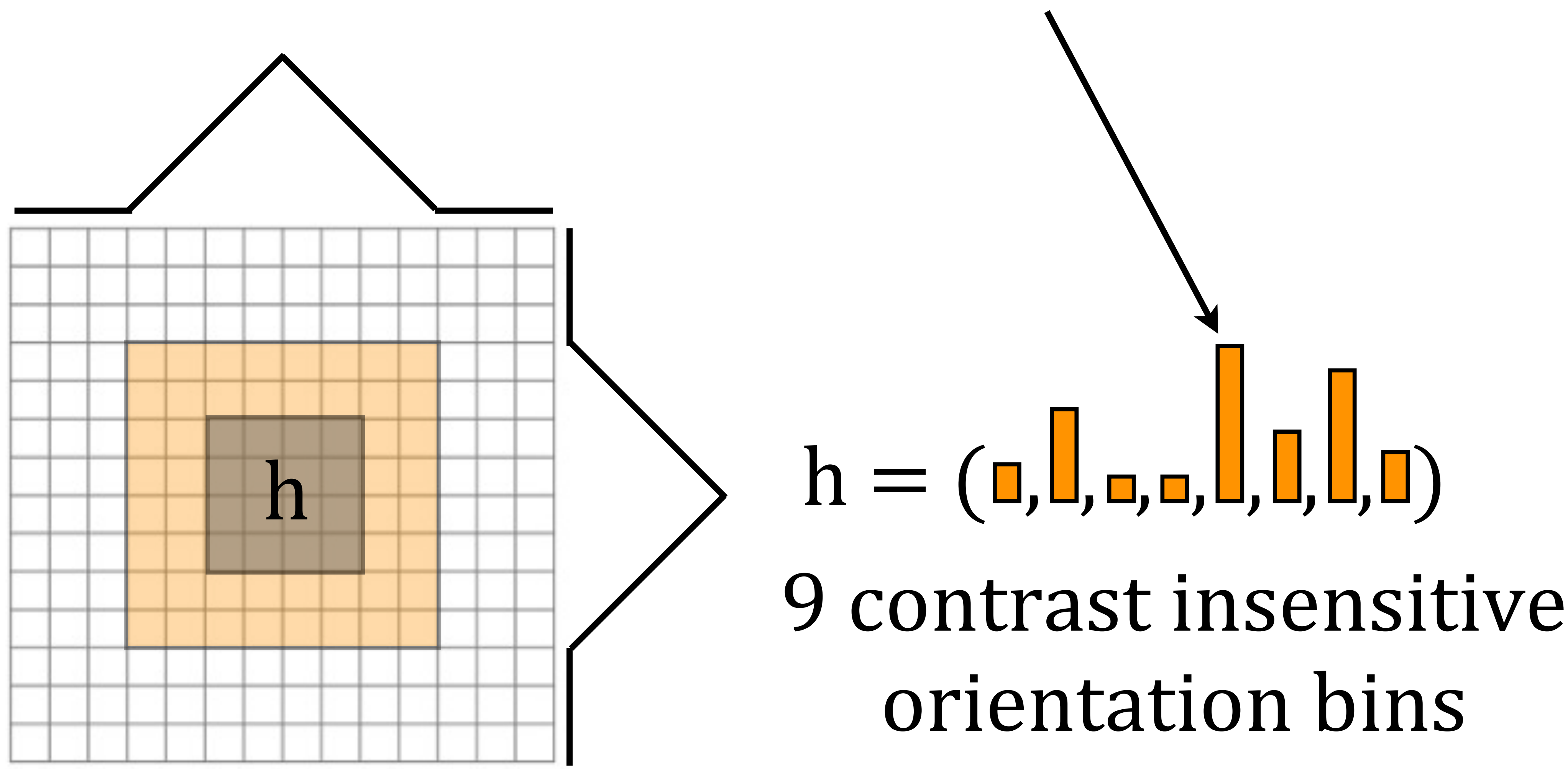


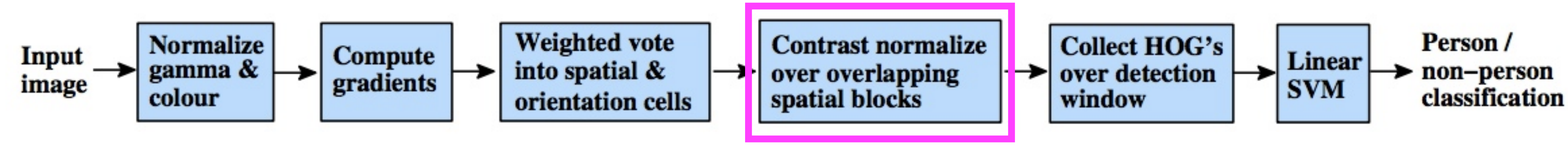
- (1, 0, -1) centered filter works best
- Alternatives: uncentered, cubic corrected, Sobel, etc.

