

CAP 5415 Computer Vision Fall 2012

Dr. Mubarak Shah

Univ. of Central Florida

Office 247-F HEC

Lecture-5





SIFT: David Lowe, UBC





SIFT - Key Point Extraction

- Stands for scale invariant feature transform
- Patented by university of British Columbia
- Similar to the one used in primate visual system (human, ape, monkey, etc.)
- Transforms image data into scale-invariant coordinates



Goal

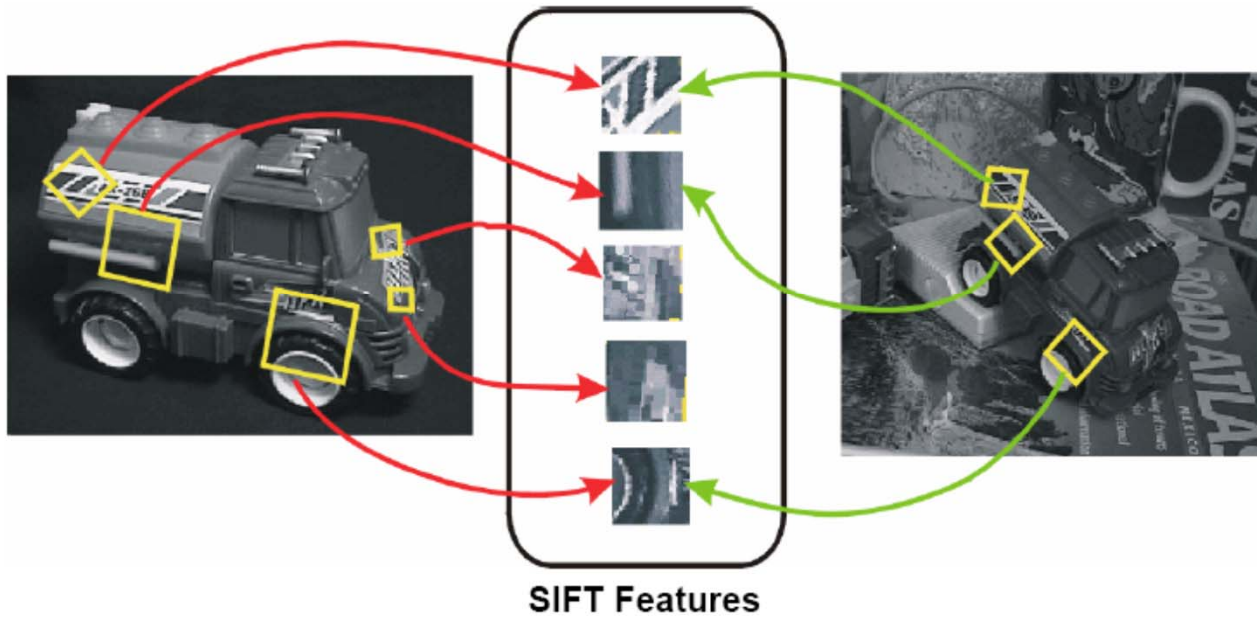
- Extracting distinctive invariant features
 - Correctly matched against a large database of features from many images
- Invariance to image scale and rotation
- Robustness to
 - Affine distortion,
 - Change in 3D viewpoint,
 - Addition of noise,
 - Change in illumination.



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



Scales

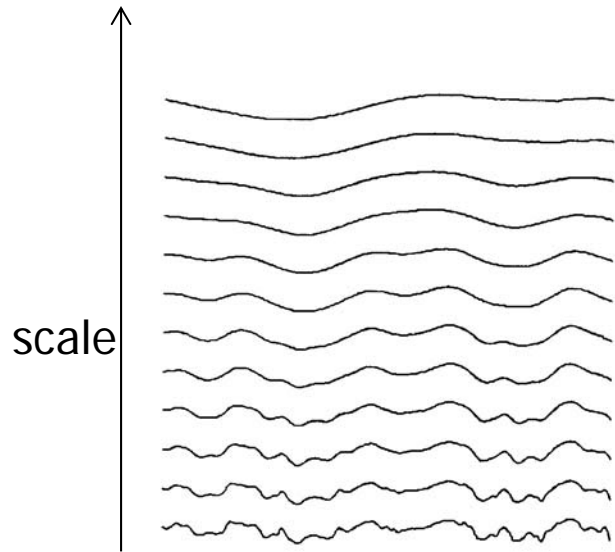
- What should be sigma value for Canny and LG edge detection?
- If use multiple sigma values (scales), how do you combine multiple edge maps?
- Marr-Hildreth:
 - *Spatial Coincidence* assumption:
 - Zerocrossings that coincide over several scales are physically significant.

Scale Space (Witkin, IJCAI 1983)

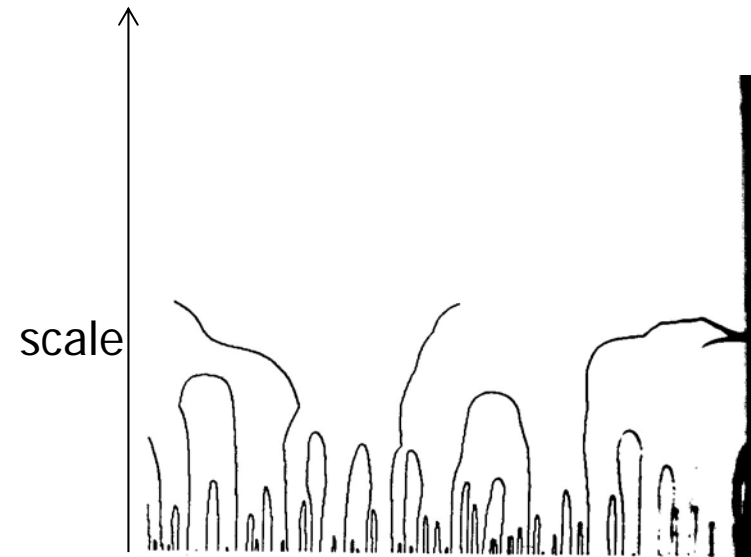
- Apply whole spectrum of scales
- Plot zerocrossings vs scales in a scale-space



