CS 6350 – COMPUTER VISION

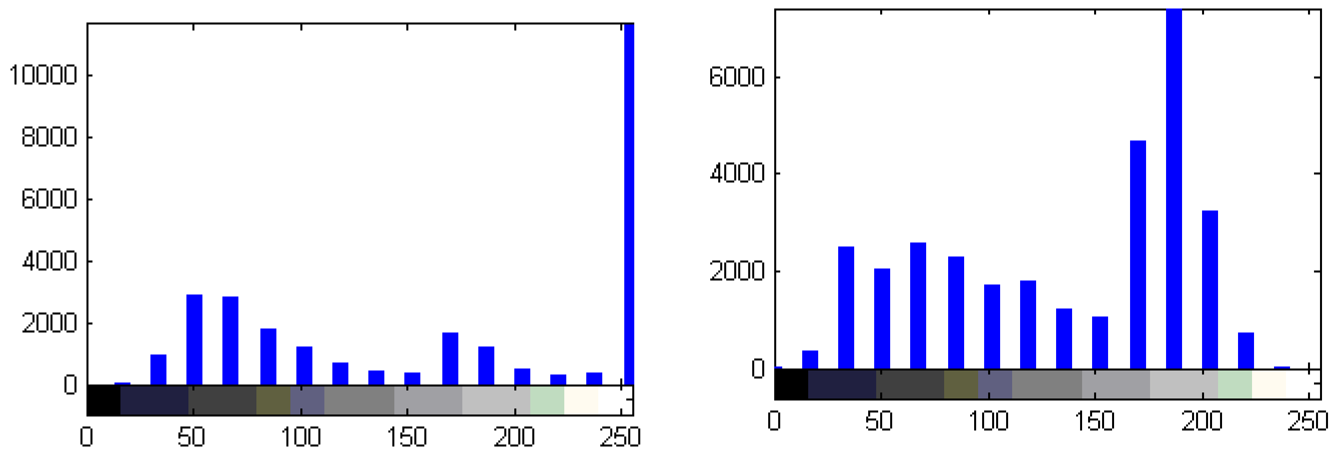
Local Feature Detectors and Descriptors

Overview

- Local invariant features
- Keypoint localization
  - Hessian detector
  - Harris corner detector
- Scale Invariant region detection
  - Laplacian of Gaussian (LOG) detector
  - Difference of Gaussian (DOG) detector
- Local feature descriptor
  - Scale Invariant Feature Transform (SIFT)
  - Gradient Localization Oriented Histogram (GLOH)
- Examples of other local feature descriptors

Motivation

- **Global feature from the whole image is often not desirable**



- **Instead match local regions which are prominent to the object or scene in the image.**
- **Application Area**
  - **Object detection**
  - **Image matching**
  - **Image stitching**

Requirements of a local feature

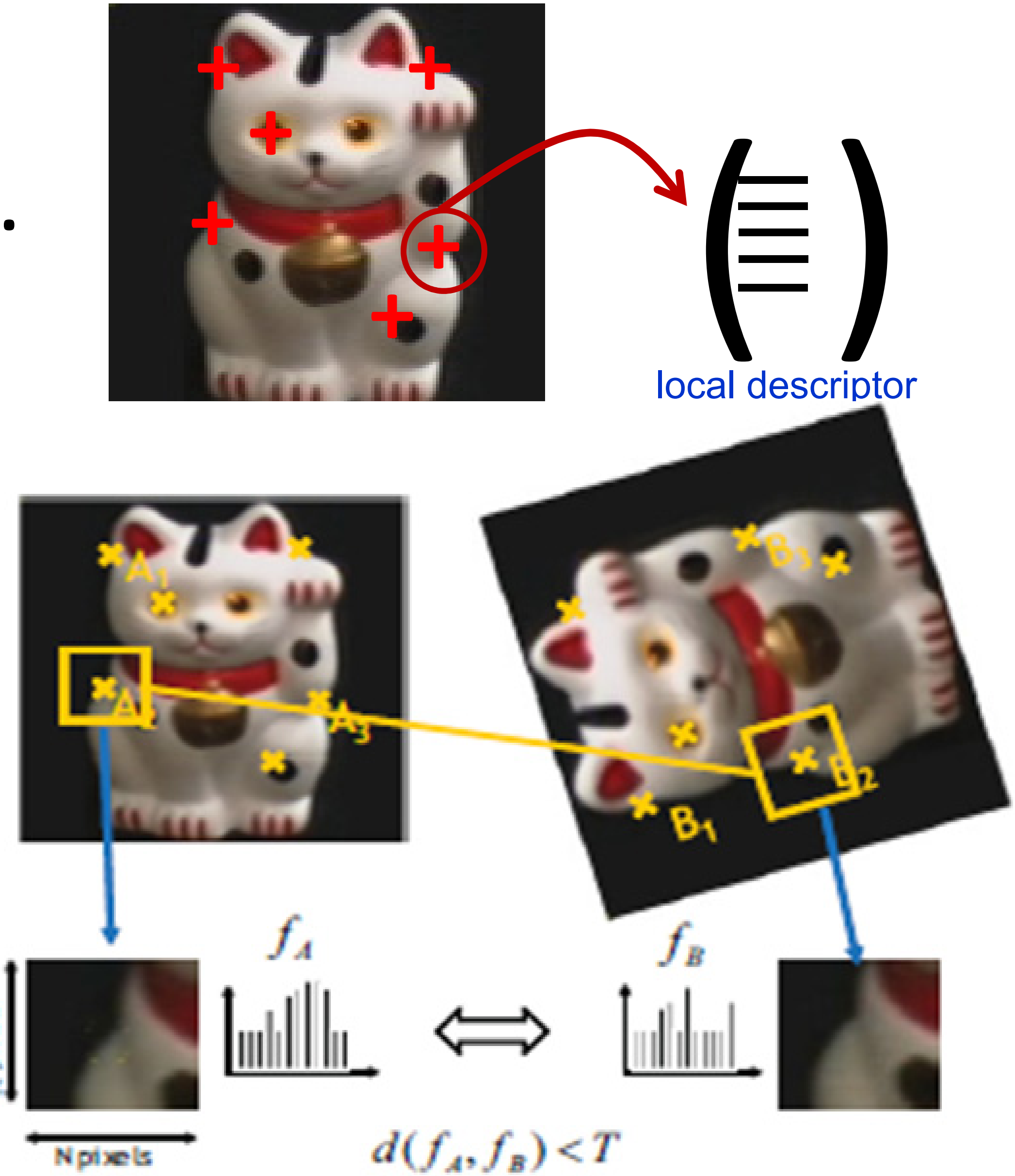
- Repetitive : Detect the same points independently in each image.
- Invariant to translation, rotation, scale.
- Invariant to affine transformation.
- Invariant to presence of noise, blur etc.
- Locality :Robust to occlusion, clutter and illumination change.
- Distinctiveness : The region should contain “interesting” structure.
- Quantity : There should be enough points to represent the image.
- Time efficient.

<u>Others preferable (but not a must):</u>

- **Disturbances, attacks,**
- **Noise**
- **Image blur**
- **Discretization errors**
- **Compression artifacts**
- **Deviations from the mathematical model (non-linearities, non-planarities, etc.)**
- **Intra-class variations**

General approach

1. Find the interest points.
2. Consider the region around each keypoint.
3. Compute a local descriptor from the region and normalize the feature.
4. Match local descriptor

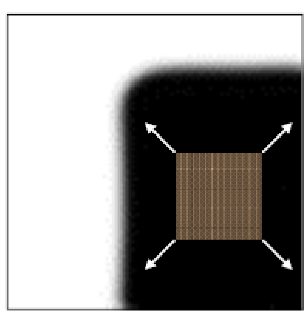


Slide credit: Bastian Leibe

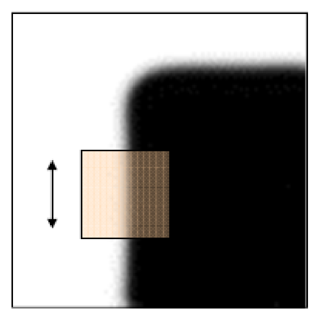
Some popular detectors

- Hessian/ Harris corner detection
- Laplacian of Gaussian (LOG) detector
- Difference of Gaussian (DOG) detector
- Hessian/ Harris Laplacian detector
- Hessian/ Harris Affine detector
- Maximally Stable Extremal Regions (MSER)
- Many others ….

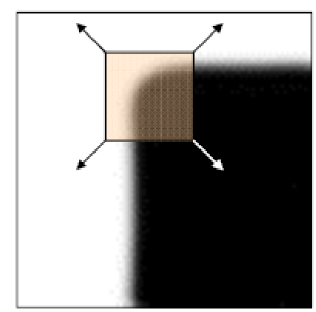
Looks for change in image gradient in two direction - CORNERS



No change in any direction



Change in one direction only



Change in both the directions

Slide credit: Fei Fei Li

Hessian Corner Detector

[Beaudet, 1978]

Searches for image locations which have strong change in gradient along both the orthogonal direction.

