Scale Invariant Feature Transform (SIFT)

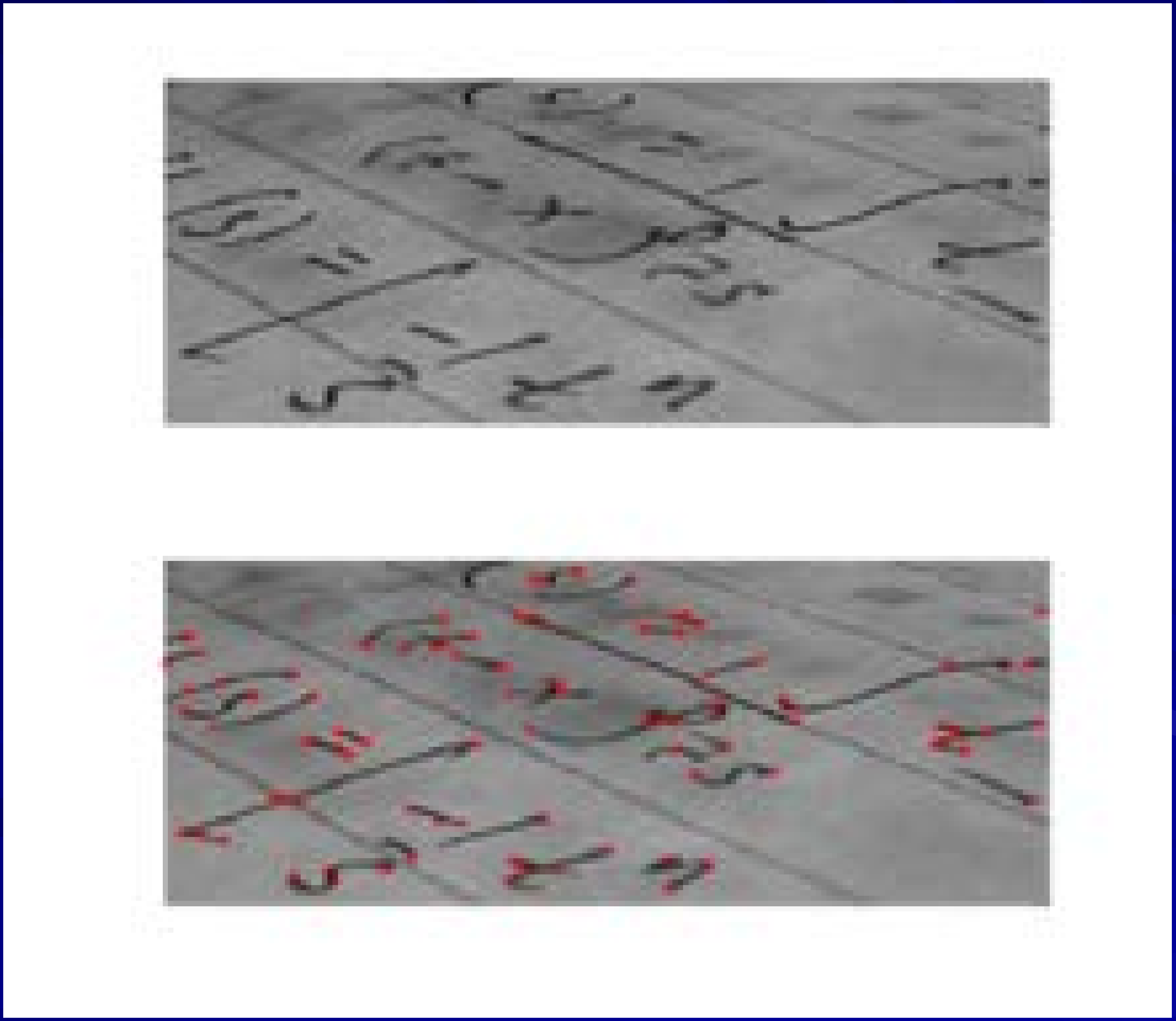
Original Source:
http://www.cs.ucf.edu/courses/cap4453/sift/

