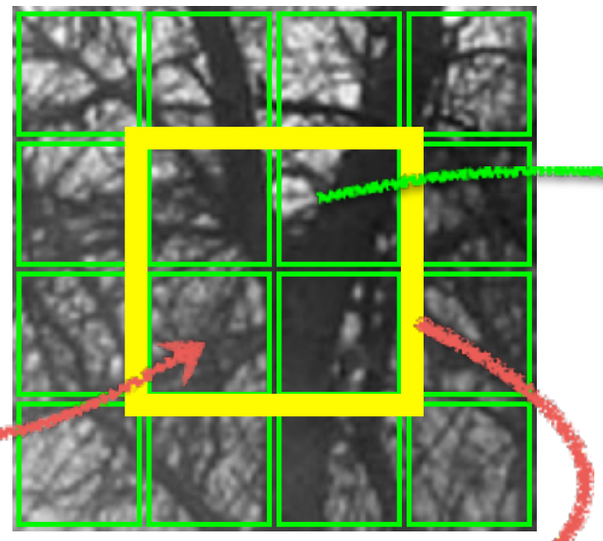
HOG descriptor

HOG

Dalal, Triggs. **Histograms of Oriented Gradients** for Human Detection. CVPR, 2005



Cell
(8x8 pixels)

Block
(2x2 cells)

histogram of ‘unsigned’ gradients

soft binning

gradient magnitude histogram
(one for each cell)

