SIFT:

Scale Invariant Feature Transform
Week - 4

Overview

- Review
- Motivation for SIFT
- SIFT Feature Detector
- SIFT Descriptor
- Application

Overview

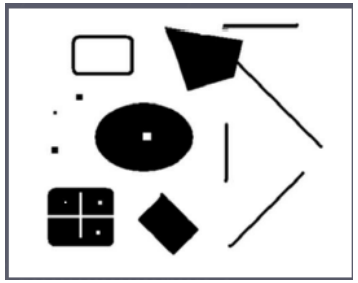
- Review
- Motivation for SIFT
- SIFT Feature Detector
- SIFT Descriptor
- Application

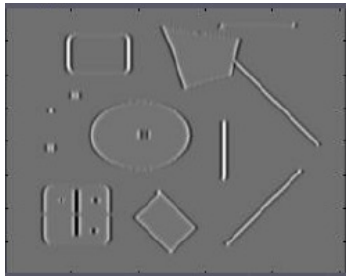
Recall: Harris Corner Detection

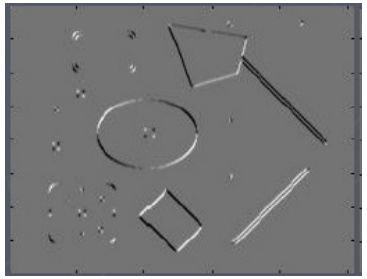
Corners as distinctive interest points

$$M = \sum w(x,y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix}$$

2 x 2 matrix of image derivatives (averaged in neighborhood of a point).





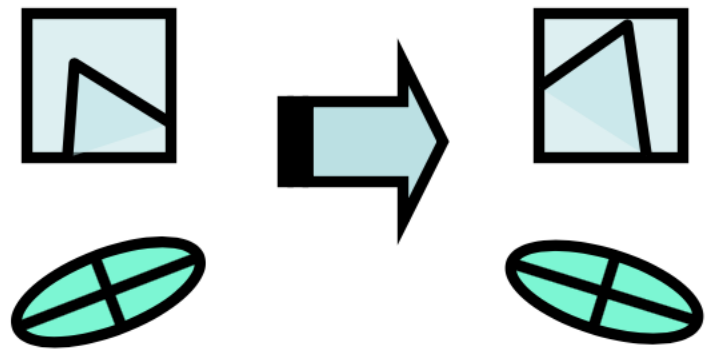


Notation:

$$I_x \Leftrightarrow \frac{\partial I}{\partial x} \qquad I_y \Leftrightarrow \frac{\partial I}{\partial y} \qquad I_x I_y \Leftrightarrow \frac{\partial I}{\partial x}\frac{\partial I}{\partial y}$$

Harris Corner: Properties

- Harris corner is rotation-invariant.



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

Corner response R is invariant to image rotation

$$\begin{pmatrix} I'_x \\ I'_y \end{pmatrix} = R_{2\times 2} \begin{pmatrix} I_x \\ I_y \end{pmatrix}$$

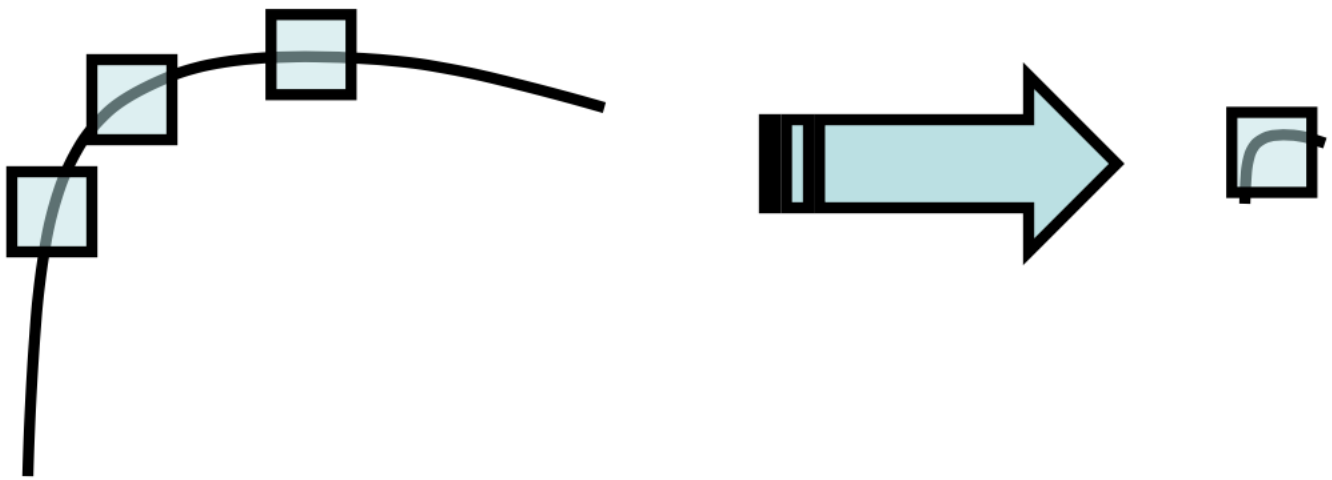
$$\begin{pmatrix} I'_x \\ I'_y \end{pmatrix} \begin{pmatrix} I'_x & I'_y \end{pmatrix} = R_{2\times 2} \begin{pmatrix} I_x \\ I_y \end{pmatrix} \begin{pmatrix} I_x & I_y \end{pmatrix} R^T_{2\times 2}$$

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

$$
\begin{aligned}
M' &= R \times M \times R^T \\
&= RX \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} (RX)^T
\end{aligned}
$$

Harris Corner: Properties

- But: it is **not invariant** to *image scale change* !



All points will be classified as edges

Corner !

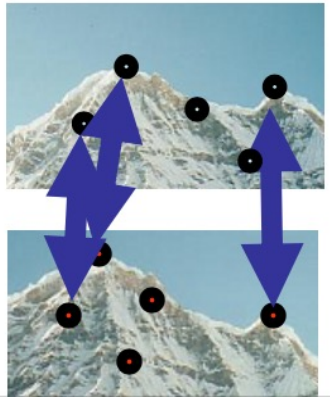
Example of scale change

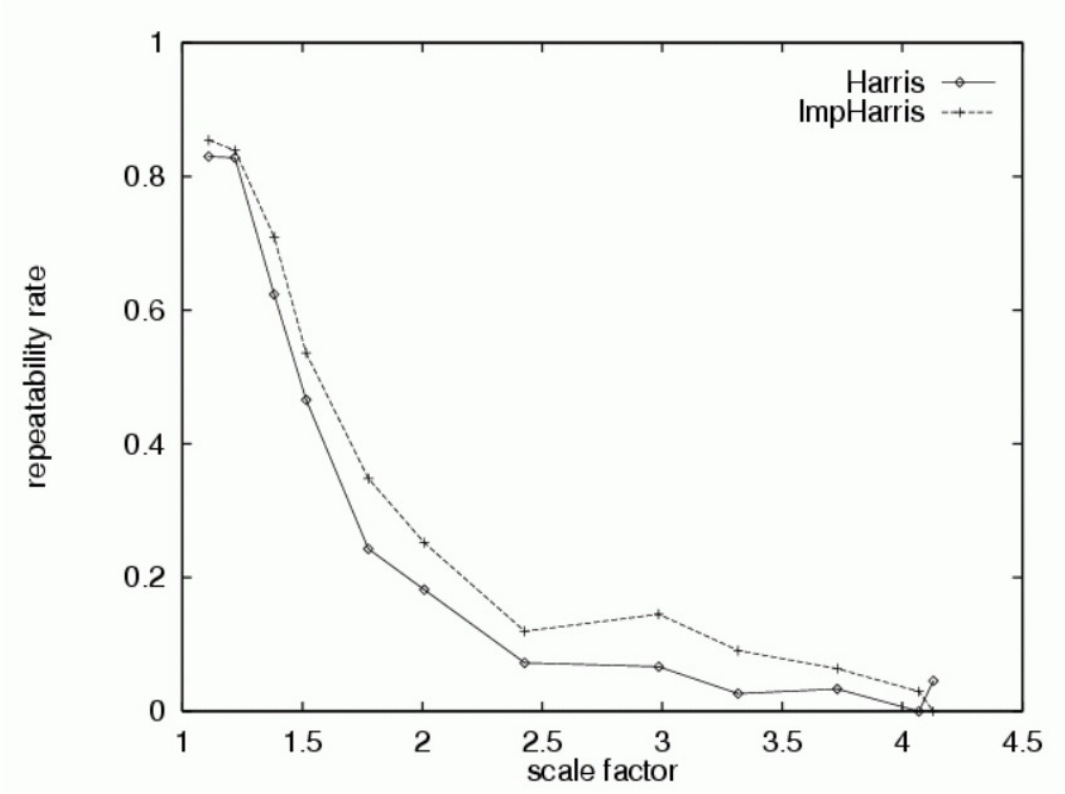
From left to right: Reference image, scale change for middle one is 1.5, scale change for right one is 4.1.

Harris Corner is not *Scale Invariant*

ε Repeatability rate:

$$\frac{\#\ \text{correspondences}}{\#\ \text{possible correspondences}}$$





Imp.Harris is a variant of Harris corner detector, which uses derivative of Gaussian instead of standard template used by Harris et al.

C.Schmid et.al. “Evaluation of Interest Point Detectors”. IJCV 2000

How to adapt to scale change ?

Goal

- To detect the same interest points regardless of image scale changes

How to adapt to scale change ?

- David Lowe's ***SIFT**** is a very efficient implementation of scale invariant distinctive image features detector.

*D. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60 (2), pp. 91-110, 2004.

SIFT: Motivation

- The Harris operator is not invariant to scale change.
- For better (more reliable, robust) image matching, Lowe's goal was to develop an interest operator that is invariant to scale and rotation.
- Also, Lowe aimed to create a descriptor that is robust to the variations in images, corresponding to typical viewing conditions.

Advantages of SIFT

- **Locality:** features are local, so robust to occlusion and clutter (no prior segmentation)
- **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
- **Extensibility:** can easily be extended to wide range of differing feature types, with each adding robustness

Overall Procedure at a High Level

1. Scale-space extrema detection

   Search over *multiple scales* and *image locations*.