With additions and modifications

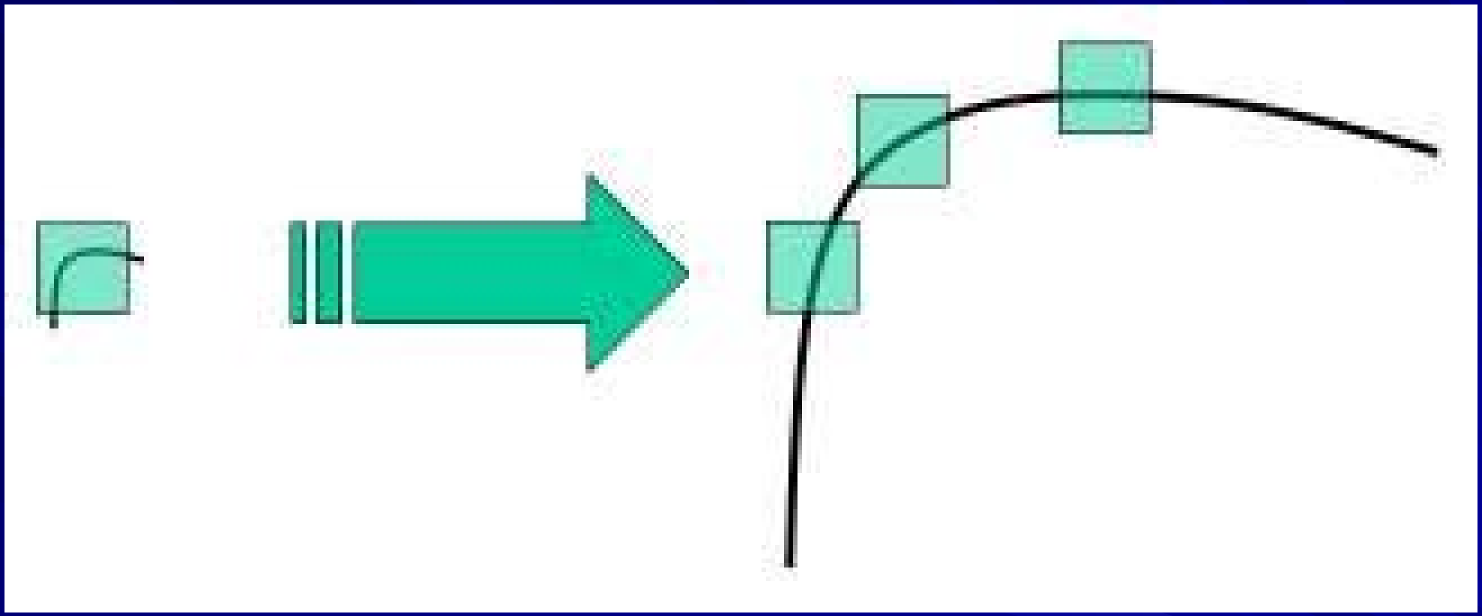
Edge/Contour, Blob and Corner Detectors

- Edge/contour detector – separator between two regions
- Blob detector – separator of a thin-line in a background region
- Corner detector – separator of a salient point in a line
  - Example: Harris corner detector

