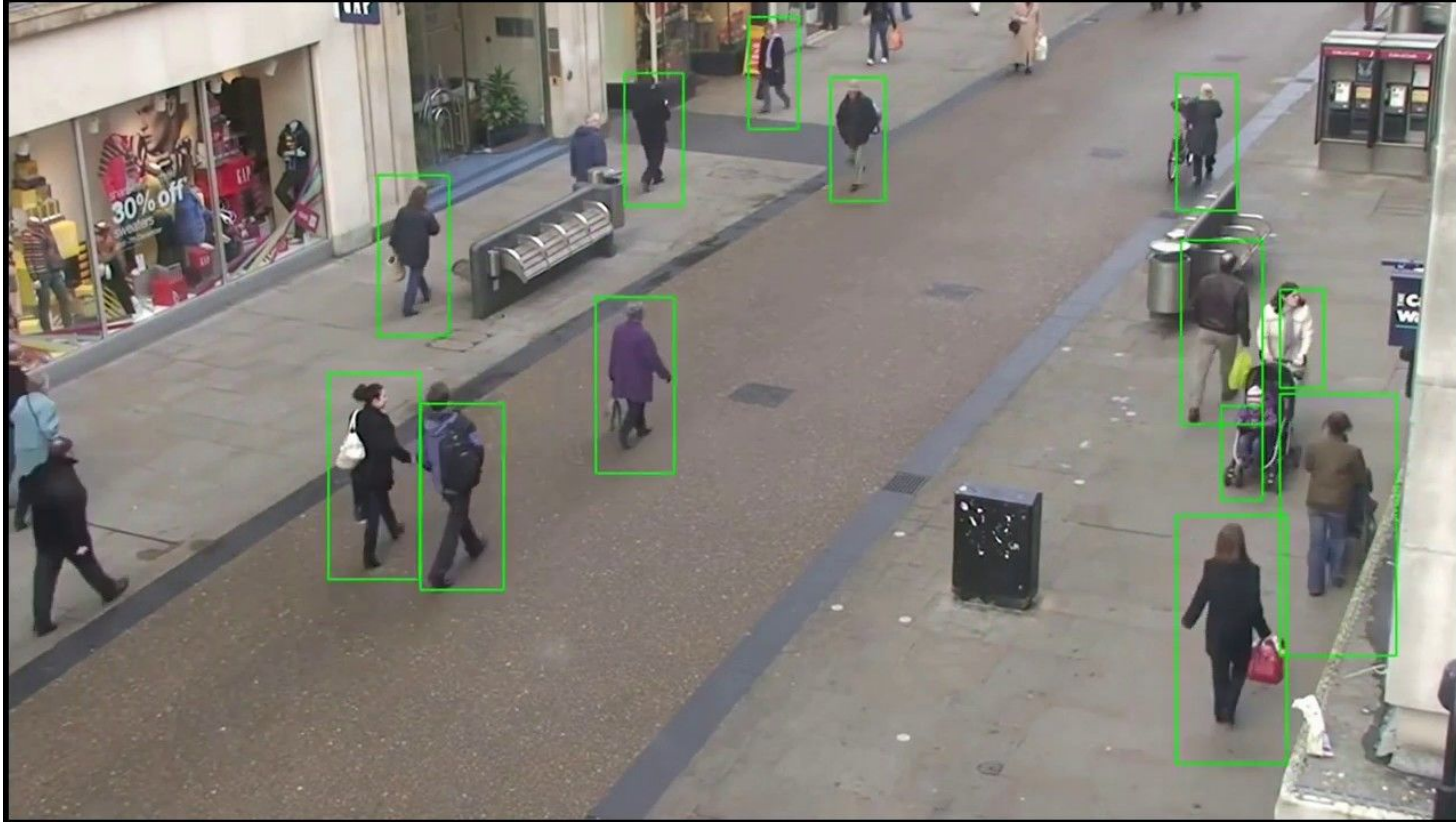
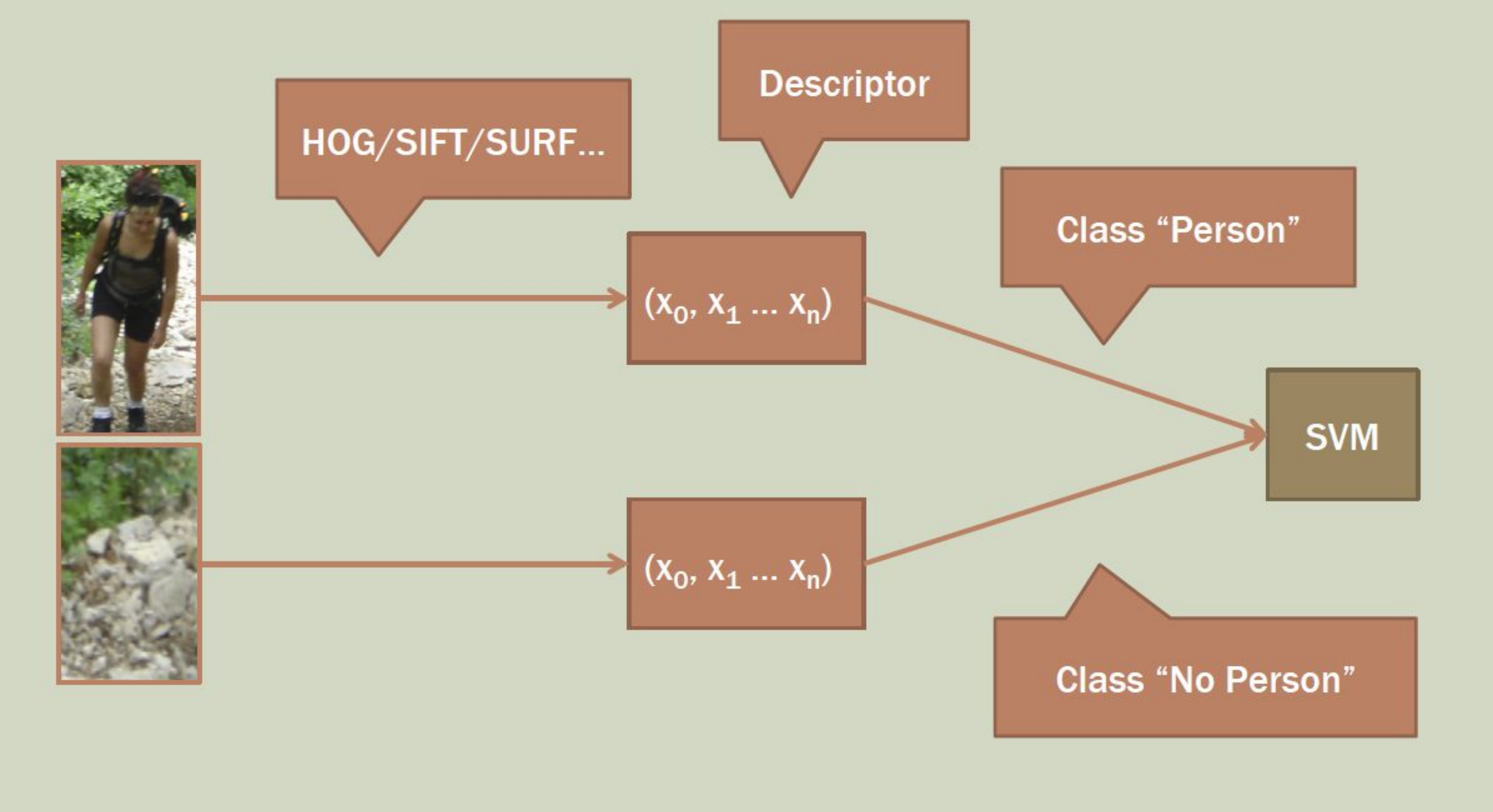
Histograms of Oriented Gradients for Human Detection (HOG)