2. Keypoint localization

   Fit a model to determine location and scale. Select keypoints based on a measure of stability.

3. Orientation assignment

   Compute best orientation(s) for each keypoint region.

4. Keypoint description

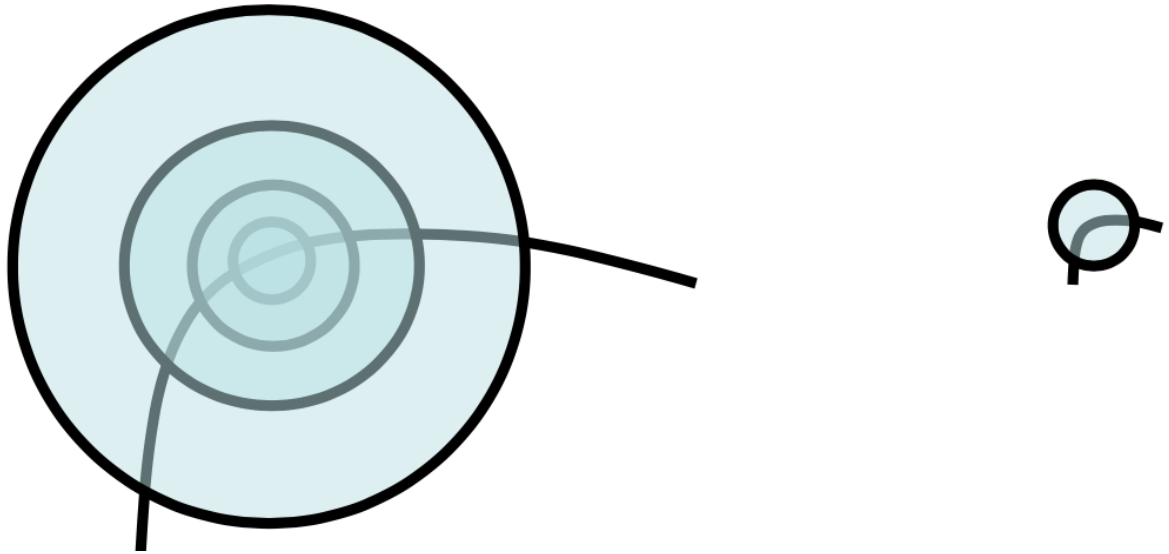
   Use local image gradients at selected scale and rotation to describe each keypoint region.

Overview

- Review
- Motivation for SIFT
- SIFT Feature Detector
- SIFT Descriptor
- Application

Scale Invariant Detection

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



Automatic Scale Selection principle [Lindeberg98*]

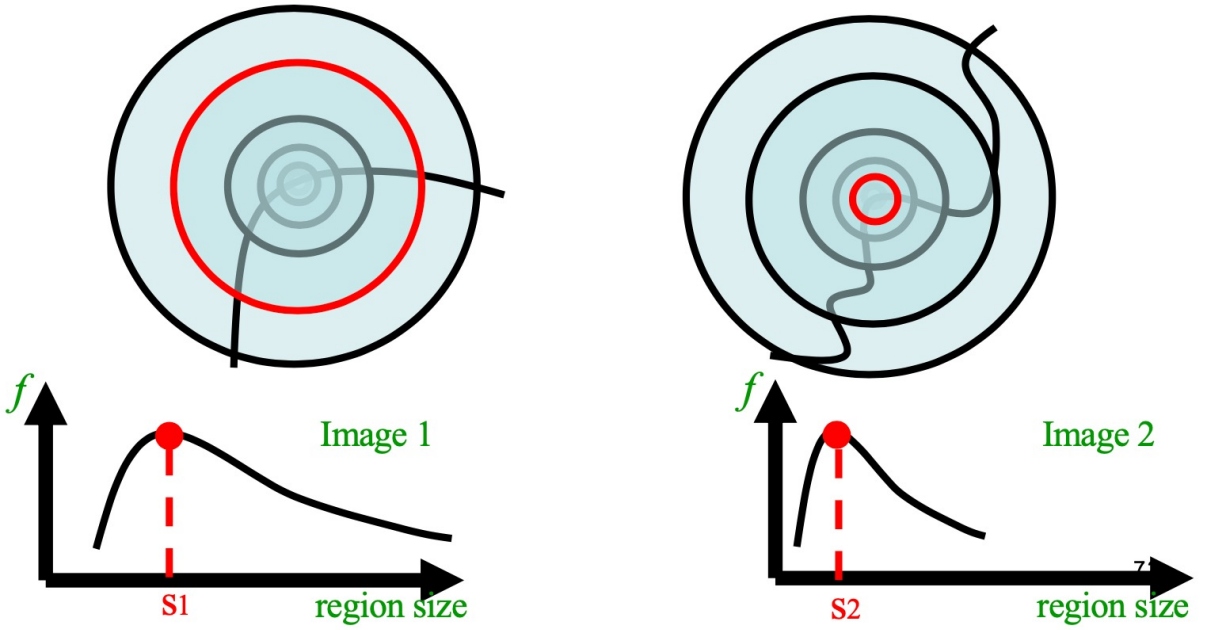
- Optimal scale selection principle:
  - The scale at which ‘some function’ (e.g. the normalized derivative) assumes that an extreme value indicates a feature containing interesting pattern/structure.

*T. Lindeberg, "Feature detection with automatic scale selection," International Journal of Computer Vision 30 (2), pp. 77-116, 1998.

Automatic Scale Selection

- **Principle:**

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

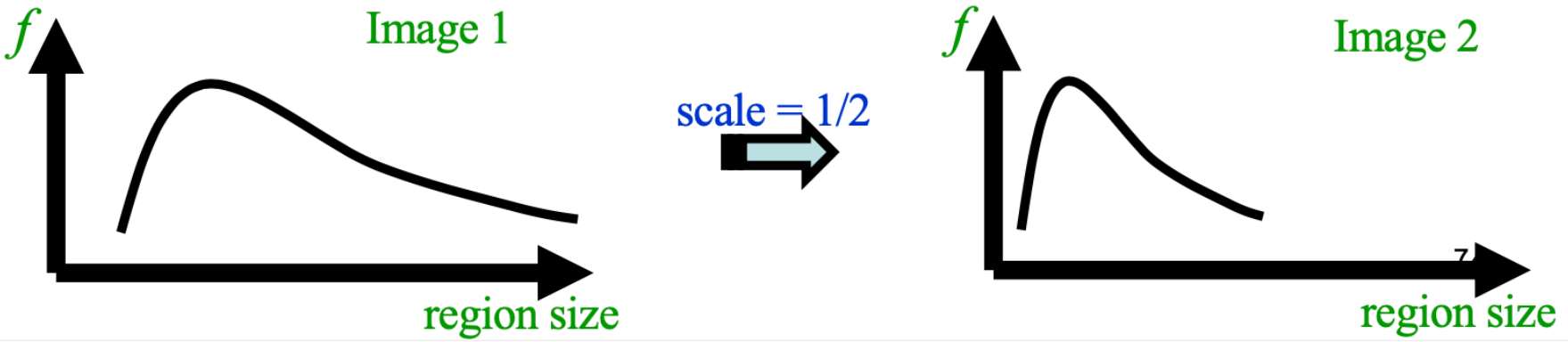


Automatic Scale Selection

- Solution:

– Design some function on the region (circle), which is “scale invariant” (the same for corresponding regions, even if they are at different scales)

– For a point in one image, we can consider **f** as a function of region size (circle radius)



- What is this ‘signature function’ f ?

Scale Invariant Function

- Functions for determining scale

$$f = \text{Kernel} * \text{Image}$$

Kernels:

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

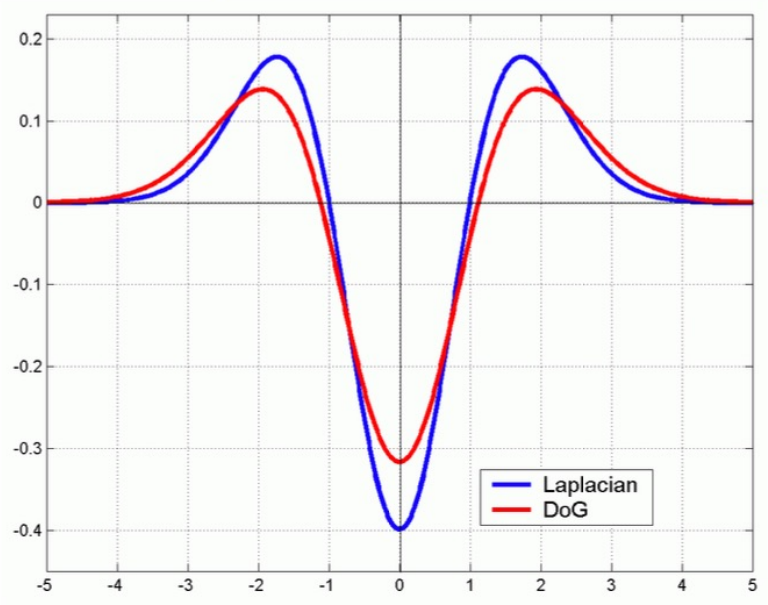
(Laplacian of Gaussian)