Concatenate and L-2 normalization

Single scale, no dominant orientation

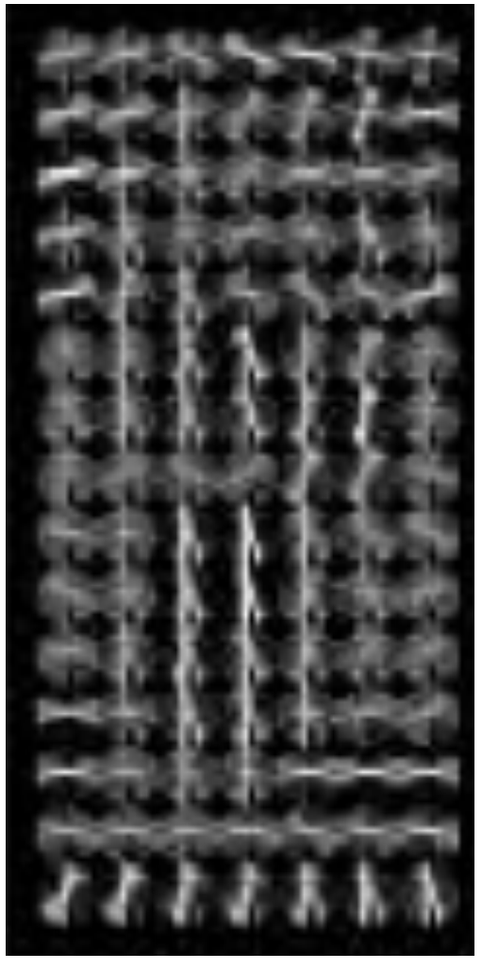
Pedestrian detection

1 cell step size

128 pixels
16 cells
15 blocks

15 x 7 x 4 x 9 =
3780

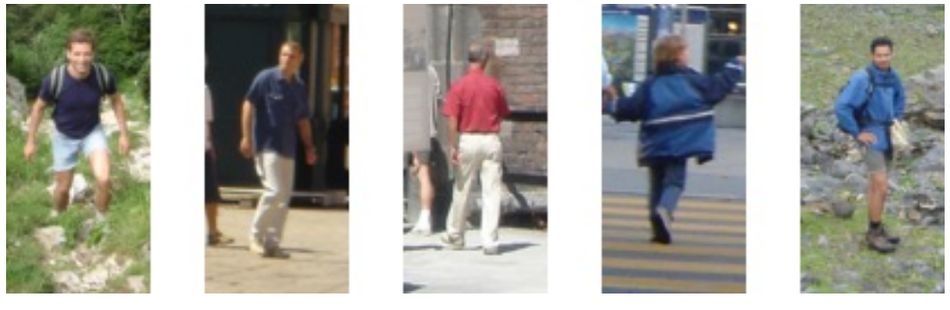
visualization



64 pixels
8 cells
7 blocks

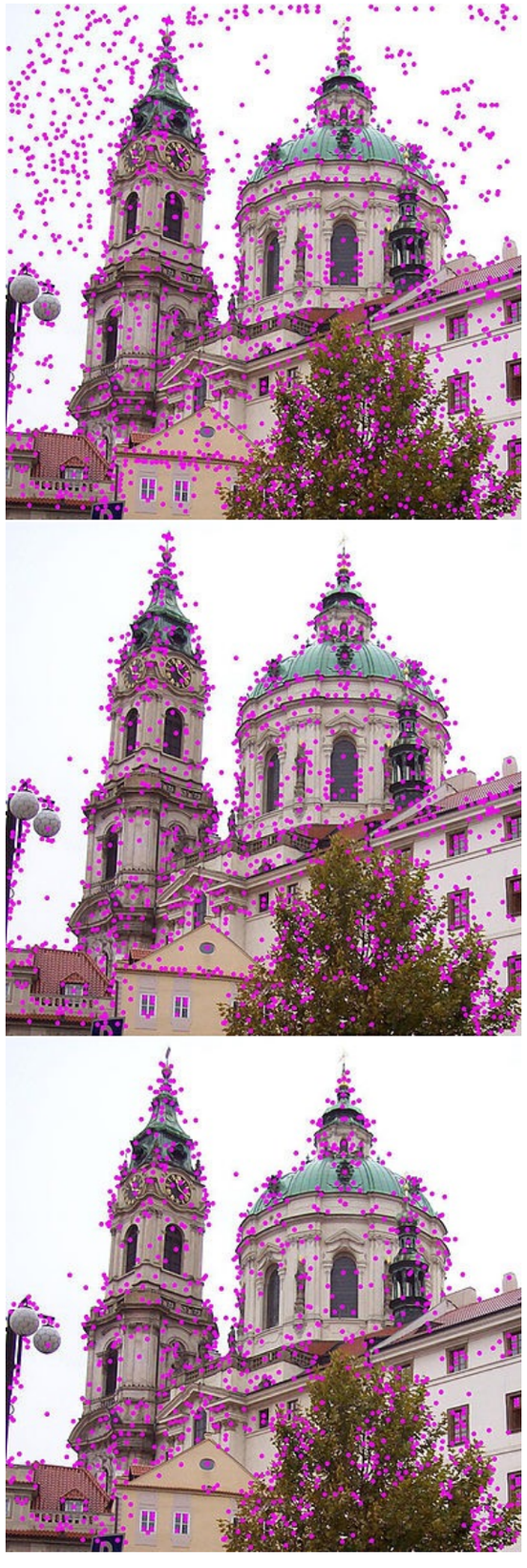
Redundant representation due to overlapping blocks

How many times is each inner cell encoded?



