SIFT:

Scale Invariant Feature Transform

SIFT Detector, SIFT Descriptor

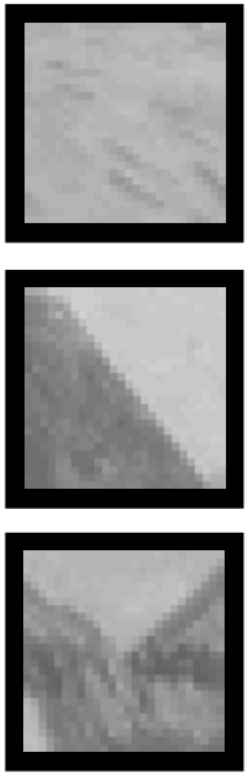
Announcement

- No Lab sessions this week
- No tutorial this week but we are going to have lectures for these two days (Tue. & Wed.)
- Lab Assignment One is due by this Sunday (due 11:59pm, 28th March 2021) (end of week 5)
- We are going to release the lab assignment two by this weekend.

Recall: Motivation

- Find “interesting” parts/pieces inside an image
  - e.g. corners, salient patches
  - Focus of attention, fixation.
  - Speed up computation.
  - Compress/extraction of information.
- Applications of interest points
  - Image Matching, Search
  - Object Detection, Object Recognition
  - Image Alignment & Stitching
  - Stereo
  - Tracking

Recall: Interest Points



isotropic structure: flat region

➡ not interesting, 0D, not useful for matching

linear structure: edges, lines

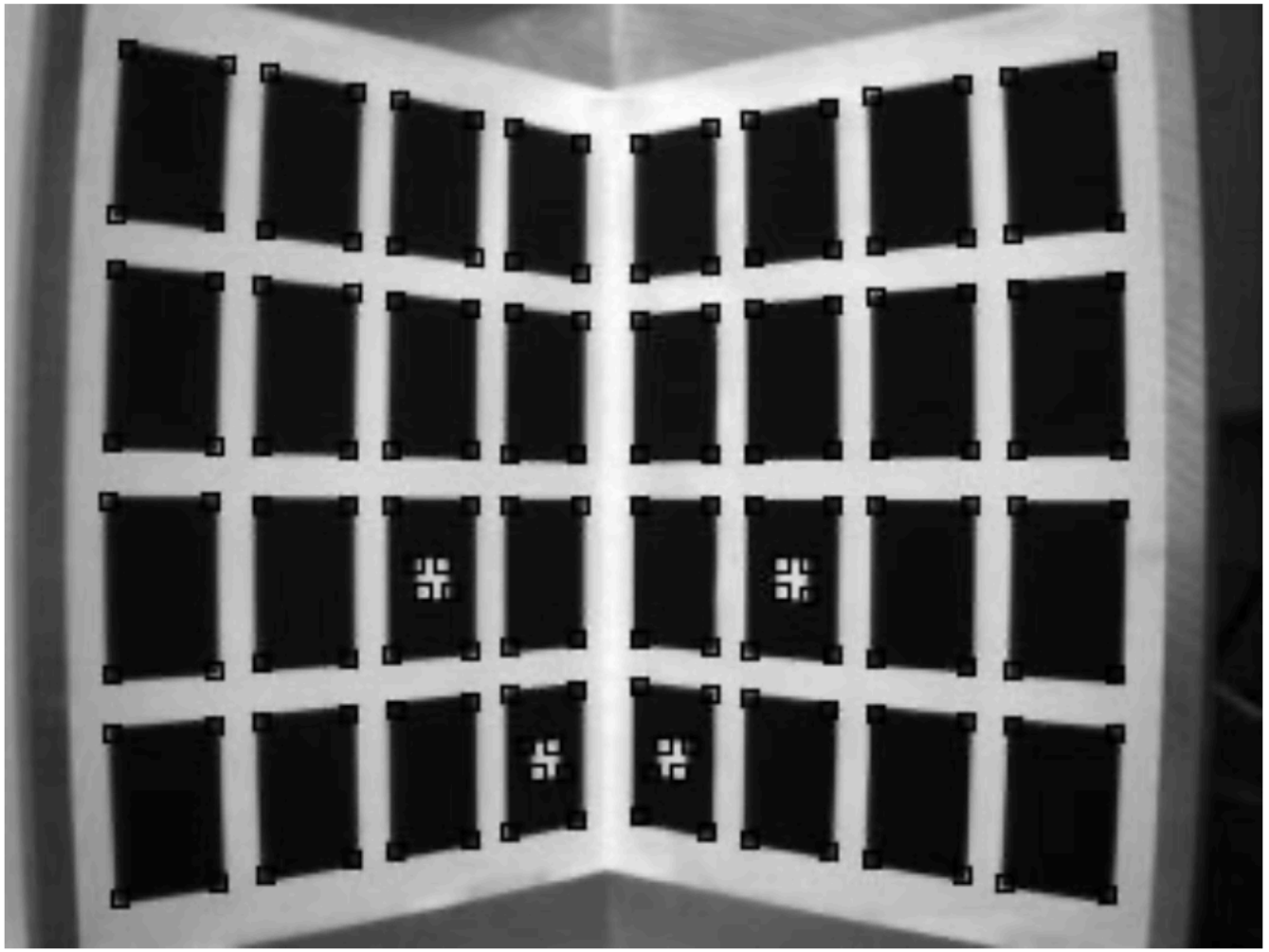
➡ edge, can be localized in 1D, subject to the aperture problem

bi-directional structure: corners

➡ corner, or **interest point**, can be localised in 2D, good for matching

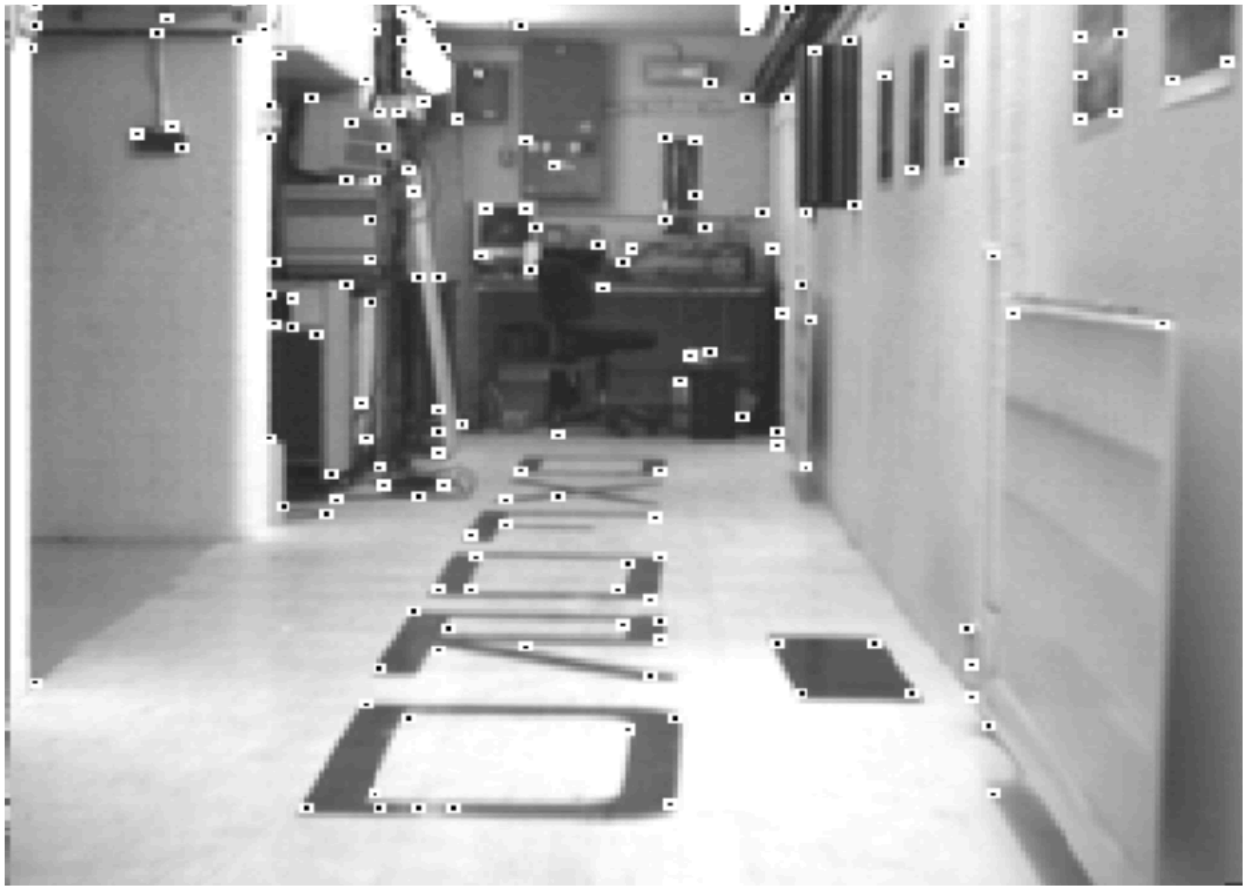
Interest Points have 2-directional structure.

Application: Corner Detection (for camera calibration)