- Discrete approx. to partial derivatives
  - $I_x = I[x+1, y] - I[x-1, y]$
  - $I_y = I[x, y+1] - I[x, y-1]$

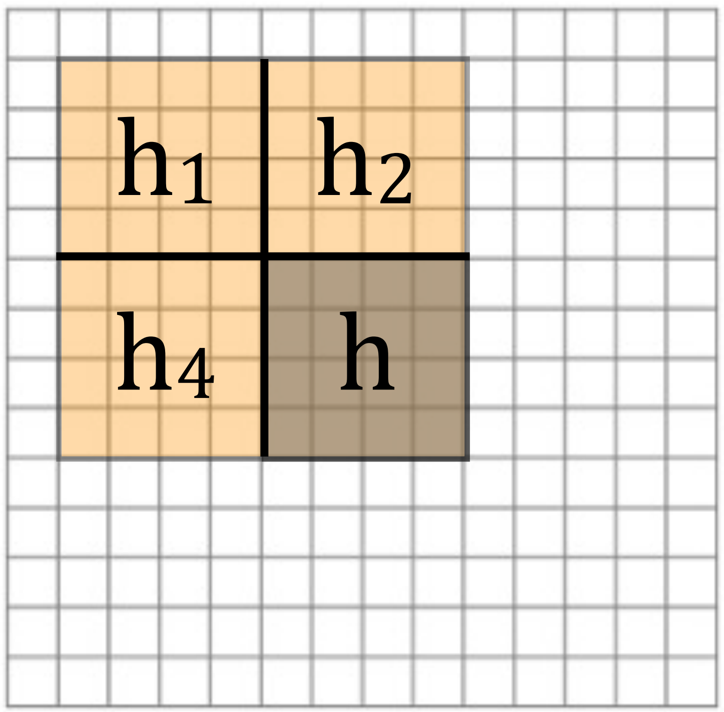


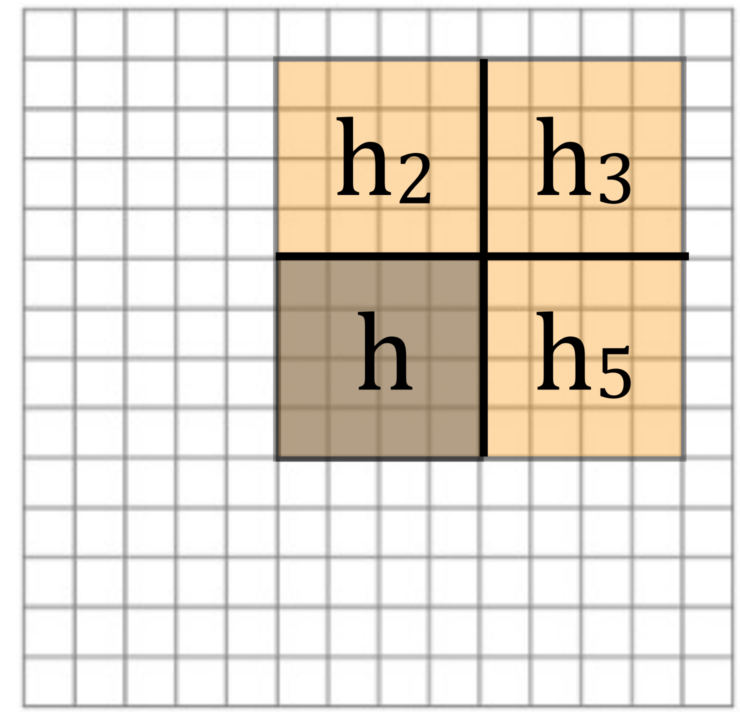
- Gradient magnitude: $m = \| (I_x, I_y) \|$
- Gradient orientation: $o = \tan^{-1}(I_y / I_x)$
- Quantize orientation; vote into bin (weighted)