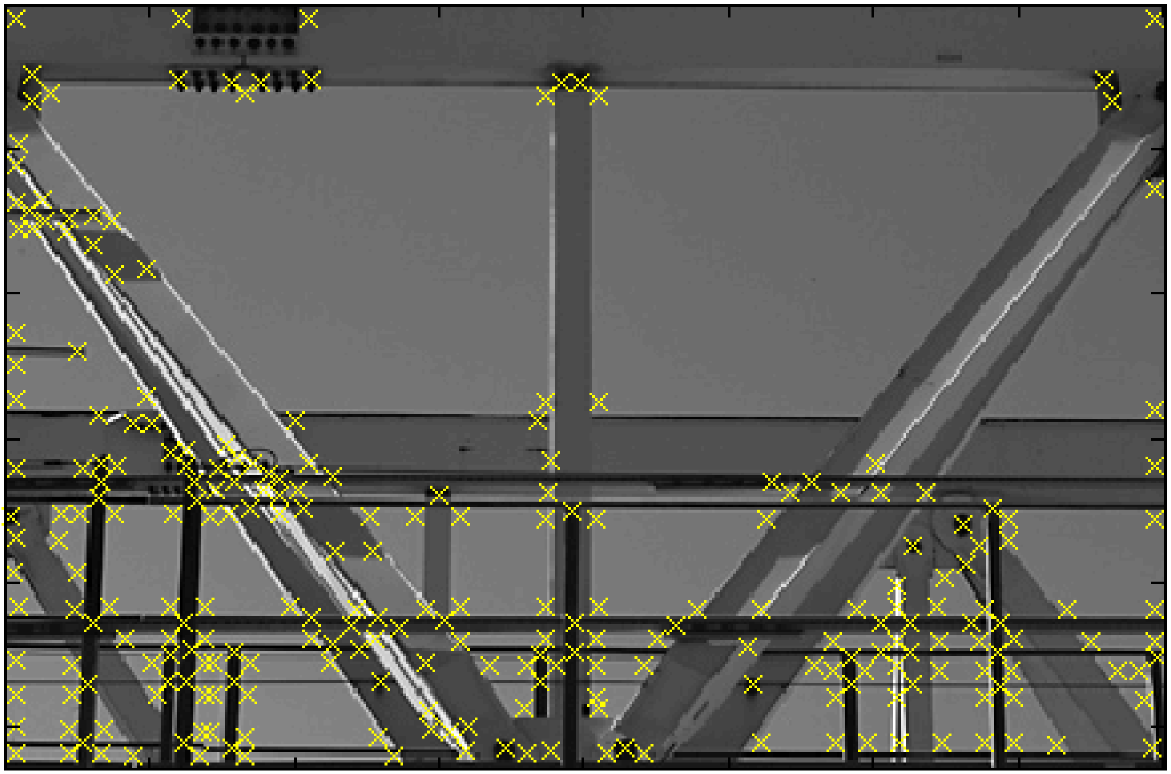
$$\mathrm{H}(\mathrm{x},\sigma) = \begin{bmatrix} \mathrm{I}_{\mathrm{xx}}(\mathrm{x},\sigma) & \mathrm{I}_{\mathrm{xy}}(\mathrm{x},\sigma) \\ \mathrm{I}_{\mathrm{xy}}(\mathrm{x},\sigma) & \mathrm{I}_{\mathrm{yy}}(\mathrm{x},\sigma) \end{bmatrix}$$

$$\boxed{\det(\mathrm{H}) = \mathrm{I}_{\mathrm{xx}}\mathrm{I}_{\mathrm{yy}} - \mathrm{I}_{\mathrm{xy}}^2}$$

- **Perform a non-maximum suppression using a 3*3 window.**
- **Consider points having higher value than its 8 neighbors.**

$$\textbf{Select points where } \det(\mathrm{H}) > \theta$$

Hessian Detector – Result



Effect: **Responses mainly on corners and strongly textured areas.**

Harris Corner

[Forstner and Gulch, 1987]

- **Search for local neighborhoods where the image content has two main directions (eigenvectors).**
- **Consider 2nd moment autocorrelation matrix**

$$\mathbf{C}(\mathbf{x},\sigma,\tilde{\sigma}) = \mathbf{G}(\mathbf{x},\tilde{\sigma}) * \begin{bmatrix} \mathbf{I}_x^2(\mathbf{x},\sigma) & \mathbf{I}_x\mathbf{I}_y(\mathbf{x},\sigma) \\ \mathbf{I}_x\mathbf{I}_y(\mathbf{x},\sigma) & \mathbf{I}_y^2(\mathbf{x},\sigma) \end{bmatrix} \qquad \tilde{\sigma} \approx 2\sigma$$

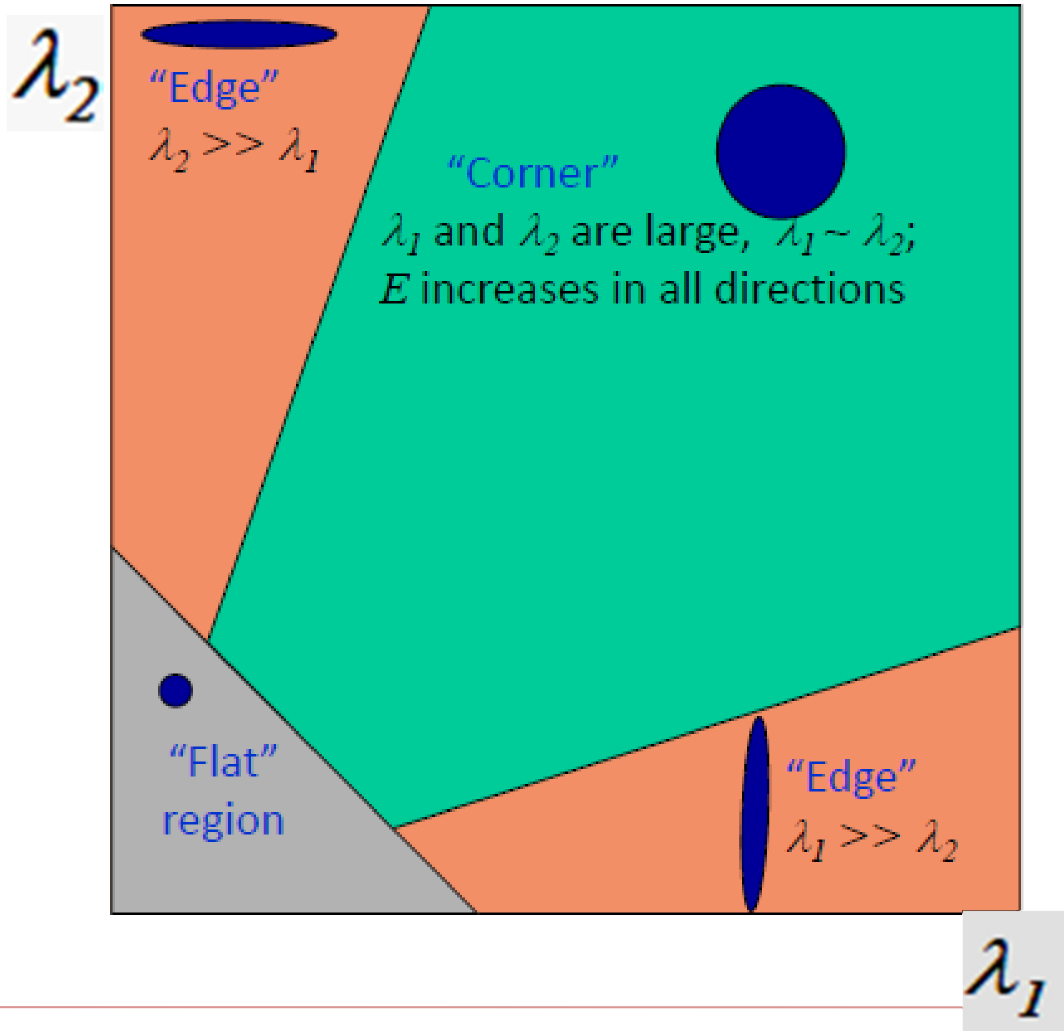
Gaussian sums over all the pixels in circular local neighborhood using weights accordingly.

$$\mathbf{C} = \begin{bmatrix} \sum \mathbf{I}_x^2 & \sum \mathbf{I}_x\mathbf{I}_y \\ \sum \mathbf{I}_x\mathbf{I}_y & \sum \mathbf{I}_y^2 \end{bmatrix} = \mathbf{R}^{-1}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\mathbf{R}$$

Symmetric Matrix

If λ_1 or λ_2 is about 0, the point is not a corner.

Harris corner
Eigen decomposition: visualization



Slide credit: K. Grauman, B. Leibe

Harris Corner: Different approach

Instead of explicitly computing the eigen values, the following equivalence are used

$$\det(\mathrm{C}) = \lambda_1 \lambda_2$$

$$\mathrm{trace}(\mathrm{C}) = \lambda_1 + \lambda_2$$

If, $r = \frac{\lambda_1}{\lambda_2} (\geq 1)$, $\quad \frac{\mathrm{trace}^2(\mathrm{C})}{\det(\mathrm{C})} =$

$$\Rightarrow H_c =$$

$$\det(C) - \alpha . trace^2(C) > threshold$$

Consider these points

α in the range 0.04 – 0.25, experimentally verified

$$\mathbf{det(C)} = \lambda_1 \lambda_2 \qquad \mathbf{trace(C)} = \lambda_1 + \lambda_2$$

$$r = \frac{\lambda_1}{\lambda_2} (\geq 1), \quad \frac{trace^2(C)}{\det(C)} = \frac{(\lambda_1 + \lambda_2)^2}{\lambda_1 \lambda_2} = \frac{(r\lambda_2 + \lambda_2)^2}{r\lambda_2^2} = \frac{(r+1)^2}{r} = r + 2 + (1/r)$$

Min. value of above, when r = 1 ??

