Feature Detection

-Harshit Sikchi

Topics:

- Harris Corner Detector
- Mar-Hildreth Edge detector
- SIFT(David Lowe)

What are Interest Points:-

They are spatial locations, or points in the image that define what is **interesting**

or what **stand out** in the image.

How do we want our detector to be:-

- Well Localized Points
- No false interest point
- Robust with respect to noise

Usage

- Object Tracking
- Stereo Calibration
- Panorama , Image Stitching
- 3D Object Reconstruction
- Robot Navigation

Figure 4.2 Two pairs of images to be matched. What kinds of feature might one use to establish a set of *correspondences* between these images?

Structure from motion

https://www.youtube.com/watch?v=i7ierVkXYa8

Reliable Descriptor

- Must be invariant to geometric and photometric differences in the two views.
- Geometric means Translation, Rotation, Affine Transform, Scaling.

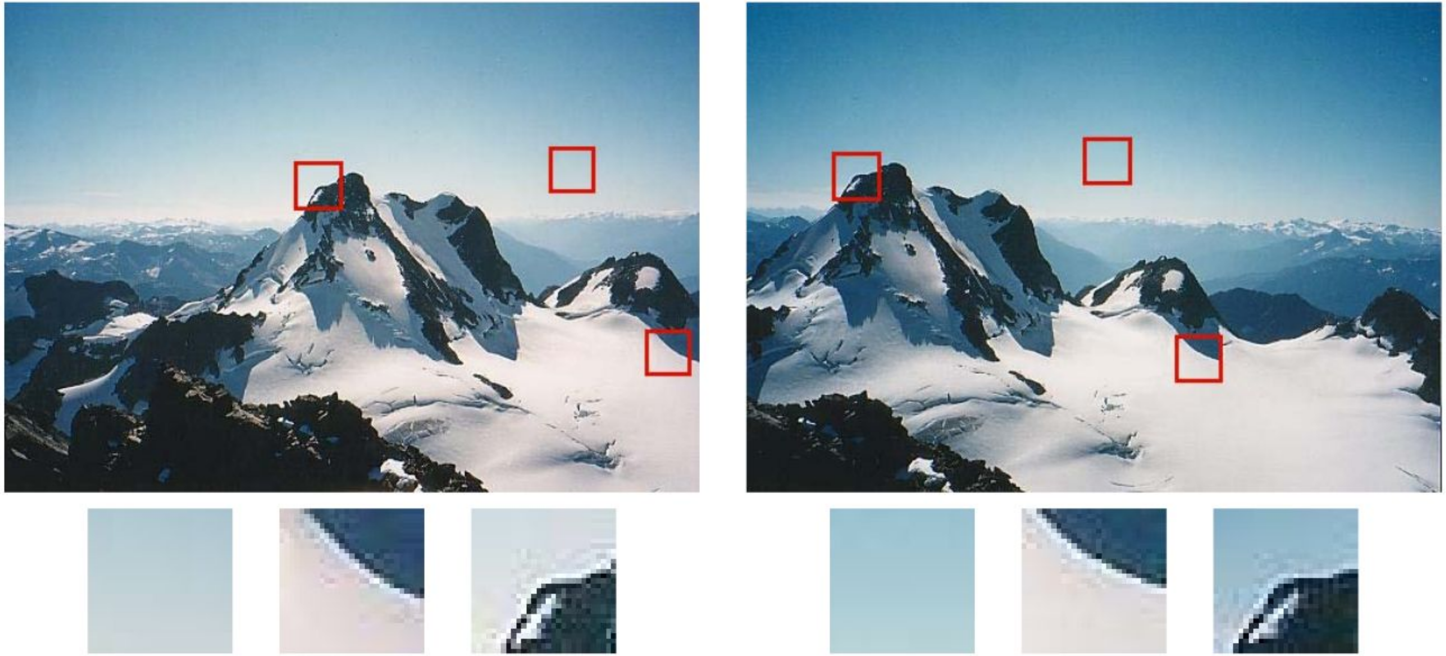


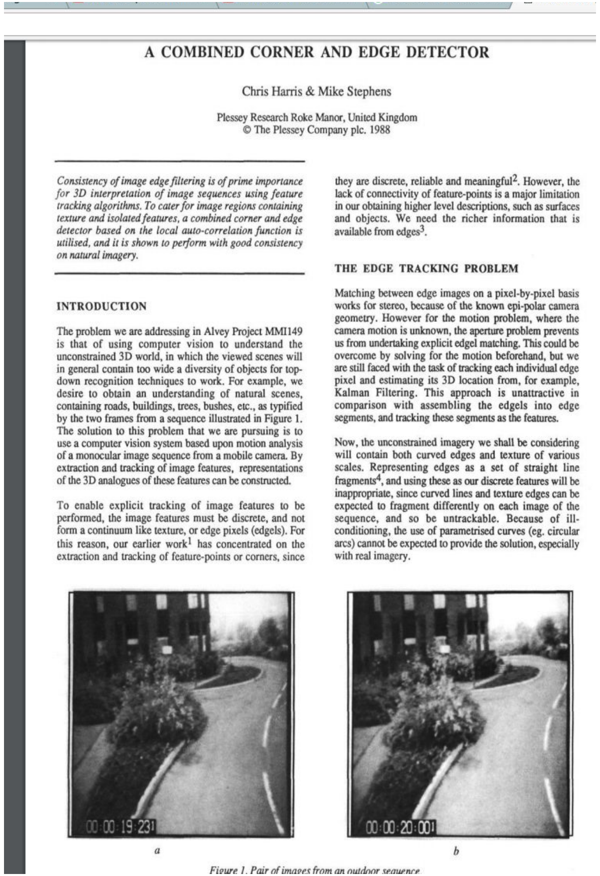
Figure 4.3 Image pairs with extracted patches below. Notice how some patches can be localized or matched with higher accuracy than others.

Approach

- Find some reliable interest points, Keypoints or Corners.
- Create a descriptor which describes those Keypoints.
- Localize them in other image.

Harris Detector

- Cited by 13663
- Probably the only paper of Harris
- http://www.bmva.org/bmvc/1988/avc-88-023.pdf



A COMBINED CORNER AND EDGE DETECTOR

Chris Harris & Mike Stephens

Plessey Research Roke Manor, United Kingdom
© The Plessey Company plc. 1988

Consistency of image edge filtering is of prime importance for 3D interpretation of image sequences using feature tracking algorithms. To cater for image regions containing texture and isolated features, a combined corner and edge detector based on the local auto-correlation function is utilised, and it is shown to perform with good consistency on natural imagery.

