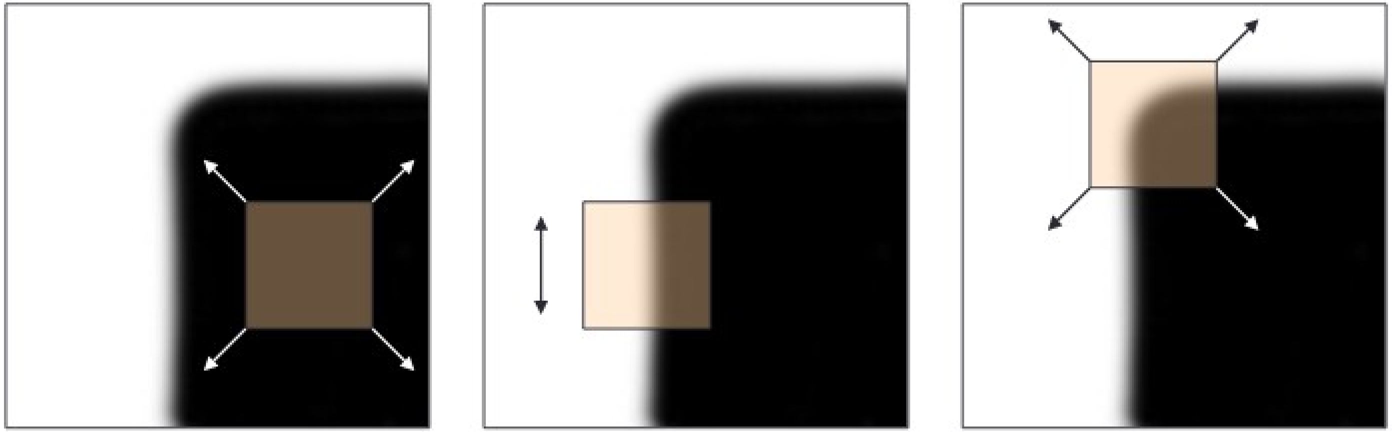
SIFT (Scale-Invariant Feature Transform)

Recap: Corner Detection: Basic Idea

- In the region around a corner, image gradient has two or more dominant directions.



Recap: Corner Detection

Change in appearance for the shift [u,v]

$$E(u,v) = \sum_{x,y} w(x,y)\left[I(x+u,y+v) - I(x,y)\right]^2$$

Window function

Shifted intensity

Intensity

We're looking for windows that produce a large E value.

Recap: Corner Detection

From taylor series we get,

$$E(u,v) \approx [u \;\; v] \begin{bmatrix} \sum_{x,y} w(x,y) I_x^2(x,y) & \sum_{x,y} w(x,y) I_x(x,y) I_y(x,y) \\ \sum_{x,y} w(x,y) I_x(x,y) I_y(x,y) & \sum_{x,y} w(x,y) I_y^2(x,y) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

Recap: Corner Detection

The quadratic expression simplifies to

$$\boxed{E(u,v) \approx [u \ \ v] \ M \begin{bmatrix} u \\ v \end{bmatrix}}$$

Where M is the second moment matrix, given computed by image derivatives

$$\boxed{M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}}$$

Recap: Corner Detection

Consider the axis aligned case where gradients are either horizontal or vertical

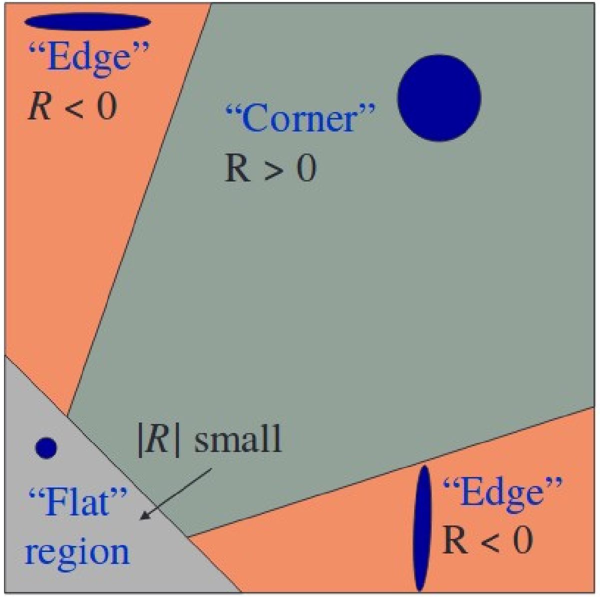
$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

If either λ is close to 0, then this is not a corner, so look for locations where both are large.

