6.869

Advances in Computer Vision

Prof. Bill Freeman

March 3, 2005

Image and shape descriptors

- Affine invariant features
- Comparison of feature descriptors
- Shape context

Readings: Mikolajczyk and Schmid; Belongie et al

Matching with Invariant Features

Darya Frolova, Denis Simakov
The Weizmann Institute of Science

March 2004

Example: Build a Panorama

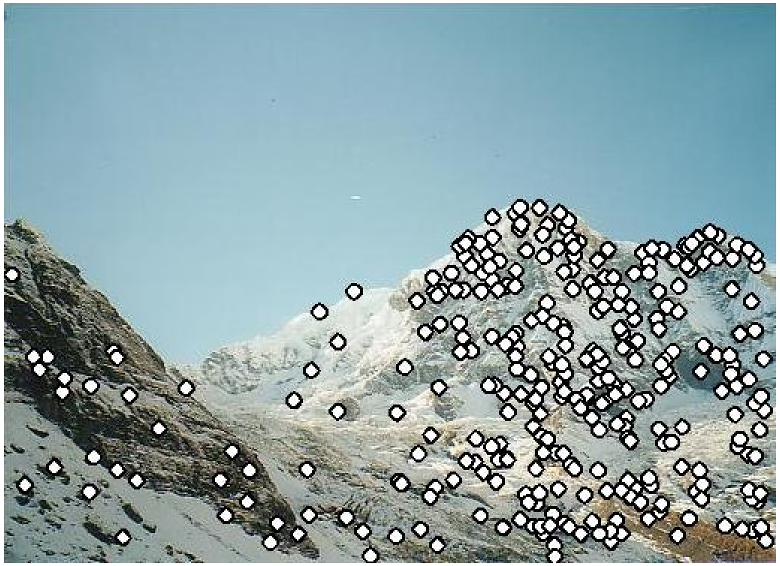
M. Brown and D. G. Lowe. Recognising Panoramas. ICCV 2003

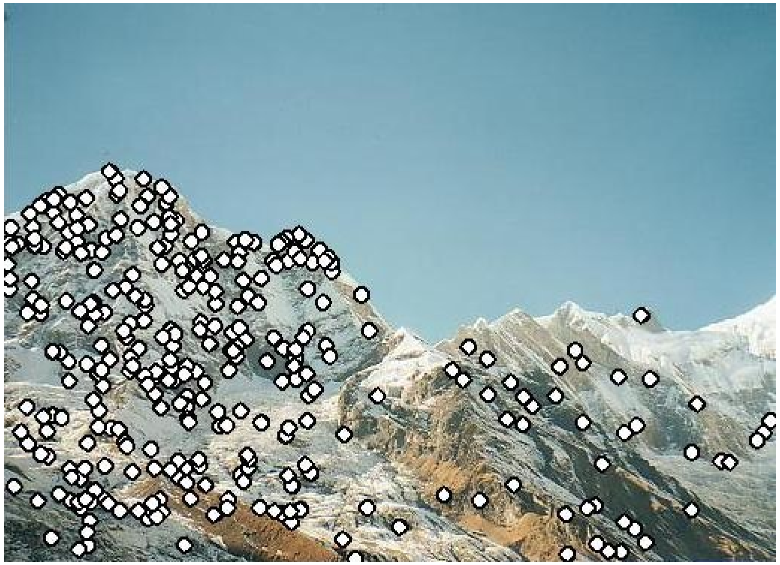
How do we build panorama?

- We need to match (align) images

Matching with Features

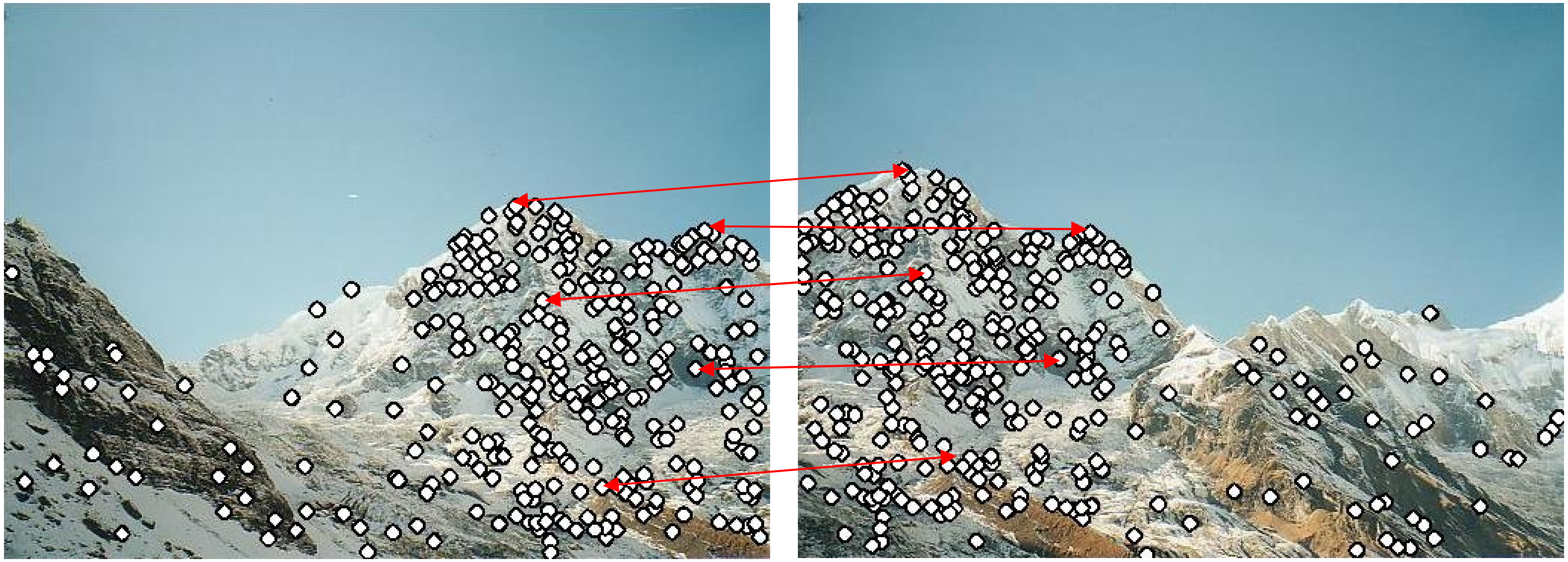
•Detect feature points in both images





Matching with Features

- Detect feature points in both images
- Find corresponding pairs

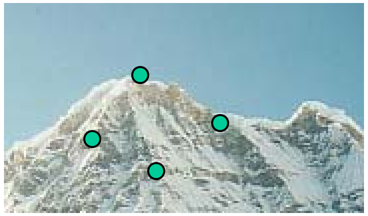


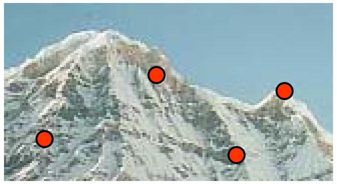
Matching with Features

- 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



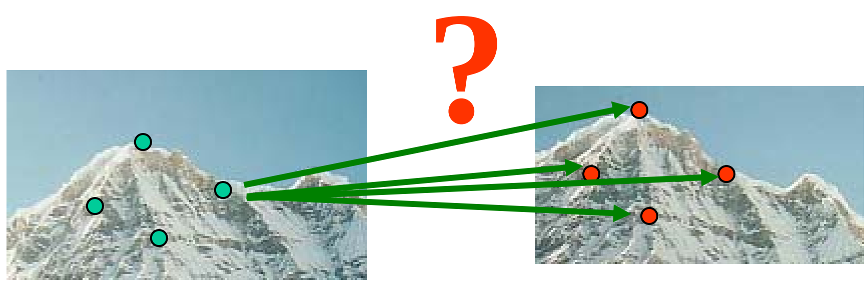


no chance to match!

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…

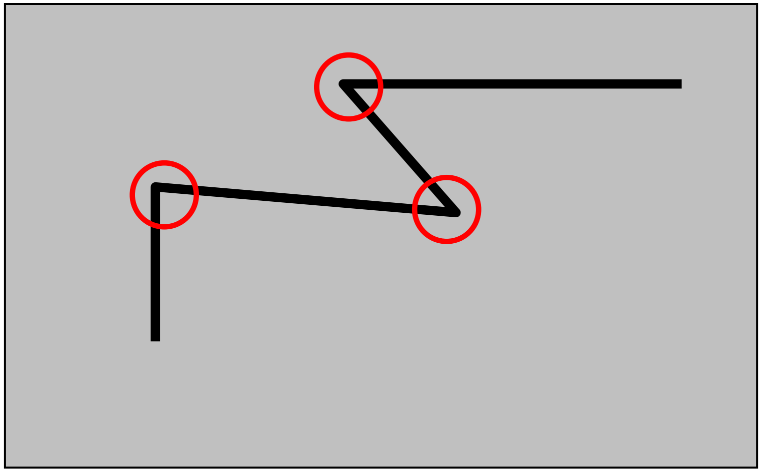
- Feature points are used also for:
  - Image alignment (homography, fundamental matrix)
  - 3D reconstruction
  - Motion tracking
  - Object recognition
  - Indexing and database retrieval
  - Robot navigation
  - … other

Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

An introductory example:

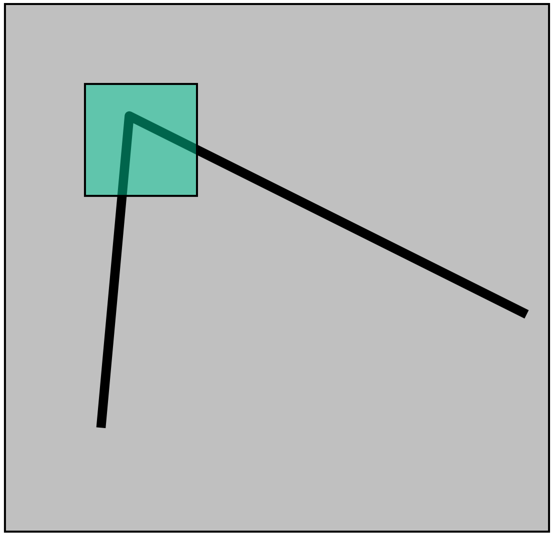
Harris corner detector



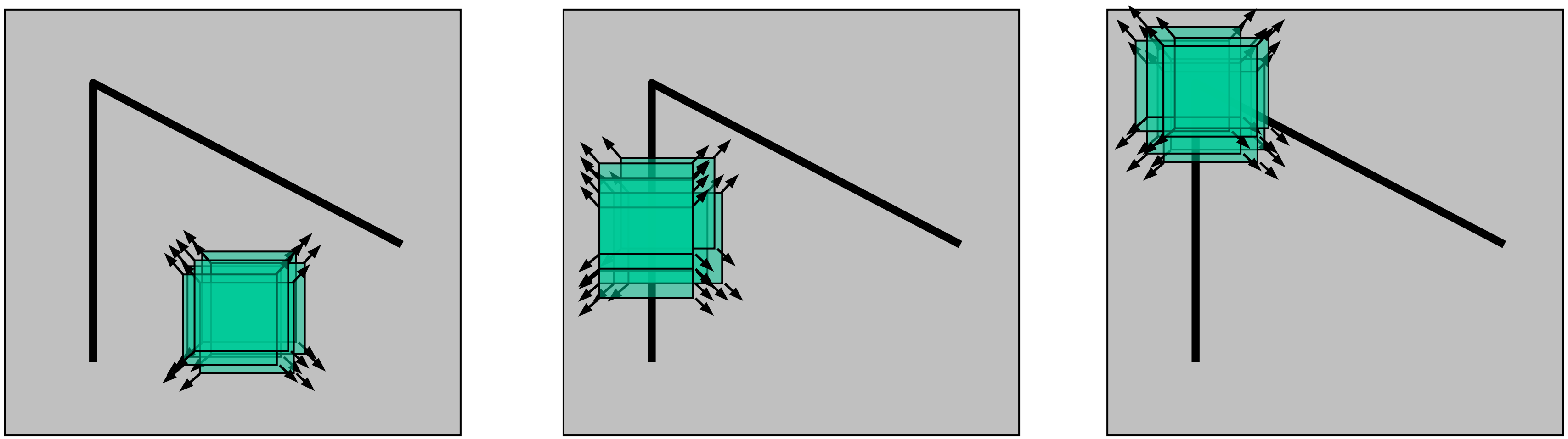
C.Harris, M.Stephens. “A Combined Corner and Edge Detector”. 1988

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

Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Harris Detector: Mathematics

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

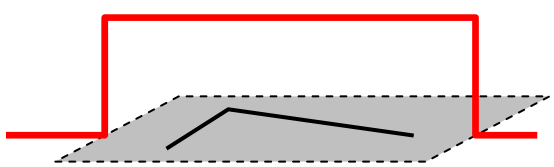
$$E(u,v) = \sum_{x,y} w(x,y)\left[I(x+u,y+v) - I(x,y)\right]^2$$

Window function

Shifted intensity

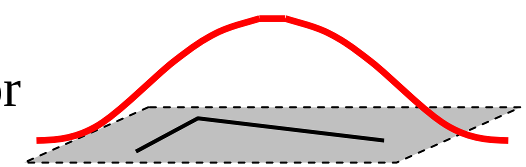
Intensity

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:

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

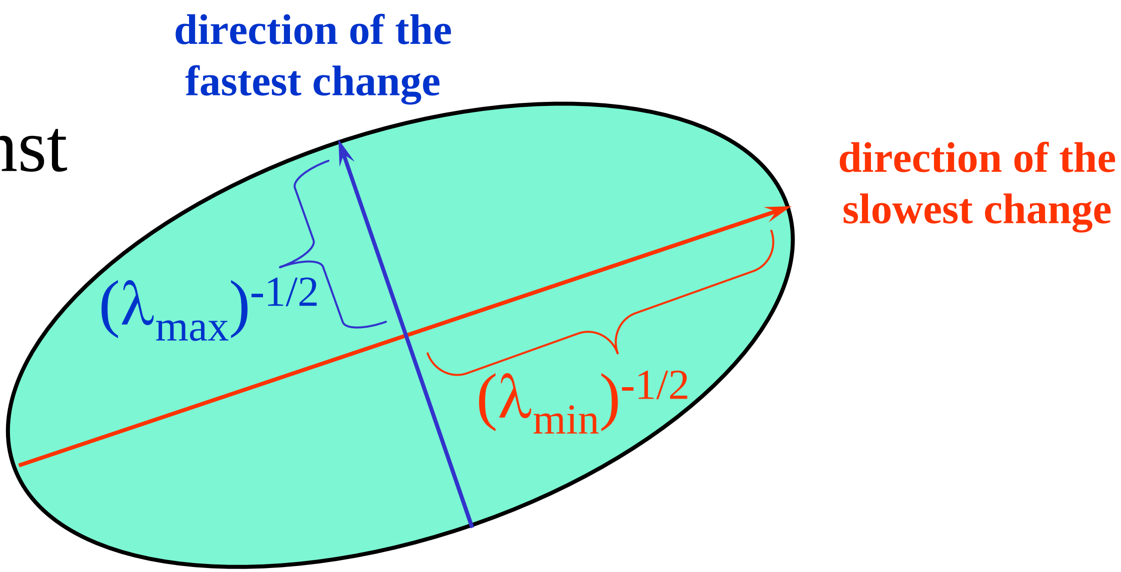
Harris Detector: Mathematics

Intensity change in shifting window: eigenvalue analysis

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

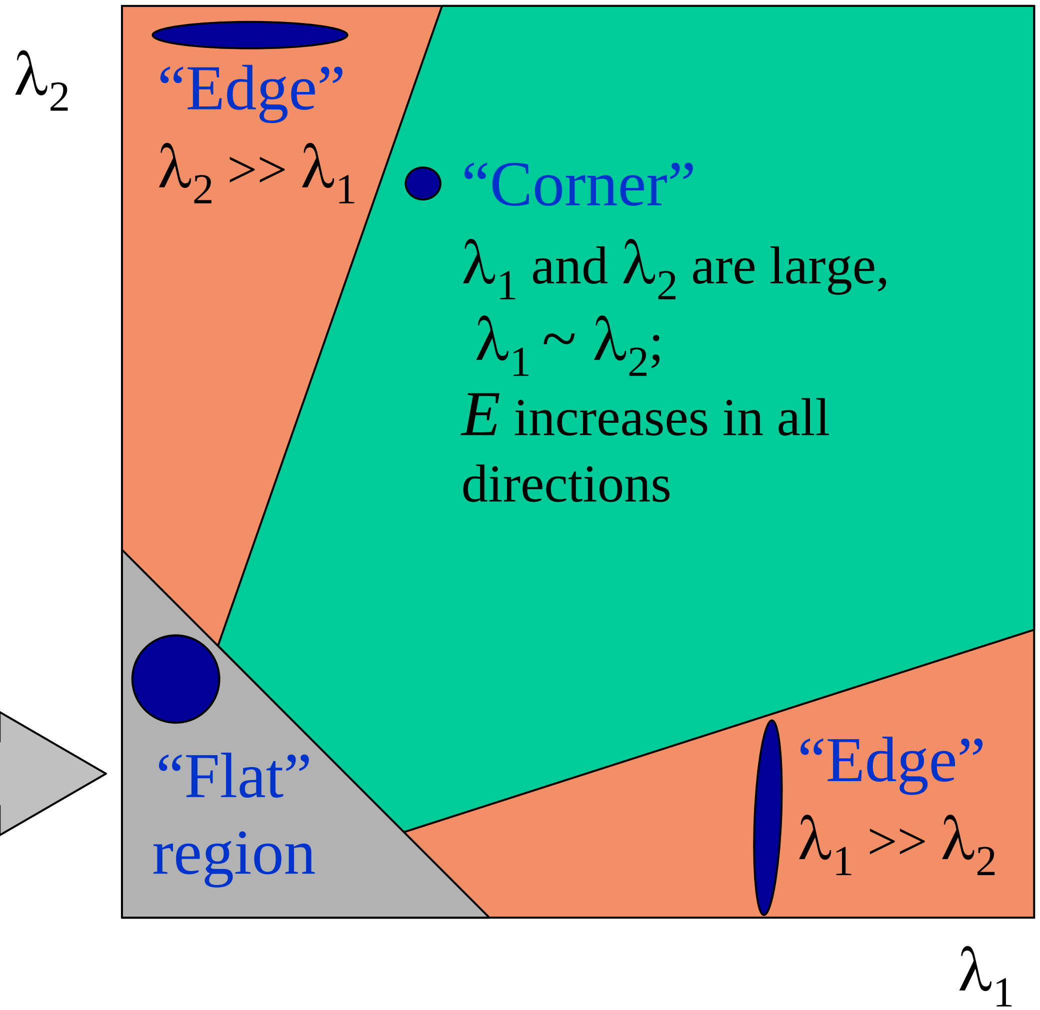
λ_1, λ_2 – eigenvalues of M

Ellipse $E(u,v)$ = const



Harris Detector: Mathematics

Classification of image points using eigenvalues of *M*:



λ_1 and λ_2 are small;
E is almost constant in all directions

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)

The principal curvatures can be computed from a 2x2 Hessian matrix, $\mathbf{H}$, computed at the location and scale of the keypoint:

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{4}$$

The derivatives are estimated by taking differences of neighboring sample points.

The eigenvalues of $\mathbf{H}$ are proportional to the principal curvatures of D. Borrowing from the approach used by Harris and Stephens (1988), we can avoid explicitly computing the eigenvalues, as we are only concerned with their ratio. Let α be the eigenvalue with the largest magnitude and β be the smaller one. Then, we can compute the sum of the eigenvalues from the trace of $\mathbf{H}$ and their product from the determinant:

$$\mathrm{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta,$$

$$\mathrm{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta.$$

In the unlikely event that the determinant is negative, the curvatures have different signs so the point is discarded as not being an extremum. Let r be the ratio between the largest magnitude eigenvalue and the smaller one, so that $\alpha = r\beta$. Then,

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r},$$

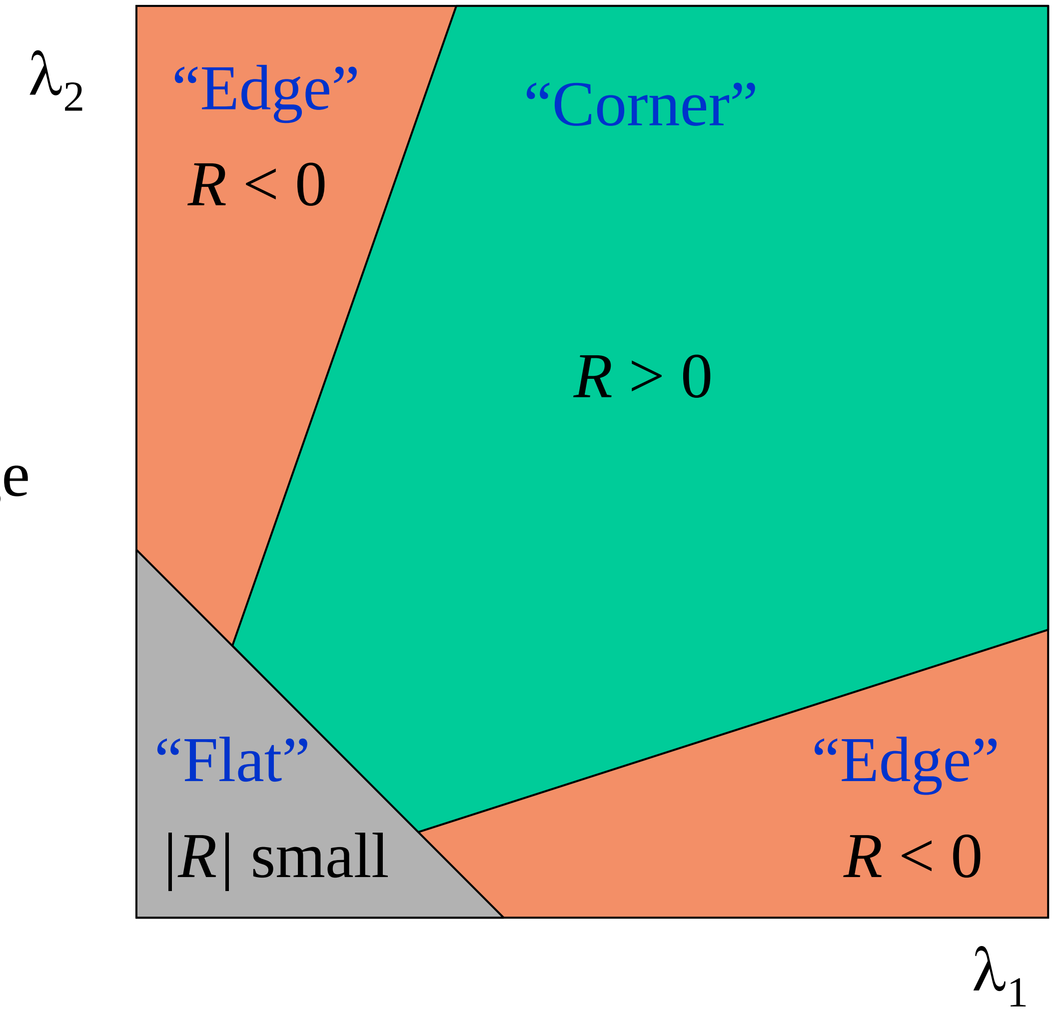
which depends only on the ratio of the eigenvalues rather than their individual values. The quantity $(r+1)^2/r$ is at a minimum when the two eigenvalues are equal and it increases with r. Therefore, to check that the ratio of principal curvatures is below some threshold, r, we only need to check

