Scale Invariant Feature Transform (SIFT)

David Lowe, University of British Columbia

Aim

- Extracting distinctive invariant features
  - Correctly matched against a large database of features from many images
- Invariance to image scale and rotation
- Robustness to
  - Affine distortion
  - Change in 3D viewpoint
  - Addition of noise
  - Change in illumination

Advantages

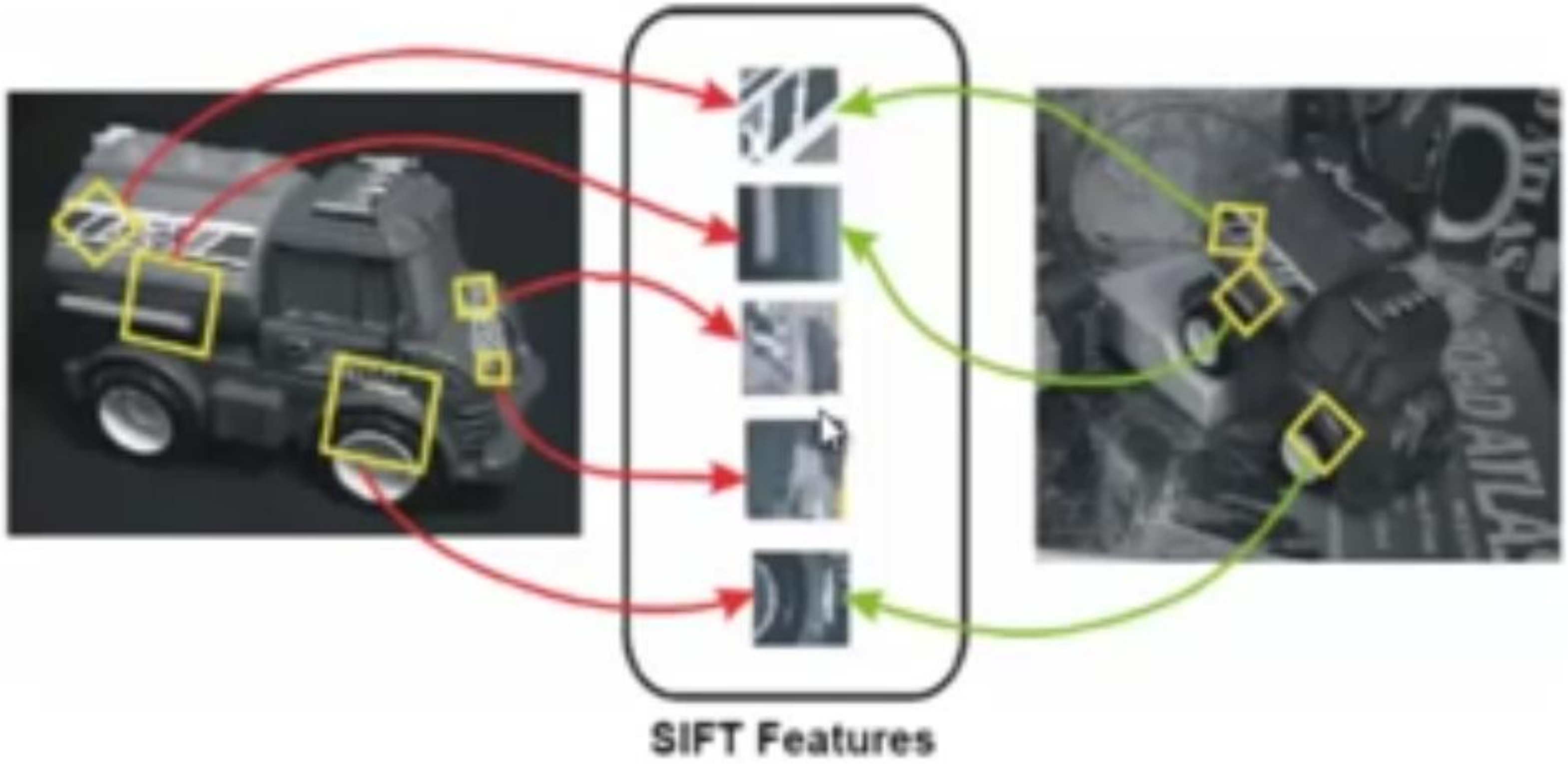
Locality: features are local, so robust to occlusion and clutter

Distinctiveness: individual features can be matched to a large database of objects