INTRODUCTION

The problem we are addressing in Alvey Project MMI149 is that of using computer vision to understand the unconstrained 3D world, in which the viewed scenes will in general contain too wide a diversity of objects for top-down recognition techniques to work. For example, we desire to obtain an understanding of natural scenes, containing roads, buildings, trees, bushes, etc., as typified by the two frames from a sequence illustrated in Figure 1. The solution to this problem that we are pursuing is to use a computer vision system based upon motion analysis of a monocular image sequence from a mobile camera. By extraction and tracking of image features, representations of the 3D analogues of these features can be constructed.

To enable explicit tracking of image features to be performed, the image features must be discrete, and not form a continuum like texture, or edge pixels (edgels). For this reason, our earlier work[1] has concentrated on the extraction and tracking of feature-points or corners, since they are discrete, reliable and meaningful[2]. However, the lack of connectivity of feature-points is a major limitation in our obtaining higher level descriptions, such as surfaces and objects. We need the richer information that is available from edges[3].

THE EDGE TRACKING PROBLEM

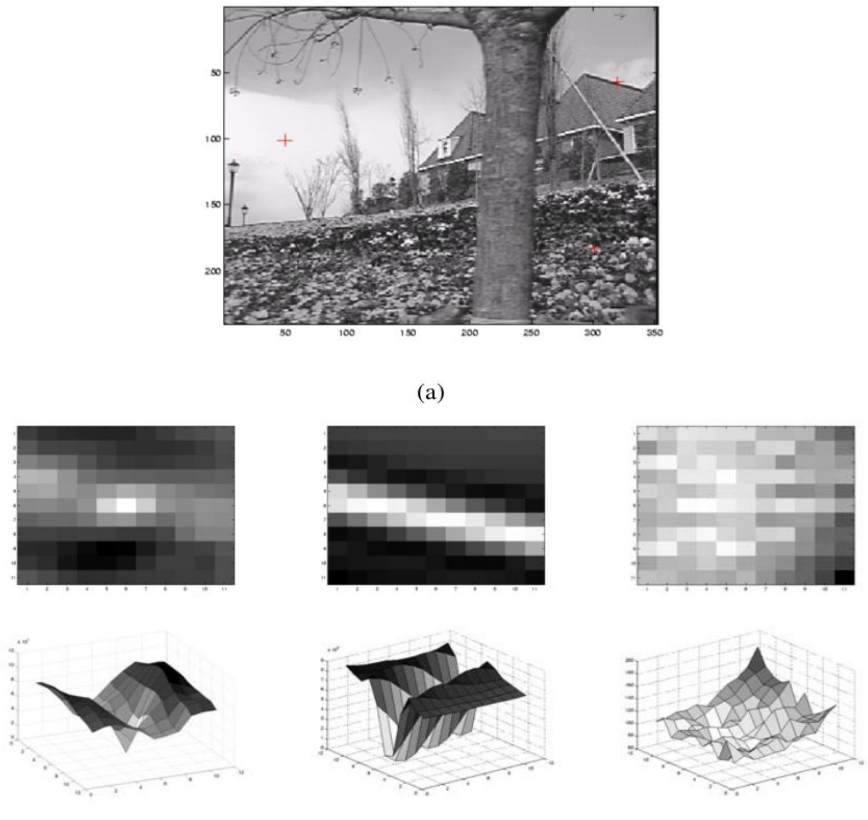
Matching between edge images on a pixel-by-pixel basis works for stereo, because of the known epi-polar camera geometry. However for the motion problem, where the camera motion is unknown, the aperture problem prevents us from undertaking explicit edgel matching. This could be overcome by solving for the motion beforehand, but we are still faced with the task of tracking each individual edge pixel and estimating its 3D location from, for example, Kalman Filtering. This approach is unattractive in comparison with assembling the edgels into edge segments, and tracking these segments as the features.

Now, the unconstrained imagery we shall be considering will contain both curved edges and texture of various scales. Representing edges as a set of straight line fragments[4], and using these as our discrete features will be inappropriate, since curved lines and texture edges can be expected to fragment differently on each image of the sequence, and so be untrackable. Because of ill-conditioning, the use of parametrised curves (eg. circular arcs) cannot be expected to provide the solution, especially with real imagery.