Output of A Corner Detector



Problem with Harris Corner Detector

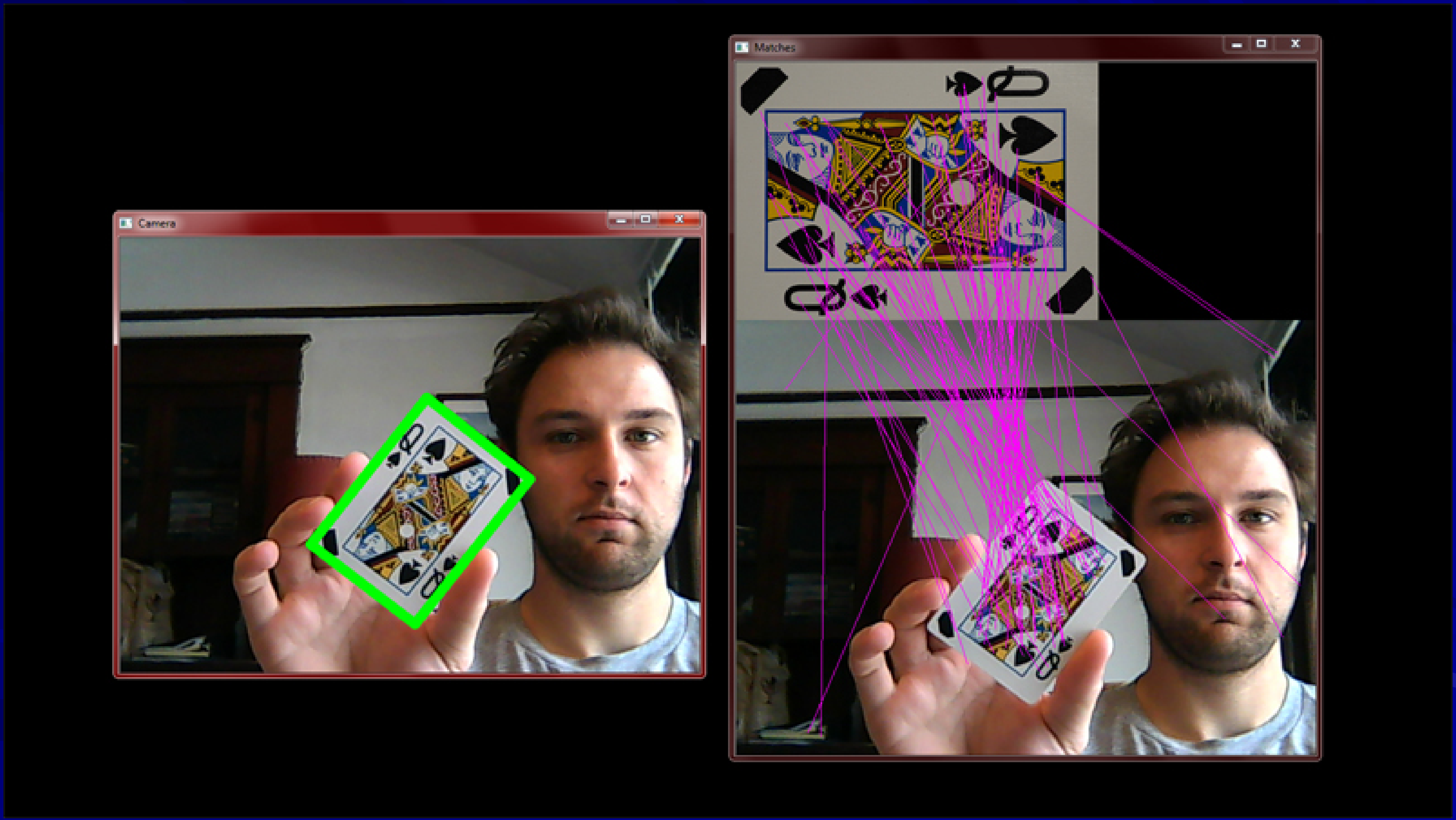


Harris corner detector is rotation-invariant but not scale-invariant

Scale-Invariant Feature Transform (SIFT)

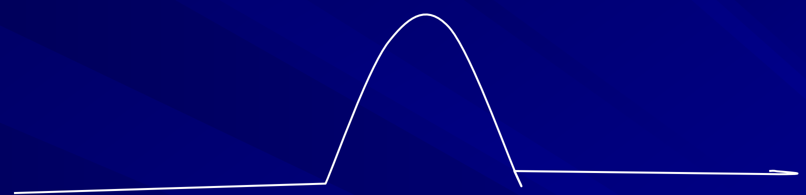
- Proposed by David Lowe in ICCV 1999
  – Two more papers, CVPR 2001 and IJCV 2004
- Generates image features, “key points”
  – Invariant to image scaling and rotation
  – Partially invariant to change in illumination and 3D camera viewpoint
  – Many can be extracted from typical images
  – Highly distinctive

Examples

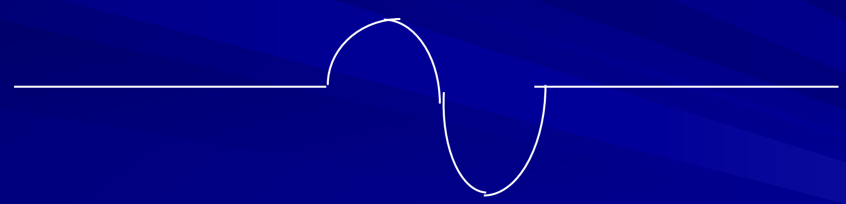


1-D Blob Manipulation

1-D Blob



1st order
derivative
(DoB)



2nd order
derivative
(LoB)

Observation1: Blob Center is located in the local extremum
Observation 2: Laplacian can be approximated by the difference of two Gaussian with different sigma