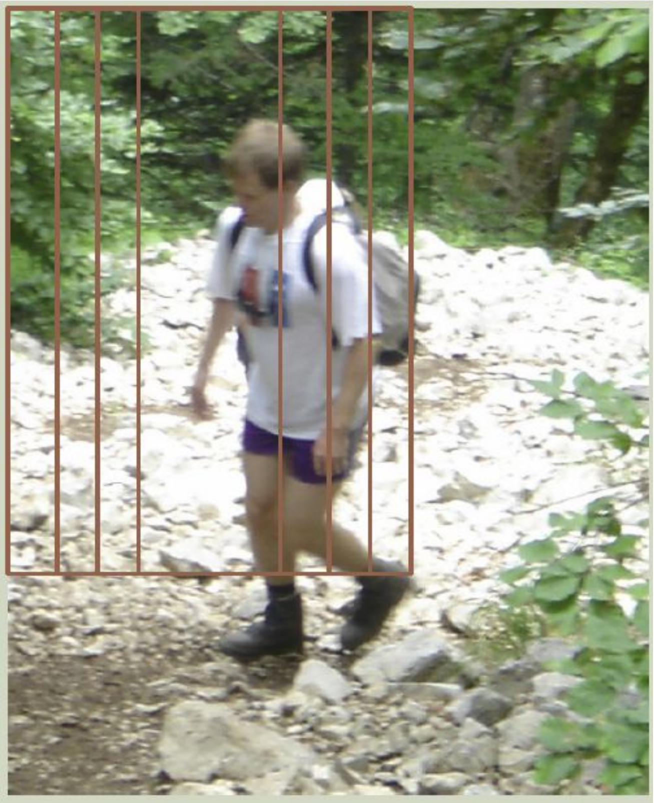
Pedestrian detection


30% off
sweaters

Human detection through a SVM classifier



Use scanning windows in the detection phase



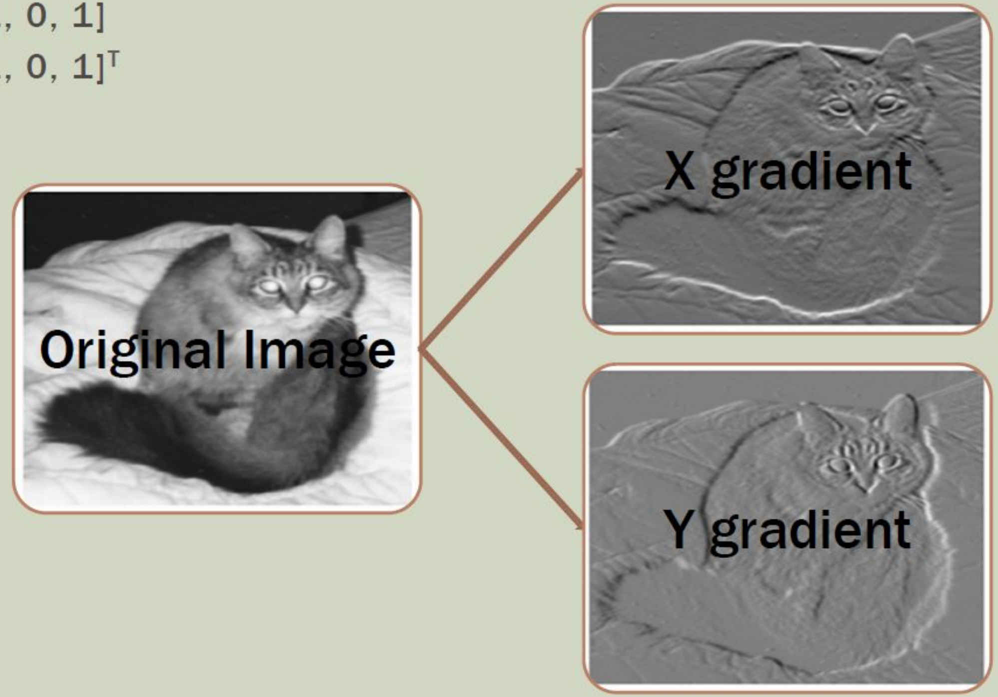
HOG Steps

HOG Steps

- HOG feature extraction
  - Compute centered horizontal and vertical gradients with no smoothing
  - Compute gradient orientation and magnitudes
    - For color image, pick the color channel with the highest gradient magnitude for each pixel.
- For a 64x128 image,
- Divide the image into 16x16 blocks of 50% overlap.
  - 7x15=105 blocks in total
- Each block should consist of 2x2 cells with size 8x8.
- Quantize the gradient orientation into 9 bins
  - The vote is the gradient magnitude
  - Interpolate votes between neighboring bin center.
  - The vote can also be weighted with Gaussian to downweight the pixels near the edges of the block.
- Concatenate histograms (Feature dimension: 105x4x9 = 3,780)