Figure 1. Pair of images from an outdoor sequence

Mathematics of Harris Detector

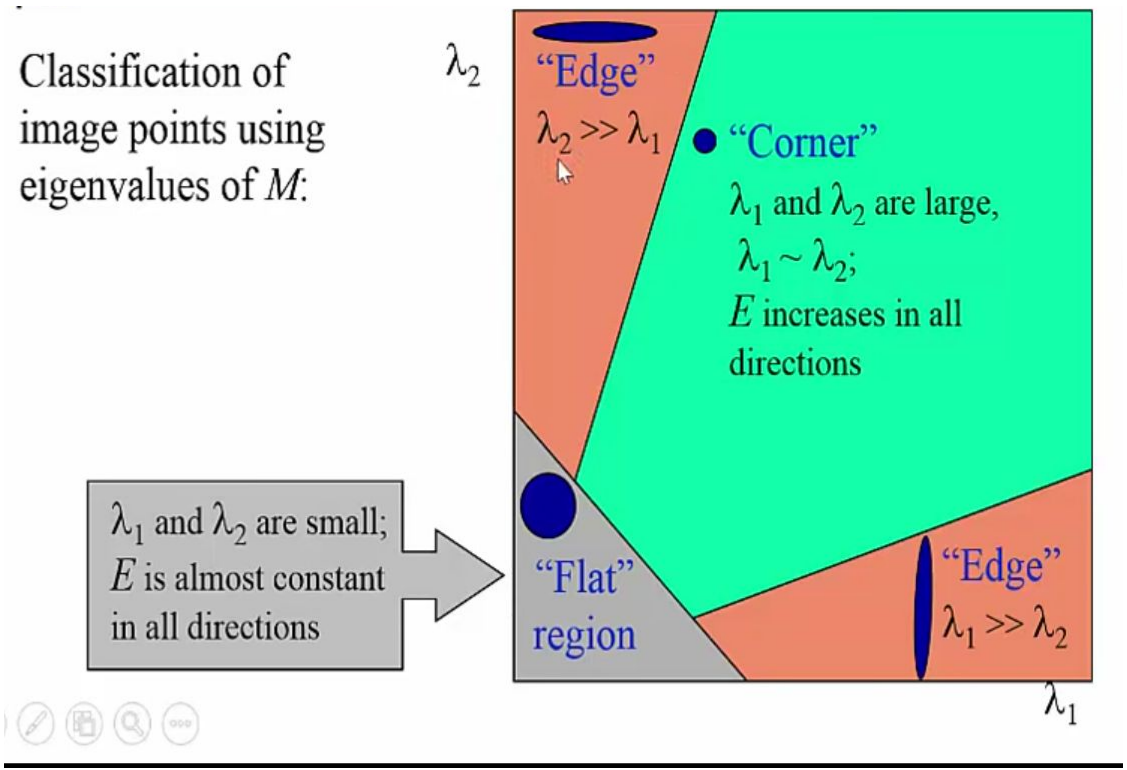
A look into mathematics



(a)

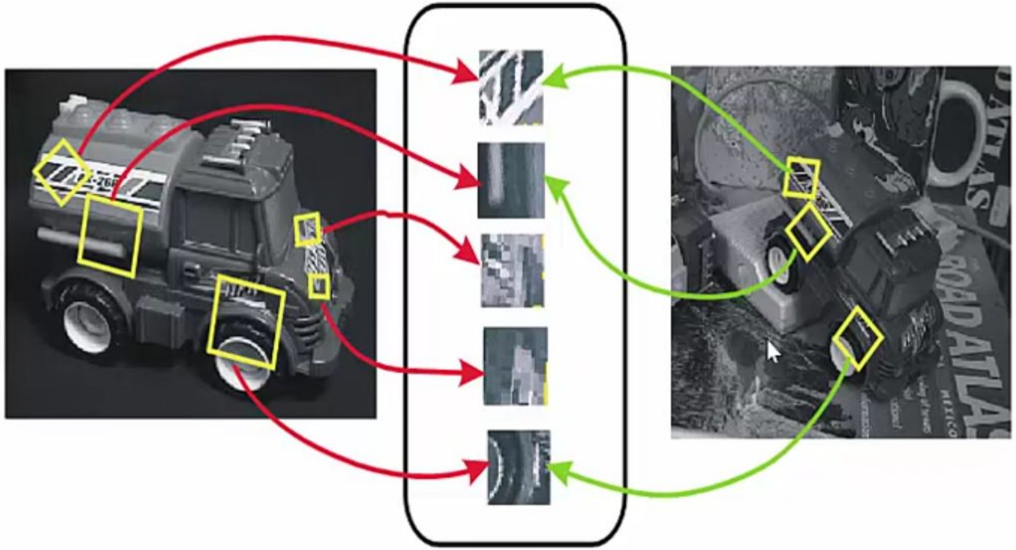
(b) (c) (d)

Mathematics(Continued)



Invariant Local Features

Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters



Features Descriptors

Overview of Keypoint Matching

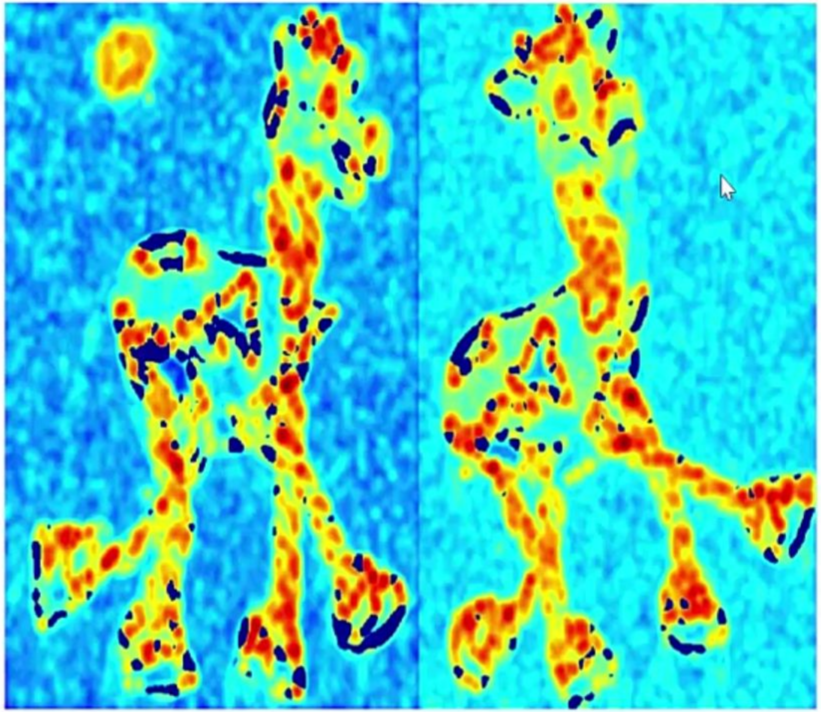
K. Grauman, B. Leibe

Overview of Keypoint Matching

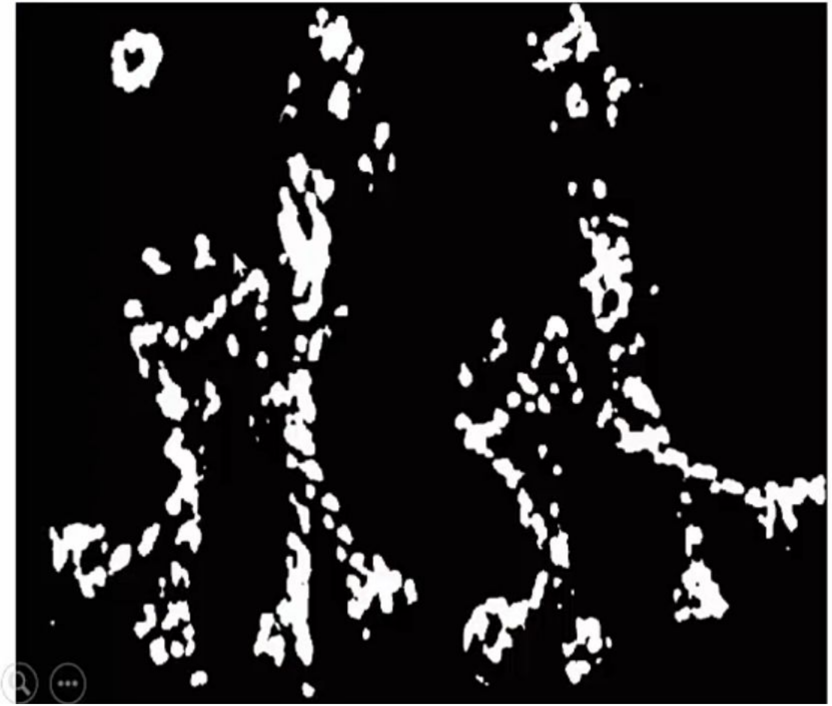
1. **Find a set of distinctive key-points**