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

(Difference of Gaussians)

where Gaussian

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

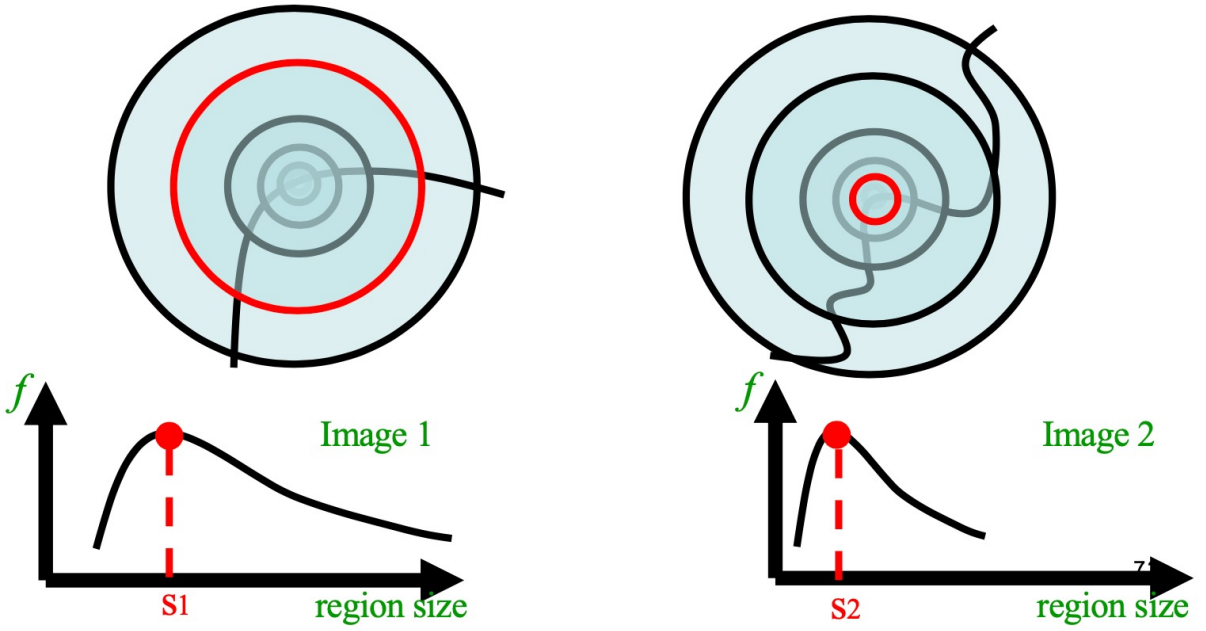


Note: both Kernels are invariant to *scale* and *rotation*

Recall: Automatic Scale Selection

- **Principle:**

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



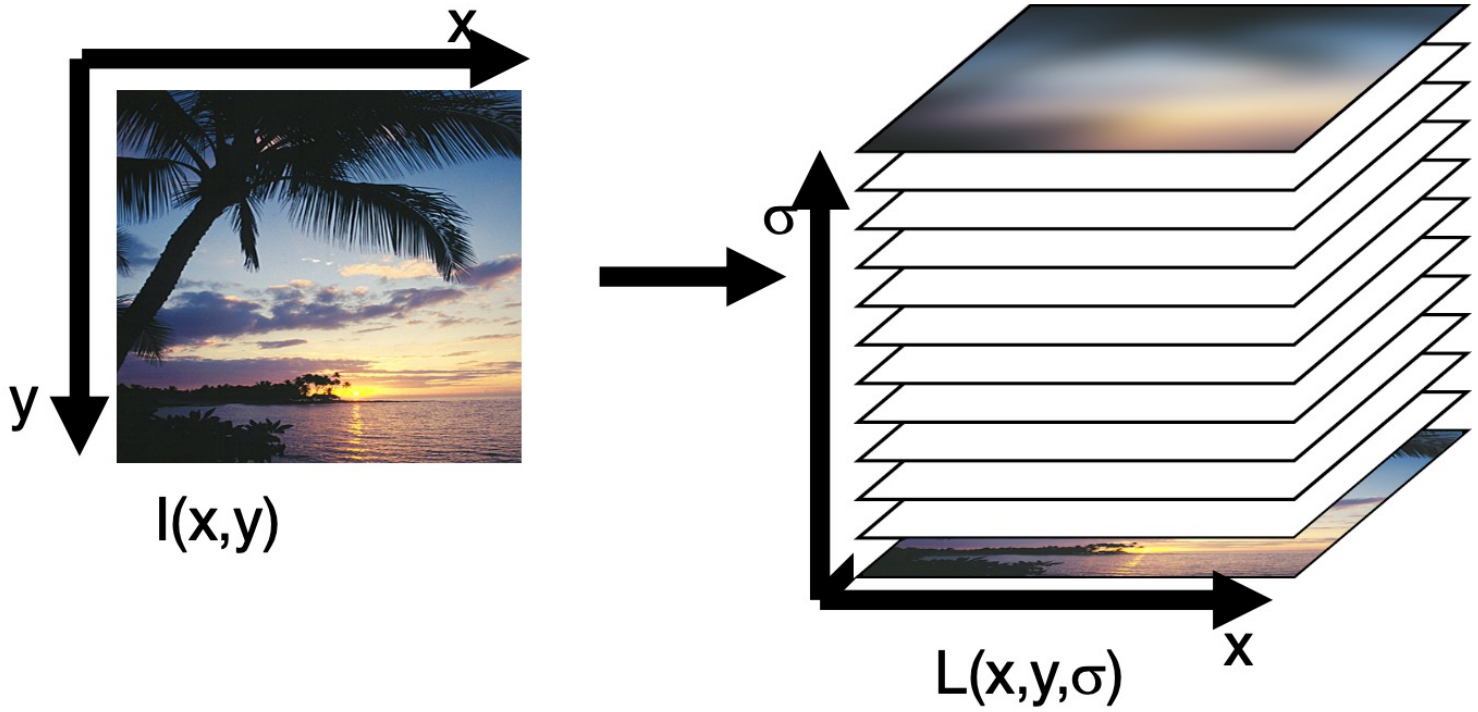
Scale space representation

$$L: R^2 \times R \rightarrow R$$

$$L(x, y; \sigma) = G(\sigma) * I(x, y)$$

- L is the scale space representation of I(x,y).
- L is obtained by smoothing I with Gaussian kernel G(σ).
- G(σ) is the natural choice for building up a scale space.

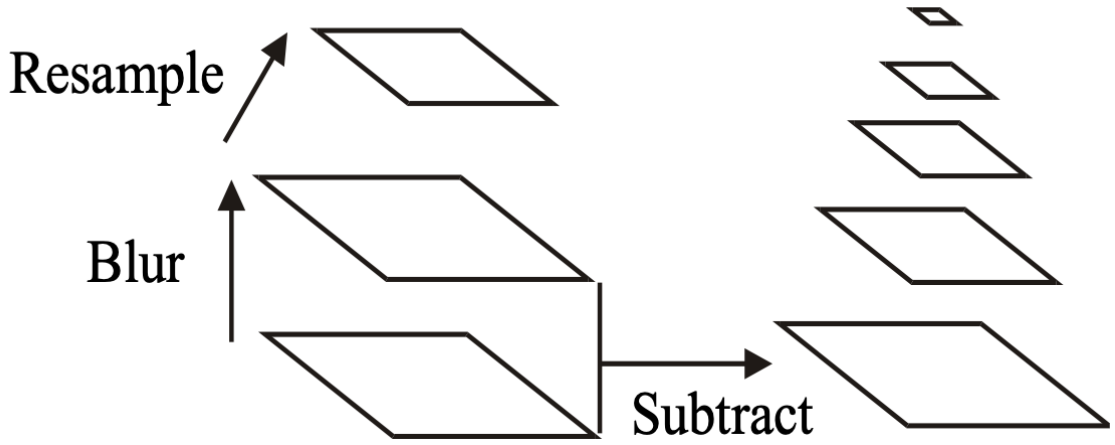
Scale space representation



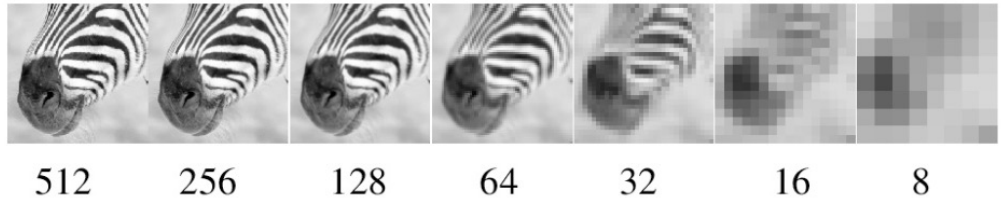
Scale-space extremum indicates the occurrence of the “optimal scale”

Build Scale-Space Pyramid

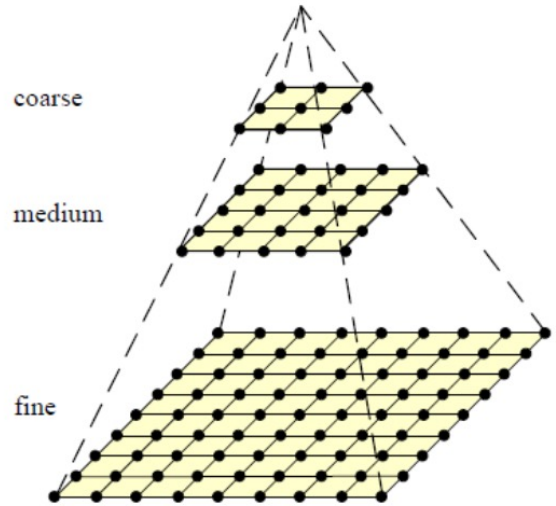
- All scales are examined to identify scale-invariant features.
- An efficient function is to compute the Difference of Gaussian (DOG) pyramid.