Scale Space



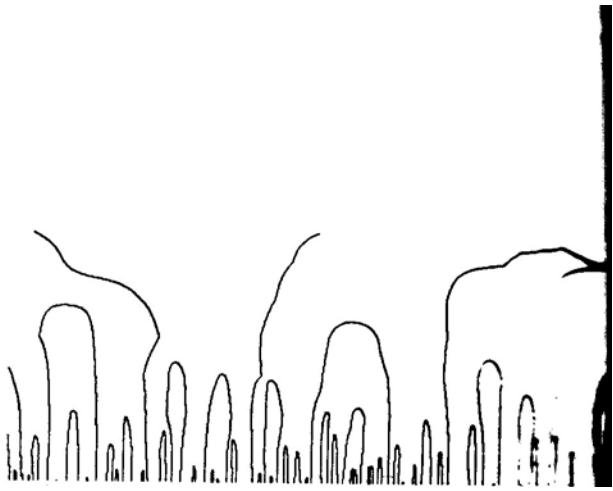
Multiple smooth versions of a signal



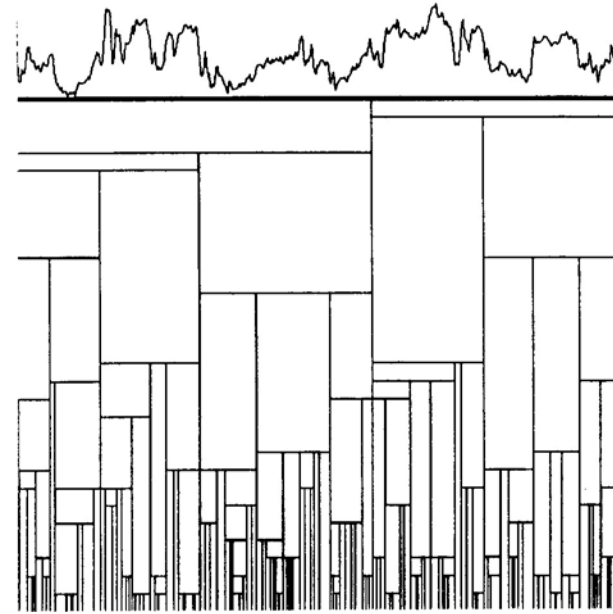
Zerocrossings at multiple scale



Scale Space



Scale Space



Interval Tree

Scale Space (Witkin, IJCAI 1983)



- Apply whole spectrum of scales
- Plot zerocrossings vs scales in a scale-space
- Interpret scale space contours
 - Contours are arches, open at the bottom, closed at the top
 - Interval tree
 - Each interval corresponds to a node in a tree, whose parent node represents larger interval, from which interval emerged, and whose off springs represent smaller intervals.
 - Stability of a node is a scale range over which the interval exists.



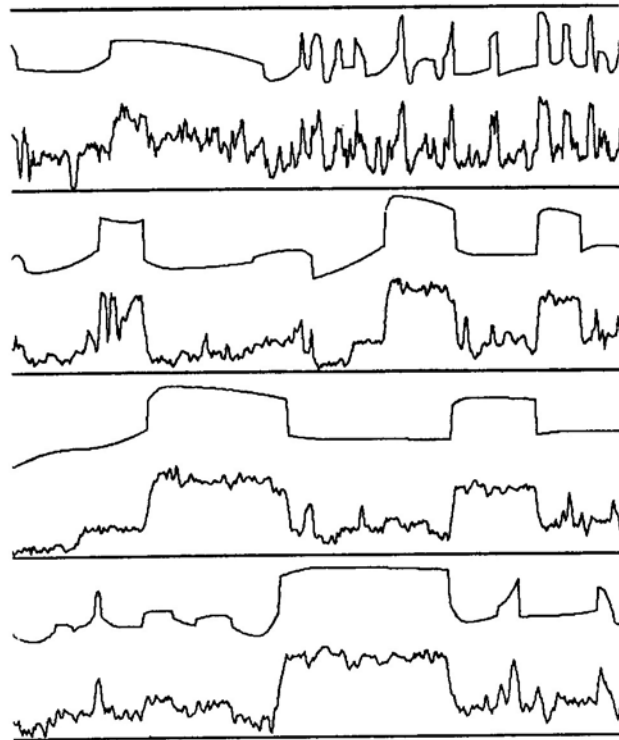
Scale Space

- Top level description
 - Iteratively remove nodes from the tree, splicing out nodes that are less stable than any of their parents and off springs



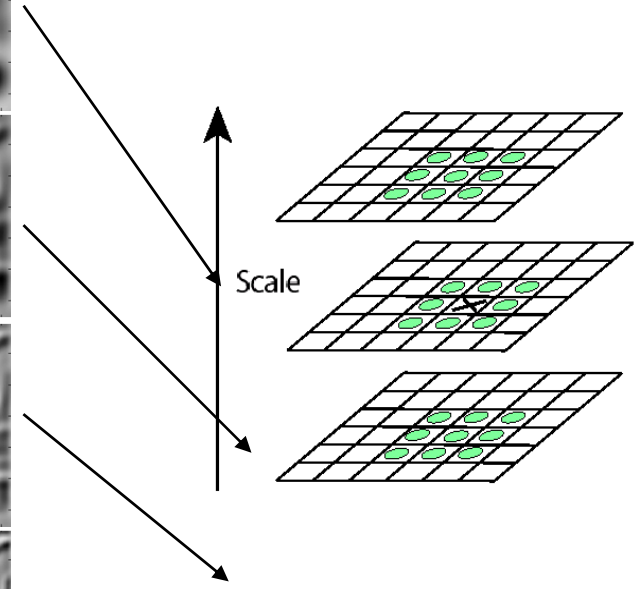
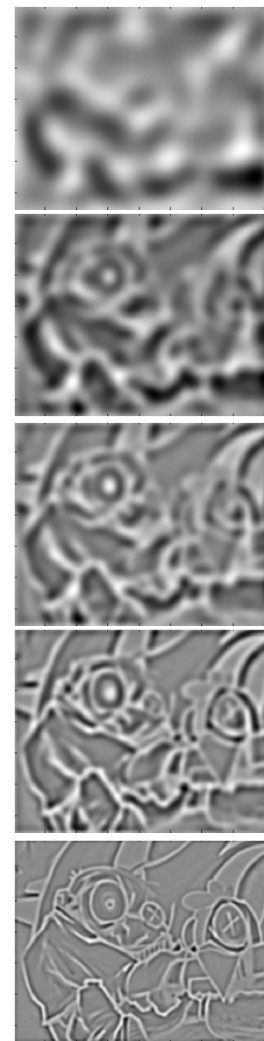
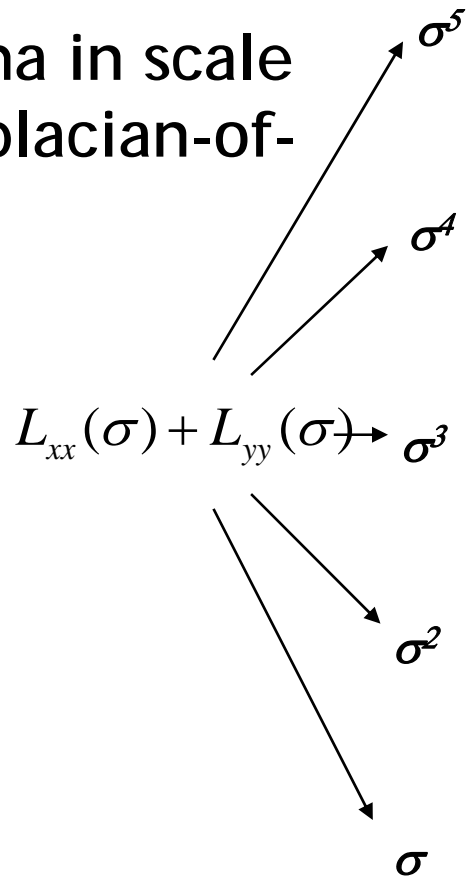
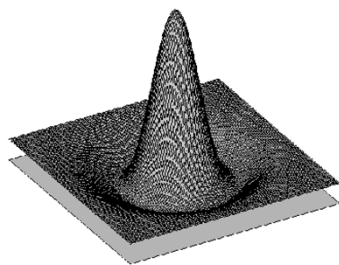
Scale Space

A top level description of several signals using stability criterion.