K. Grauman, B. Leibe

Compute corner response



Response>Threshold



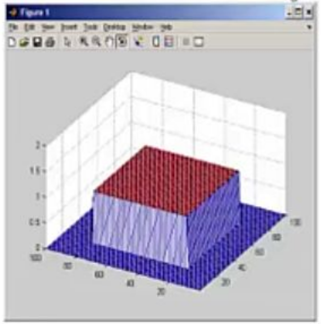
Final Result

- Change of intensity for the shift (u,v)

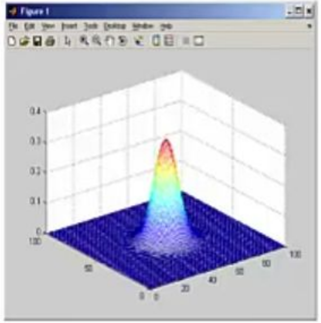
$$E(u,v) = \sum_{x,y} \underbrace{w(x,y)}_{\text{window function}} \, [\underbrace{I(x+u,y+v)}_{\text{shifted intensity}} - \underbrace{I(x,y)}_{\text{intensity}}]^2$$

Auto-correlation

Window functions →

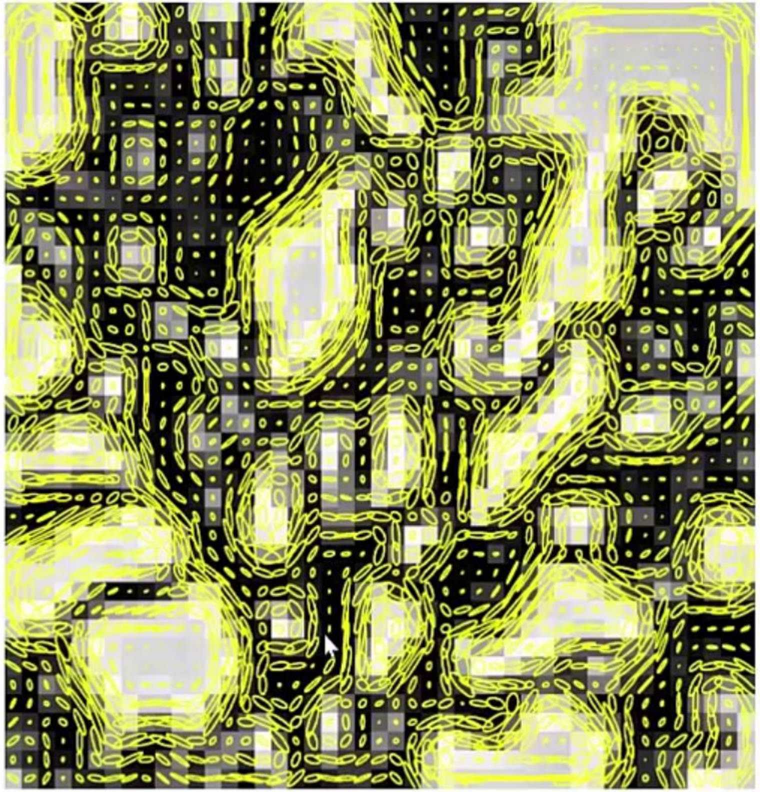


UNIFORM



GAUSSIAN

Visualization of second moment matrices



Slide Credit: James Hays

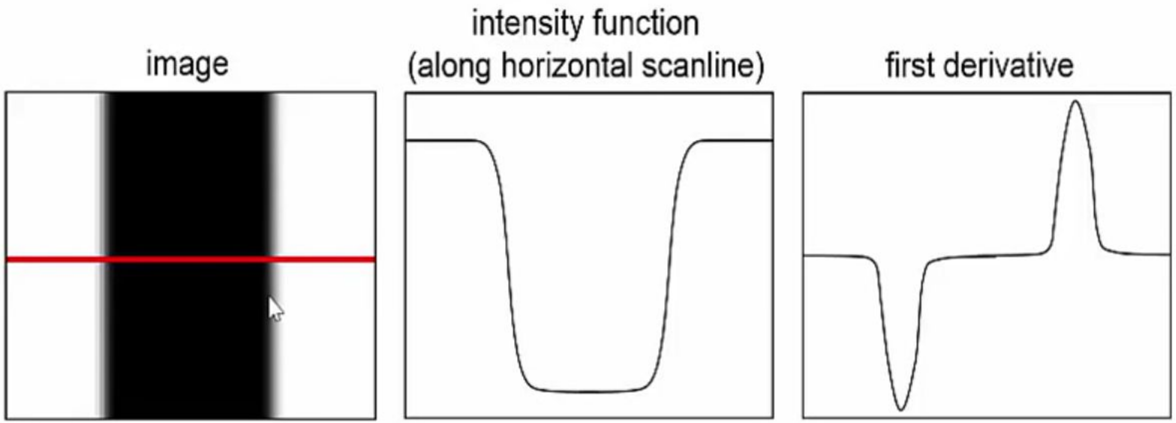
Marr-Hildreth Edge Detector

Steps:

- Smooth image by a gaussian filter ->S
- Apply Laplacian to S
- Find Zero Crossing