Let, r = 2; $trc^2 = dc * (4.5)$

$$\Rightarrow H_c =$$

For Edge: *r* >>1, say 5

$$H_c = dc(1 - 7.2 * 0.1);$$

$$= 0.3 * dc;$$

For, *r* = 10:

$$H_c = dc(1 - 12.1 * 0.05);$$

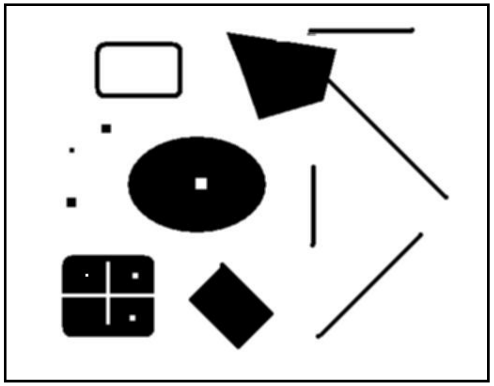
$$= 0.4 * dc;$$

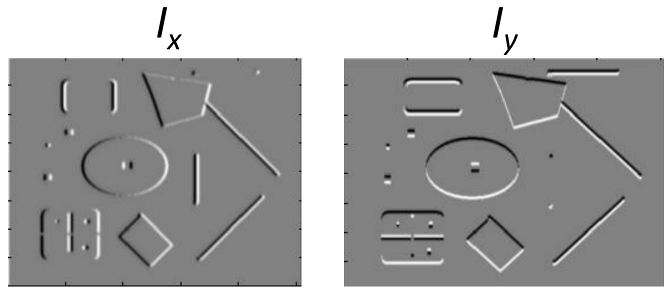
For Corners, *r* = 2

$$H_c = dc(1 - 4.5 * 0.1);$$

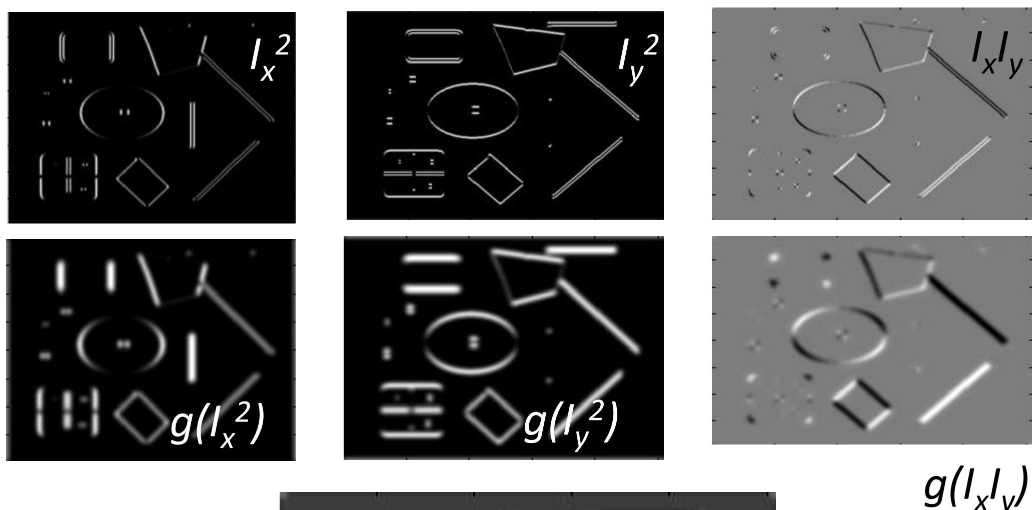
$$= dc * 0.55$$

Harris Corner : Example





1. Image derivatives



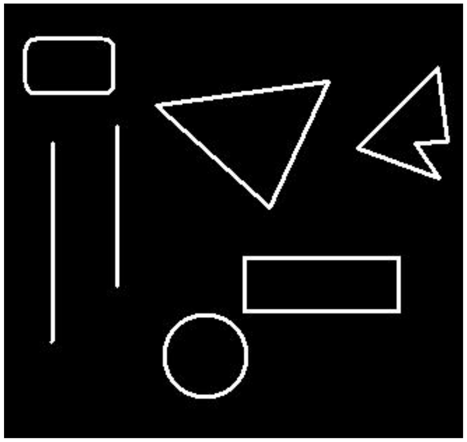
2. Square of derivatives

3. Gaussian filter $G(\sigma_I)$

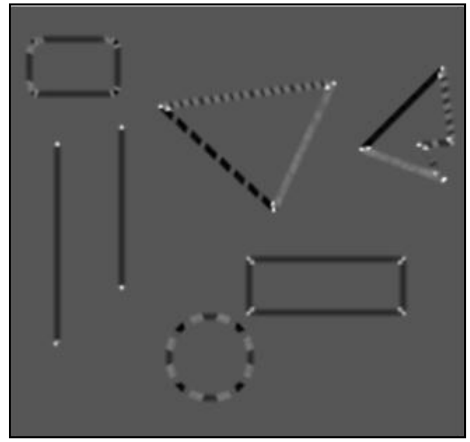
4. Cornerness function – both eigenvalues are strong