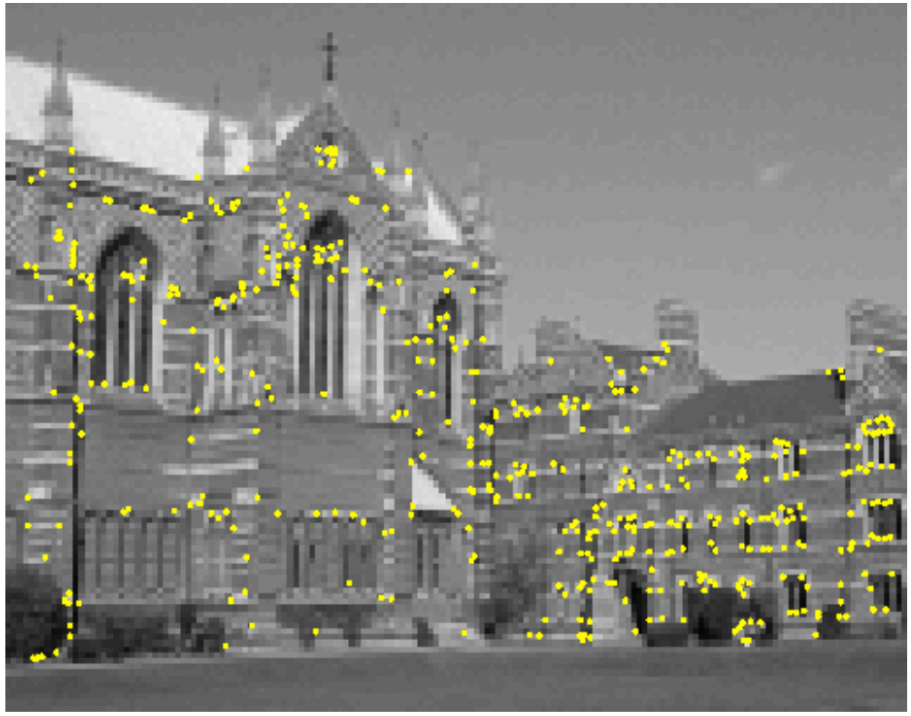
courtesy of B. Wilburn

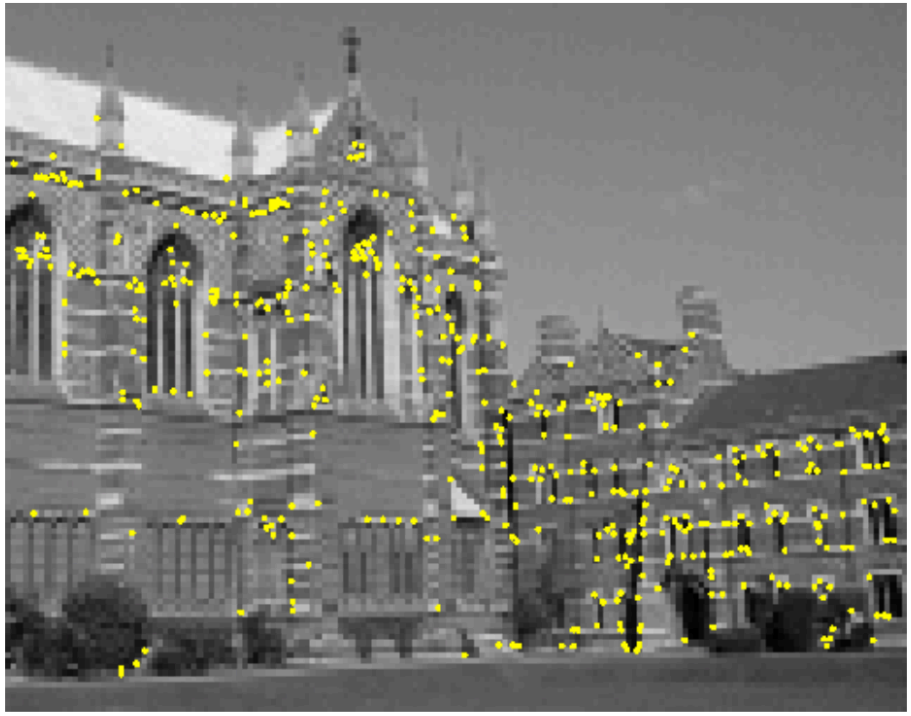
Application: Robot navigation



courtesy of S. Smith

Application: Matching between two images

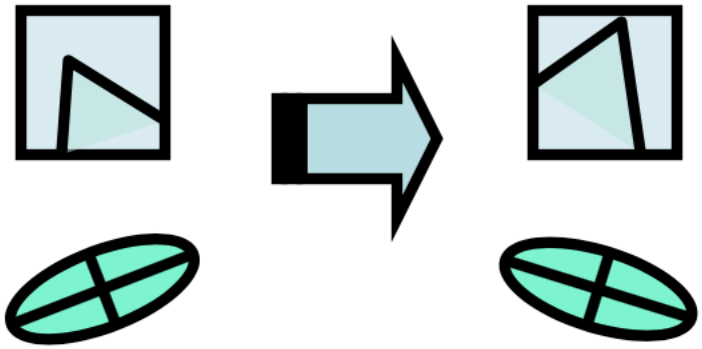




Interest points extracted with Harris (~ 500 points)

Harris Corner: Properties

- Harris corner is rotation-invariant.

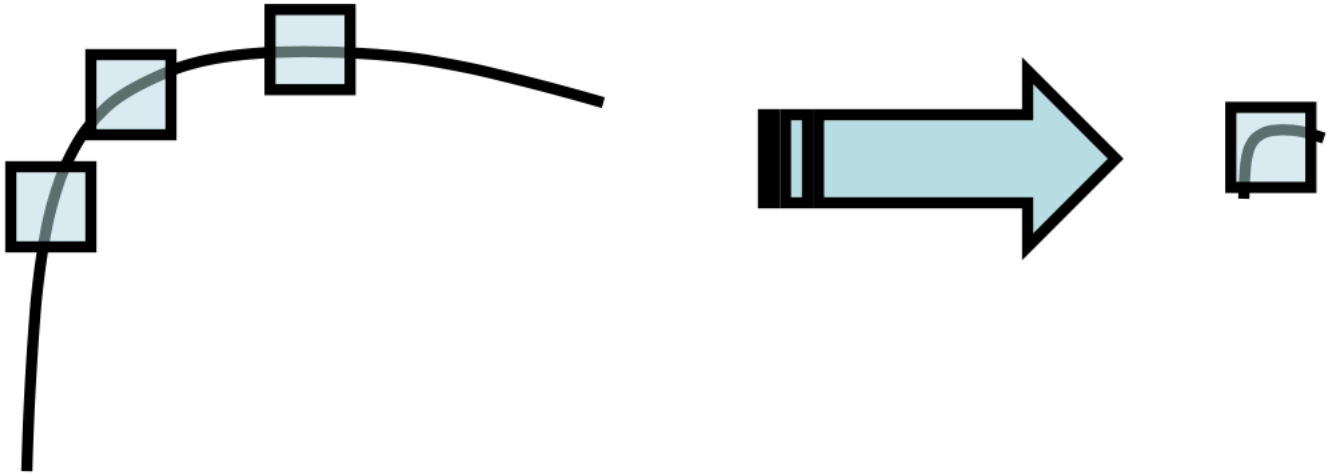


Ellipse rotates but its shape (i.e. eigenvalues) remains the same

Corner response R is invariant to image rotation

Harris Corner: Properties

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



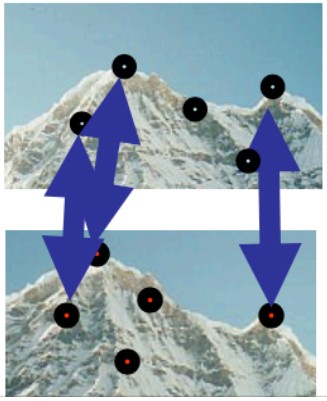
All points will be classified as edges

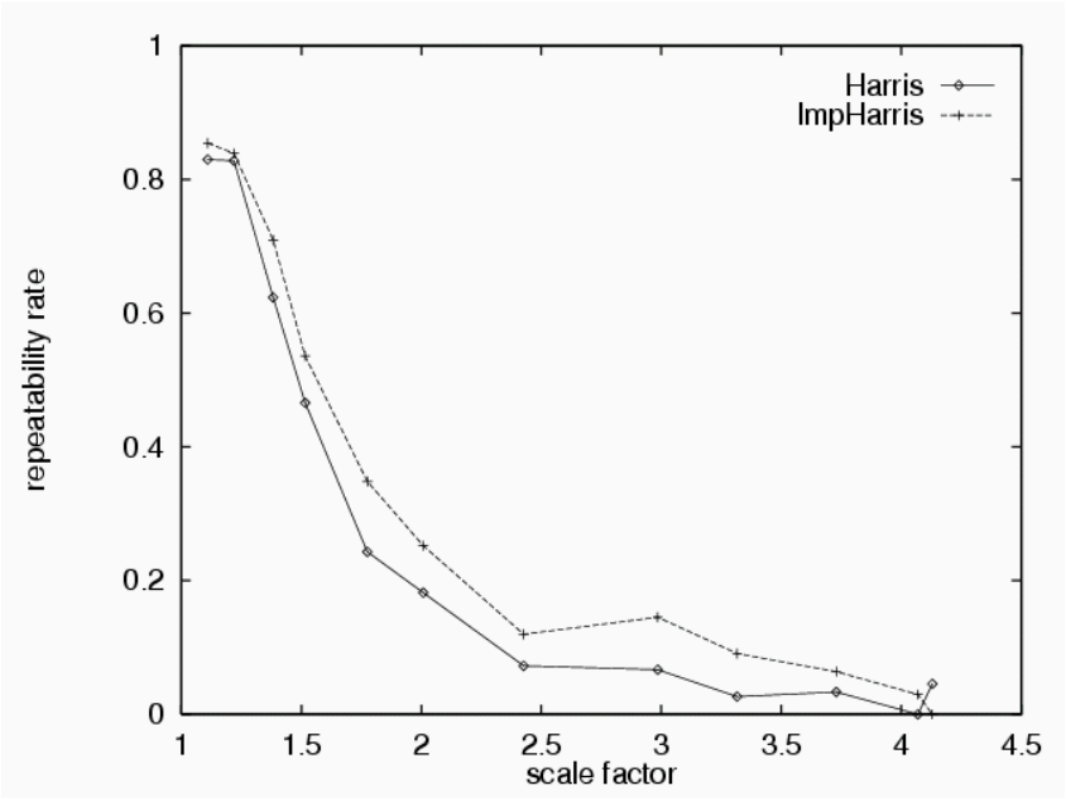
Corner !

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

Figure 10: Scale change sequence. The left image is the reference image. The scale change for the middle one is 1.5 and for the right one 4.1.

How to adapt to scale change ?

We want to:

detect *the same* interest points regardless of *image scale changes*

- Our goal is to be able to match an object in different images where the object appears in different scale, rotation, viewpoints, etc. How?

image 1

image 2

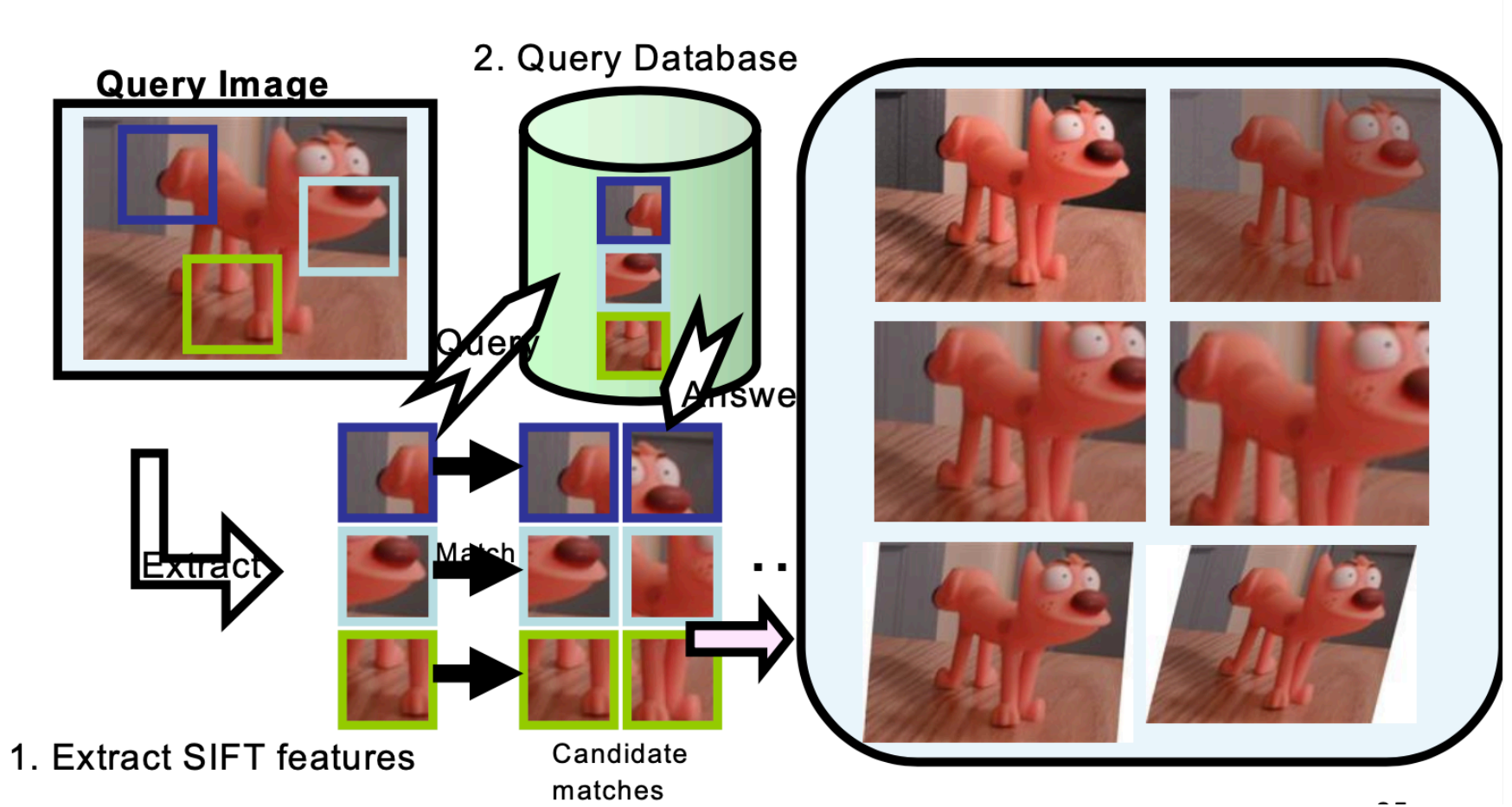
Figure: We want to be able to match these two objects / images

Cred: Sanja Fidler

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.