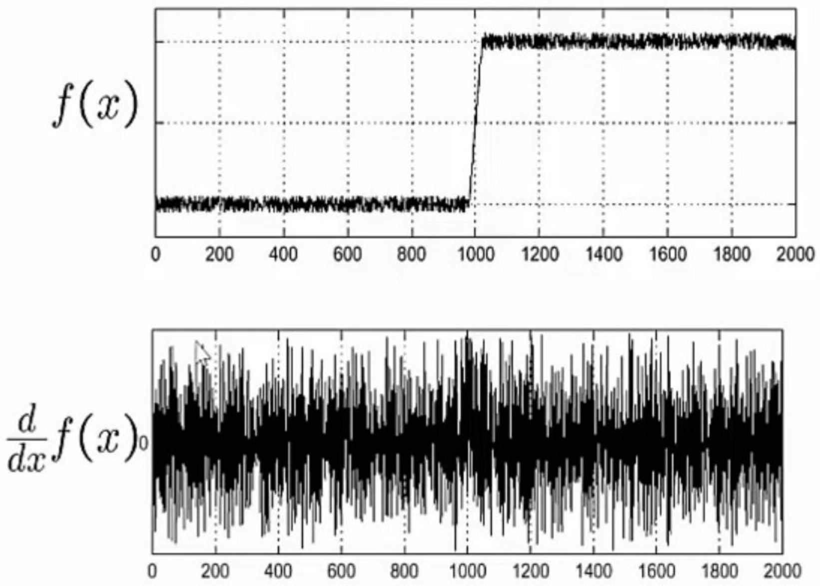
Characterizing edges

- An edge is a place of rapid change in the image intensity function



Slide Credit: James Hays

Why smoothing?

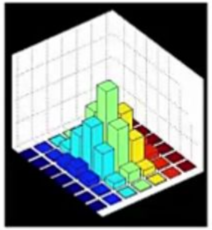


Marr Hildreth Edge Detector

- **Gaussian smoothing**

$$\overbrace{S}^{\text{smoothed image}} = \overbrace{g}^{\text{Gaussian filter}} * \overbrace{I}^{\text{image}}$$

$$g = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



- **Find Laplacian**

$$\Delta^2 S = \overbrace{\frac{\partial^2}{\partial x^2} S}^{\text{secondorder derivative in } x} + \overbrace{\frac{\partial^2}{\partial y^2} S}^{\text{secondorder derivative in } y}$$

- ∇ is used for gradient (first derivative)
- Δ^2 is used for Laplacian (Secondt derivative

Gaussian Mask

$$g(x,y) = e^{\frac{-(x^2+y^2)}{2o^2}}$$

0	0	0	0	1	2	2	2	1	0	0	0	0
0	0	1	3	6	9	11	9	6	3	1	0	0
0	1	4	11	20	30	34	30	20	11	4	1	0
0	3	11	26	50	73	82	73	50	26	11	3	0
1	6	20	50	93	136	154	136	93	50	20	6	1
2	9	30	73	136	198	225	198	136	73	30	9	2
2	11	34	82	154	225	255	225	154	82	34	11	2
2	9	30	73	136	198	225	198	136	73	30	9	2
1	6	20	50	93	136	154	136	93	50	20	6	1
0	3	11	26	50	73	82	73	50	26	11	3	0
0	1	4	11	20	30	34	30	20	11	4	1	0
0	0	1	3	6	9	11	9	6	3	1	0	0
0	0	0	0	1	2	2	2	1	0	0	0	0

$\sigma = 2$

LOG

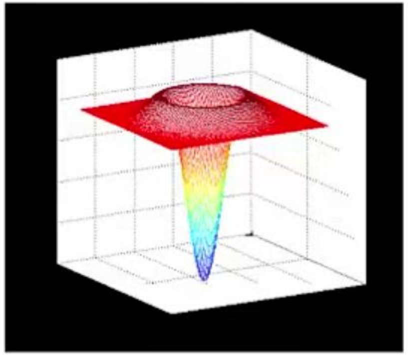
- Deriving the Laplacian of Gaussian (LoG)

$$\Delta^2 S = \Delta^2(g * I) = (\Delta^2 g) * I$$

$$g = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

$$g_x = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}\left(-\frac{2x}{2\sigma^2}\right)$$

$$\Delta^2 g = -\frac{1}{\sqrt{2\pi}\sigma^3}\left(2-\frac{x^2+y^2}{\sigma^2}\right)e^{-\frac{x^2+y^2}{2\sigma^2}}$$



SIFT(Scale Invariant Feature Transform)

Facts:-

- 40,106 Citations !!
- Patented by University of British Columbia. (Well,If ideas start getting patented, Scientists would be the richest community)
- Transform image into scale invariant coordinates.

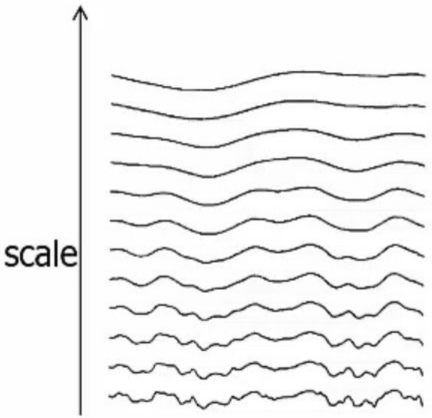
Robustness to

- Affine distortion(rotation , shear , scale) distortion
- Change in 3D viewpoint
- Addition of Noise
- Change in Illumination

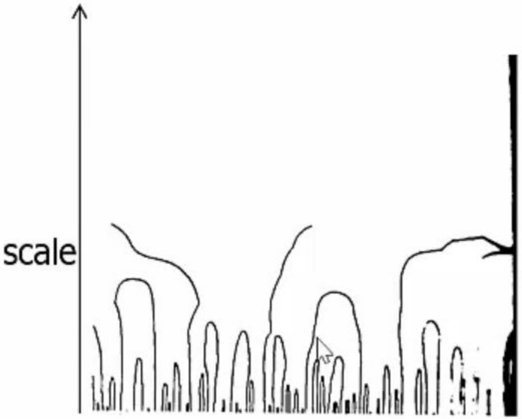
Steps

- Scale-space Peak selection
  - Potential Key point locations
- Key Point Localization
  - Accurately locating feature key points
- Orientation Assignment
  - Assigning orientation to those key points
- Key Point Descriptor
  - Describing the key point as a high dimensional vector<128>.