Quantity: many features can be generated for even small objects

Efficiency: close to real-time performance

Invariant Local features



Steps for Extracting Key Points

Scale space peak selection: Potential locations for finding features

Key point localization: Accurately locating the feature key points

Orientation Assignment: Assigning orientation to the key points

Key point descriptor: Describing the key point as a high dimensional vector

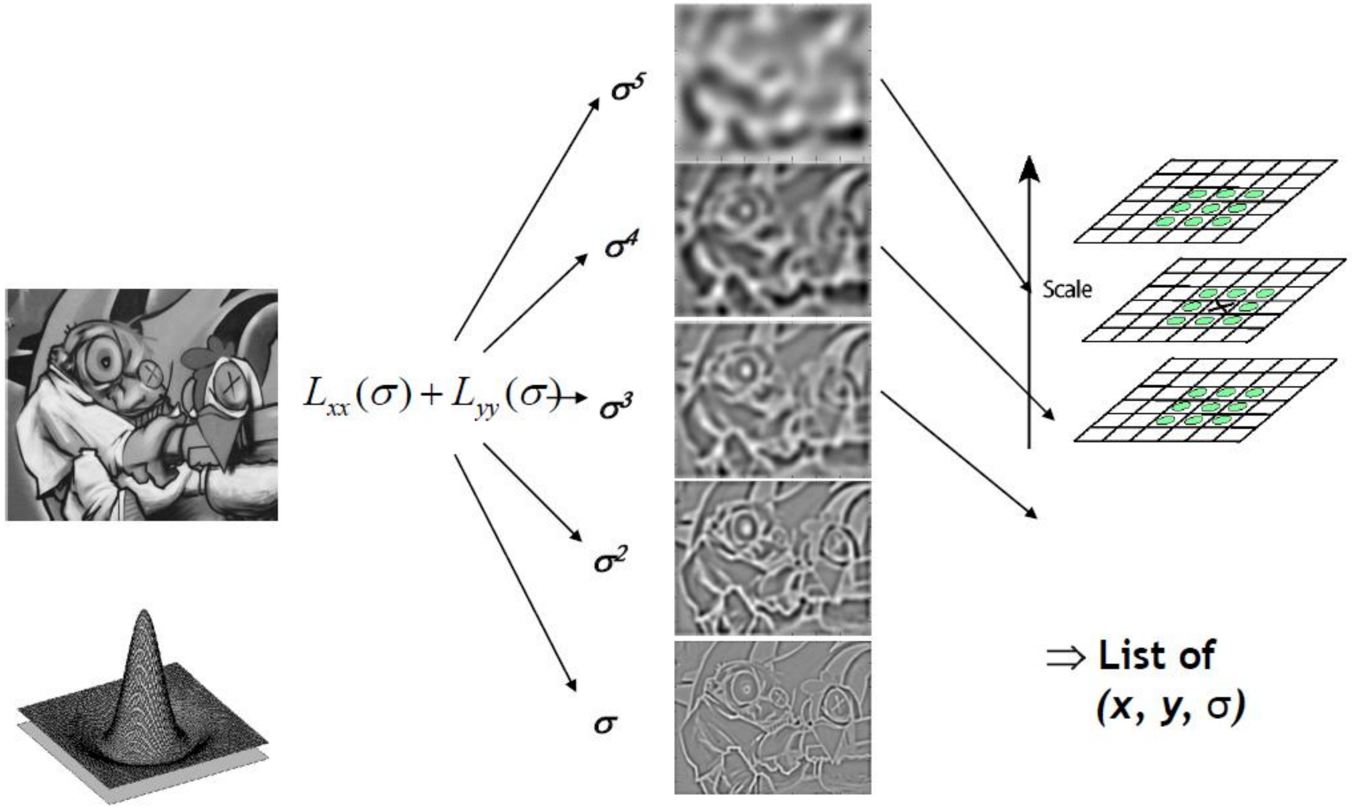
Scale

What should be sigma value for Canny and LoG edge detection?

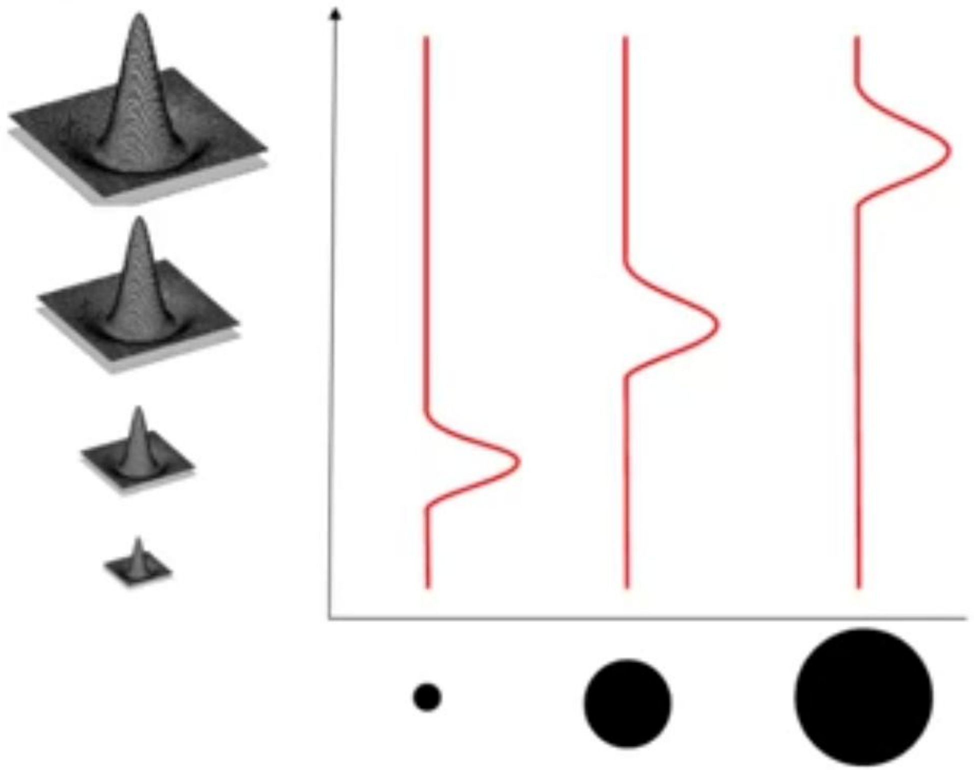
If use multiple sigma values (scales), how do you combine multiple edge maps?

Laplacian of Gaussian (LoG)