Gaussian Pyramid

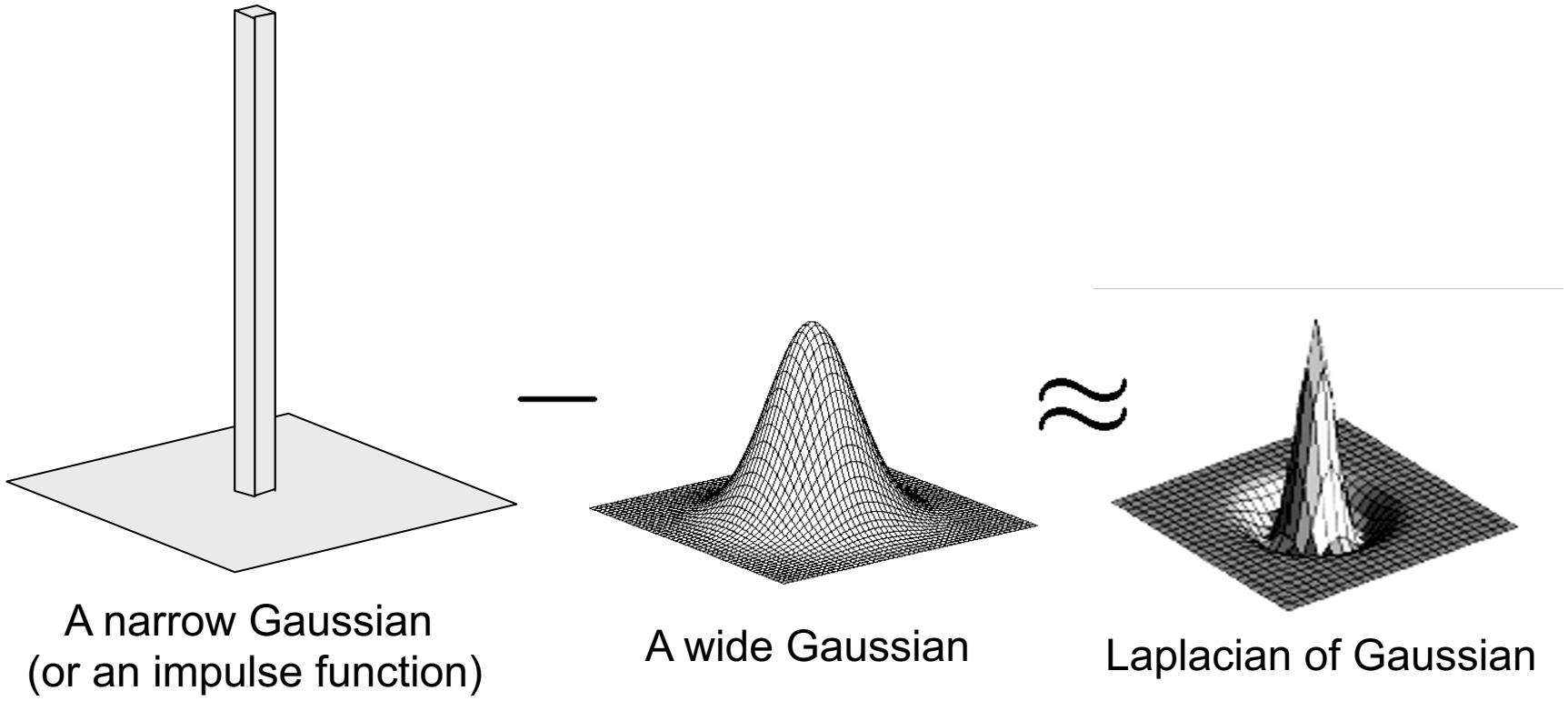


512 256 128 64 32 16 8

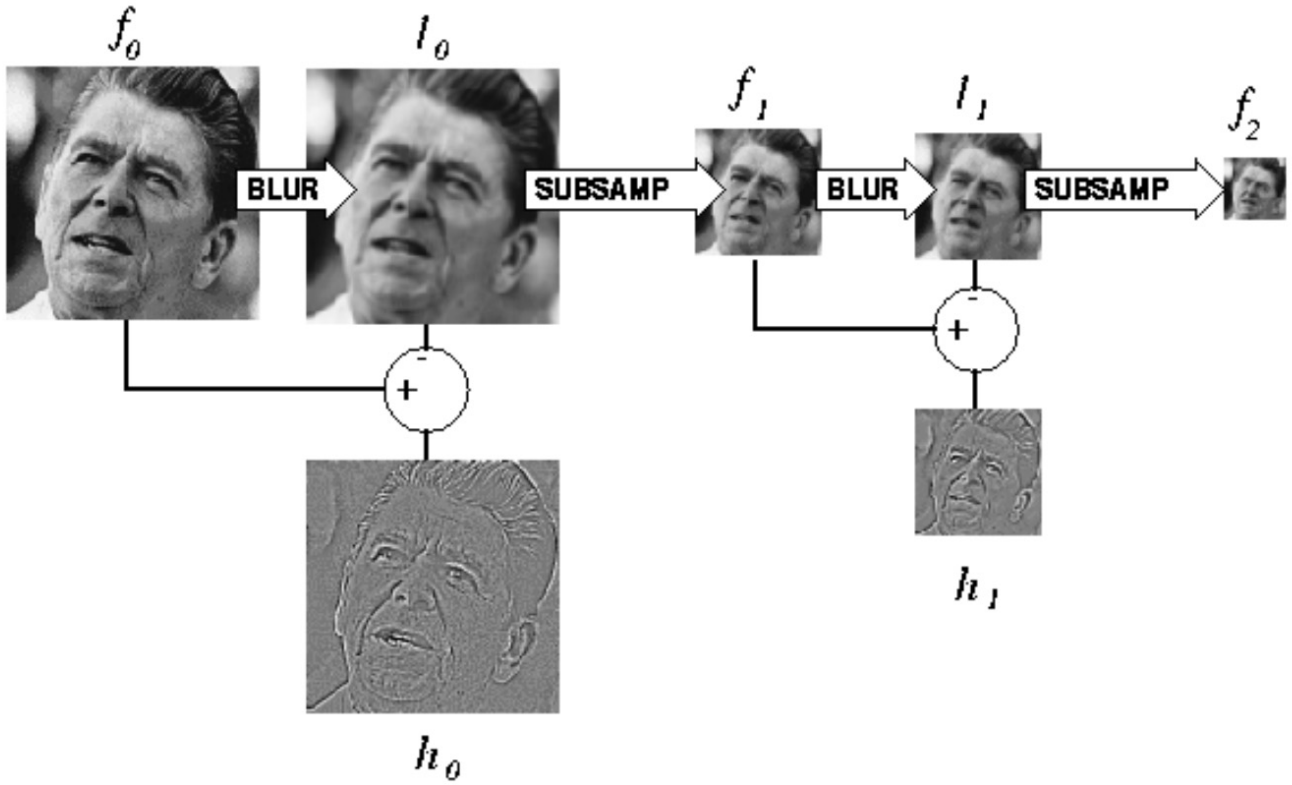


From **Forsyth**

“Laplacian of Gaussian” (LoG) can be approximately computed as “Difference of Gaussians (DoG).

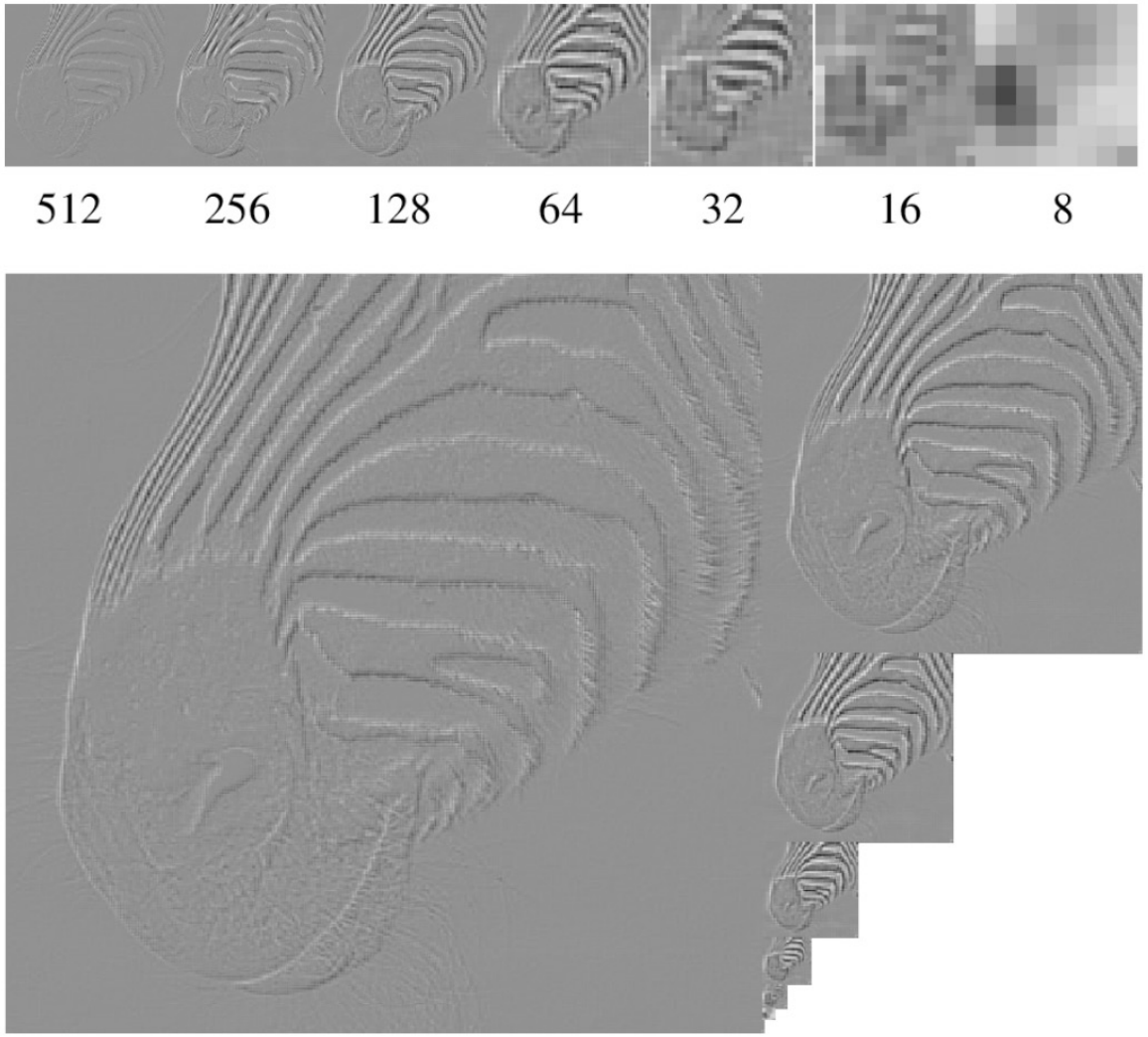


Compute Gaussian and Laplacian Pyramid



http://sepwww.stanford.edu/data/media/public/sep/morgan/texturematch/paper_html/node3.html#SECTION00012000000000000000