Laplacian-of-Gaussian (LoG)

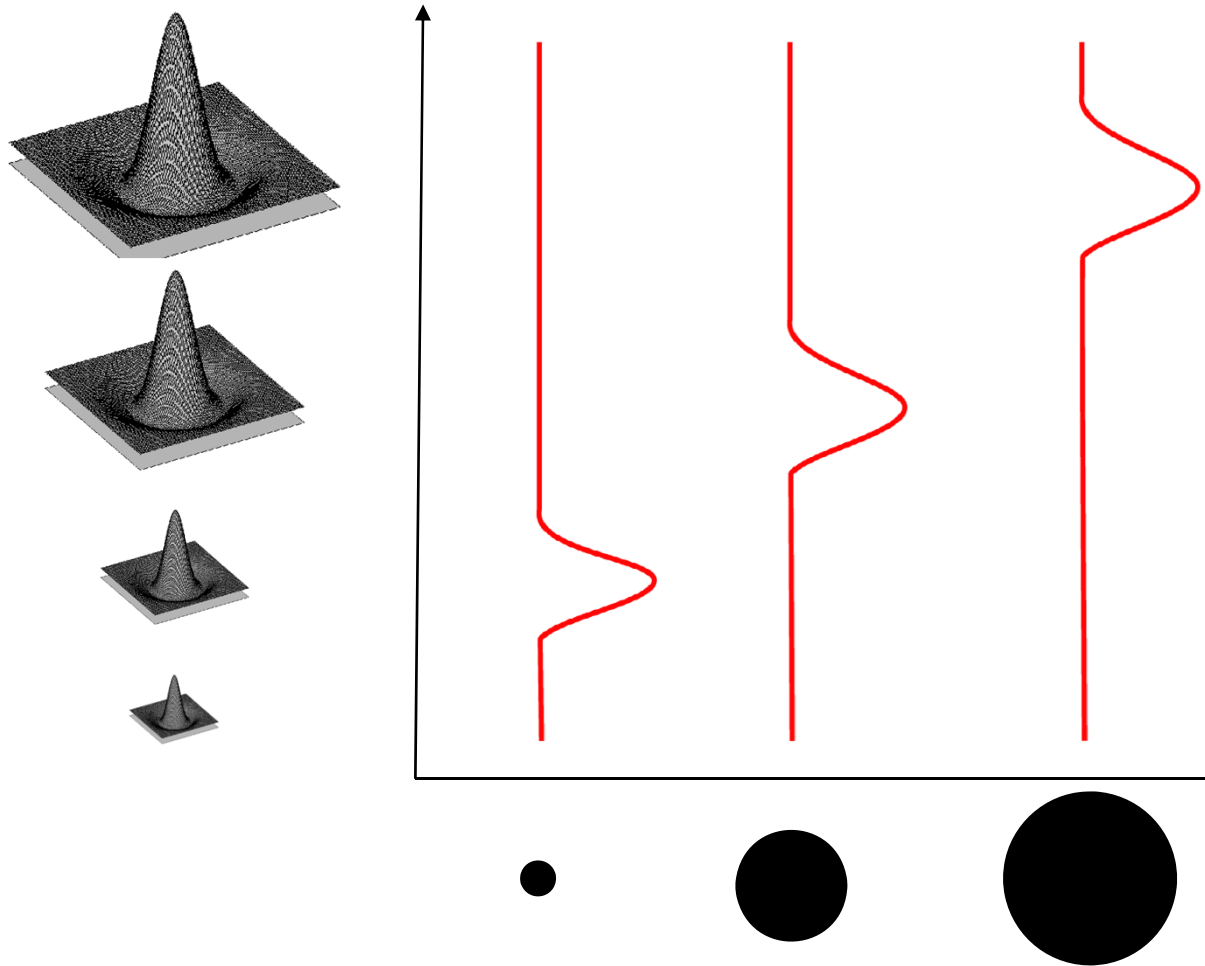
- Interest points:
Local maxima in scale space of Laplacian-of-Gaussian



⇒ List of (x, y, σ)

What Is A Useful Signature Function?