Interest points: Local maxima in scale space of Laplacian-of-Gaussian



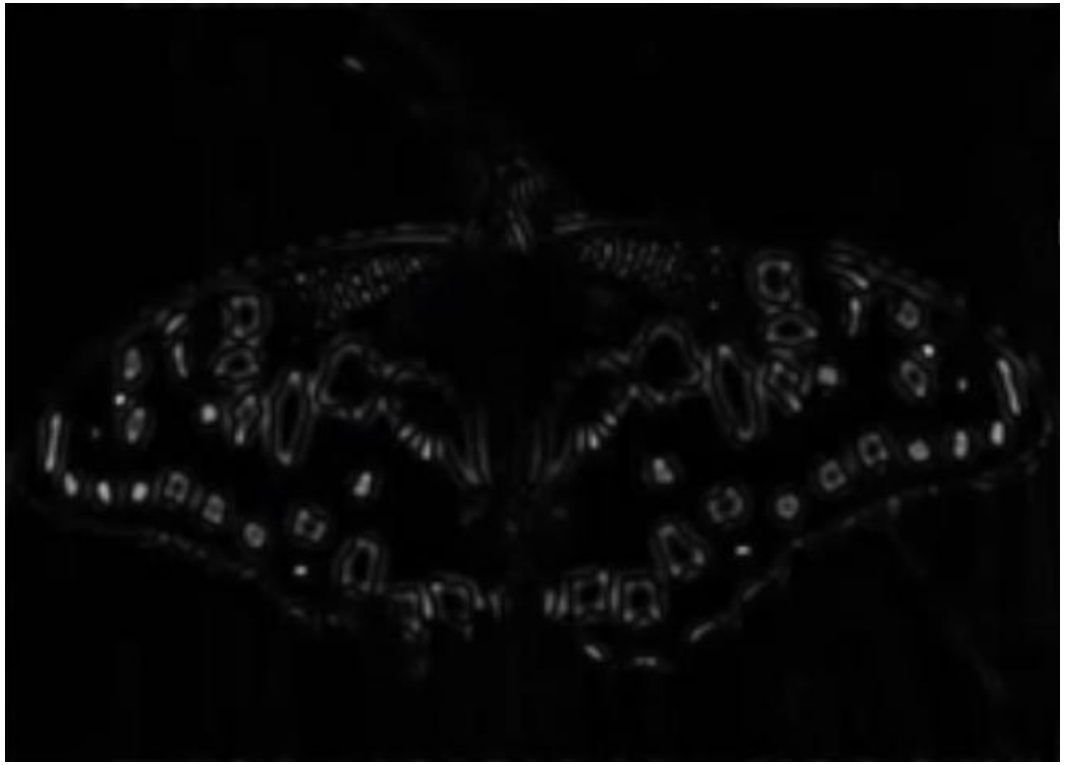
Laplacian of Gaussian (LoG)



