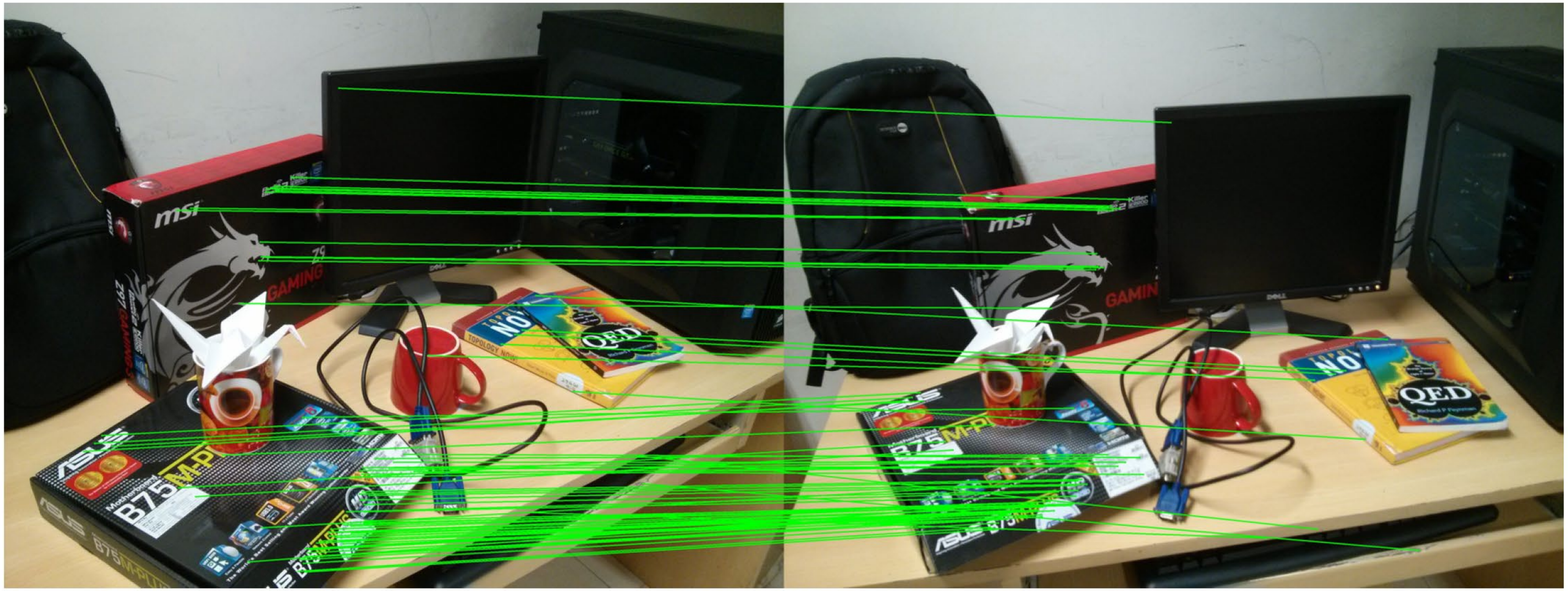
Scale Invariance and SIFT

CS 4391 Introduction Computer Vision

Professor Yu Xiang

The University of Texas at Dallas

Feature Detection and Matching

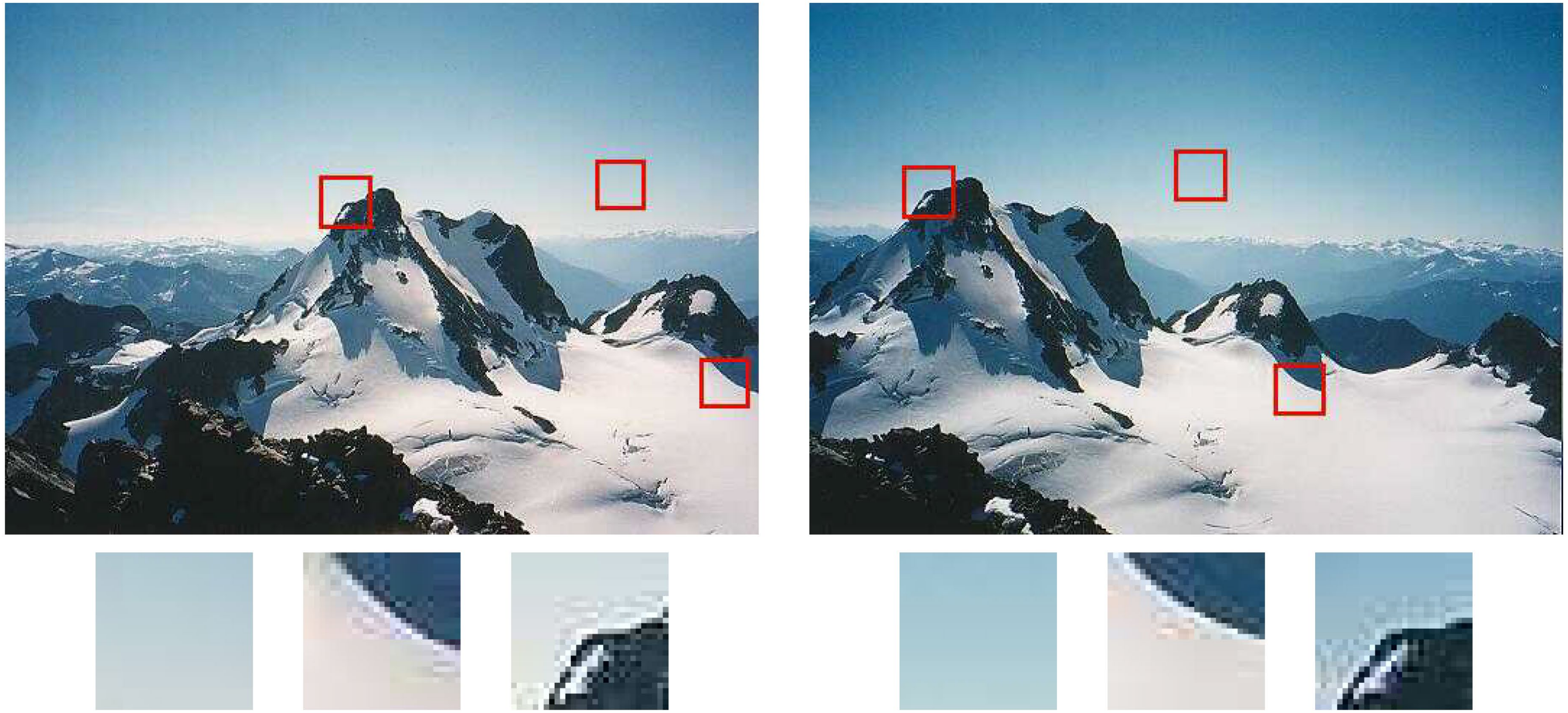


Geometry-aware Feature Matching for Structure from Motion Applications. Shah et al, WACV’15