- Local contrast normalization and clipping





$$h^1 = \max[\, 0.2, h/||(h; h_1; h_2; h_4)|| \,]$$

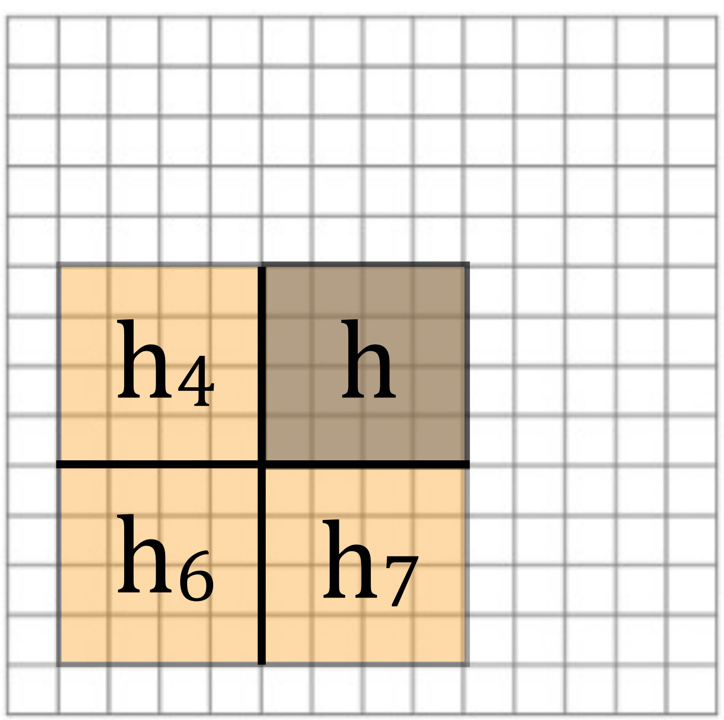
$$h^2 = \max[\, 0.2, h/||(h; h_2; h_3; h_5)|| \,]$$

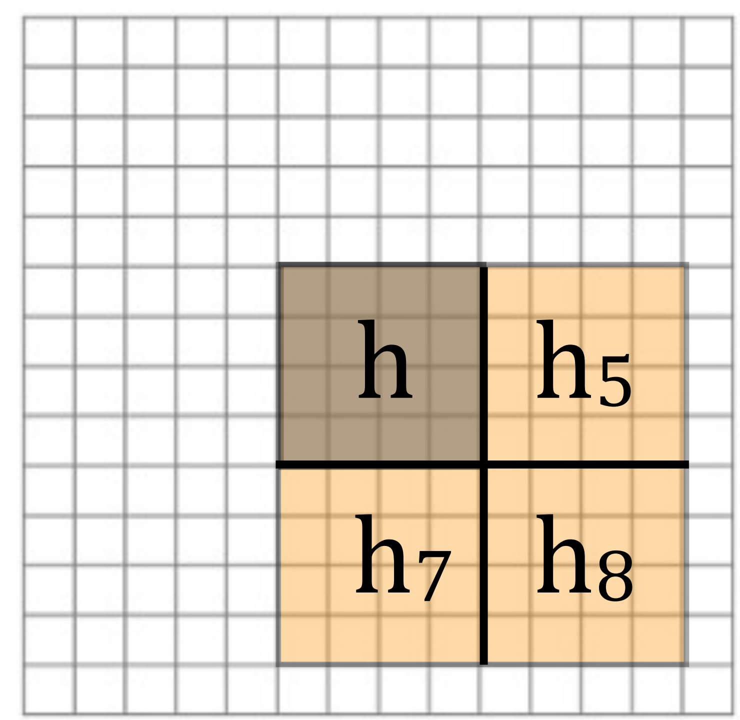
$$h^3 = \max[\, 0.2, h/||(h; h_4; h_6; h_7)|| \,]$$