Scale Space

Scale Space: Sigma value 2

Scale Space: Sigma value ~2.5



Scale Space: Sigma value ~3

Scale Space: Sigma value ~4

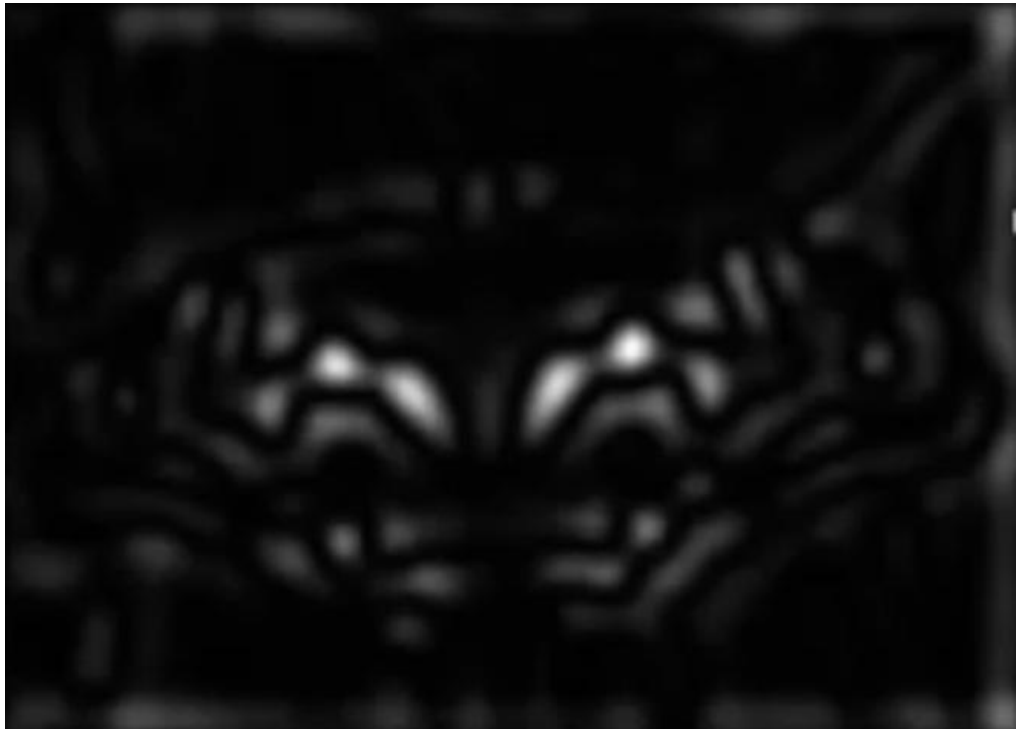
Scale Space: Sigma value ~5

Scale Space: Sigma value ~6

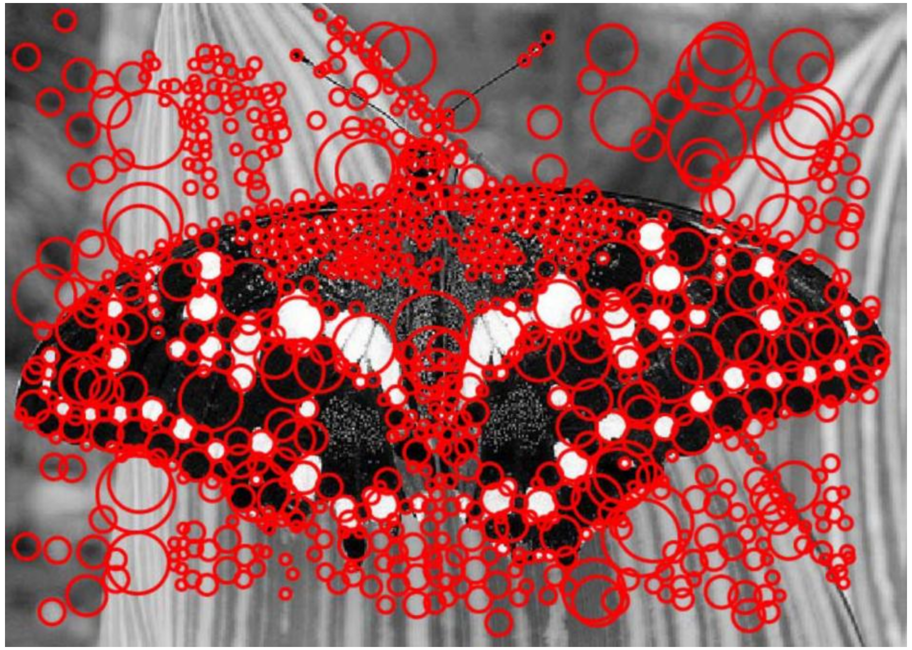
Scale Space: Sigma value ~7.5

Scale Space: Sigma value ~9.5

Scale Space: Sigma value ~11



Scale Space



Building a Scale Space

- All scales must be examined to identify scale-invariant features
- An efficient function is to compute the Laplacian Pyramid (Difference of Gaussian) (Burt & Adelson, 1983)

LoG and DoG

$$\frac{\partial G}{\partial \sigma} = \sigma \Delta^2 G$$

Heat Equation

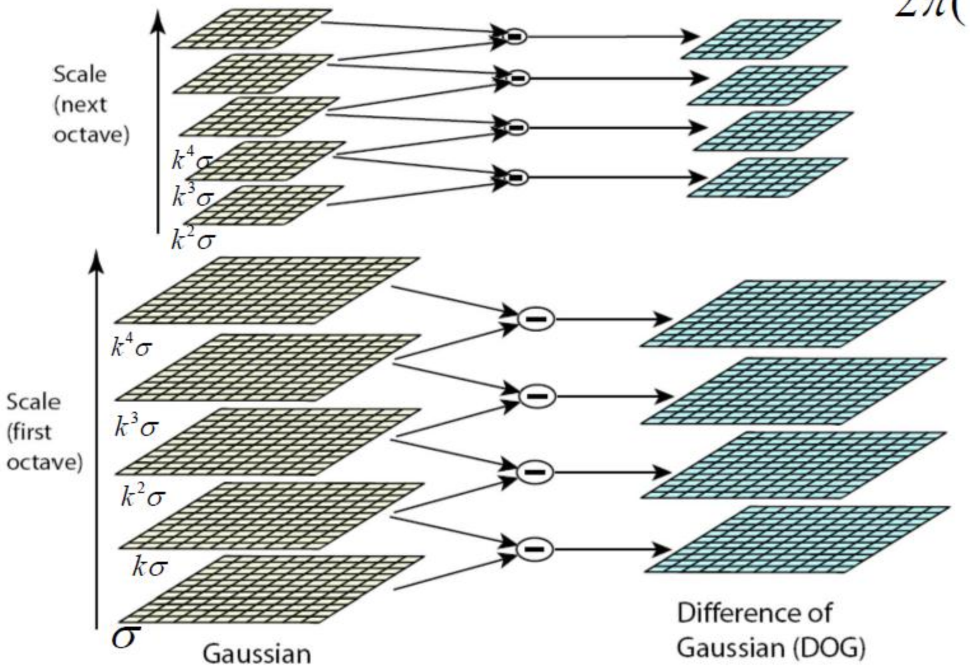
$$\sigma \Delta^2 G = \frac{\partial G}{\partial \sigma} = \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