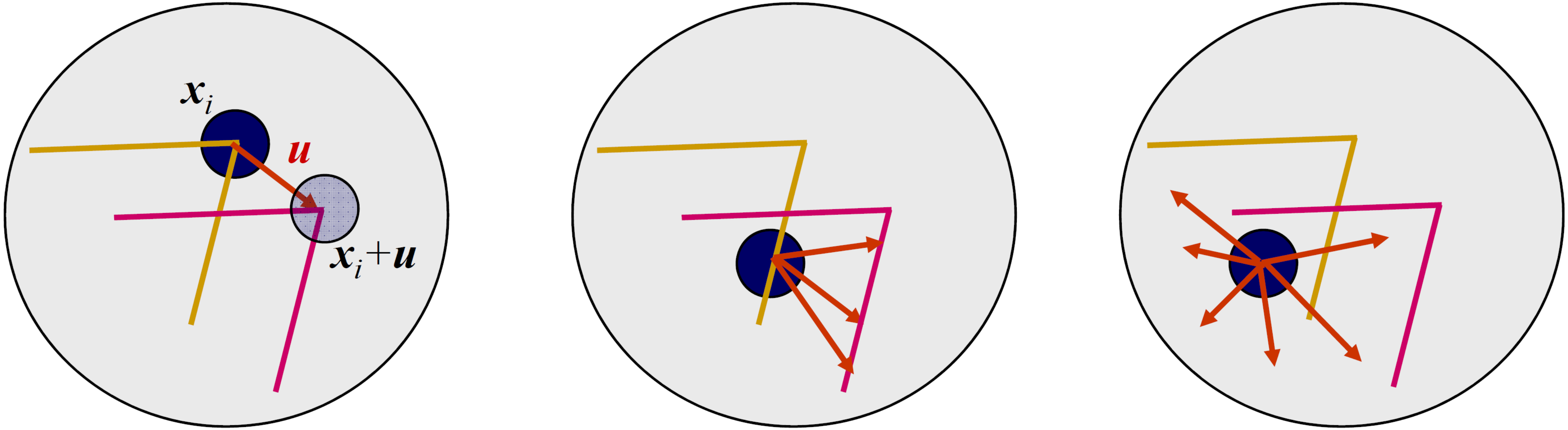
Applications: stereo matching, image stitching, 3D reconstruction, camera pose estimation, object recognition

Feature Detectors

- How to find image locations that can be reliably matched with images?



Feature Detectors



(a) Corner

(b) Edge

(c) Textureless region

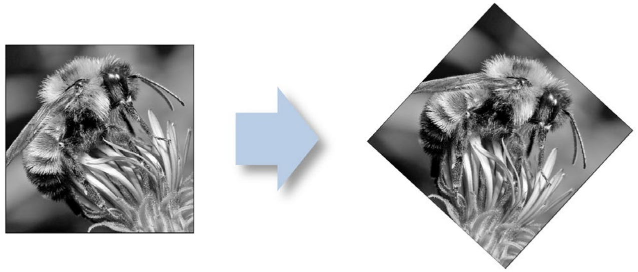
Invariance

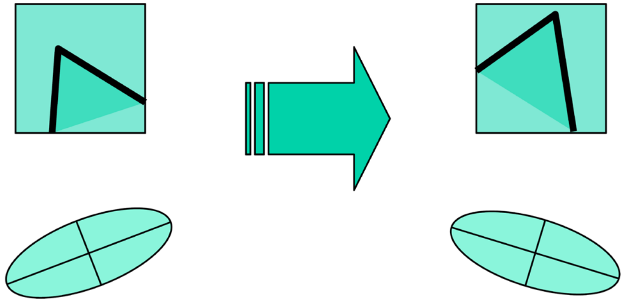
- Can the same feature point be detected after some transformation?
  - Translation invariance

    Are Harris corners translation invariance?

  - 2D rotation invariance

    Are Harris corners rotation invariance?

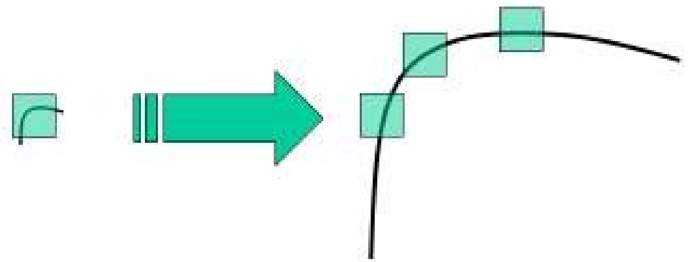


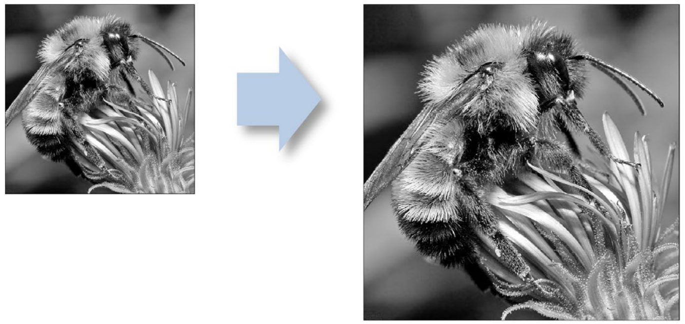


  - Scale invariance

    Are Harris corners scale invariance?

No





Scale Invariance

- Solution 1: detection features in all scales, matching features in corresponding scale (for small scale change)