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

SIFT



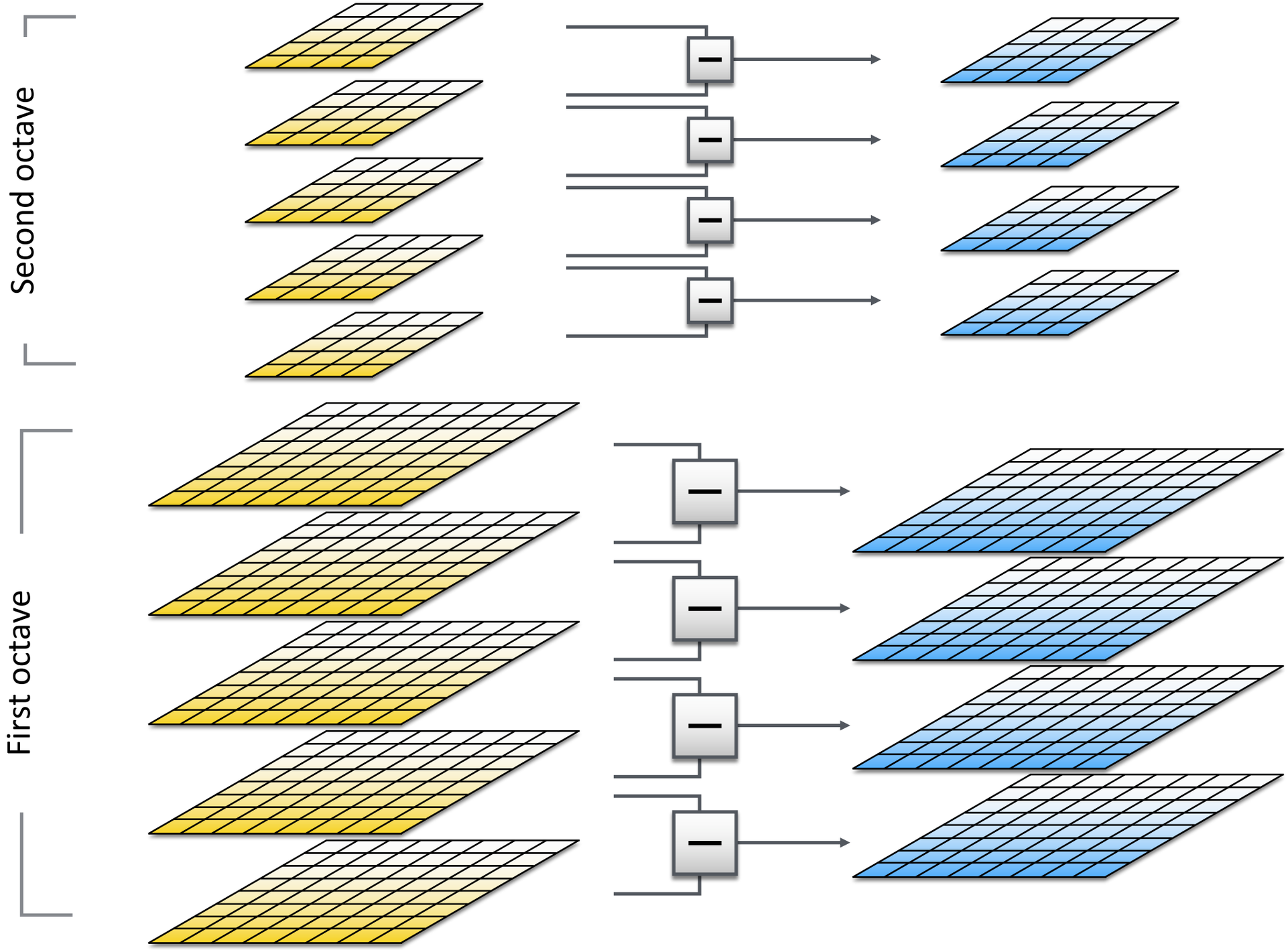