Content Based Image Retrieval (CBIR)



Reference (paper reading)

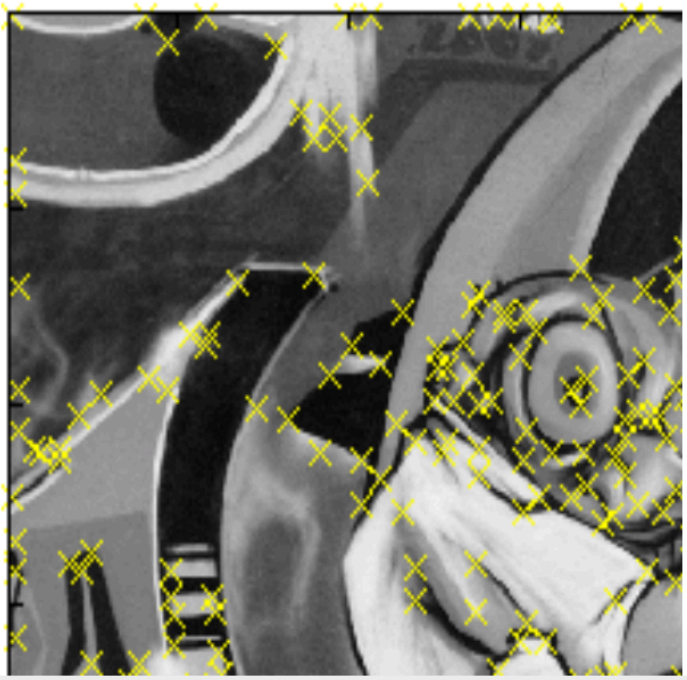
Journal paper (extension of the conference version):

Distinctive Image Features from Scale-Invariant Keypoints,

David G. Lowe, IJCV04

Comparison

Harris



SIFT

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

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 key point region.

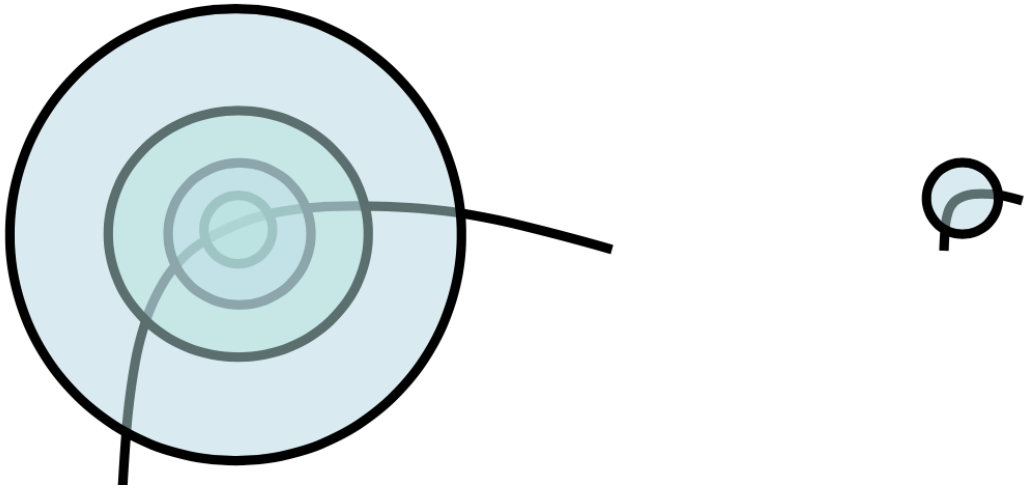
4. Keypoint description

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

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

Scale Invariant Detection

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



Automatic Scale Selection principle [Lindeberg98]

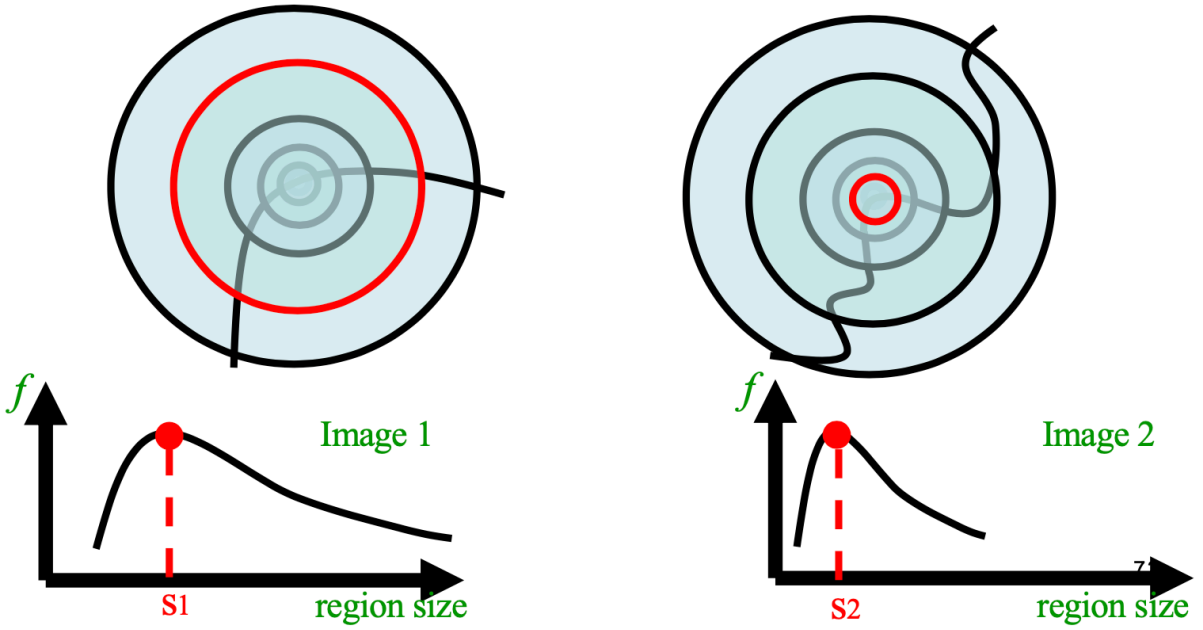
- Optimal scale selection principle:

– The scale at which some function (e.g. the normalized derivative) assumes an extreme value indicates a feature containing interesting pattern/structure.

Automatic Scale Selection

- **Principle:**

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

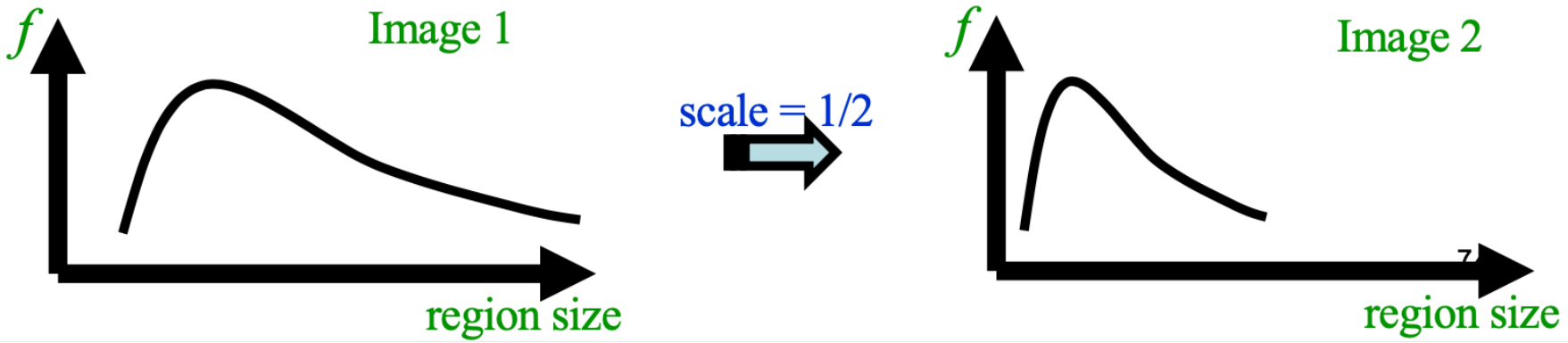


Automatic Scale Selection

- 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)



Scale Invariant Function

- Functions for determining scale

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

Kernels:

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

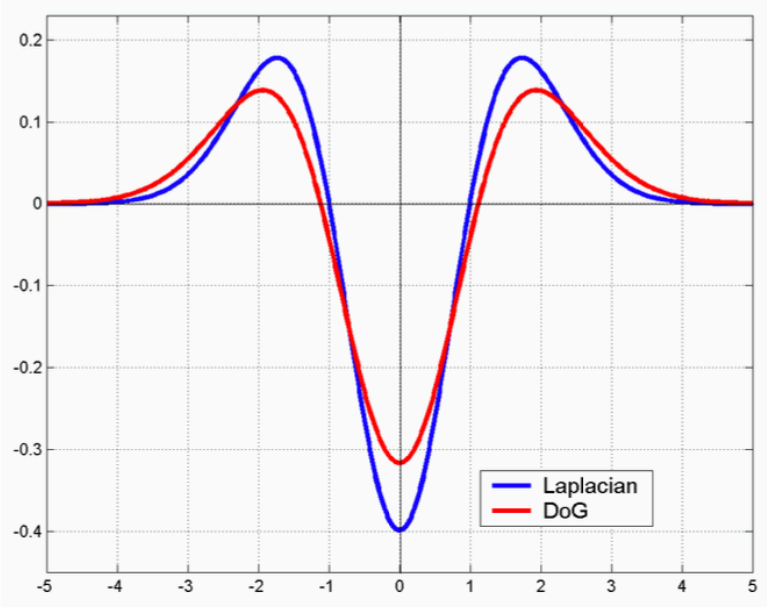
(Laplacian of Gaussian)