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r+1)^2}{r}.$$

cf David Lowe's analysis

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



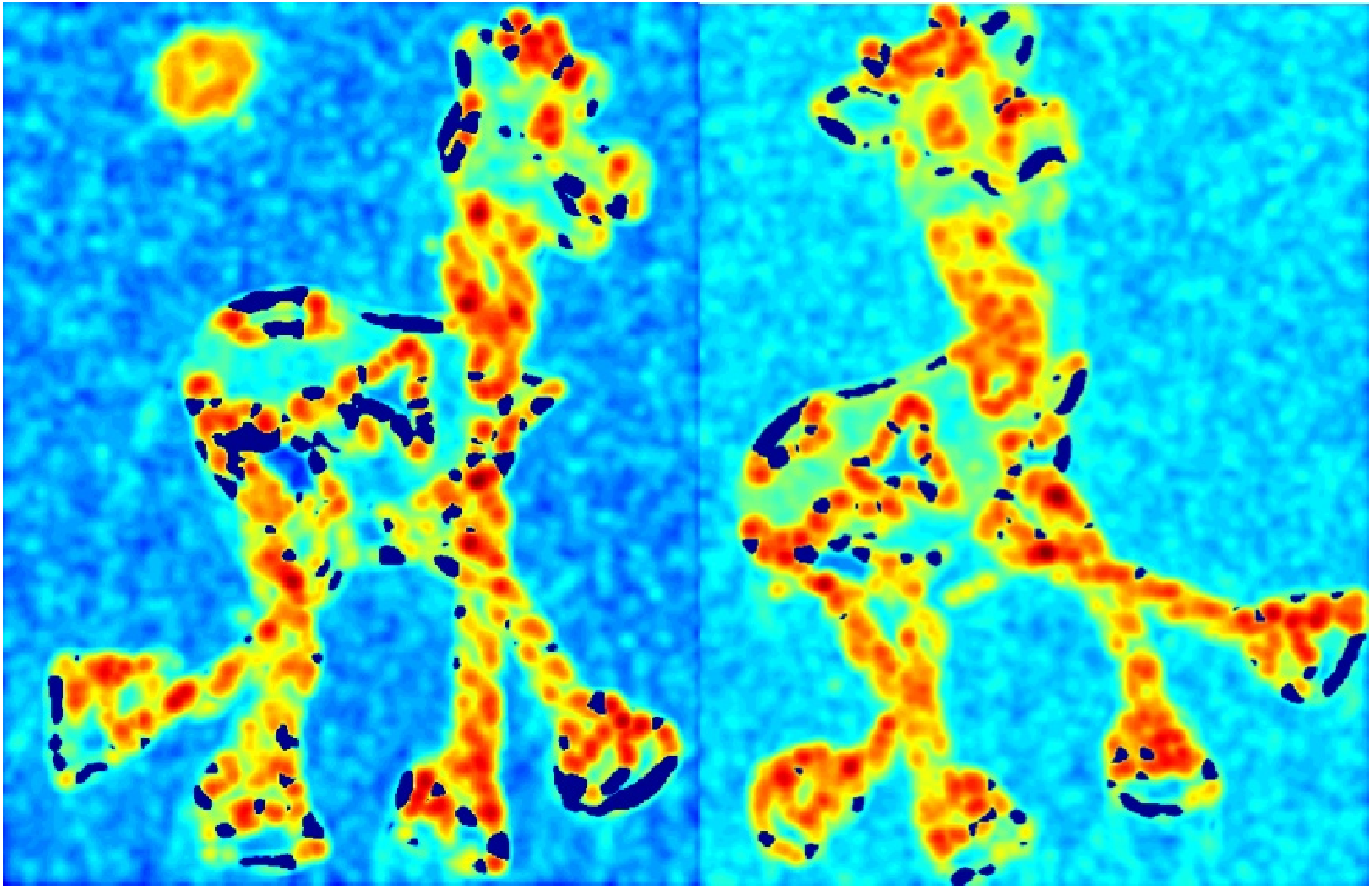
Harris Detector

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

Harris Detector: Workflow

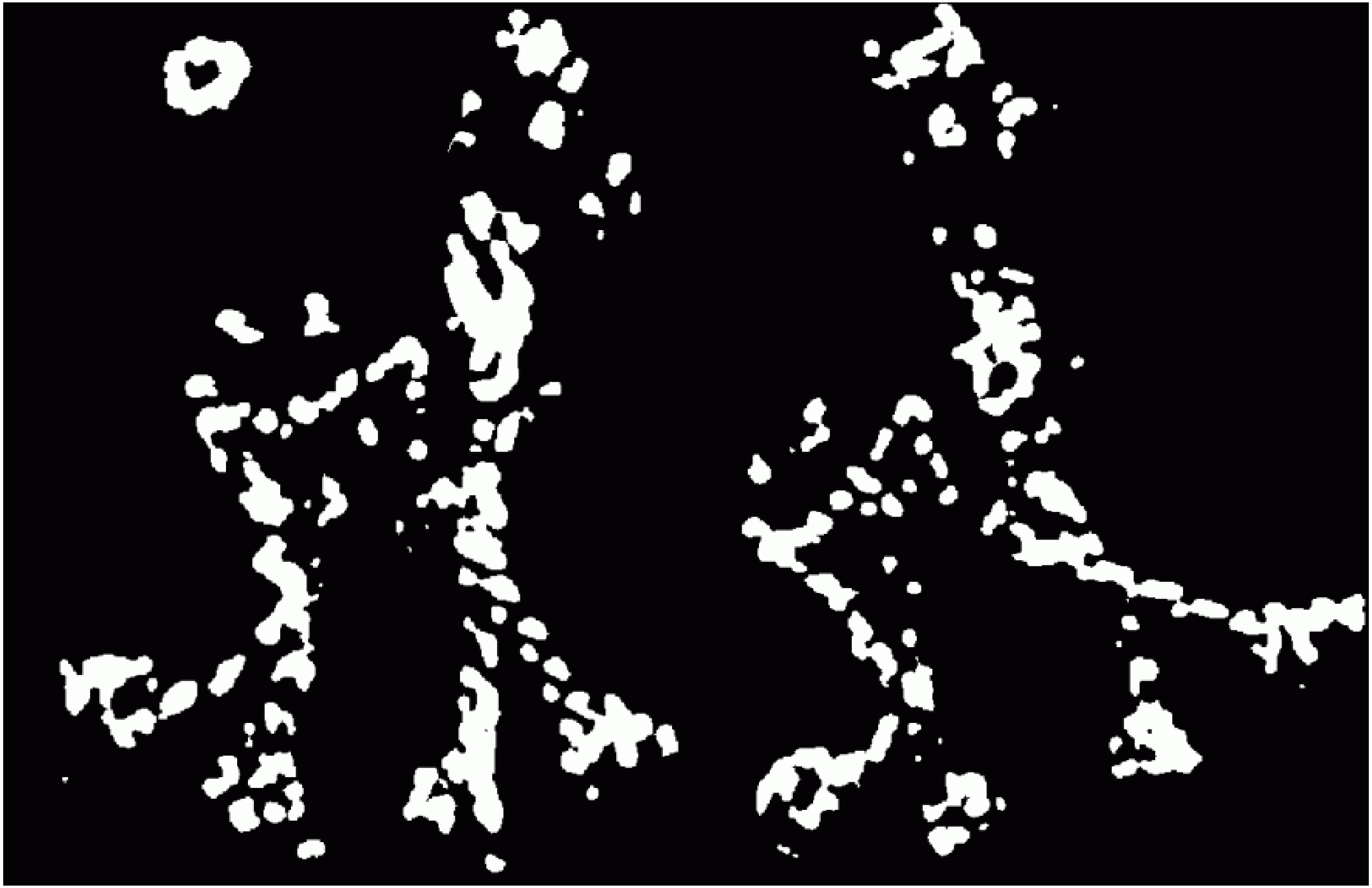
Harris Detector: Workflow

Compute corner response R



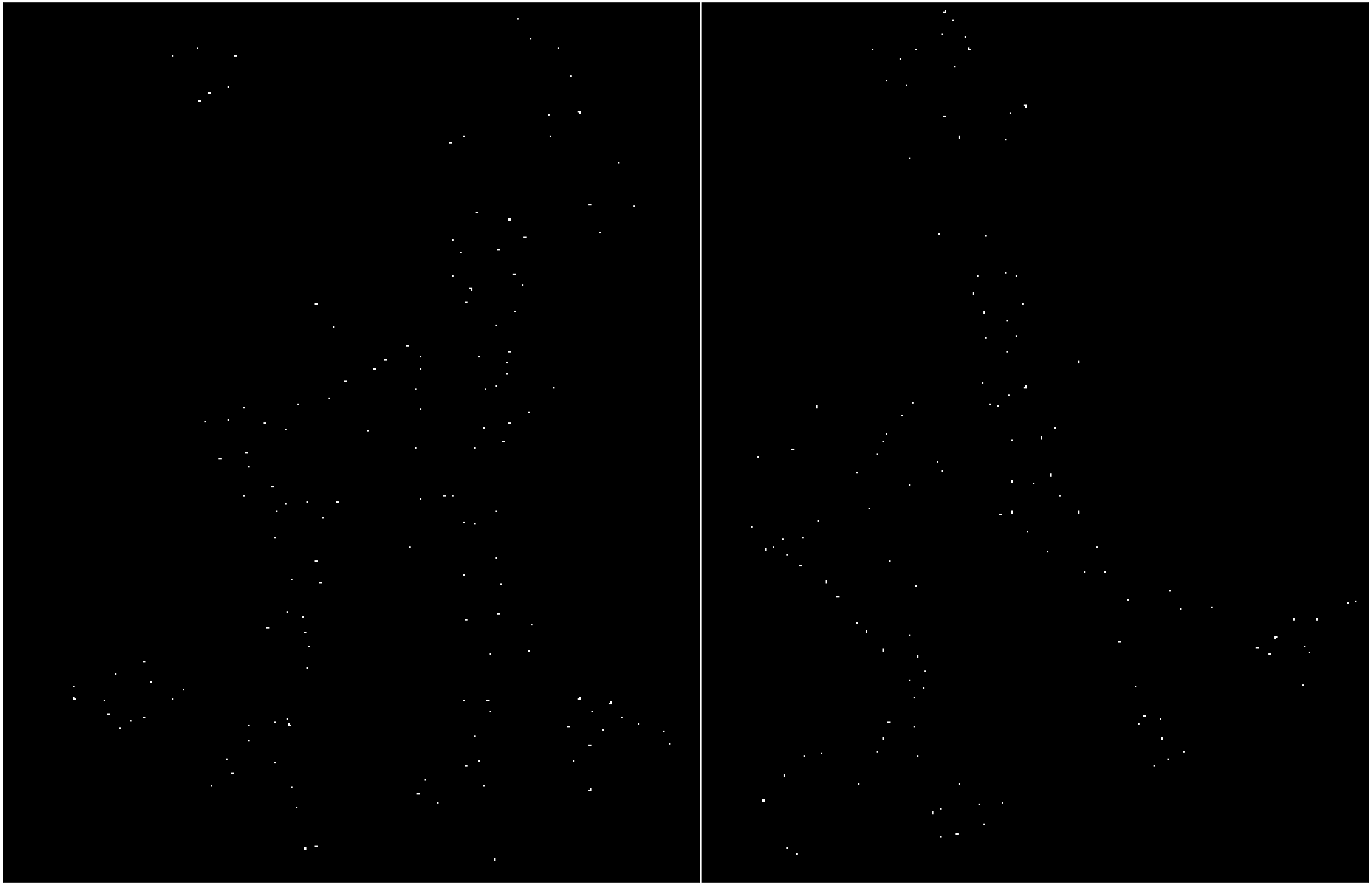
Harris Detector: Workflow

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*

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

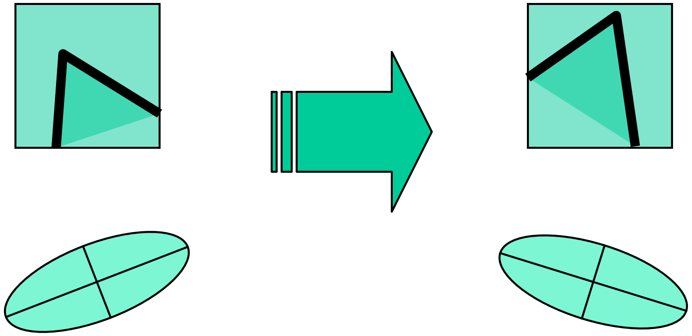
- A good (corner) point should have a *large intensity change* in *all directions*, i.e. R should be large positive

Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Harris Detector: Some Properties

- Rotation invariance



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

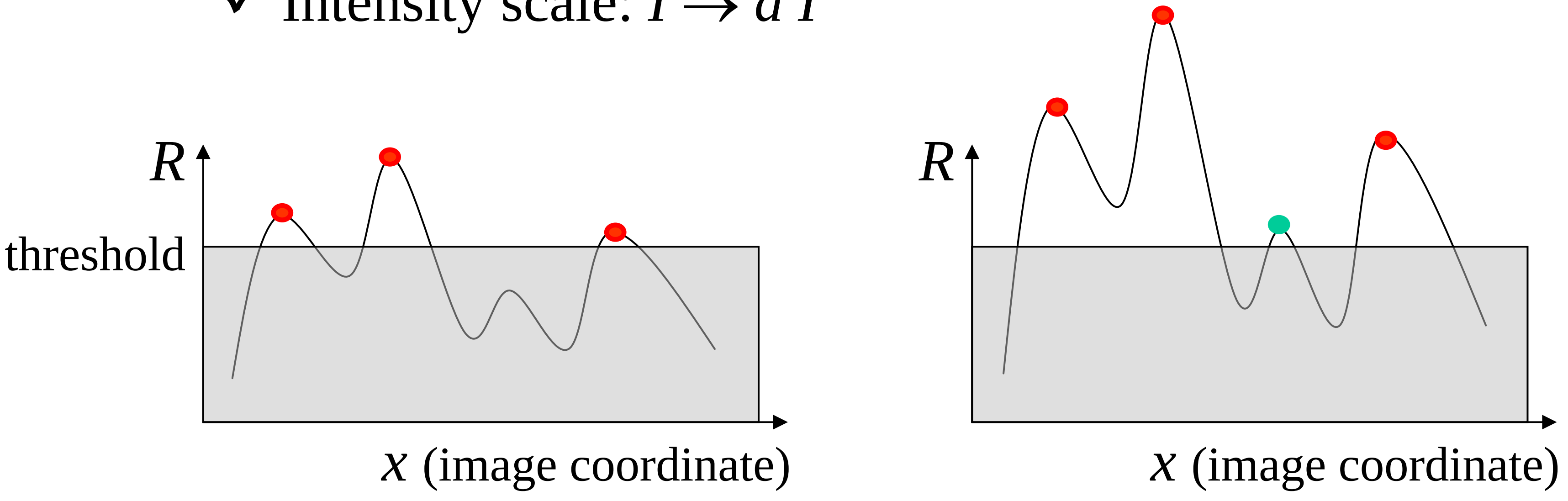
Corner response R is invariant to image rotation

Harris Detector: Some Properties

- Partial invariance to *affine intensity* change

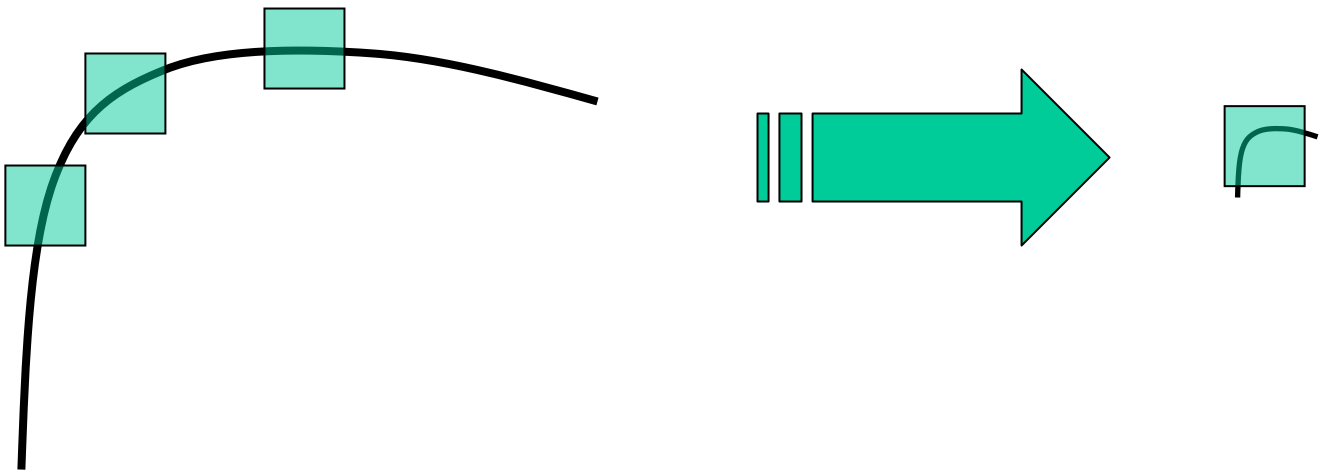
  ✓ Only derivatives are used => invariance to intensity shift $I \rightarrow I + b$

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



Harris Detector: Some Properties

- But: non-invariant to *image scale*!



All points will be classified as edges

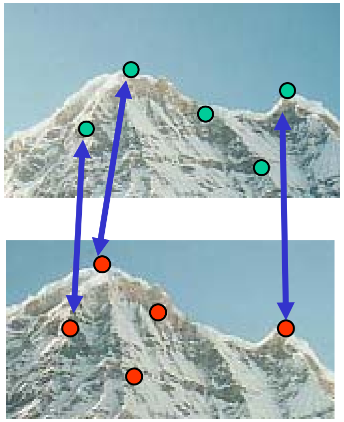
Corner !

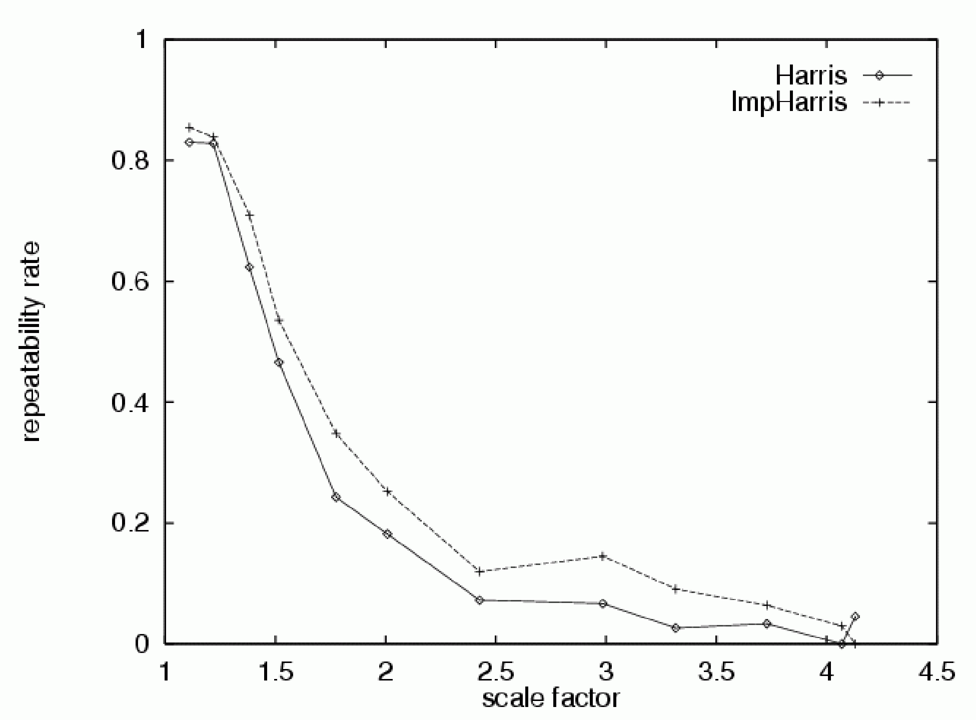
Harris Detector: Some Properties

- 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

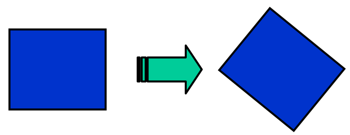
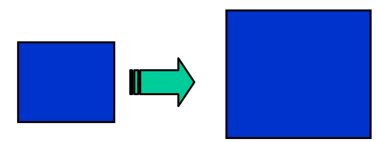
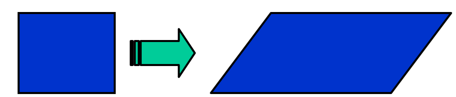
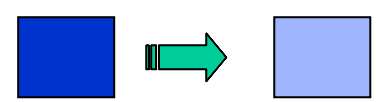
Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

We want to:

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

Models of Image Change

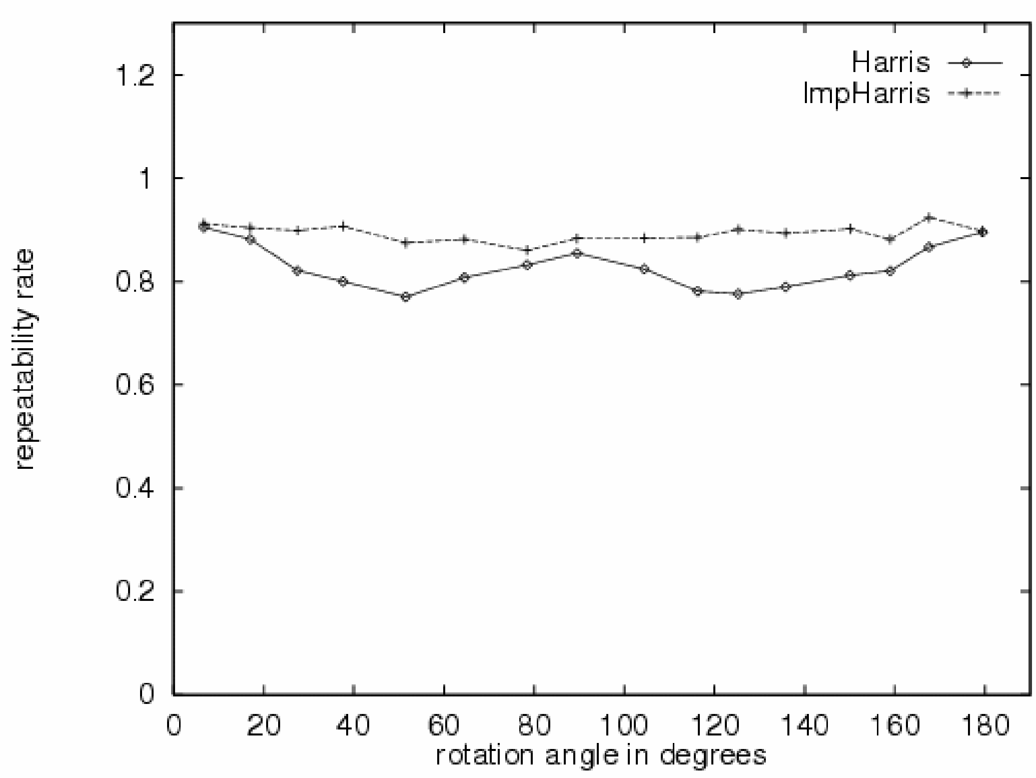
- Geometry
  - Rotation 
  - Similarity (rotation + uniform scale) 
  - Affine (scale dependent on direction) 
    valid for: orthographic camera, locally planar object
- Photometry
  - Affine intensity change ($I \rightarrow a\,I + b$) 

Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Rotation Invariant Detection

- Harris Corner Detector



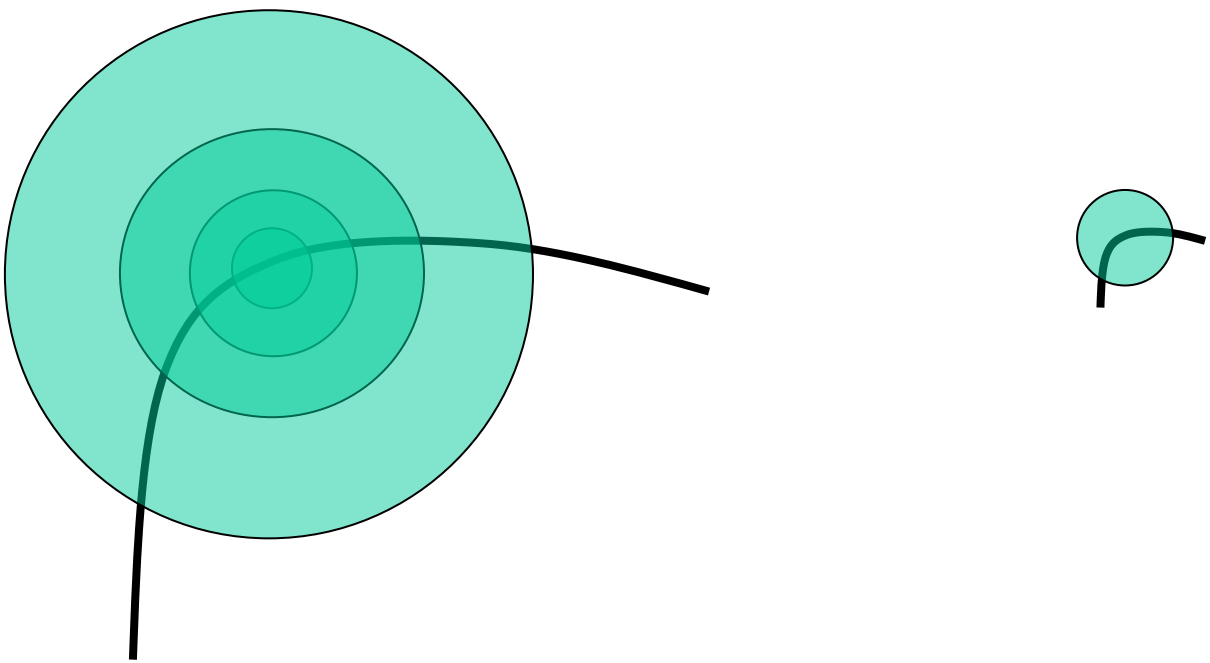
C.Schmid et.al. “Evaluation of Interest Point Detectors”. IJCV 2000

Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

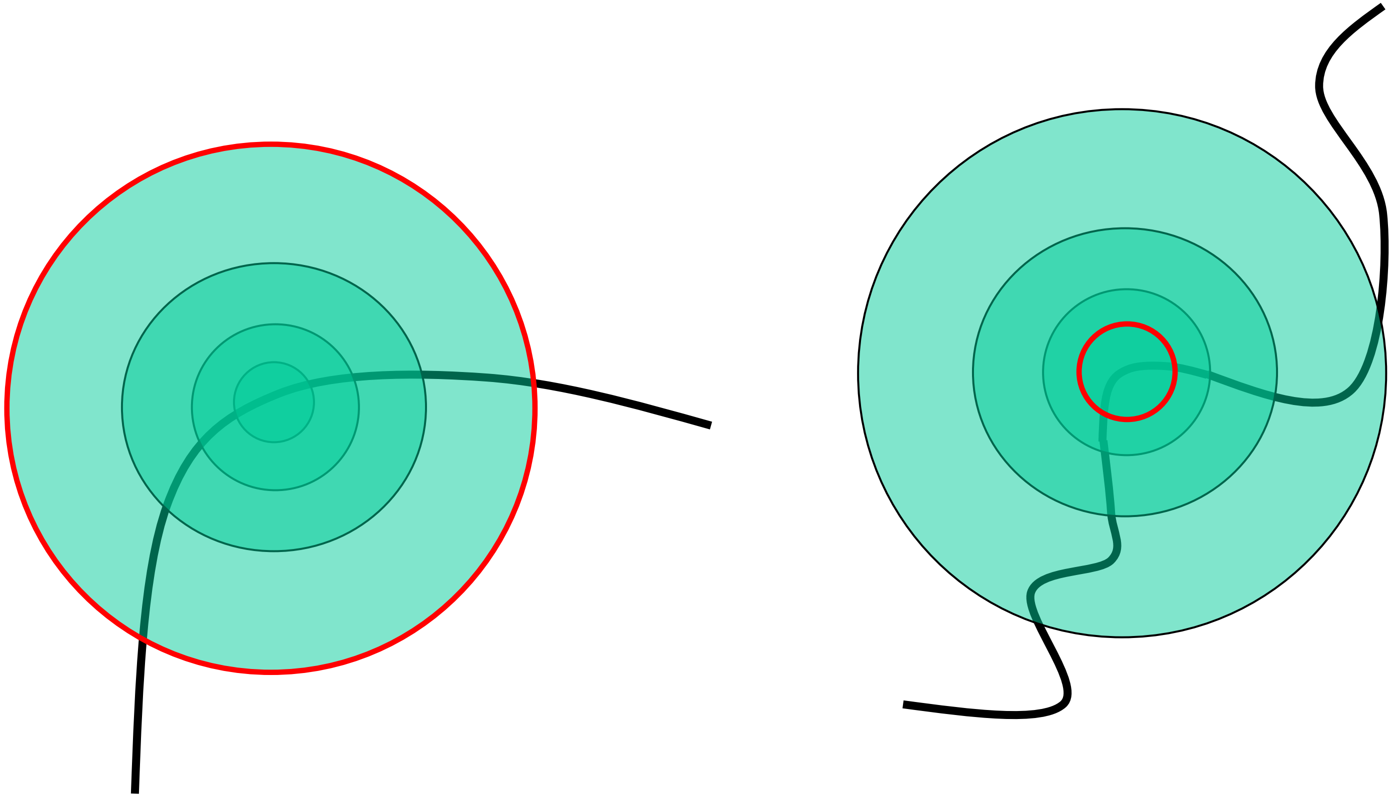
Scale Invariant Detection

- 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?

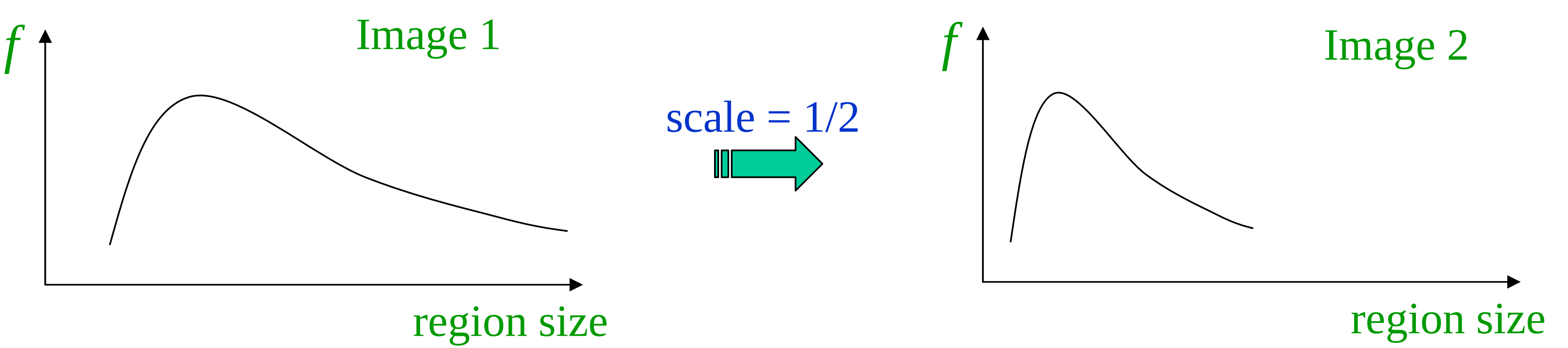


Scale Invariant Detection

- 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)



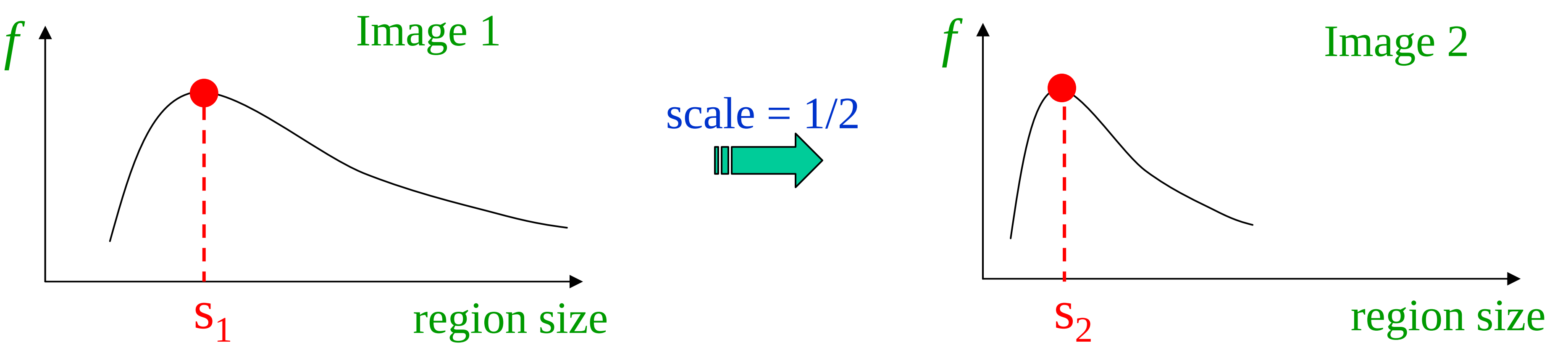
Scale Invariant Detection

- Common approach:

Take a local maximum of this function

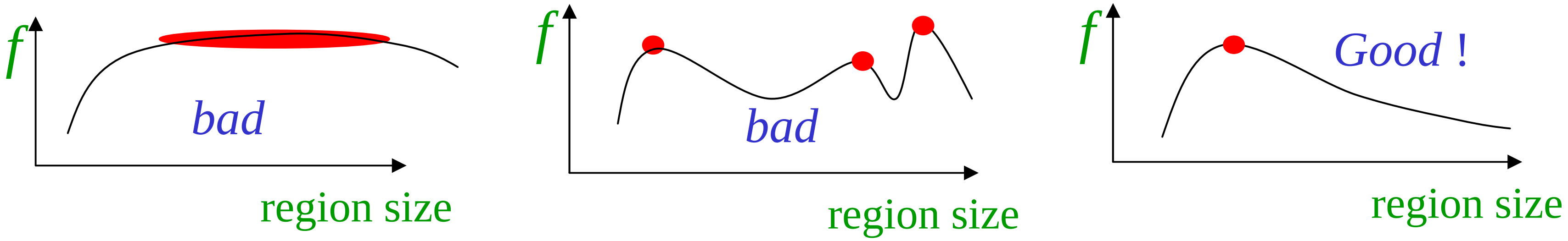
Observation: region size, for which the maximum is achieved, should be *invariant* to image scale.

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:

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

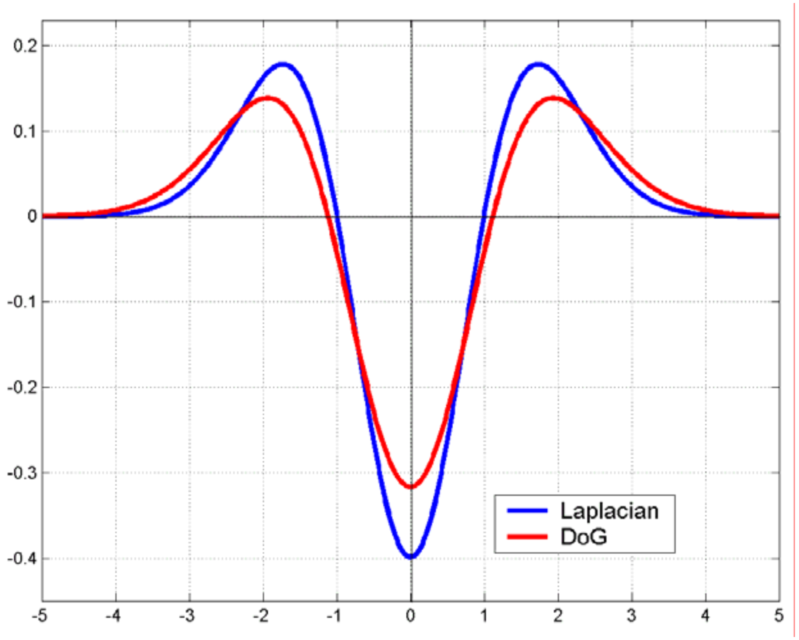
(Laplacian)

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

(Difference of Gaussians)

where Gaussian

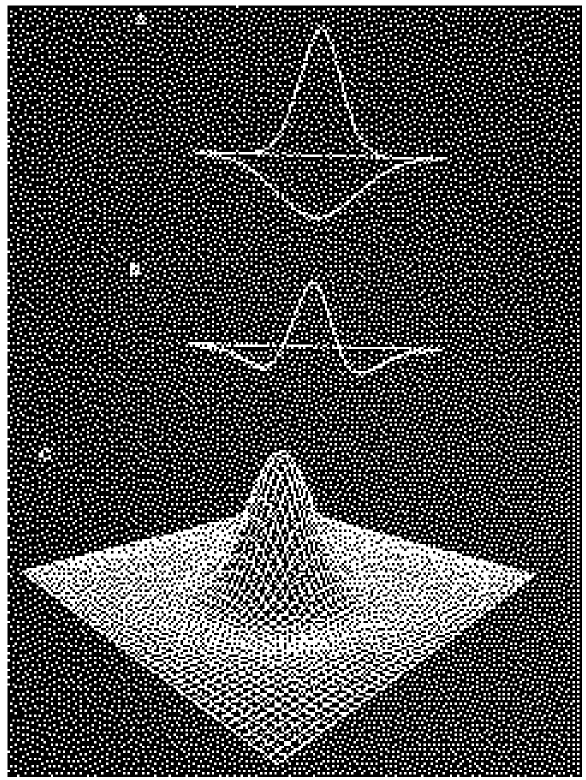
$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



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

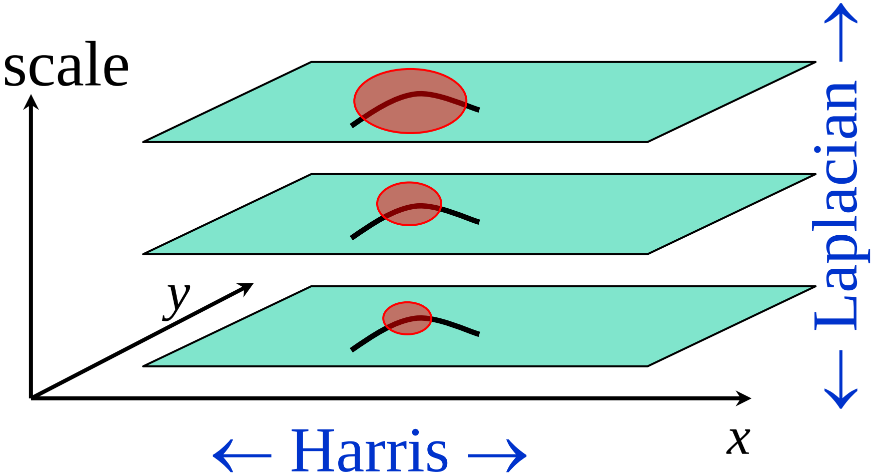
Scale Invariant Detection

- Compare to human vision: eye’s response

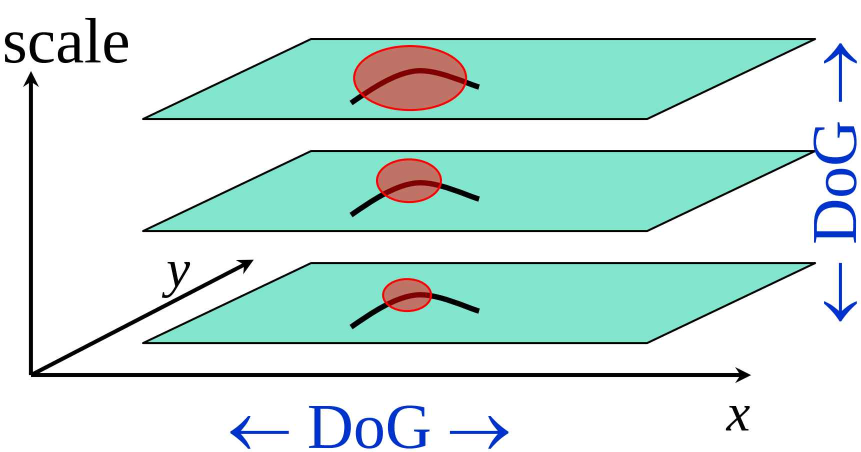


Scale Invariant Detectors

- 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

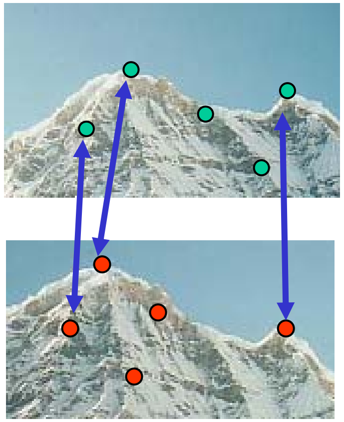
[2] D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". Accepted to IJCV 2004

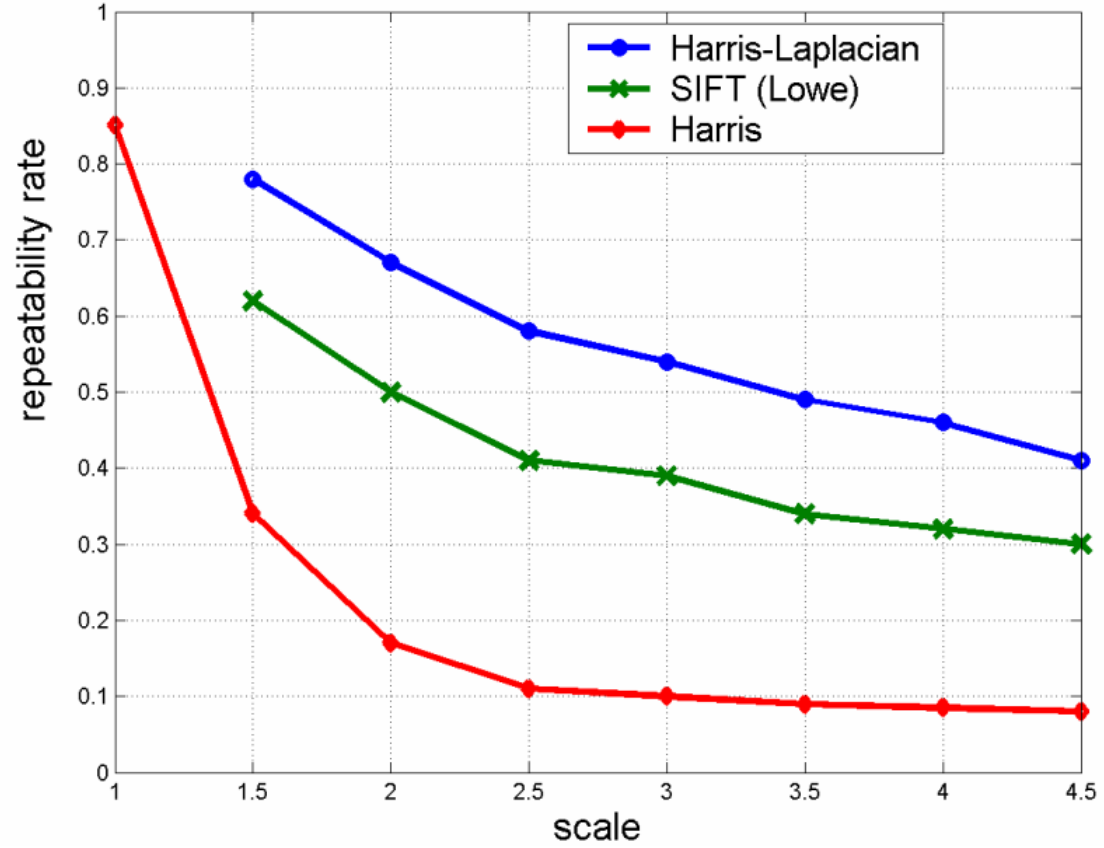
Scale Invariant Detectors

- 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

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:

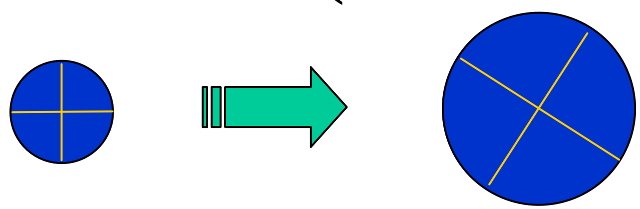
1. Harris-Laplacian [Mikolajczyk, Schmid]: maximize Laplacian over scale, Harris' measure of corner response over the image
2. SIFT [Lowe]: maximize Difference of Gaussians over scale and space

Contents

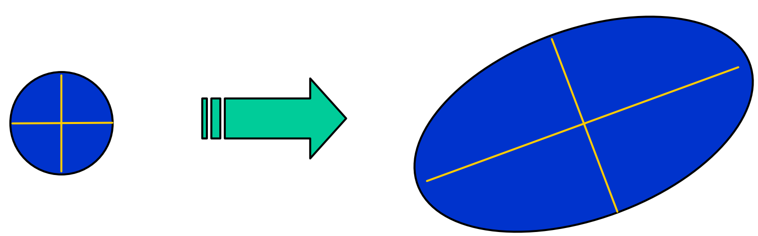
- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Affine Invariant Detection

- Above we considered:
  Similarity transform (rotation + uniform scale)

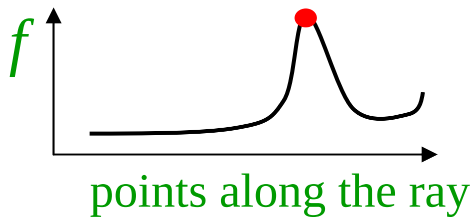


- Now we go on to:
  Affine transform (rotation + non-uniform scale)



Affine Invariant Detection

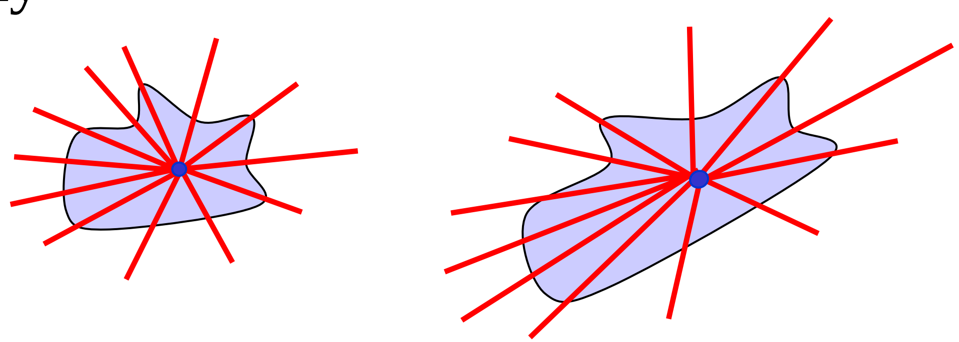
- Take a local intensity extremum as initial point
- Go along every ray starting from this point and stop when extremum of function f is reached



$$f(t) = \frac{\left| I(t) - I_0 \right|}{\frac{1}{t} \int_o^t \left| I(t) - I_0 \right| dt}$$

- We will obtain approximately corresponding regions

Remark: we search for scale in every direction



T.Tuytelaars, L.V.Gool. “Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions”. BMVC 2000.

Affine Invariant Detection

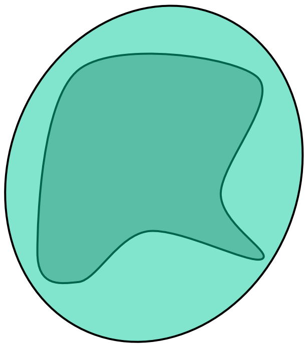
- The regions found may not exactly correspond, so we approximate them with ellipses

- Geometric Moments:

$$m_{pq} = \int_{\mathbb{R}^2} x^p y^q f(x, y)\,dxdy$$

Fact: moments m_{pq} uniquely determine the function f

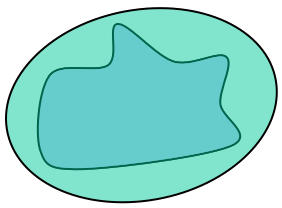
Taking f to be the characteristic function of a region (1 inside, 0 outside), moments of orders up to 2 allow to approximate the region by an ellipse



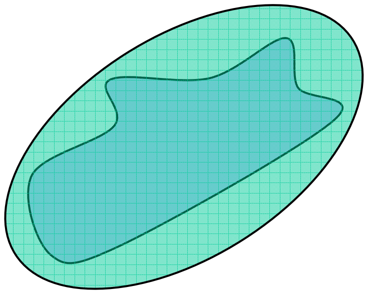
This ellipse will have the same moments of orders up to 2 as the original region

Affine Invariant Detection

- Covariance matrix of region points defines an ellipse:



$$q = Ap$$



$$p^T \Sigma_1^{-1} p = 1$$

$$q^T \Sigma_2^{-1} q = 1$$

$$\Sigma_1 = \left\langle pp^T \right\rangle_{\text{region 1}}$$

$$\Sigma_2 = \left\langle qq^T \right\rangle_{\text{region 2}}$$

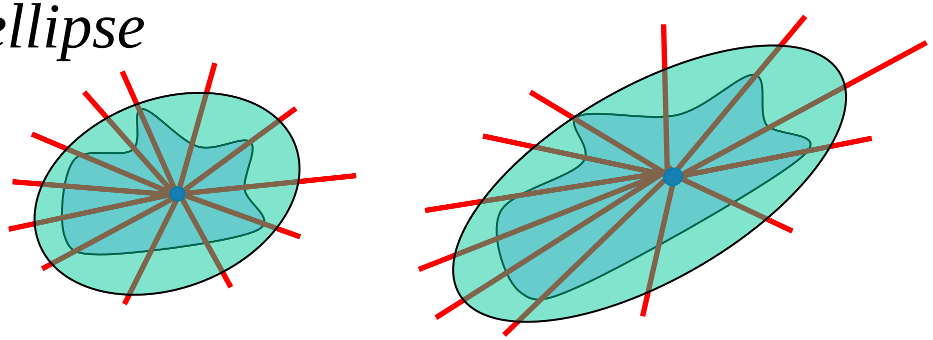
($p = [x, y]^T$ is relative to the center of mass)

$$\Sigma_2 = A \Sigma_1 A^T$$

Ellipses, computed for corresponding regions, also correspond!

Affine Invariant Detection

- Algorithm summary (detection of affine invariant region):
  - Start from a *local intensity extremum* point
  - Go in *every direction* until the point of extremum of some function f
  - Curve connecting the points is the region boundary
  - Compute *geometric moments* of orders up to 2 for this region
  - Replace the region with *ellipse*



T.Tuytelaars, L.V.Gool. “Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions”. BMVC 2000.

Affine Invariant Detection

- Maximally Stable Extremal Regions
  - *Threshold* image intensities: $I > I_0$
  - Extract *connected components* (“Extremal Regions”)
  - Find a threshold when an extremal region is “Maximally Stable”, i.e. *local minimum* of the relative growth of its square
  - Approximate a region with an *ellipse*

J.Matas et.al. “Distinguished Regions for Wide-baseline Stereo”. Research Report of CMP, 2001.

Affine Invariant Detection : Summary

- Under affine transformation, we do not know in advance shapes of the corresponding regions
- Ellipse given by geometric covariance matrix of a region robustly approximates this region
- For corresponding regions ellipses also correspond

Methods:

1. Search for extremum along rays [Tuytelaars, Van Gool]:
2. Maximally Stable Extremal Regions [Matas et.al.]

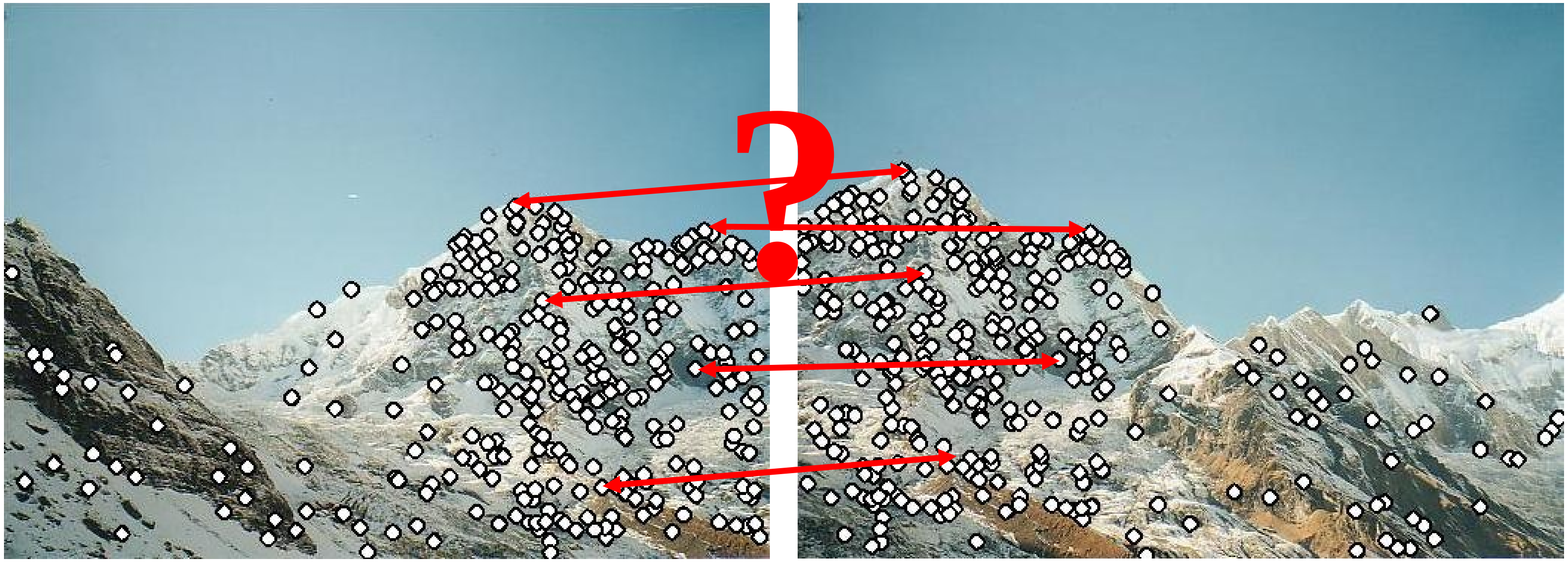
Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Point Descriptors

- We know how to detect points
- Next question:

How to match them?