Algorithmic Overview

- Scale-space extrema detection
  - Uses difference-of-Gaussian function to approximate the Laplacian operator
- Keypoint localization
  - Sub-pixel location and scale fit to a model
- Orientation assignment
  - 1 or more for each keypoint
- Keypoint descriptor
  - Created from local image gradients

Scale Space

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

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

Scale Space

- Keypoints are detected using scale-space extrema in difference-of-Gaussian function *D*
- *D* definition:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y)$$

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

- Efficient to compute

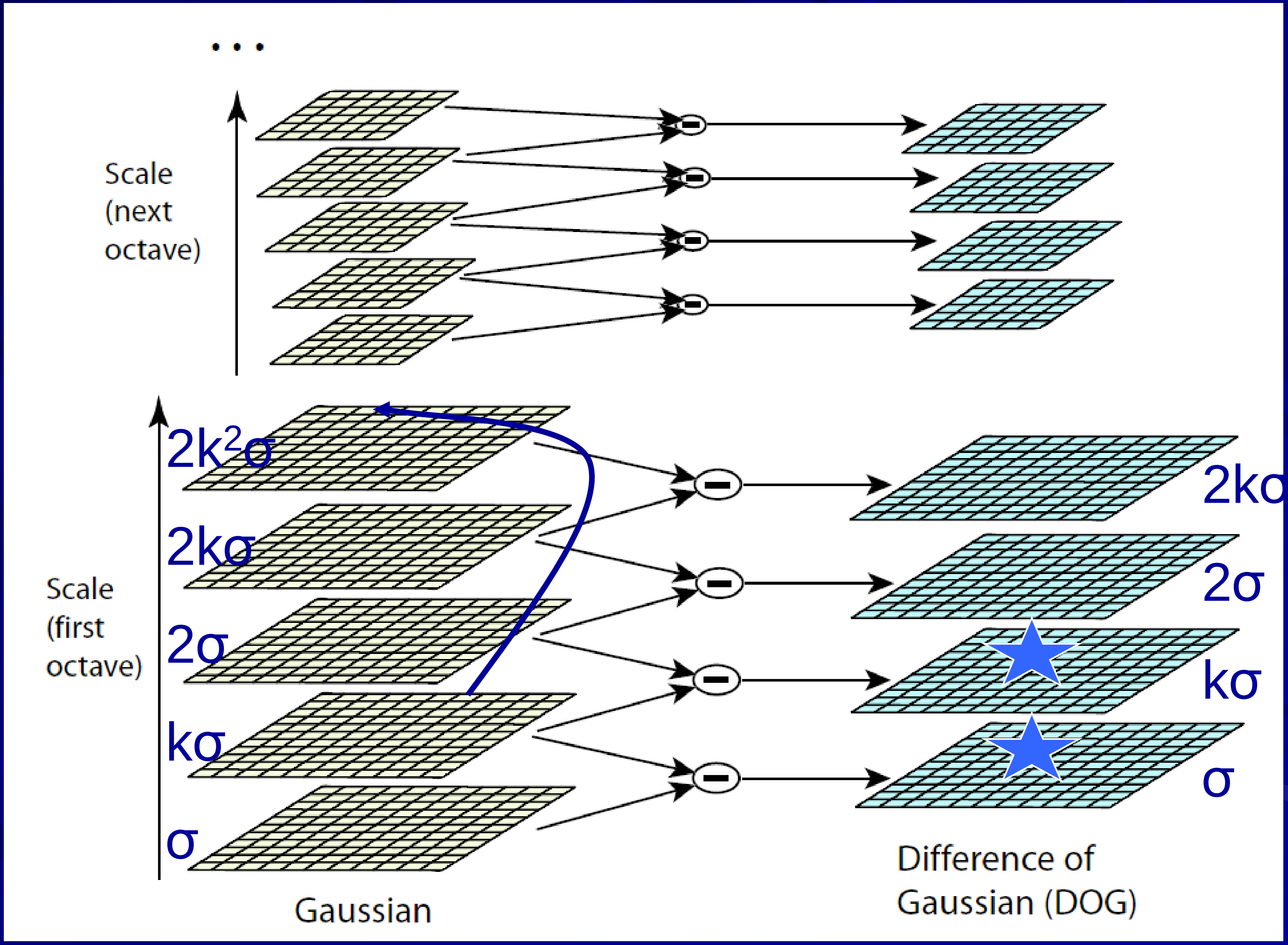
Relationship of *D* to $\sigma^2\nabla^2 G$