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

(Difference of Gaussians)

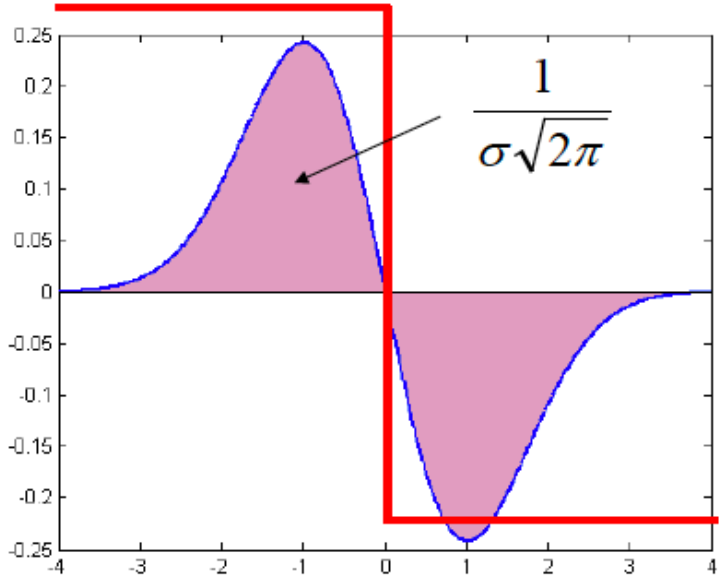
where Gaussian

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



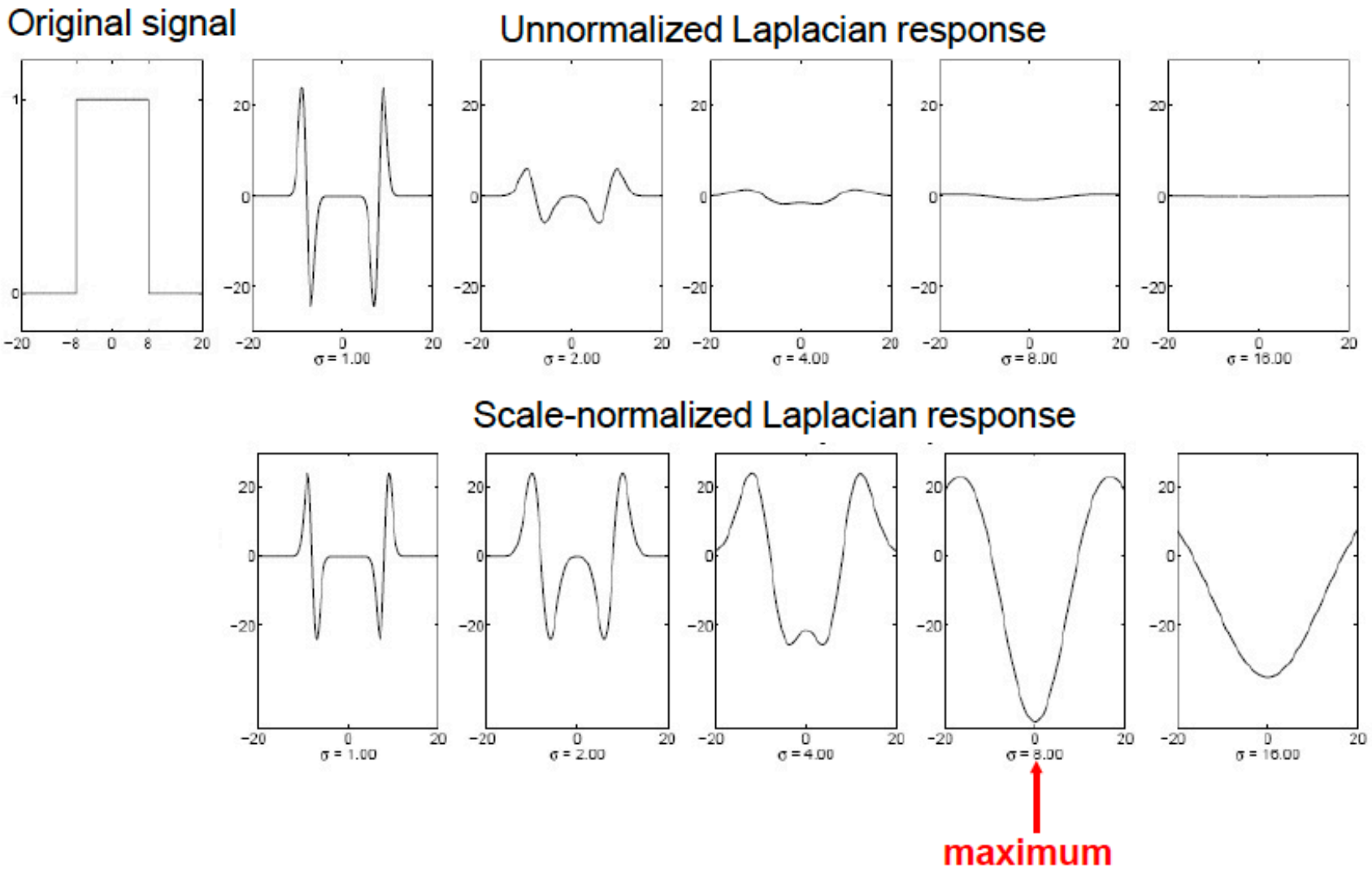
Scale invariant function

- The response of a derivative of Gaussian filter to a perfect step edge decreases as σ increases



- The response of a derivative of Gaussian filter to a perfect step edge decreases as σ increases
- To keep response the same (scale-invariant), must multiply Gaussian derivative by σ
- Laplacian is the second Gaussian derivative, so it must be multiplied by σ^2

Scale invariant function



Scale Invariant Function

- Functions for determining scale

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

Kernels:

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

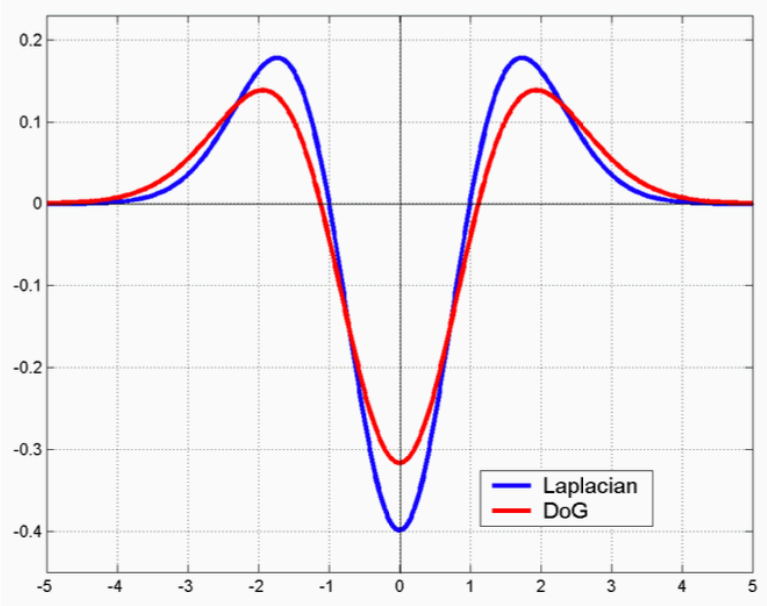
(Laplacian of Gaussian)

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

(Difference of Gaussians)

where Gaussian

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



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

Scale Invariant Function

- Laplacian of Gaussian is expensive
- DOG for approximation.

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)

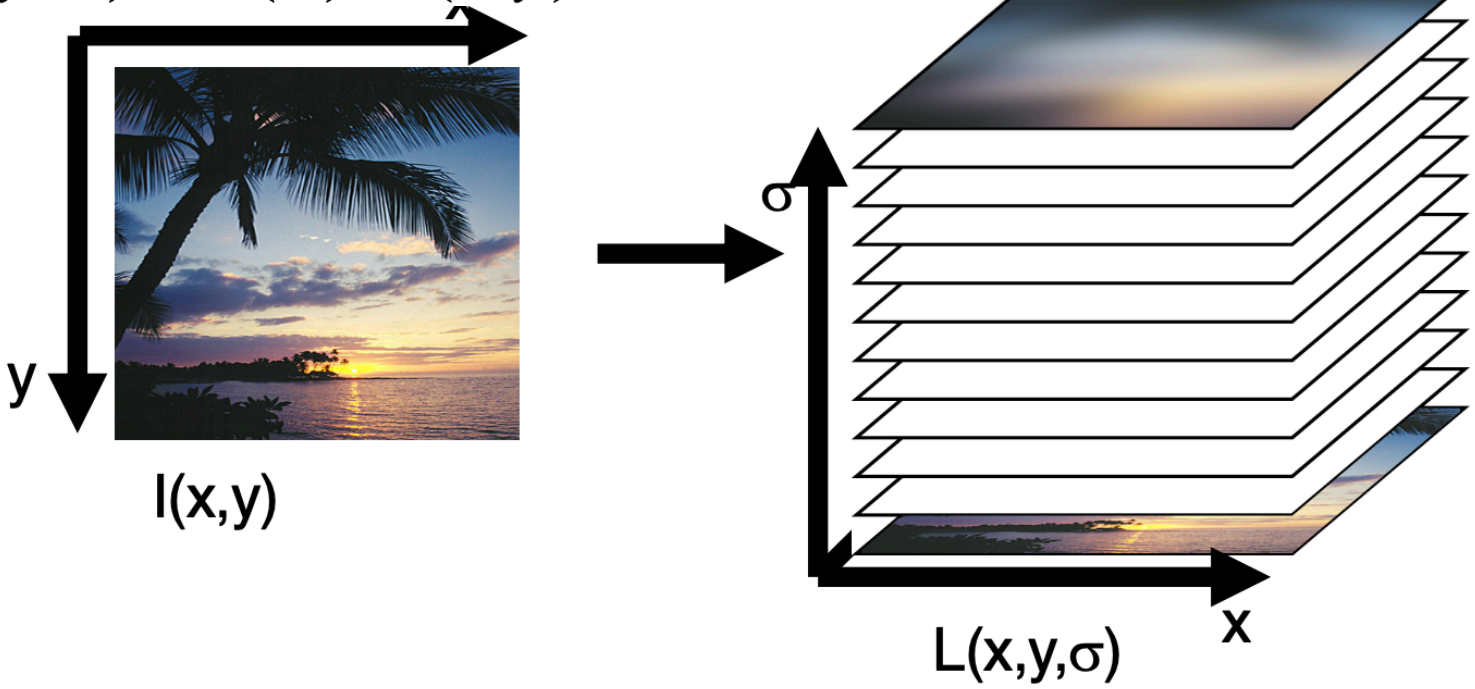
where Gaussian

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

Scale Space

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

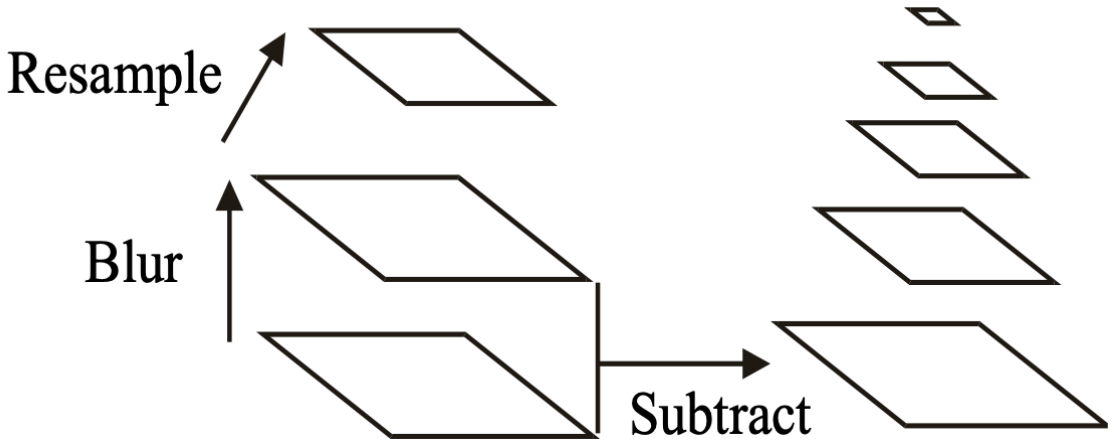
$$L(x, y; \sigma) = G(\sigma) * I(x, y)$$