SIFT

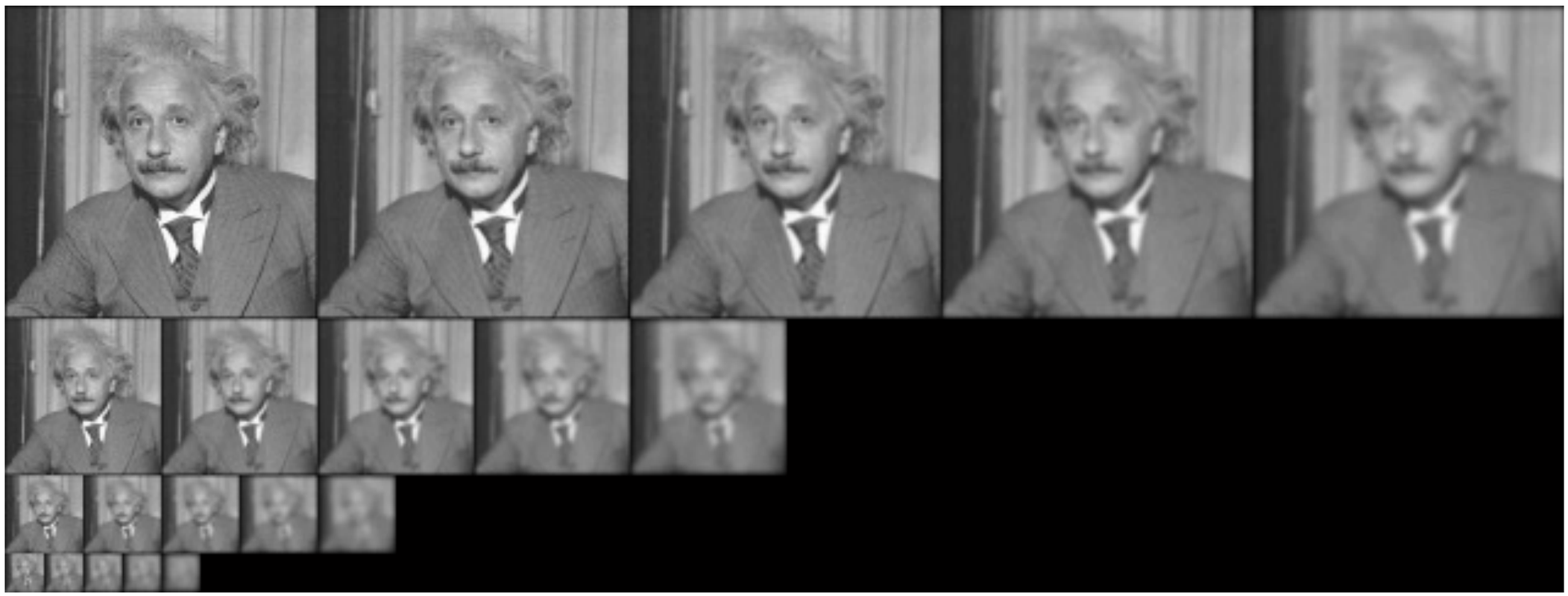
(Scale Invariant Feature Transform)

