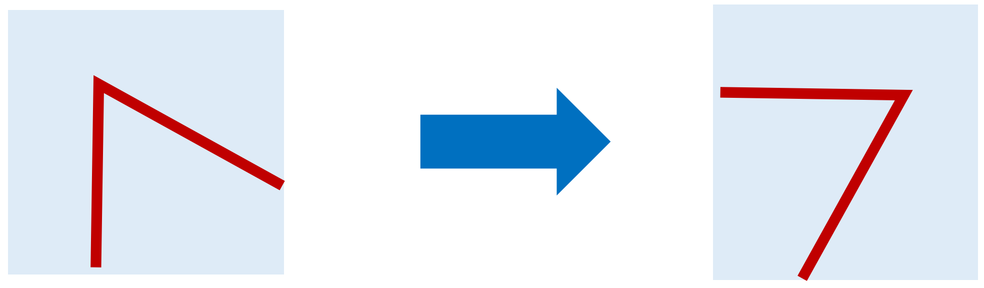
Features and Scale Space

Chetan Arora

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Harris Corner Detector: Invariances

- Translation Invariant?
- Rotation Invariant?
- Scale Invariant?
- Illumination Invariant?



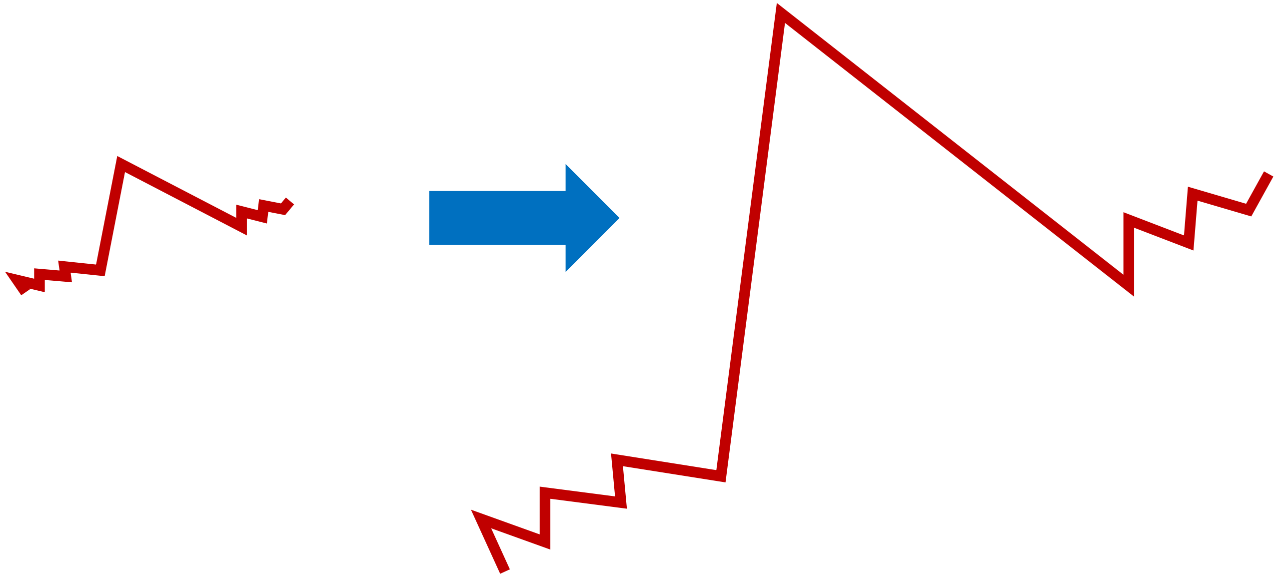


Image Pyramids

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Recall: Cascaded Convolutions

$$[1 \quad 1]^n$$

$$[1 \quad 1]$$

$$[1 \quad 2 \quad 1]$$

$$[1 \quad 3 \quad 3 \quad 1]$$

$$[1 \quad 4 \quad 6 \quad 4 \quad 1]$$

Interesting Aside

$$[1 \quad 1]^n$$

$$[1 \quad 1]$$

$$[1 \quad 2 \quad 1]$$

$$[1 \quad 3 \quad 3 \quad 1]$$

$$[1 \quad 4 \quad 6 \quad 4 \quad 1]$$

$$a_{nr} = \frac{n!}{r!\,(n-r)!} = \binom{n}{r}$$

n = number of elements in the filter minus 1

r = position of element in the filters (starting at 0)

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Interesting Aside

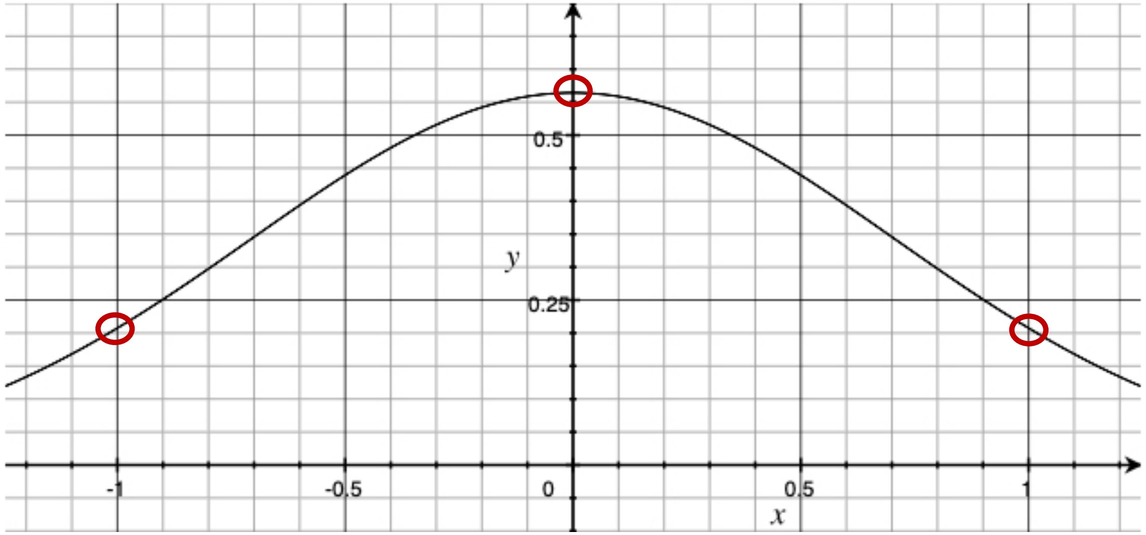
$$\frac{1}{2}[1 \quad 1]^n$$

$$\frac{1}{2}[1 \quad 1]$$

$$\frac{1}{4}[1 \quad 2 \quad 1]$$

$$\frac{1}{8}[1 \quad 3 \quad 3 \quad 1]$$

$$\frac{1}{16}[1 \quad 4 \quad 6 \quad 4 \quad 1]$$



$$\frac{1}{4}[1 \quad 2 \quad 1]$$

approximates Gaussian with $\sigma = \frac{1}{\sqrt{2}}$

$$\frac{1}{16}[1 \quad 4 \quad 6 \quad 4 \quad 1]$$

approximates Gaussian with $\sigma = 1$

Why?

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Interesting Aside

Fun facts:

- The distribution of the sum of two random variables $X + Y$ is the convolution of their two distributions. **Why?**

- Recall Convolution: $(f * g)(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$

- $z = x + y \Rightarrow y = z - x$

- $p_{X+Y}(z) = \int_{-\infty}^{\infty} p_y(z-x)\, p_x(x)\, dx$

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Interesting Aside

$X\colon\ G_{\sigma_1}(x)$ and $Y\colon\ G_{\sigma_2}(y)$

$$G_\sigma(x+y) = (G_{\sigma_1} * G_{\sigma_2})(\mathrm{z})$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2$$

$$\frac{1}{4}[1 \quad 2 \quad 1]$$

approximates Gaussian with $\sigma = \frac{1}{\sqrt{2}}$

$$\frac{1}{16}[1 \quad 4 \quad 6 \quad 4 \quad 1]$$

approximates Gaussian with $\sigma = 1$

Gaussian Smoothing at Different Scales

$\sigma = 3$

$\sigma = 10$

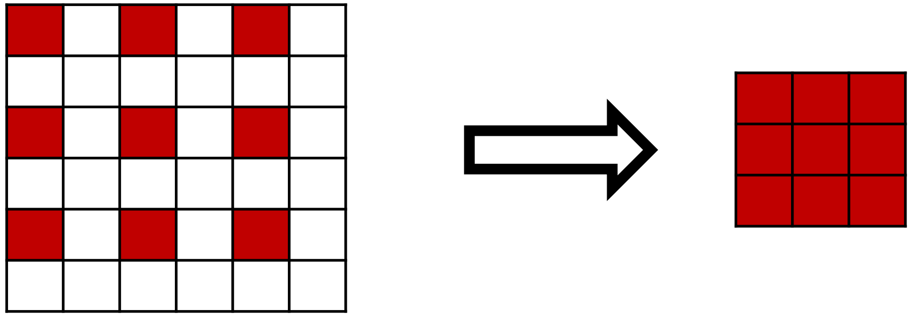
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Subsampling

- A large amount of smoothing reduces the rate of change of neighboring pixels (frequency of features) in the image.

- We do not need to keep all the pixels around. Can progressively reduce the number of pixels as we smooth more and more.

- Leads to a “pyramid” representation if we ‘subsample’ at each level.