Scale Space

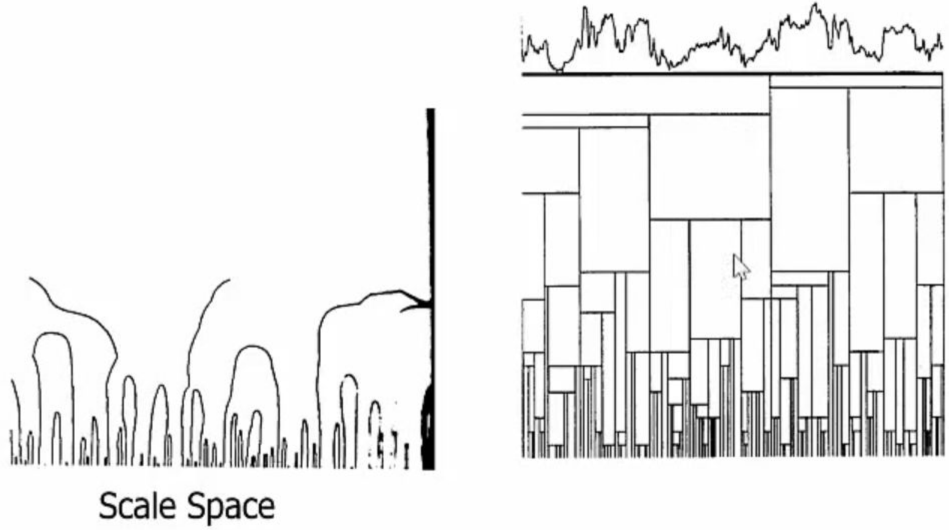


Multiple smooth versions of a signal



Zerocrossings at multiple scale

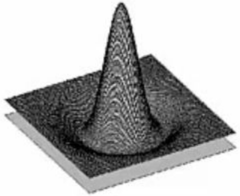
Scale Space(Atkins):-

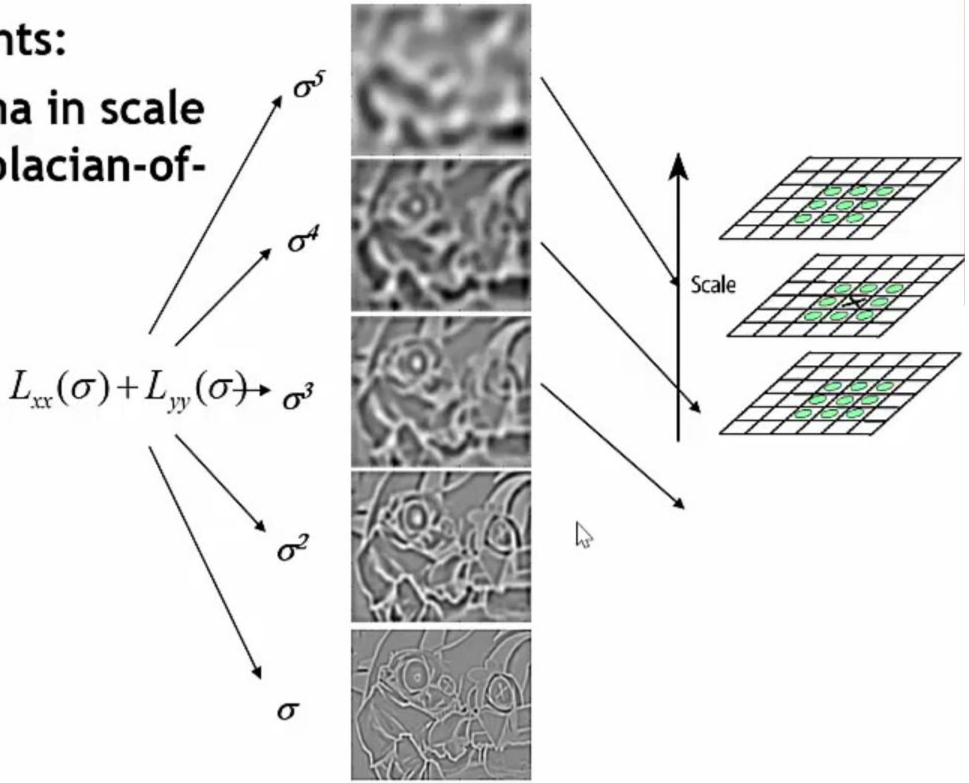


Laplacian-of-Gaussian (LoG)

- Interest points:

  Local maxima in scale space of Laplacian-of-Gaussian

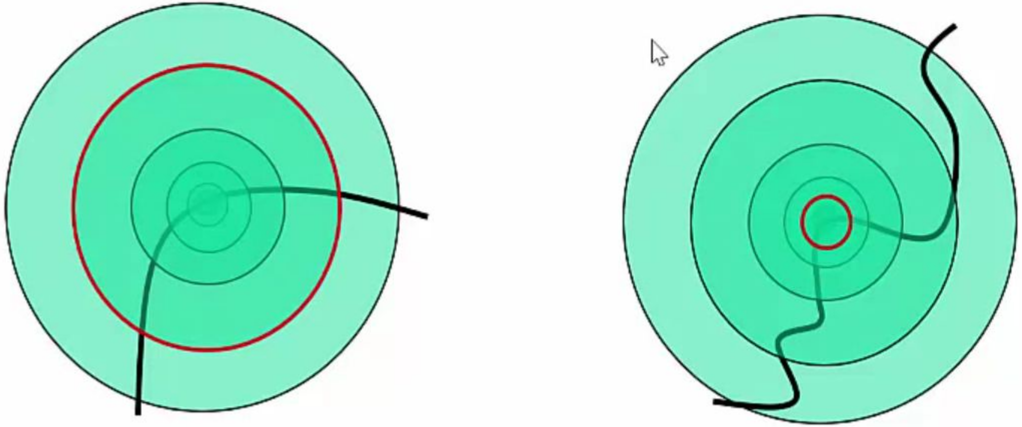




Automatic Scale Selection

Intuition:

- Find scale that gives local maxima of some function *f* in both position and scale.