SIFT describes both a **detector** and **descriptor**

1. Multi-scale extrema detection
2. Keypoint localization
3. Orientation assignment
4. Keypoint descriptor

1. Multi-scale extrema detection

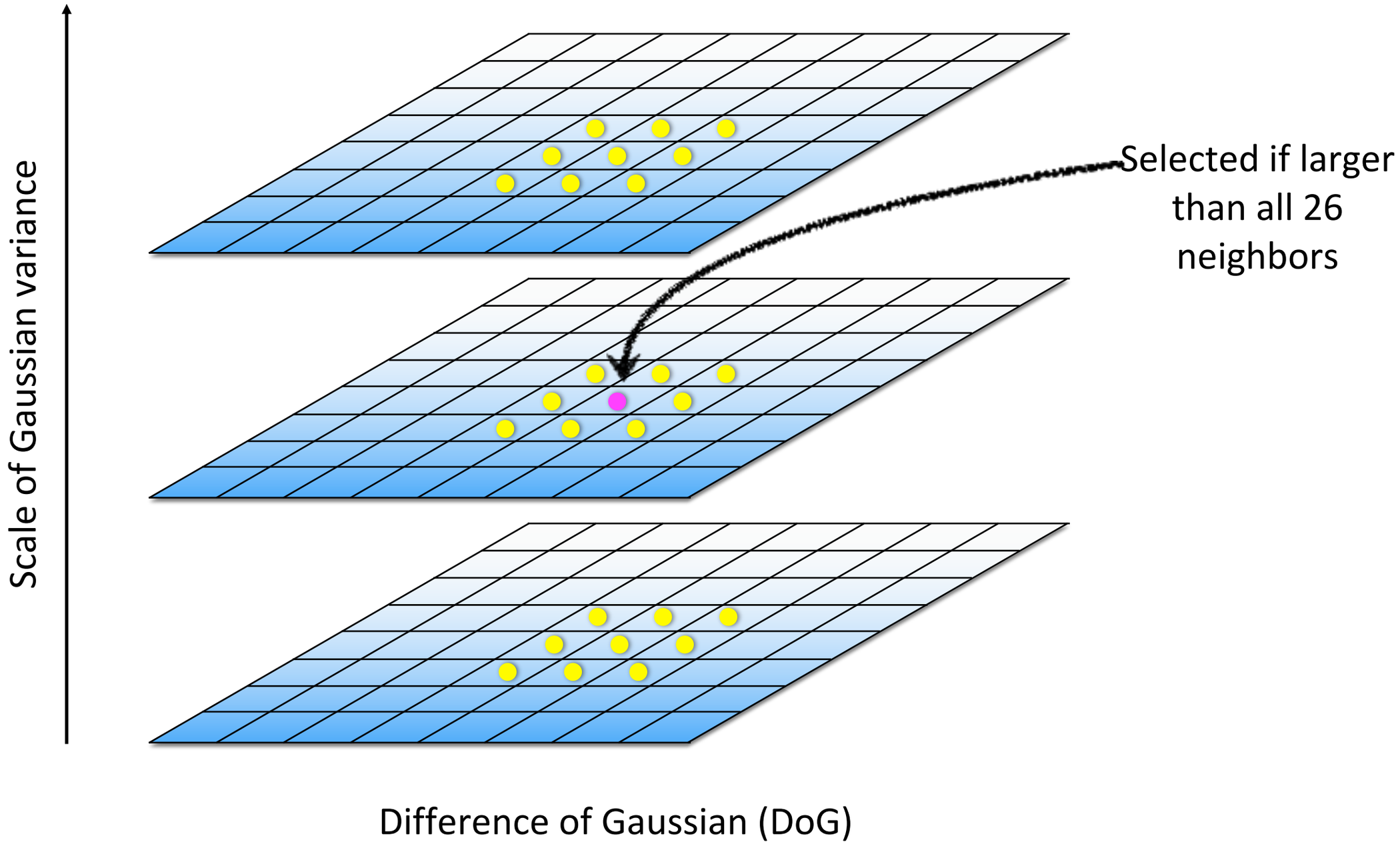




Gaussian

Laplacian

Scale-space extrema



2. Keypoint localization

2nd order Taylor series approximation of DoG scale-space

$$f(\mathbf{x}) = f + \frac{\partial f}{\partial \mathbf{x}}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 f}{\partial \mathbf{x}^2}\mathbf{x}$$

$$\mathbf{x} = \{x, y, \sigma\}$$

Take the derivative and solve for extrema

$$\mathbf{x}_m = -\frac{\partial^2 f}{\partial \mathbf{x}^2}^{-1} \frac{\partial f}{\partial \mathbf{x}}$$