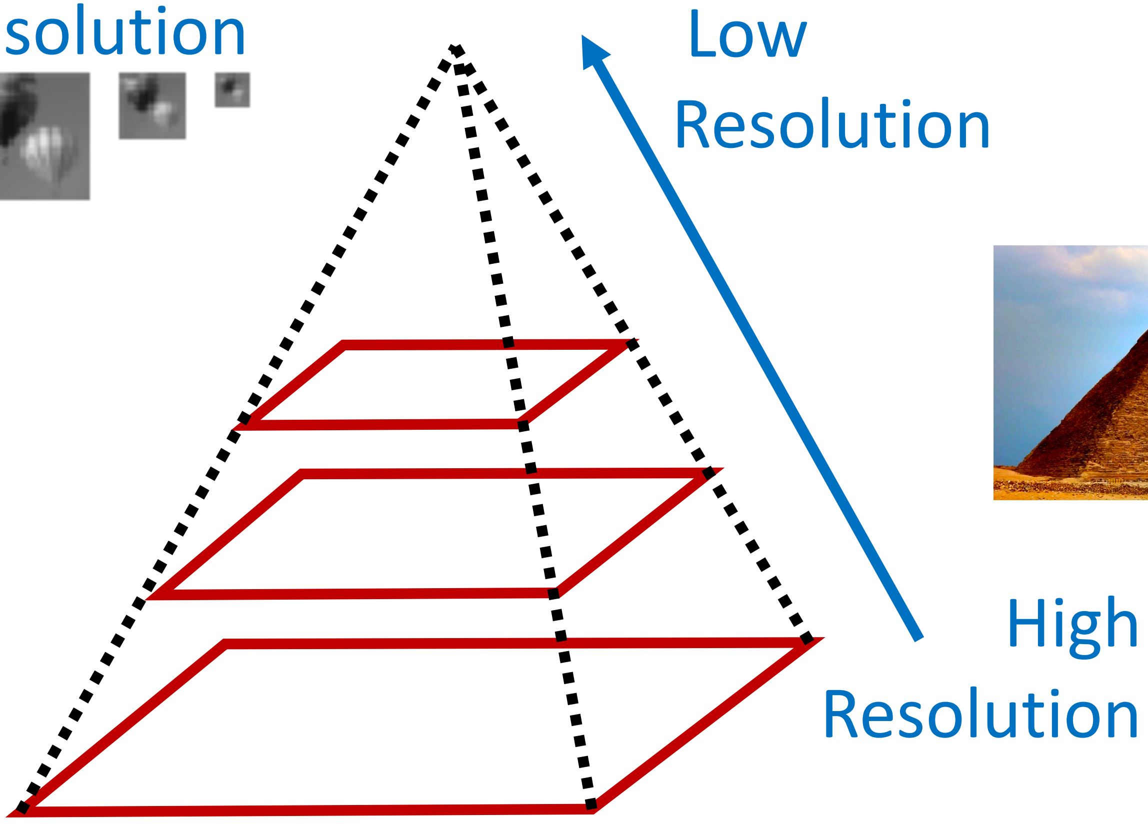


Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Gaussian Pyramid

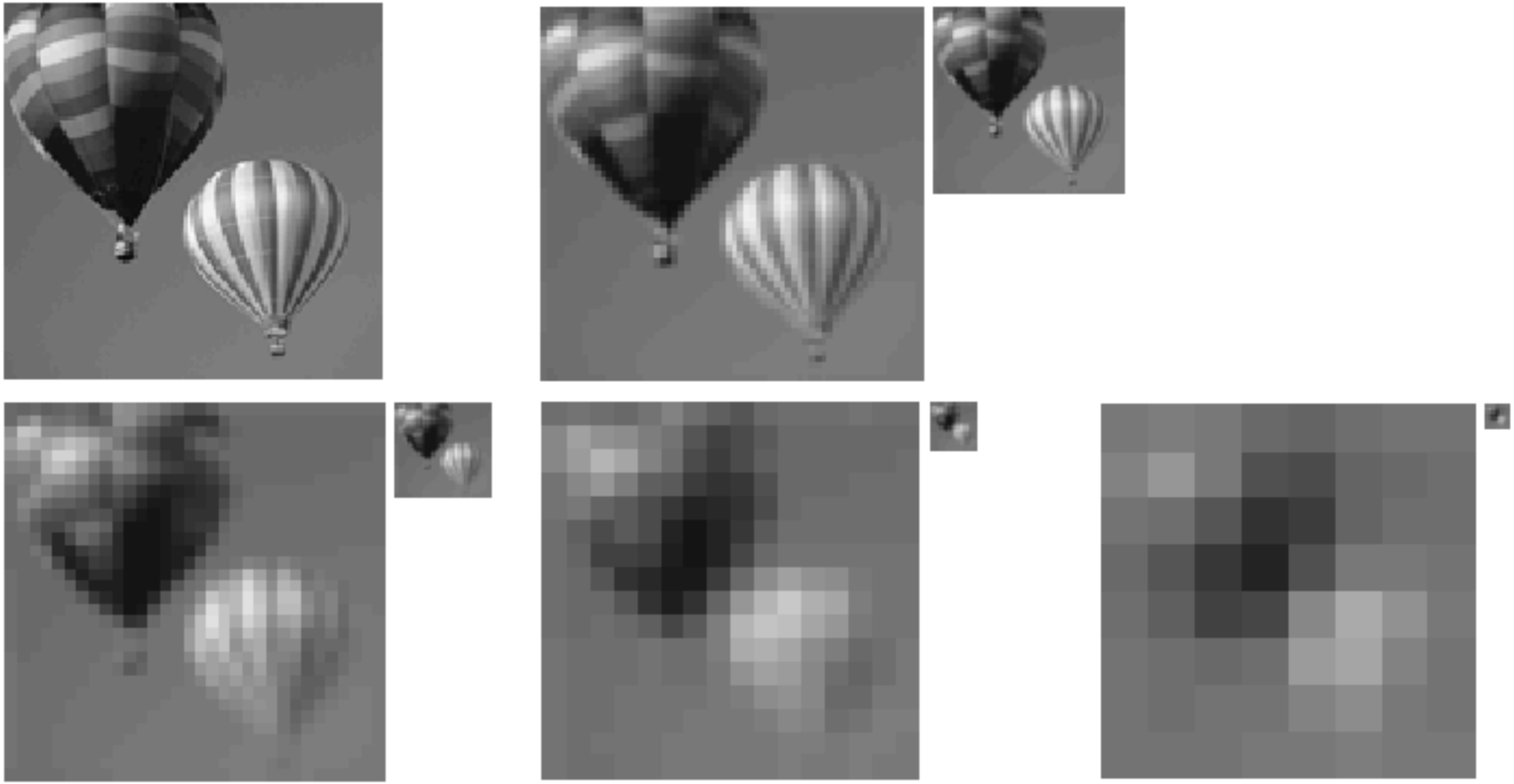
- Synthesis: Smooth image with a Gaussian. Downsample. Repeat.

High Resolution → Low Resolution



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Smaller Image → Lower Resolution



Interesting Aside

- Subsampling without prior blurring is a bad idea (introduces aliasing)!

Downsampled

Blurred and Downsampled

Interesting Aside

- Subsampling without prior blurring is a bad idea (introduces aliasing)!

Downsampled

Blurred and Downsampled

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Interesting Aside

- Subsampling without prior blurring is a bad idea (introduces aliasing)!

Downsampled

Blurred and Downsampled

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

What is Aliasing?

- A digital photograph of a striped shirt, or a brick wall (shown above) with high frequencies, with small distance between the stripes/bricks, can cause aliasing of the shirt/wall when it is sampled by the camera's image sensor. The aliasing appears as a moiré pattern.

- The “solution” to higher sampling in the spatial domain for this case would be to move closer to the shirt, use a higher resolution sensor, or to optically blur the image before acquiring it with the sensor.

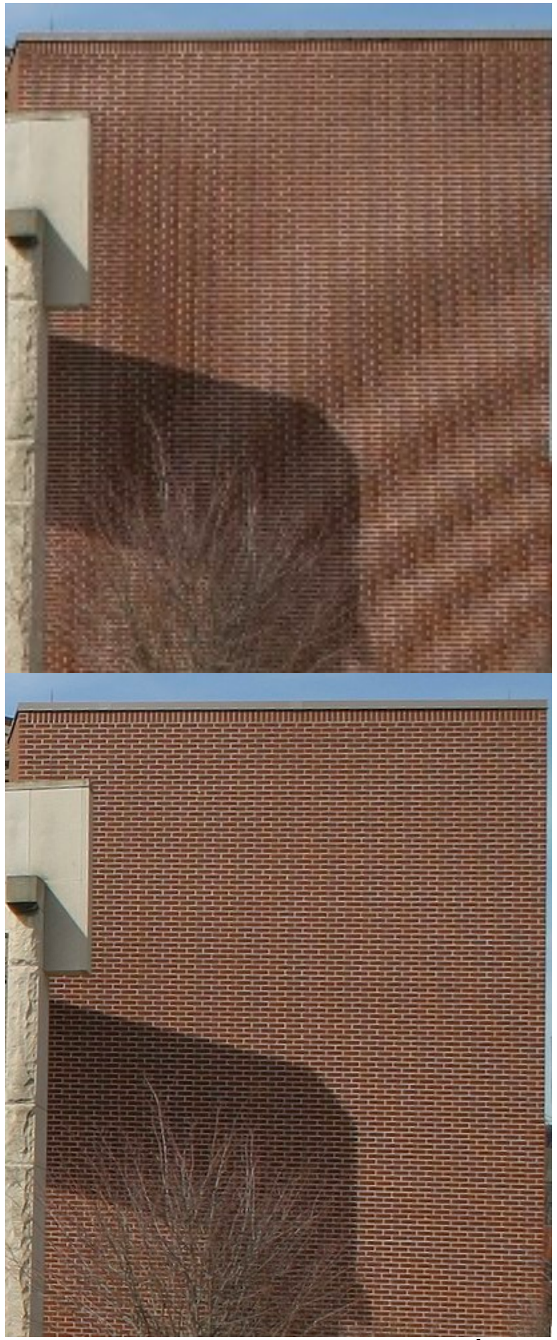
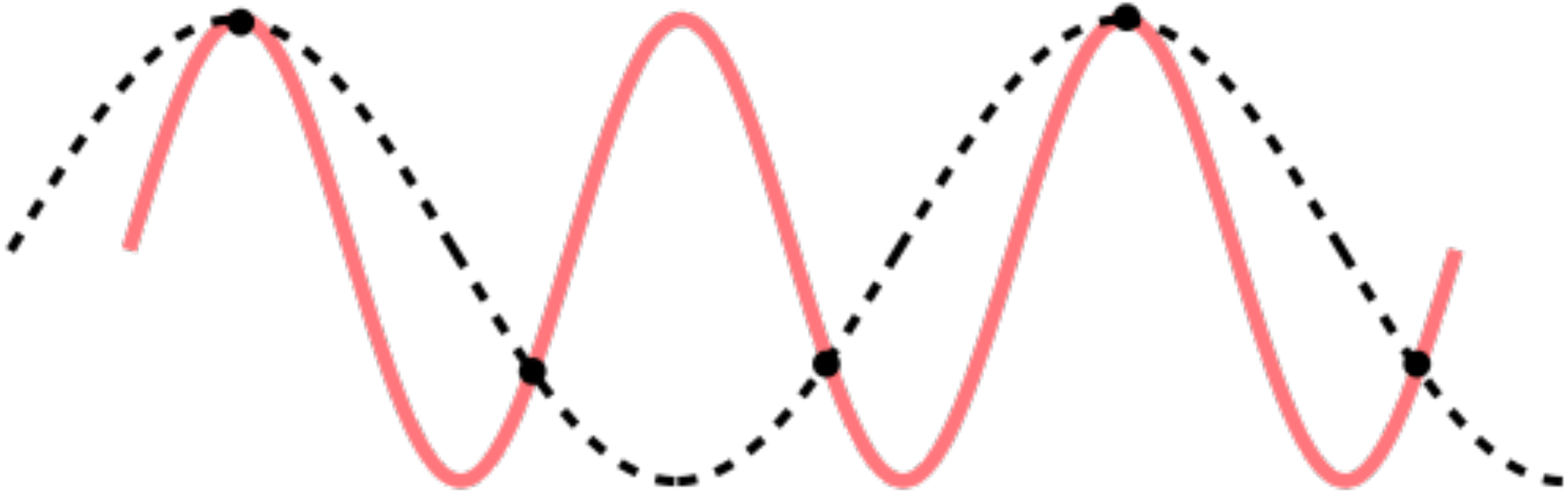


Image source: Wikipedia

What is Aliasing?



The samples of two sine waves can be identical when at least one of them is at a frequency above half the sample rate.

Image source: Wikipedia

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Constructing Gaussian Pyramid

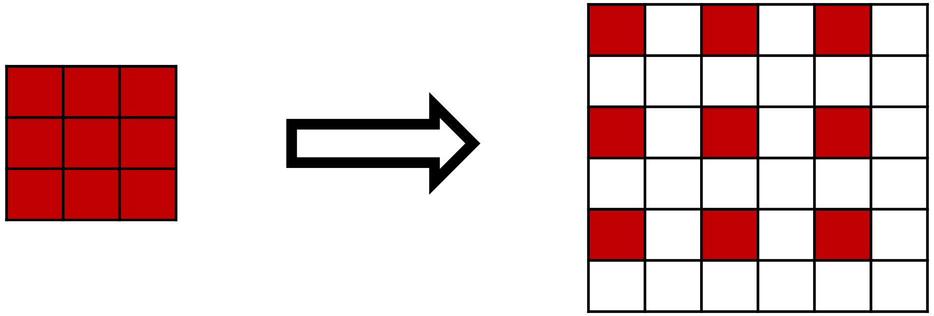
Input Image $*\ g(\sigma = 1) *\ g(\sigma = 1) *\ g(\sigma = 1) *\ g(\sigma = 1)$ Subsample Image $*\ g(\sigma = 1) *\ g(\sigma = 1) *\ g(\sigma = 1) *\ g(\sigma = 1)$ Subsample Image

Effective σ $\quad \sqrt{2} \quad 2 \quad 2\sqrt{2} \quad 4$

$\mathbf{1}$

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Upsampling



How to fill in the empty values?

- Interpolation:
  - Initially set empty pixels to zero
  - Convolve upsampled image with a Gaussian filter.
  - E.g. 5x5 kernel with sigma = 1.
  - Must also multiply by 4. **Why?**

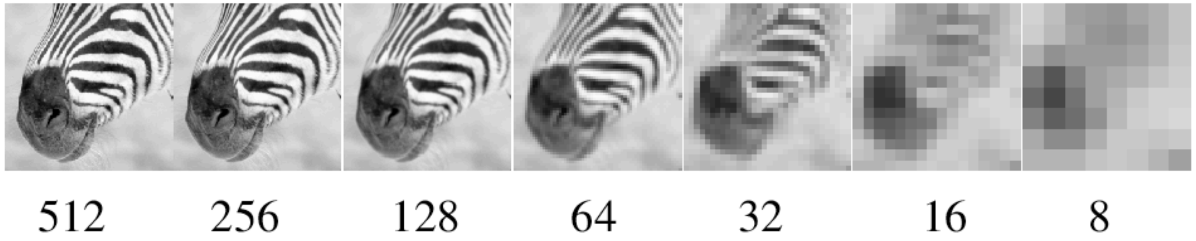
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Scale Space

- Different scales are appropriate for describing different objects in the image

- We do not know the correct scale/size ahead of time.

- Image Pyramids are a useful representation for multi-resolution analysis of images.