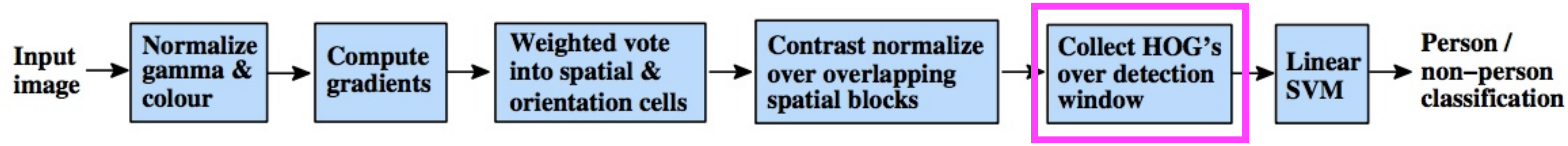
$$h^4 = \max[\, 0.2, h/||(h; h_5; h_7; h_8)|| \,]$$

$$f = (h^1; h^2; h^3; h^4)$$

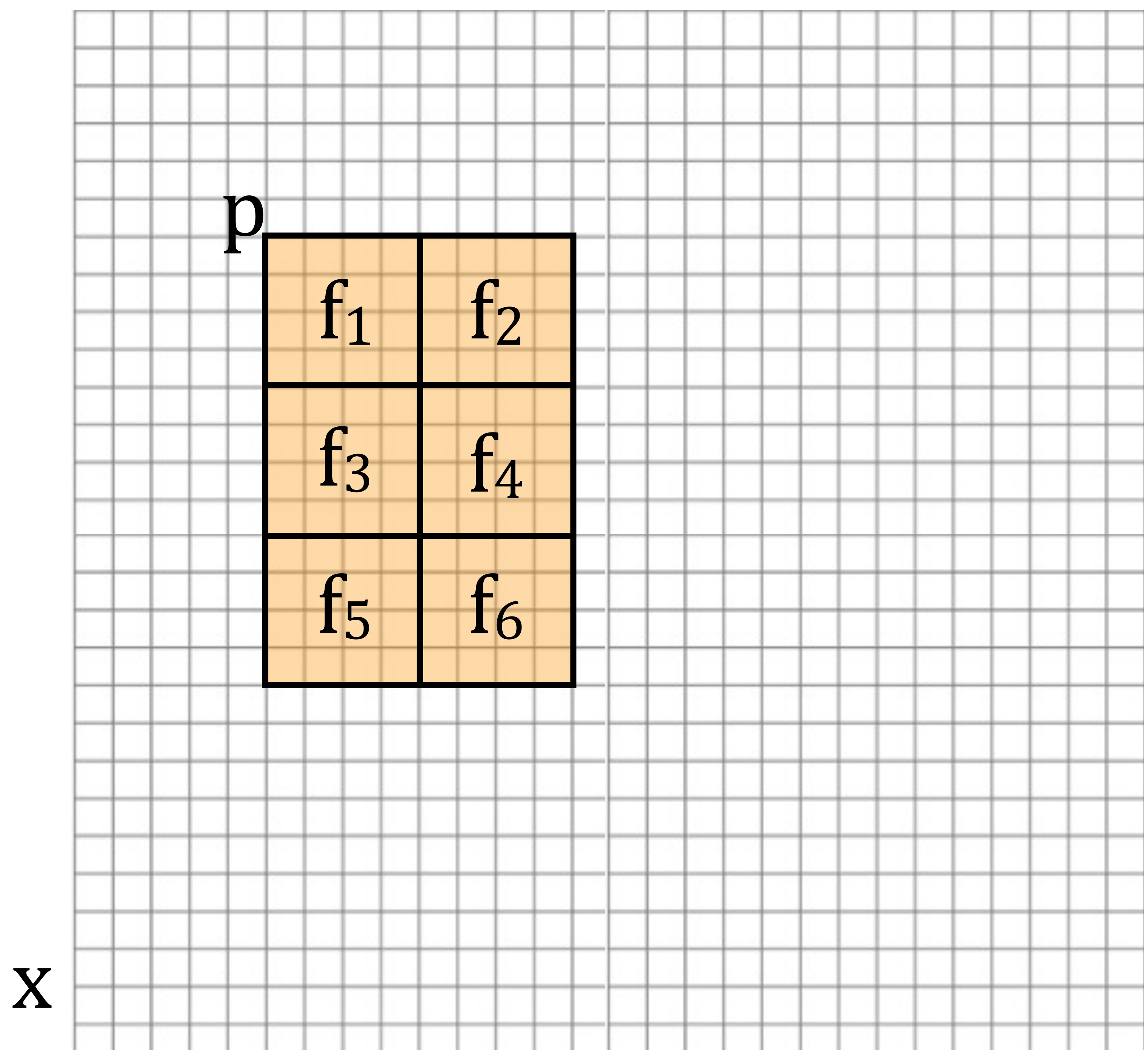