Additional tests to retain only strong features

3. Orientation assignment

For a keypoint, **L** is the **Gaussian-smoothed** image with the closest scale,

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

x-derivative

y-derivative

$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$

Detection process returns

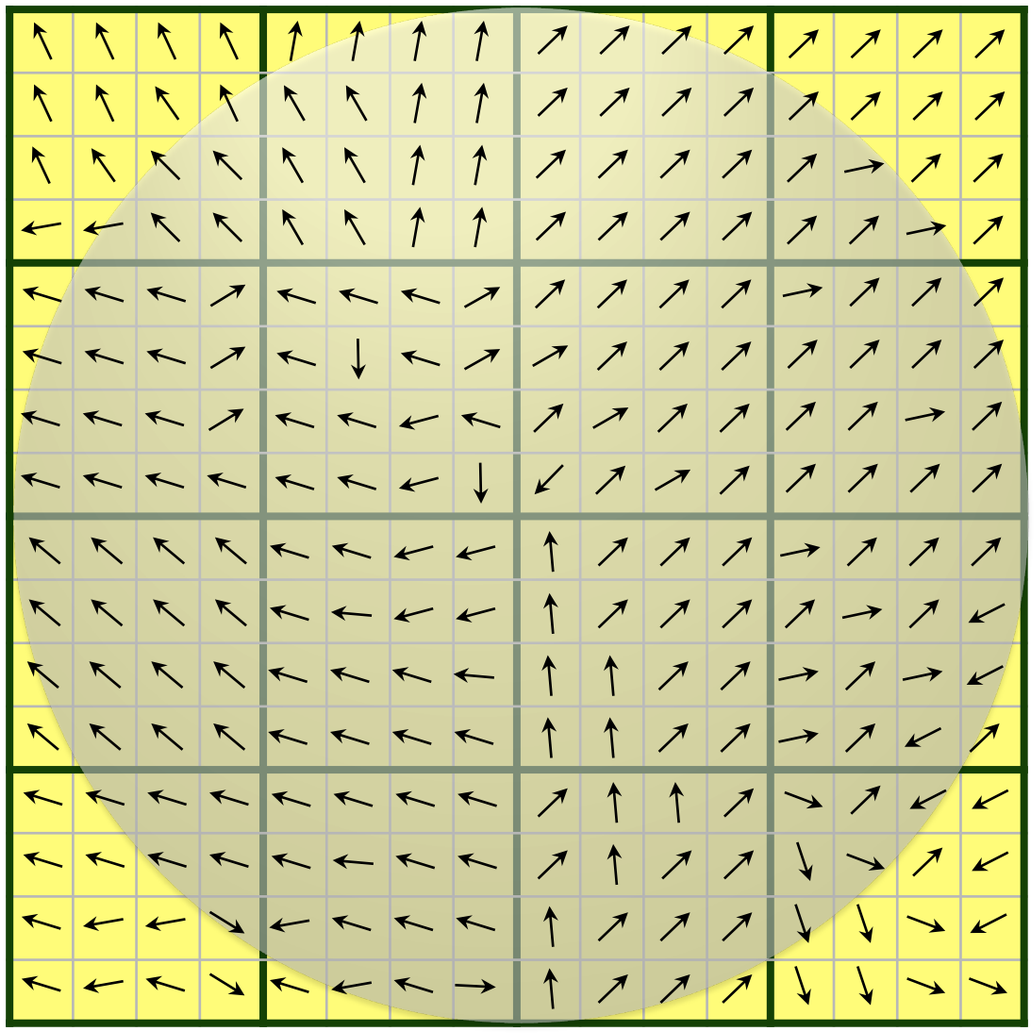
$$\{x, y, \sigma, \theta\}$$

location scale orientation

4. Keypoint descriptor

Image Gradients

(4 x 4 pixel per cell, 4 x 4 cells)