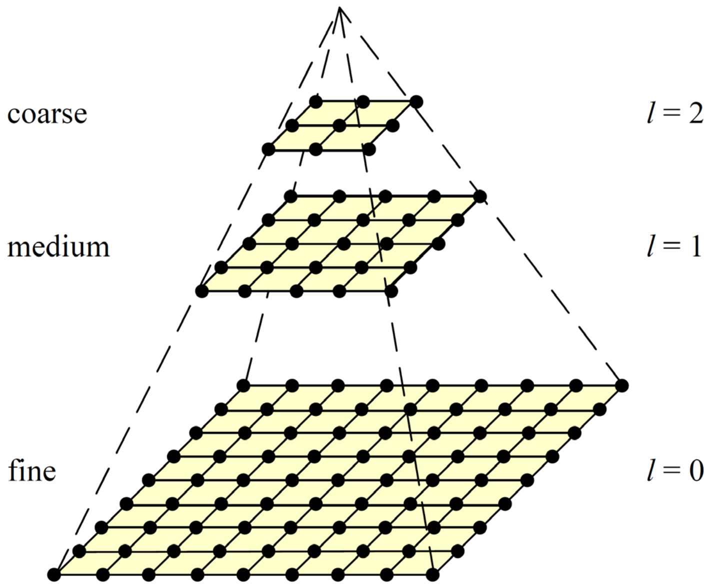
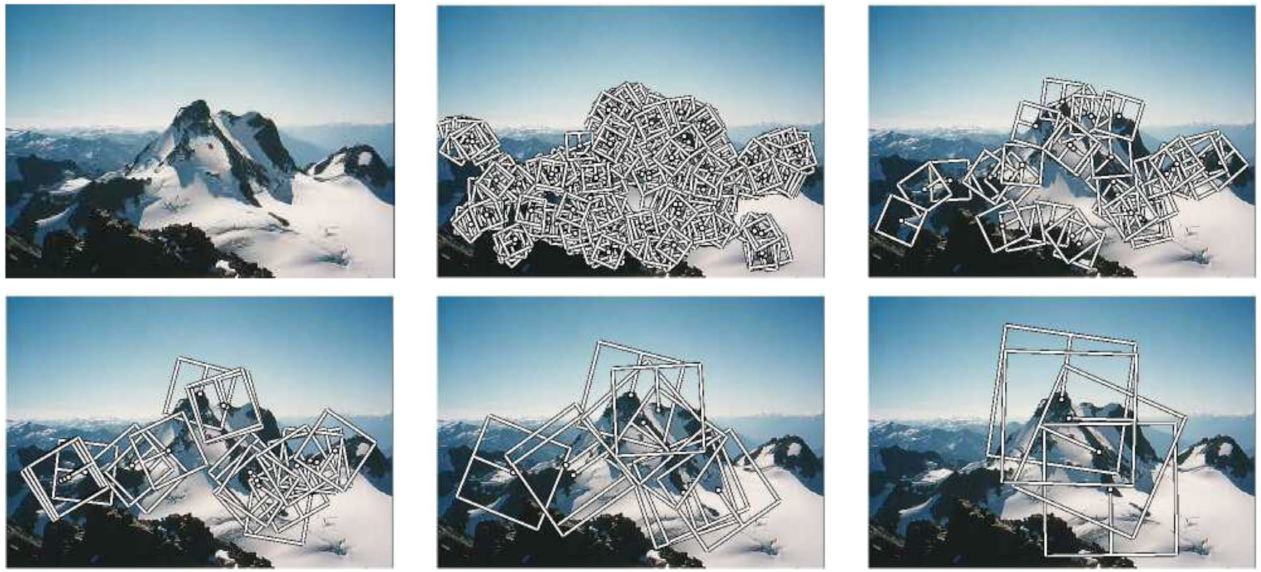


Image pyramid



Multi-scale oriented patches (MOPS) extracted at five pyramid levels (Brown, Szeliski, and Winder 2005)