Harris corner response function

$$R = \det(M) - \alpha \, \mathrm{trace}(M)^2 = \lambda_1 \lambda_2 - \alpha(\lambda_1 + \lambda_2)^2$$

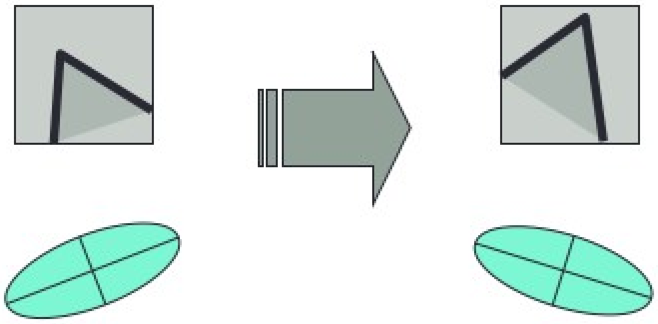
- R is large for a corner
- R is negative with large magnitude for an edge
- |R| is small for a flat region



Harris Detector Properties

- Rotational invariance

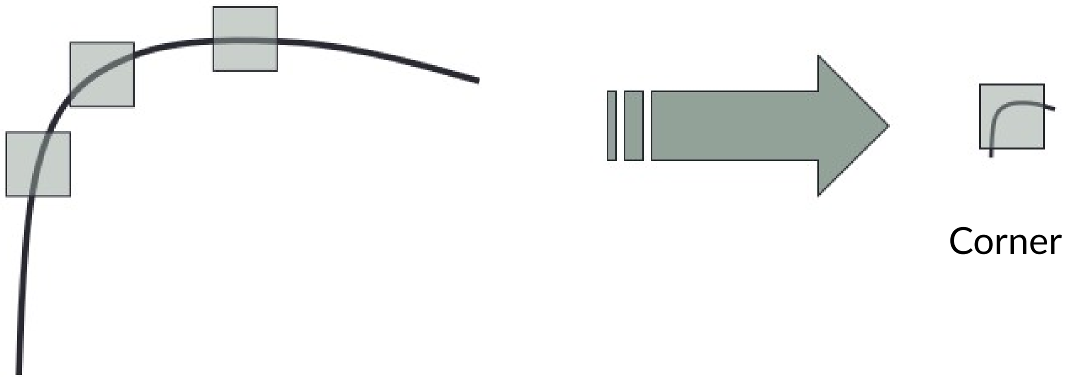
Ellipse rotates but its shape (i.e. eigenvalues) remains
the same.