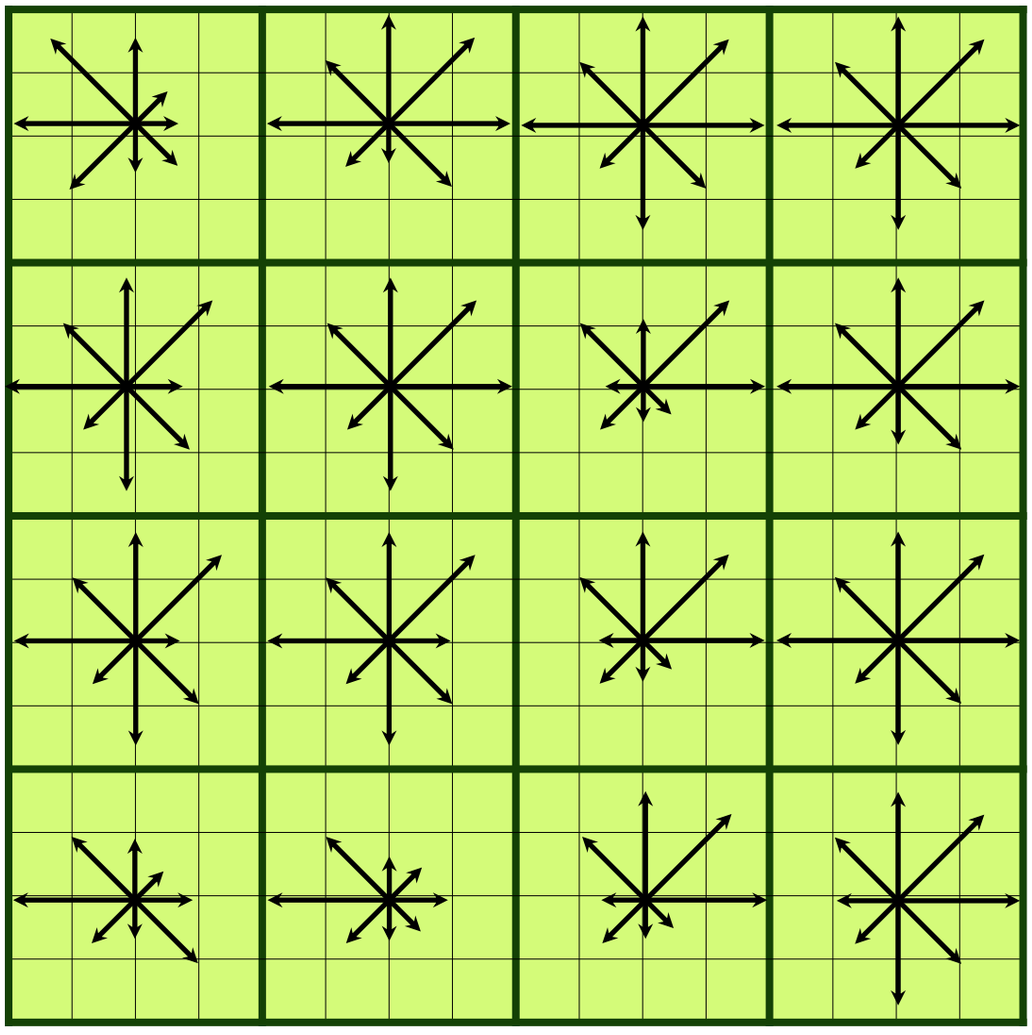


Gaussian weighting

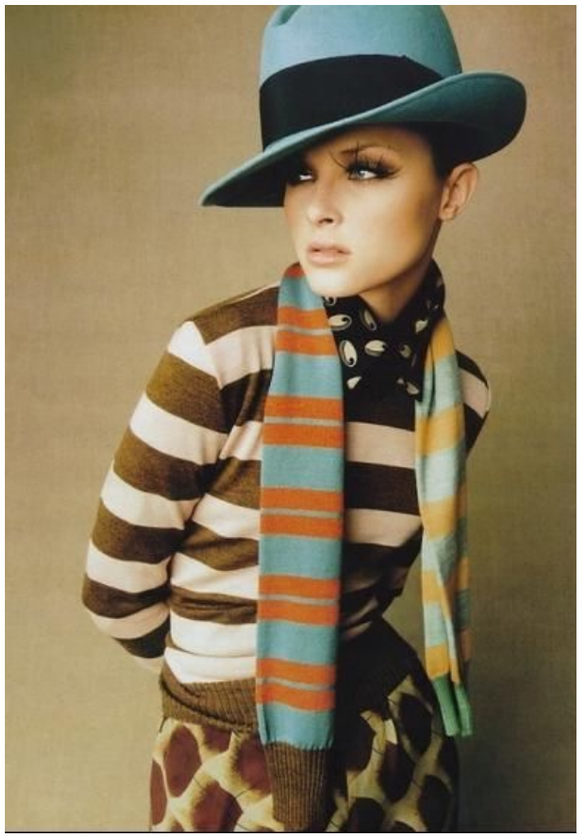
(sigma = half width)