Laplacian of Gaussian (LoG)

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Recall: Second Derivative

$$\frac{\partial}{\partial x} f(i,j) \cong f(i,j) - f(i-1,j)$$

$$\frac{\partial}{\partial y} f(i,j) \cong f(i,j) - f(i,j-1)$$

$$\begin{bmatrix} 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\frac{\partial^2}{\partial x^2} f(i,j) \cong \frac{\partial}{\partial x} f(i+1,j) - \frac{\partial}{\partial x} f(i,j)$$

$$= f(i-1,j) + f(i+1,j) - 2f(i,j)$$

$$\begin{bmatrix} 1 & -1 \end{bmatrix} * \begin{bmatrix} 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -2 & 1 \end{bmatrix}$$

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Recall: Laplacian

Equation:

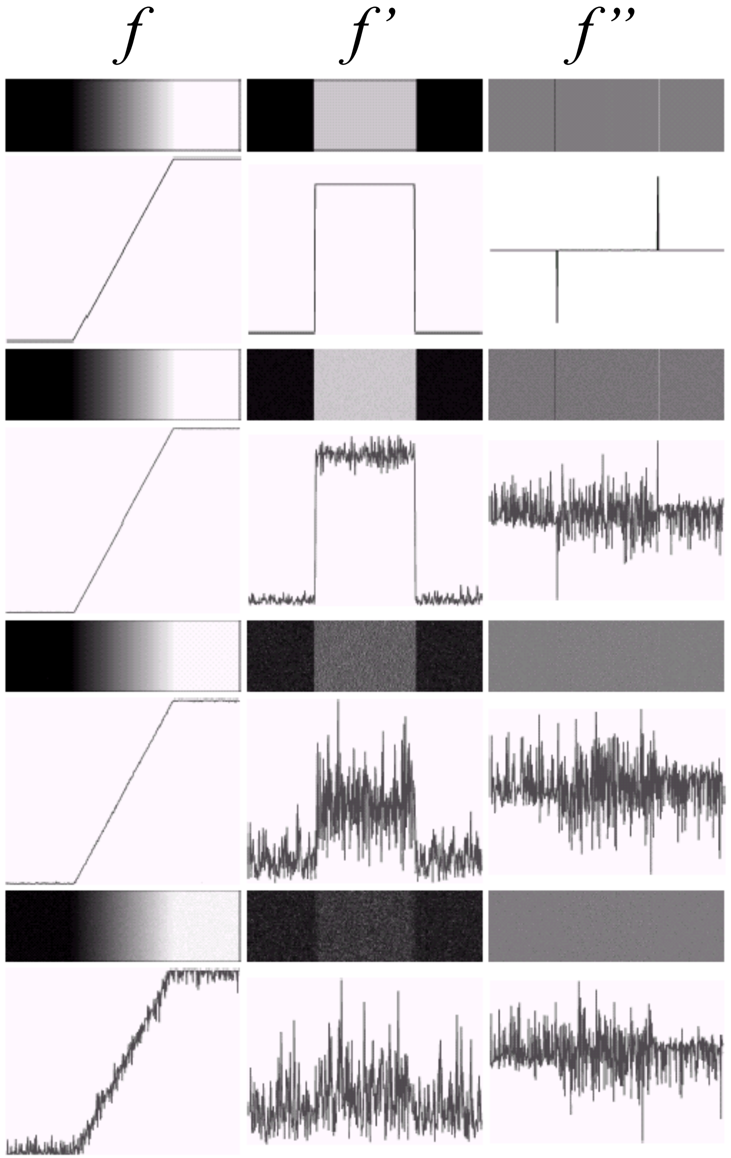
$$\nabla^2 f = \frac{\partial}{\partial x^2} f + \frac{\partial}{\partial y^2} f$$

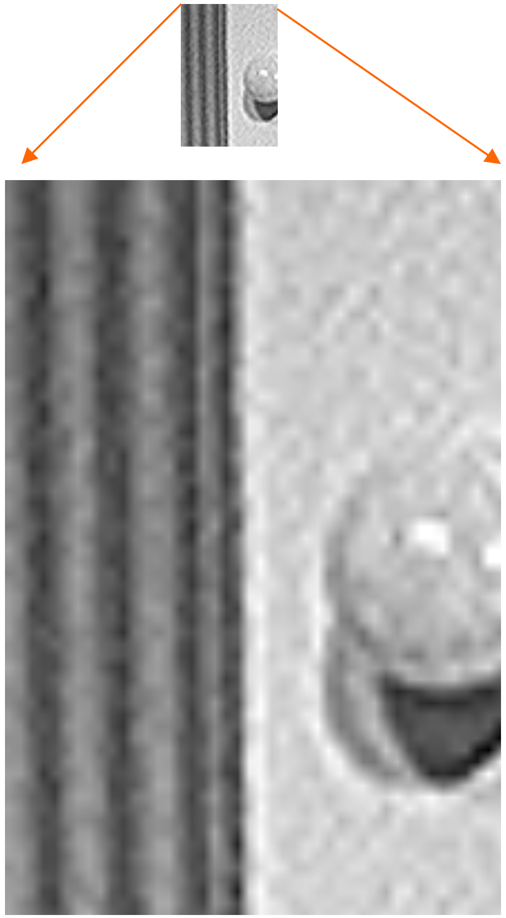
Convolution:

$$\begin{bmatrix} 1 & -2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Recall: Effect of Noise on Derivatives





Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Recall: Zero Crossing After Smoothing

Input

Laplacian after increasingly strong smoothing

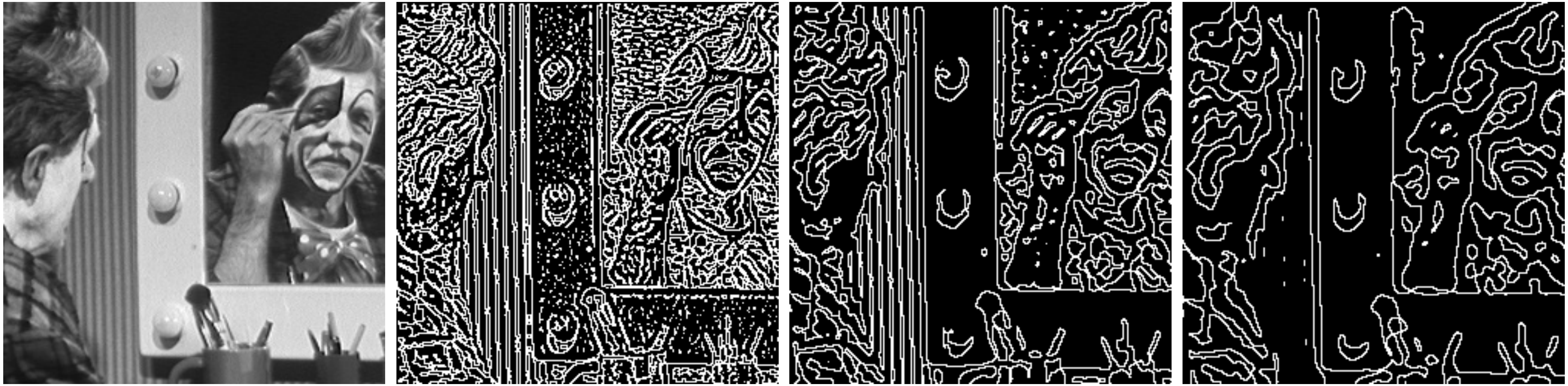


Image ⇨ Gaussian ⇨ Laplacian ⇨ Output

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

LoG Filter

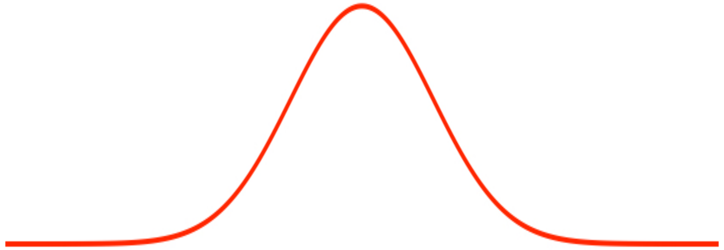
Image $\Rightarrow$ Gaussian $\Rightarrow$ Laplacian $\Rightarrow$ Output