Slide credit: K. Grauman, B. Leibe

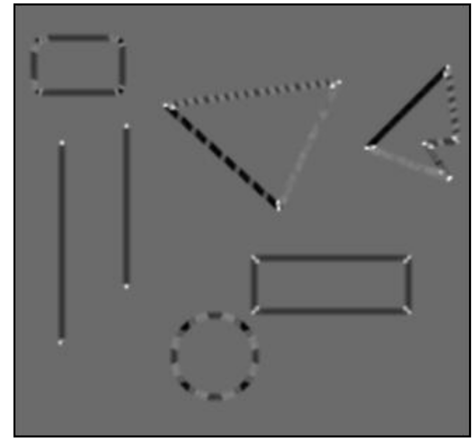
CORNERNESS – HARRIS CORNER



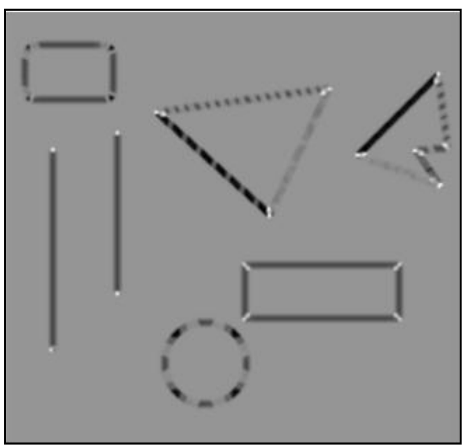
α = .04



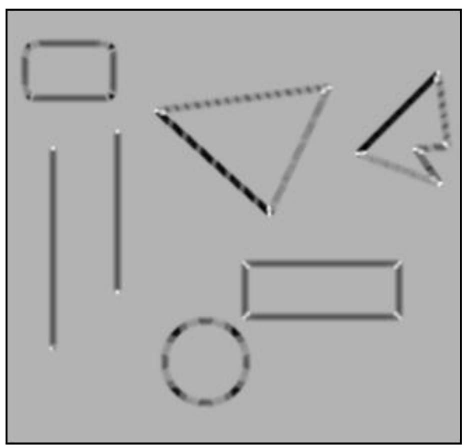
α = .08



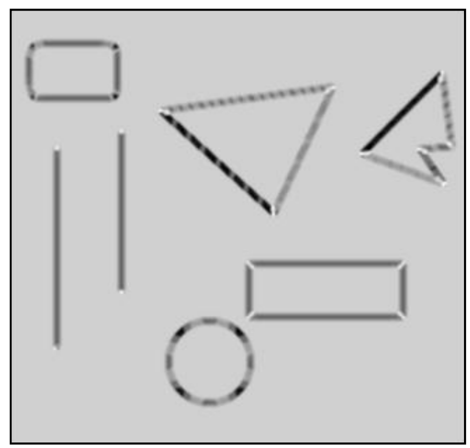
α = .1



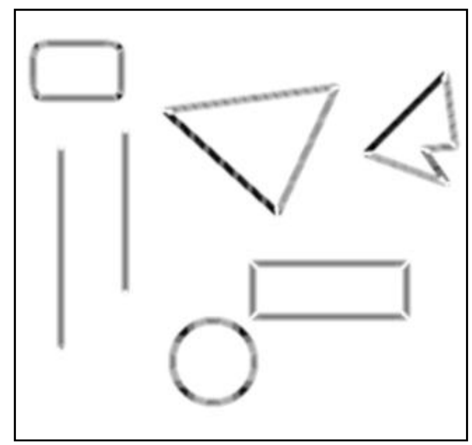
α = .14



α = .17

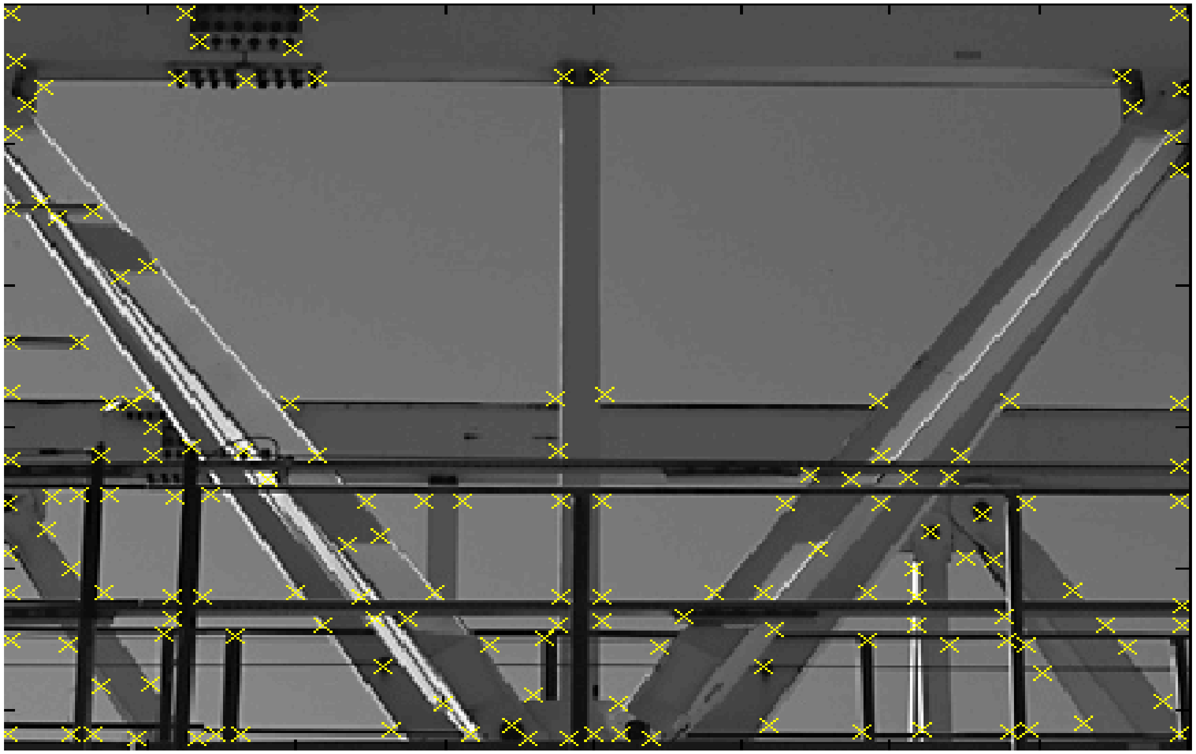


α = .2



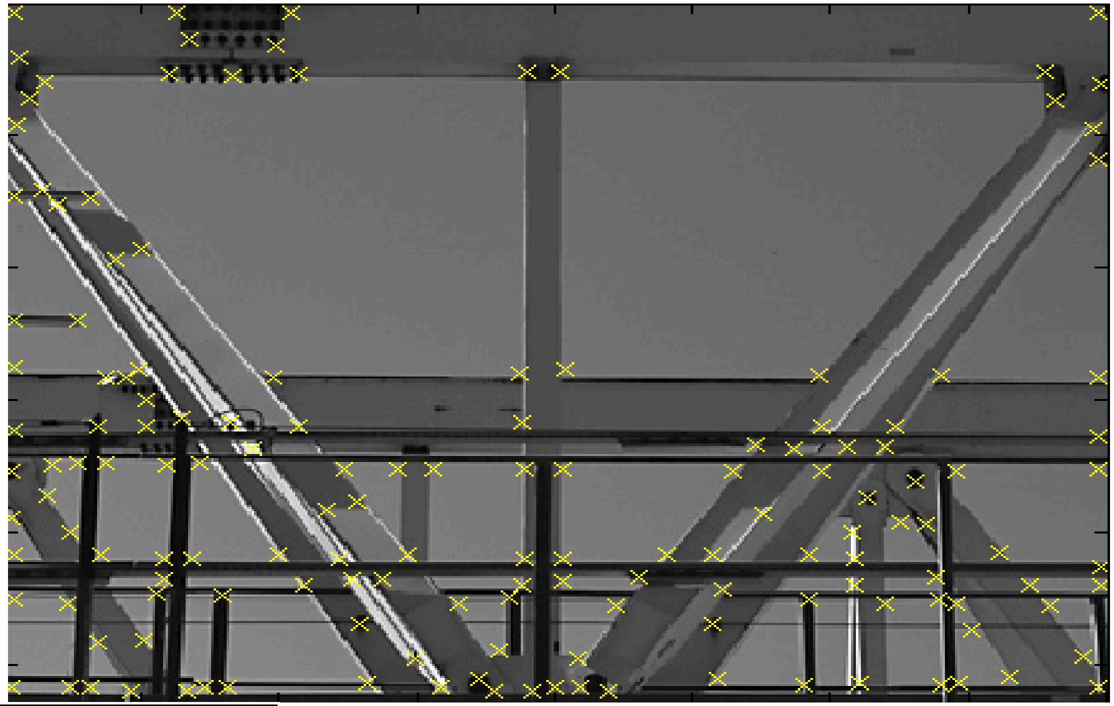
α = .25

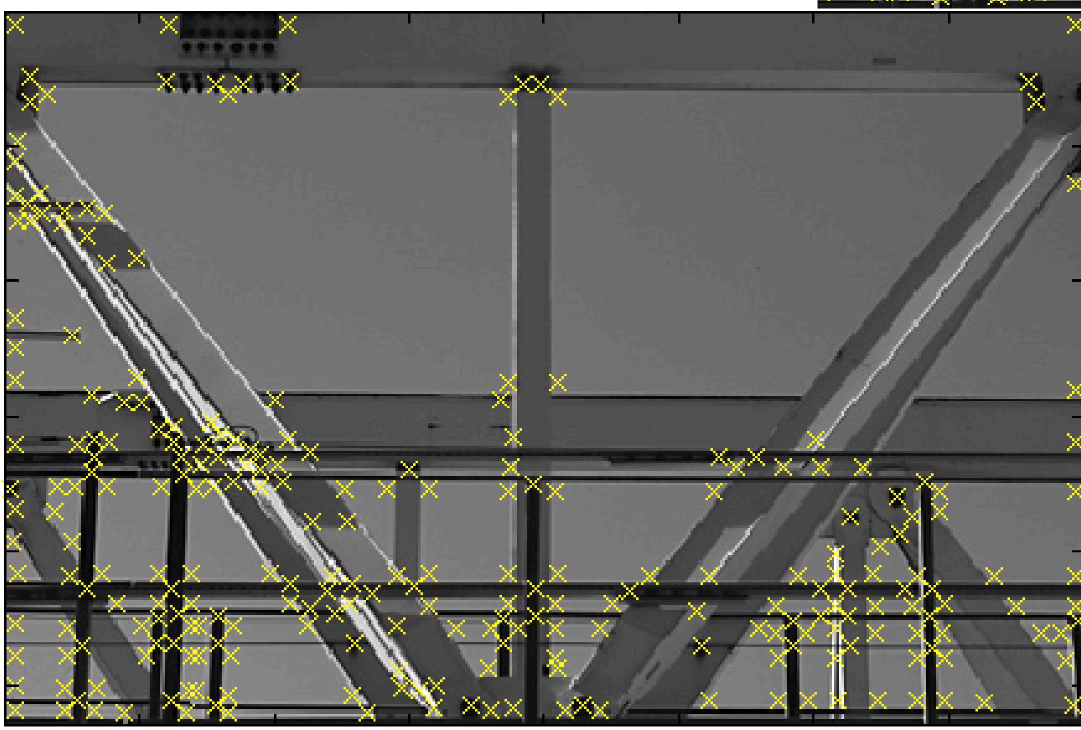
Harris Corner : Result



Effect: **A very precise corner detector.**

Harris Corner



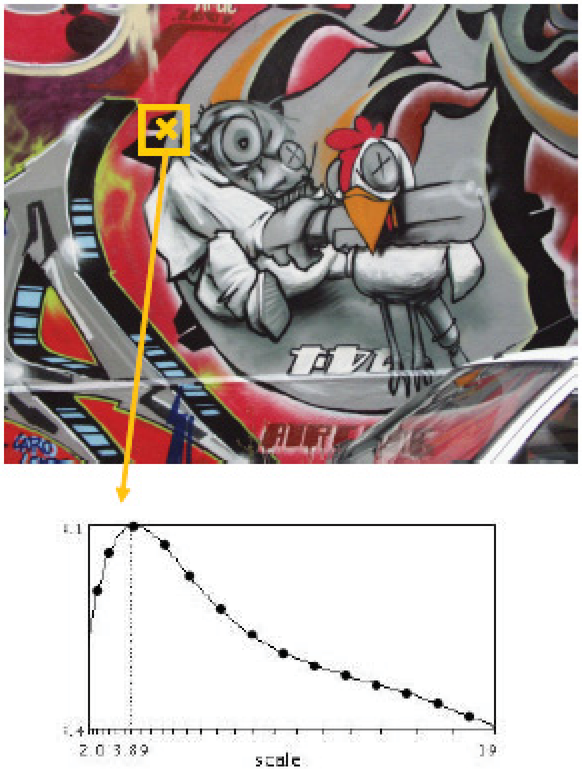


Hessian Detector

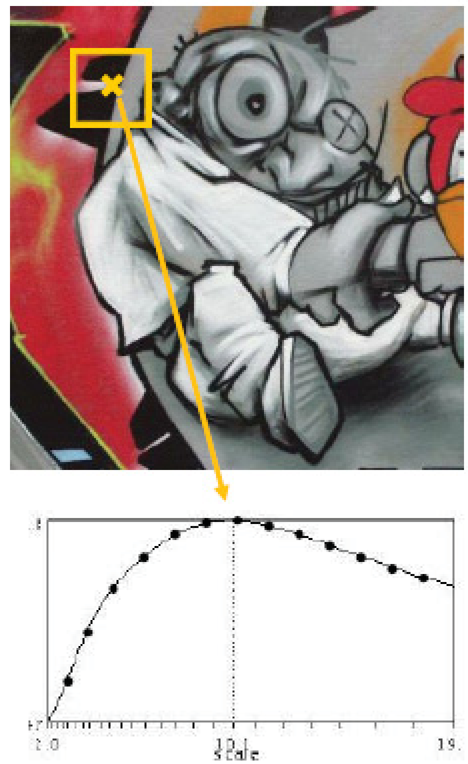
Scale Invariant region detection

Hessian and Harris corner detectors are not scale invariant.

$$\left|LoG(x,\sigma_n)\right| = \sigma_n^2 \left|L_{xx}(x,\sigma_n) + L_{yy}(x,\sigma_n)\right|$$



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



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

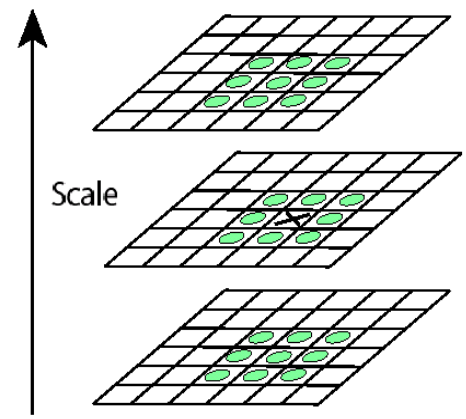
Solution:
Use the concept of Scale Space

Laplacian of Gaussian (LOG) detector [Lindeberg, 1998]