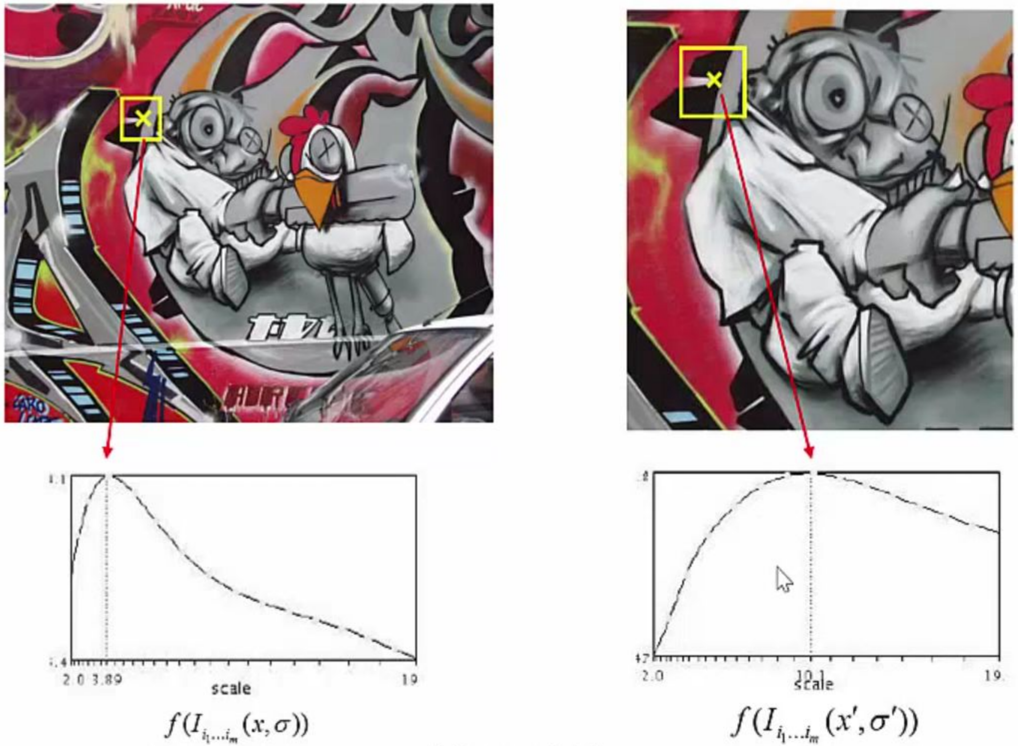
K. Grauman,

How to do this?

$$f(I_{i_1 \ldots i_m}(x, \sigma)) = f(I_{i_1 \ldots i_m}(x', \sigma'))$$

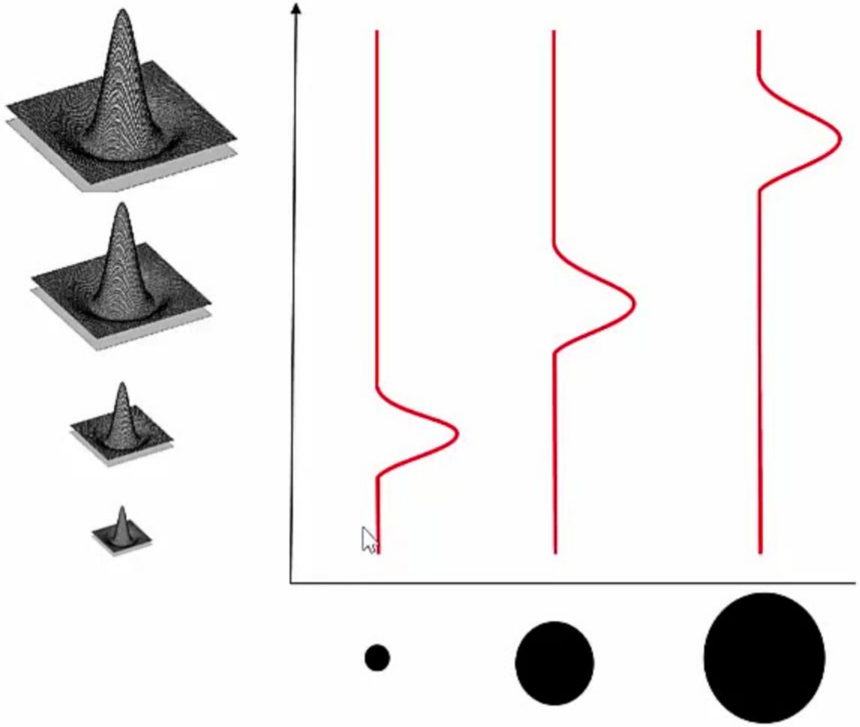
Signature Function

- Function responses for increasing scale (scale signature)