$$\nabla^2(f * G)(x) \equiv (\nabla^2 G * f)(x)$$

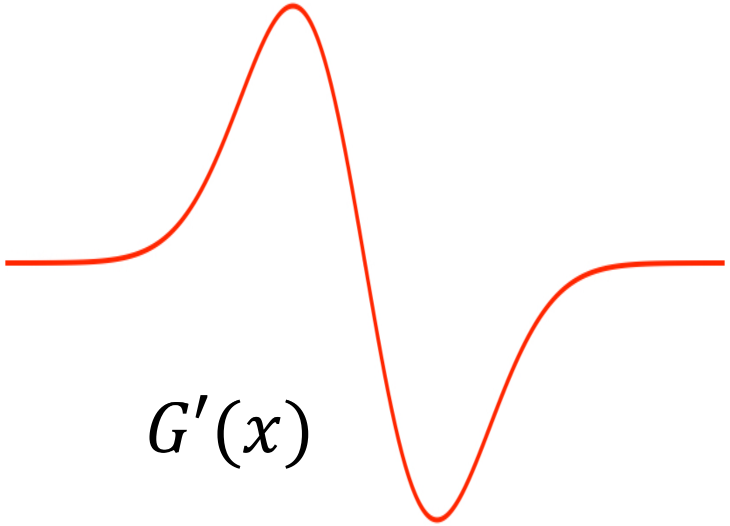
Laplacian of Gaussian
(LoG) Filter

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

LoG Filter

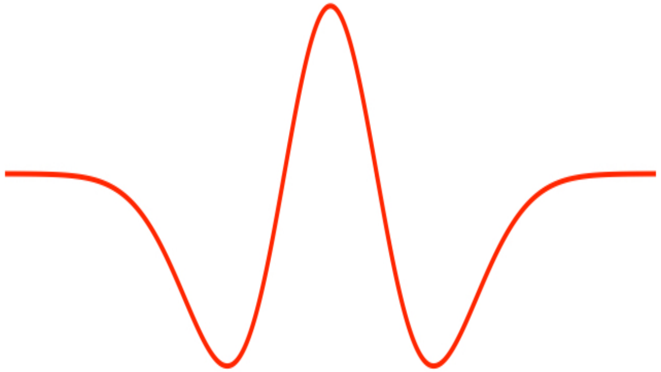


$G(x)$



$G'(x)$

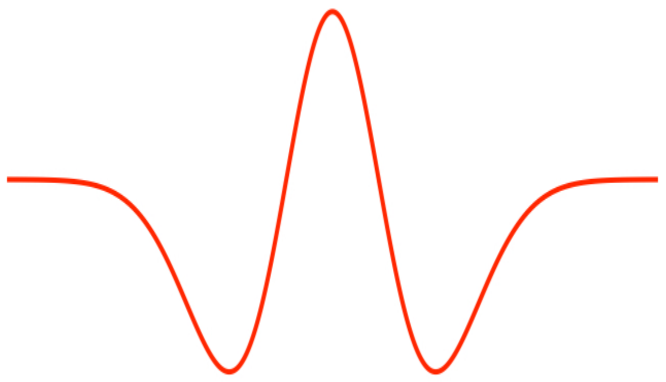
Mexican Hat

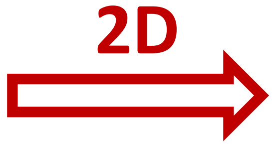


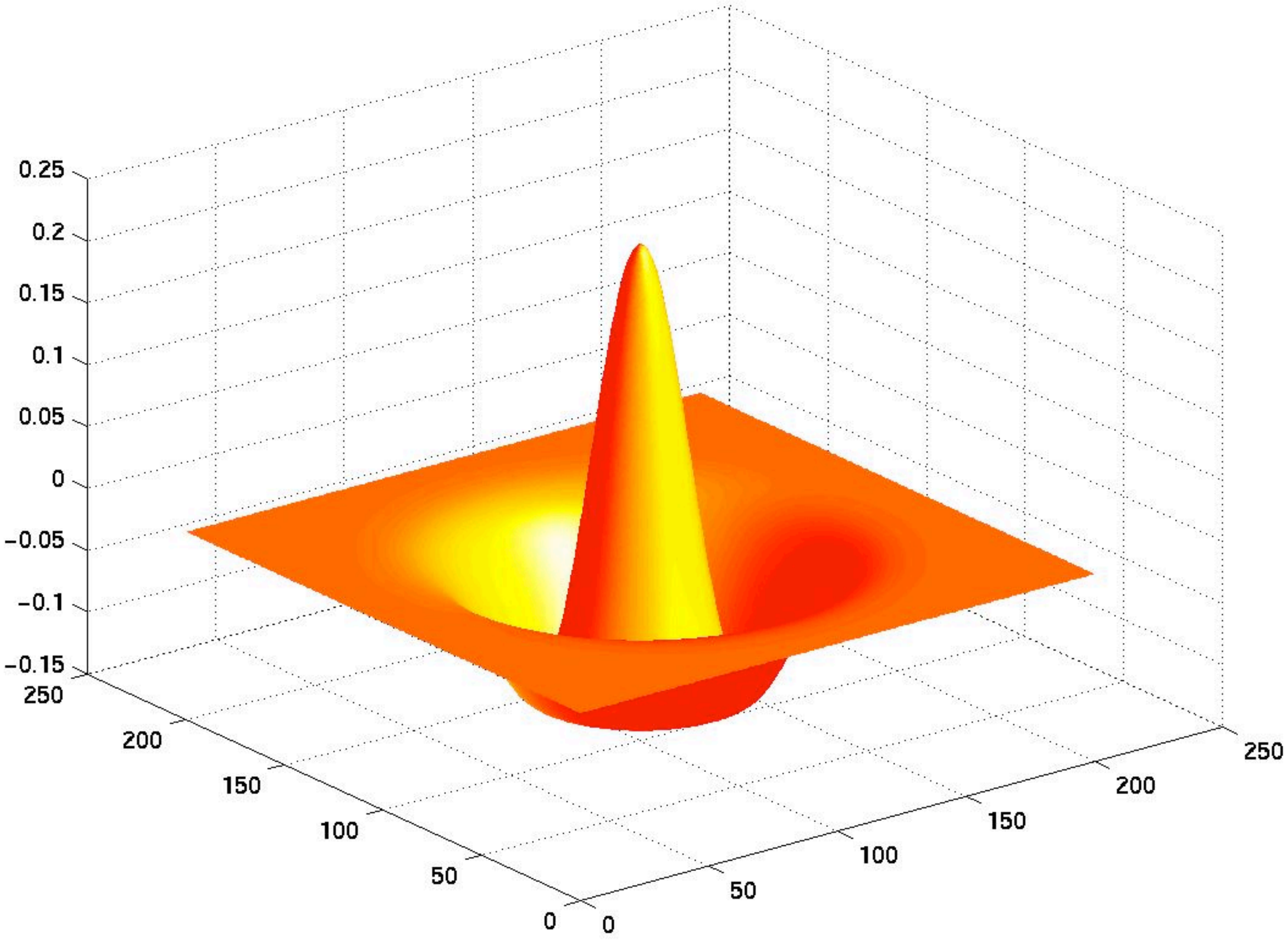
$-G''(x)$

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

LoG Filter



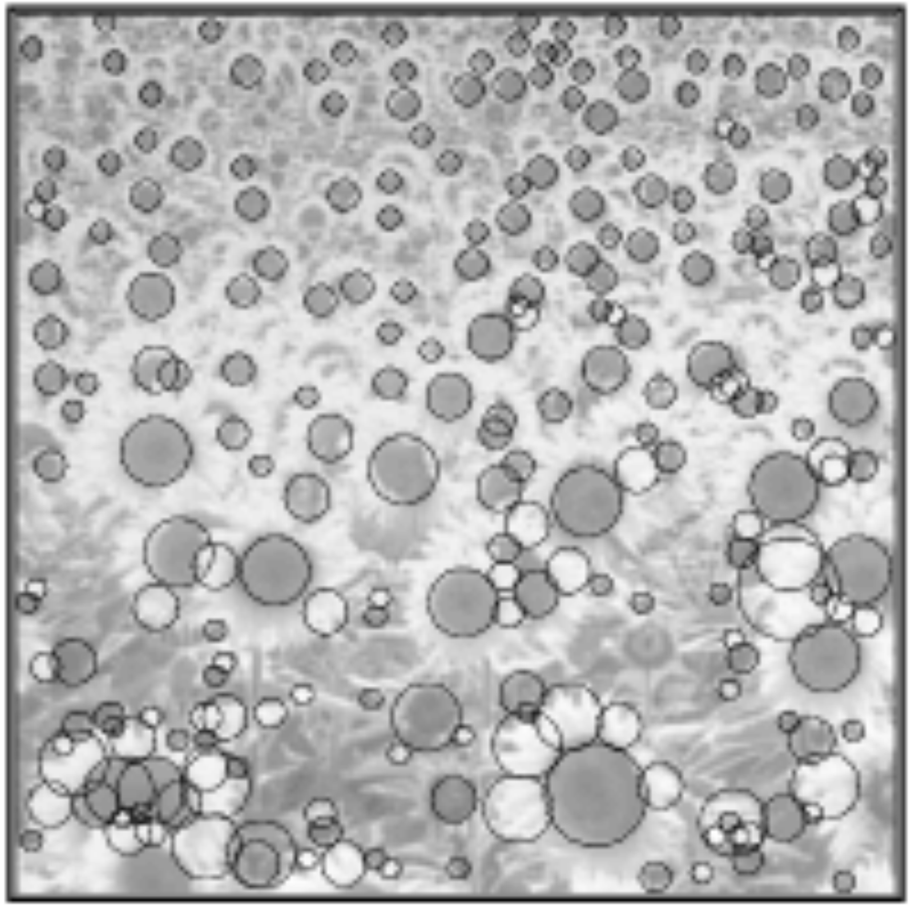




Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Blob Detection using LoG

- How can an edge finder be used to find blobs in an image?

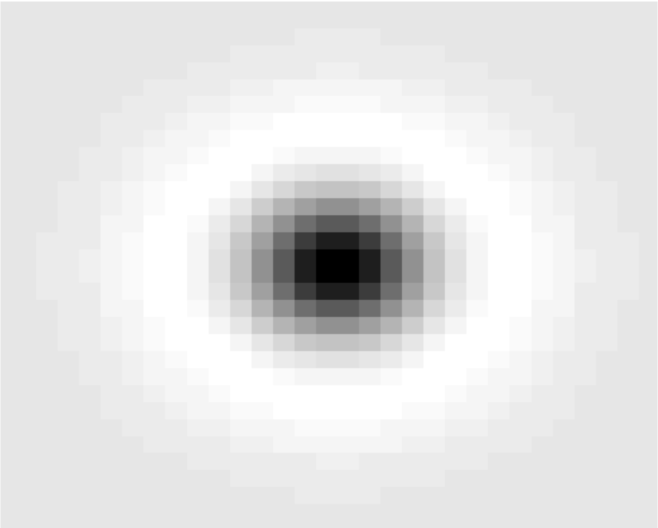


Lindeberg. IJCV 1998.
Feature detection with automatic scale selection

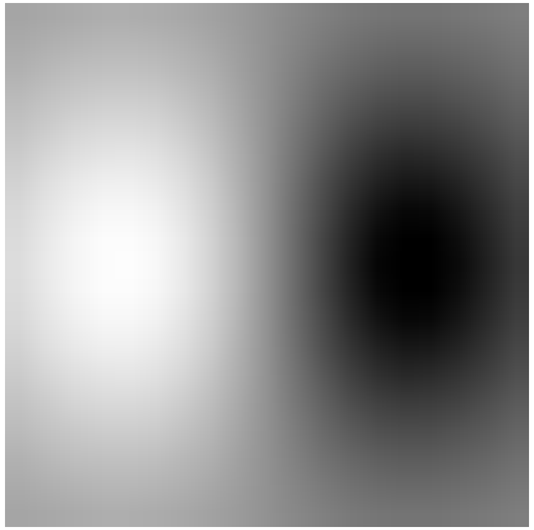
Blob Detection using LoG

Key Idea

- Cross correlation with a flipped filter (in a convolution) can be viewed as comparing a little “picture” of what you want to find against all local regions in the image.



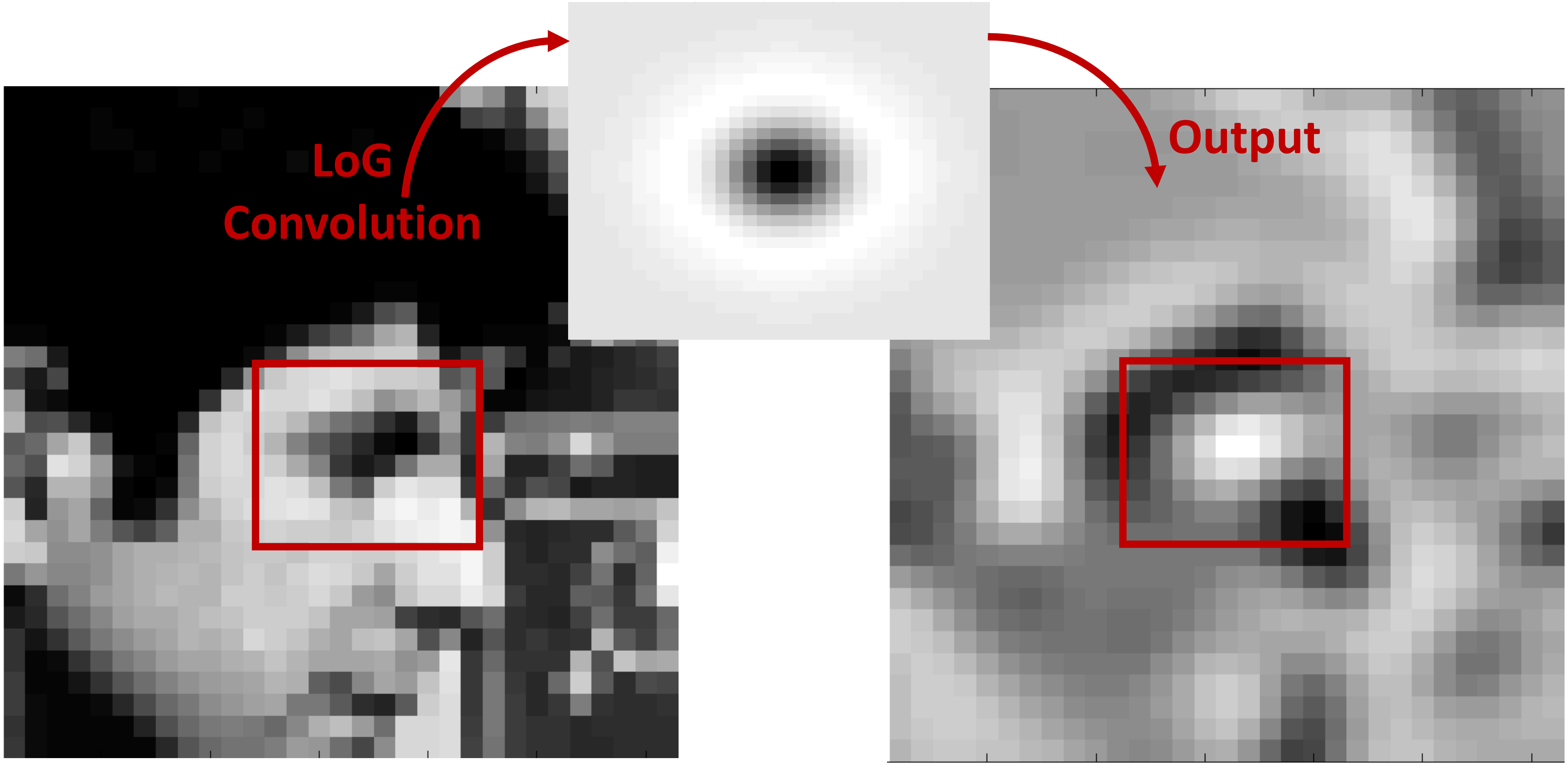
Maximum response
Dark blob on light background



Maximum response
Vertical edge; lighter on left

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

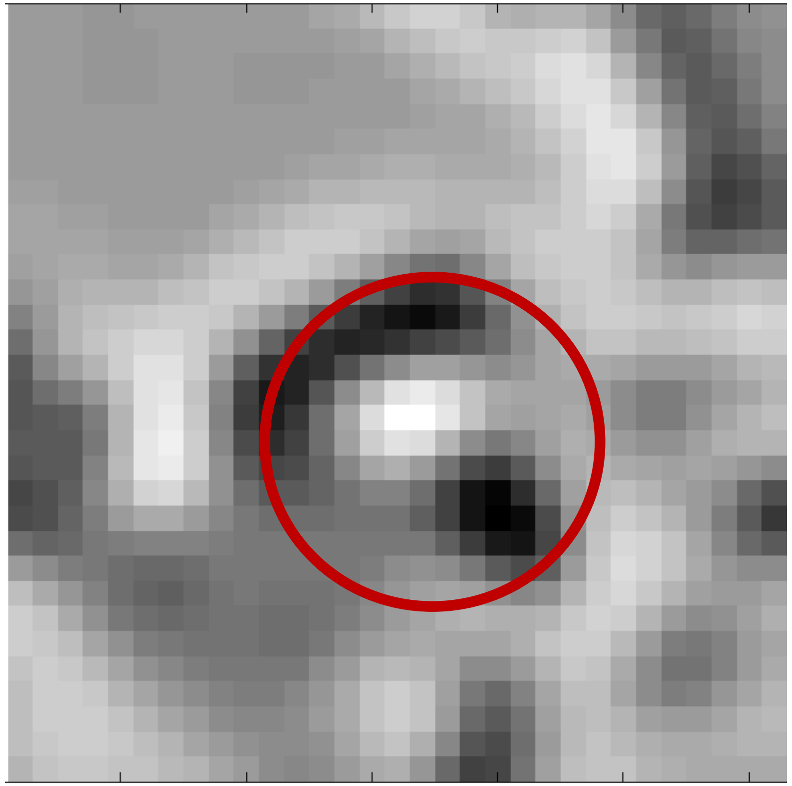
Blob Detection using LoG



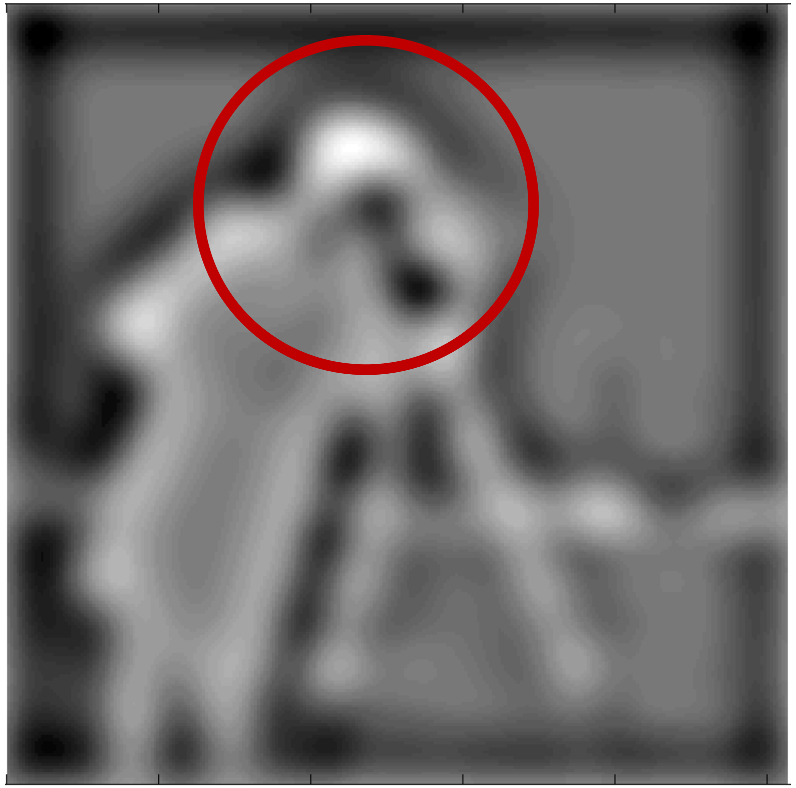
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Blob Detection using LoG

- Scale of blob (size ; radius in pixels) is determined by the sigma parameter of the LoG filter.