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

$$\text{Typical values}: \sigma = 1.6;\ k = \sqrt{2}$$

Building a Scale Space

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



$$k = \sqrt{2}$$

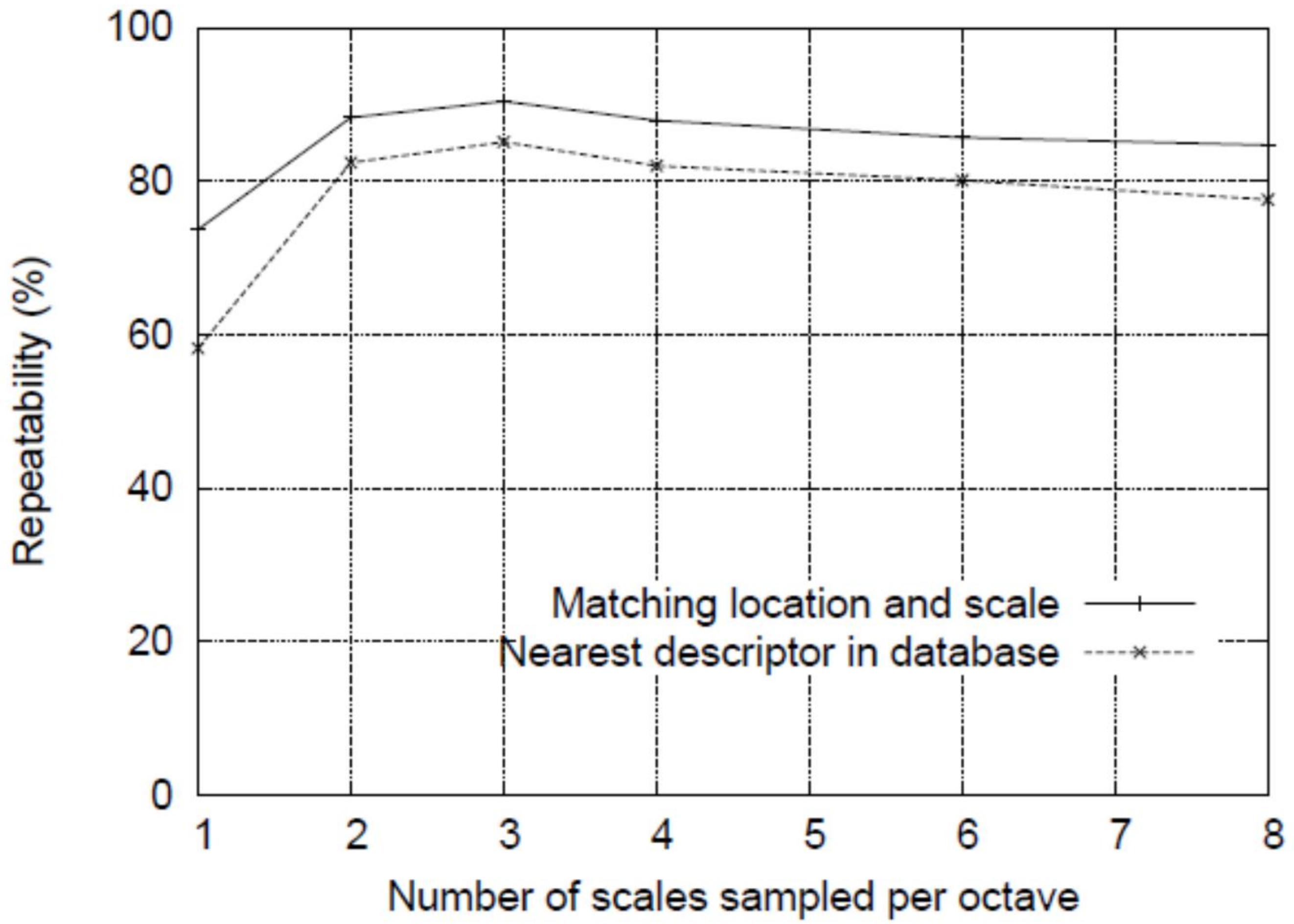
Scale & Octave

scale →

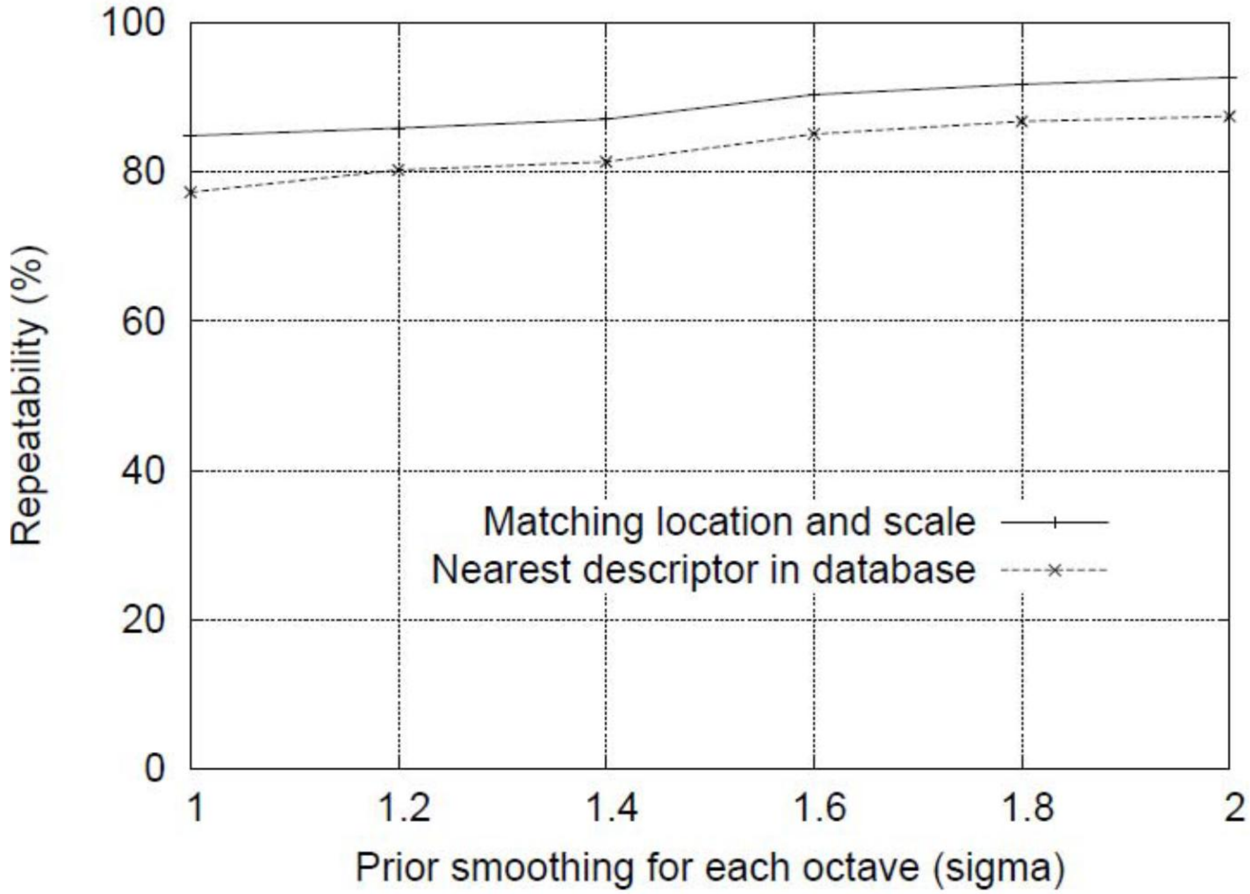
octave

0.707107	1.000000	1.414214	2.000000	2.828427
1.414214	2.000000	2.828427	4.000000	5.656854
2.828427	4.000000	5.656854	8.000000	11.313708
5.656854	8.000000	11.313708	16.000000	22.627417

How many Scale per Octave?