- Laplacian-of-Gaussian = “blob” detector



Scale-space blob detector: Example

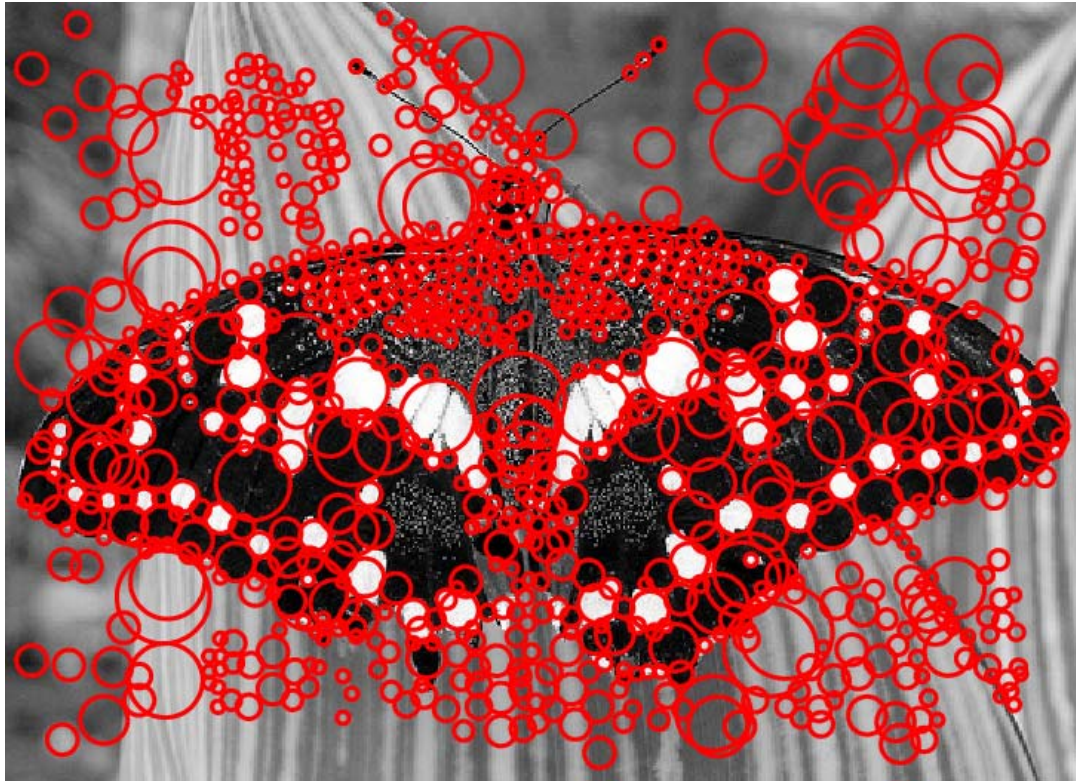


Scale-space blob detector: Example



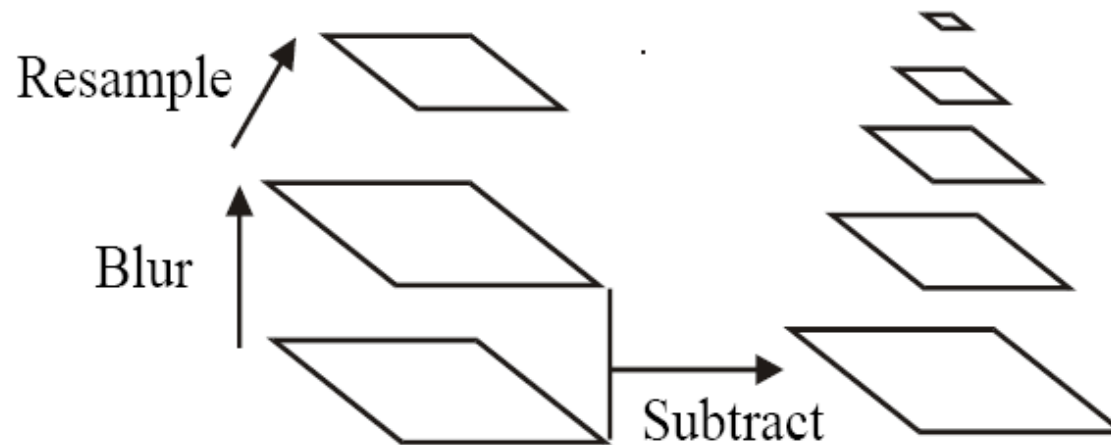
sigma = 11.9912

Scale-space blob detector: Example



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)



Approximation of LoG by Difference of Gaussians

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

Heat Equation

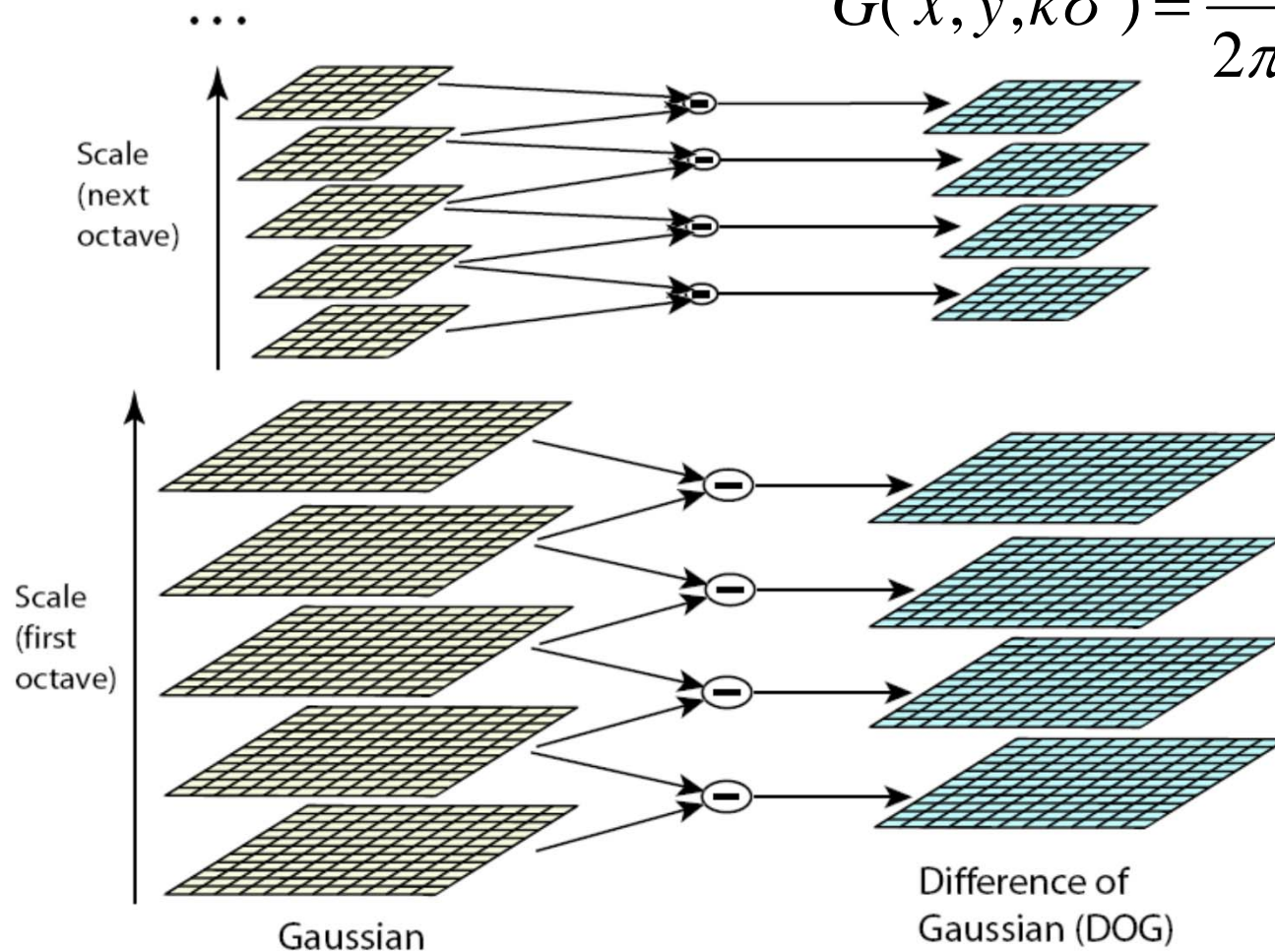
$$\sigma \Delta^2 G = \frac{\partial G}{\partial \sigma} = \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