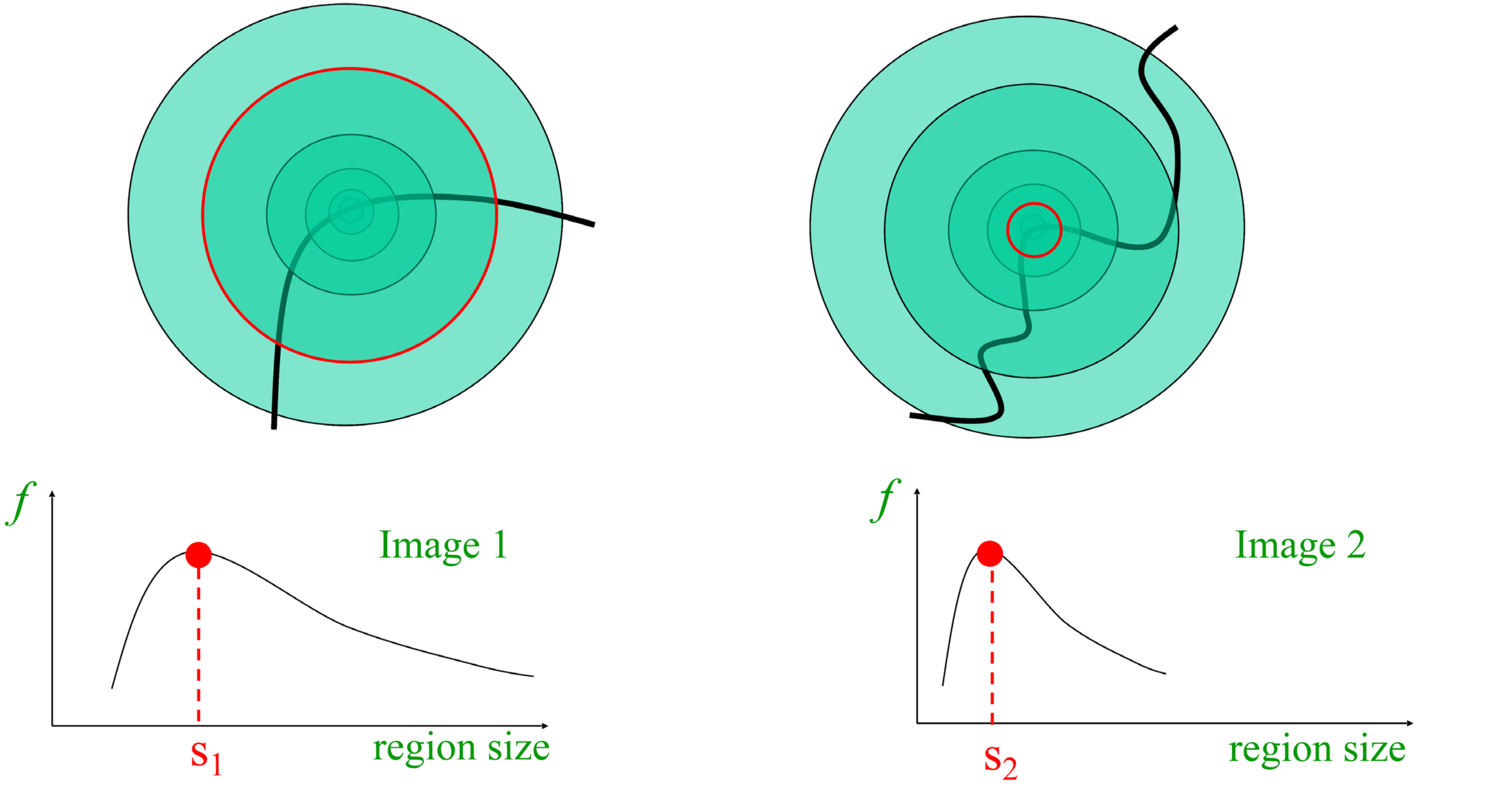
Scale Invariance

- Solution 2: detect features that are stable in both location and scale

Consider Harris corner detector

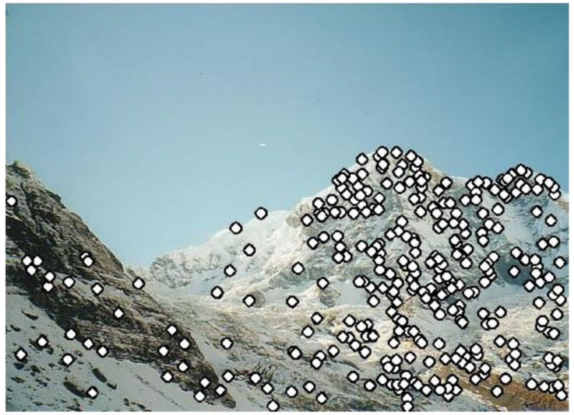
Intuition: Find local maxima in both position and scale

What filter can we use for scale selection?

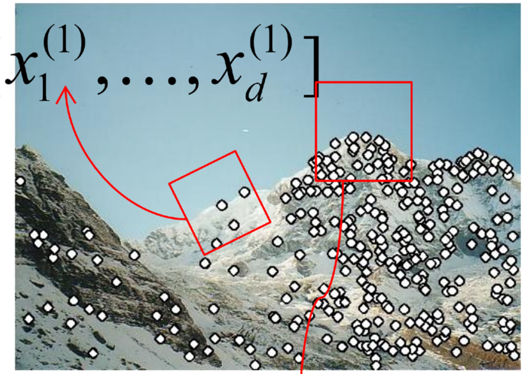


Scale Invariance Feature Transform (SIFT)

- Keypoint detection



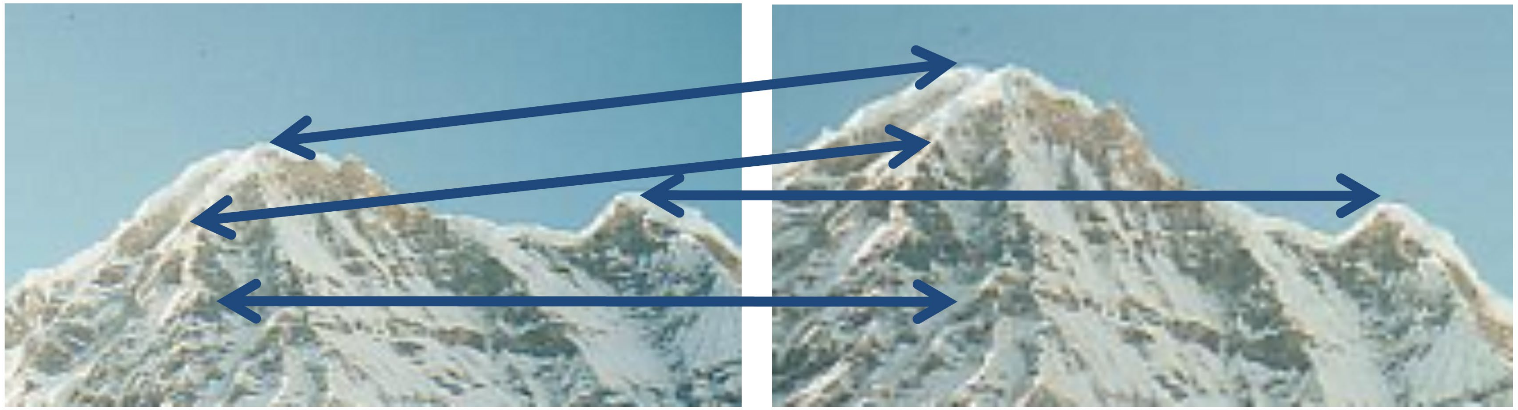
$$\mathbf{x}_1 = [x_1^{(1)}, \dots, x_d^{(1)}]$$



$$\mathbf{x}_2 = [x_1^{(2)}, \dots, x_d^{(2)}]$$

- Compute descriptors

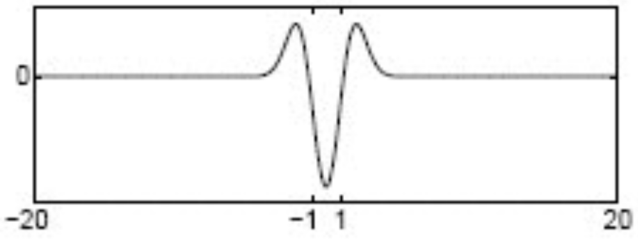
- Matching descriptors



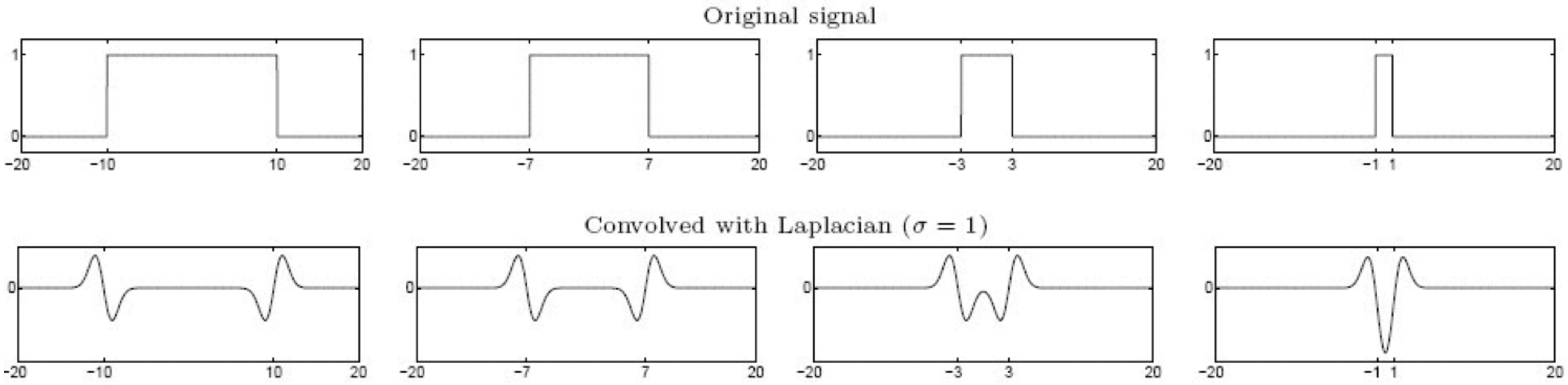
David Lowe, Distinctive Image Features from Scale-Invariant Keypoints. IJCV, 2004