Corner response R is invariant to image rotation

Harris Detector Properties

- Not invariant to image scale

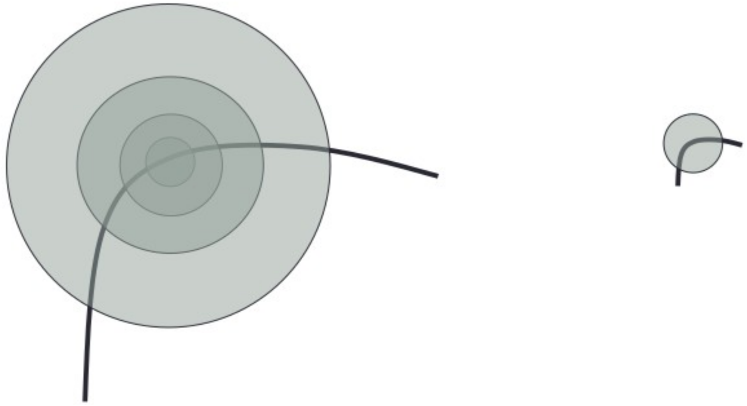


All points will be
classified as edges

Corner

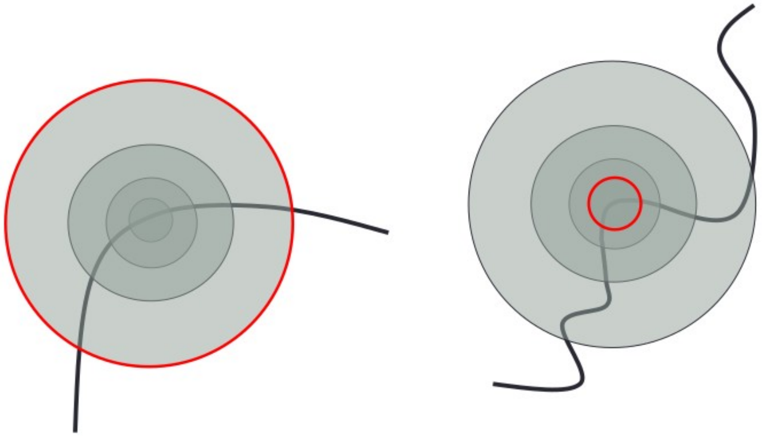
Scale Invariant Detection

- Consider regions (e.g. circles) of different sizes around a point
- Regions of corresponding sizes will look the same in both images



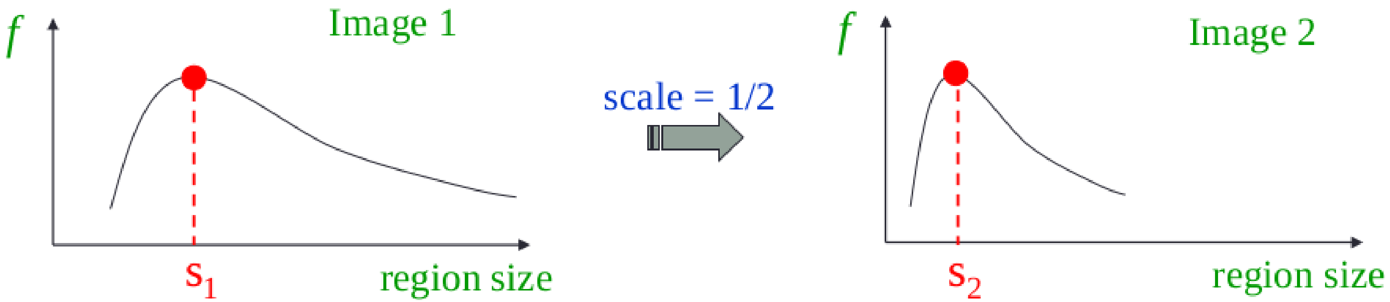
Scale Invariant Detection

- The problem: how do we choose corresponding circles independently in each image?



Scale Invariant Detection

- Solution:
  - Design a function on the region (circle), which is “scale invariant” (the same for corresponding regions, even if they are at different scales)
  - Take a local maximum of this function
  - Observation: region size, for which the maximum is achieved, should be invariant to image scale.



Goal

- Extracting distinctive invariant features
- Invariance to image scale and rotation
- Robust to
    - Distortion,
    - Change in 3D viewpoint,
    - Addition of noise,
    - Change in illumination.

SIFT (Scale Invariant Feature Transform)