Laplacian Pyramid



From Forsyth

Two Popular Scale Invariant Detectors

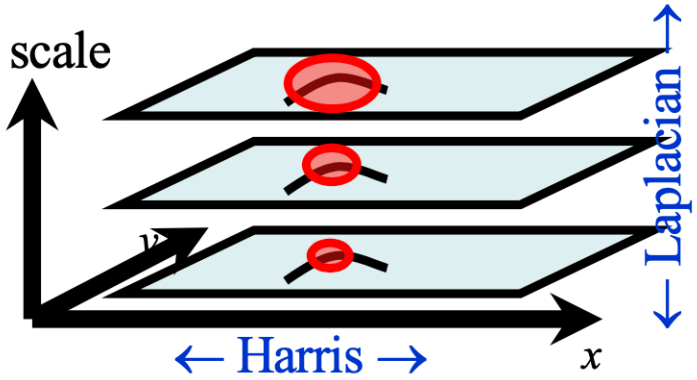
- Harris-Laplacian[1]

Find local maximum of:

- Harris corner detector

in space (image coordinates)

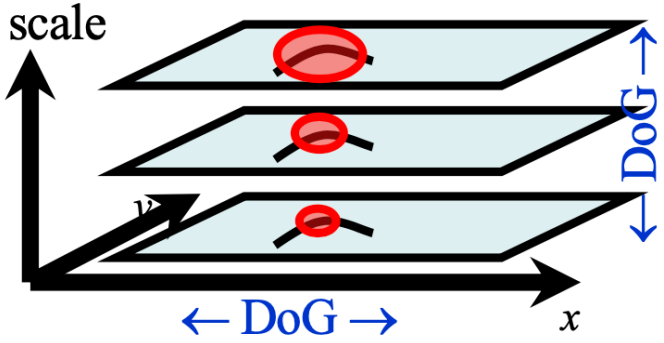
- Laplacian in scale



- SIFT[2]

Find local maximum of:

- Difference of Gaussians in space and scale



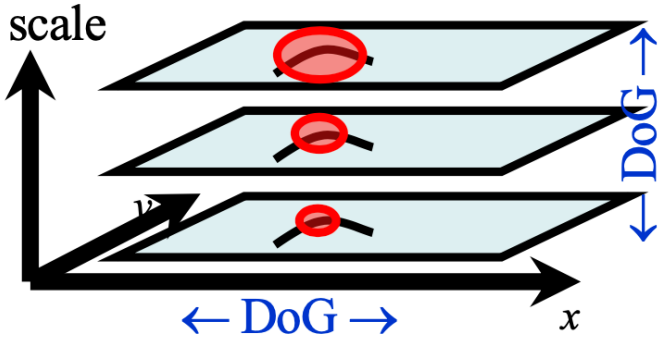
[1]K.Mikolajczyk, C.Schmid. “Indexing Based on Scale Invariant Interest Points”. ICCV 2001
[2]D.Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. IJCV 2004

SIFT

- SIFT

Find local maximum of:

  - Difference of Gaussians in space and scale

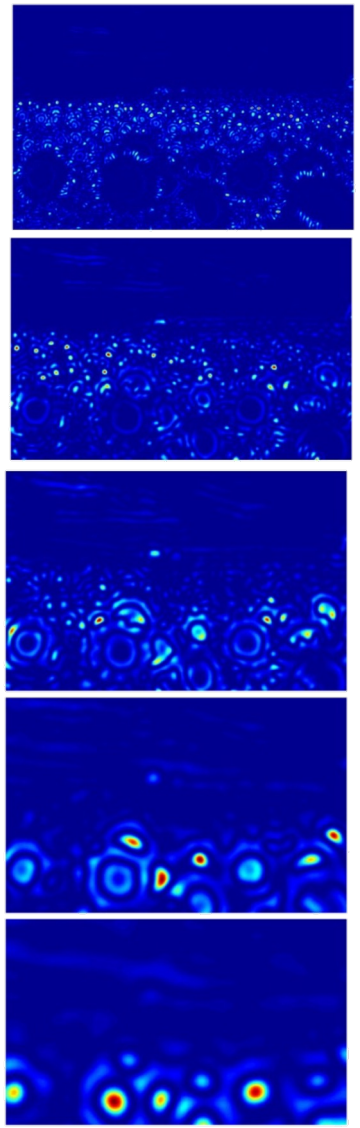


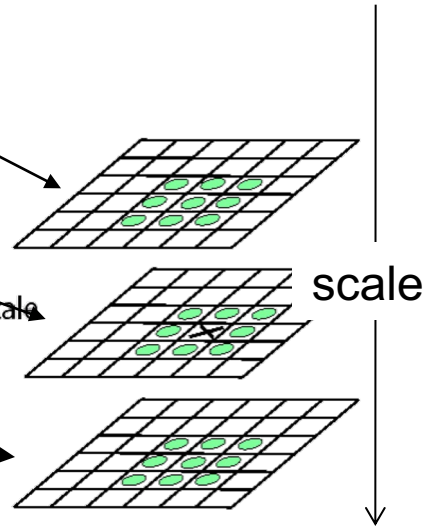
- Interest points are local maxima in both position and scale.

Scale invariant interest points

Interest points are local maxima in both position and scale.

$$L_{xx}(\sigma) + L_{yy}(\sigma)$$





⇒ **List of**
(x, y, σ***)***

Squared filter response maps

Slide credit: Kristen Grauman

Keypoint Spatial Localization

- There are still a lot of points, some of them are not good enough.
- The locations of keypoints may not be accurate.
- Eliminating edge points.

Accurate Localization of Key Points

- Discard points with low contrast
  - Step one: Second order Taylor expansion of the LOG (DOG(s)):

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

  - Step two: Taking the derivative w.r.t x and set it to zero to refine the optimum.

  - Step three: Evaluate D at $\hat{x}$

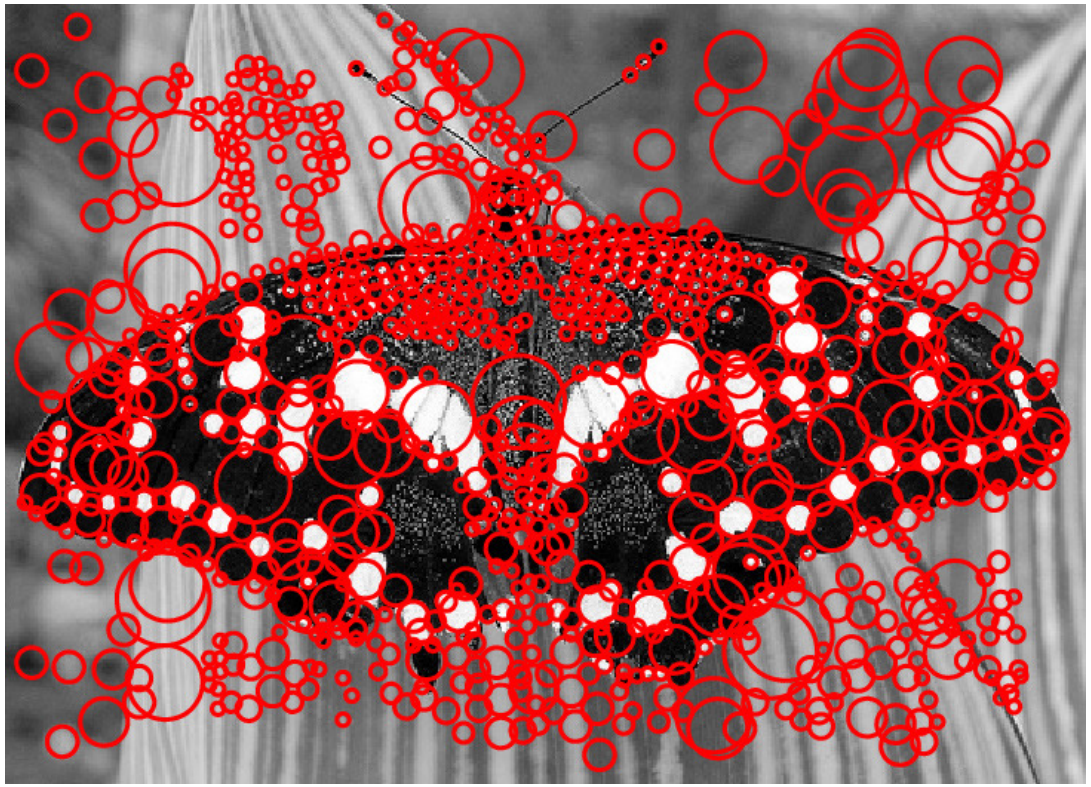
$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1}\frac{\partial D}{\partial \mathbf{x}}$$

  - Reject points with D less than a threshold (e.g 0.03 in the paper)

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

Eliminating edge responses

- Laplacian has strong response along edges



Removing Edge Points

- Such a point has large principal curvature across the edge but a small on in the perpendicular direction
- The principal curvatures can be calculated from a hessian function

$$H = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{pmatrix}$$

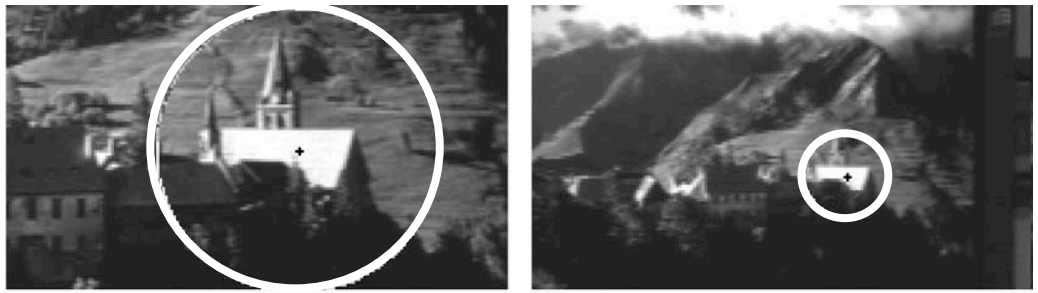
- The eigenvalues of H are proportional to the principal curvatures. So two eigenvalues shouldn't have much difference if it belongs to an interest point. Otherwise, it belongs to edges.