Image → **Gradients** → Cells → Blocks → Descriptor

- Convolve the image with discrete derivative mask
  - [-1, 0, 1]
  - $[-1, 0, 1]^T$

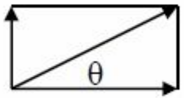


Computing Gradients

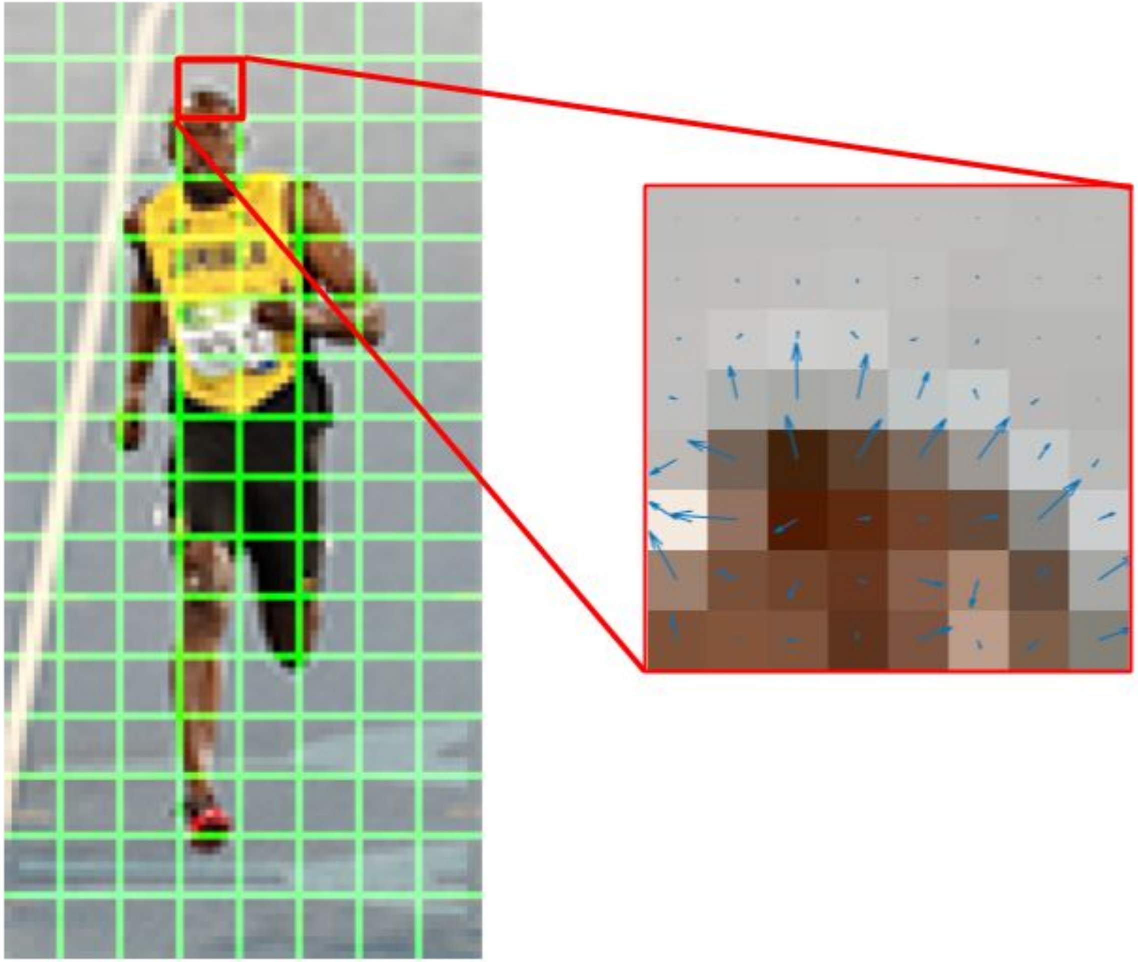
- Centered: $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}$

- Filter masks in x and y directions
  - Centered: $\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$ $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