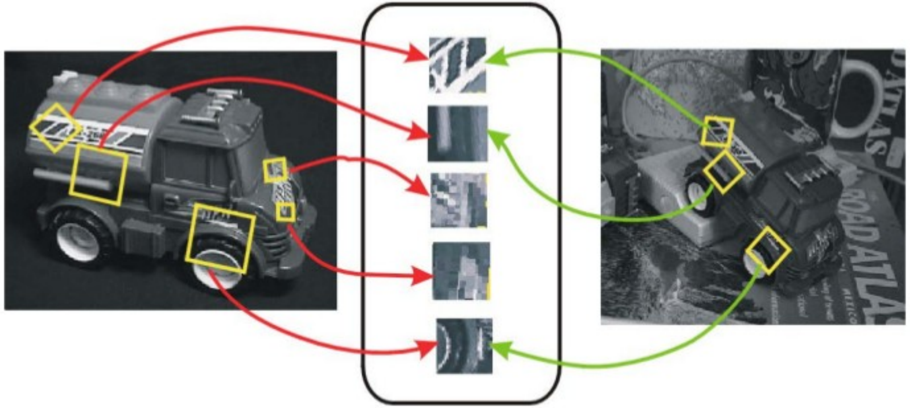
D.Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. IJCV 2004

Motivation

- The Harris operator is not invariant to scale
- For better image matching, Lowe's goal was to develop an interest operator – a detector – that is invariant to scale and rotation.
- Also, Lowe aimed to create a descriptor that was robust to the variations corresponding to typical viewing conditions. The descriptor is the most-used part of SIFT.

Idea of SIFT

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



SIFT Features

Overall Procedure of SIFT

- Scale-space extrema detection
  - Search over multiple scales and image locations
- Keypoint localization
  - Select keypoints based on a measure of stability.
- Orientation assignment
  - Compute best orientation(s) for each keypoint region.
- Keypoint description
  - Use local image gradients at selected scale and rotation to describe each keypoint region.

Scale-space extrema detection

- Find LoG for each image which is equivalent to find difference of gaussians(DoG) for two different blurred image (Computationally effective)

$$L = \sigma^2 \left(G_{xx}(x, y, \sigma) + G_{yy}(x, y, \sigma) \right)$$

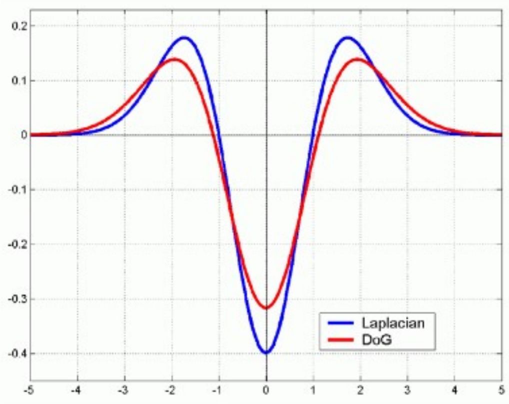
(Laplacian)

$$DoG = G(x, y, k\sigma) - G(x, y, \sigma)$$

(Difference of Gaussians)