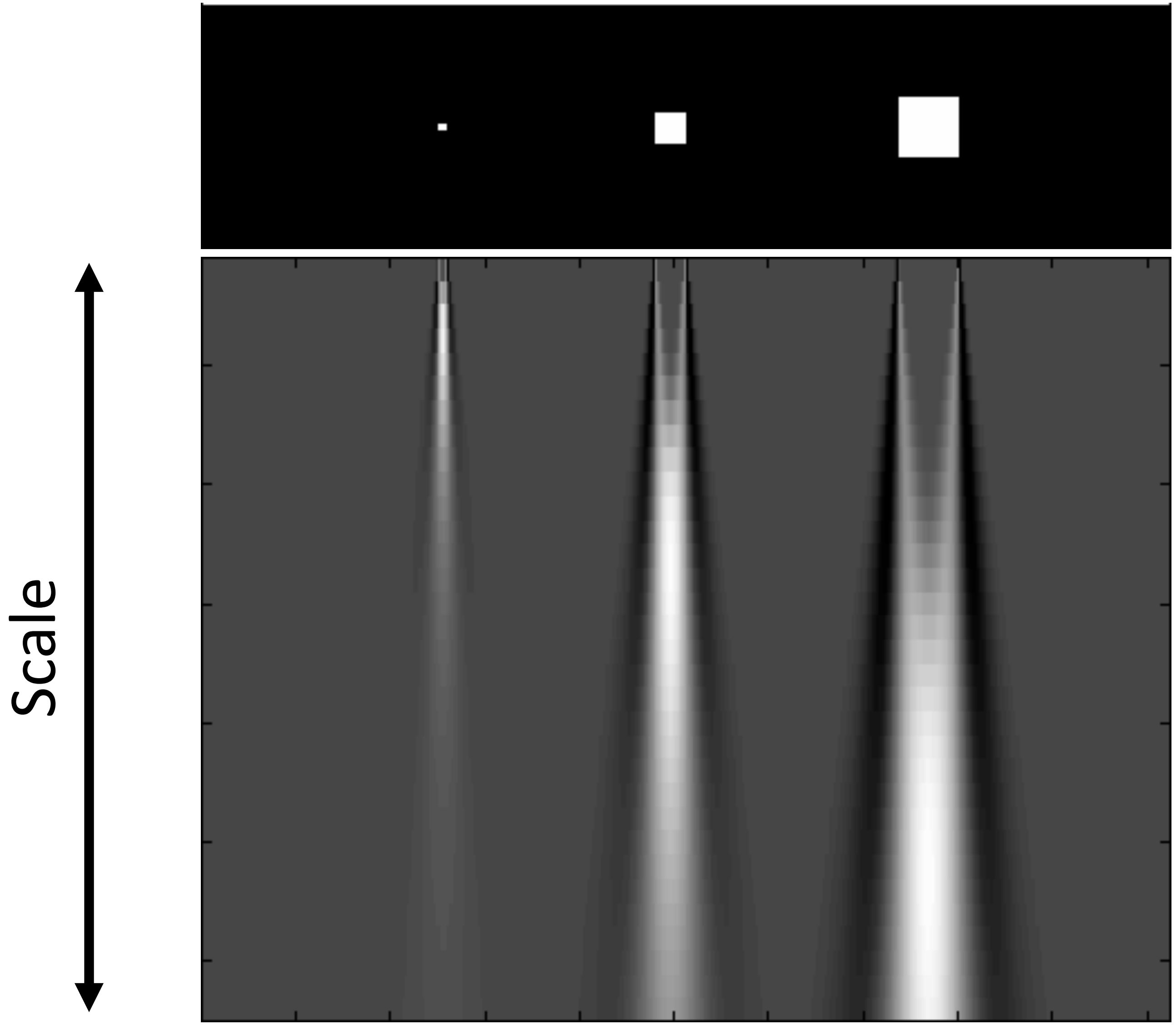
LoG $\sigma = 2$



LoG $\sigma = 10$

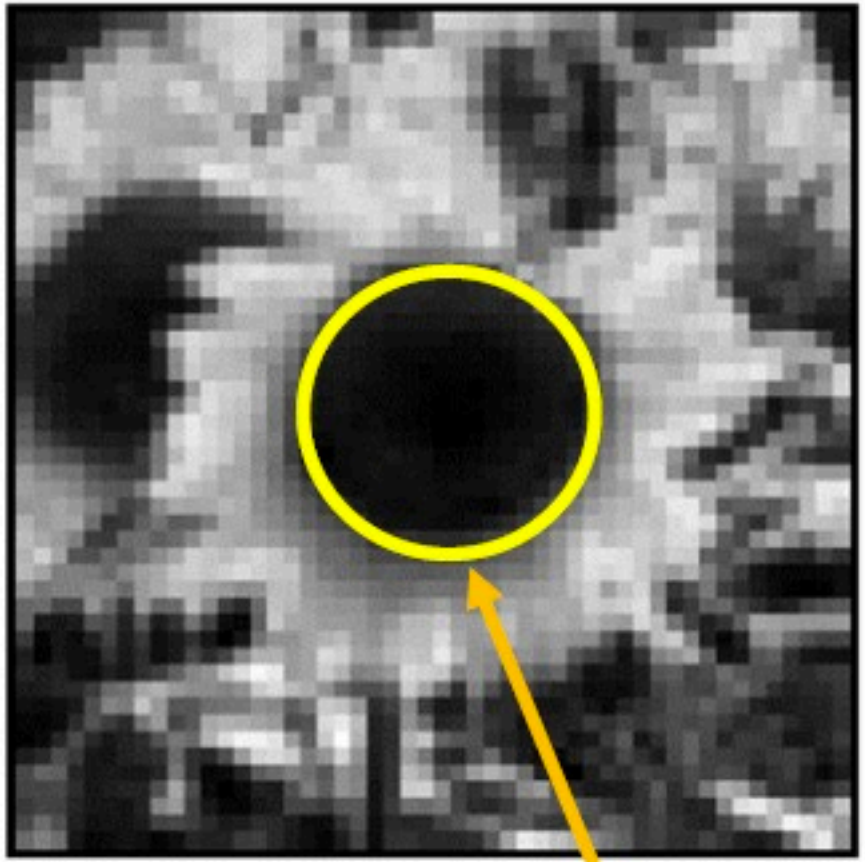
Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

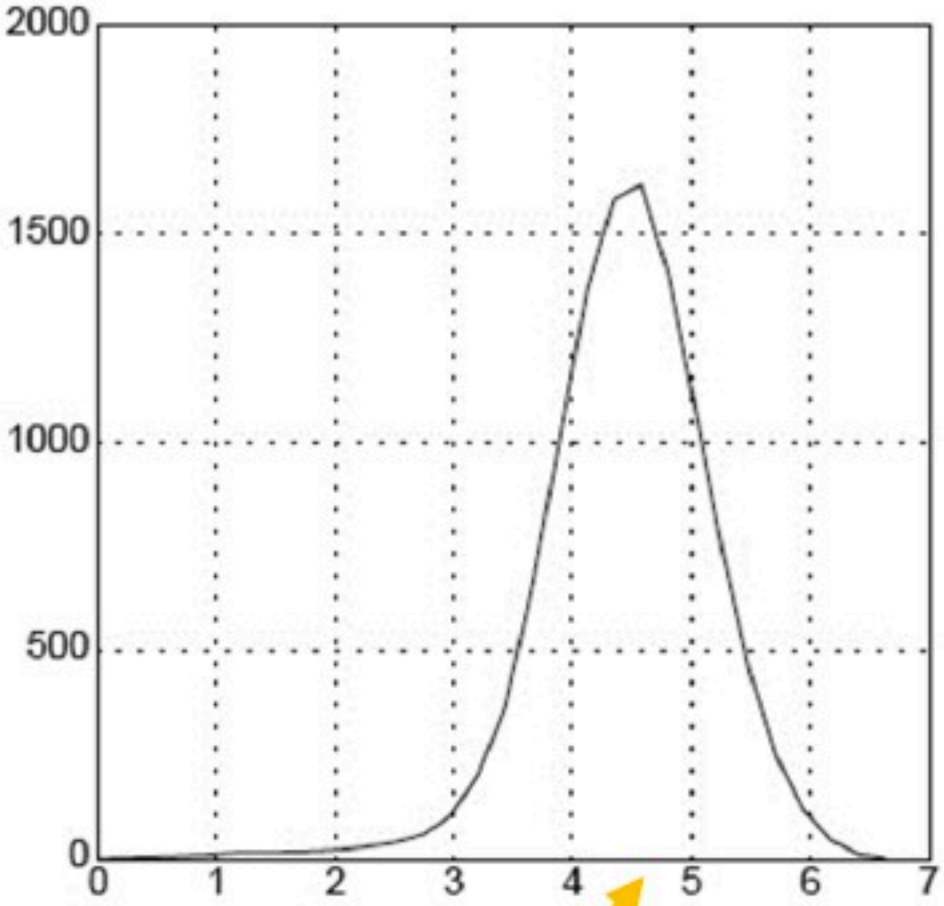
Local Scale Space Maxima



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Local Scale Space Maxima



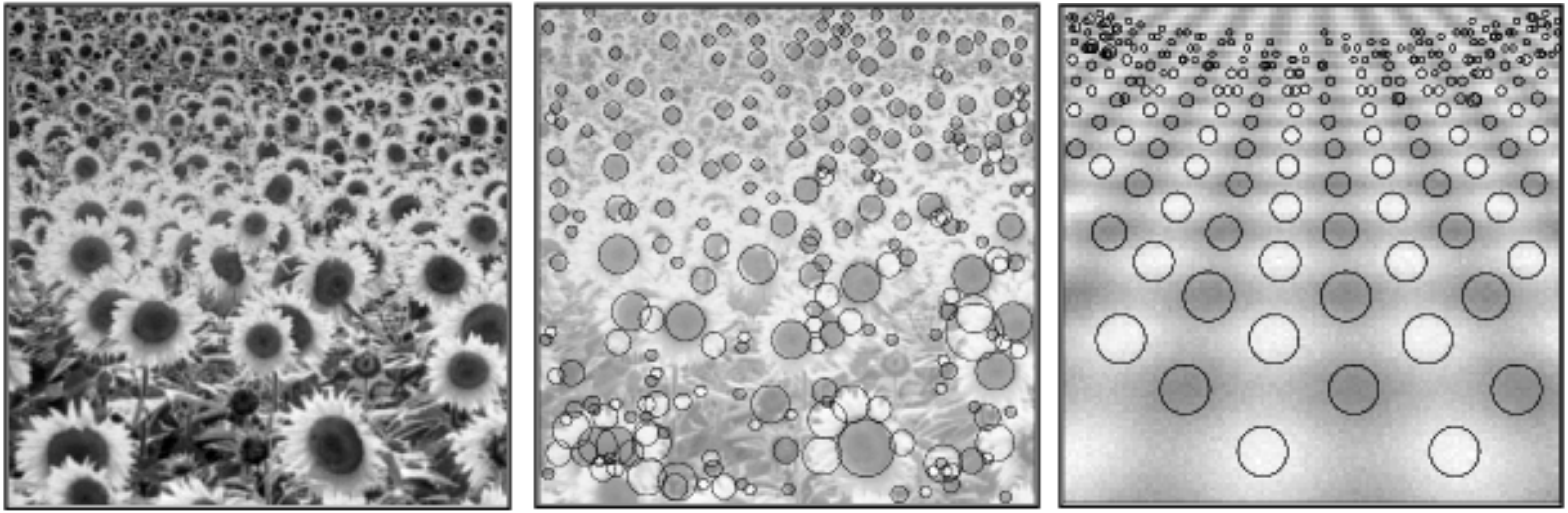


Characteristic scale

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Blob Detection using LoG

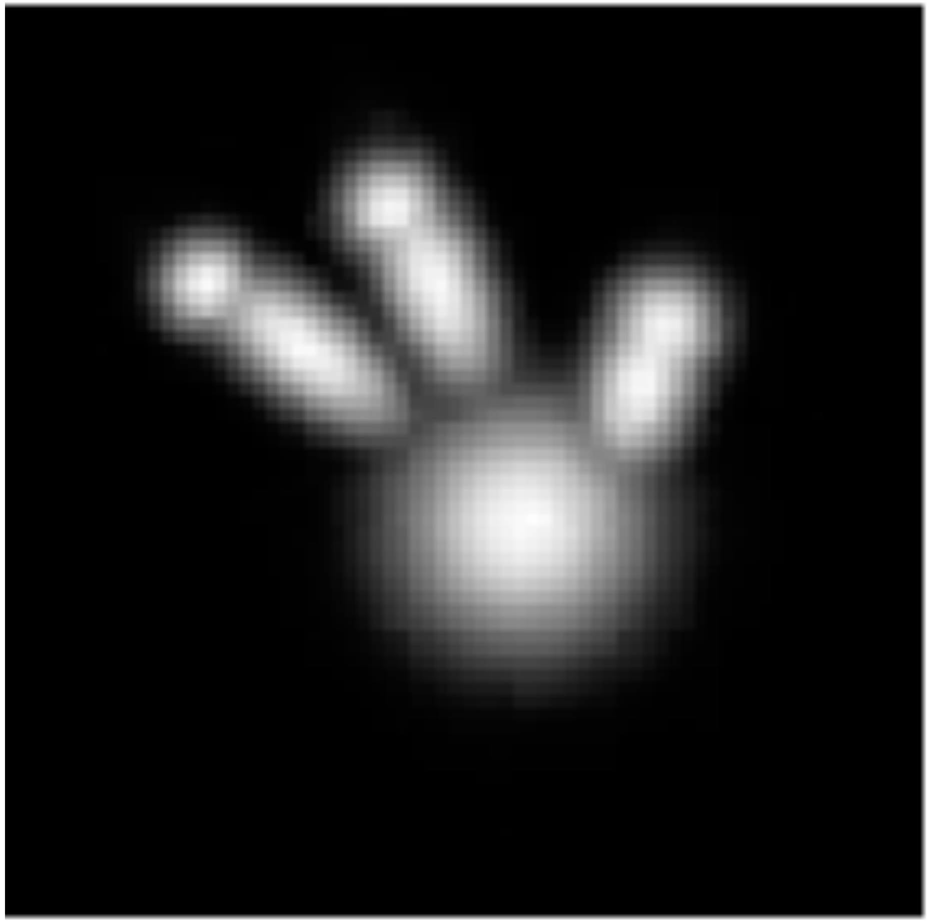
- Blobs are detected as local extrema in space and scale, within the LoG scale-space volume.