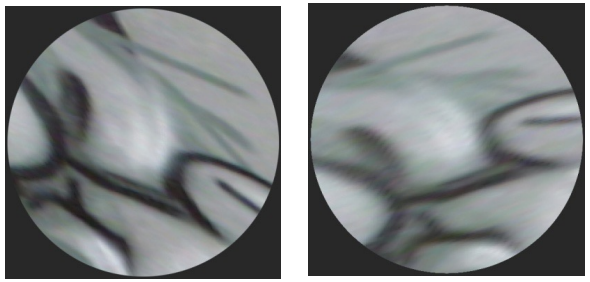
Overview

- Motivation for SIFT
- SIFT Feature Detector
- SIFT Descriptor
- Application

From feature detection to feature description

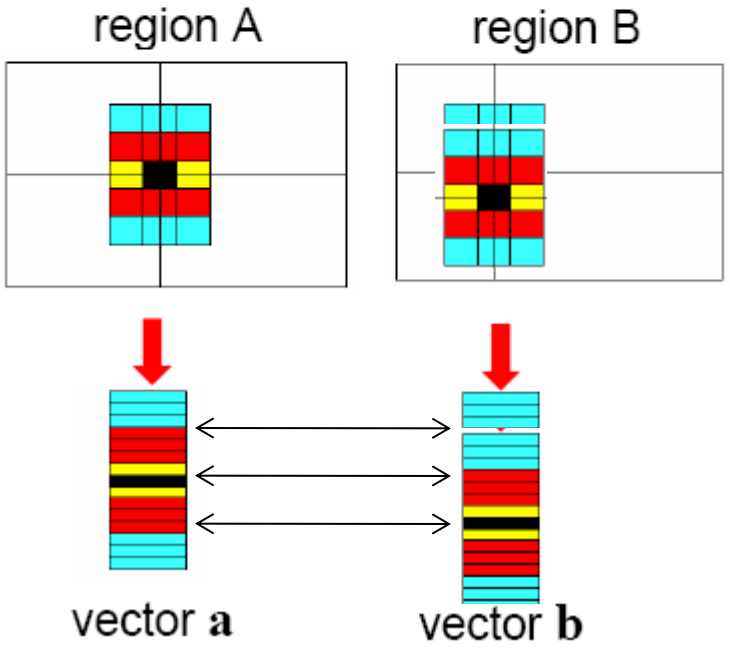
- To recognize the same pattern in multiple images, we need to match appearance “signatures” in the neighborhoods of extracted keypoints
  - But corresponding neighborhoods can be related by a scale change or rotation
  - We want to *normalize* neighborhoods to make signatures invariant to these transformations





Slide from Svetlana Lazebnik

Raw patches as local descriptors



The simplest way to describe the neighborhood around an interest point is to write down the list of intensities to form a feature vector.

But this is very sensitive to even small shifts, rotations.

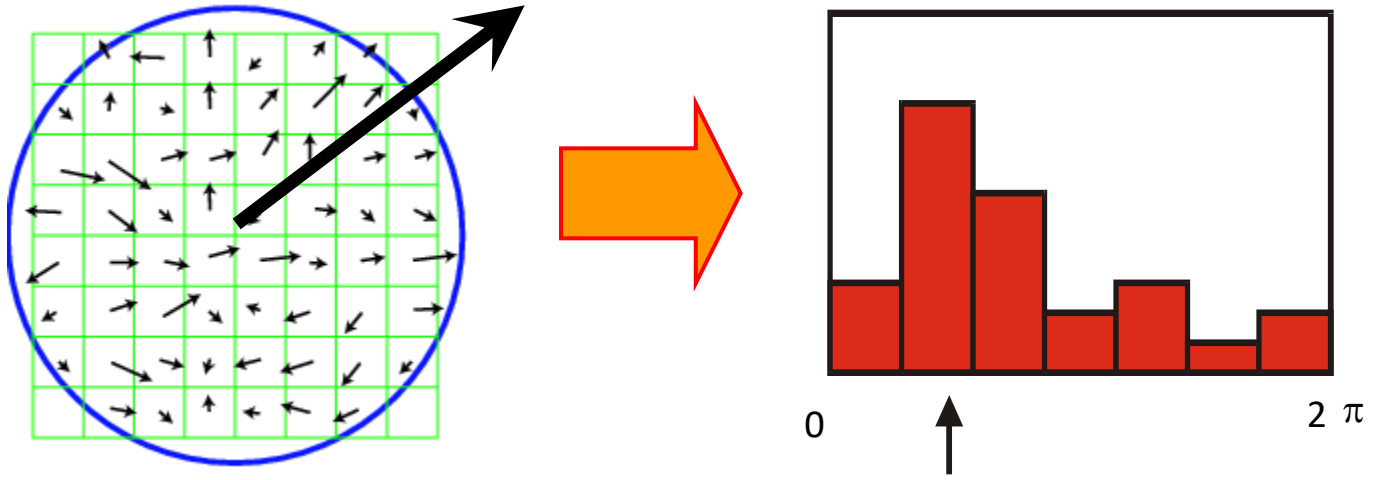
Figure: Andrew Zisserman

Photometric transformations

Figure from T. Tuytelaars ECCV 2006 tutorial

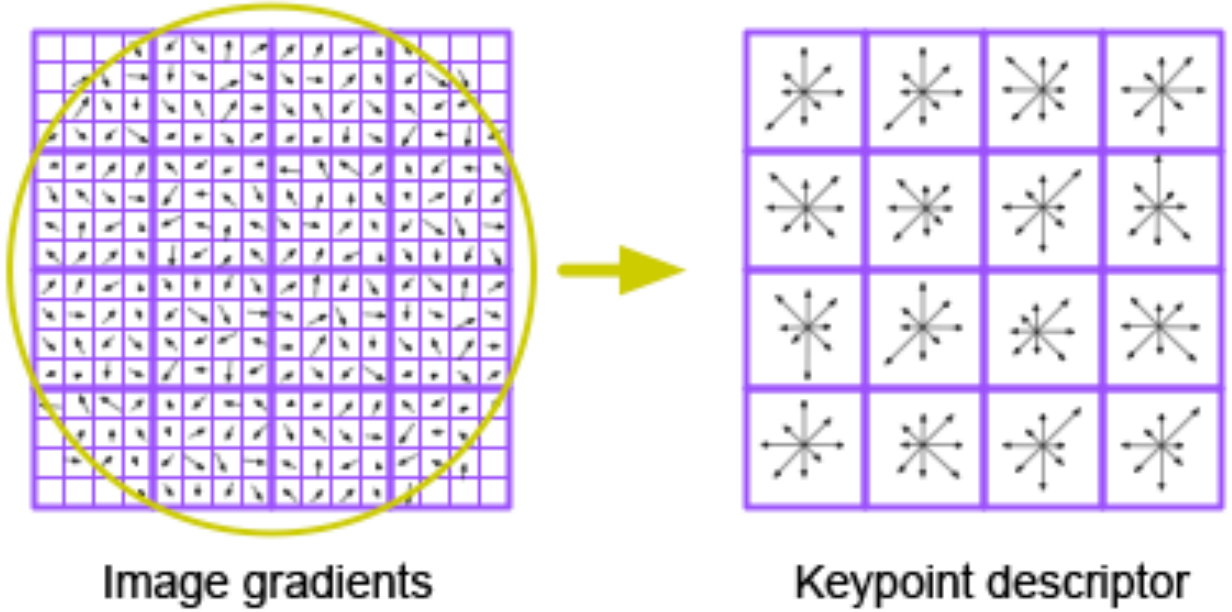
Finding a reference orientation

- Create histogram of local gradient directions in the patch
- Assign reference orientation at peak of smoothed histogram



SIFT descriptors

- Inspiration: complex neurons in the primary visual cortex



D. Lowe, Distinctive image features from scale-invariant keypoints, *IJCV* 60 (2), pp. 91-110, 2004

SIFT Descriptor

- Based on 16x16 image patch
- 4x4 subregions
- 8 bins in each subregion
- 4x4x8=128 dimensions in total

128-D Descriptor

- 16x16 Gradient window is taken. Partitioned into 4x4 subwindows.
- Histogram of 4x4 samples in 8 directions
- Gaussian weighting around center(σ is 1.5 imes that of the scale of a keypoint)
- 4x4x8 = 128 dimensional feature vector

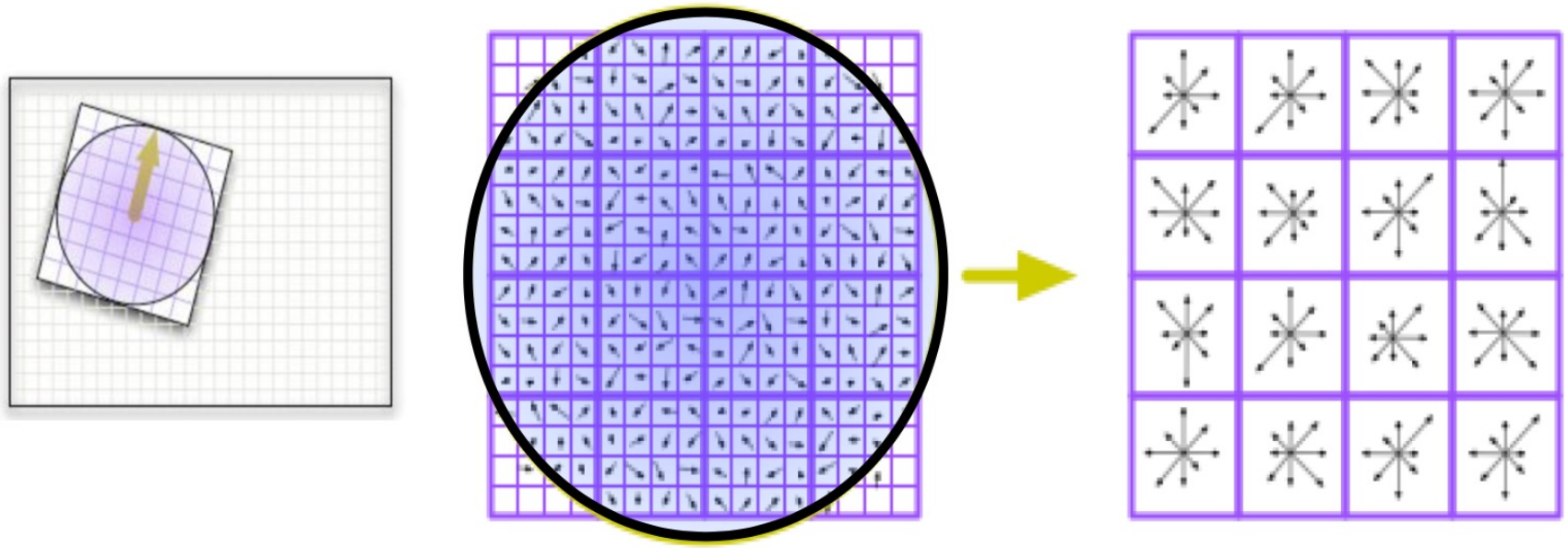
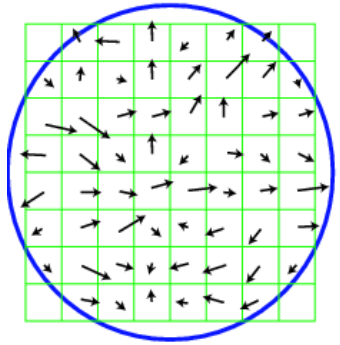