- Close approximation to scale-normalized Laplacian of Gaussian, $\sigma^2\nabla^2 G$
- Diffusion equation: $\frac{\partial G}{\partial \sigma} = \sigma\nabla^2 G$
- Approximate *∂G/∂σ*: $\frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma}$
  - giving, $\frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma} \approx \sigma\nabla^2 G$

$$G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2\nabla^2 G$$

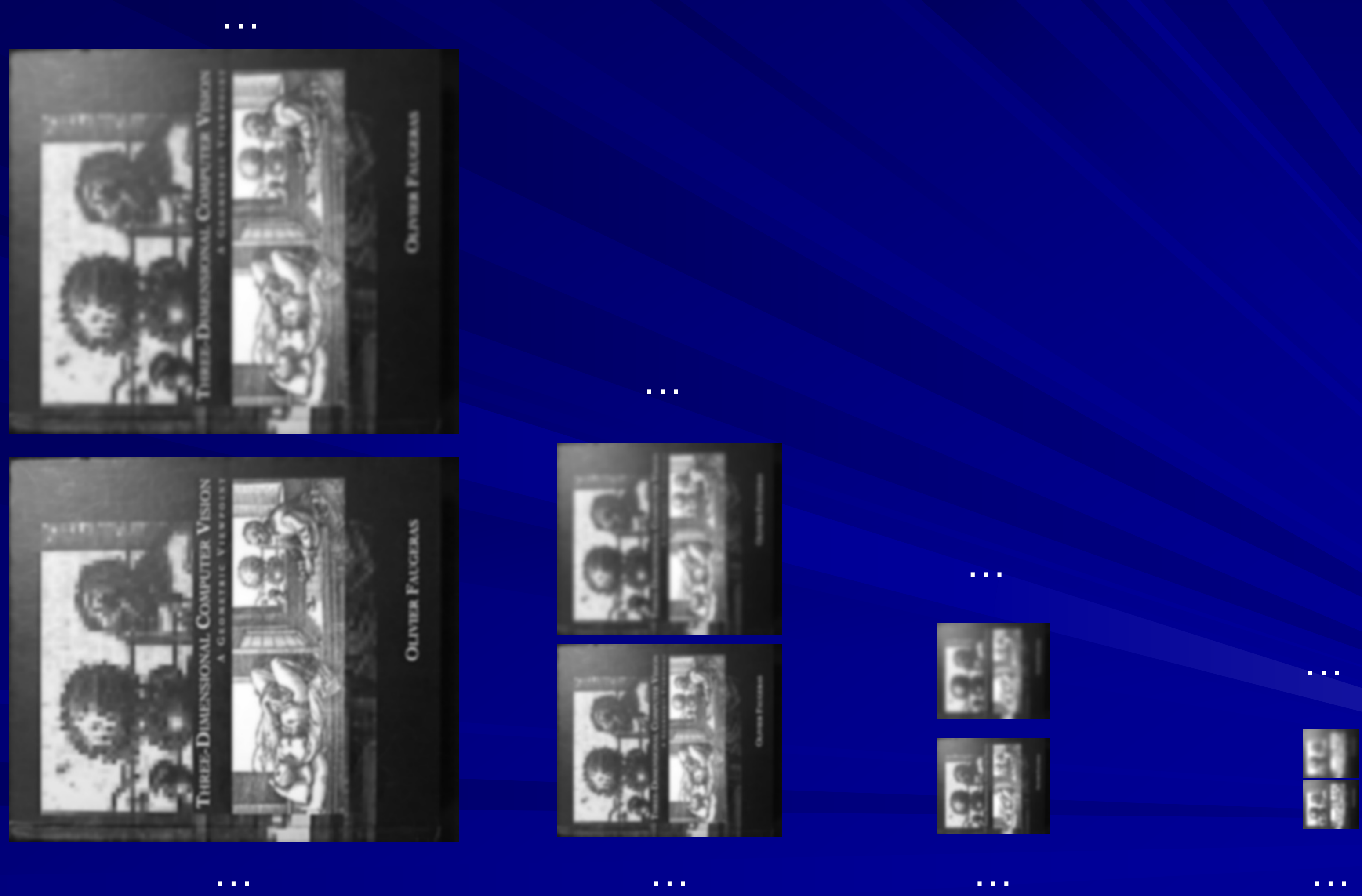
- When *D* has scales difference by a constant factor it already incorporates the σ^2 scale normalization required for scale-invariance