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

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

Building a Scale Space

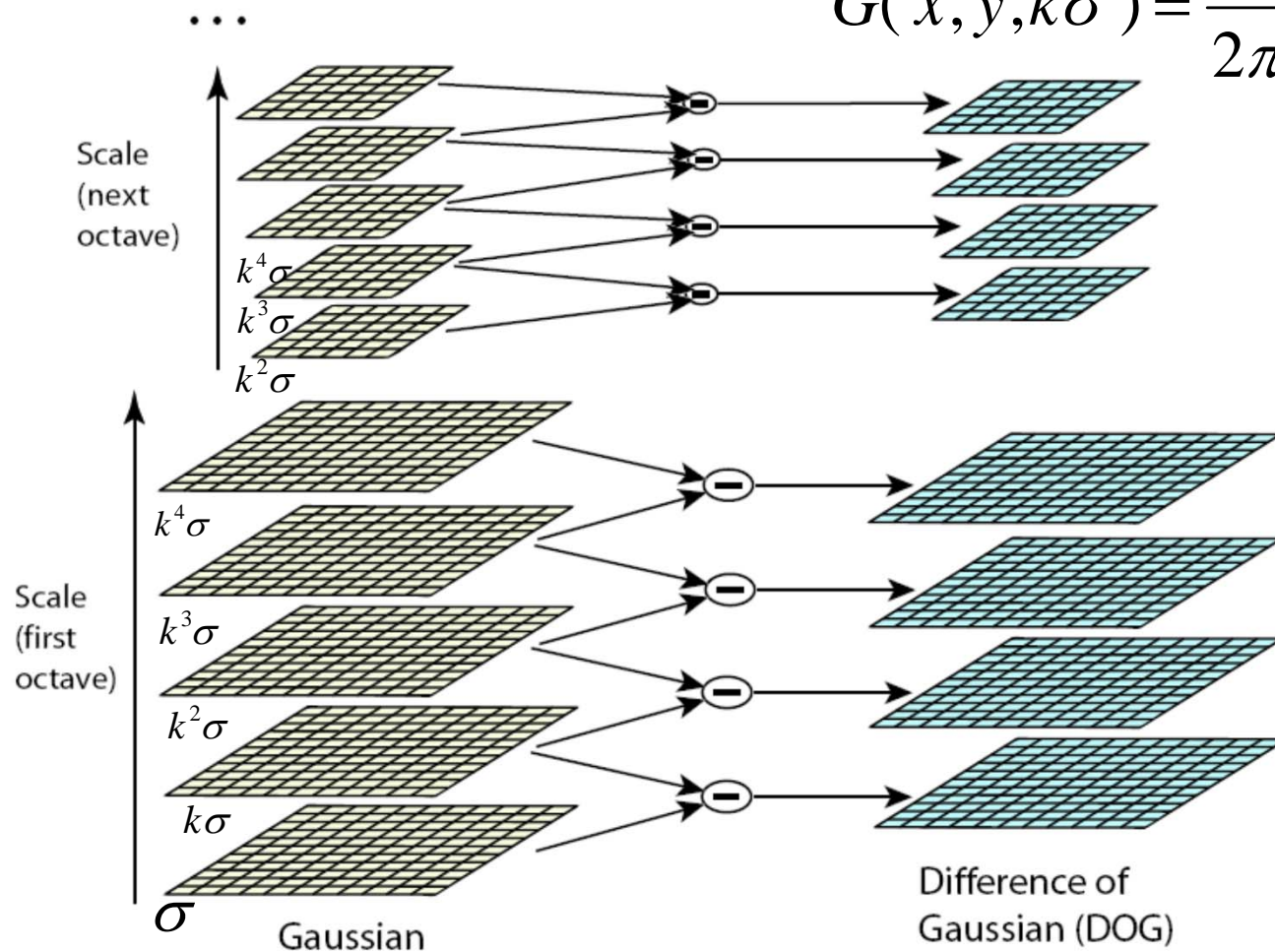
$$G(x, y, k\sigma) = \frac{1}{2\pi(k\sigma)^2} e^{-(x^2 + y^2) / 2k^2\sigma^2}$$