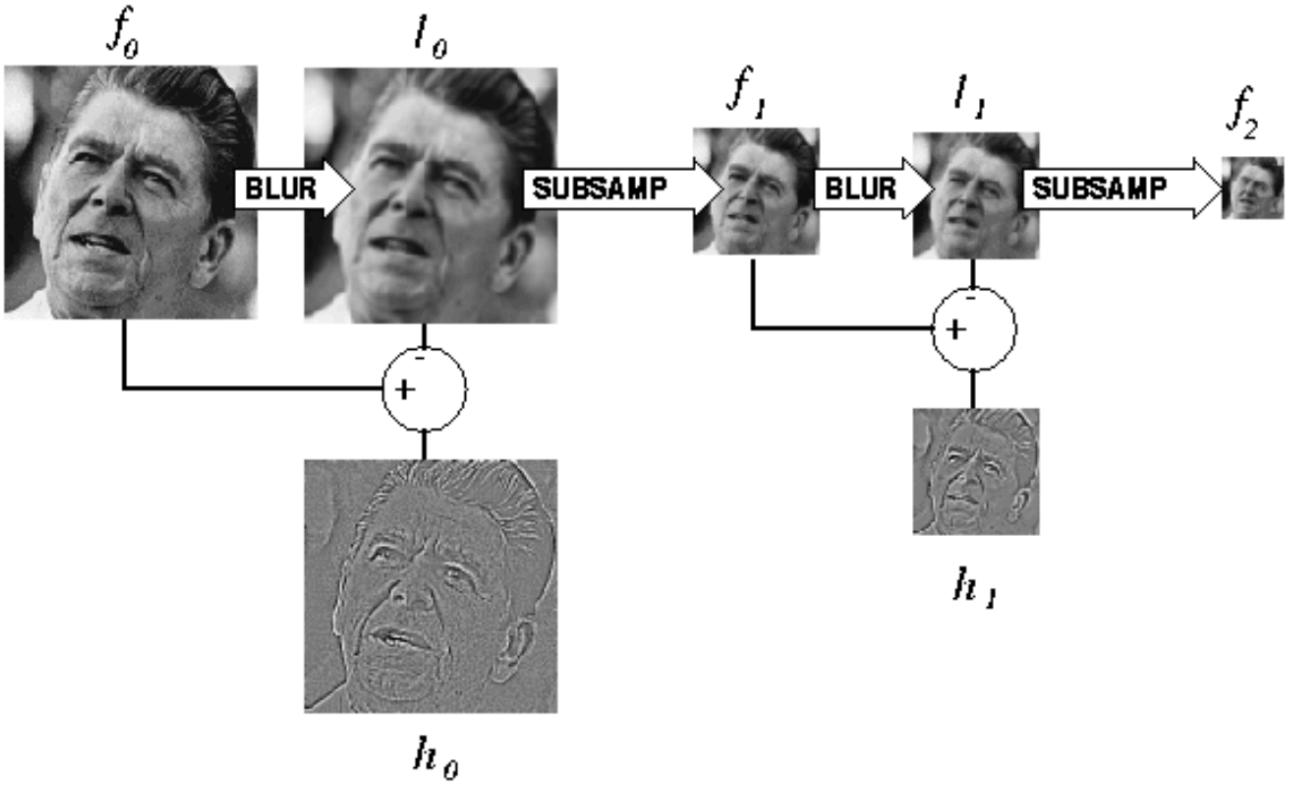
- Scale space is built on applying Gaussian kernel of varying standard deviation to the image.

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.

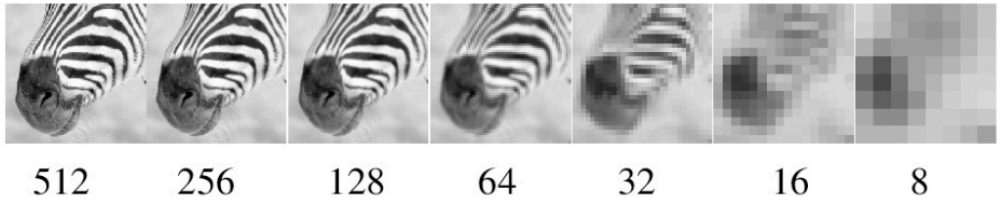


Compute Gaussian and Laplacian Pyramid

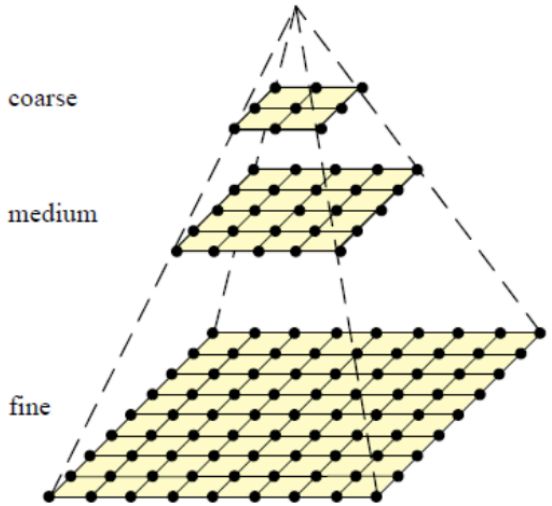


http://sepwww.stanford.edu/data/media/public/sep/morgan/texturematch/paper_html/node3.html#SECTION00012000000000000000

Gaussian Pyramid

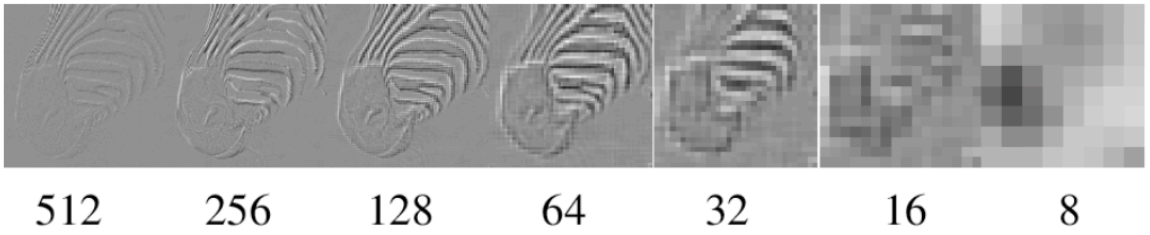


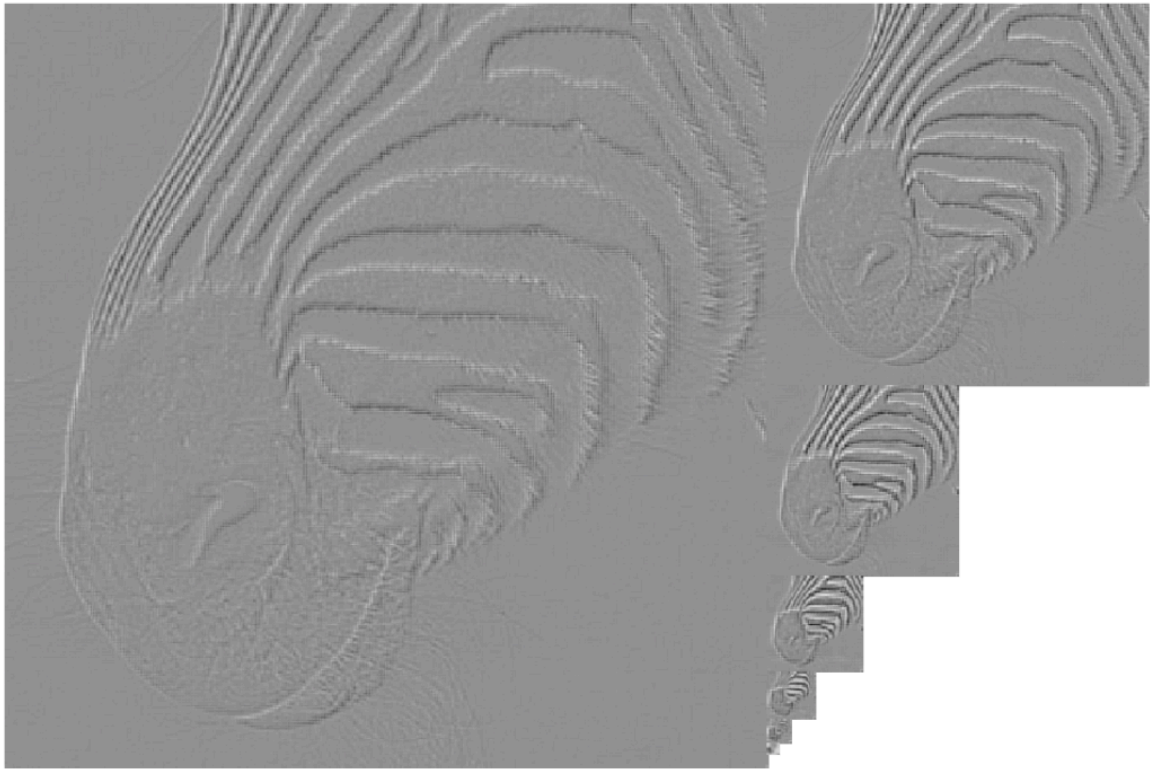
512 256 128 64 32 16 8



From **Forsyth**

Laplacian Pyramid





From Forsyth

Two Popular Scale Invariant Detectors

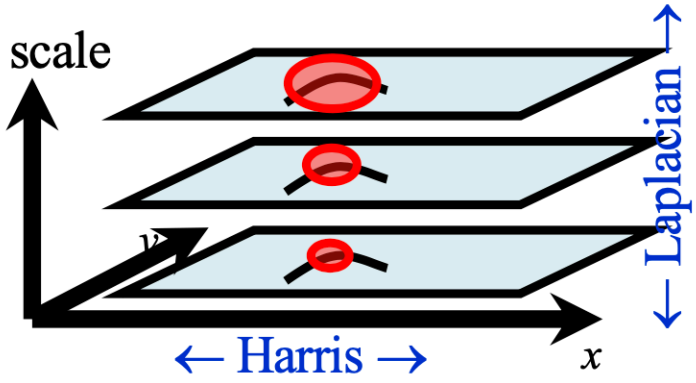
- Harris-Laplacian[1]