Laplacian of Gaussian for Scale Selection

Laplacian filter

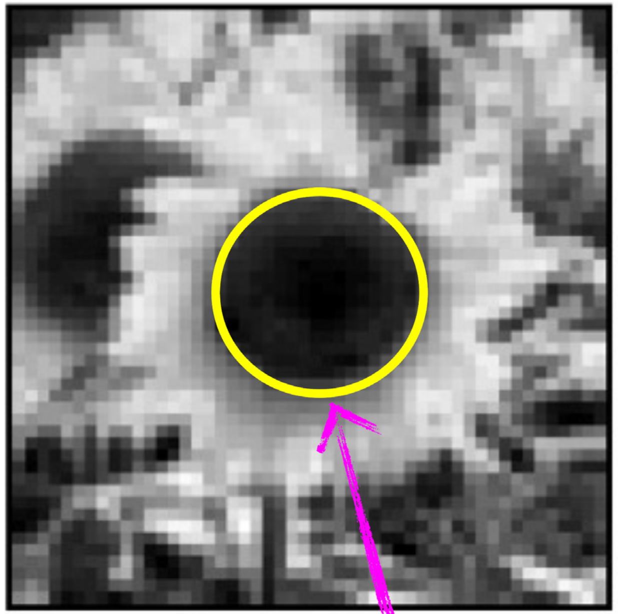


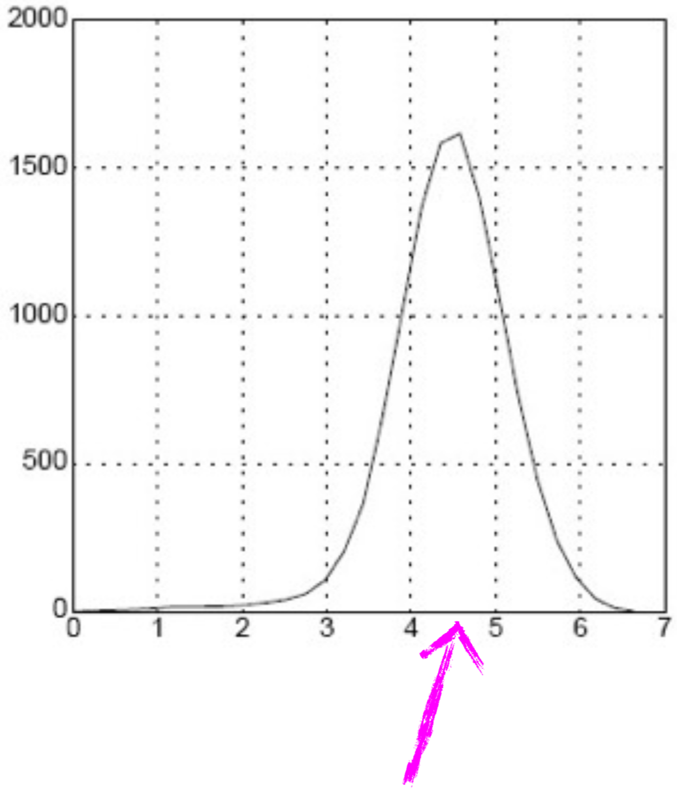
$$g''(x) = \left(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}\right)e^{-\frac{x^2}{2\sigma^2}}$$

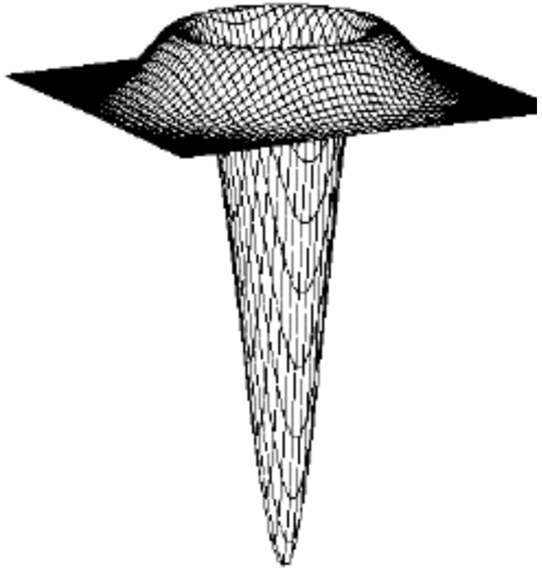


Highest response when the signal has the same **characteristic scale** as the filter