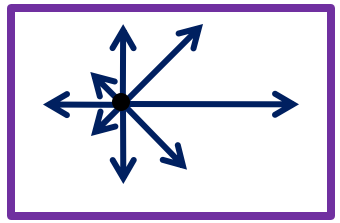
Image from: Jonas Hurrelmann

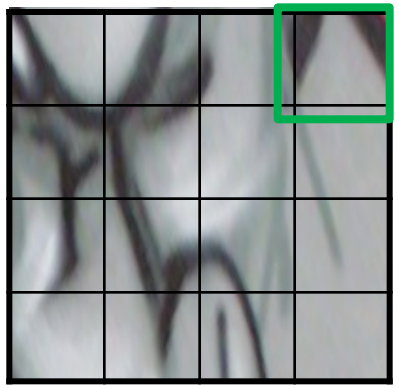
Scale Invariant Feature Transform (SIFT) descriptor [Lowe 2004]

- Use histograms to bin pixels within sub-patches according to their orientation.

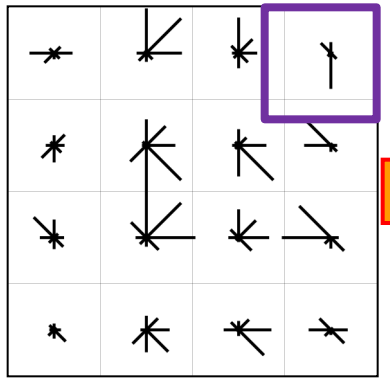


gradients





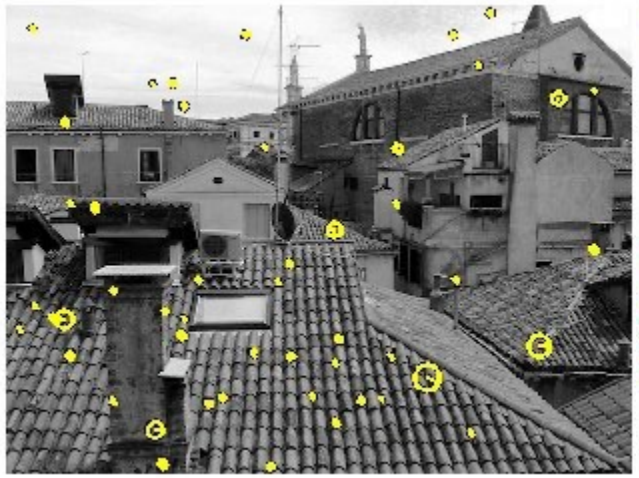
subdivided local patch



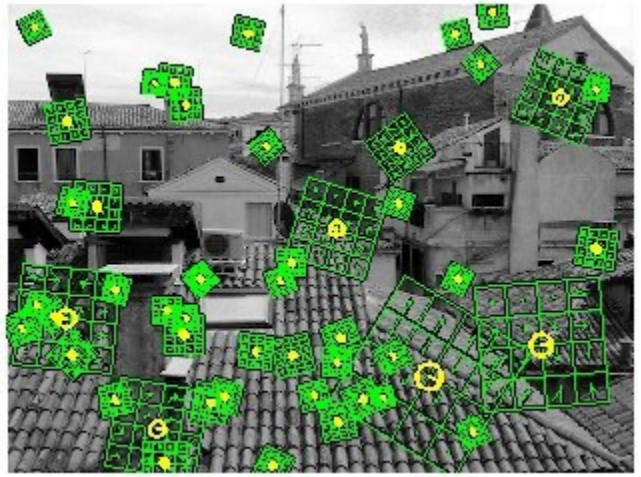
histogram per grid cell

Final descriptor = concatenation of all histograms, normalize

Scale Invariant Feature Transform (SIFT) descriptor [Lowe 2004]



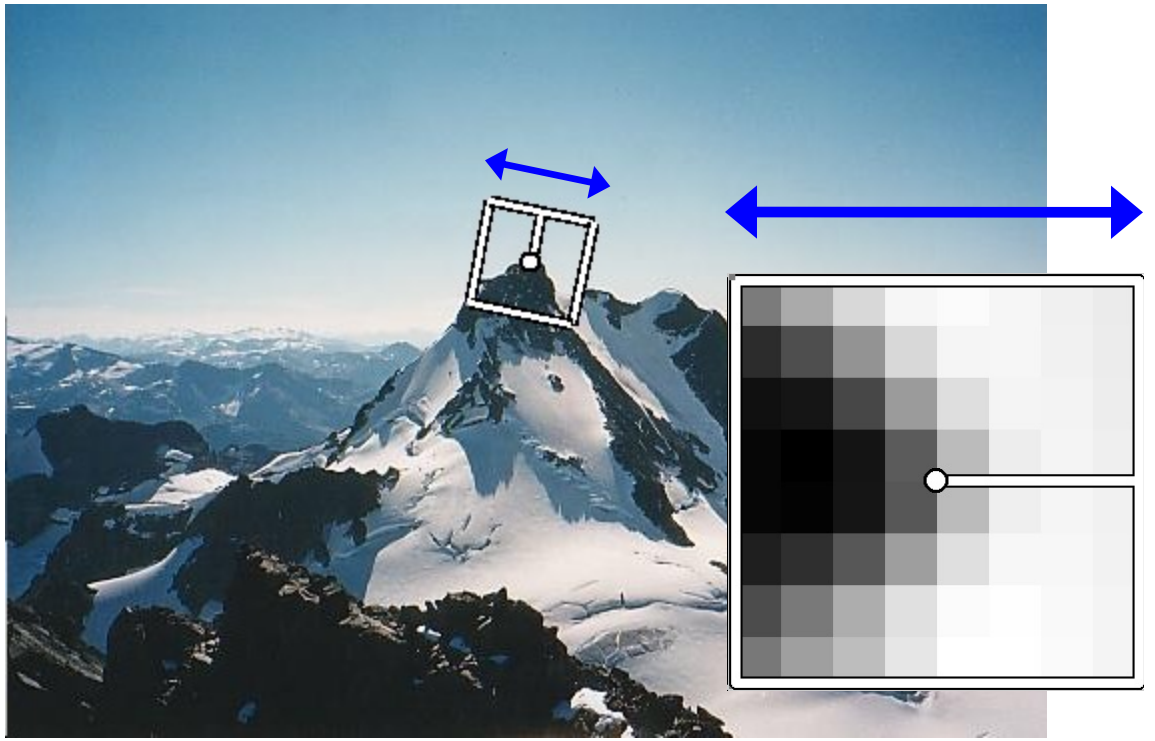
Interest points and their scales and orientations (random subset of 50)



SIFT descriptors

http://www.vlfeat.org/overview/sift.html

Making descriptor rotation invariant

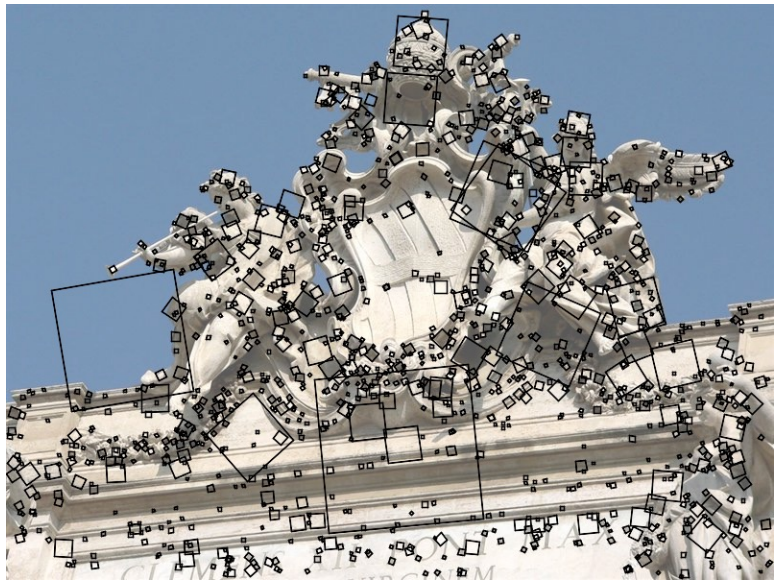


- Rotate patch according to its dominant gradient orientation
- This puts the patches into a canonical orientation.