Lindeberg. IJCV 1998.
Feature detection with automatic scale selection

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Gesture Recognition via Blob Detection

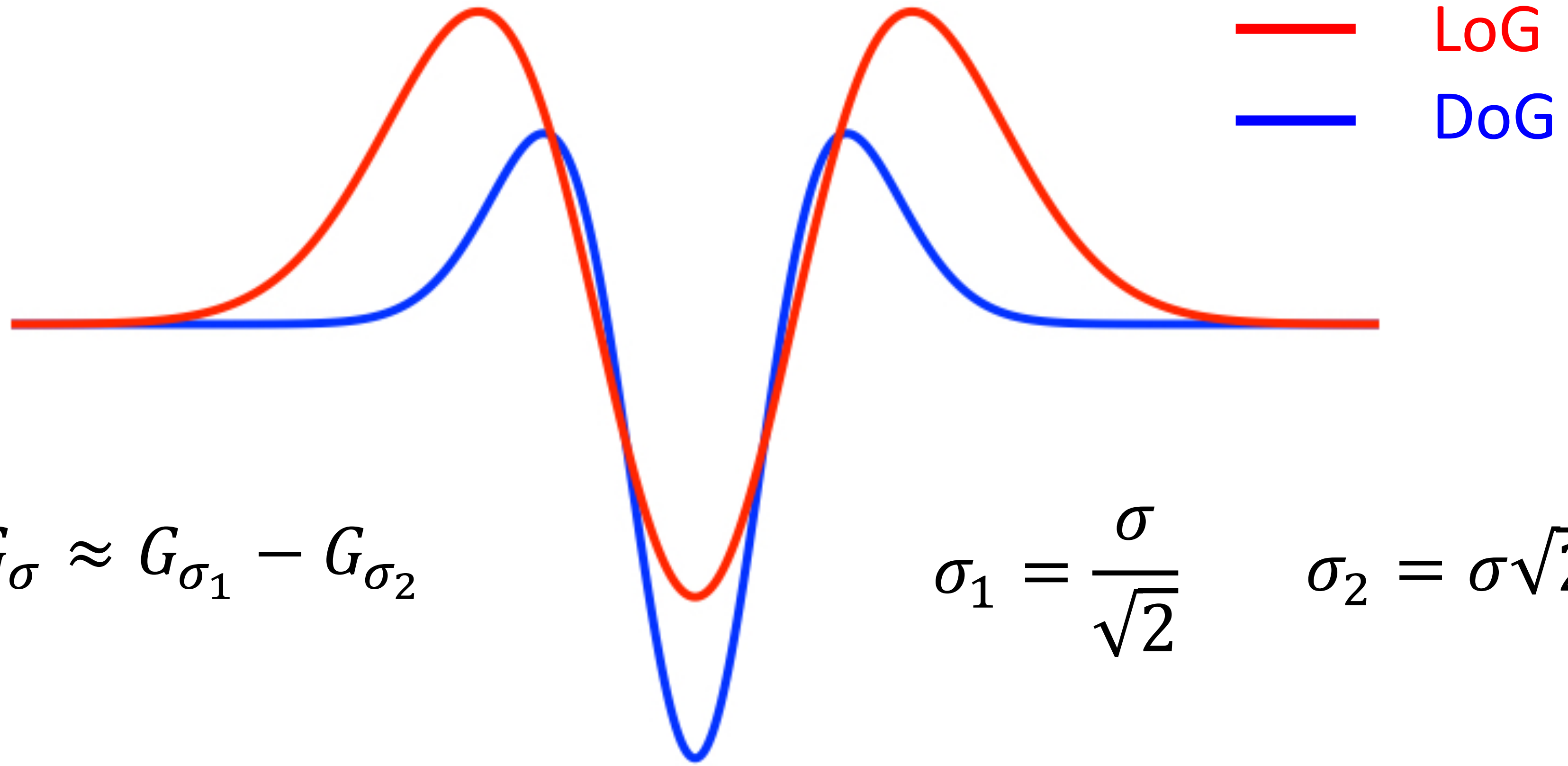


Lars Bretzner, Ivan Laptev, and Tony Lindeberg. Automatic Face and Gesture Recognition, 2002
Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

LoG → DoG

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



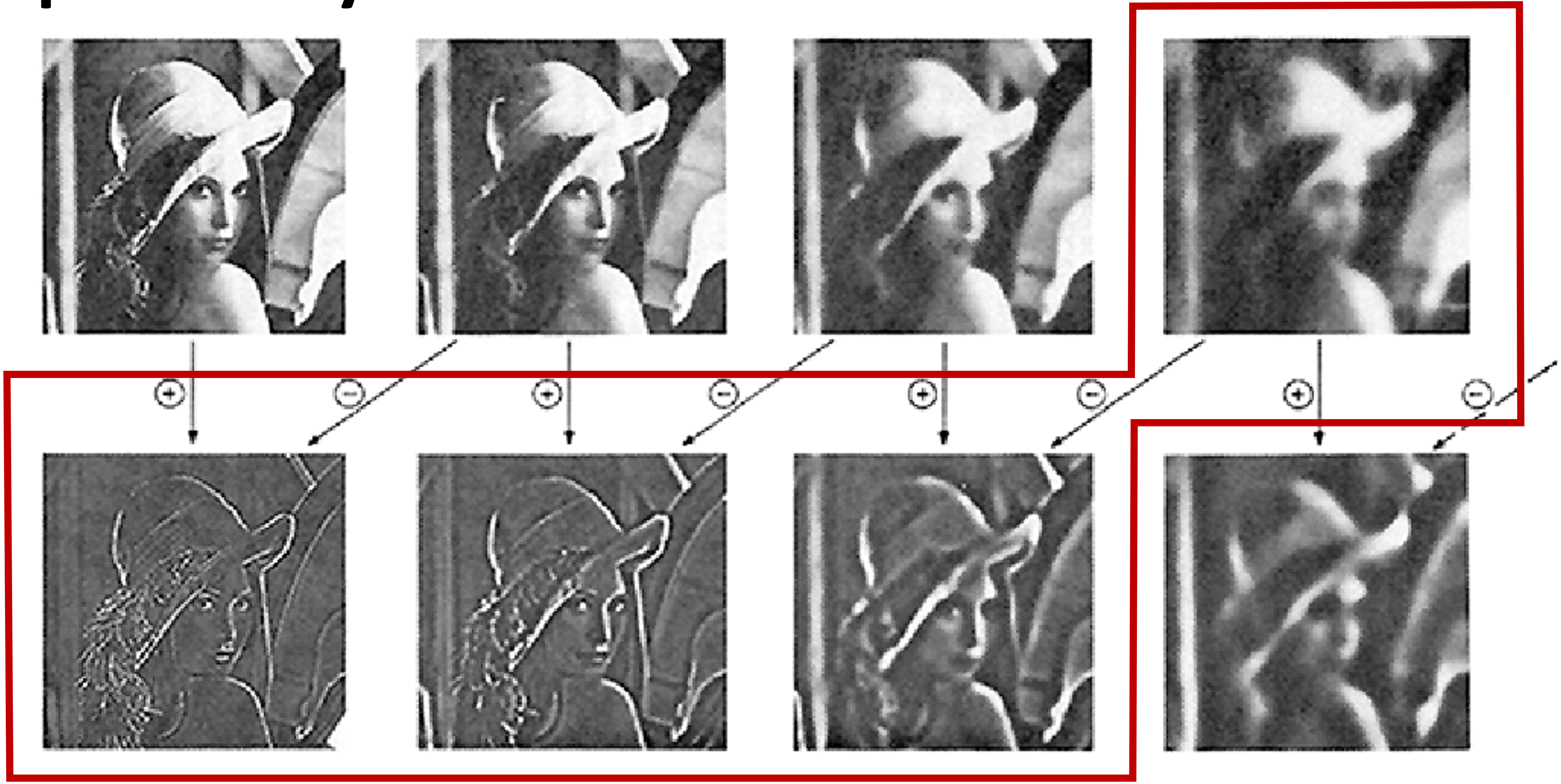
$$\nabla^2 G_\sigma \approx G_{\sigma_1} - G_{\sigma_2}$$

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

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Laplacian Pyramids

Store only these



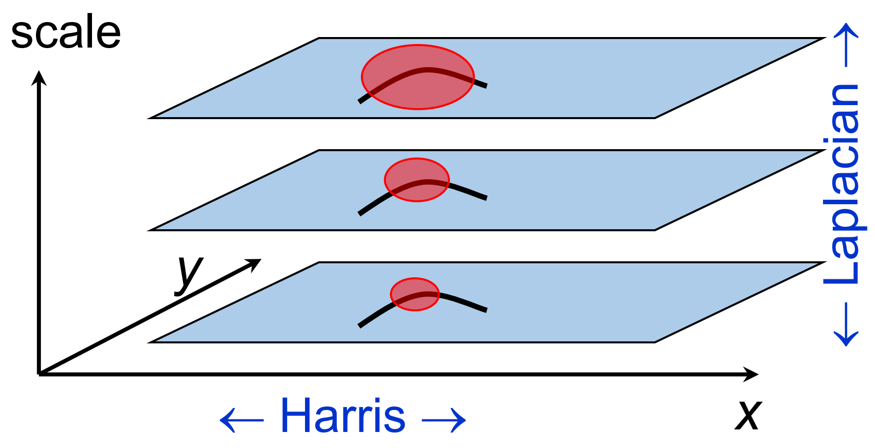
Scale Invariant Detectors

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Scale Invariant Detectors

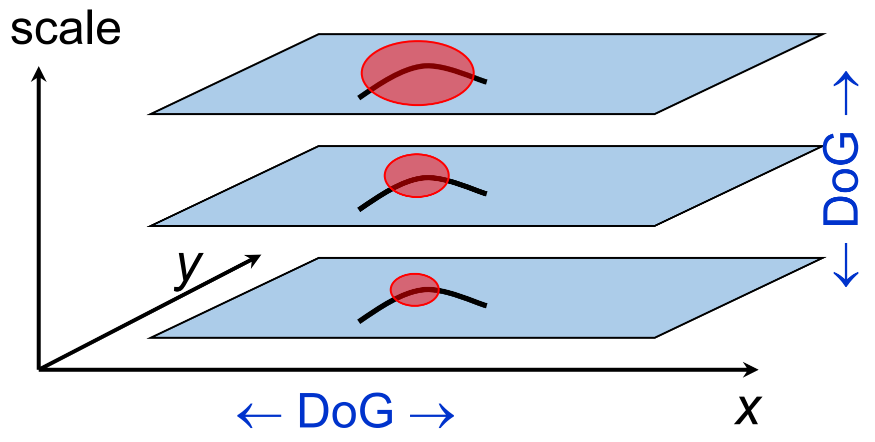
Harris-Laplacian

- Find local maximum of:
  - Harris corner detector in space (image coordinates)
  - Select points which are also maxima of Laplacian in scale