Final dimensionality per cell: 36



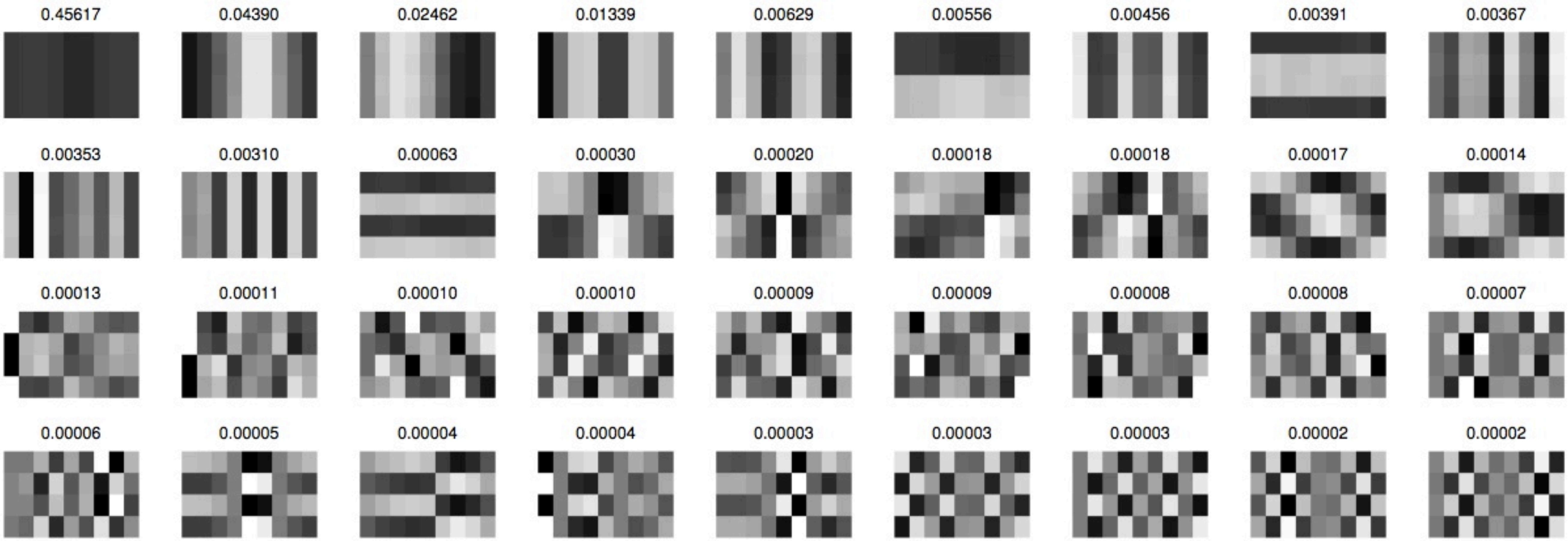




- Sliding window feature vector
  - $\Phi(x, p) = (f_1; f_2; f_3; f_4; f_5; f_6)$