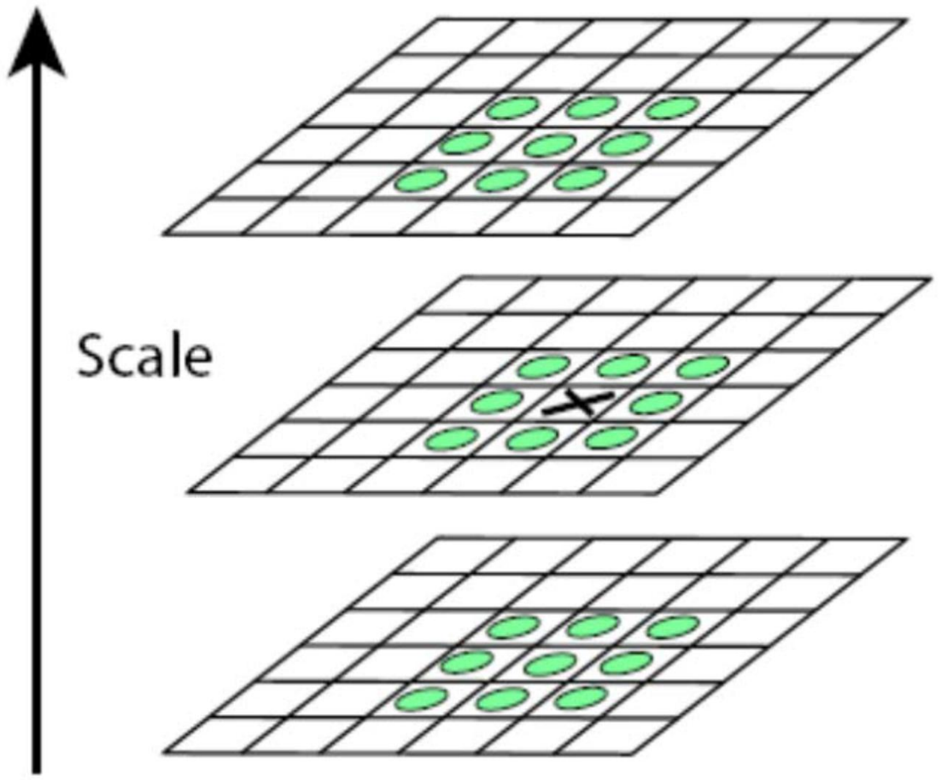


Initial value of Sigma?



Scale Space Peak Detection

- Compare a pixel (X) with 26 pixels in current and adjacent scales (Green Circles)
- Select a pixel (X) if larger/smaller than all 26 pixels
- Large number of extrema, computationally expensive
  - Detect the most stable subset with a coarse sampling of scales



Key Point Localization

Initial Outlier Rejection

Low contrast candidates
Poorly localized candidates along an edge
Taylor series expansion of DoG, **D**

$$\boxed{D(\mathbf{x}) = D + \frac{\partial D}{\partial \mathbf{x}}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}} \quad \mathbf{x} = (x, y, \sigma)^T$$

Extrema is located at

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

Value of D(x) at extrema must be large, |D(x)|>Th

Initial Outlier Rejection

832 interest points to 729 interest points

Th = 0.03

Further Outlier Rejection

DoG has strong response along edge

Assume DoG as a surface

- Compute principle curvature (PC)
- Along the edge, one of the PC is very low, across the edge high

Further Outlier Rejection

Compute Hessian of D

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

$$Tr(H) = D_{xx} + D_{yy} = \lambda_1 + \lambda_2$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \lambda_1 \lambda_2$$

Remove outliers by evaluating

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

$$r = \frac{\lambda_1}{\lambda_2}$$

Further Outlier Rejection

Eliminate interest point if r > 10

$$\frac{Tr(H)^2}{Det(H)} = \frac{(r+1)^2}{r} \qquad r = \frac{\lambda_1}{\lambda_2}$$

Further Outlier Rejection

729 interest points to 536 interest points

Orientation Assignment

To achieve rotation invariance

Compute central derivatives, gradient magnitude and direction of L at the scale of interest point (x,y)

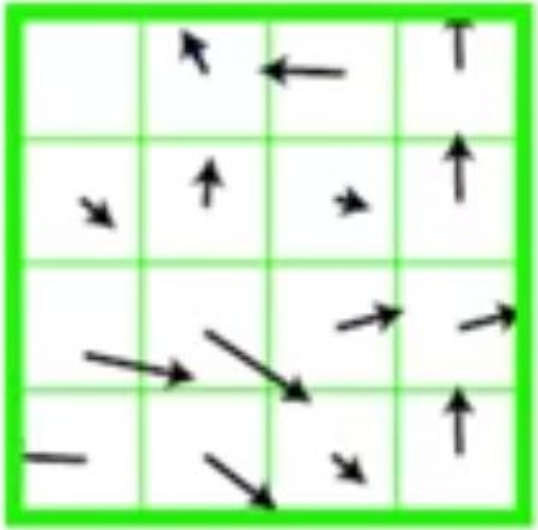
$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

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

Orientation Assignment

Create a weighted direction histogram in a neighbourhood of an interest point

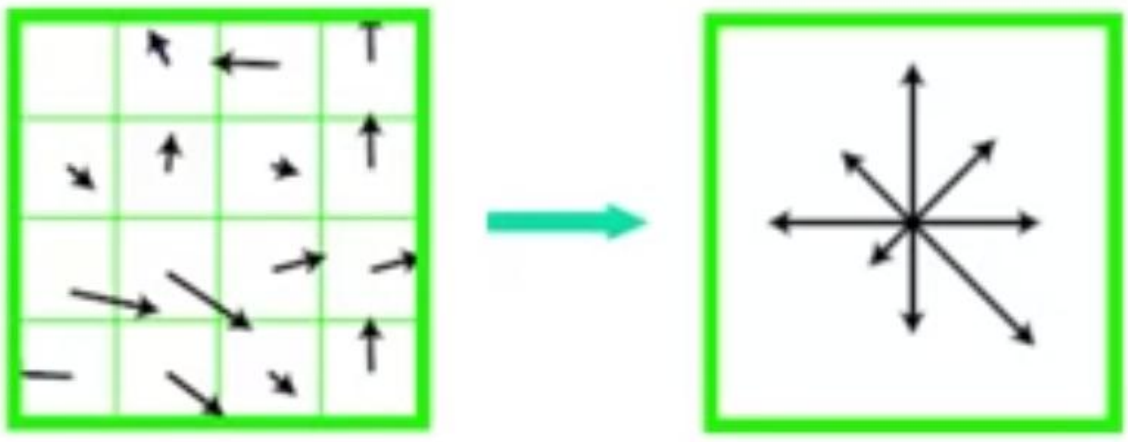
Weights are Gradient magnitude and spatial Gaussian