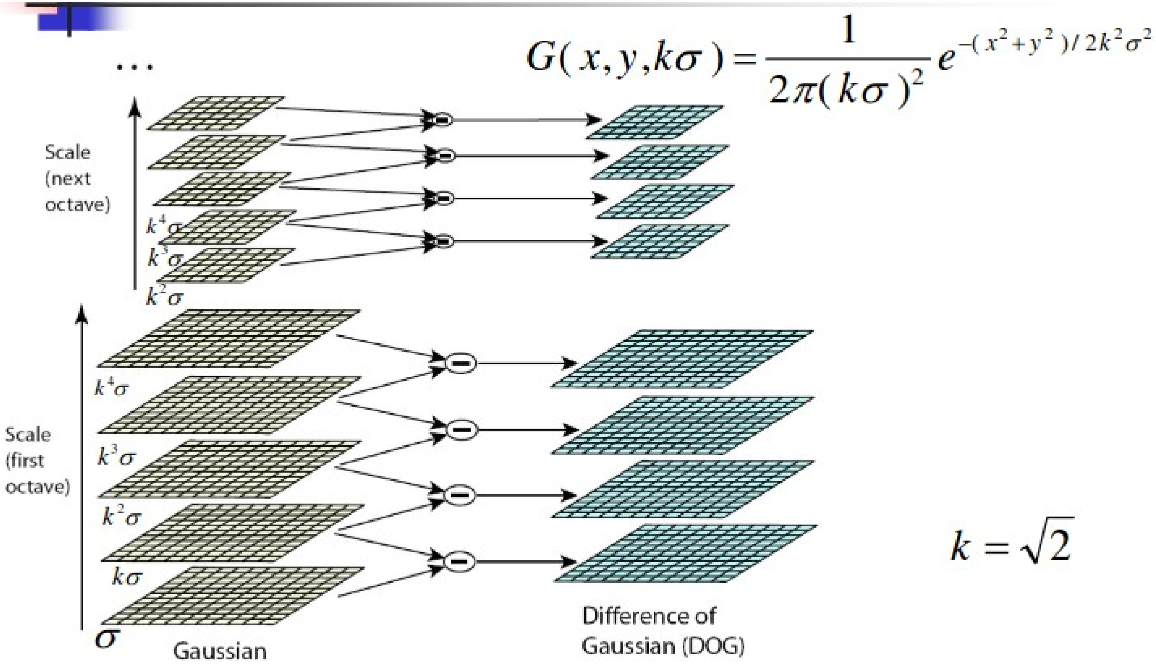
where Gaussian

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

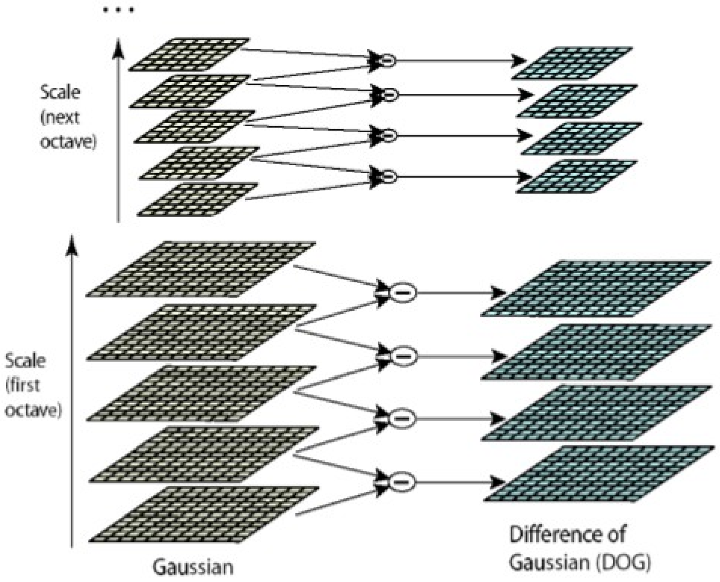


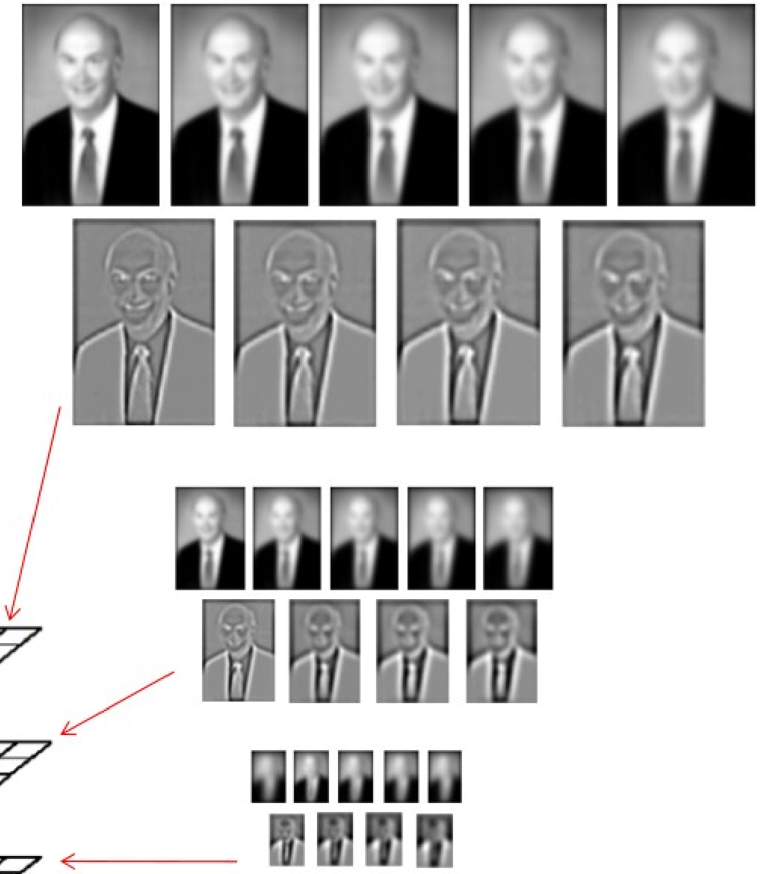
Note: both kernels are invariant to *scale* and *rotation*