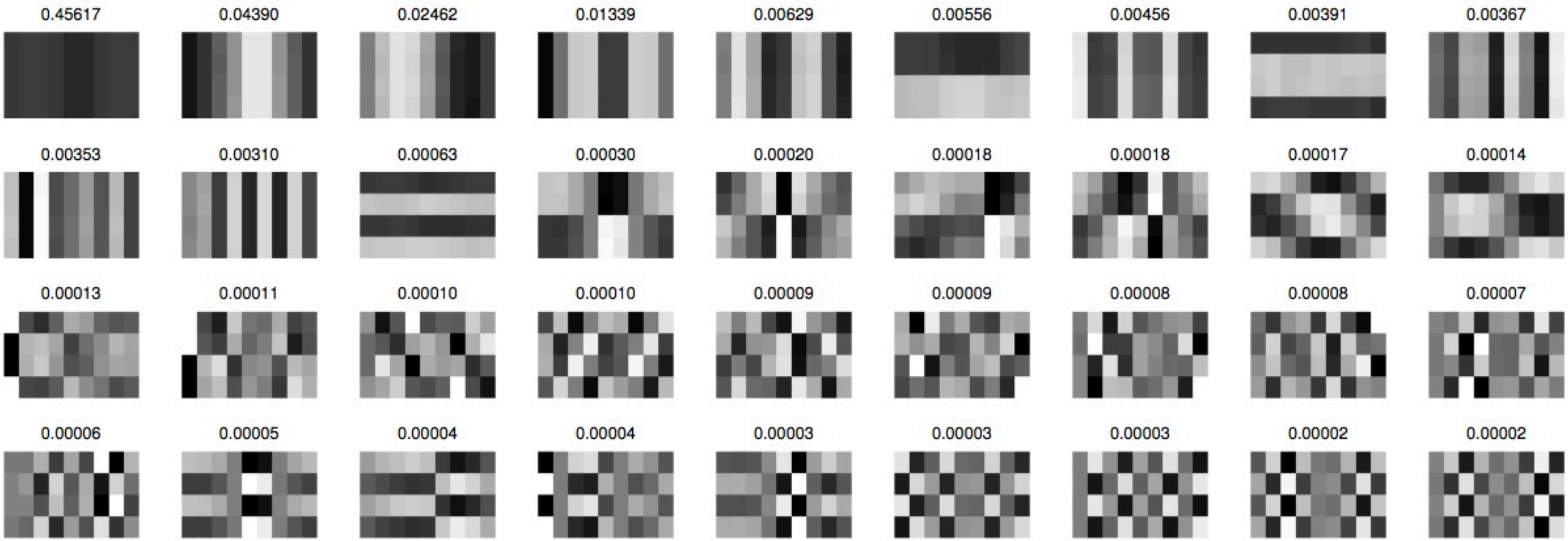
HOG reformulation



- PCA of HOG features
  - Eigenvectors have a strong structure
  - Dim. reduction to top 12 with no loss in performance

Felzenszwalb et al., PAMI 10

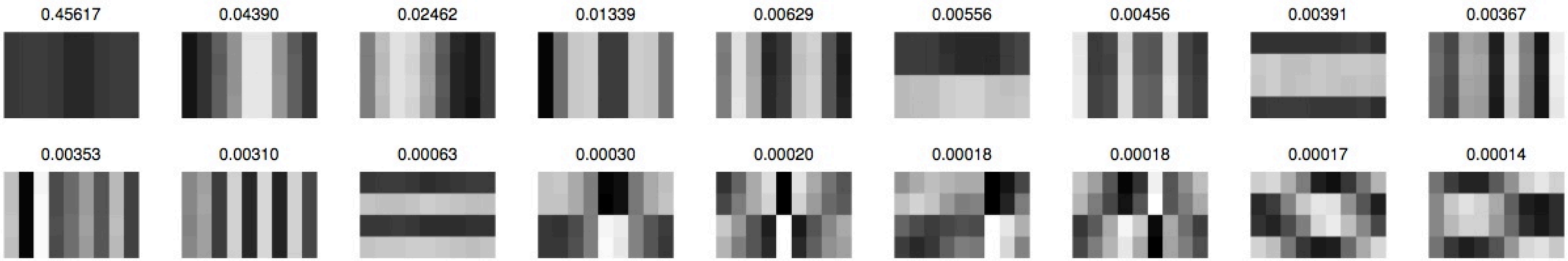
HOG PCA eigenvectors



- Eigenvector structure
  - All rows or columns are (approximately) constant in the top 12 eigenvectors
  - Suggests a different basis

Felzenszwalb et al., PAMI 10

V, a sparse basis for HOG



1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0

u_1

0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0

u_2

...

0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	1

u_9

1	1	1	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

v_1

...

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1

v_4

New basis $V = \{u_1, \ldots, u_9\} \cup \{v_1, v_2, v_3, v_4\}$

Interpretation of V

Original HOG

$f = (h^1; h^2; h^3; h^4)$

Final dimensionality per cell: 36

Analytic projection

$$f = (h^1 + h^2 + h^3 + h^4; 1 \cdot h^1; 1 \cdot h^2; 1 \cdot h^3; 1 \cdot h^4)$$

Final dimensionality per cell: 13

Felzenszwalb et al., PAMI 10

HOG summary

- There's no one true HOG feature
  - Large number of parameters and design choices (see Dalal's thesis)
  - Typical settings
    - cells: 6-8 pixels wide
    - cell blocks: 2x2 or 3x3 rectangular
- The original formulation contains redundant information
  - Efficient and intuitive dimensionality reduction by analytic projection