Scale Space Construction



Typically, k=sqrt{2}=1.414 and σ=1.6, no. of scale = 5 and no. of octaves = 4

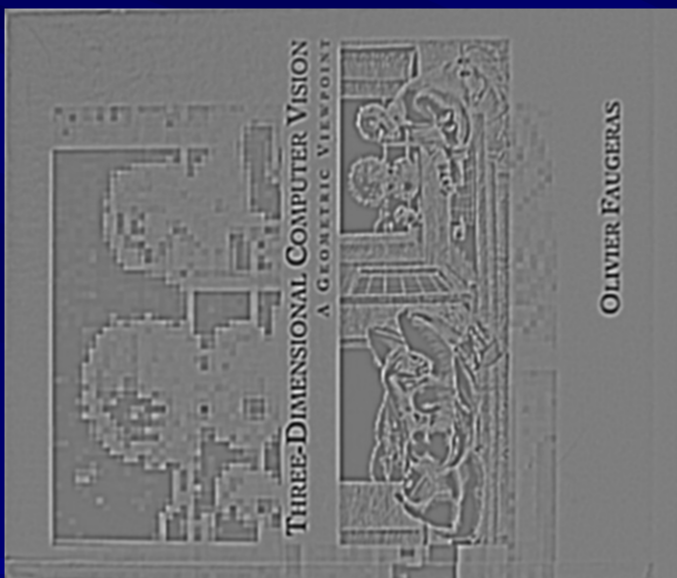
Gaussian Convolved Images



first octave second octave third octave fourth octave

Difference-of-Gaussian Images

…



…

first octave

…

…

second octave

…

…

third octave

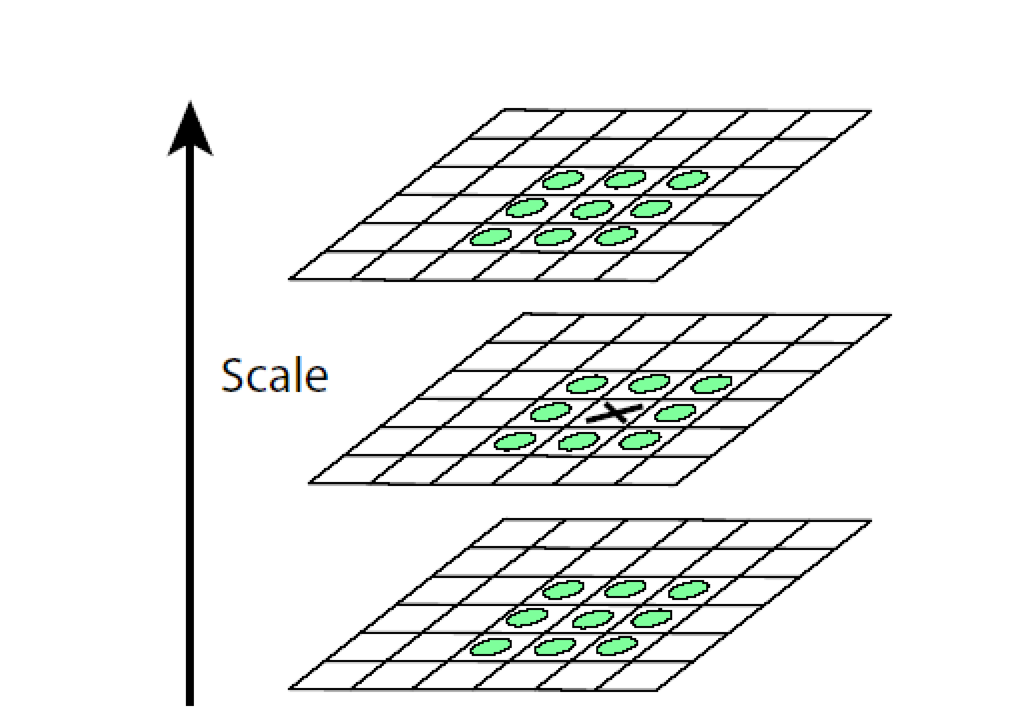
…

…

fourth octave

Key Point Selection

- A key point is selected only if it is a local minimum or maximum in the 3D DoG scale space