SIFT

- Find local maximum of:
  - Difference of Gaussians in space and scale

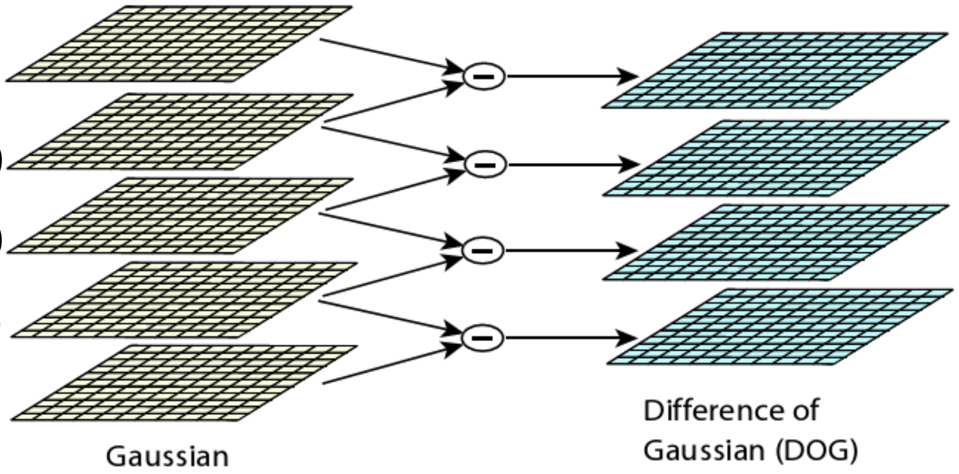


Difference of Gaussians

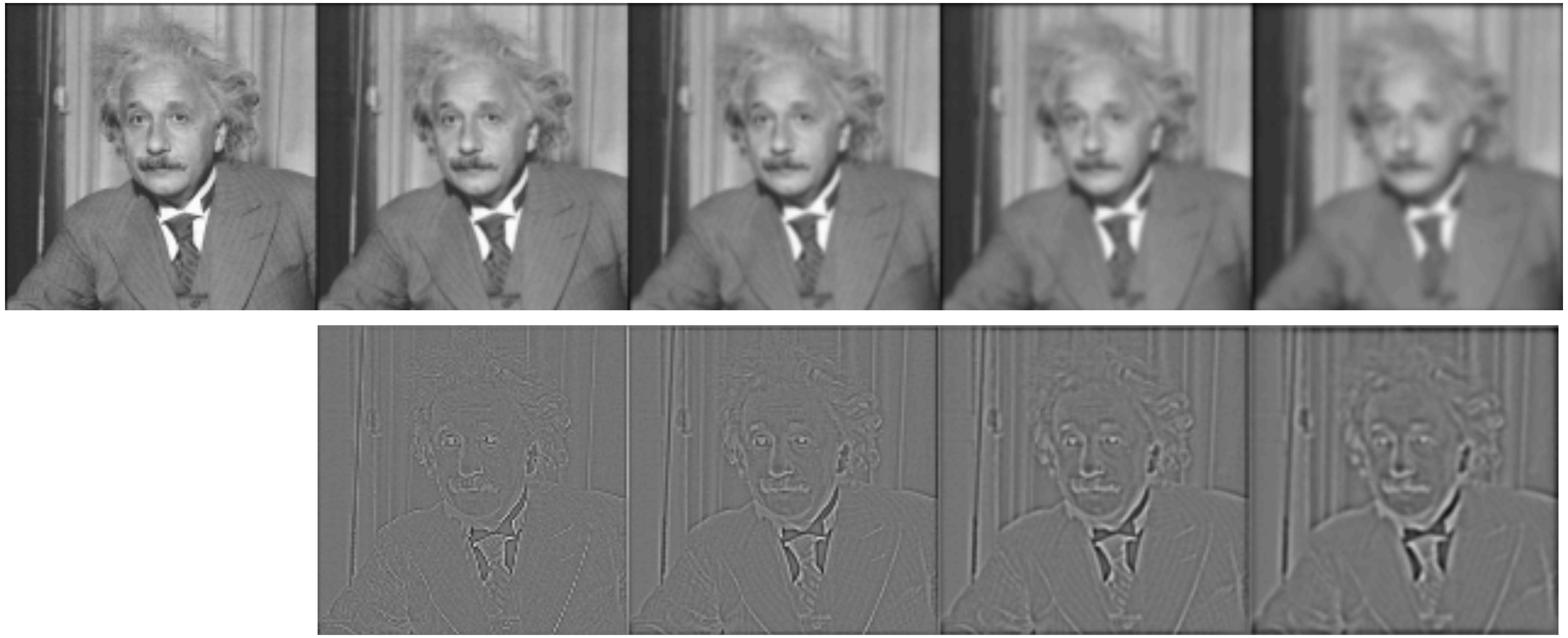
$I * G(k^2\sigma)$

$I * G(k\sigma)$

$I * G(\sigma)$



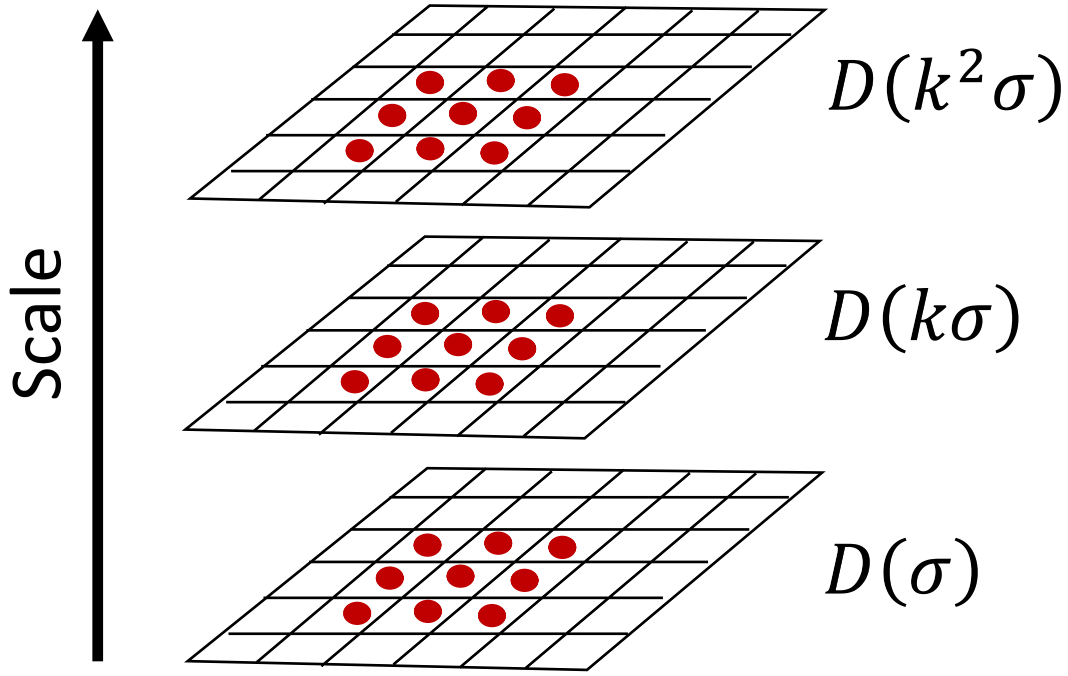
$$D(\sigma) \equiv I * (G(k\sigma) - G(\sigma))$$



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Scale Space Extrema

- Choose all extrema within 3×3×3 neighborhood.

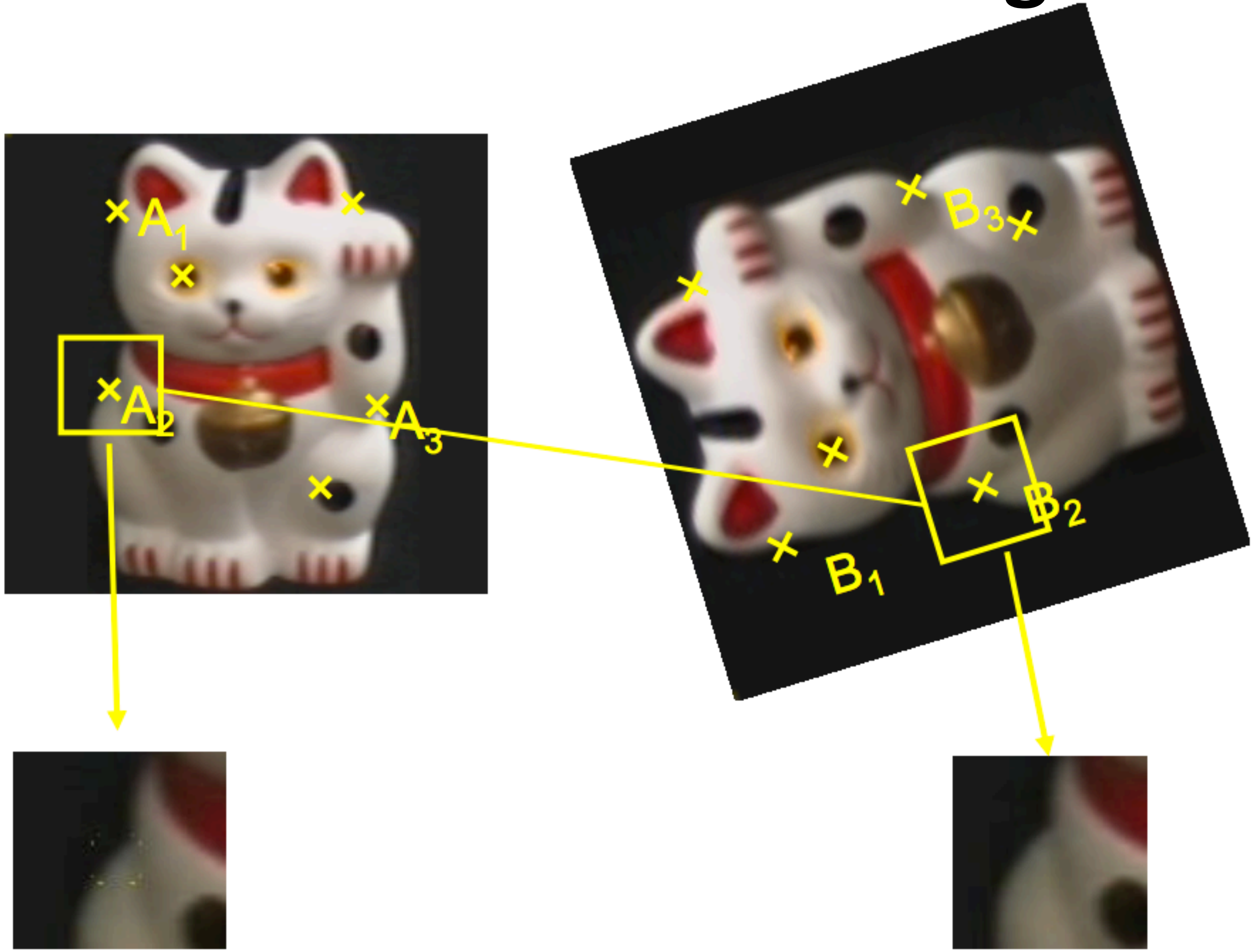


X is selected if it's response is larger or smaller than all 26 neighbors

Rotation Invariant Matching (Descriptors)

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

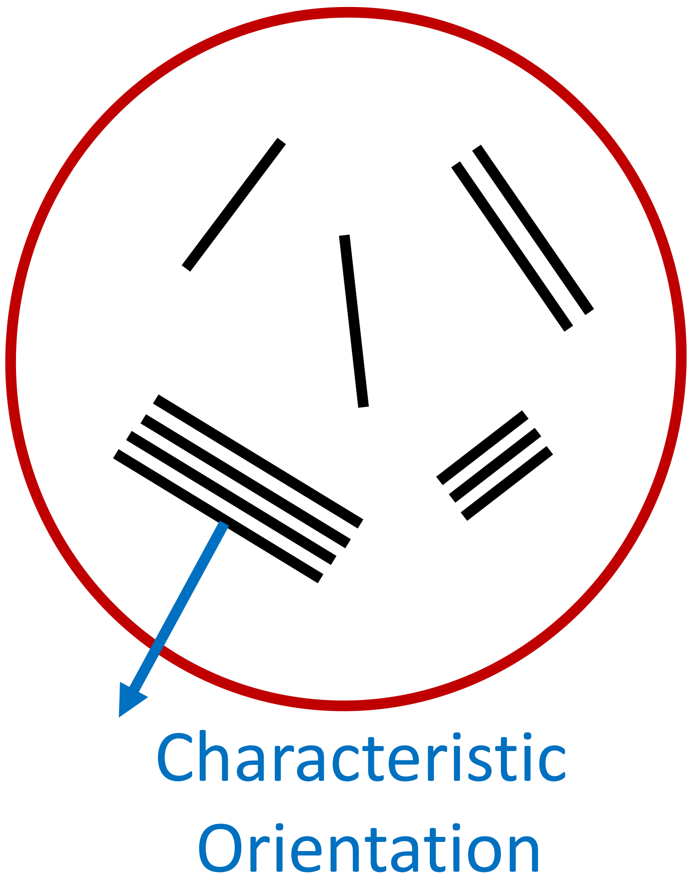
What is Rotation Invariant Matching?