Point descriptor should be:

1. Invariant
2. Distinctive

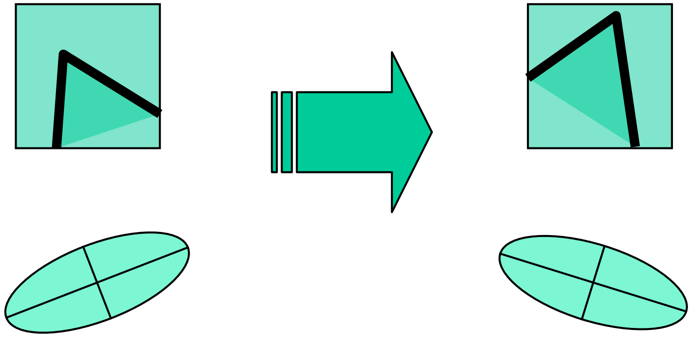
Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Descriptors Invariant to Rotation

- Harris corner response measure:
  depends only on the eigenvalues of the matrix *M*

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



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

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:

$$|m_{kl}|$$

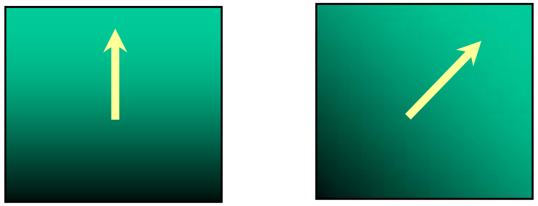
Matching is done by comparing vectors $[|m_{kl}|]_{k,l}$

J.Matas et.al. “Rotational Invariants for Wide-baseline Stereo”. Research Report of CMP, 2003

Descriptors Invariant to Rotation

- Find local orientation

Dominant direction of gradient



- Compute image derivatives relative to this orientation

[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

Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

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

Contents

- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Affine Invariant Descriptors

- Affine invariant color moments

$$m_{pq}^{abc} = \int_{region} x^p y^q R^a(x,y) G^b(x,y) B^c(x,y) dxdy$$

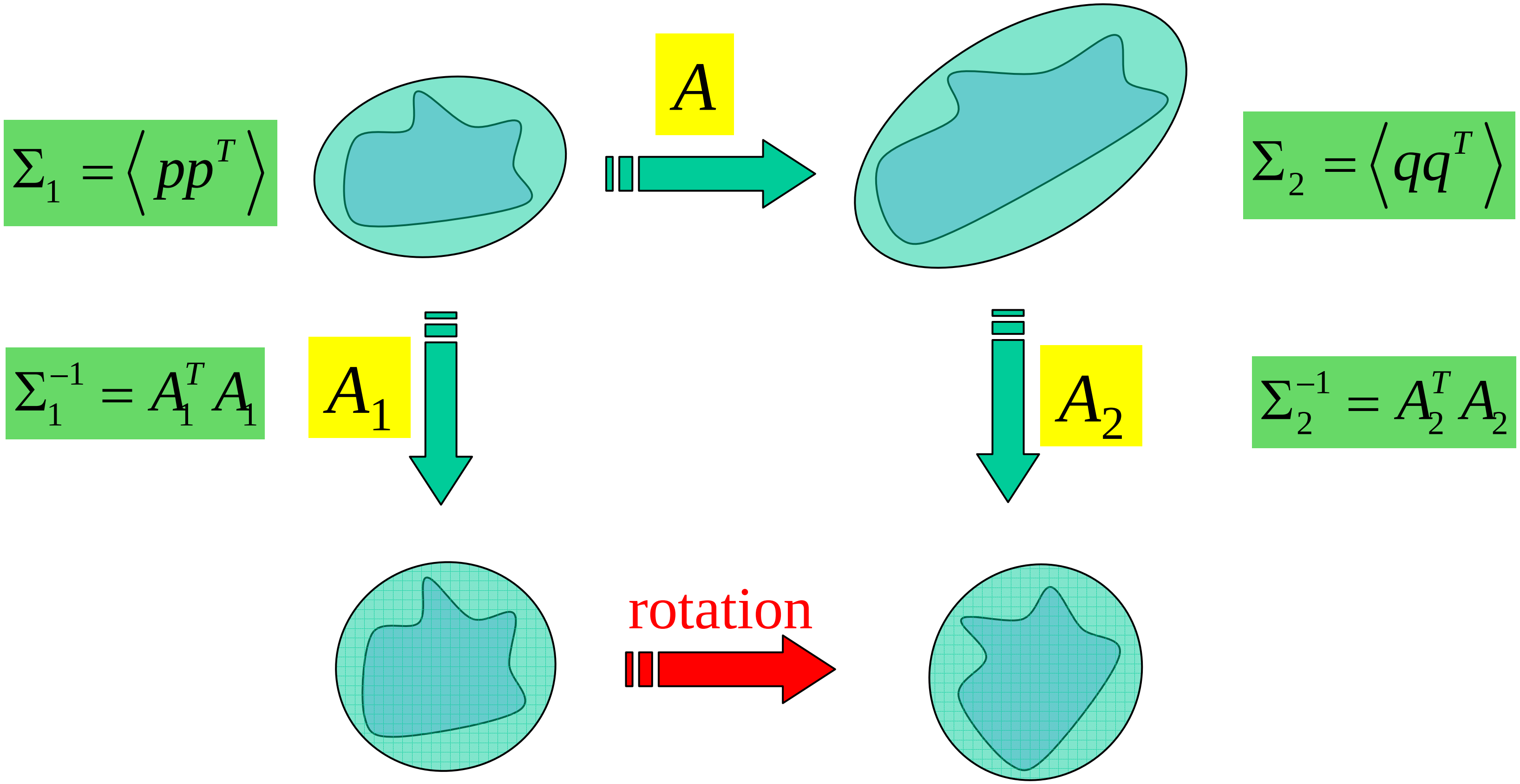
Different combinations of these moments are fully affine invariant

Also invariant to affine transformation of intensity $I \rightarrow a\,I + b$

F.Mindru et.al. "Recognizing Color Patterns Irrespective of Viewpoint and Illumination". CVPR99

Affine Invariant Descriptors

- Find affine normalized frame



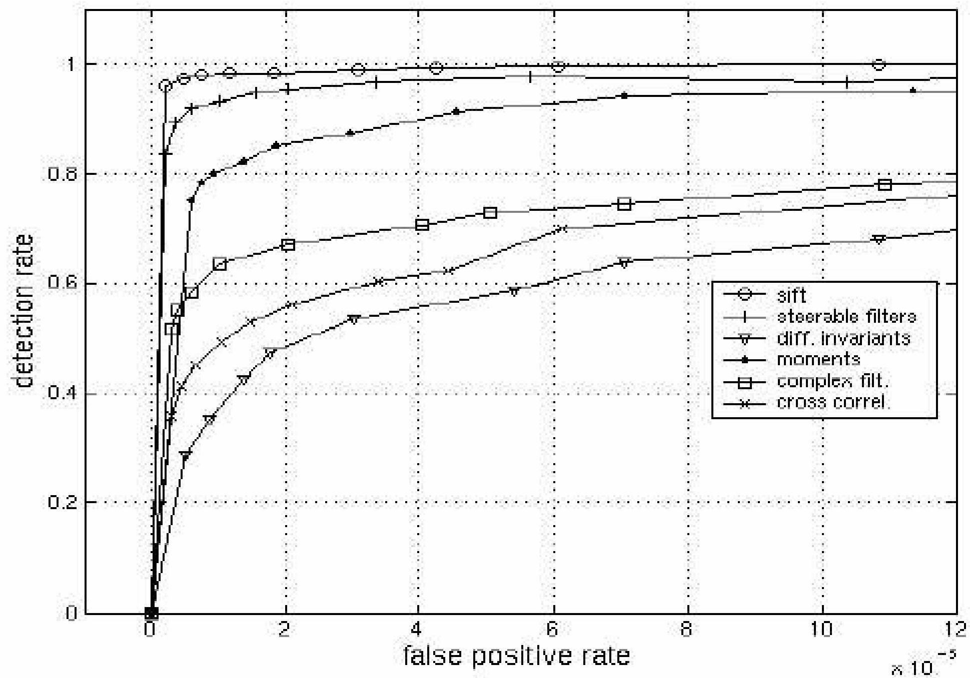
- Compute rotational invariant descriptor in this normalized frame

J.Matas et.al. “Rotational Invariants for Wide-baseline Stereo”. Research Report of CMP, 2003

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

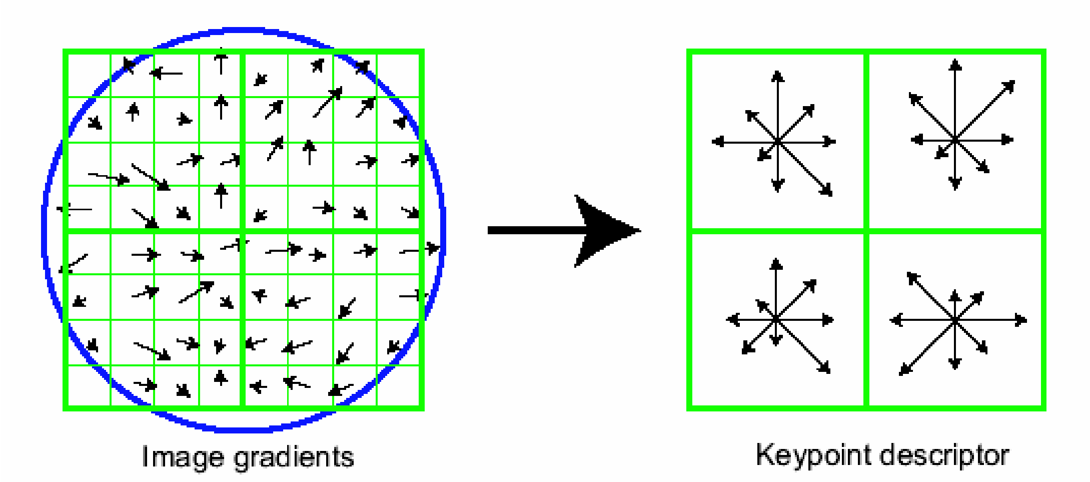
Scale = 2.5
Rotation = 45^0



[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

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

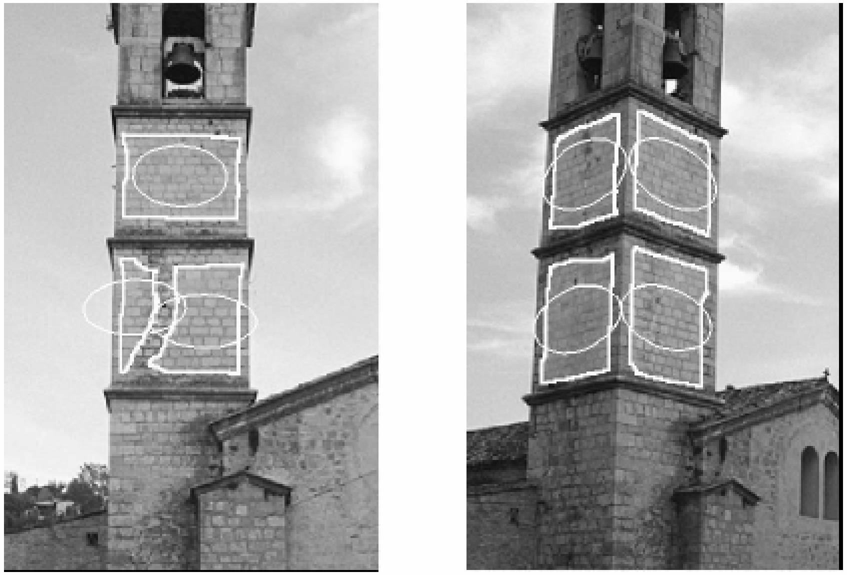


D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". Accepted to IJCV 2004

Affine Invariant Texture Descriptor

- Segment the image into regions of different textures (by a non-invariant method)
- Compute matrix *M* (the same as in Harris detector) over these regions

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



- This matrix defines the ellipse

$$[x, y] M \begin{bmatrix} x \\ y \end{bmatrix} = 1$$

- Regions described by these ellipses are invariant under affine transformations
- Find affine normalized frame
- Compute rotation invariant descriptor

F.Schaffalitzky, A.Zisserman. "Viewpoint Invariant Texture Matching and Wide Baseline Stereo". ICCV 2003

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)

Talk Resume

- Stable (repeatable) feature points can be detected regardless of image changes
  - Scale: search for correct scale as *maximum* of appropriate function
  - Affine: approximate regions with *ellipses* (this operation is affine invariant)
- Invariant and distinctive descriptors can be computed
  - Invariant *moments*
  - *Normalizing* with respect to scale and affine transformation

Evaluation of interest points and descriptors

Cordelia Schmid

CVPR’03 Tutorial

Introduction

- Quantitative evaluation of interest point detectors
  - points / regions at the same relative location