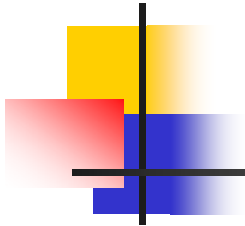
$$k = \sqrt{2}$$

Building a Scale Space

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



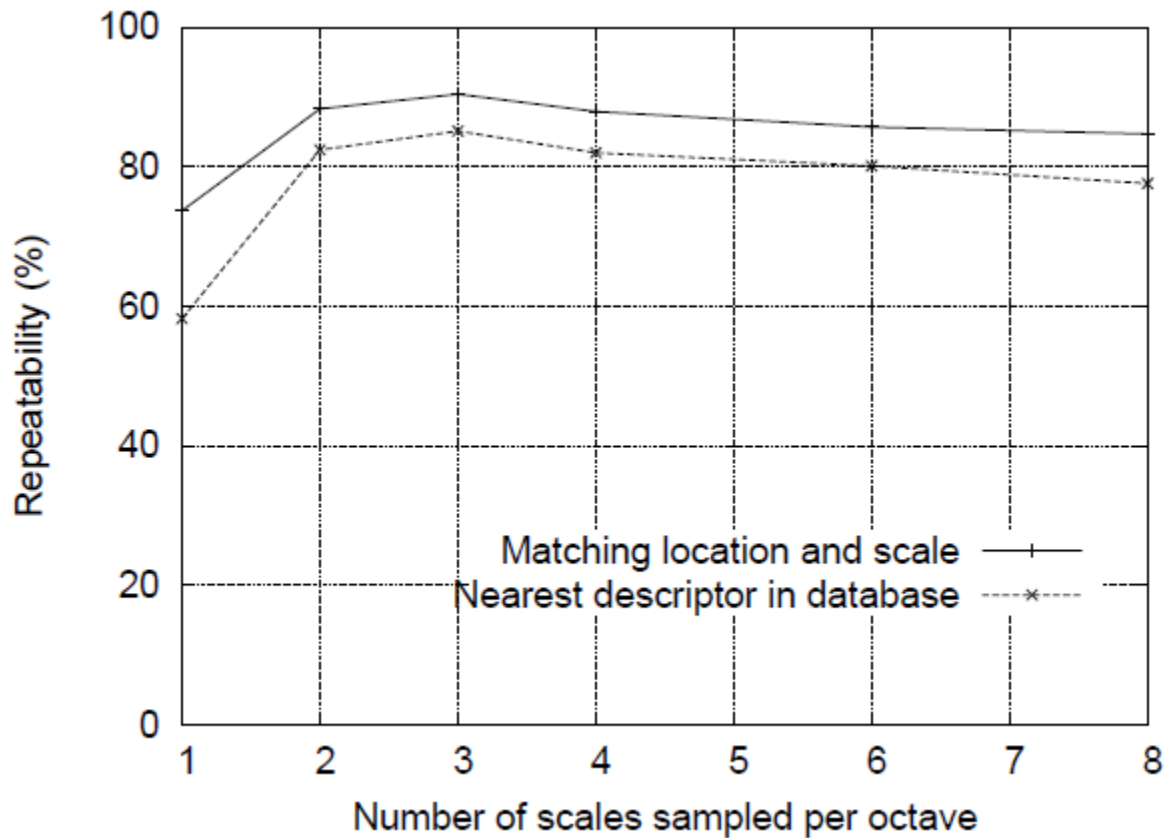
$$k = \sqrt{2}$$



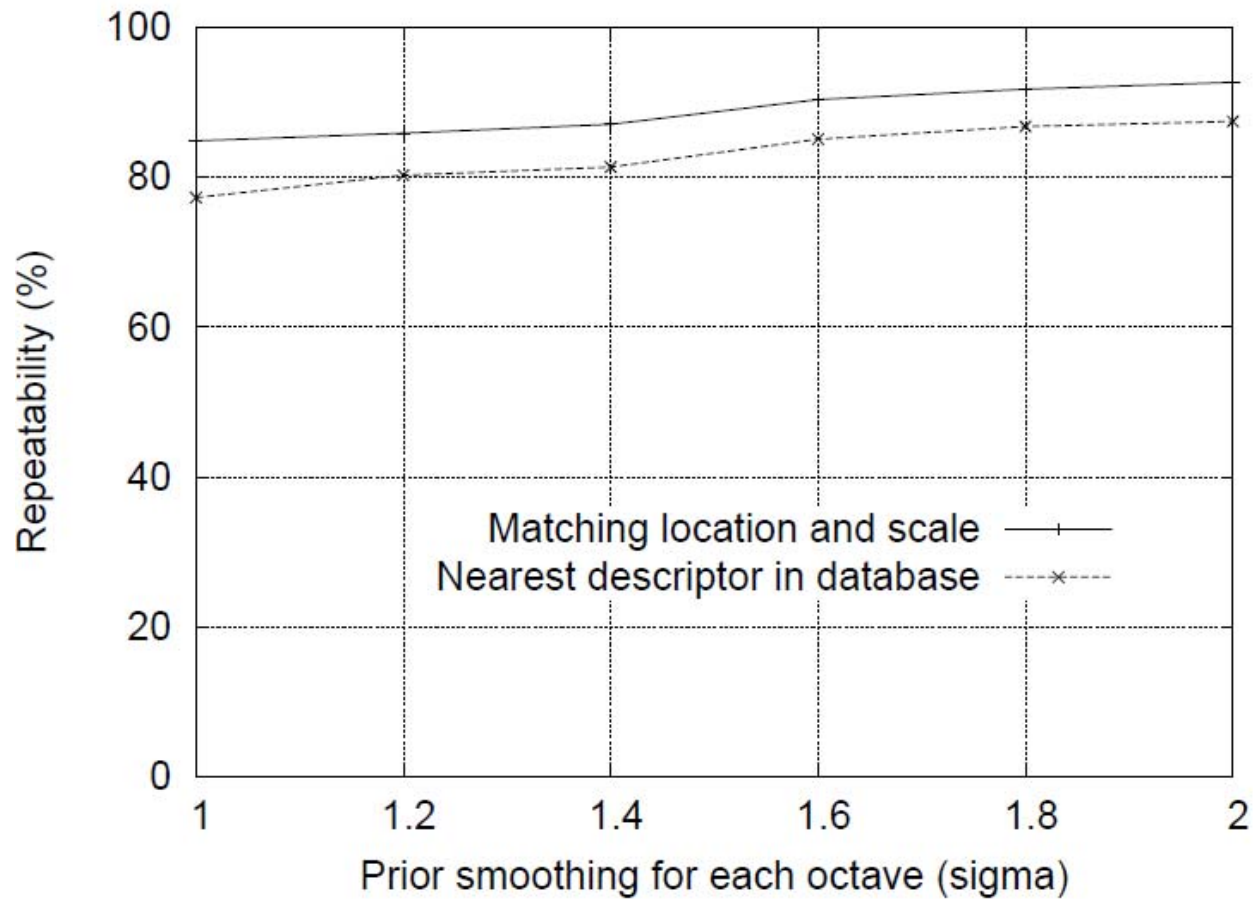
	scale →				
octave	0.707107	1.000000	1.414214	2.000000	2.828427
	1.414214	2.000000	2.828427	4.000000	5.656854
	2.828427	4.000000	5.656854	8.000000	11.313708
	5.656854	8.000000	11.313708	16.000000	22.627417

$$\sigma = .707187.6; k = \sqrt{2}$$

How many scales per octave?

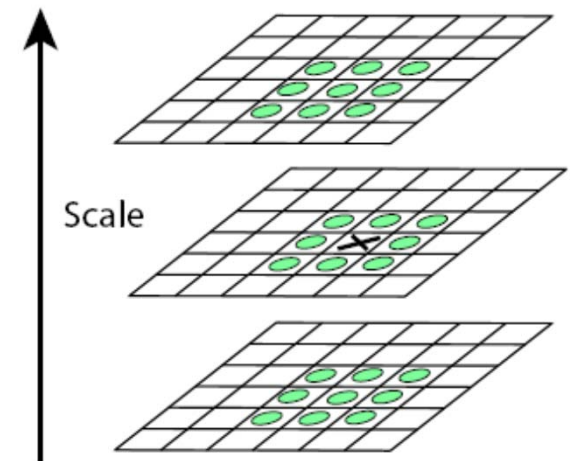


Initial value of sigma



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

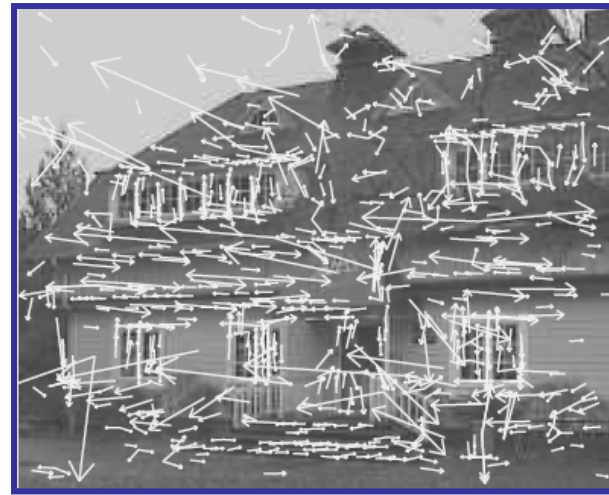


Key Point Localization

- Candidates are chosen from extrema detection



original image



extrema locations



Initial Outlier Rejection

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

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

Homework

- Minima or maxima is located at $\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}$
- Value of $D(\mathbf{x})$ at minima/maxima must be large, $|D(\mathbf{x})| > th$.

Initial Outlier Rejection



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



Further Outlier Rejection

- DOG has strong response along edge
- Assume DOG as a surface
 - Compute principal curvatures (PC)
 - Along the edge one of the PC is very low, across the edge is high



Further Outlier Rejection

- Analogous to Harris corner detector
- Compute Hessian of D

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad \begin{aligned} \text{Tr}(\mathbf{H}) &= D_{xx} + D_{yy} = \lambda_1 + \lambda_2 \\ \text{Det}(\mathbf{H}) &= D_{xx}D_{yy} - (D_{xy})^2 = \lambda_1\lambda_2 \end{aligned}$$

- Remove outliers by evaluating

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



Further Outlier Rejection

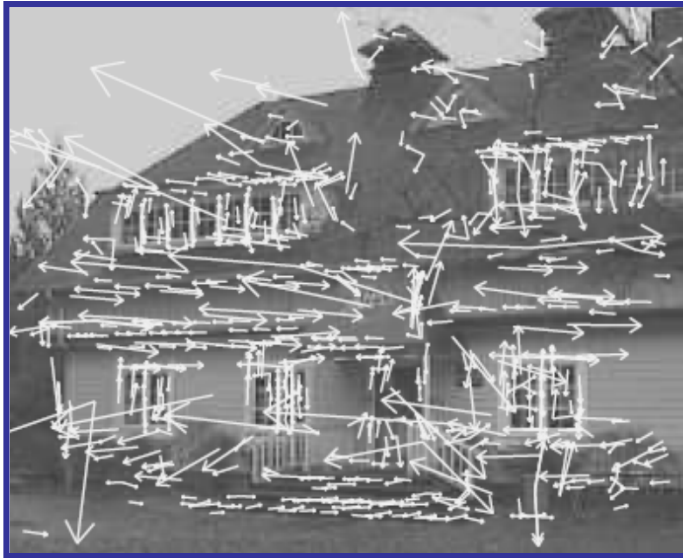
- Following quantity is minimum when $r=1$
- It increases with r

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

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

- Eliminate key points if $r > 10$

Further Outlier Rejection



from 729 key points to 536 key points.