Find local maximum of:

  - Harris corner detector

in space (image coordinates)

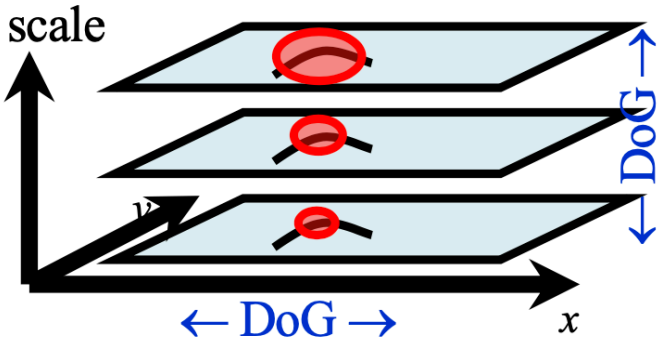
  - Laplacian in scale



- 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

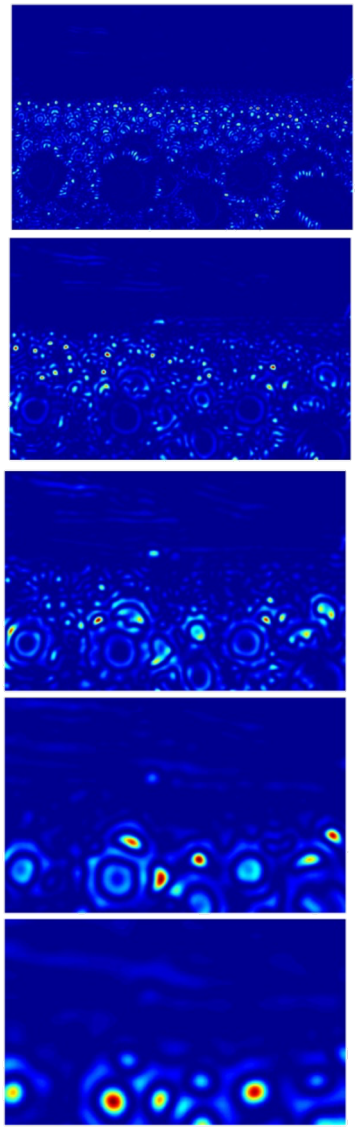
SIFT

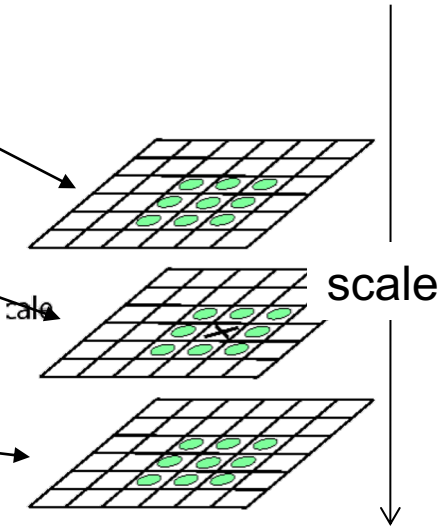
- Interest points are local maxima in both position and scale.

Scale invariant interest points

Interest points are local maxima in both position and scale.

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



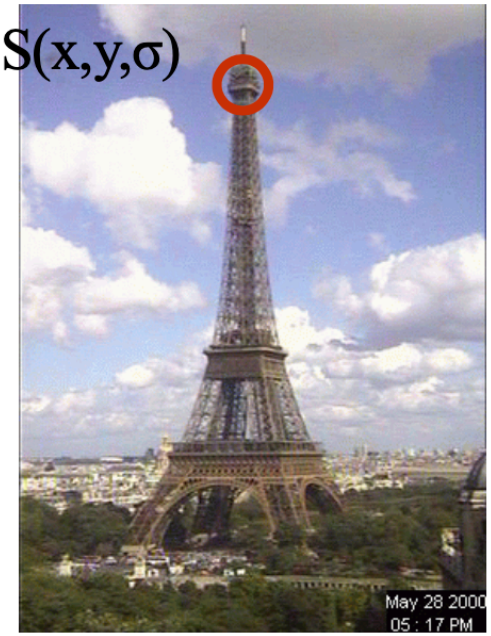


⇒ **List of** ***(x, y,*** σ***)***

Squared filter response maps

Slide credit: Kristen Grauman

Optimal Scale



Keypoint Spatial Localization

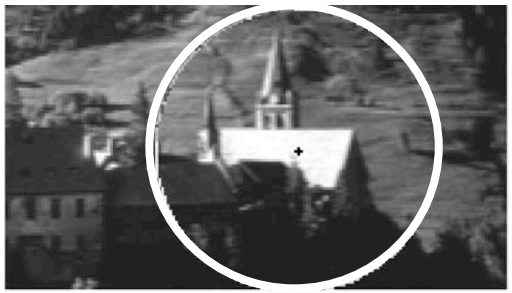
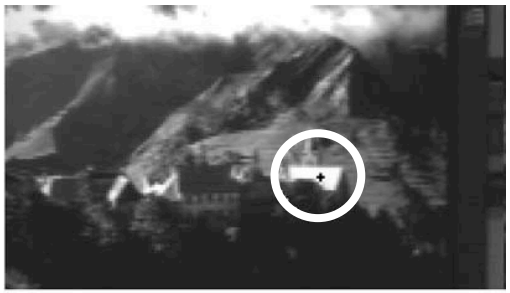
- There are still a lot of points, some of them are not good enough.
- The locations of keypoints may be not accurate.
- Eliminating edge points.

Removing Edge Points

- Such a point has large principal curvature acrossthe edge but a small one in the perpendicular direction
- The principal curvatures can be calculated from a Hessian function $H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$
- The eigenvalues of H are proportional to the principal curvatures, so two eigenvalues shouldn't have much difference

From feature detection to feature description

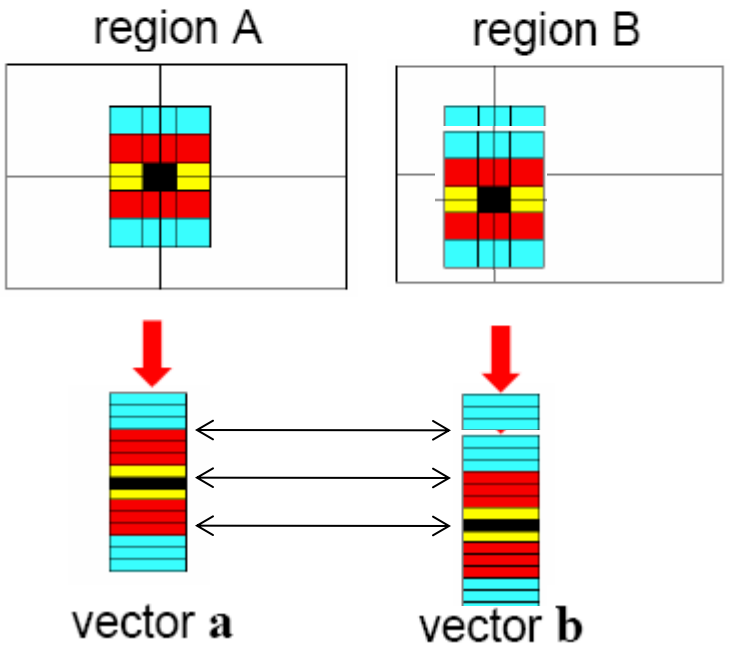
- To recognize the same pattern in multiple images, we need to match appearance "signatures" in the neighborhoods of extracted keypoints
  - But corresponding neighborhoods can be related by a scale change or rotation
  - We want to *normalize* neighborhoods to make signatures invariant to these transformations

Photometric transformations

Figure from T. Tuytelaars ECCV 2006 tutorial

Raw patches as local descriptors



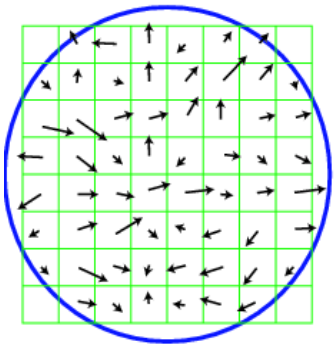
The simplest way to describe the neighborhood around an interest point is to write down the list of intensities to form a feature vector.

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

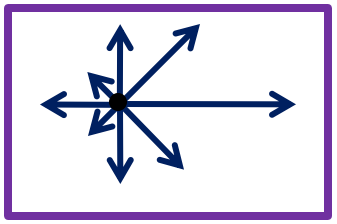
Figure: Andrew Zisserman

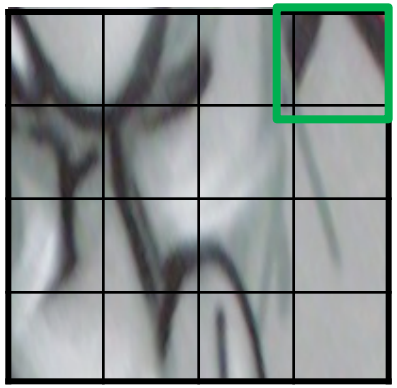
Scale Invariant Feature Transform (SIFT) descriptor [Lowe 2004]