DoG scale space

Extrema in this image

Key Point Localization (1)

- 3D quadratic function is fit to the local sample points
- Start with Taylor expansion with sample point as the origin $D(\mathrm{X}) = D + \frac{\partial D^T}{\partial \mathrm{X}}\mathrm{X} + \frac{1}{2}\mathrm{X}^T \frac{\partial^2 D}{\partial \mathrm{X}^2}\mathrm{X}$
  – where $\mathrm{X} = (x, y, \sigma)^T$
- Take the derivative with respect to *X*, and set it to 0, giving $0 = \frac{\partial D}{\partial X} + \frac{\partial^2 D}{\partial X^2}\hat{X}$
- $\hat{\mathrm{X}} = -\frac{\partial^2 D^{-1}}{\partial \mathrm{X}^2}\frac{\partial D}{\partial \mathrm{X}}$ is the location of the key point
- This is a 3x3 linear system

Key Point Localization (2)

$$\begin{bmatrix} \frac{\partial^2 D}{\partial \sigma^2} & \frac{\partial^2 D}{\partial \sigma y} & \frac{\partial^2 D}{\partial \sigma x} \\ \frac{\partial^2 D}{\partial \sigma y} & \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial yx} \\ \frac{\partial^2 D}{\partial \sigma x} & \frac{\partial^2 D}{\partial yx} & \frac{\partial^2 D}{\partial x^2} \end{bmatrix} \begin{bmatrix} \sigma \\ y \\ x \end{bmatrix} = - \begin{bmatrix} \frac{\partial D}{\partial \sigma} \\ \frac{\partial D}{\partial y} \\ \frac{\partial D}{\partial x} \end{bmatrix}$$

- Derivatives approximated by finite differences,
  - example:

$$\frac{\partial D}{\partial \sigma} = \frac{D_{k+1}^{i,j} - D_{k-1}^{i,j}}{2}$$

$$\frac{\partial^2 D}{\partial \sigma^2} = \frac{D_{k-1}^{i,j} - 2D_{k}^{i,j} + D_{k+1}^{i,j}}{1}$$

$$\frac{\partial^2 D}{\partial \sigma y} = \frac{(D_{k+1}^{i+1,j} - D_{k-1}^{i+1,j}) - (D_{k+1}^{i-1,j} - D_{k-1}^{i-1,j})}{4}$$

- If *X* is > 0.5 in any dimension, process repeated

Post-Processing via Filtering

- Contrast (use prev. equation): $D(\hat{X}) = D + \frac{1}{2}\frac{\partial D^T}{\partial X}\hat{X}$
  - If | *D(X)* | < 0.03, throw it out
- Edge point removal:
  - Use ratio of principal curvatures to throw out poorly defined peaks
  - Curvatures come from Hessian: $H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$
  - Ratio of $Trace(H)^2$ and *Determinant(H)*

$$Tr(H) = D_{xx} + D_{yy}$$
$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2$$

If ratio > $(r+1)^2/(r)$, throw it out (SIFT uses r=10)

Keypoint Orientation Assignment (1)

- An orientation is assigned to each keypoint to achieve invariance to image rotation.
- A neigborhood is taken around the keypoint location depending on the scale, and the gradient magnitude and direction is calculated in that region.
- An orientation histogram with 36 bins covering 360 degrees is created. It is weighted by gradient magnitude and Gaussian-weighted circular window with its sigma equal to 1.5 times the scale of the keypoint.