Efficient DoG Computation using Gaussian Scale Pyramid

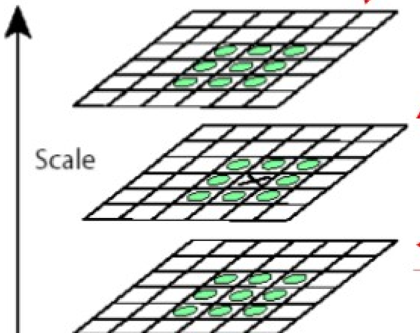


DOG detector : Flowchart



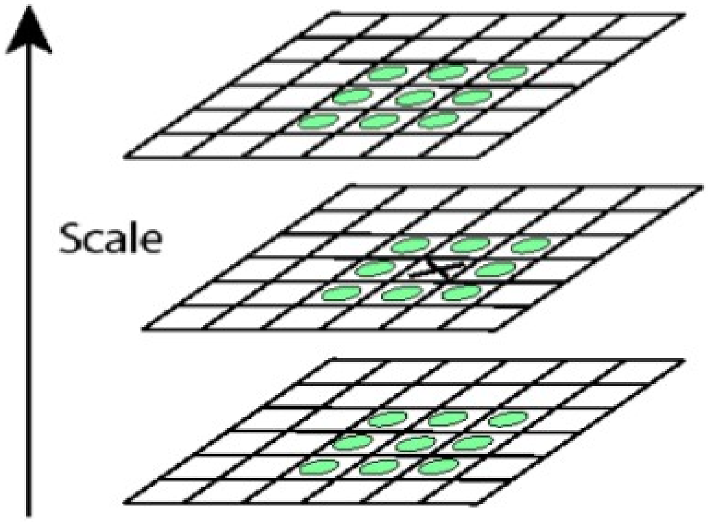


X is selected if it is larger or smaller than all neighbors.



Local Extrema in DoG Images

- Minima
- Maxima
- 26 neighbourhood



Key Point Localization

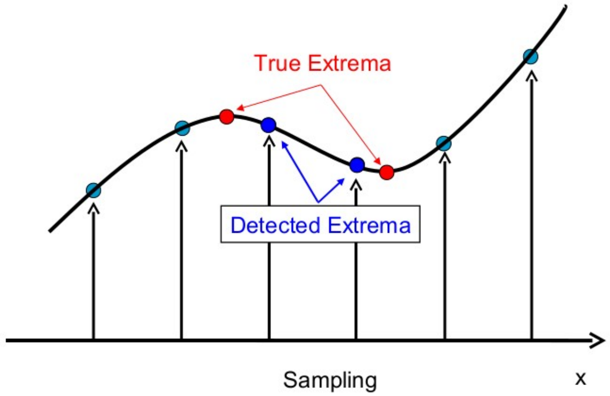
- Candidates are chosen from extrema detection

original image

extrema locations

Initial Outlier Rejection

- Low contrast points
- Poorly localized candidates along an edge
- Taylor series expansion of DOG, D



Initial Outlier Rejection

$$D(X) = D(0) + \frac{\partial D(0)}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D(0)}{\partial X^2} X$$

Consider first order approximate

$$D(X) = D(0) + \frac{\partial D(0)}{\partial X} X$$

$$\frac{\partial D(X)}{\partial X} = \frac{\partial^2 D(0)}{\partial X^2} X + \frac{\partial D(0)}{\partial X}$$