Orientation Assignment

Select the peak as direction of the interest point

Introduce additional interest points (same location) at local peaks (within 80% of the max peak) of the histogram with different directions



SIFT Descriptor

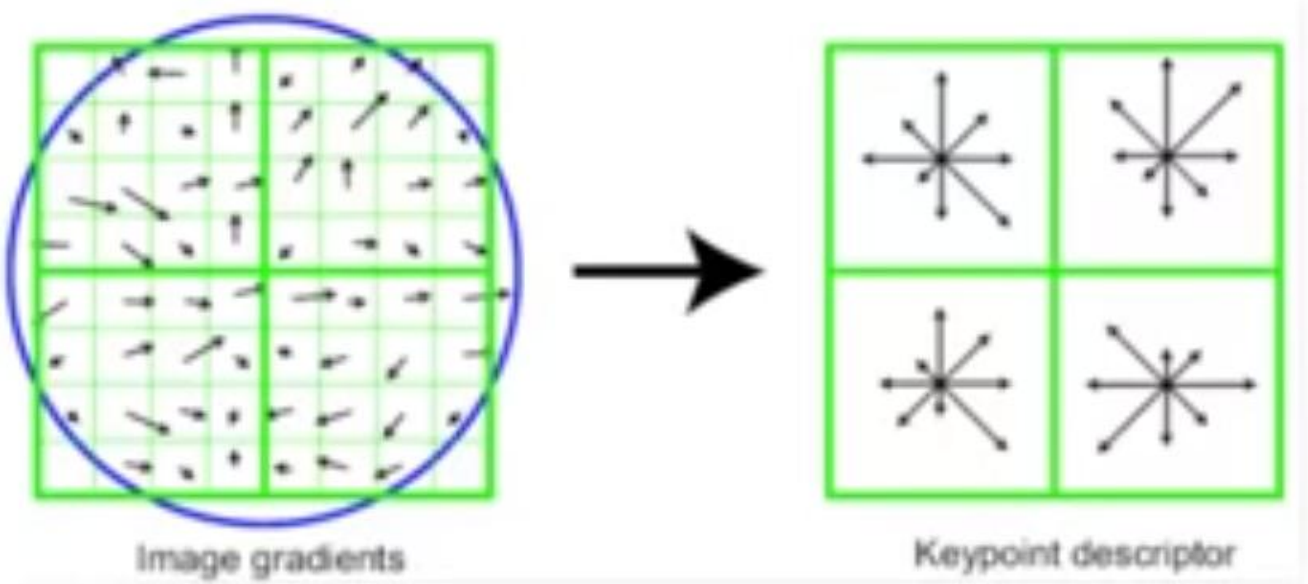
Possible descriptor

- Store intensity sample in the neighbourhood
- Sensitive to lighting changes, 3D transformations