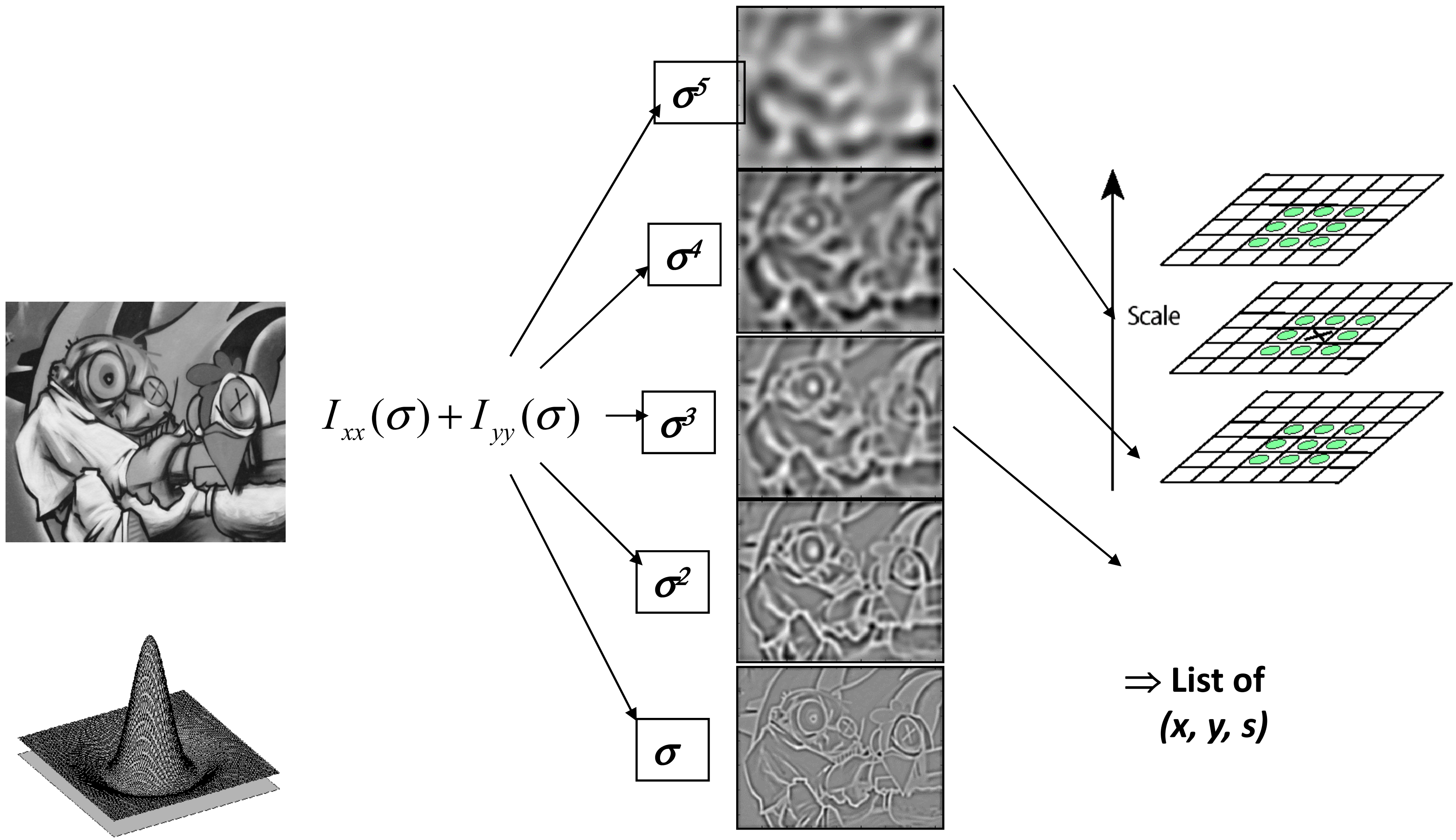
- Using the concept of Scale Space.
- Instead of taking zero crossing (for edge detection), consider the point which is maximum among its 26 neighbors (9+9+8).

$$L(x,\sigma) = \sigma^2 (I_{xx}(x,\sigma) + I_{yy}(x,\sigma))$$

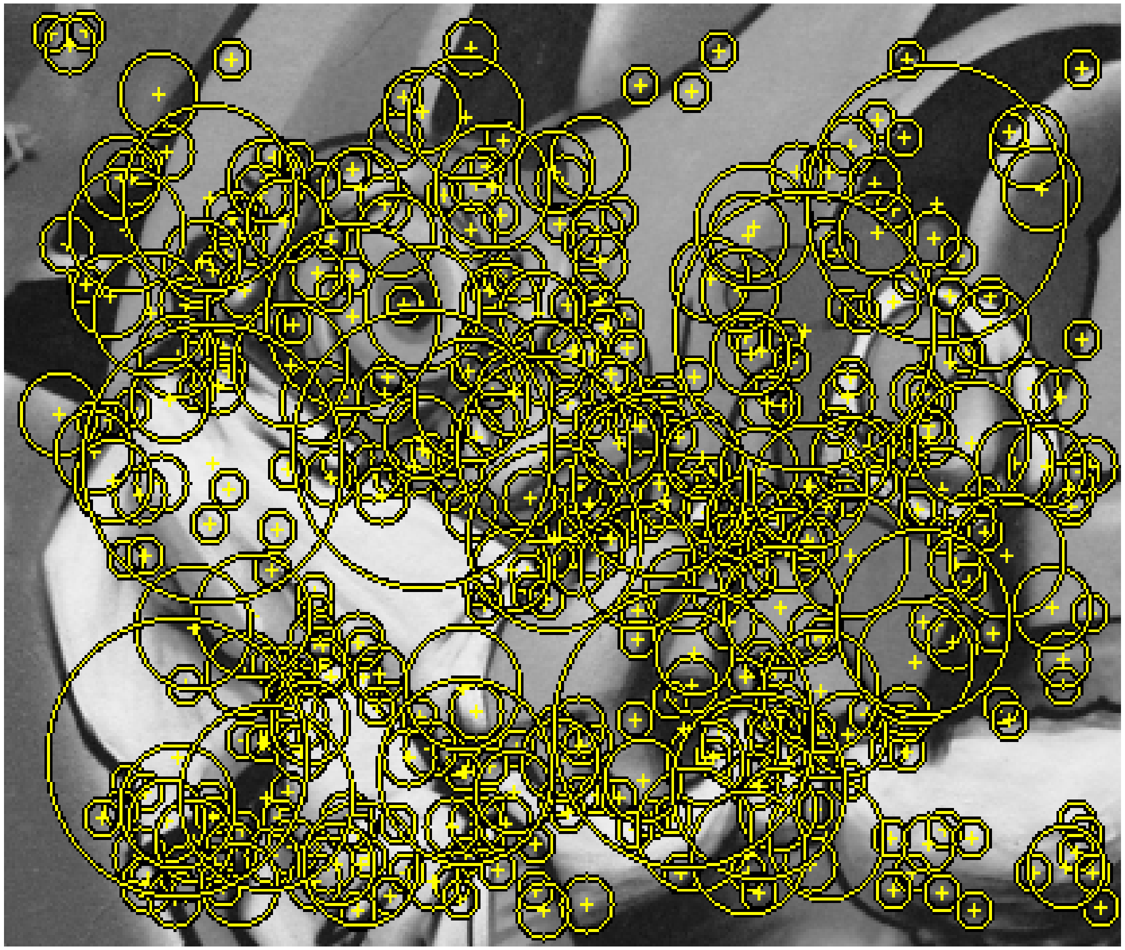
- LOG can be used for finding the characteristic scale for a given image location.
- LOG can be used for finding scale invariant regions by searching 3D (location + scale) extrema of the LOG.
- LOG is also used for edge detection.



LOG detector : Flowchart

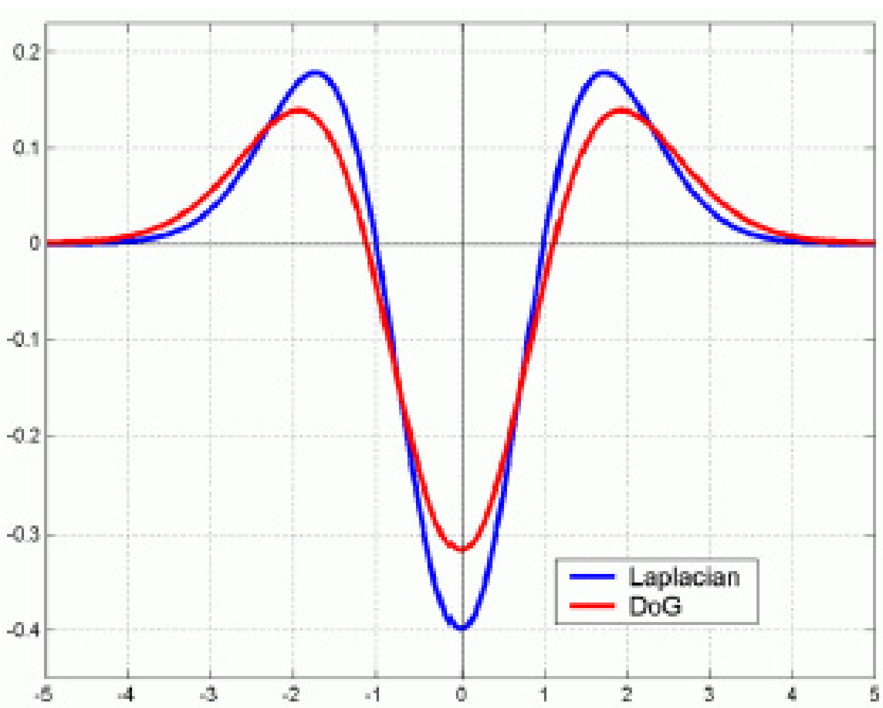


LOG detector : Result



Difference of Gaussian (DOG) Detector [Lowe, 2004]