To maximize D(X) set $\frac{\partial D(X)}{\partial X} = 0$

$$0 = \frac{\partial^2 D(0)}{\partial X^2} X + \frac{\partial D(0)}{\partial X}$$

=> Minima or maxima is located at $X = -\left(\frac{\partial^2 D(0)}{\partial X^2}\right)^{-1} \frac{\partial D(0)}{\partial X}$

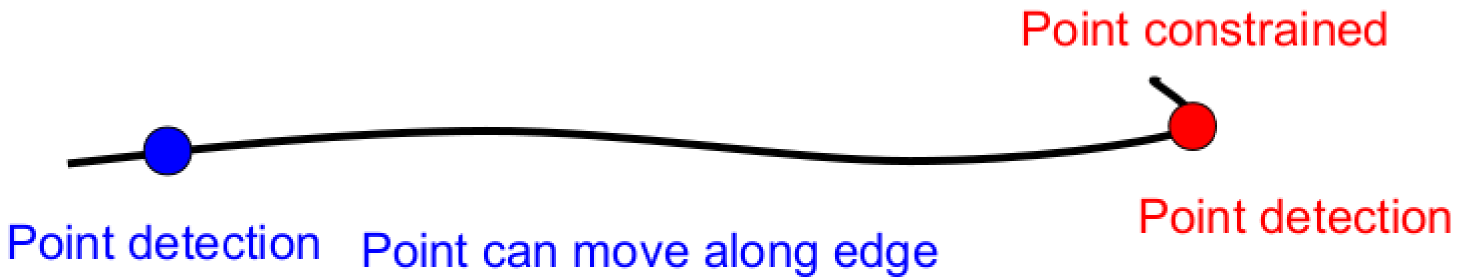
=> Value of `D(x)`at minima/maxima must be large, |`D(x)`|>`th`

Initial Outlier Rejection

from 832 key points to 729 key points, th=0.03.

Further Outlier Rejection

- Reject points with strong edge response in one direction only
- Use Harris - using Trace and Determinant of Hessian



Further Outlier Rejection

from 729 key points to 536 key points.

Orientation Assignment

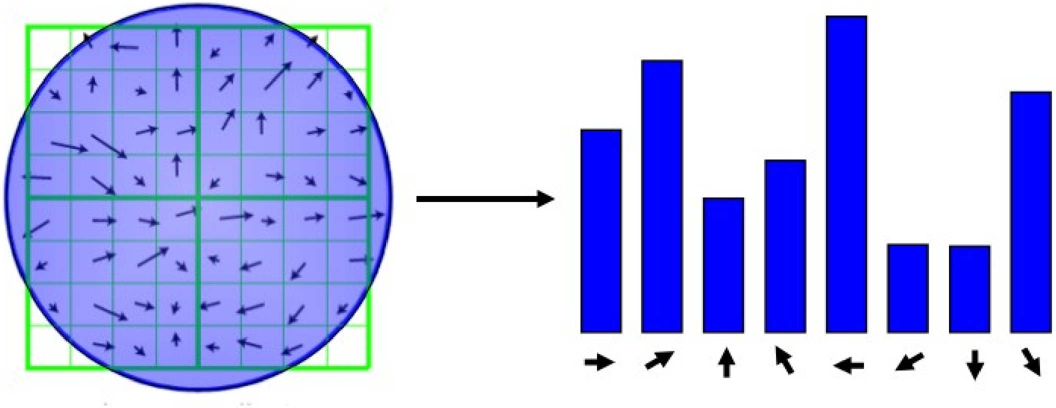
- Aim : Assign constant orientation to each keypoint based on local image property to obtain rotational invariance.
- Compute gradient magnitudes and orientations of L (Smooth image) at the scale of key point (x,y)