SIFT descriptor

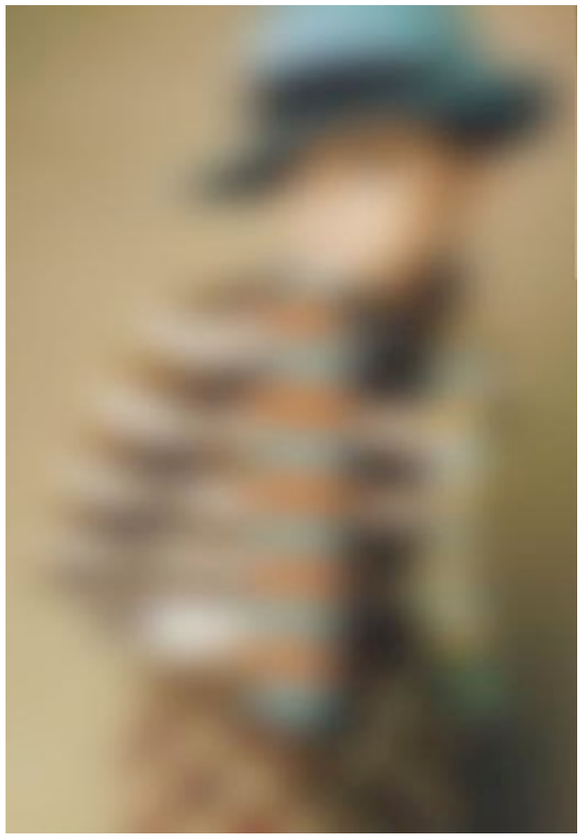
(16 cells x 8 directions = 128 dims)



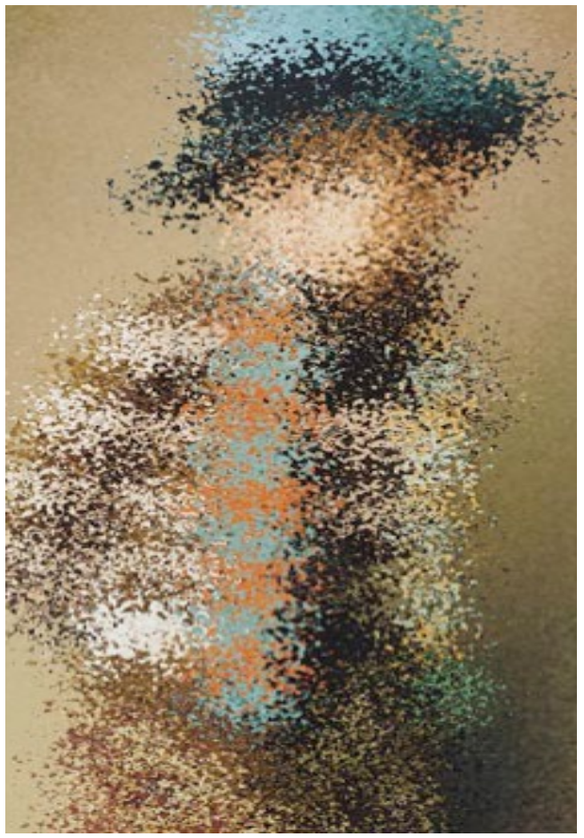
Discriminative power



Raw pixels



Sampled



Locally orderless

Global histogram

Generalization power