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Rotation Invariance: Key Idea

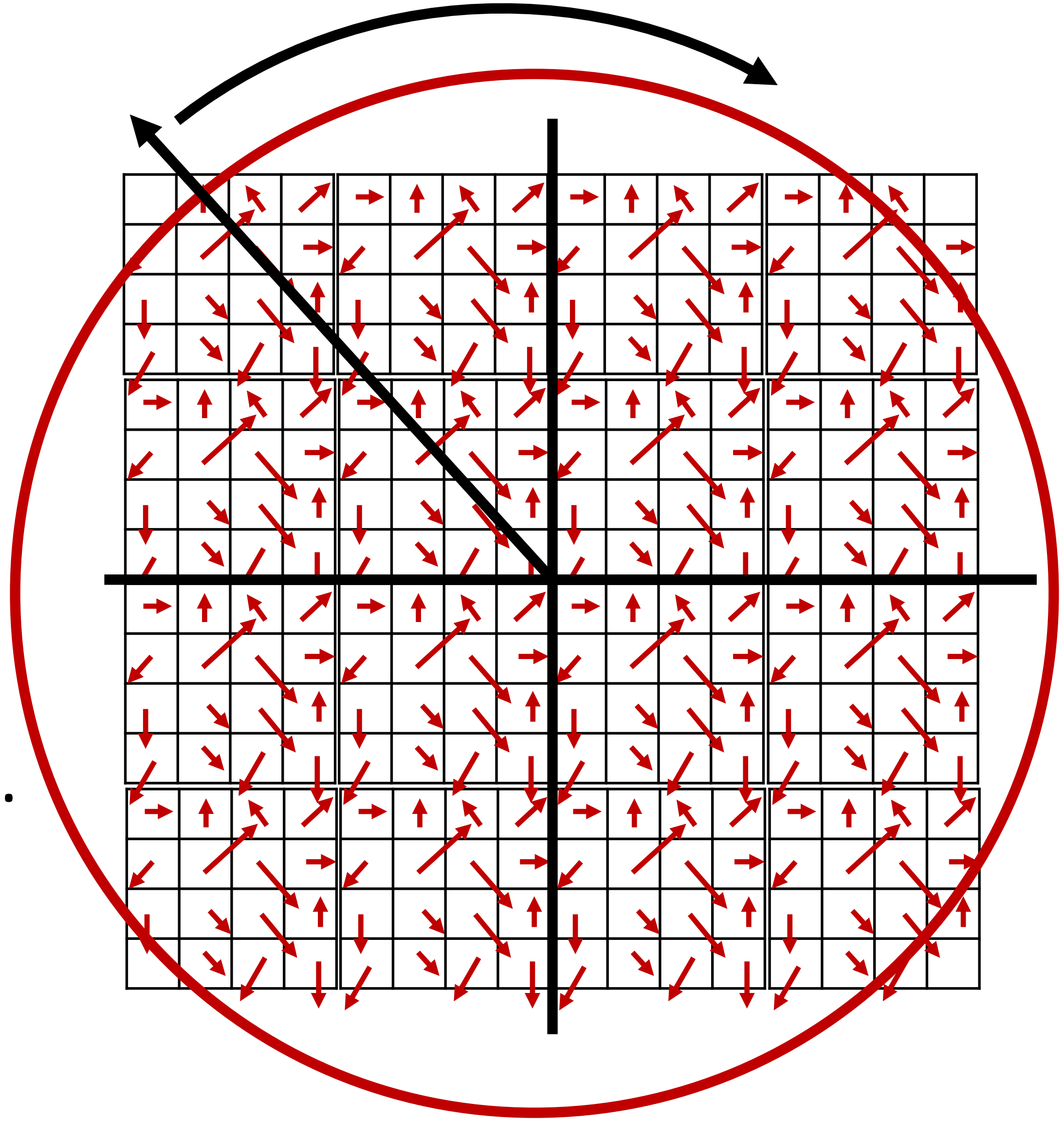
- We are given a keypoint and its scale from DoG
- Select a characteristic orientation for the keypoint (based on the most prominent gradient there)
- Describe all features **relative** to this orientation
- Causes descriptor to be rotation invariant!
  - If the keypoint is rotated in another image, the descriptor will be the same, since it is **relative** to the characteristic orientation



SIFT Descriptor

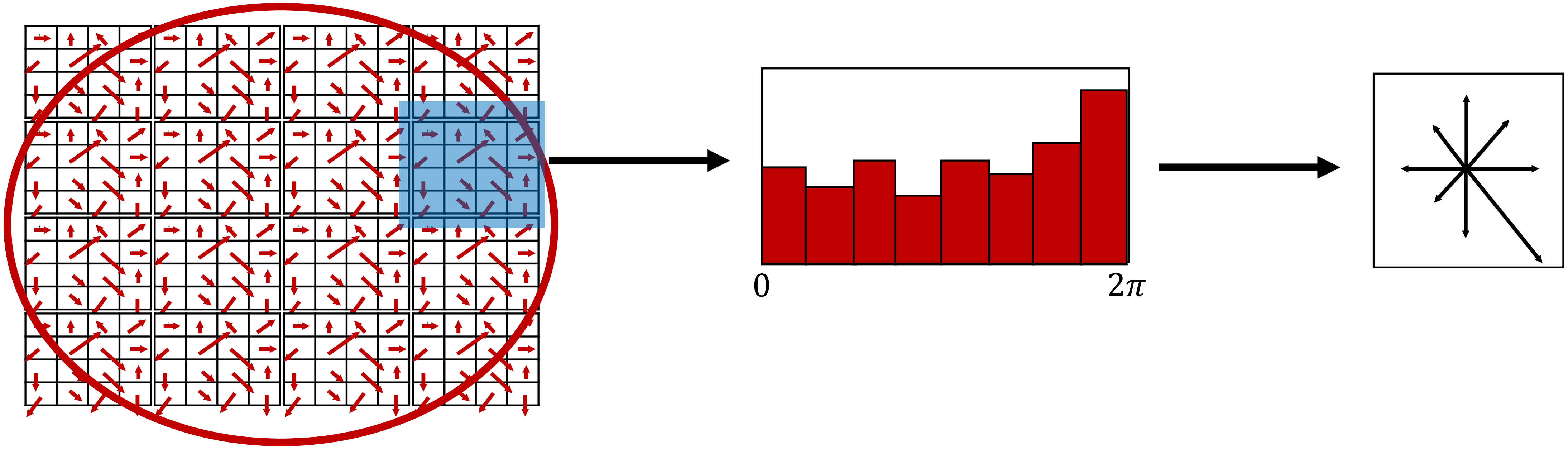
- Use the blurred image associated with the keypoint's scale
- Take image gradients over the keypoint neighborhood.
- To become rotation invariant, rotate the gradient directions by keypoint orientation.
- We could have rotated the whole image as well (slower)

Rotate Gradient Directions



Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

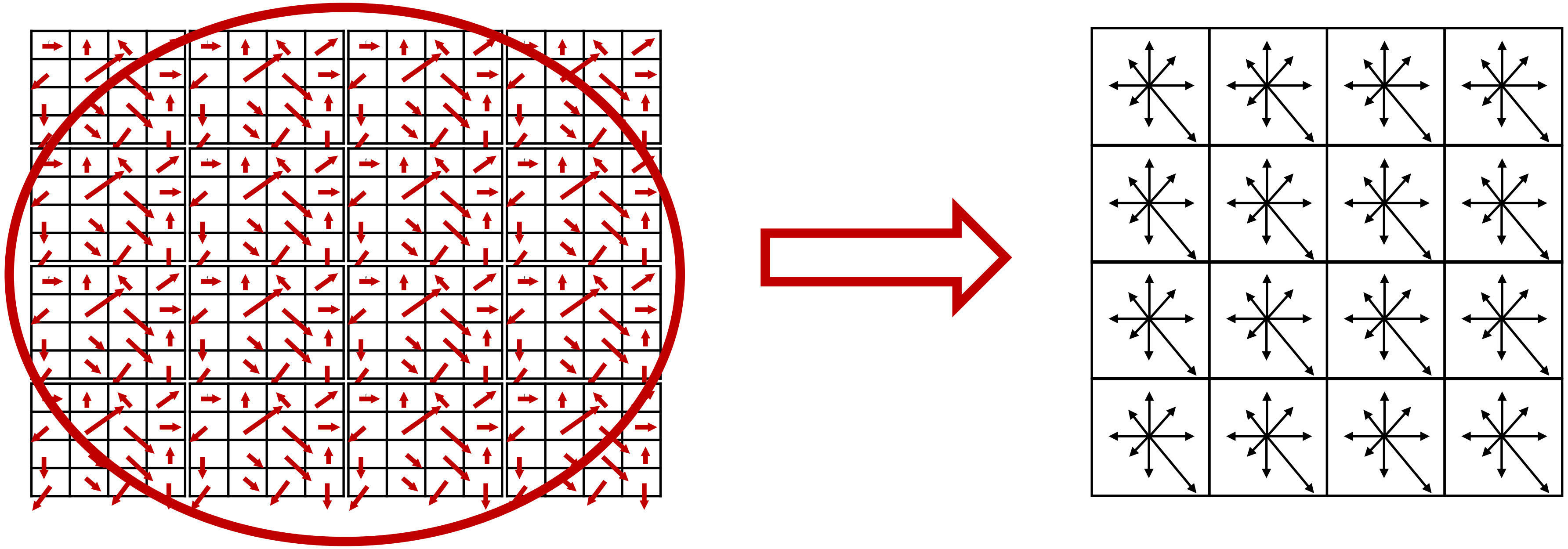
SIFT Descriptor



- Using precise gradient locations is fragile. Instead, create an orientation histogram of a 4x4 cells.
- A gradient's contribution is divided among the nearby histograms based on distance. Halfway between two histogram locations = Half contribution to both. Also, scales down gradient contributions far from the center.

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

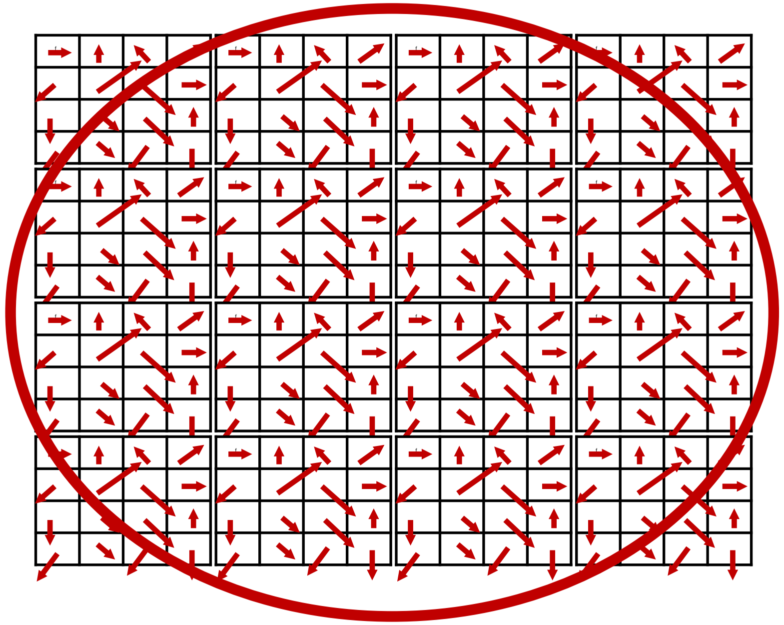
SIFT Descriptor

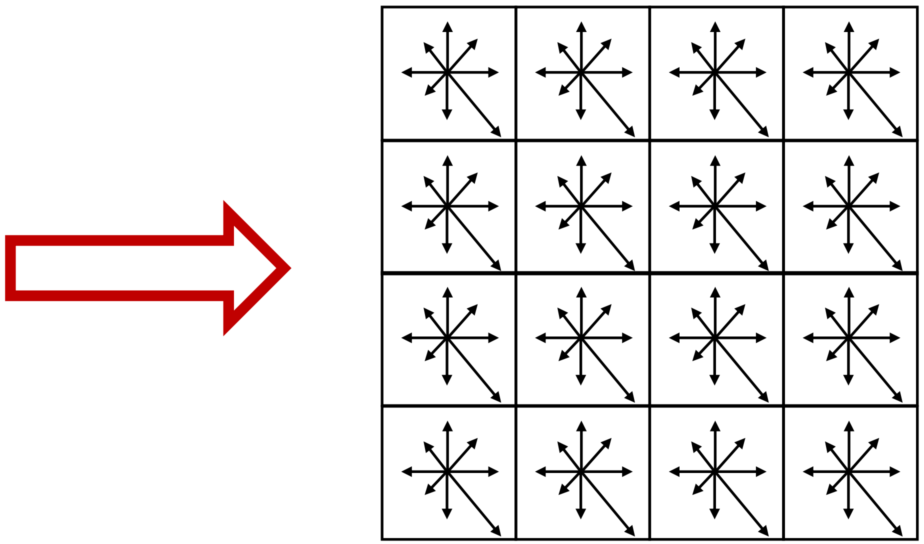


- SIFT uses 8 orientation bins per histogram, and a 4x4 histogram array, yielding 8 x 4x4 = 128 numbers, for a 16x16 window.
- A SIFT descriptor is a 128 dimensional vector, invariant to rotation (we rotated the gradients) and scale (worked with the scaled image from DoG).
- Use Euclidean distance between the two descriptor vectors to match.

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

SIFT Descriptor: Illumination Robustness





- SIFT descriptor is made of gradients. Invariant to small changes in brightness
- A higher-contrast in an image increases the magnitude of gradients linearly. To correct for contrast changes, normalize the vector (unit magnitude)
- Large gradients from 3D illumination effects effect robustness. Clamp all values in the vector to be ≤ 0.2. Then normalize the vector again.

Chetan Arora
Department of Computer Science and Engineering, IIT Delhi

Histogram of Oriented Gradients

- Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions.