Image Features: Descriptors and Matching

adapted from CSE 576
by Richard Szeliski

Motivation

Features are used for:

- Image alignment (homography, fundamental matrix)
- 3D reconstruction
- Motion tracking
- Object recognition
- Indexing and database retrieval
- Robot navigation
- … other

OBS: Computer Vision and Machine Learning…

Outline

- Feature detectors

- Feature descriptors

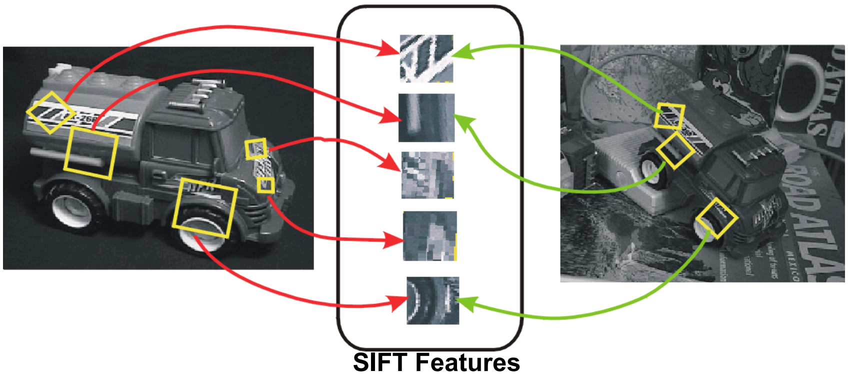
Outline

- Feature detectors
  - scale and affine invariant (points, regions)
  - selection of features

- Feature descriptors
  - patches, oriented patches
  - SIFT (orientations)

Invariant Local Features

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



Advantages of local features

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

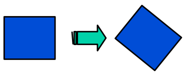
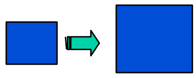
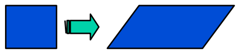
Quantity: many features can be generated for even small objects

Efficiency: close to real-time performance

Extensibility: can easily be extended to wide range of differing feature types, with each adding robustness

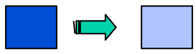
Models of Image Change

Geometry

- Rotation 
- Similarity (rotation + uniform scale) 
- Affine (scale dependent on direction) 
  valid for: orthographic camera, locally planar object

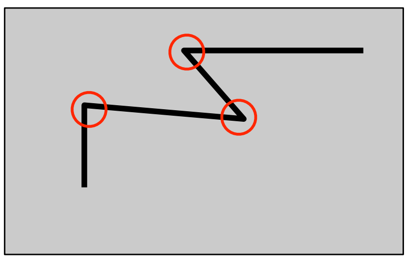
Photometry

- Affine intensity change ($I \rightarrow a\,I + b$)



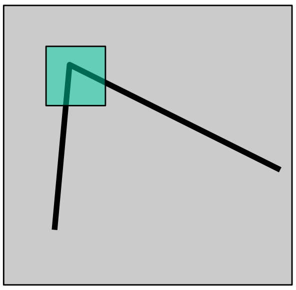
Harris corner detector

C.Harris, M.Stephens.
“A Combined Corner and Edge Detector”.
1988

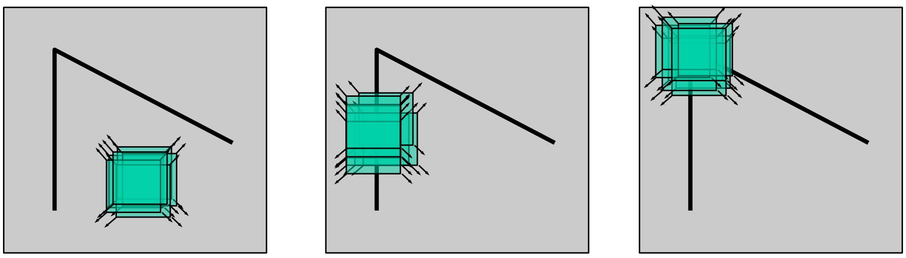


The Basic Idea

- We should easily recognize the point by looking through a small window
- Shifting a window in *any direction* should give *a large change* in intensity



Harris Detector: Basic Idea



"flat" region:
no change in
all directions

"edge":
no change along
the edge direction

"corner":
significant change
in all directions

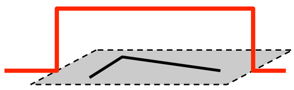
Harris Detector: Mathematics

Change of intensity for the shift $[u,v]$:

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

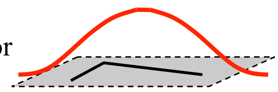
Window function   Shifted intensity   Intensity

Window function $w(x,y)$ =



1 in window, 0 outside

or



Gaussian

Harris Detector: Mathematics

For small shifts $[u,v]$ we have a *bilinear* approximation:

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

where M is a 2×2 matrix computed from image derivatives:

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

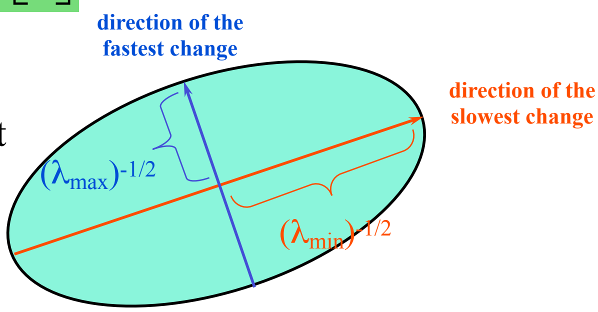
Harris Detector: Mathematics

Intensity change in shifting window: eigenvalue analysis

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

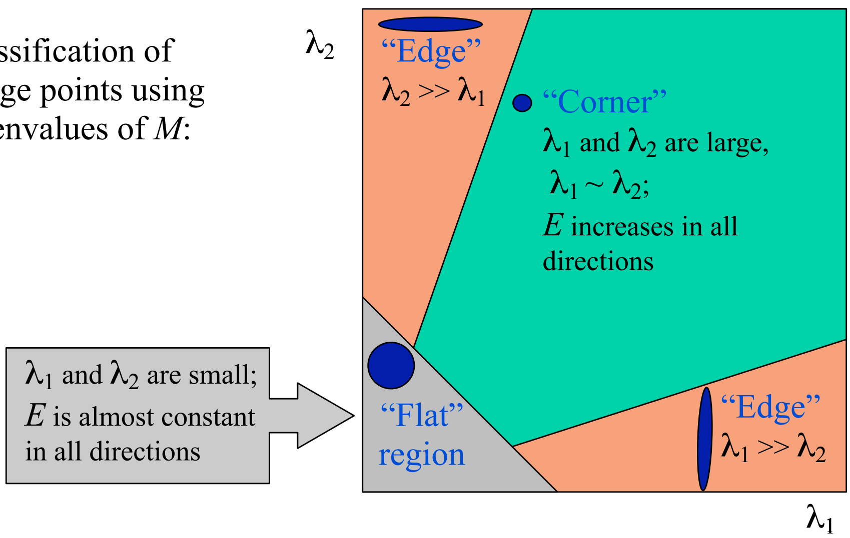
λ_1, λ_2 – eigenvalues of M

Ellipse $E(u,v)$ = const



Harris Detector: Mathematics

Classification of image points using eigenvalues of *M*:



Harris Detector: Mathematics

Measure of corner response:

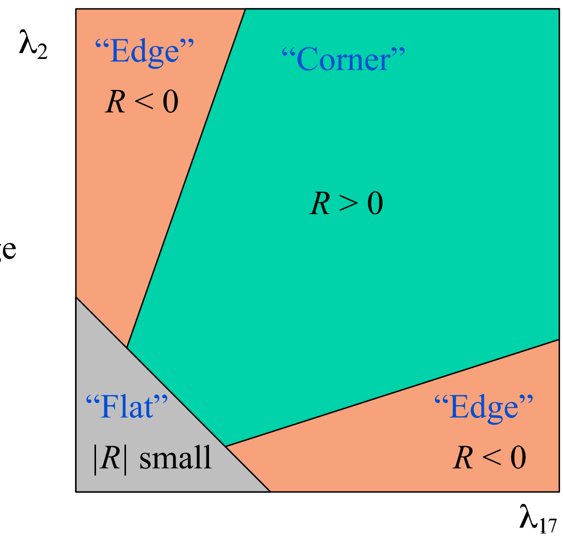
$$R = \det M - k\left(\operatorname{trace} M\right)^2$$

$$\det M = \lambda_1 \lambda_2$$
$$\operatorname{trace} M = \lambda_1 + \lambda_2$$

(k – empirical constant, k = 0.04-0.06)

Harris Detector: Mathematics

- R depends only on eigenvalues of M
- R is large for a corner
- R is negative with large magnitude for an edge
- $|R|$ is small for a flat region



Harris Detector

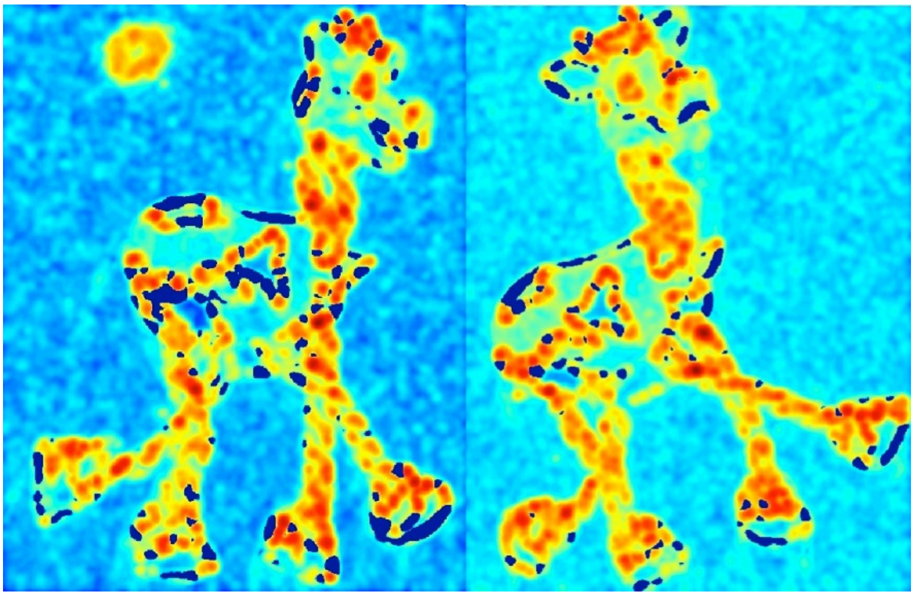
The Algorithm:

- Find points with large corner response function R (R > threshold)
- Take the points of local maxima of R

Harris Detector: Workflow

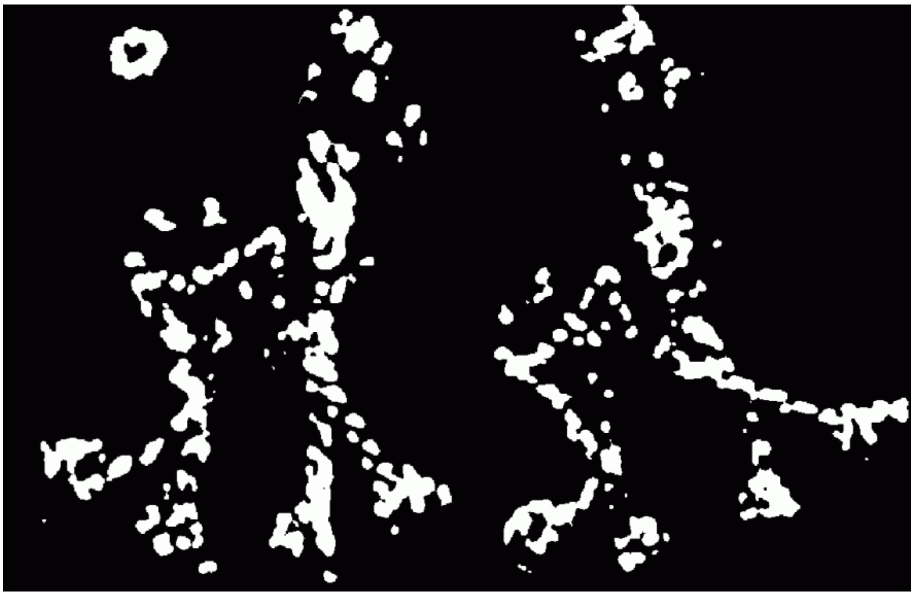
Harris Detector: Workflow

Compute corner response R



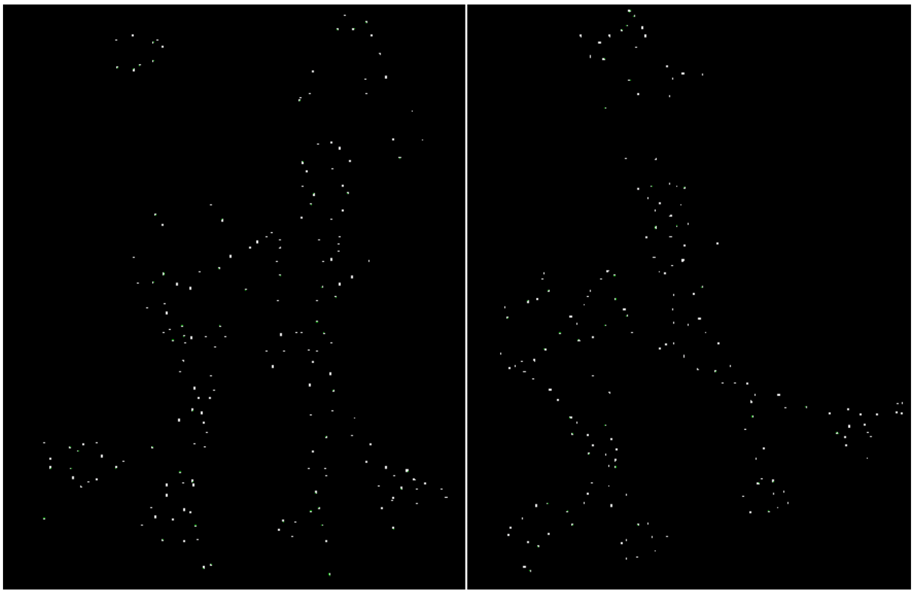
Harris Detector: Workflow

Find points with large corner response: R>threshold

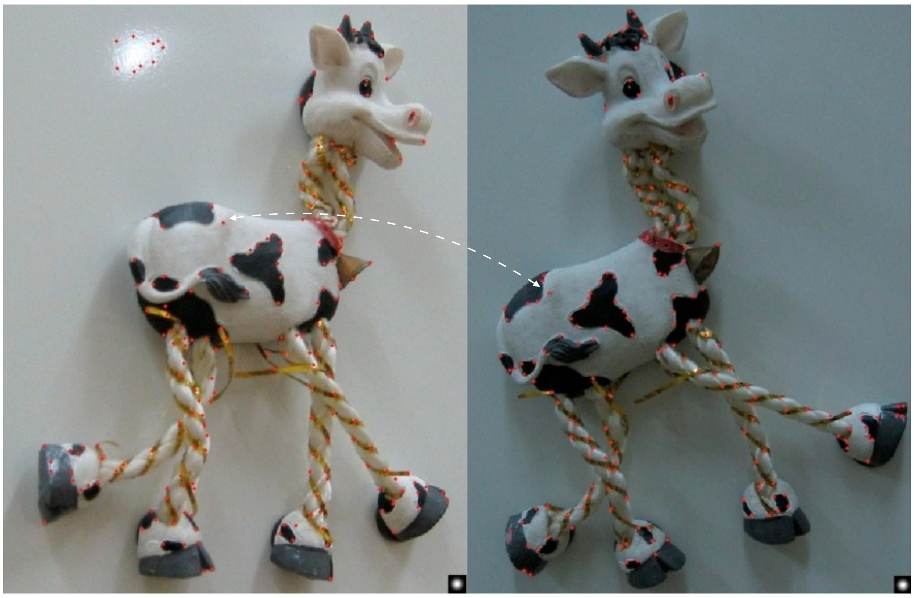


Harris Detector: Workflow

Take only the points of local maxima of R



Harris Detector: Workflow



Harris Detector: Summary

Average intensity change in direction $[u,v]$ can be expressed as a bilinear form:

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

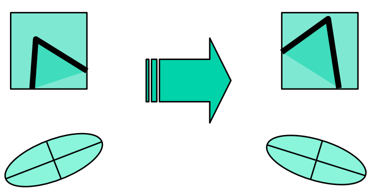
Describe a point in terms of eigenvalues of M:
measure of corner response

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2$$

A good (corner) point should have a *large intensity change* in *all directions*, i.e. R should be large positive

Harris Detector: Some Properties

Rotation invariance



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

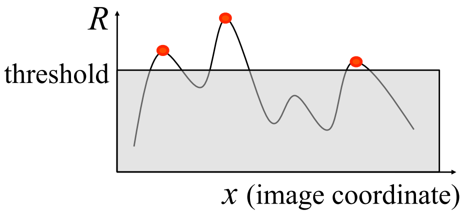
Corner response R is invariant to image rotation

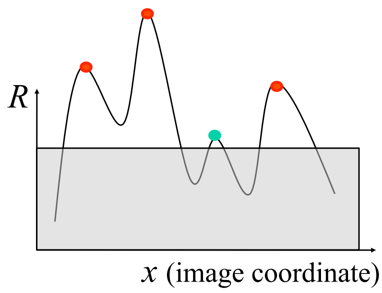
Harris Detector: Some Properties

Partial invariance to *affine intensity* change

✓ Only derivatives are used => invariance to intensity shift $I \rightarrow I + b$

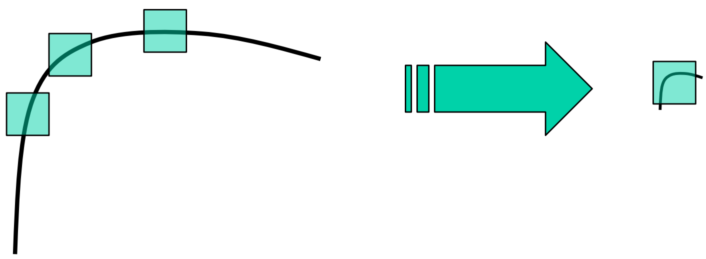
✓ Intensity scale: $I \rightarrow a\, I$





Harris Detector: Some Properties

But: <u>non-invariant to *image scale*</u>!



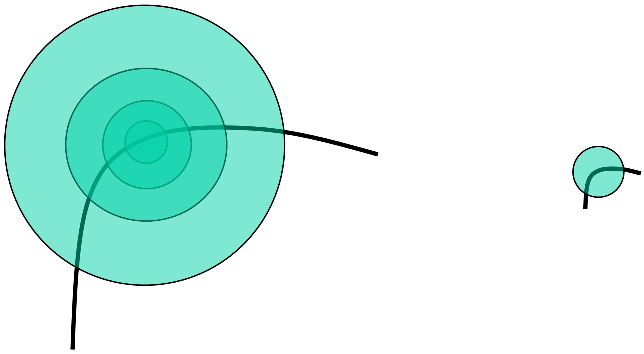
All points will be classified as edges

Corner !

Scale Invariant Detection

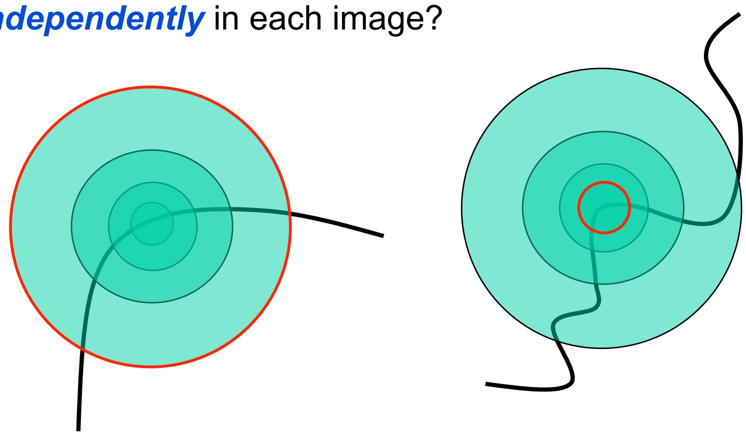
Consider regions (e.g. circles) of different sizes around a point

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 invariance

Requires a method to repeatably select points in location and scale:

The only reasonable scale-space kernel is a Gaussian (Koenderink, 1984; Lindeberg, 1994)

Efficient choice is to detect peaks in the difference of Gaussian pyramid (Burt & Adelson, 1983 – but examining more scales)

Difference-of-Gaussian with constant ratio of scales is a close approximation to Lindeberg's scale-normalized Laplacian (can be shown from the heat diffusion equation)

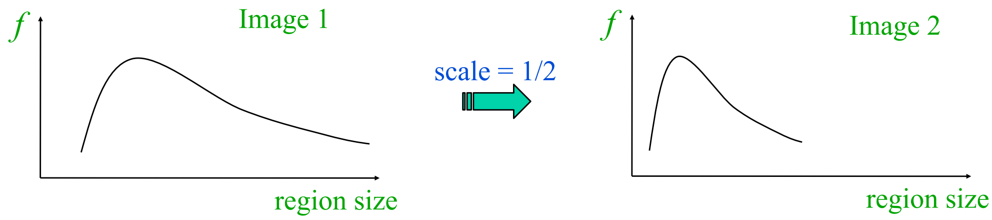
Scale Invariant Detection

Solution:

- Design a function on the region (circle), which is "scale invariant" (the same for corresponding regions, even if they are at different scales)

  Example: average intensity. For corresponding regions (even of different sizes) it will be the same.

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



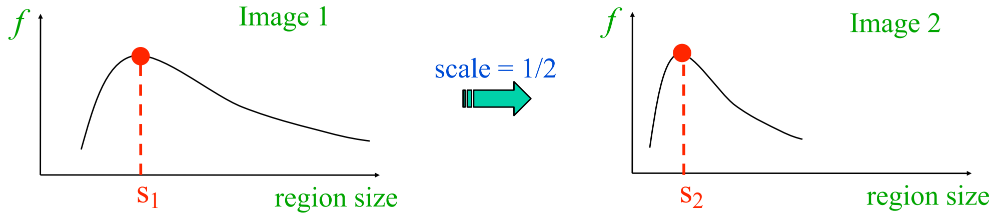
Scale Invariant Detection

Common approach:

Take a local maximum of this function

Observation: region size, for which the maximum is achieved, should be *invariant* to image scale.

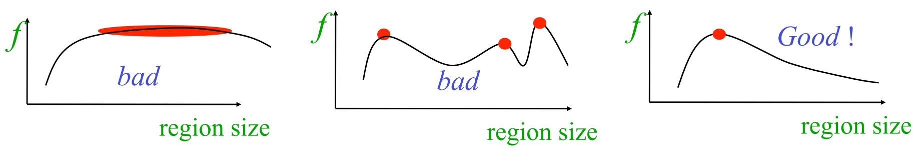
Important: this scale invariant region size is found in each image independently!



Scale Invariant Detection

A "good" function for scale detection:

has one stable sharp peak



- For usual images: a good function would be a one which responds to contrast (sharp local intensity change)

Scale Invariant Detection

Functions for determining scale

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

Kernels:

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

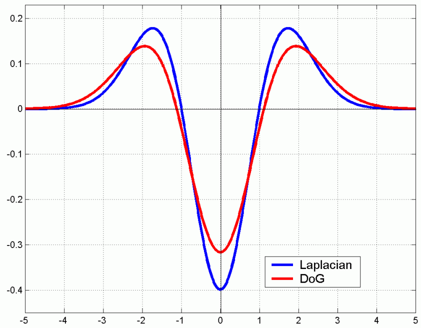
(Laplacian)

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

(Difference of Gaussians)

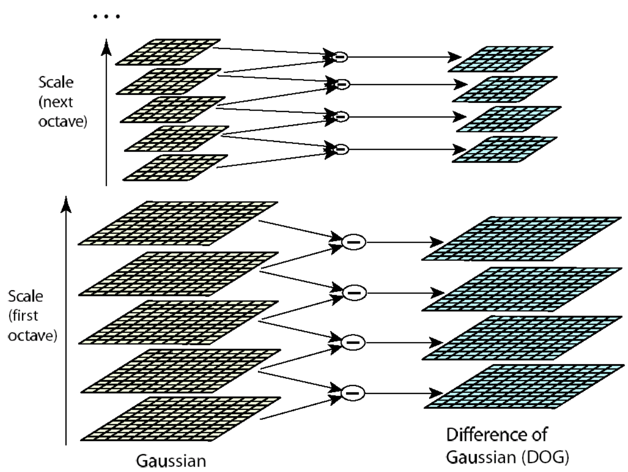
where Gaussian

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



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

Scale space: one octave at a time

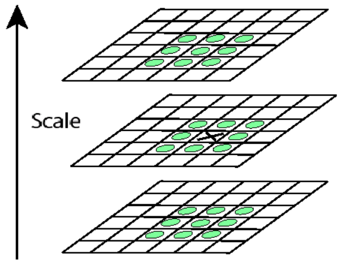


Key point localization

Detect maxima and minima of difference-of-Gaussian in scale space

Fit a quadratic to surrounding values for sub-pixel and sub-scale interpolation (Brown & Lowe, 2002)

Taylor expansion around point:



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

Offset of extremum (use finite differences for derivatives):

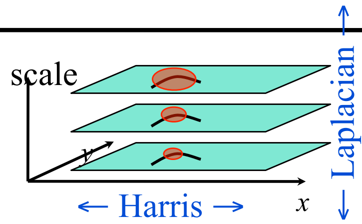
$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}$$

Scale Invariant Detectors

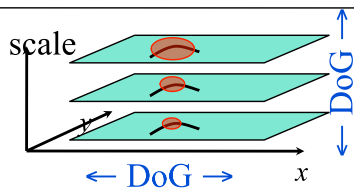
Harris-Laplacian[1]

Find local maximum of:

- Harris corner detector in space (image coordinates)
- Laplacian in scale



- SIFT (Lowe)[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

Scale Invariant Detection: Summary

Given: two images of the same scene with a large *scale difference* between them

Goal: find *the same* interest points *independently* in each image

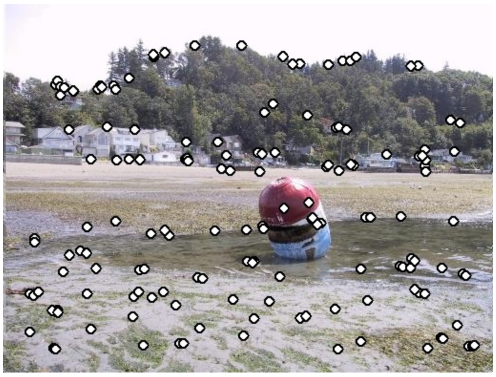
Solution: search for *maxima* of suitable functions in *scale* and in *space* (over the image)

Methods:

1. Harris-Laplacian [Mikolajczyk, Schmid]: maximize Laplacian over scale, Harris' measure of corner response over the image
2. SIFT [Lowe]: maximize Difference of Gaussians over scale and space

Feature selection

Distribute points evenly over the image



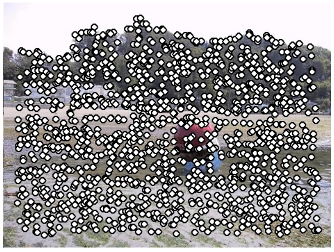
Adaptive Non-maximal Suppression

Desired: Fixed # of features per image

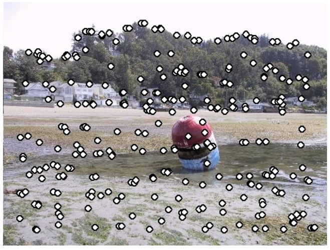
- Want evenly distributed spatially…
- Search over non-maximal suppression radius

$$r_i = \min_j |\mathbf{x}_i - \mathbf{x}_j|, \text{ s.t. } f(\mathbf{x}_i) < c_{\text{robust}} f(\mathbf{x}_j), \ \mathbf{x}_j \in \mathcal{I}$$

[Brown, Szeliski, Winder, CVPR'05]



r = 8, n = 1388



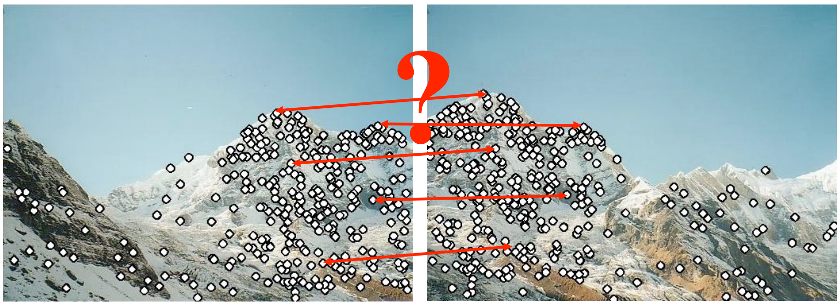
r = 20, n = 283

Outline

- Feature detectors
  - scale and affine invariant (points, regions)
  - selection of features
- Feature descriptors
  - patches, oriented patches
  - SIFT (orientations)

Feature descriptors

We know how to detect points
Next question: **How to match them?**

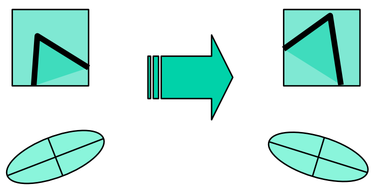


Point descriptor should be:
1. Invariant
2. Distinctive

Descriptors invariant to rotation

Harris corner response measure:
depends only on the eigenvalues of the matrix *M*
Careful with window effects! (Use circular)

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



Descriptors Invariant to Rotation

Image moments in polar coordinates

$$m_{kl} = \iint r^k e^{-i\theta l} I(r,\theta)\, dr d\theta$$

Rotation in polar coordinates is translation of the angle:

$$\theta \rightarrow \theta + \theta_0$$

This transformation changes only the phase of the moments, but not its magnitude

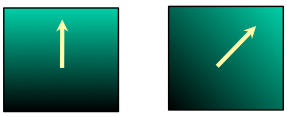
Rotation invariant descriptor consists of magnitudes of moments: $|m_{kl}|$

Matching is done by comparing vectors $[|m_{kl}|]_{k,l}$

Descriptors Invariant to Rotation

Find local orientation

Dominant direction of gradient



- Compute image derivatives relative to this orientation

Descriptors Invariant to Scale

Use the scale determined by detector to compute descriptor in a normalized frame

For example:
- moments integrated over an adapted window
- derivatives adapted to scale: sI_x

Invariance to Intensity Change

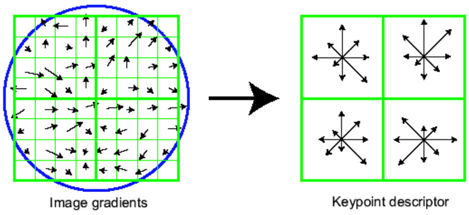
Detectors
- mostly invariant to affine (linear) change in image intensity, because we are searching for *maxima*

Descriptors
- Some are based on derivatives => invariant to intensity shift
- Some are normalized to tolerate intensity scale
- Generic method: pre-normalize intensity of a region (eliminate shift and scale)

SIFT – Scale Invariant Feature Transform

Descriptor overview:

- Determine scale (by maximizing DoG in scale and in space), local orientation as the dominant gradient direction. Use this scale and orientation to make all further computations invariant to scale and rotation.
- Compute gradient orientation histograms of several small windows (128 values for each point)
- Normalize the descriptor to make it invariant to intensity change



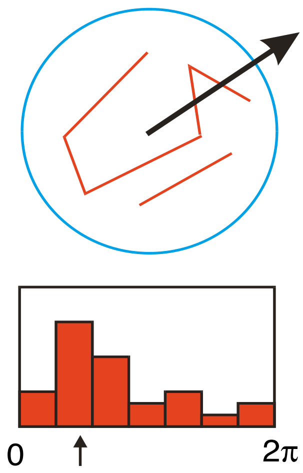
D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". IJCV 2004

Select canonical orientation

Create histogram of local gradient directions computed at selected scale

Assign canonical orientation at peak of smoothed histogram

Each key specifies stable 2D coordinates (x, y, scale, orientation)

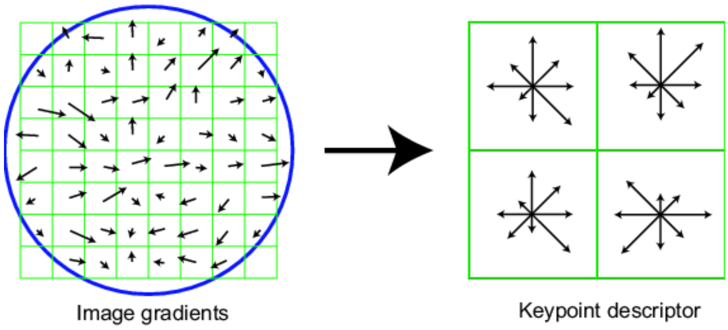


SIFT vector formation

Thresholded image gradients are sampled over 16x16 array of locations in scale space

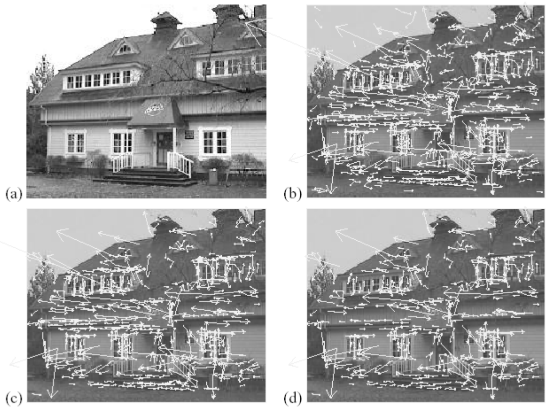
Create array of orientation histograms

8 orientations x 4x4 histogram array = 128 dimensions



Example of keypoint detection

Threshold on value at DOG peak and on ratio of principle curvatures (Harris approach)

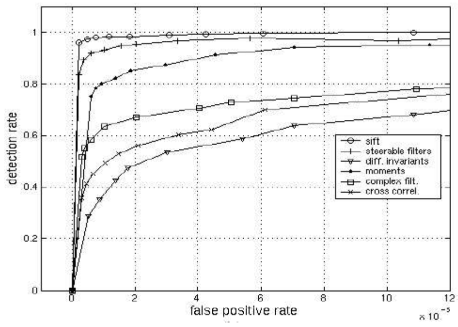


(a) 233x189 image
(b) 832 DOG extrema
(c) 729 left after peak value threshold
(d) 536 left after testing ratio of principle curvatures

SIFT – Scale Invariant Feature Transform[1]

Empirically found[2] to show very good performance, invariant to *image rotation*, *scale*, *intensity change*, and to moderate *affine* transformations

Scale = 2.5
Rotation = 45^0



[1] D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". Accepted to IJCV 2004
[2] K.Mikolajczyk, C.Schmid. "A Performance Evaluation of Local Descriptors". CVPR 2003