Image from Matthew Brown

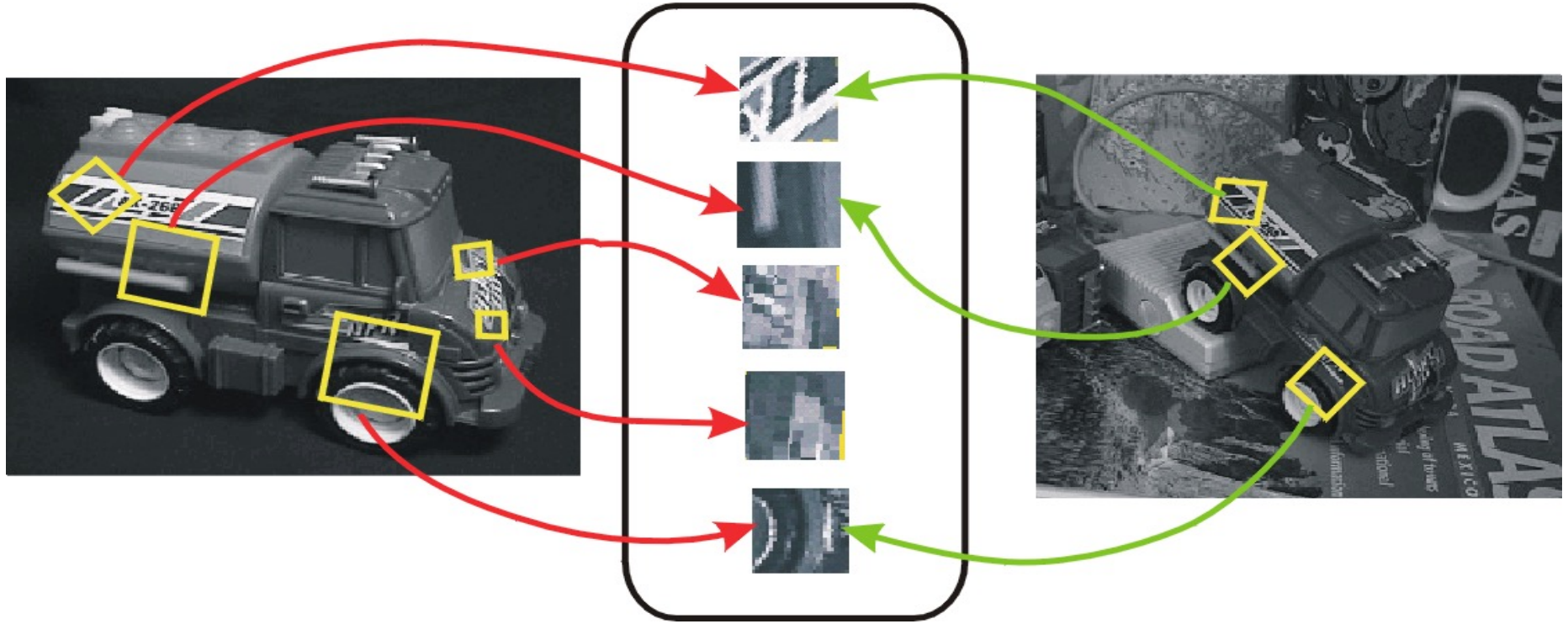
SIFT features

- Detected features with characteristic scales and orientations:



David G. Lowe. **"Distinctive image features from scale-invariant keypoints."** *IJCV* 60 (2), pp. 91-110, 2004.

From keypoint detection to feature description



- Detection is *covariant*:

  features(transform(image)) = transform(features(image))

- Description is *invariant*:

  features(transform(image)) = features(image)

SIFT descriptor performance

- Very robust to view angle chages.
  - 80% repeatability at:
    - 10% image noise
    - 45° viewing angle
    - 1k-100k keypoints in database

- Best *descriptor* according to [Mikolajczyk & Schmid 2005]'s extensive survey

Properties of SIFT

Extraordinarily robust detection and description technique

- Can handle changes in viewpoint
  - Up to about 60 degree out-of-plane rotation
- Can handle significant changes in illumination
  - Sometimes even day vs. night
- Fast and efficient—can run in real time
- Lots of code available

Other Scale/View-angle Invariant Features

- SURF
- Harris-Affine
- Ahrris-Laplacian
- MSER
- Daisy
- BRIEF
- ORB
- BRISK
- ……

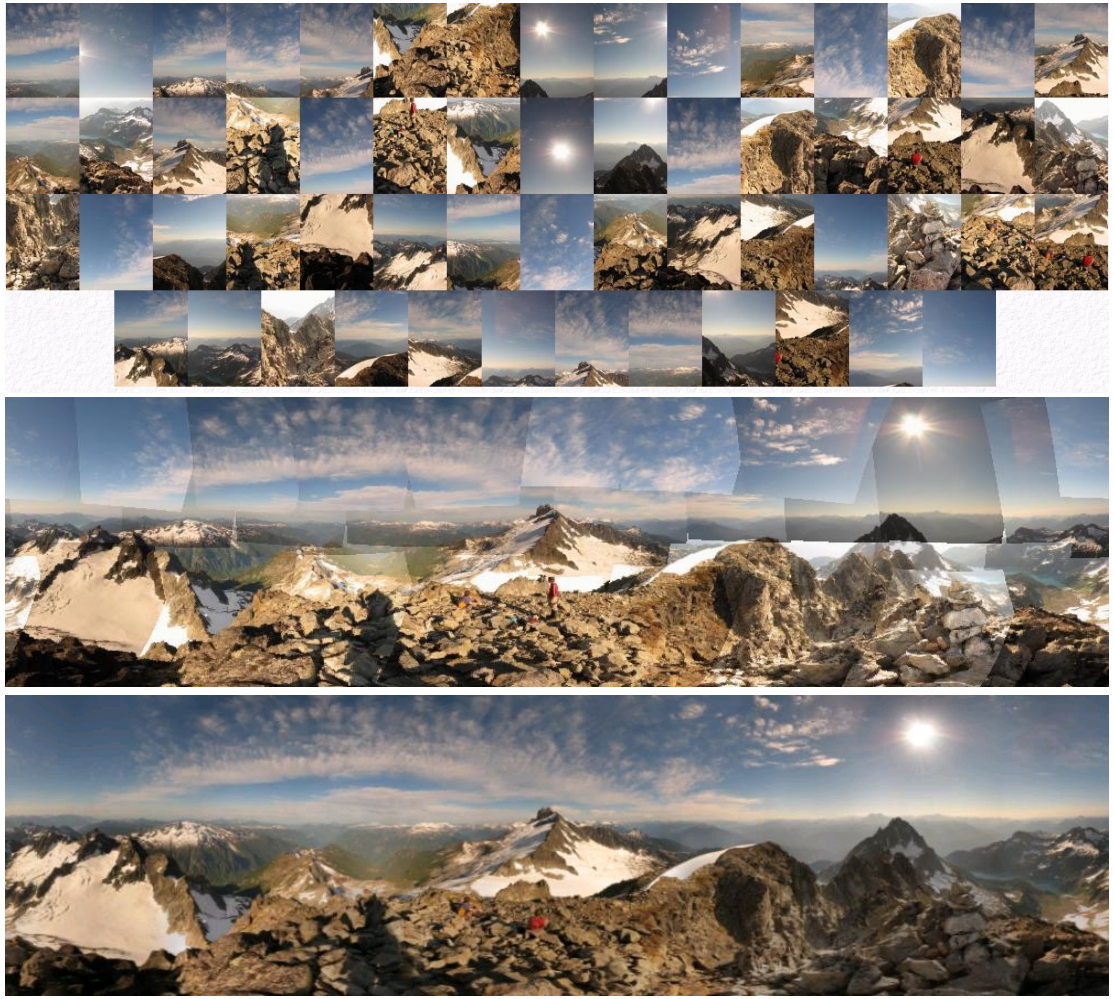
- For most local invariant feature detectors, executables are available online:

  – http://robots.ox.ac.uk/~vgg/research/affine

Applications of local invariant features

- Wide baseline stereo
- Motion tracking
- Panoramas
- Mobile robot navigation
- 3D reconstruction
- Recognition
- …

Automatic mosaicing

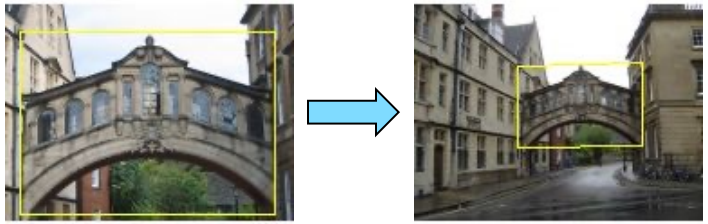


Matthew Brown
http://matthewalunbrown.com/autostitch/autostitch.html

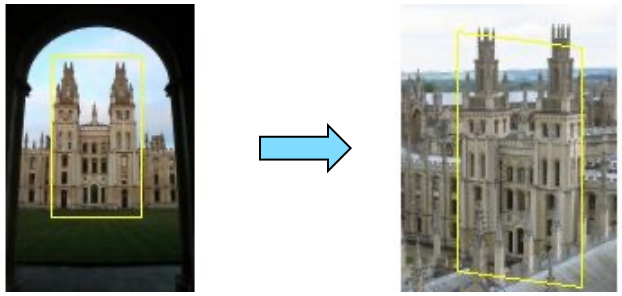
Wide baseline stereo

[Image from T. Tuytelaars ECCV 2006 tutorial]

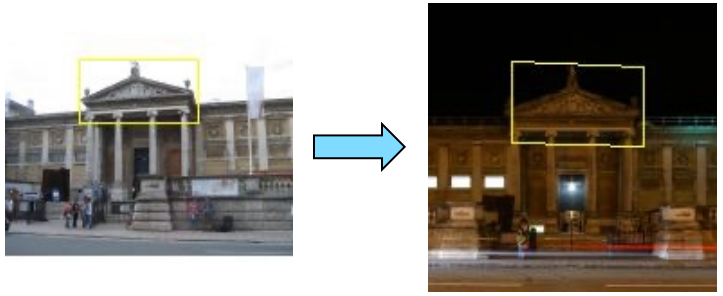
Recognition of specific objects, scenes



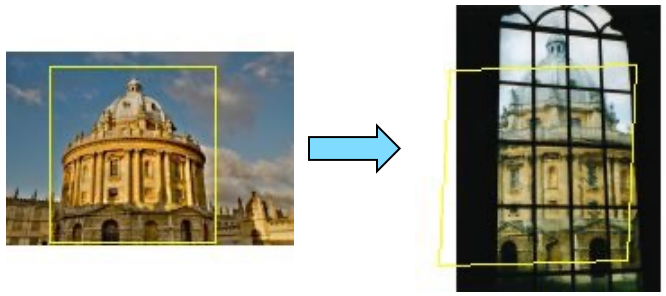
Scale



Viewpoint



Lighting



Occlusion

Slide credit: J. Sivic

Reading

- SIFT on Wikipedia.
- D. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60 (2), pp. 91-110, 2004. This paper contains details about efficient implementation of a Difference-of-Gaussians scale space.
- T. Lindeberg, "Feature detection with automatic scale selection," International Journal of Computer Vision 30 (2), pp. 77-116, 1998. This is advanced reading for those of you who are *really* interested in the mathematical details.