Approximate LOG using DOG for computational efficiency



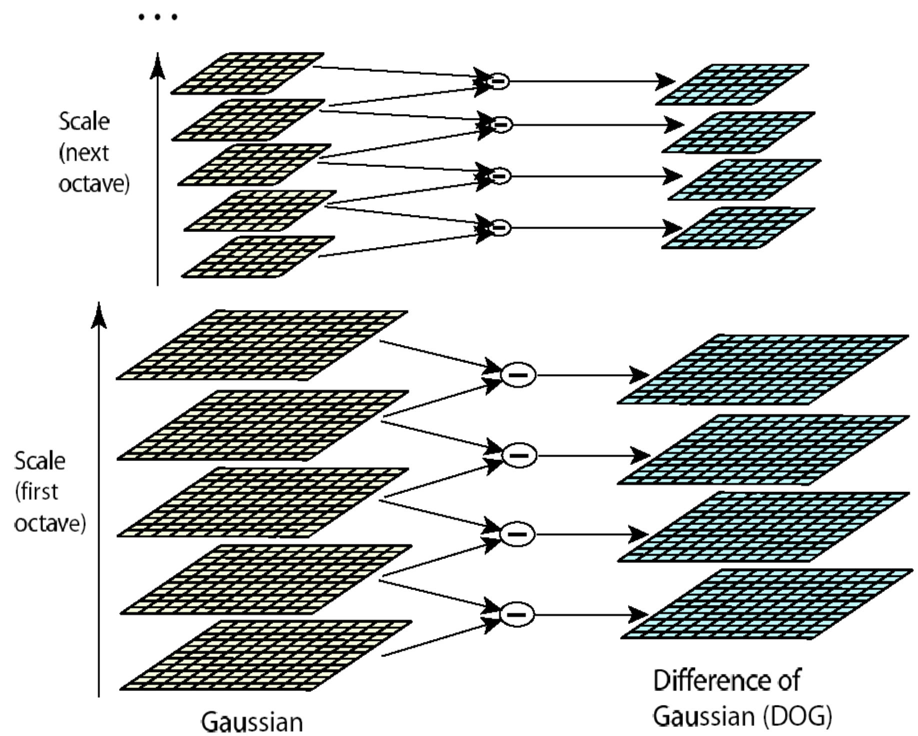
$$D(x,\sigma) = (G(x,k\sigma) - G(x,\sigma)) * I(x)$$

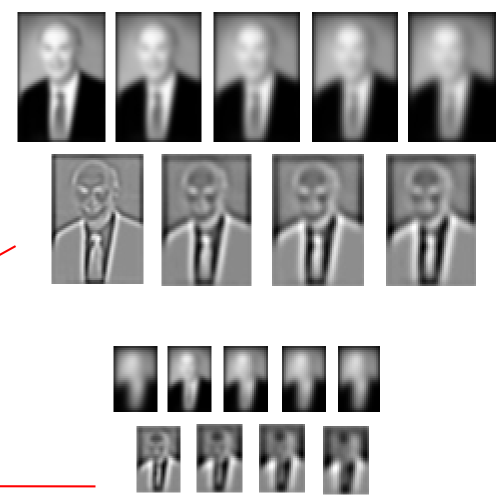
$k = 2^{1/K}$

K = 0, 1, 2, … , constant

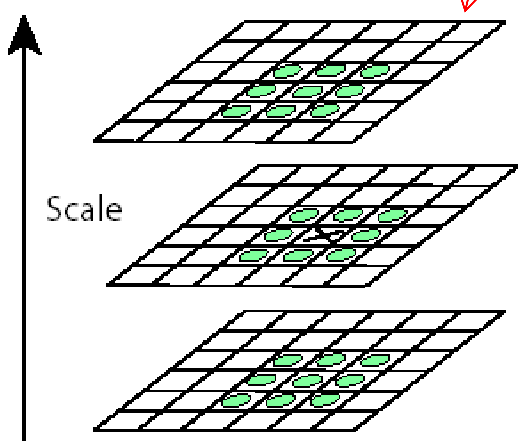
Consider the region where the DOG response is greater than a threshold and the scale lies in a predefined range $[s_{min}, s_{max}]$

DOG detector : Flowchart





X is selected if it is larger or smaller than all neighbors.

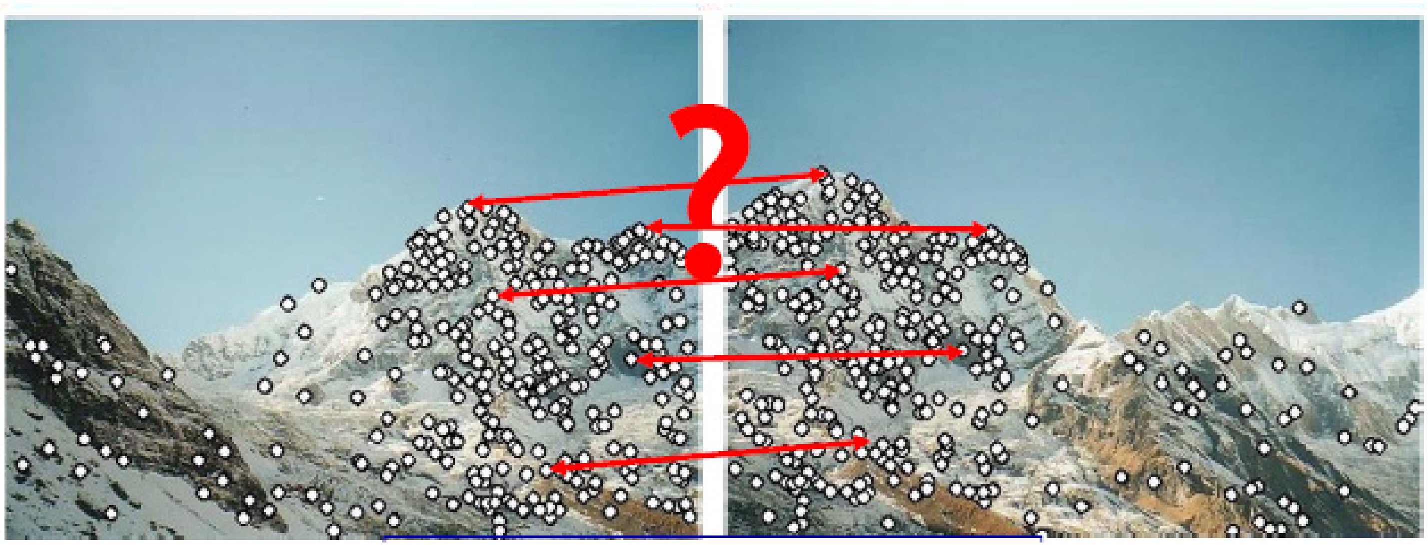


DOG detector : Result

Feature detector	Edge	Corner	Blob
Canny	X		
Sobel	X		
Harris & Stephens / Plessey	X	X	
SUSAN	X	X	
Shi & Tomasi		X	
Level curve curvature		X	
FAST		X	X
Laplacian of Gaussian		X	X
Difference of Gaussians		X	X
Determinant of Hessian		X	X
MSER			X
PCBR			X
Grey-level blobs			X

Local Descriptors

- **We have detected the interest points in an image.**
- **How to match the points across different images of the same object?**



Use Local Descriptors

Slide credit: Fei Fei Li

List of local feature descriptors

- Scale Invariant Feature Transform (SIFT)
- Speed-Up Robust Feature (SURF)
- Histogram of Oriented Gradient (HOG)
- Gradient Location Orientation Histogram (GLOH)
- PCA-SIFT
- Pyramidal HOG (PHOG)
- Pyramidal Histogram Of visual Words (PHOW)
- Others....(shape Context, Steerable filters, Spin images).

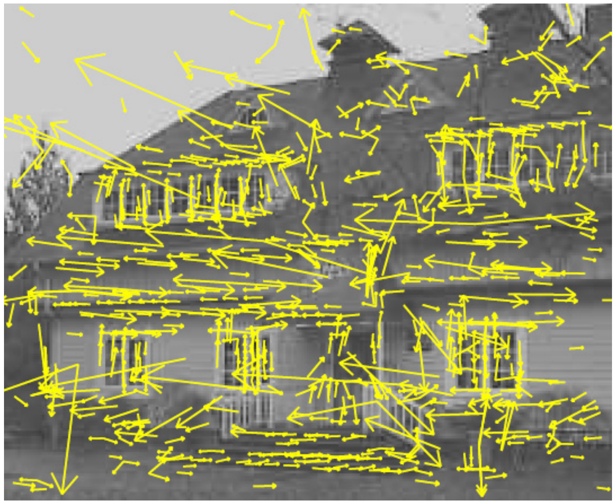
Should be robust to viewpoint change or illumination change

SIFT [Lowe, 2004]

- **Step 1: Scale-space extrema Detection** - Detect interesting points (invariant to scale and orientation) using DOG.
- **Step 2: Keypoint Localization** – Determine location and scale at each candidate location, and select them based on stability.
- **Step 3: Orientation Estimation** – Use local image gradients to assigned orientation to each localized keypoint. Preserve theta, scale and location for each feature.
- **Step 4: Keypoint Descriptor -** Extract local image gradients at selected scale around keypoint and form a representation invariant to local shape distortion and illumination them.

SIFT [Lowe, 2004]

Step 1: Detect interesting points using DOG.



832 DOG extrema

SIFT : Step 2

Step 2: Accurate keypoint localization

- Aim : reject the low contrast points and the points that lie on the edge.

Low contrast points elimination:

Fit keypoint at $\underline{x}$ to nearby data using quadratic approximation.

$$D(\underline{x}) = D + \frac{\partial D^T}{\partial \underline{x}}\underline{x} + \frac{1}{2}\underline{x}^T \frac{\partial^2 D^T}{\partial \underline{x}^2}\underline{x}$$

Where,

$$\mathbf{D(x,\sigma) = (G(x,k\sigma) - G(x,\sigma)) * I(x)}$$

Calculate the local maxima of the fitted function.