Gradient Orientation histogram

- Robust representation

SIFT Descriptor

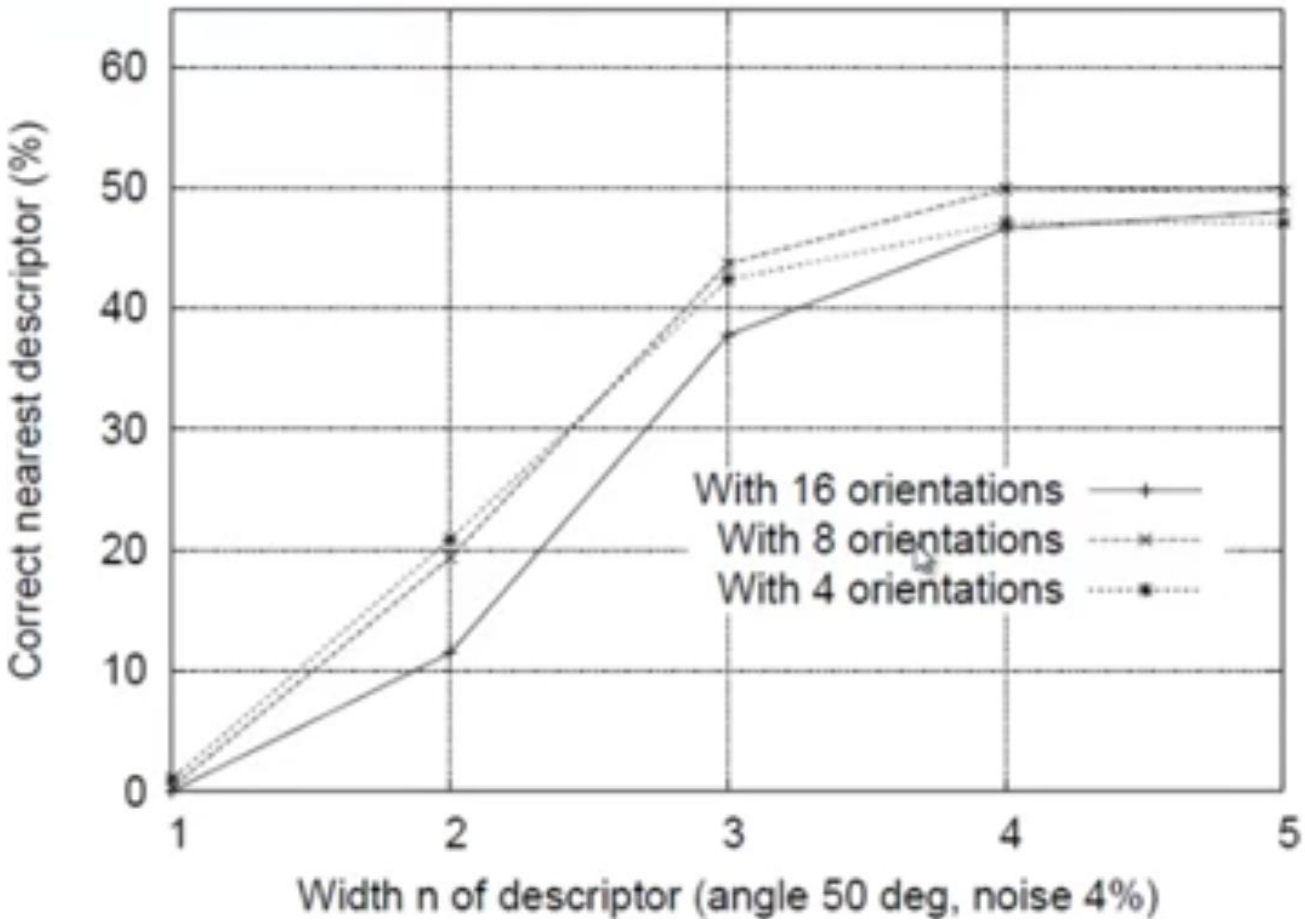


Compute relative orientation and magnitude in a 16*16 neighbourhood at interest point

Form weighted histogram (8 bin) for 4*4 regions

- Weight by magnitude and spatial Gaussian
- Concatenate 16 histograms in one long vector of 128 dimensions

SIFT Descriptor regions



SIFT Descriptor

Store numbers in a vector

Normalize to unit vector

Bound unit vector items to maximum 0.2 and renormalize to unit vector

Matching

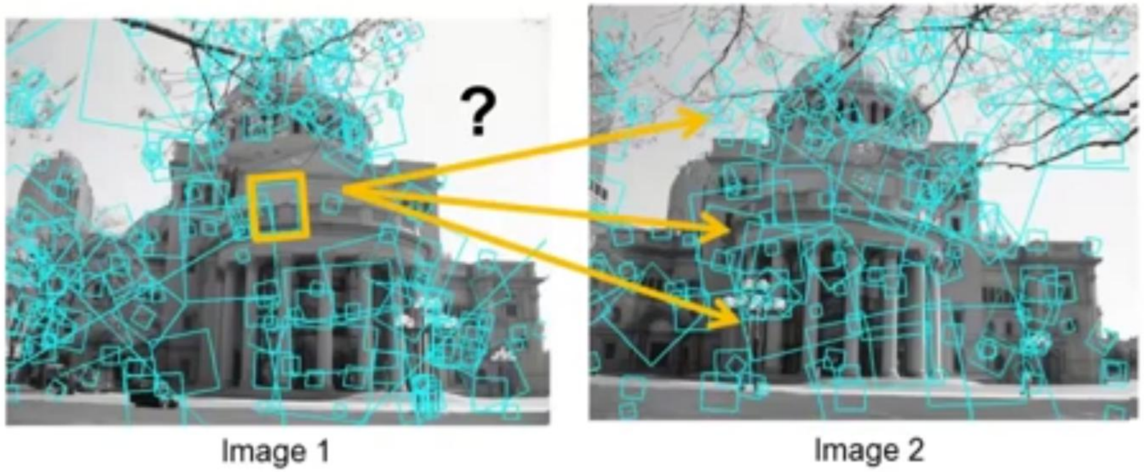
Find nearest neighbour i.e. an interest point with minimum Euclidean distance

Look at ration of distance between best and 2nd best match (0.8)

Matching

Candidate match using SIFT descriptor

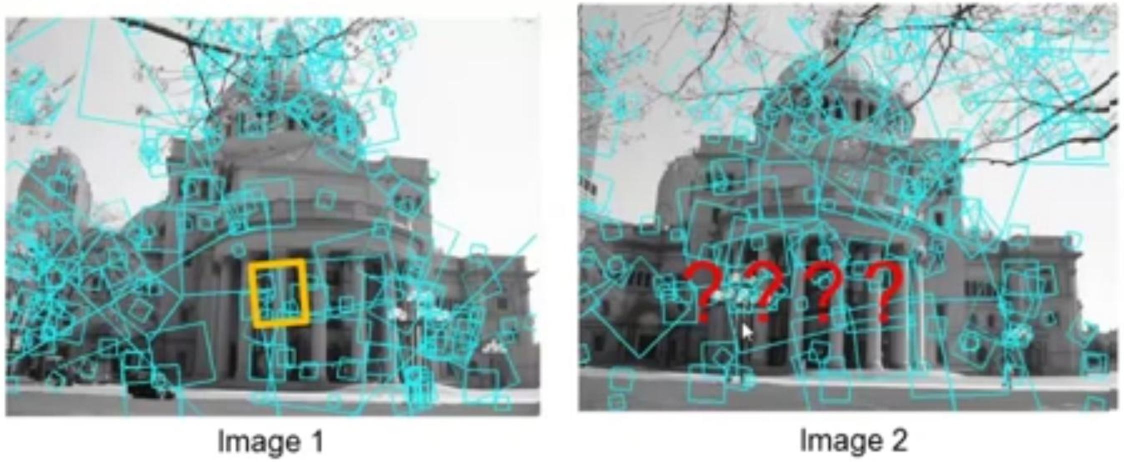
Compare them all and pick closest (or closest k, or within a threshold distance)



Matching

For robustness consider ratio of distance to best match and second-best match

If ratio is low, then select first match otherwise ambiguous match



Ratio of distance to best match and second-best match