- Gradient
  - Magnitude: $s = \sqrt{s_x^2 + s_y^2}$



  - Orientation: $\theta = \arctan(\frac{s_y}{s_x})$



2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

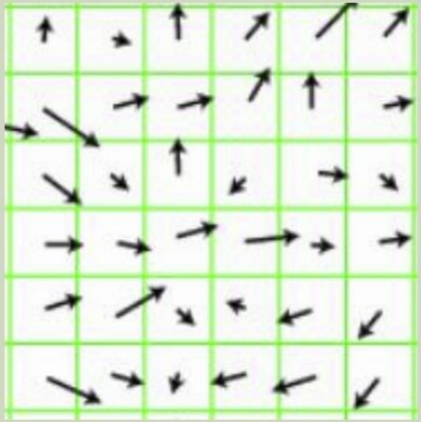
Gradient Magnitude

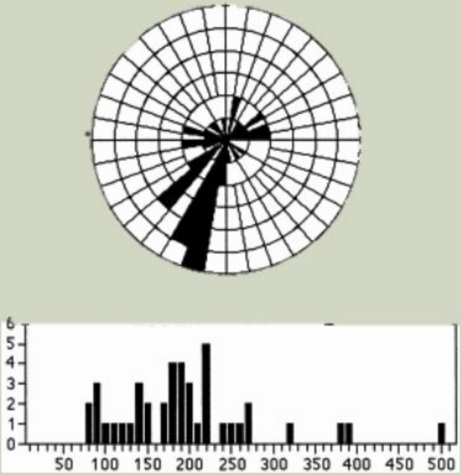
80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Gradient Direction

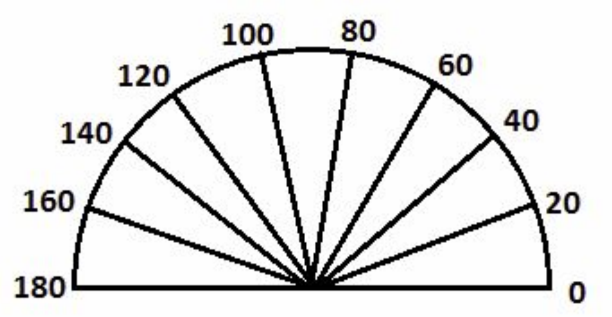
Image → Gradients → **Cells** → Blocks → Descriptor

- Now we count up the gradient angles in 8x8 cells
  - Vote weight = magnitude = $\sqrt{dx^2 + dy^2}$
  - Who you vote for ~ angle = $\arctan(dy/dx)$

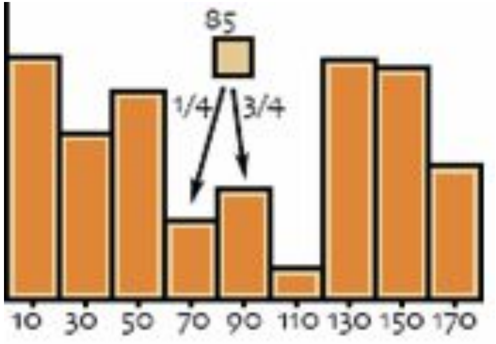




Votes



- Each cell has 8 x 8 pixels
- Quantize the gradient orientation into 9 bins (0-180)
    - The vote is the gradient magnitude
    - Interpolate votes linearly between neighboring bin centers.
        - Example: if θ=85 degrees.
        - Distance to the bin center Bin 70 and Bin 90
        - are 15 and 5 degrees, respectively.
        - Hence, ratios are 5/20=1/4, 15/20=3/4.