Laplacian of Gaussian for Scale Selection







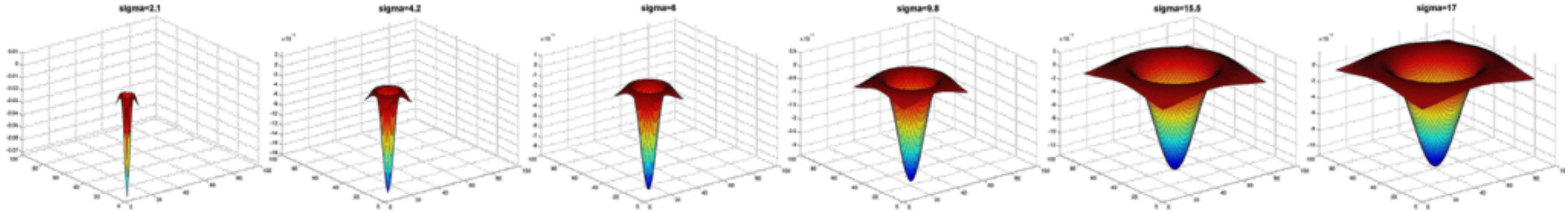
$\nabla^2 h_\sigma(u, v)$

Laplacian of Gaussian

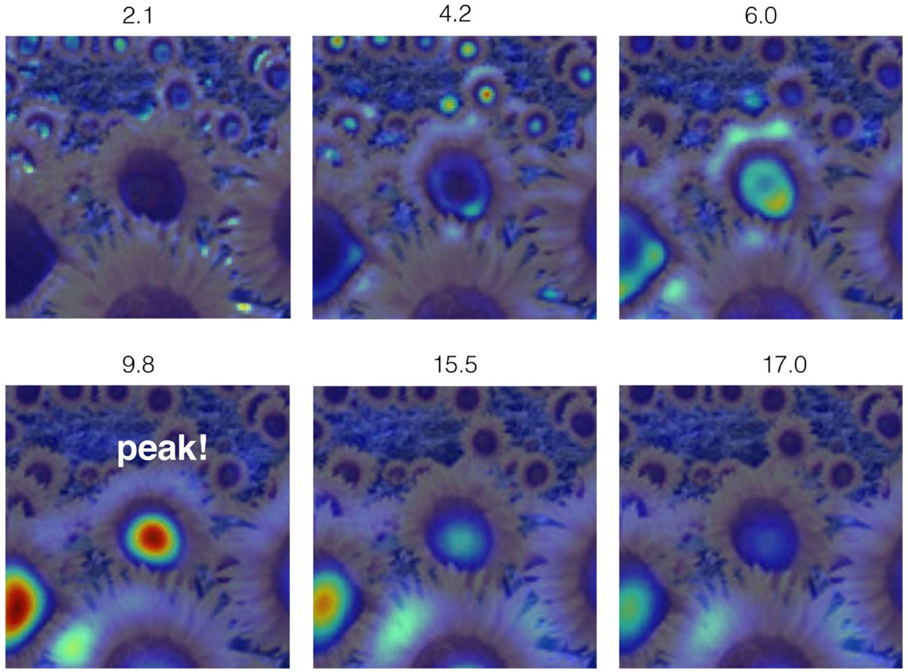
characteristic scale

Search over different scales σ

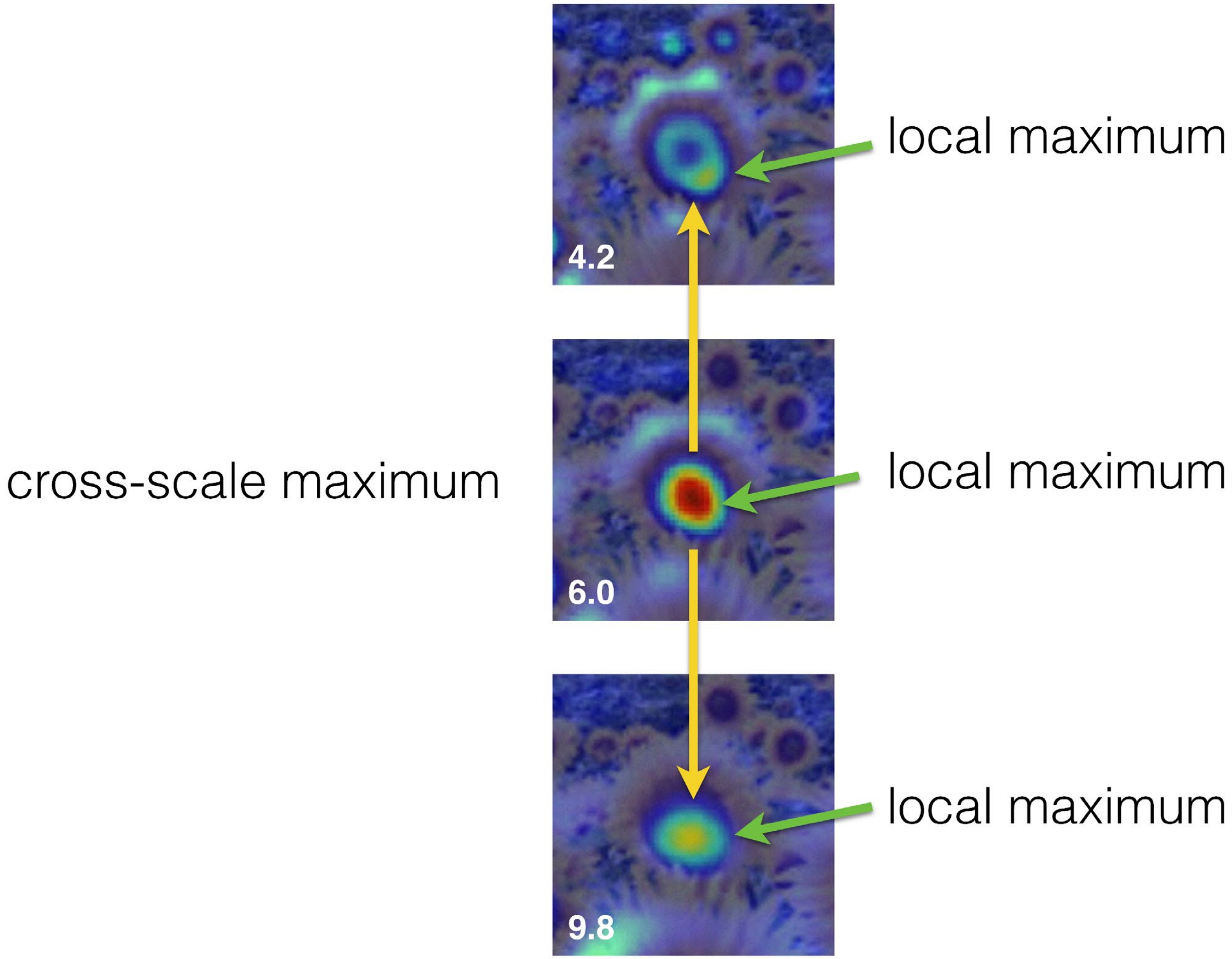
Laplacian of Gaussian for Scale Selection



Multi-scale
2D Blob detection



Laplacian of Gaussian for Scale Selection



Cascaded Gaussians

- Repeated convolution by a smaller Gaussian to simulate effects of a larger one

$$G*(G*f) = (G*G)*f \quad \text{[associative]}$$

$$G_{\sigma_1} * G_{\sigma_2} = G_{\sigma} \qquad \sigma^2 = \sigma_1{}^2 + \sigma_2{}^2$$

- Explanation sketch: convolution in spatial domain is multiplication in frequency domain (Fourier space). Fourier transform of Gaussian is

$$\mathcal{F}_x\left[e^{-\frac{x^2}{2\sigma^2}}\right] = e^{-2\pi^2\sigma^2u^2}$$

$$e^{-2\pi^2\sigma_1^2u^2}e^{-2\pi^2\sigma_2^2u^2} = e^{-2\pi^2(\sigma_1^2+\sigma_2^2)u^2}$$