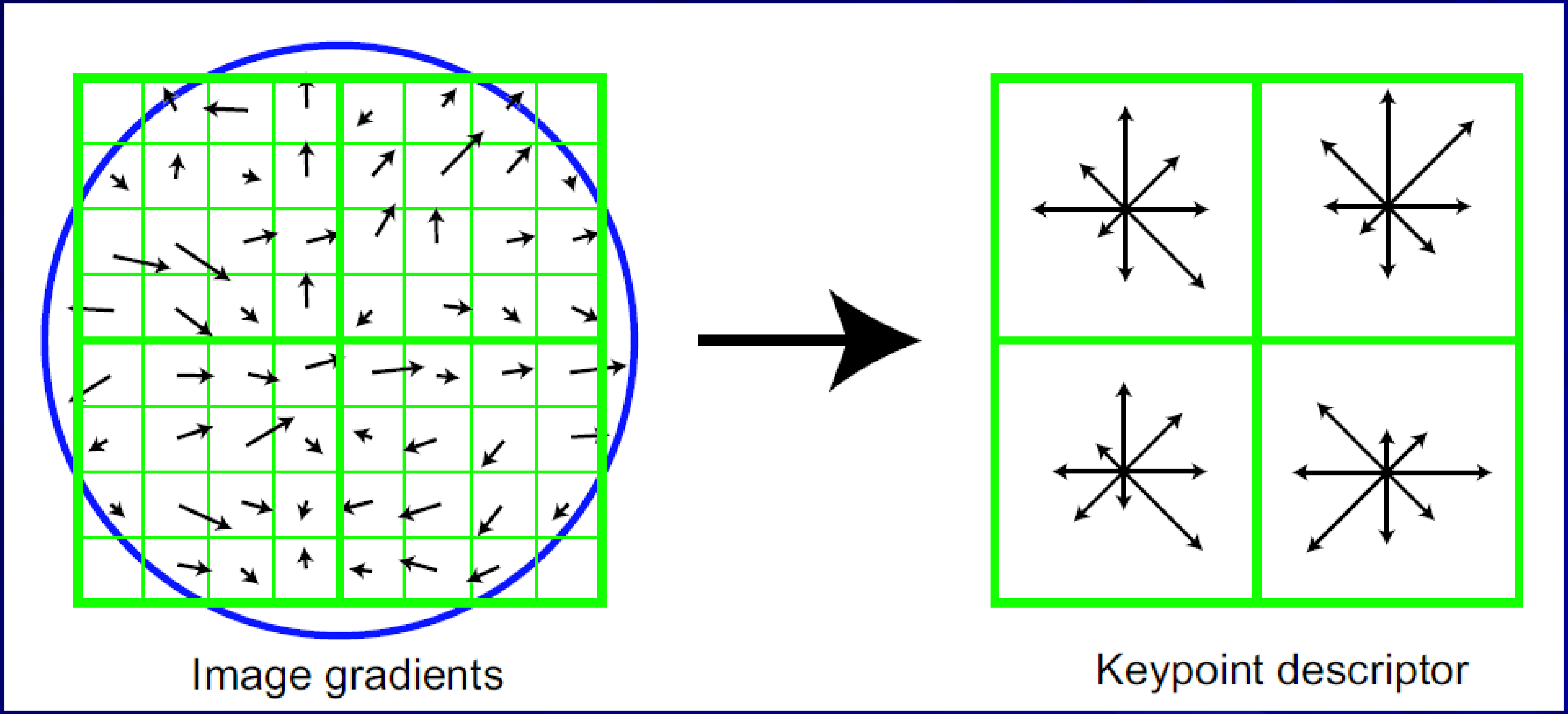
Key Point Orientation Assignment (2)

- The highest peak in the histogram is taken and any peak above 80% of it is also considered to calculate the orientation.
  - It creates keypoints with the same location and scale, but different directions
- It contributes to stability of matching

SIFT Descriptor (1)

- A 16x16 neighborhood around the keypoint
  - Divided into 16 sub-blocks of 4x4 size
- For each sub-block, 8 bin orientation histogram is created
  - A total of 128 (=8x16) bin values are available
  - It is represented as a vector to form keypoint descriptor
  - Besides, several measures are taken to achieve robustness against illumination changes, rotation, etc

SIFT Descriptor (2)



- Descriptor has 3 dimensions *(x,y,θ)*
- Position and orientation of each gradient sample rotated relative to keypoint orientation

SIFT Descriptor (3)

- Best results achieved with 4x4x8 = 128 descriptor size
- Normalize to unit length
  - Reduces effect of illumination change
- Cap each element to 0.2, normalize again
  - Reduces non-linear illumination changes
  - 0.2 determined experimentally

Object Detection

- Create a database of keypoints from training images
- Match keypoints to a database
  - Nearest neighbor search

PCA-SIFT

- Different descriptor (same keypoints)
- Apply PCA to the gradient patch
- Descriptor size is 20 (instead of 128)
- More robust, faster

Summary

- Scale space
- Difference-of-Gaussian
- Localization
- Filtering
- Orientation assignment
- Descriptor, 128 elements