SIFT Overview

Local features

- …as opposed to “global features” such a color histogram
- ..describe “robust” properties of “points of interest” in the image
- .. “ robust” means invariant to
  - illumination differences
  - scaling,
  - rotation,
  - changes in viewing angle, and
  - noise

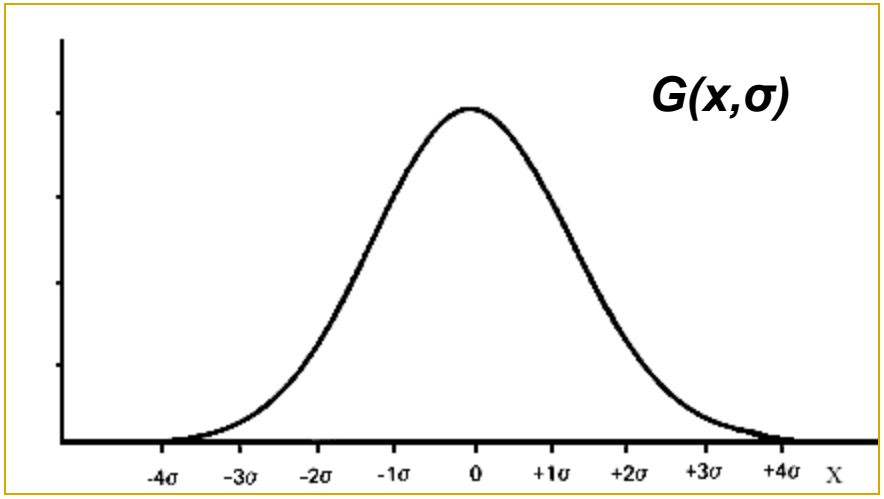
Scale Invariant Feature Transform (SIFT)

- Scale-space extrema detection
  - Identify points of interest across multiple scales of the image
- Keypoint localization
  - Pick most robust points
- Orientation assignment
  - Normalize the orientation to ensure rotation-invariance
- Generation of keypoint descriptors.
  - Extract an illumination invariant descriptor around each keypoint

Scale space extrema detection

- Blur the image using Gaussian (normal) filters

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}}$$



$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y)$$

Convolution

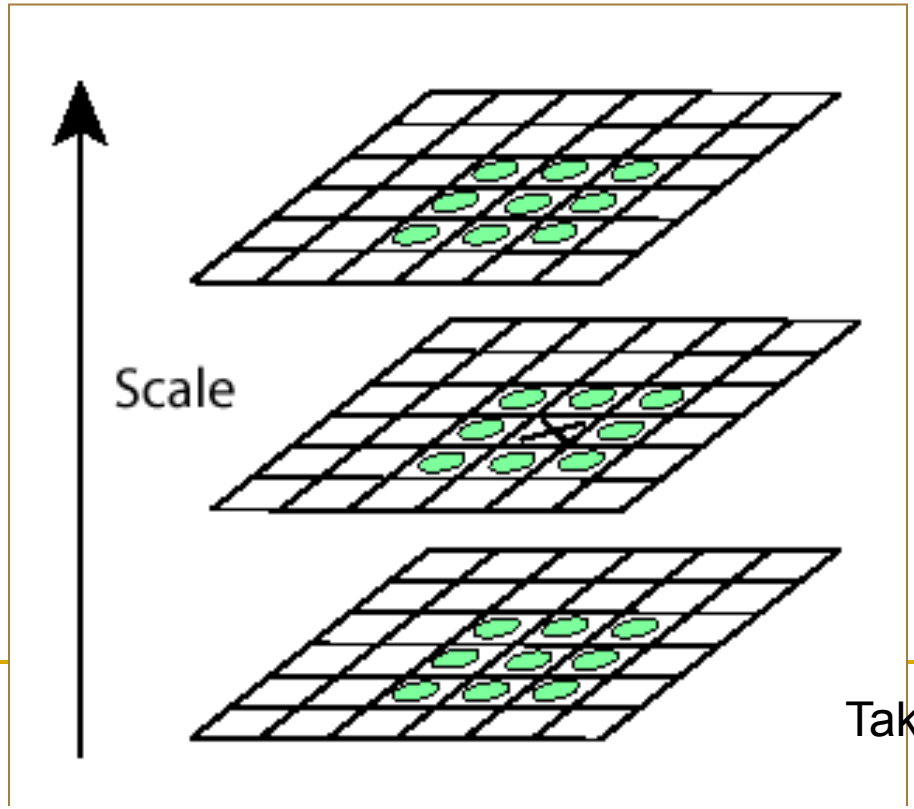
Scale space extrema detection

- Take the difference between the adjacent scales

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Scale space extrema detection

- Each pixel is compared against it 26 neighbors
- Points that are local maxima or minima are selected as points of interest



Taken from [Lowe04]

Keypoint localization

- Keypoints
  - with low contrast
    - sensitive to noise
  - poorly localized along edges
    - unstable due to noise

  are eliminated

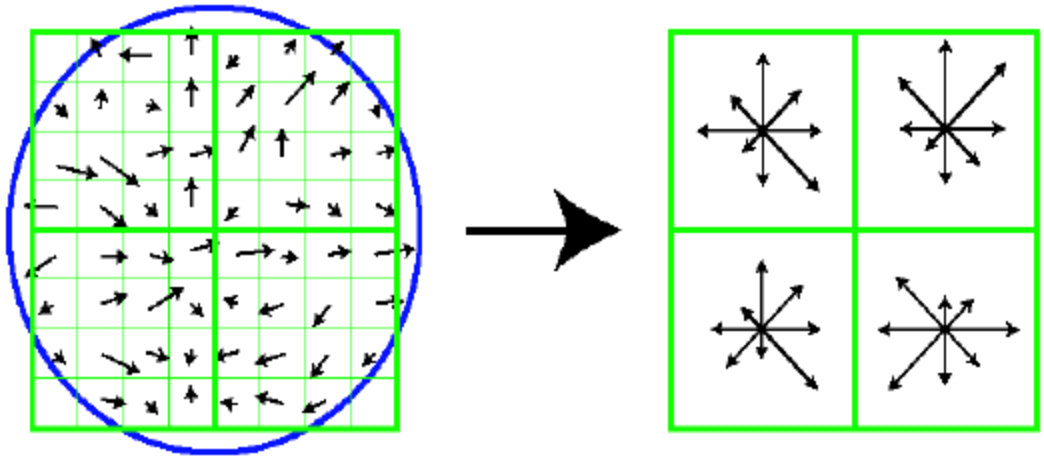
Orientation assignment

- A gradient orientation histogram is computed in the neighborhood of the keypoint in the appropriate scale, σ
  - Contributions of neighboring pixels are weighted by the gradient magnitude and a Gaussian window with 1.5σ scale
- Peaks in the histogram correspond to dominant orientations.
  - A distinct keypoint is created for each dominant direction

Keypoint Descriptor Computation

- Relative to the keypoint orientation to ensure rotation invariance

Taken from [Lowe04]



  - This example shows a 2x2 descriptor array computed from an 8x8 set of samples; actual implementation uses 4x4 descriptors from 16x16 samples