Matching features

Computational Photography, 6.882

Prof. Bill Freeman
April 11, 2006

Image and shape descriptors: Harris corner detectors and SIFT features.

Suggested readings: Mikolajczyk and Schmid, David Lowe IJCV.

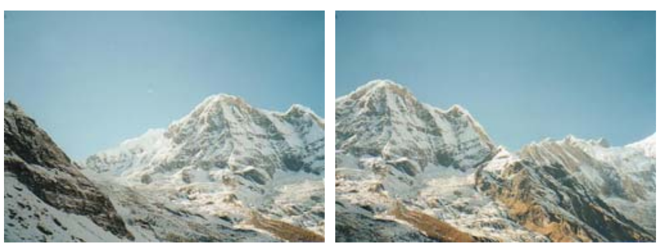
Modifications to slides by Allan$_+$ Jepson, Oct. 2009

Building a Panorama

M. Brown and D. G. Lowe. Recognising Panoramas. ICCV 2003

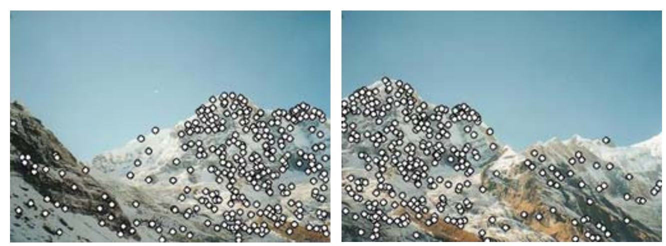
How do we build a panorama?

- We need to match (align) images



Matching with Features

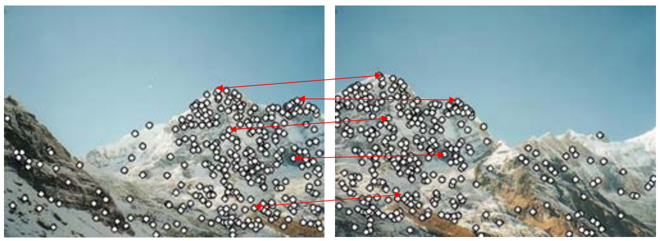
•Detect feature points in both images



Matching with Features

•Detect feature points in both images

•Find corresponding pairs



Matching with Features

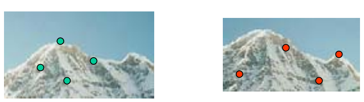
•Detect feature points in both images

•Find corresponding pairs

•Use these pairs to align images

Matching with Features

- Problem 1:
  - Detect the *same* point *independently* in both images

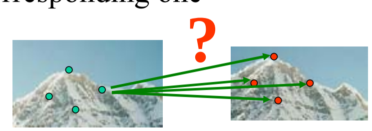


no chance to match!

We need a repeatable detector

Matching with Features

- Problem 2:
  - For each point correctly recognize the corresponding one



We need a reliable and distinctive descriptor

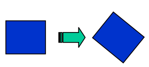
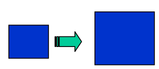
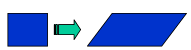
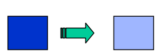
More motivation…

- Feature points are used also for:
  - Image alignment (homography, fundamental matrix)
  - 3D reconstruction
  - Motion tracking
  - Object recognition
  - Indexing and database retrieval
  - Robot navigation
  - … other

We want to:

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

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 ($I \rightarrow a\, I + b$) 

Selecting Good Features

What's a "good feature"?

- Distinctive Image Location, Scale and Orientation:
  - image landmark.
  - pose can be repeatably identified from the image itself.
- Descriptive:
  - Provides distinctive information about the image structure in a neighbourhood of the landmark point.
- Stable under viewpoint changes:
  - information is stable under common changes in noise, orientation, scale, 3D pose, view, lighting.

Contents

Harris Detector: Summary

- Spatially averaged outer product of image gradient:

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

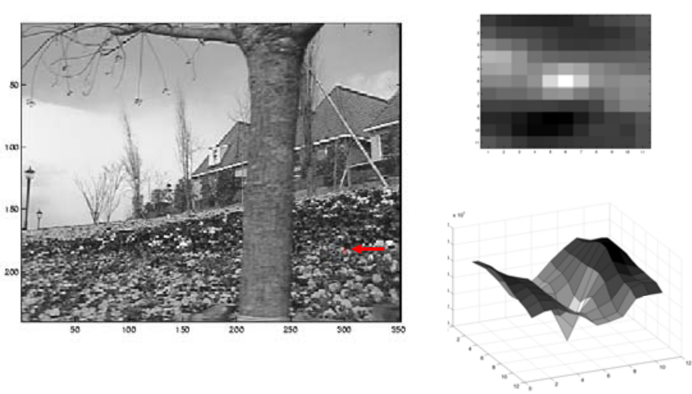
- Eigenvalues of M indicate texture/oriented/blank regions

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

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

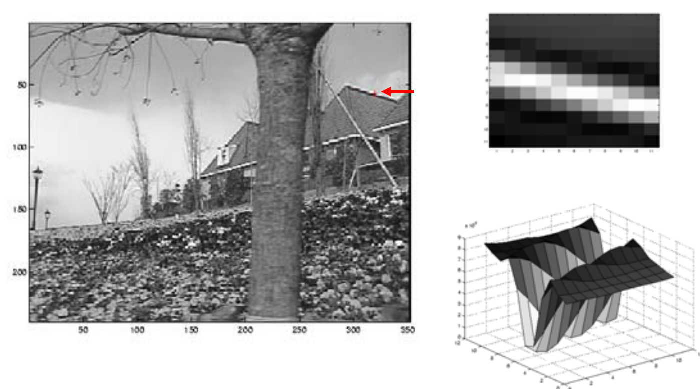
- A good (textured) point should have a *large intensity change* in *all directions*, i.e. R should be large positive

Selecting Good Features



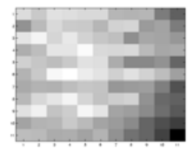
λ_1 and λ_2 are large

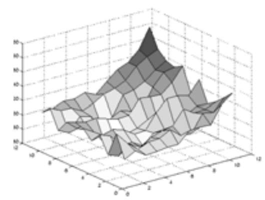
Selecting Good Features



large λ_1, small λ_2

Selecting Good Features





small λ_1, small λ_2

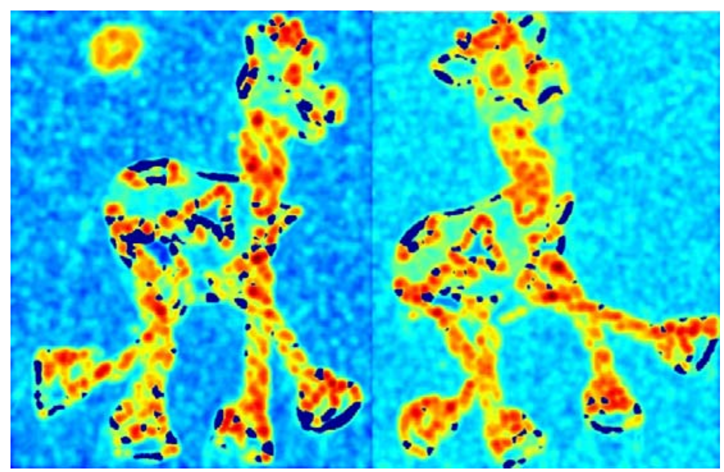
Harris Detector

- 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

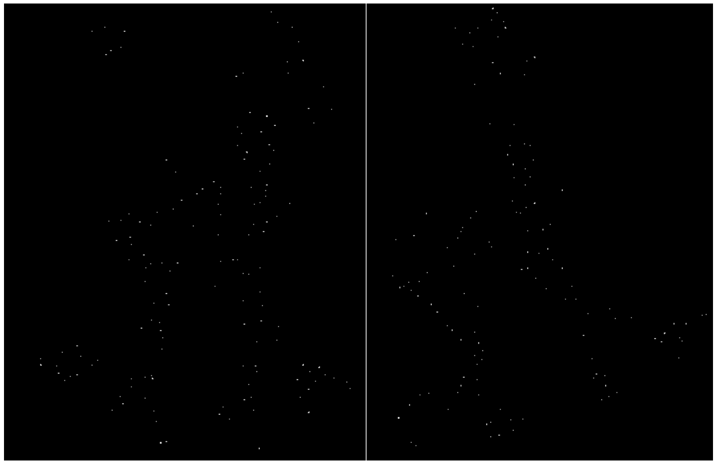


Harris Detector: Workflow

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: Some Properties

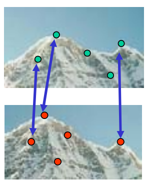
- Rotation invariance.
- Scaling I → aI,
  - R → aR, R > threshold varies,
  - but local spatial peaks remain peaks.
- Not scale invariant.

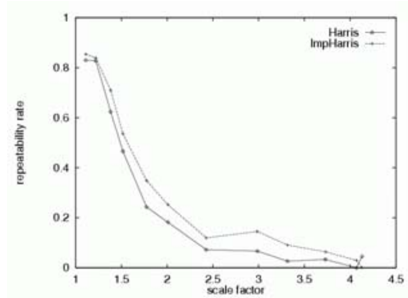
Harris Detector: Some Properties

- Quality of Harris detector for different scale changes

Repeatability rate:

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



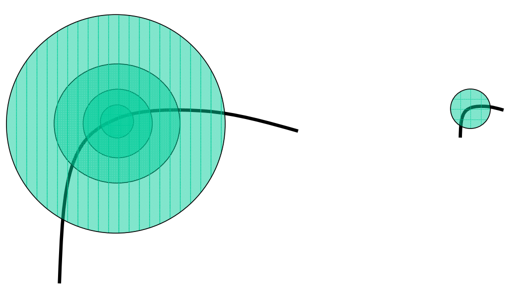


C.Schmid et.al. “Evaluation of Interest Point Detectors”. IJCV 2000

Contents

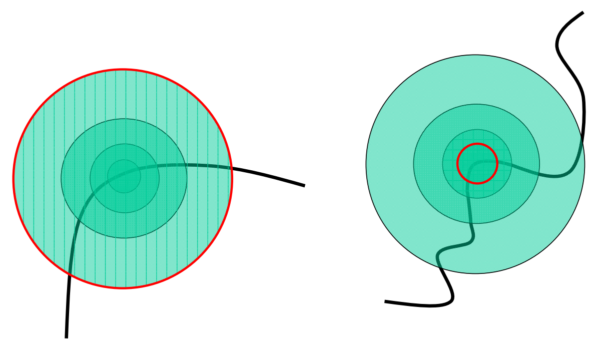
Scale Invariant Detection

- 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?



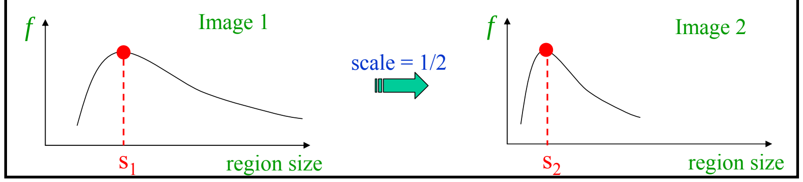
Scale Invariant Detection

- Common approach:

  Take a isotropic bandpass filter, vary σ (aka region size)

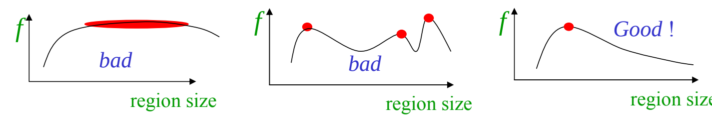
  Size for which the maximum is achieved, should be *invariant* to image scale.

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 variations.

Scale Invariant Detection

- Functions for determining scale $f = \text{Kernel} * \text{Image}$

Kernels:

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

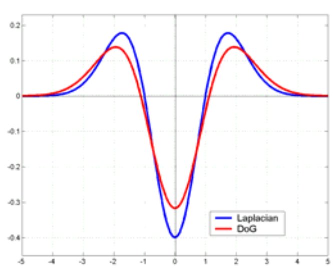
(Laplacian)

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

(Difference of Gaussians)

where Gaussian

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

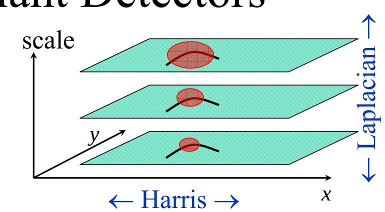


Note: both kernels are invariant to *scale* and *rotation*

Scale Invariant Detectors

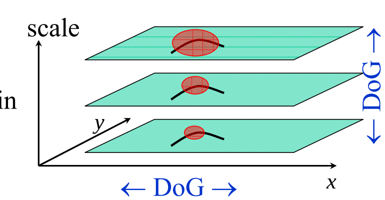
- Harris-Laplacian[1]

  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

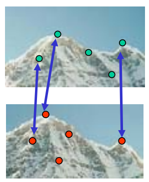
[2] D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". Accepted to IJCV 2004

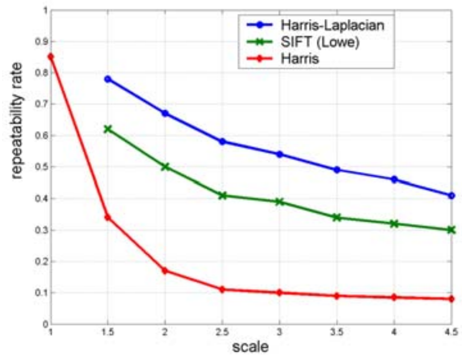
Scale Invariant Detectors

- Experimental evaluation of detectors w.r.t. scale change

Repeatability rate:

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



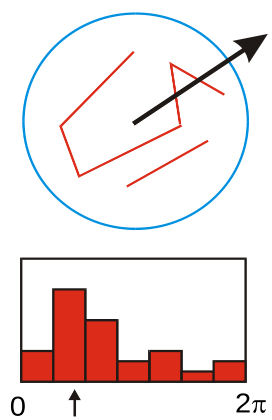


K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001

Contents

Select canonical orientation

- Create histogram of local gradient directions computed at selected scale
- Assign canonical orientation(s) at peak(s) of smoothed histogram
- Each key-point then specifies stable 2D coordinates (x, y, scale, orientation)

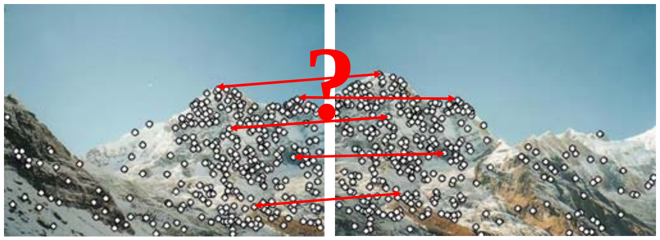


Contents

Point Descriptors

- We know how to detect points
- Next question:

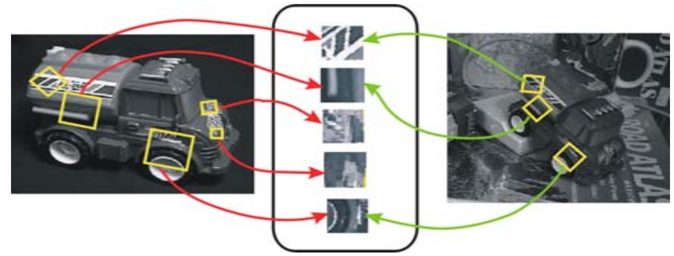
How to match them?



Point descriptor should be:
1. Insensitive to image deformations.
2. Distinctive

Invariant Local Features

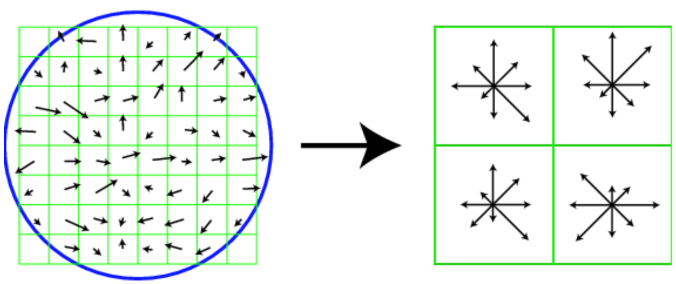
- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters



SIFT Features (Scale Invariant Feature Transform)

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



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

Distinctiveness of features

- Vary size of database of features, with 30 degree affine change, 2% image noise
- Measure % correct for single nearest neighbor match

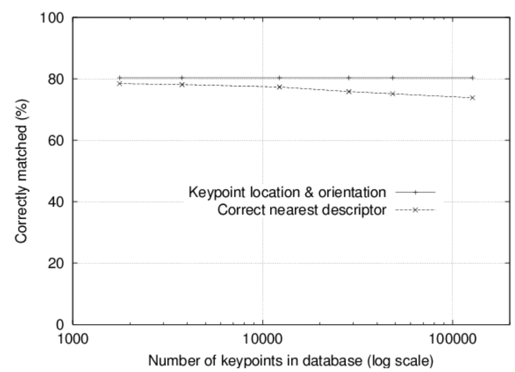


Figure 12: The training images for two objects are shown on the left. These can be recognized in a cluttered image with extensive occlusion, shown in the middle. The results of recognition are shown on the right. A parallelogram is drawn around each recognized object showing the boundaries of the original training image under the affi ne transformation solved for during recognition. Smaller squares indicate the keypoints that were used for recognition.

Figure 13: This example shows location recognition within a complex scene. The training images for locations are shown at the upper left and the 640x315 pixel test image taken from a different viewpoint is on the upper right. The recognized regions are shown on the lower image, with keypoints shown as squares and an outer parallelogram showing the boundaries of the training images under the affi ne transform used for recognition.

A good SIFT features tutorial

http://www.cs.toronto.edu/~jepson/csc2503/tutSIFT04.pdf

By Estrada, Jepson, and Fleet.

The Matlab SIFTtutorial in utvisToolbox is also pretty cool.