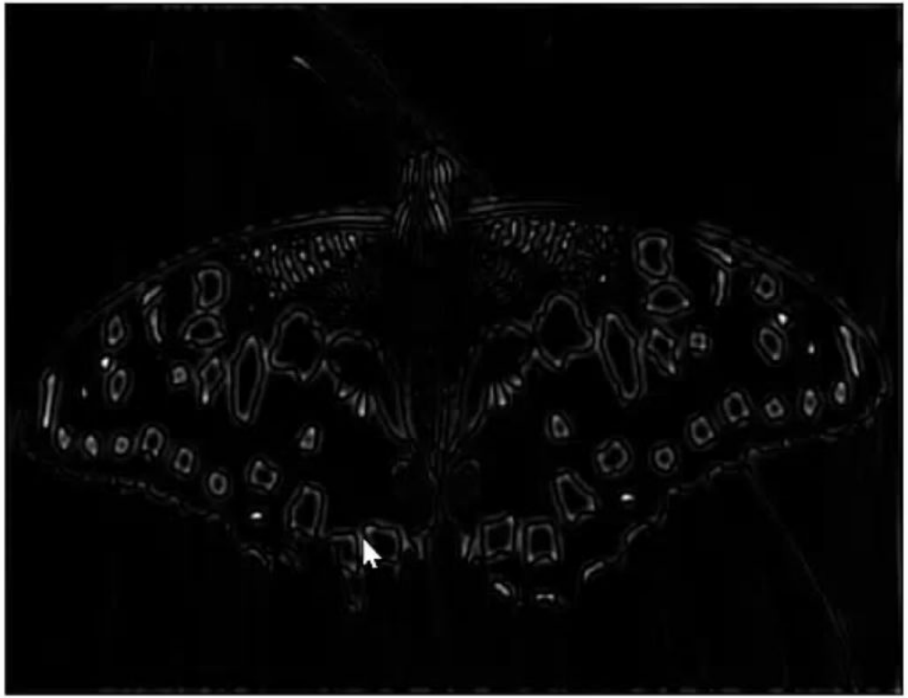


K. Grauman, B. Leibe

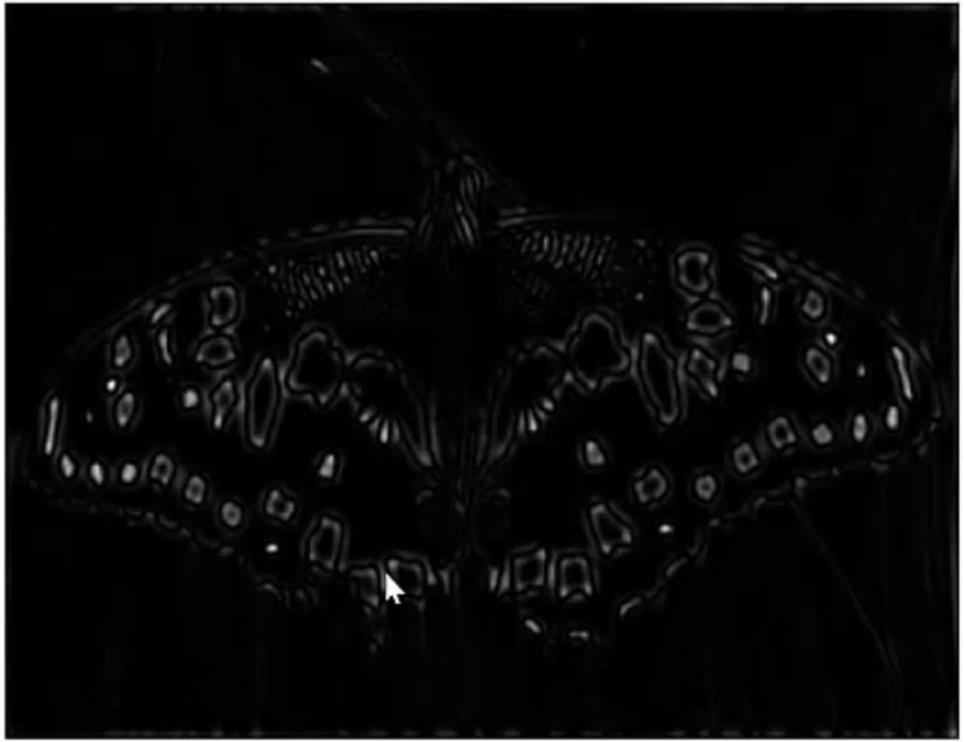
- **Laplacian-of-Gaussian = "blob" detector**



K. Grauman, B. Leibe



sigma = 2

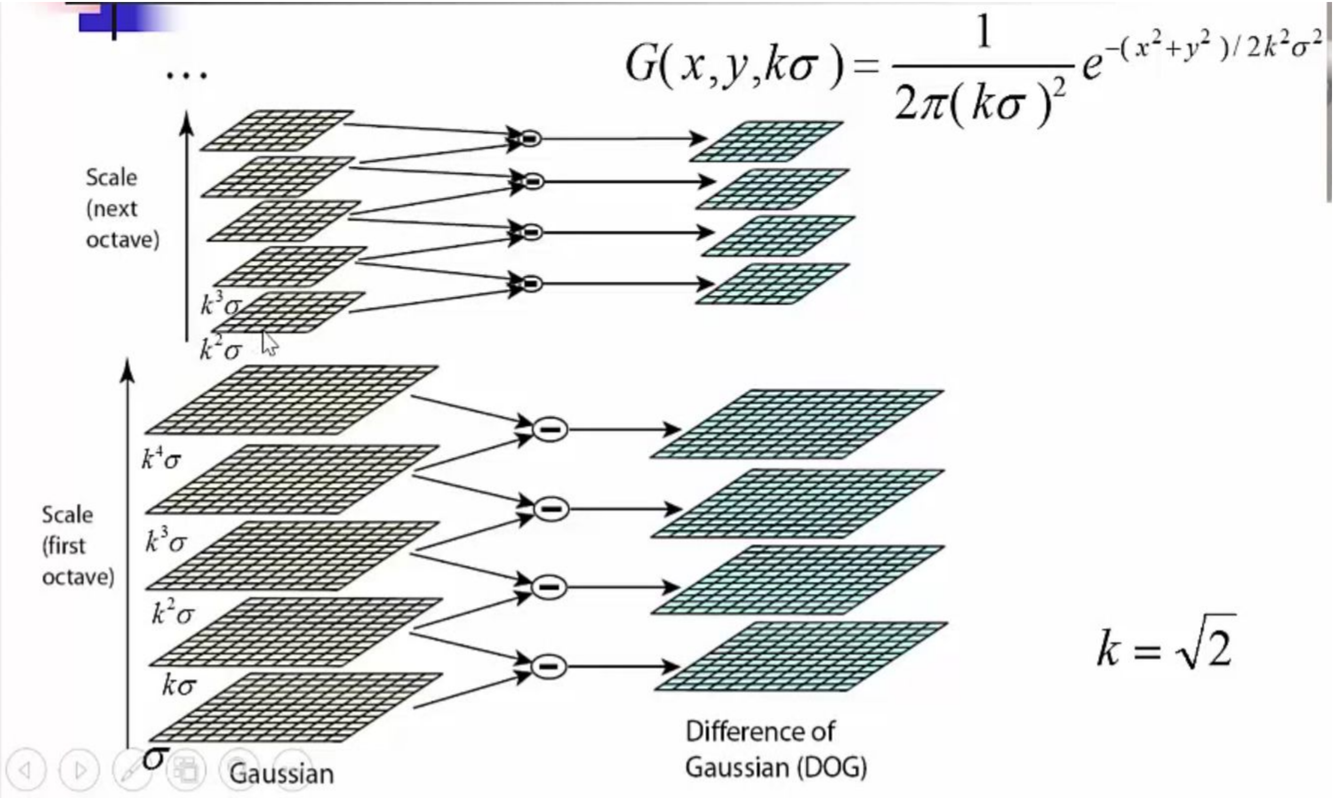


sigma = 2.5018

sigma = 3.1296

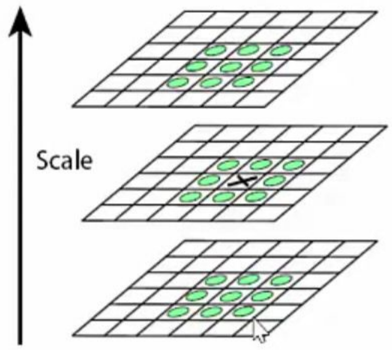
sigma = 3.9149

Building the scale space:-



Finding out interest point:-

- Compare a pixel (**X**) with 26 pixels in current and adjacent scales (Green Circles)



Thank you