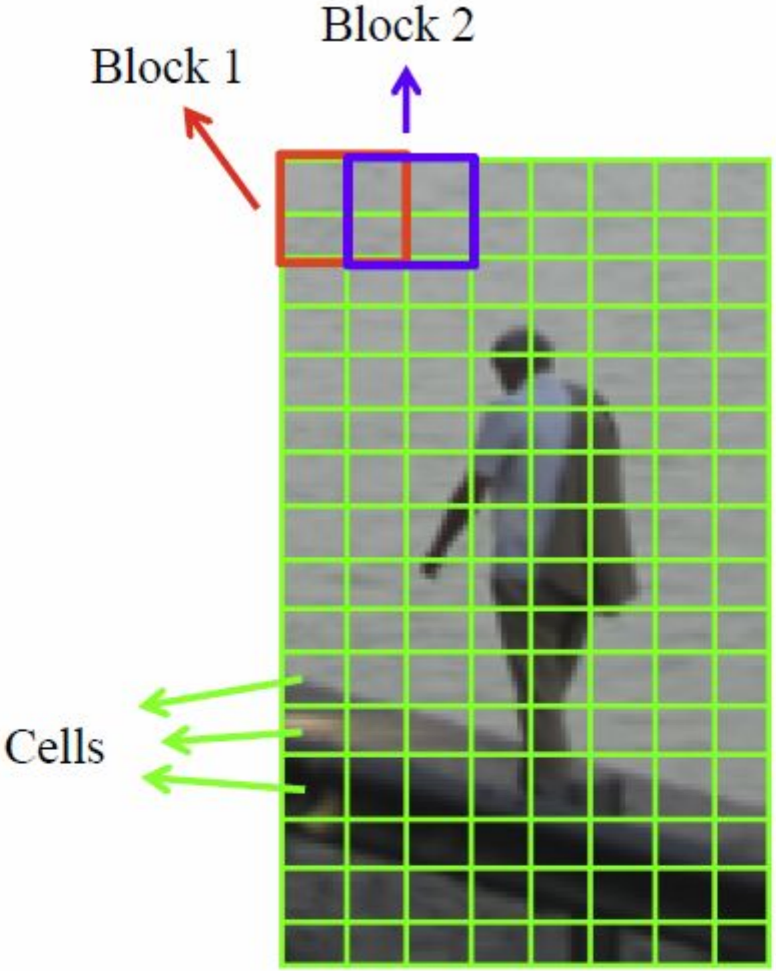
    - The vote can also be weighted with Gaussian to down weight the pixels near the edges of the block.

Blocks and Cells

For a 64x128 image

- 16x16 blocks of 50% overlap.
    - 7x15=105 blocks in total

- Each block should consist of 2x2 cells with size 8x8.



Block normalization

- Group the cells into overlapping blocks of 2 × 2 cells each
- Concatenate the four cell histograms in each block into a single block feature b and normalize the block feature by its Euclidean norm

$$\mathbf{b} \leftarrow \frac{\mathbf{b}}{\sqrt{\|\mathbf{b}\|^2 + \epsilon}}$$

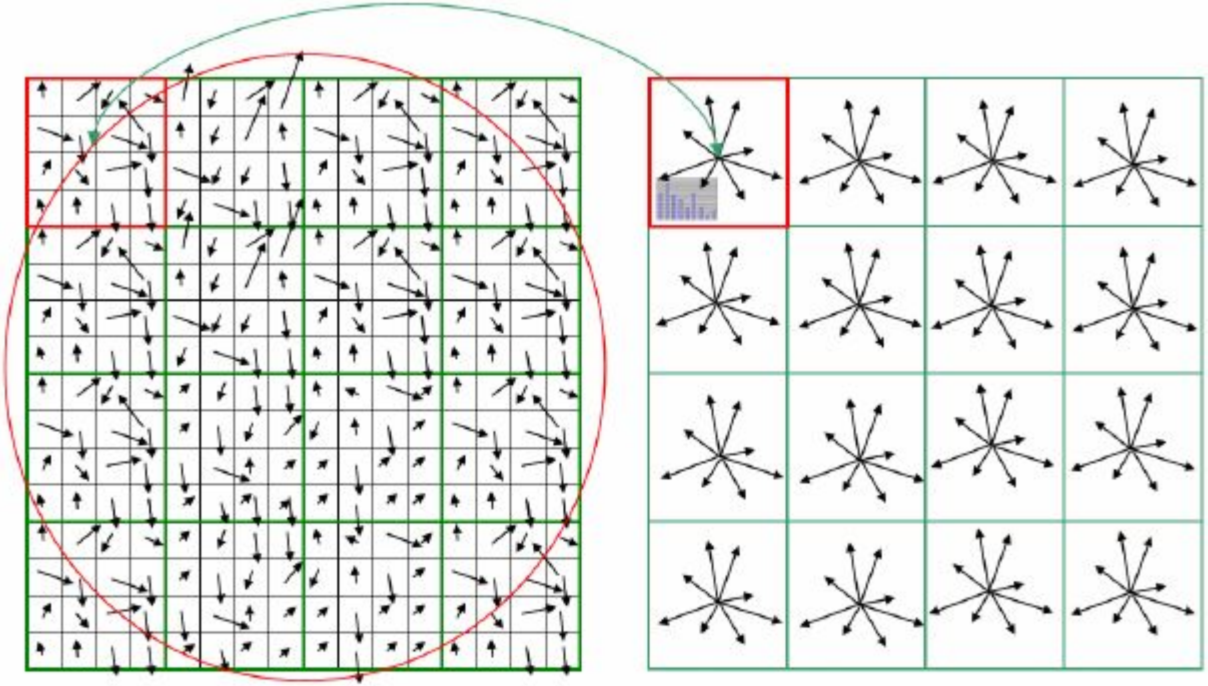
- Block normalization is a compromise: On one hand, cell histograms need to be normalized to reduce the effect of changes in contrast between images of the same object. On the other hand, overall gradient magnitude does carry some information, and normalization over a block—a region greater than a single cell—preserves some of this information, namely, the relative magnitudes of gradients in cells within the same block. Since each cell is covered by up to four blocks, each histogram is represented up to four times with up to four different normalizations.