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

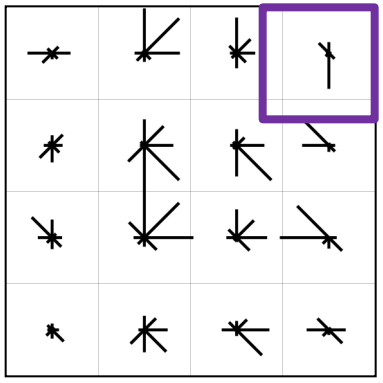


gradients





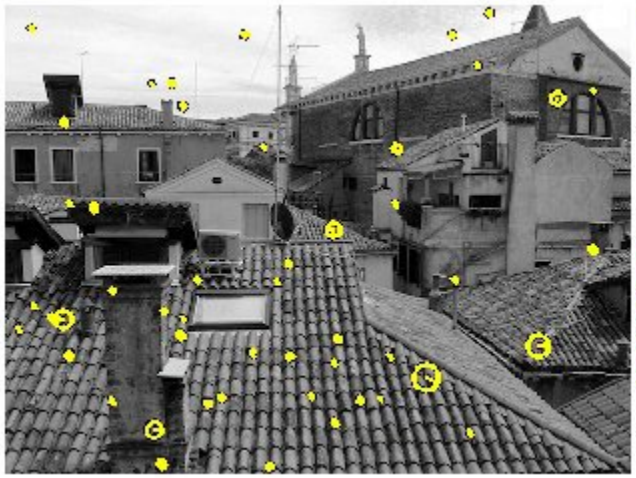
subdivided local patch



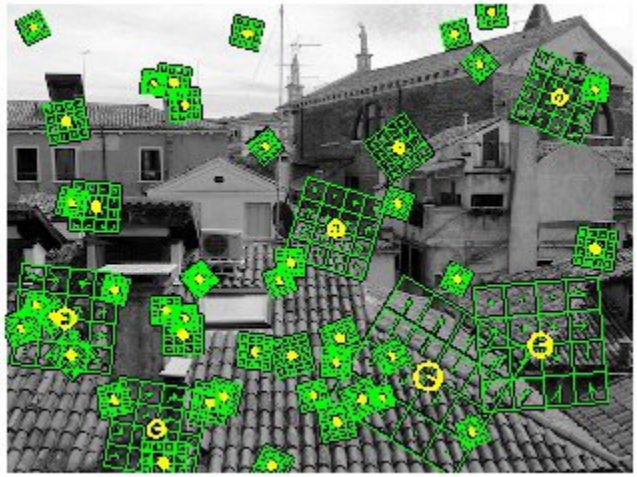
histogram per grid cell

Final descriptor = concatenation of all histograms, normalize

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



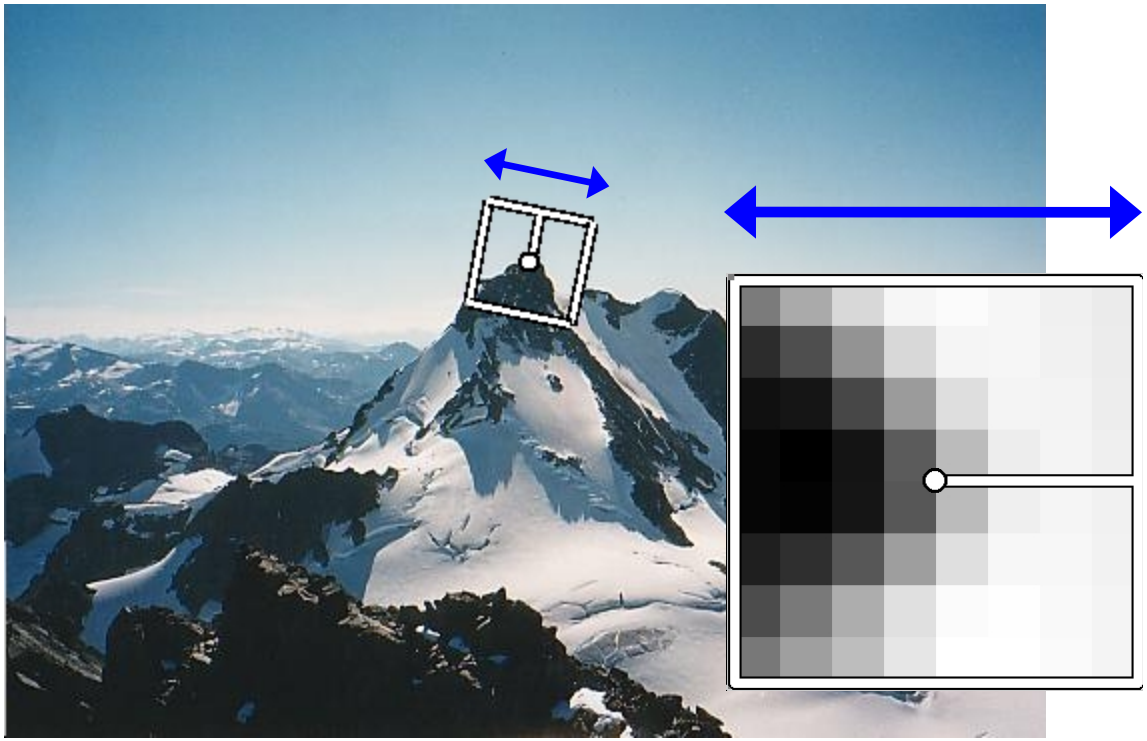
Interest points and their scales and orientations (random subset of 50)



SIFT descriptors

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

Making descriptor rotation invariant

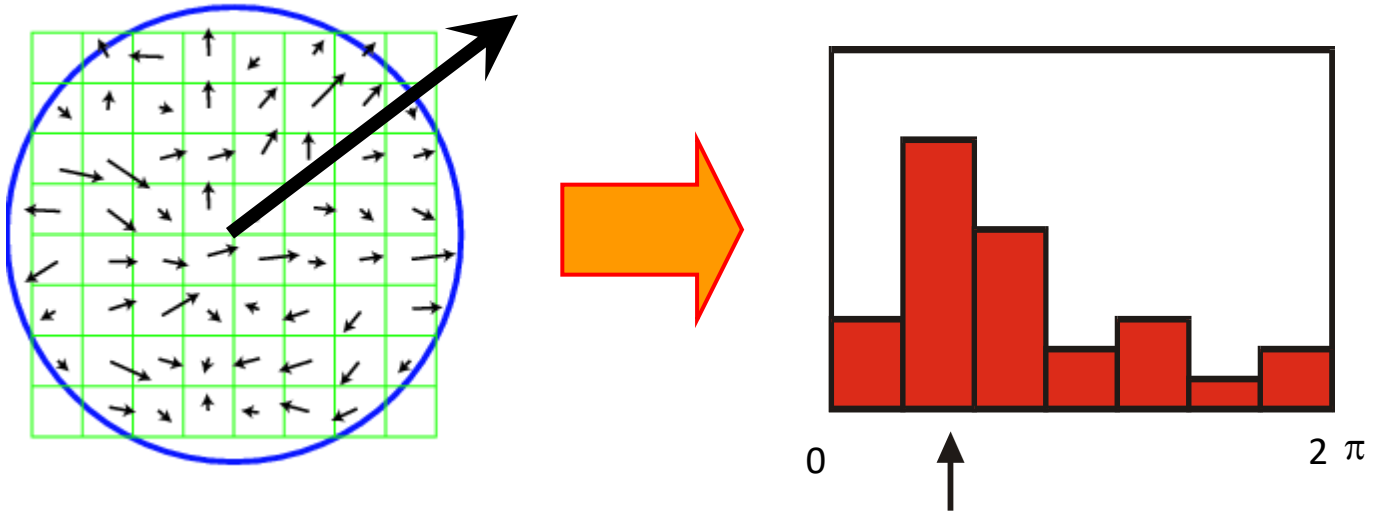


- Rotate patch according to its dominant gradient orientation
- This puts the patches into a canonical orientation.

Image from Matthew Brown

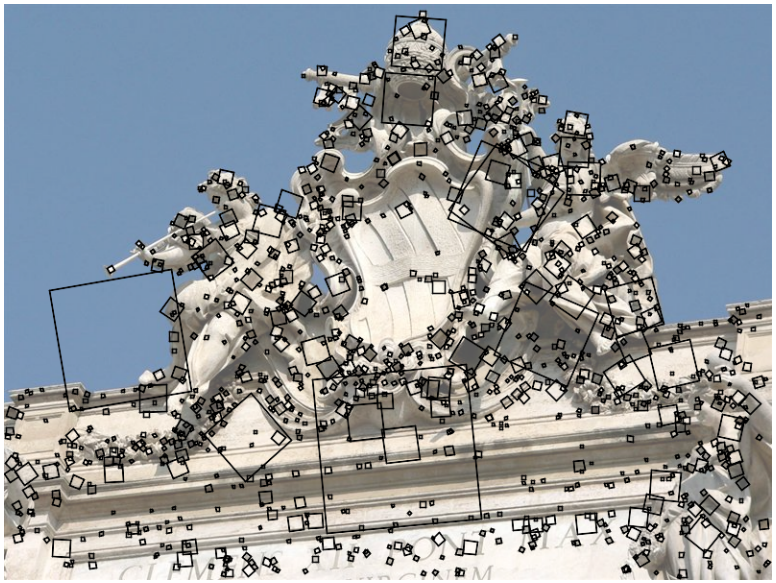
Finding a reference orientation

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



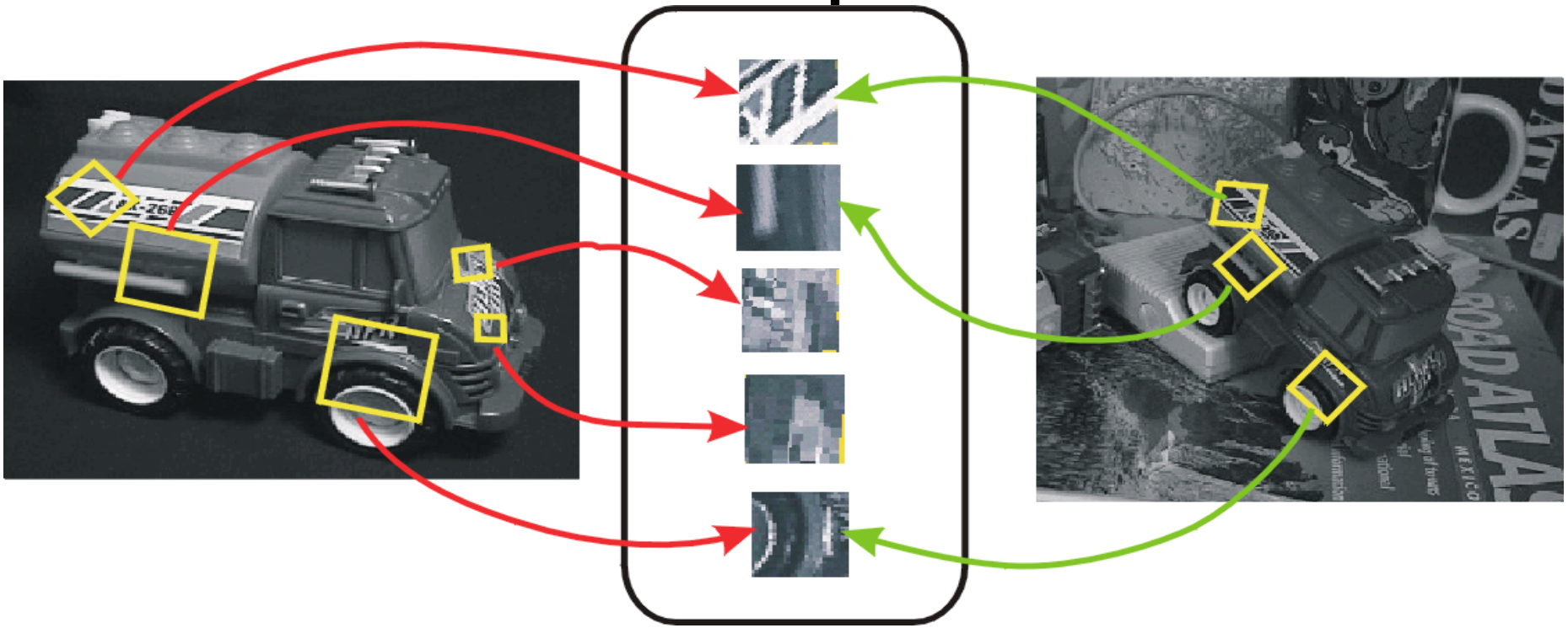
SIFT features

- 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*:

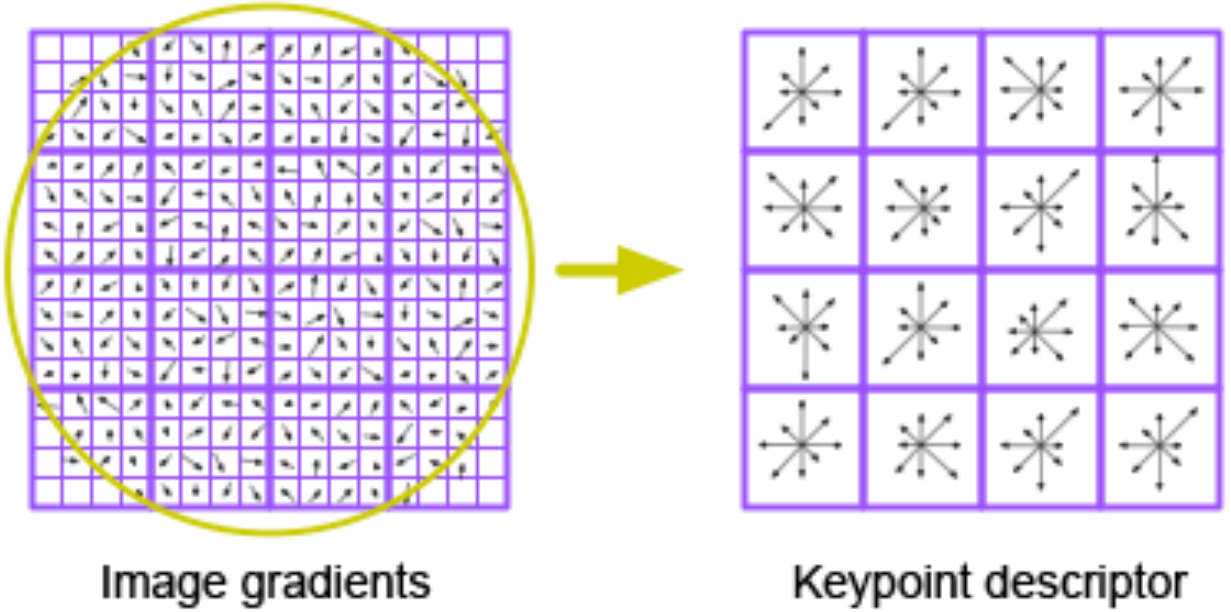
  features(transform(image)) = transform(features(image))

- Description is *invariant*:

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

SIFT 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 patches
- 4x4 subregions, each region is of 16x16 image patch
- 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 0.5 times that of the scale of a keypoint)
- 4x4x8 = 128 dimensional feature vector

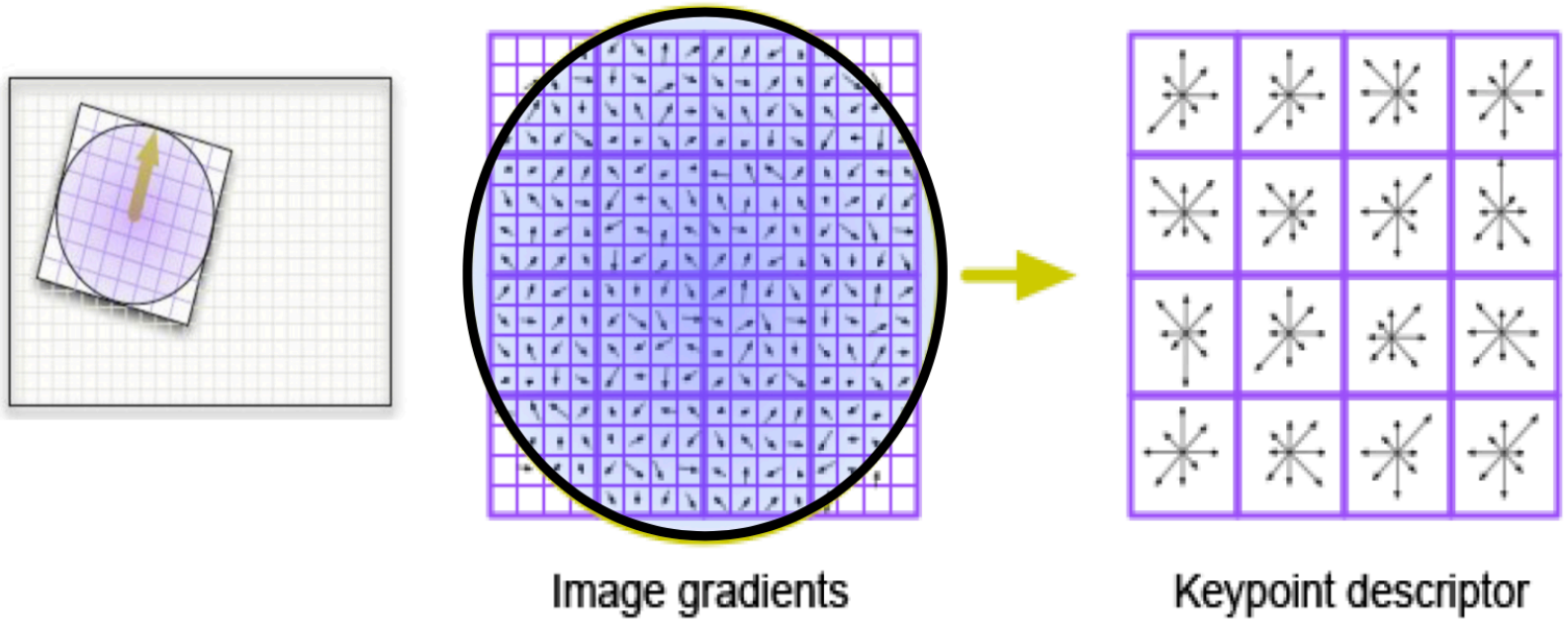


Image from: Jonas Hurrelmann

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
- ……

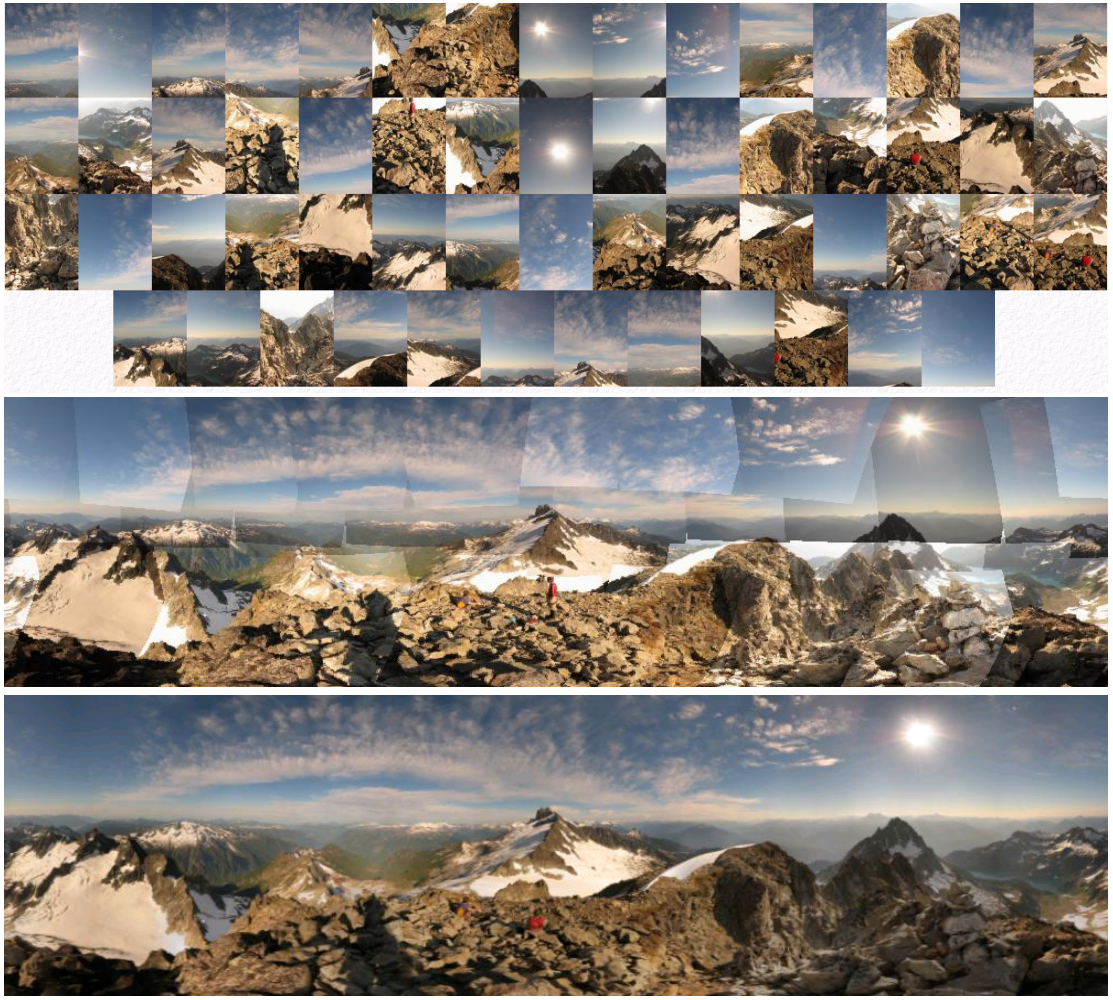
- For most local invariant feature detectors, executables are available online:

  - 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

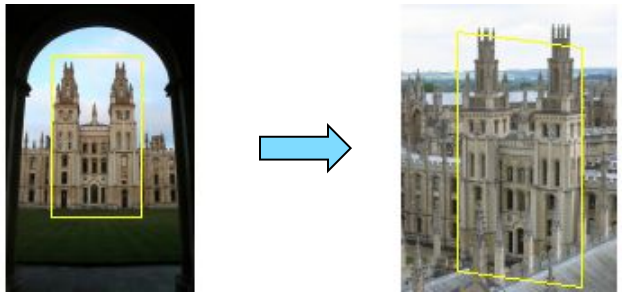
http://matthewalunbrown.com/autostitch/autostitch.html

Wide baseline stereo

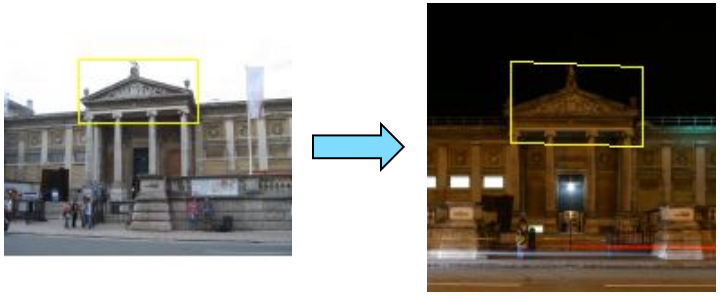
[Image from T. Tuytelaars ECCV 2006 tutorial]

Recognition of specific objects, scenes

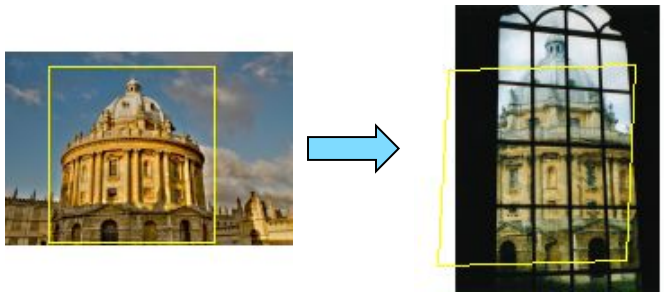
Scale



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.