Discard local minima (for contrast) $D(\hat{\underline{x}}) < 0.03$

Low contrast points elimination:

Fit keypoint at $\underline{x}$ to nearby data using quadratic approximation.

$$D(\underline{x}) = D + \frac{\partial D^T}{\partial \underline{x}}\underline{x} + \frac{1}{2}\underline{x}^T \frac{\partial^2 D^T}{\partial \underline{x}^2}\underline{x}$$

Calculate the local maxima of the fitted function { $\underline{X}$ = (x, y, σ)}.

$$\partial D / \partial \underline{x} = \frac{\partial\left[D + \frac{\partial D^T}{\partial \underline{x}}\underline{x} + \frac{1}{2}\underline{x}^T \frac{\partial^2 D}{\partial \underline{x}^2}\underline{x}\right]}{\partial \underline{x}} = \qquad = 0$$

$$\Rightarrow$$

$$\underline{\hat{x}} = -\frac{\partial^2 D}{\partial \underline{x}^2}^{-1}\frac{\partial D}{\partial \underline{x}}$$

SIFT : Step 2

Eliminating edge response:

DOG gives strong response along edges – Eliminate those responses

Solution: check "cornerness" of each keypoint.

- **On the edge one of principle curvatures is much bigger than another.**
- **High cornerness ⇔ No dominant principle curvature component.**
- **Consider the concept of Hessian and Harris corner**

Hessian Matrix

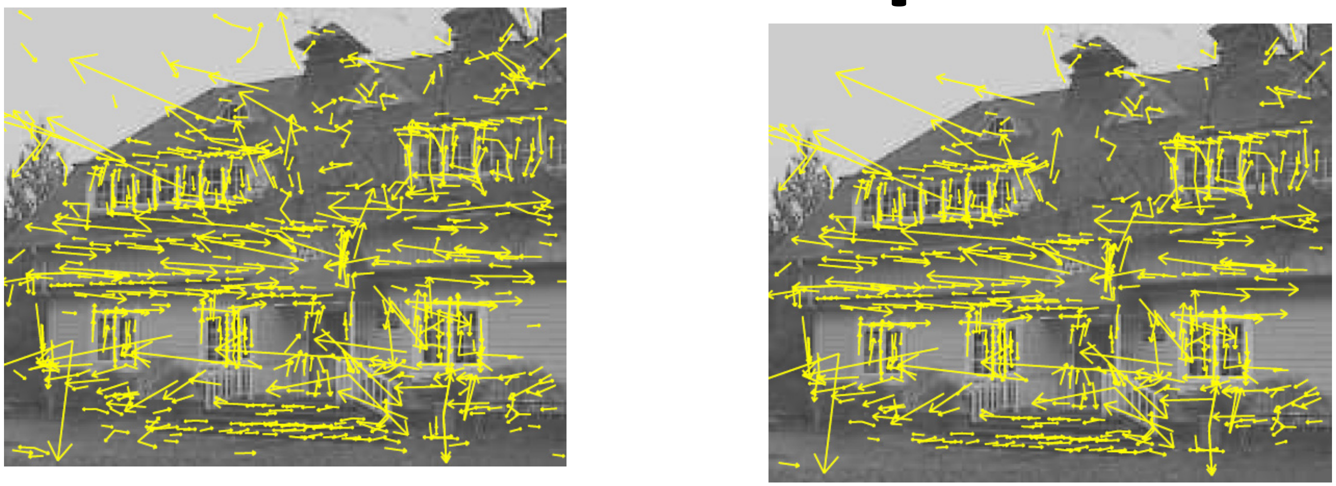
$$H = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}$$

Harris corner criterion

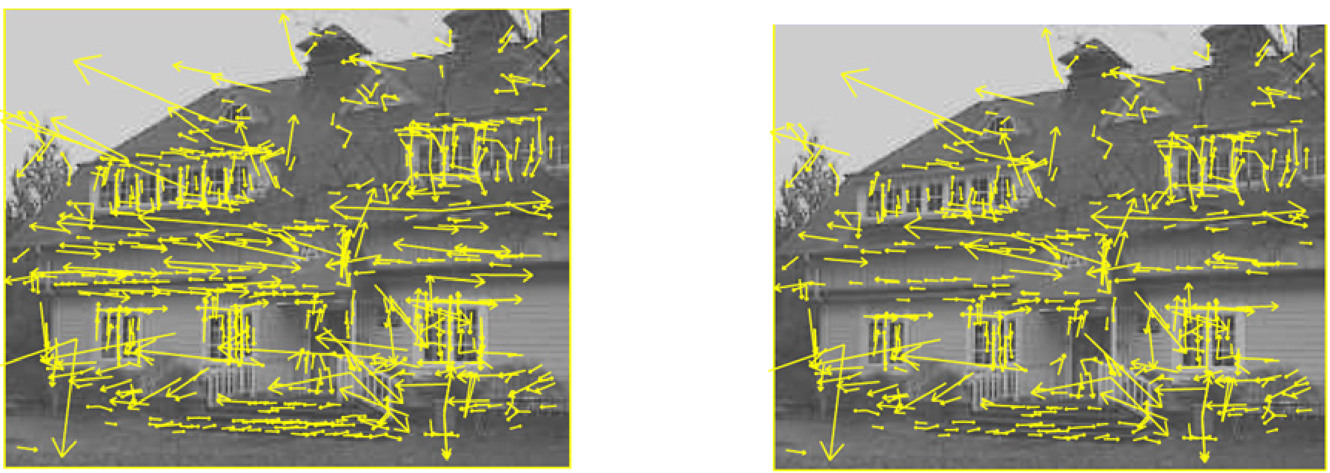
$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$

Discard points with response below threshold; Value of r = 10, is used;

SIFT : Step 2



729 out of 832 are left after contrast thresholding



536 out of 729 are left after cornerness thresholding

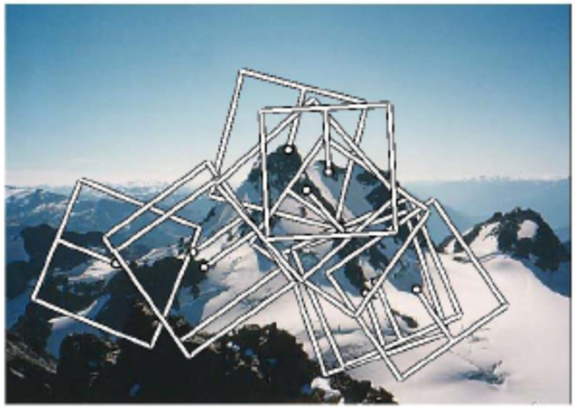
Slide credit: David Lowe

SIFT : Step 3

Step 3: Orientation Assignment

- Aim : Assign constant orientation to each keypoint based on local image property to obtain rotational invariance.

To transform relative data accordingly



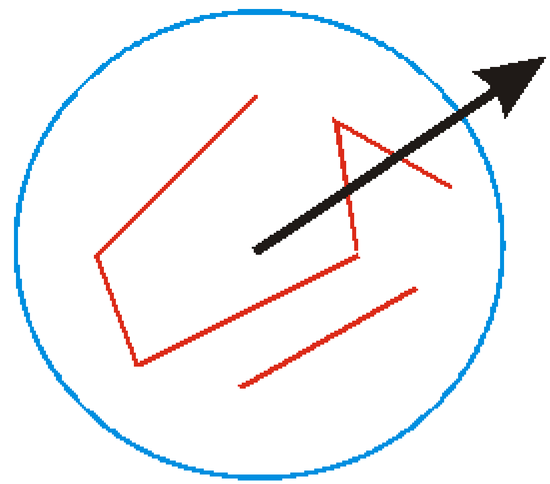
The magnitude and orientation of gradient of an image patch I(x,y) at a particular scale is:

$$\mathbf{m}(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{I}(\mathbf{x}+1,\mathbf{y})-\mathbf{I}(\mathbf{x}-1,\mathbf{y}))^2 + (\mathbf{I}(\mathbf{x},\mathbf{y}+1)-\mathbf{I}(\mathbf{x},\mathbf{y}-1))^2}$$

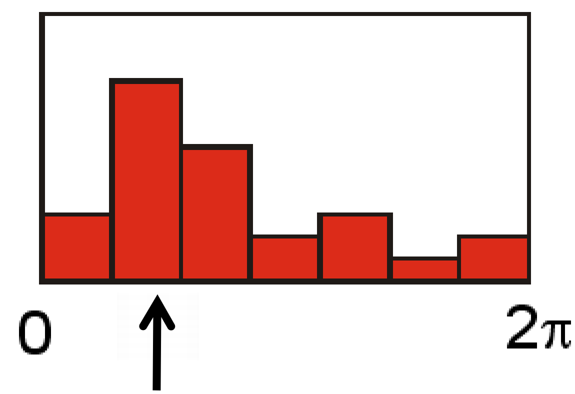
$$\theta(\mathbf{x},\mathbf{y}) = \tan^{-1}\frac{\mathbf{I}(\mathbf{x},\mathbf{y}+1)-\mathbf{I}(\mathbf{x},\mathbf{y}-1)}{\mathbf{I}(\mathbf{x}+1,\mathbf{y})-\mathbf{I}(\mathbf{x}-1,\mathbf{y})}$$

SIFT : Step 3

Step 3: Orientation Assignment



- Create weighted (magnitude + Gaussian) histogram of local gradient directions computed at selected scale



- Assign dominant orientation of the region as that of the peak of smoothed histogram
- For multiple peaks create multiple key points