HOG Features

The normalized block features are concatenated into a single HOG feature vector h, which is normalized as follows

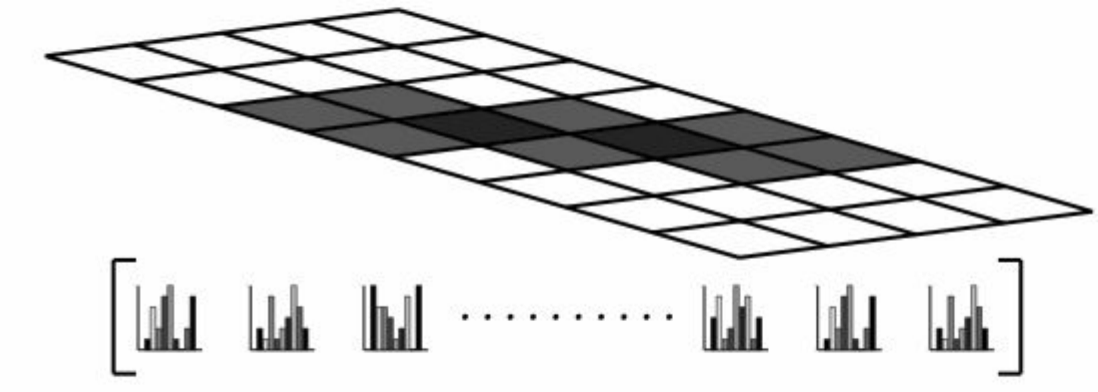
$$
\begin{aligned}
\mathbf{h} &\leftarrow \frac{\mathbf{h}}{\sqrt{\|\mathbf{h}\|^2 + \epsilon}} \\
h_n &\leftarrow \min(h_n, \tau) \\
\mathbf{h} &\leftarrow \frac{\mathbf{h}}{\sqrt{\|\mathbf{h}\|^2 + \epsilon}}.
\end{aligned}
$$

Here, h_n is the n-th entry of h and is a positive threshold (= 0.2). Clipping the entries of h to be no greater than (after the first normalization) ensures that very large gradients do not have too much influence—they would end up washing out all other image detail. The final normalization makes the HOG feature independent of overall image contrast.

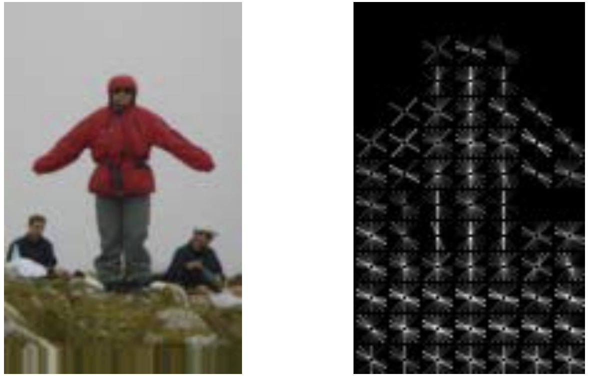


Final Feature Vectors

- Concatenate histograms
  - Make it a 1D vector of length 3780.



- Visualization



References

https://ieeexplore.ieee.org/document/1467360/9411012

https://www.learnopencv.com/histogram-of-oriented-gradients/

http://mccormickml.com/2013/05/09/hog-person-detector-tutorial/