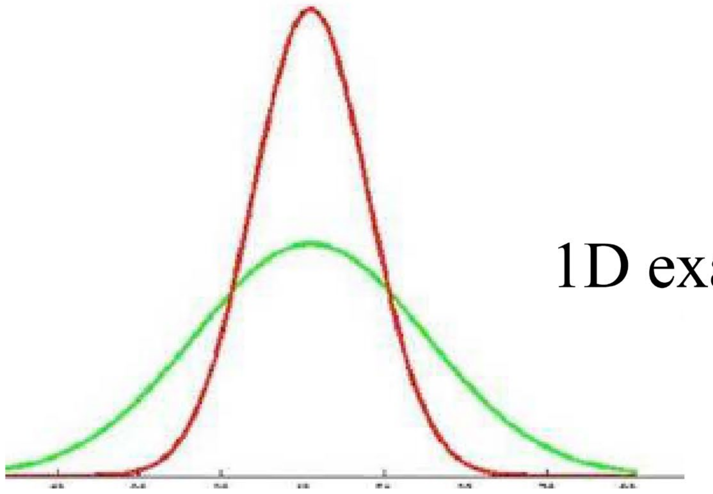
Approximating LoG with DoG

- LoG can be approximate by a Difference of two Gaussians (DoG) at different scales

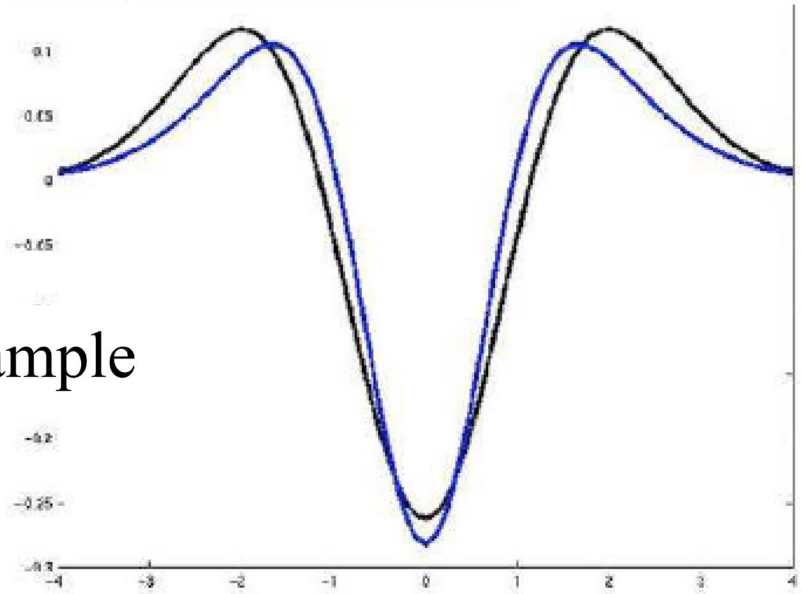
$$\nabla^2 G_\sigma \approx G_{\sigma_1} - G_{\sigma_2}$$

Best approximation when:

$\sigma_1 = \frac{\sigma}{\sqrt{2}}, \ \sigma_2 = \sqrt{2}\sigma$



1D example



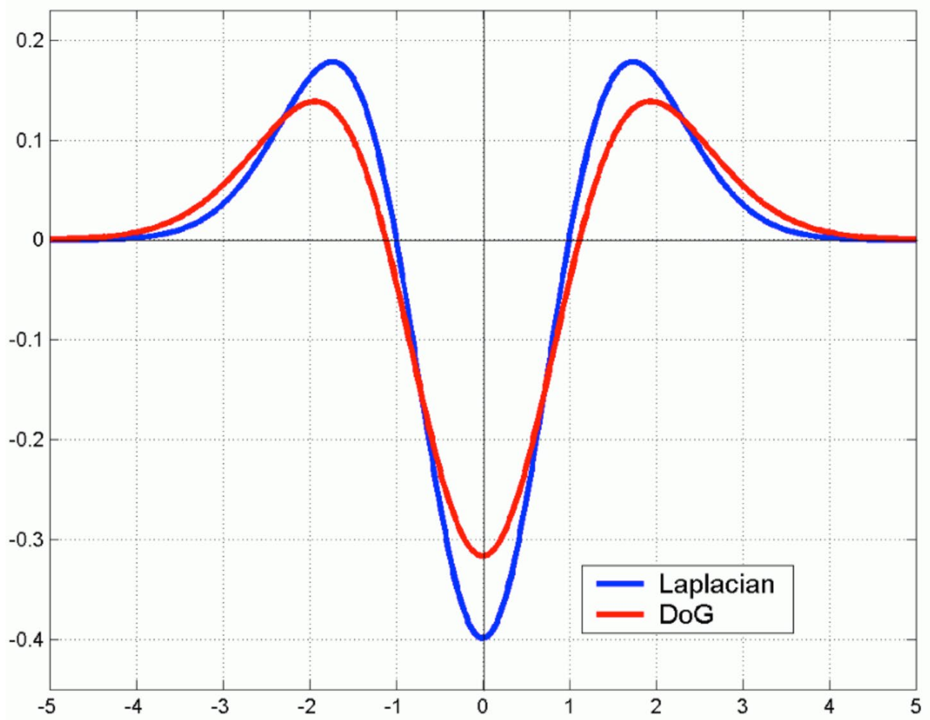
SIFT: Scale-space Extrema Detection

- Difference of Gaussian (DoG)