Orientation Assignment

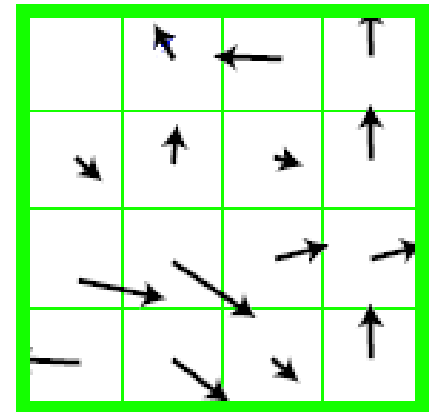
- To achieve rotation invariance
- Compute central derivatives, gradient magnitude and direction of L (smooth image) at the scale of key point (x,y)

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

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

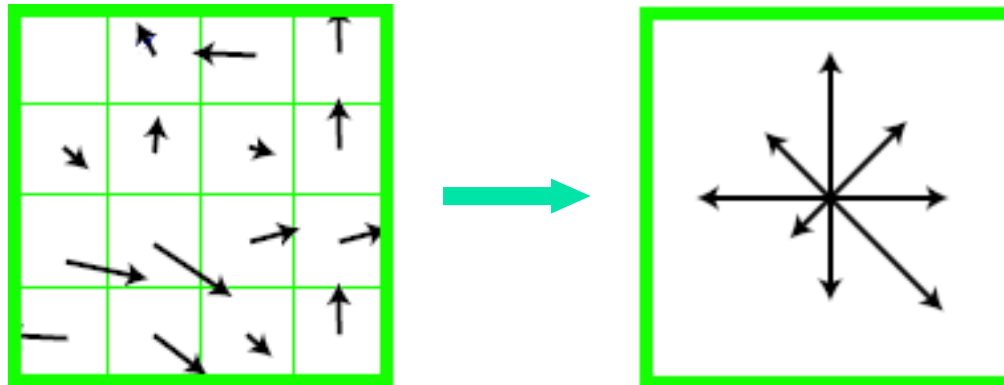
Orientation Assignment

- Create a weighted direction histogram in a neighborhood of a key point (36 bins)
- Weights are
 - Gradient magnitudes
 - Spatial gaussian filter with $\sigma = 1.5 \times \langle \text{scale of key point} \rangle$



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





Local Image Descriptors at Key Points

- Possible descriptor
 - Store intensity samples in the neighborhood
 - Sensitive to lighting changes, 3D object transformation
- Use of gradient orientation histograms
 - Robust representation

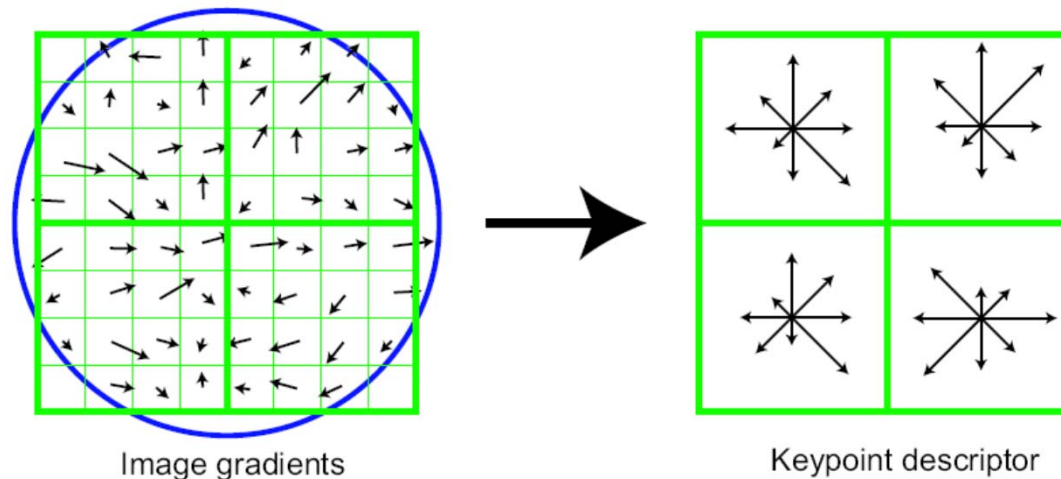


Similarity to IT cortex

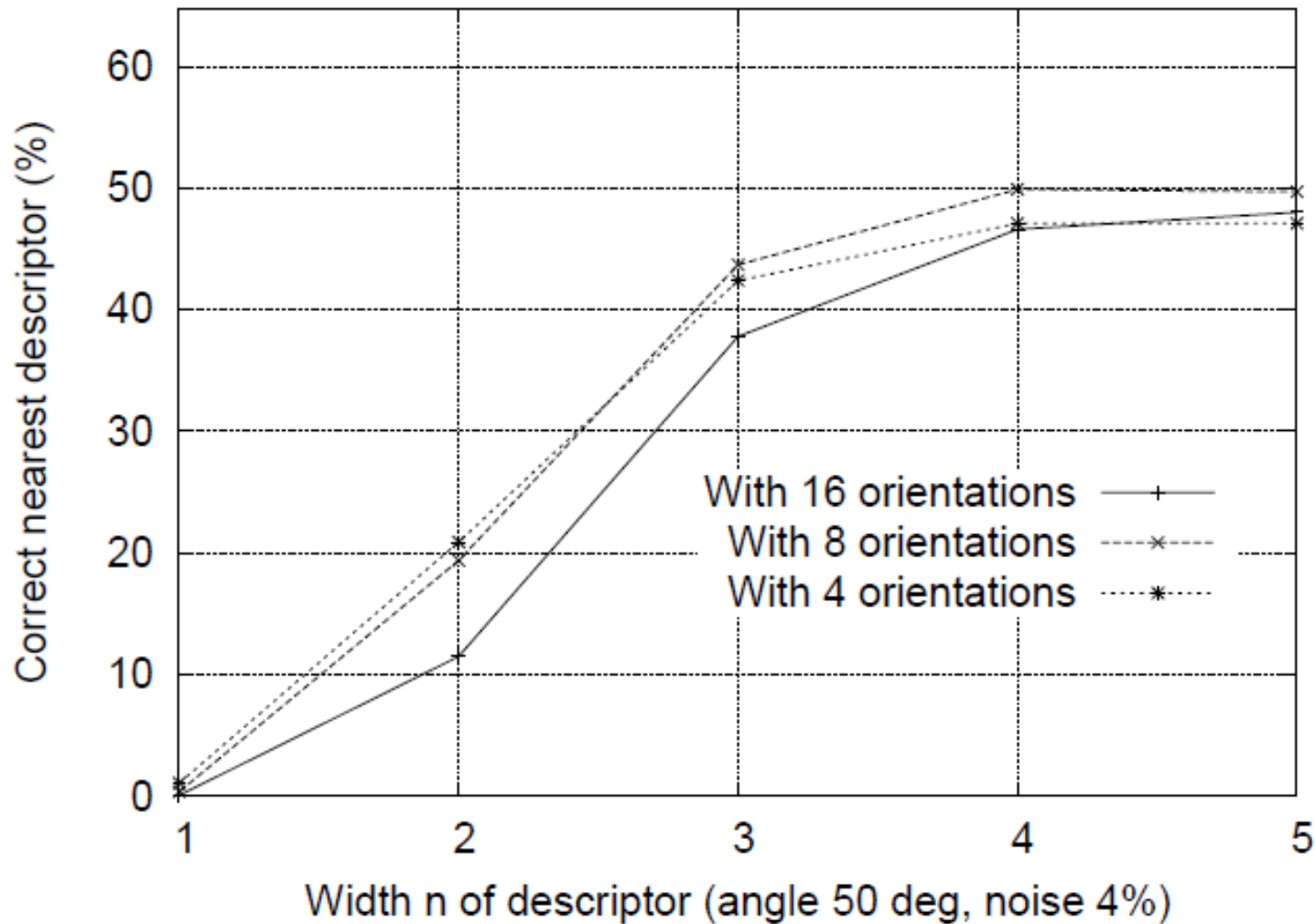
- Complex neurons respond to a gradient at a particular orientation.
- Location of the feature can shift over a small receptive field.
- Edelman, Intrator, and Poggio (1997)
 - The function of the cells allow for matching and recognition of 3D objects from a range of view points.
- Experiments show better recognition accuracy for 3D objects rotated in depth by up to 20 degrees