Slide credit: David Lowe

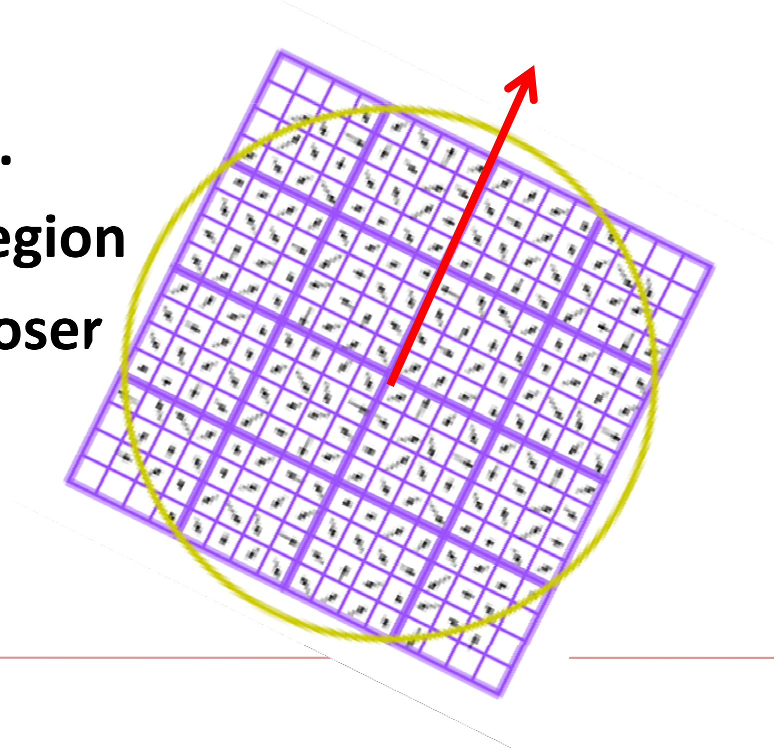
SIFT : Step 4

Already obtained precise location, scale and orientation to each keypoin

Step 4: Local image descriptor

Aim – Obtain local descriptor that is highly distinctive yet invariant to variation like illumination and affine change

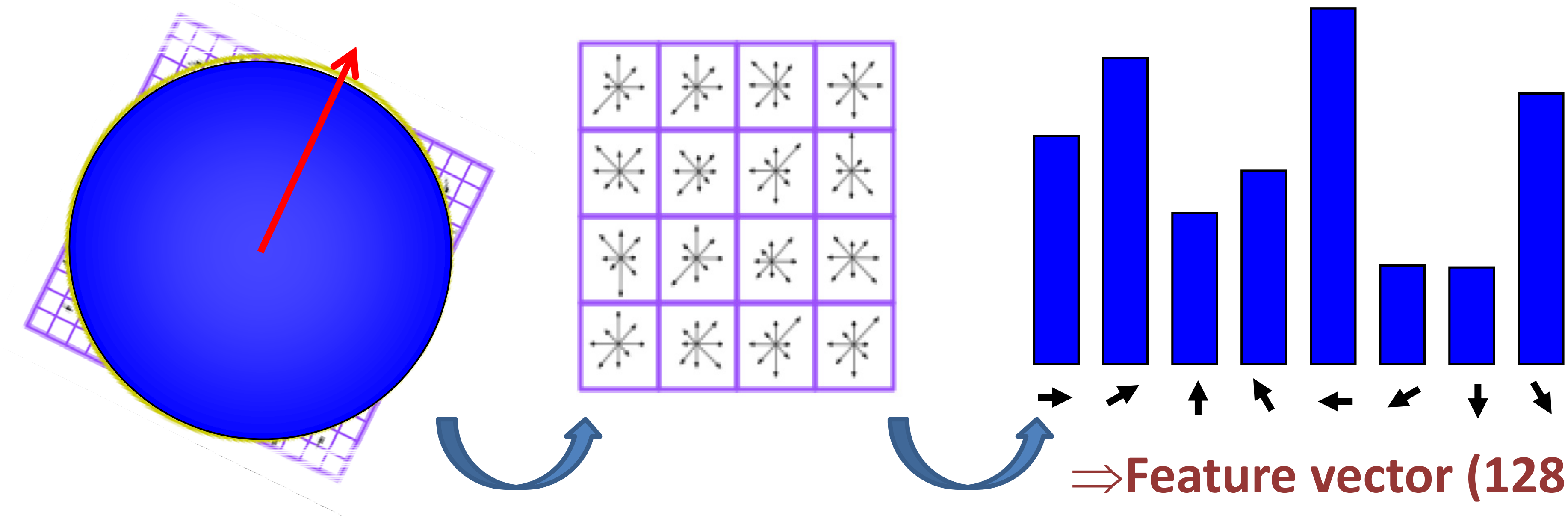
- Consider a rectangular grid 16*16 in the direction of the dominant orientation of the region.
- Divide the region into 4*4 sub-regions.
- Consider a Gaussian filter above the region which gives higher weights to pixel closer to the center of the descriptor.



SIFT : Step 4

Step 4: Local image descriptor

- Create 8 bin gradient histograms for each sub-region
  Weighted by magnitude and Gaussian window (σ is half the window size)



⇒Feature vector (128)
8*4*4 = 128

Finally, normalize 128 dim vector to make it illumination invariant

SIFT : Some Result

Object detection

SIFT : Some Result

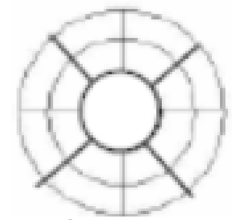
Panorama

GLOH

First 3 steps – same as SIFT

Step 4 – Local image descriptor

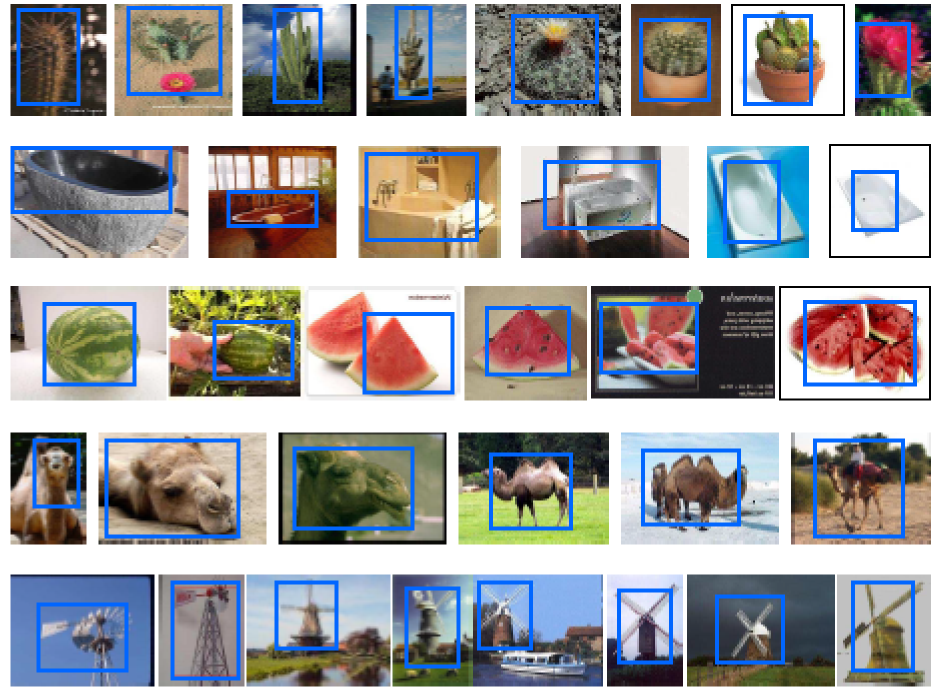
- Consider log-polar location grid with 3 different radii and 8 angular direction for two of them, in total 17 location bin
- Form histogram of gradients having 16 bins
- Form a feature vector of 272 dimension (17*16)
- Perform dimensionality reduction and project the features to a 128 dimensional space.



192 correct matches (yellow) and 208 false matches (blue).

Some other examples

SURF



PHOW

HOG

Other Feature descriptors - old and new:

- **LBP, LTP and variants, HAAR;**
- **PCA-SIFT, VLAD, MOSIFT,**
- **deep features, CNN, Fisher vector,**
- **SV-DSIFT, BF-DSIFT, LL-MO1SIFT, 1SIFT, VM1SIFT, VLADSIFT,**
- **DECAF, Fisher vector pyramid, IFV**
- **Dirichlet Histogram**
- **Simplex based STV (3-D), MSDR;**

BOV-W, Steak flow, tracklets, spatio-temporal gradients, LCS, LTDS, MRF, LDA, RFT, LCSS, MDA, DFM, Dynamic textures, BOAW, HFST, SRC based MHOF, LBPTOPS, HOP

Reference

1. Kristen Grauman and Bastian Leibe, Visual Object Recognition, Synthesis Lectures on Artificial Intelligence and Machine Learning, April 2011, Vol. 5, No. 2, Pages 1-181.
2. Beaudet, "Rotationally invariant image operators", in International Joint Conference on Pattern Recognition, pp. 579-583., 1978.
3. Förstner, W. and Gülch, E., "A fast operator for detection and precise location of distinct points, corners and centers of circular features", in ISPRS Inter commission Workshop', pp. 281-305, 1987.
4. Harris, C. and Stephens, M., "A combined corner and edge detector", *in* 'Alvey Vision Conference', pp. 147–151, 1988.
5. Lindeberg, T., 'Scale-space theory: A basic tool for analyzing structures at different scales', Journal of Applied Statistics 21(2), pp. 224–270, 1994.
6. **Lowe, D., 'Distinctive image features from scale-invariant keypoints', International Journal of Computer Vision 60(2), pp. 91–110, 2004.**
7. Mikolajczyk, K. and Schmid, C., 'A performance evaluation of local descriptors', IEEE Transactions on Pattern Analysis & Machine Intelligence 27(10), 31–37, 2005.

THANK YOU