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

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

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned}$$

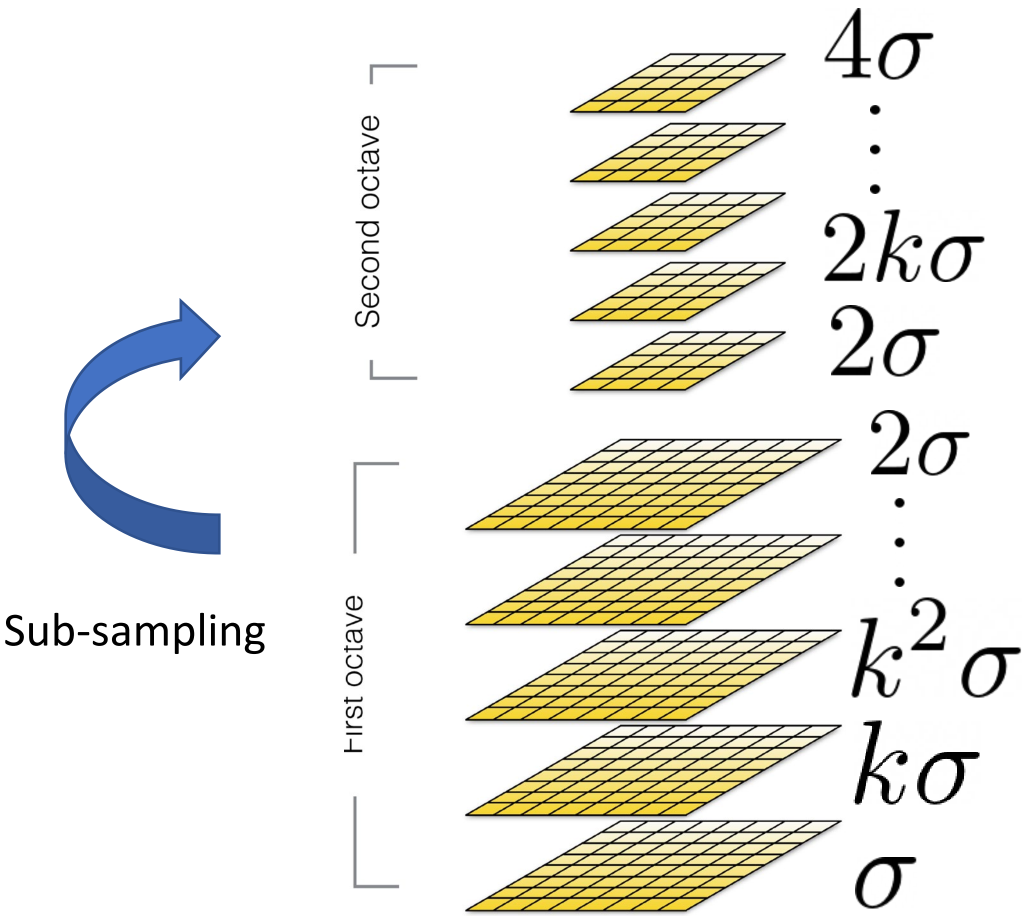


Approximate of Laplacian of Gaussian
(efficient to compute)

k is a constant

SIFT: Scale-space Extrema Detection

- Gaussian pyramid



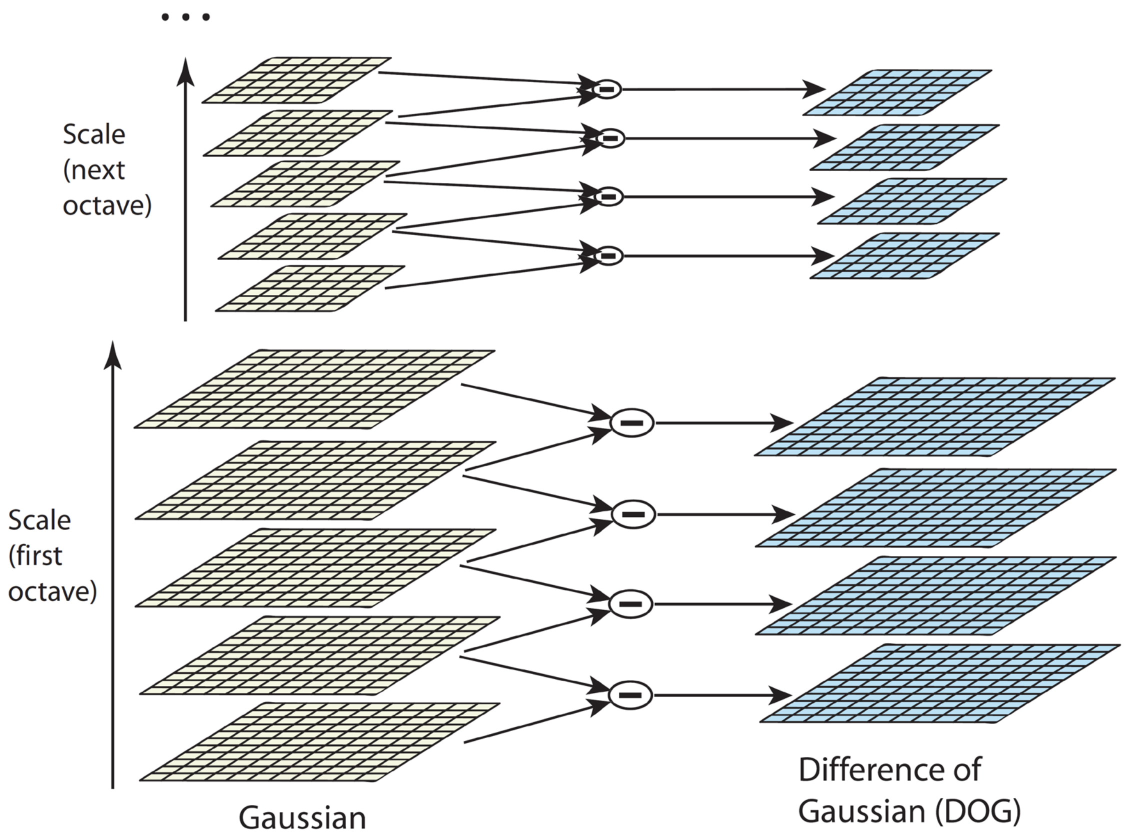
- Gaussian filters

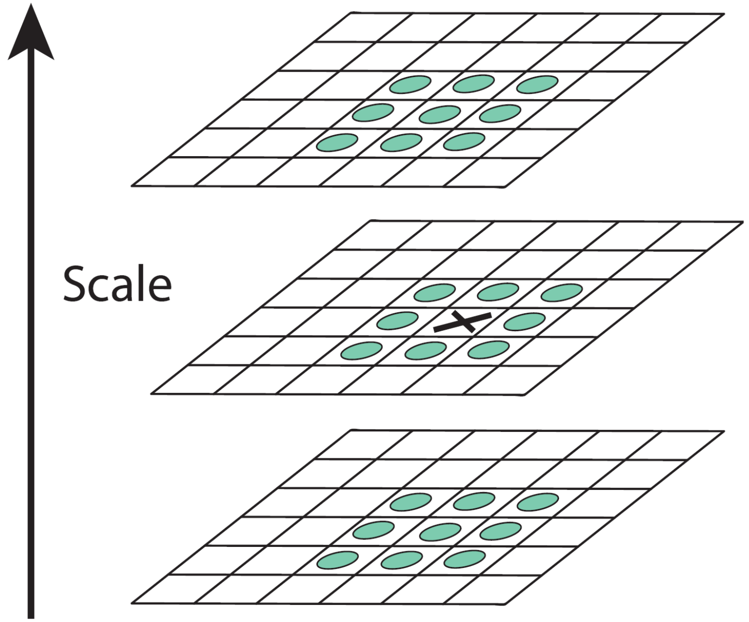
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

$$G_{\sigma_1} * G_{\sigma_2} = G_{\sigma} \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2$$

- Sub-sampling by a factor of 2
  - Multiple the Gaussian kernel deviation by 2

SIFT: Scale-space Extrema Detection





Maxima and minima of DoG images

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

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

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned}$$

Further Reading

- Section 7.1, Computer Vision, Richard Szeliski
- David Lowe, Distinctive Image Features from Scale-Invariant Keypoints. IJCV, 2004 https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf
- ORB: An efficient alternative to SIFT or SURF. Rublee et al., ICCV, 2011