Extraction of Local Image Descriptors at Key Points

- Compute relative orientation and magnitude in a 16x16 neighborhood at key point
- Form weighted histogram (8 bin) for 4x4 regions
 - Weight by magnitude and spatial Gaussian
 - Concatenate 16 histograms in one long vector of 128 dimensions
- Example for 8x8 to 2x2 descriptors



Descriptor Regions (n by n)





Extraction of Local Image Descriptors at Key Points

- Store numbers in a vector
- Normalize to unit vector (**UN**)
 - Illumination invariance (affine changes)
- For non-linear intensity transforms
 - Bound **Unit Vector** items to maximum 0.2 (remove larger gradients)
 - Renormalize to unit vector



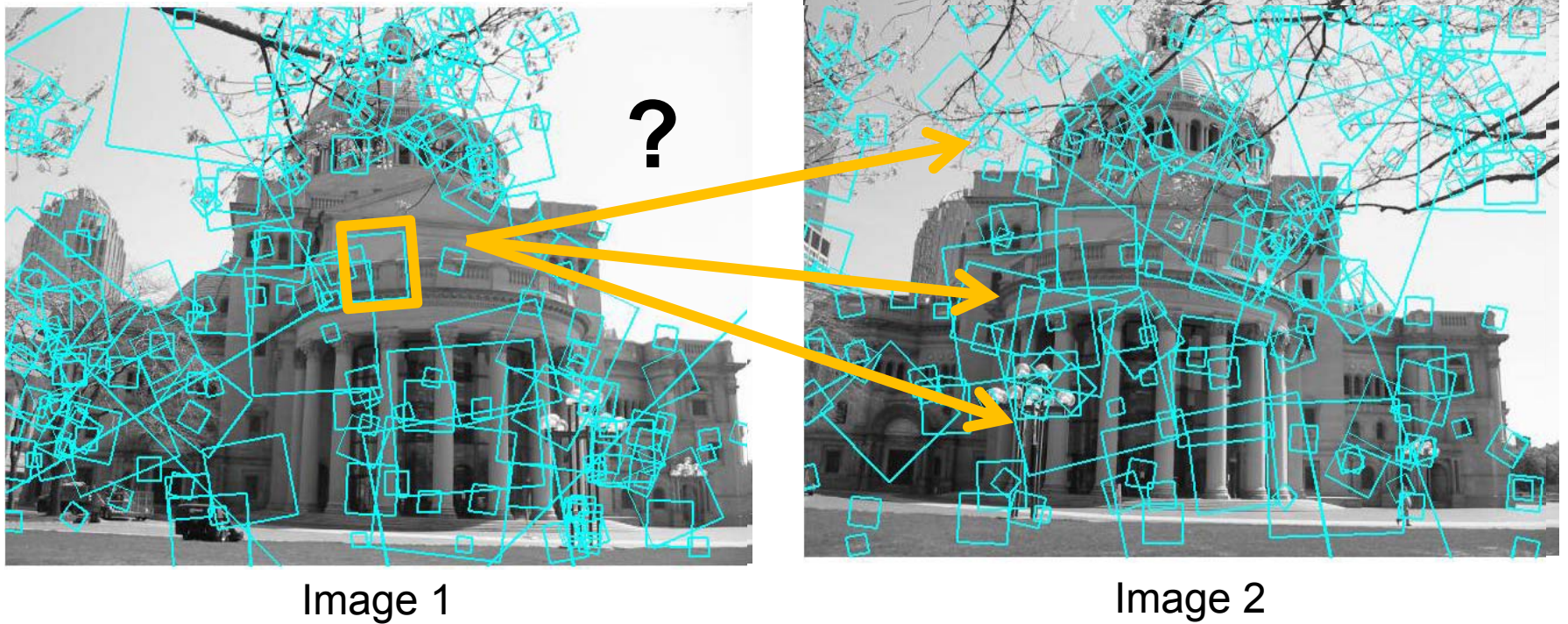
Key point matching

- 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)

Matching local features



Matching local features



- To generate **candidate matches**, find patches that have the most similar appearance or SIFT descriptor
- Simplest approach: compare them all, take the closest (or closest k, or within a thresholded distance)

Ambiguous matches



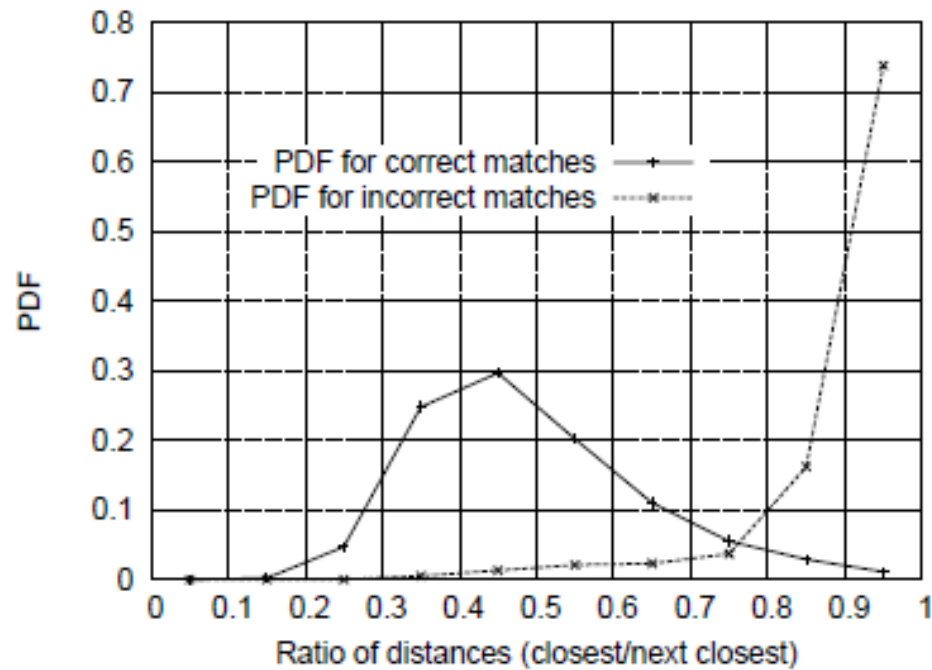
Image 1



Image 2

- At what distance do we have a good match?
- To add robustness to matching, can consider **ratio** : distance to best match / distance to second best match
- If low, first match looks good.
- If high, could be ambiguous match.

The ratio of distance from the closest to the distance of the second closest





Reference

D. Lowe. *Distinctive image features from scale-invariant key points.*, International Journal of Computer Vision 2004.