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

$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$

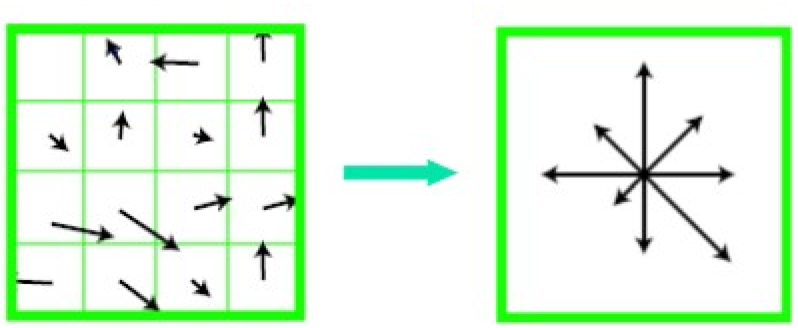
Orientation Assignment

- Create a weighted direction histogram in a neighborhood of a key point (36 bins)
  - Weighted by magnitude and Gaussian window (that of the scale of a keypoint)



Orientation Assignment

- Select the peak as direction of the key point
- Introduce additional key points (same location) at local peaks (within 80% of max peak) of the histogram with different directions



Keypoint Descriptors

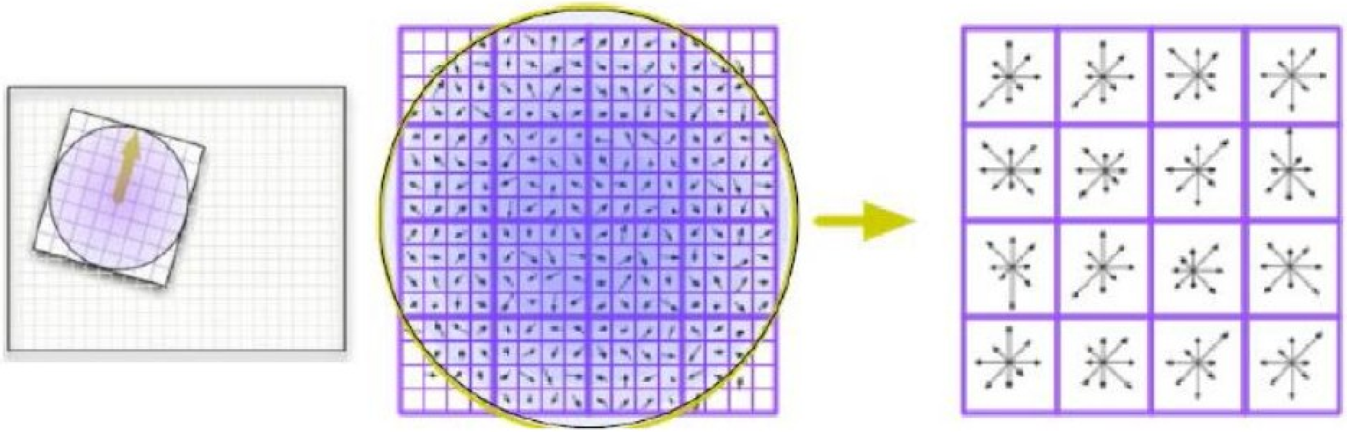
- At this point, each keypoint has
  - Location
  - Scale
  - Orientation
- Next is to compute a descriptor for the local image region about each keypoint that is
  - highly distinctive
  - invariant as possible to variations such as changes in viewpoint and illumination

Keypoint Descriptor

- Rotate the window to standard orientation
- Scale the window size based on the scale at which the point was found.
- Compute relative orientation and magnitude in a 16x16 neighborhood at key point
- Form weighted histogram (8 bin) for 4x4 regions
  - Weight by magnitude and Gaussian
  - Concatenate 16 histograms in one long vector of 128 dimensions

SIFT vector formation

- 4x4 array of gradient orientation histograms over 4x4 pixels
- 8 orientations x 4x4 array = 128 dimensions
- 128-dim vector normalized to unit length to reduce the effect of illumination



Key point matching

- Match the key points against a database of that obtained from training images.
- Find the nearest neighbor i.e. a key point with minimum Euclidean distance
  - Efficient Nearest Neighbor matching
    - Looks at ratio of distance between best and 2nd best match (.8)

Thank you!