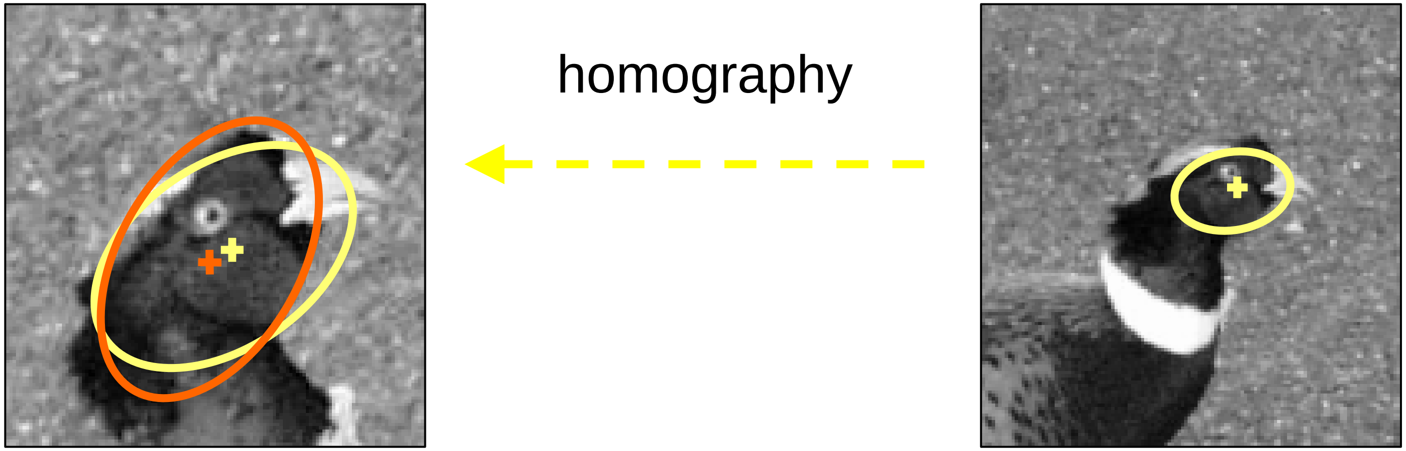
  => repeatability rate

- Quantitative evaluation of descriptors
  - distinctiveness

  => detection rate with respect to false positives

Quantitative evaluation of detectors

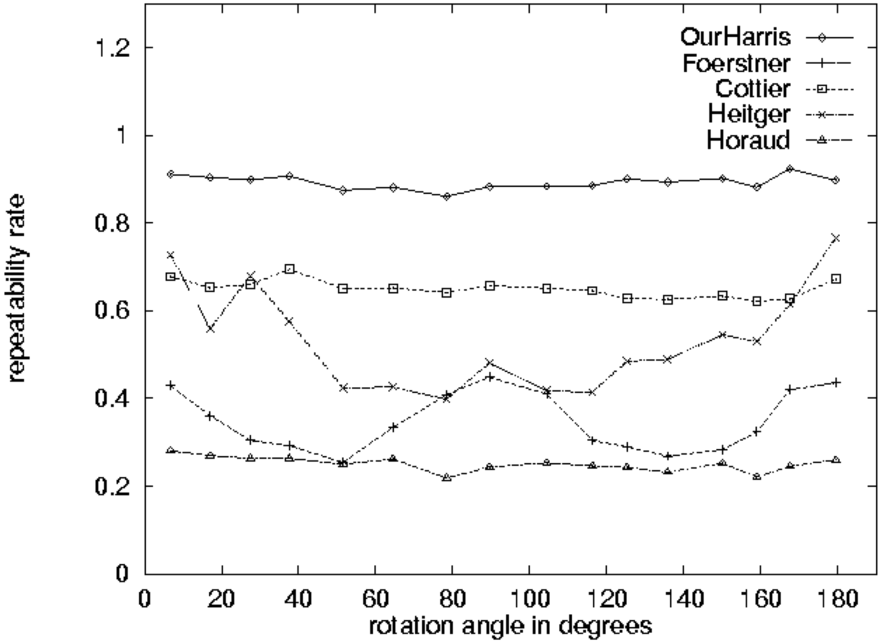
- Repeatability rate : percentage of corresponding points



- Two points are corresponding if
  1. The location error is less than 1.5 pixel
  2. The intersection error is less than 20%

Comparison of different detectors

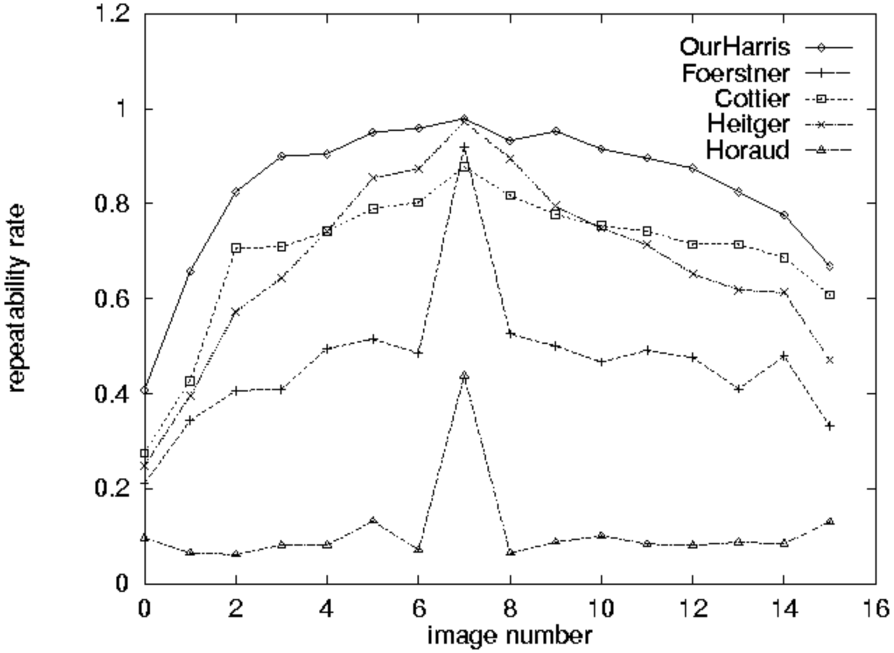
repeatability - image rotation



[Comparing and Evaluating Interest Points, Schmid, Mohr & Bauckhage, ICCV 98]

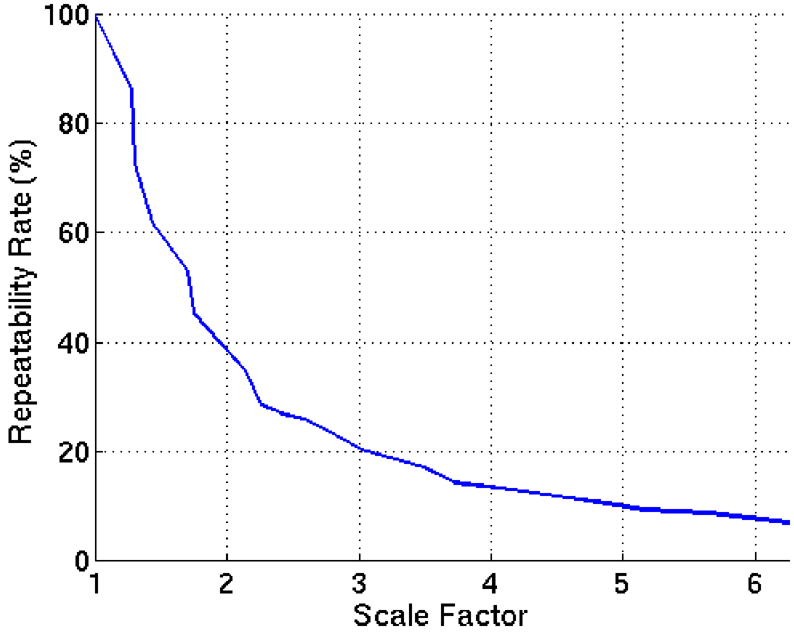
Comparison of different detectors

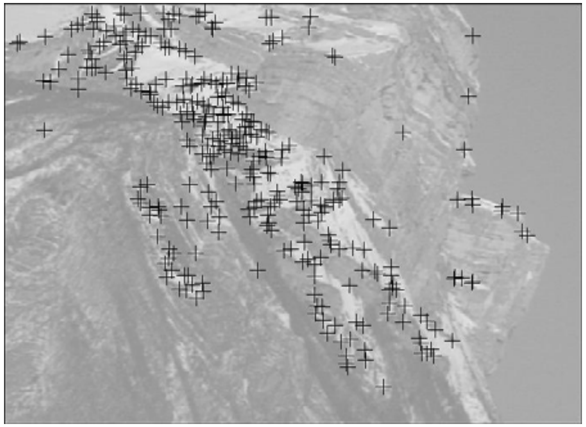
repeatability – perspective transformation

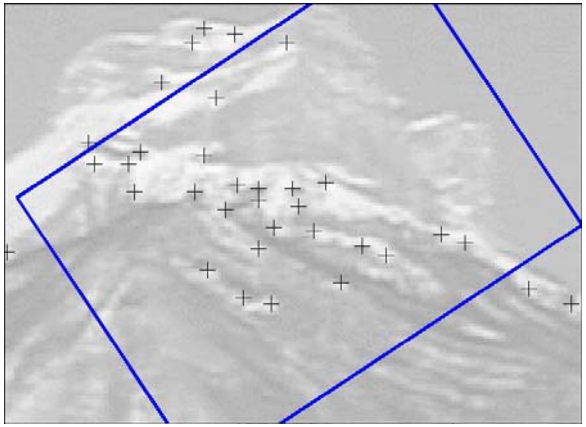


[Comparing and Evaluating Interest Points, Schmid, Mohr & Bauckhage, ICCV 98]

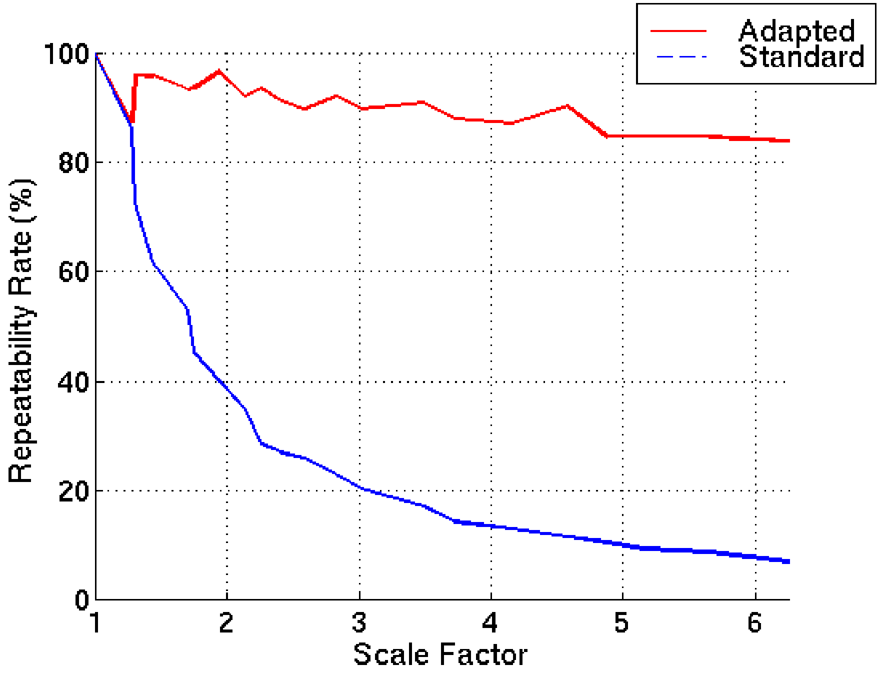
Harris detector + scale changes

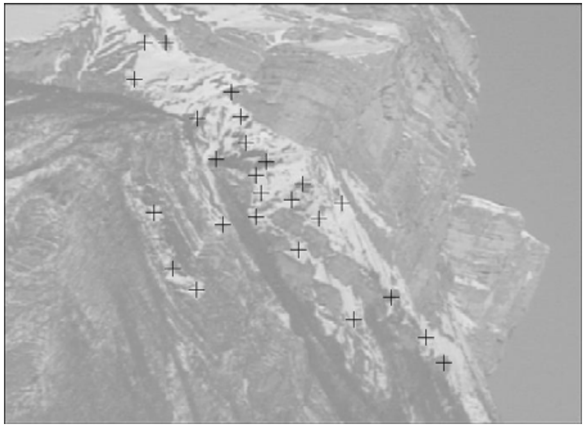


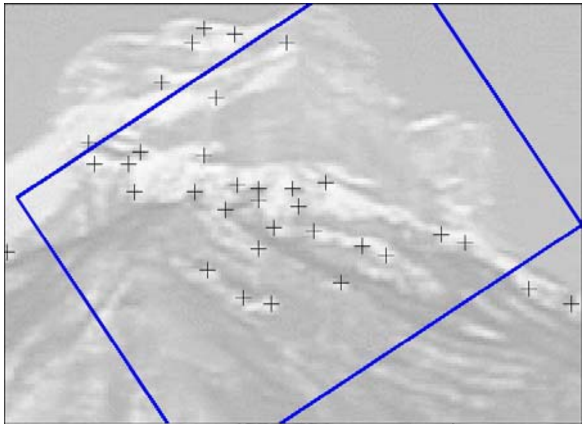




Harris detector – adaptation to scale

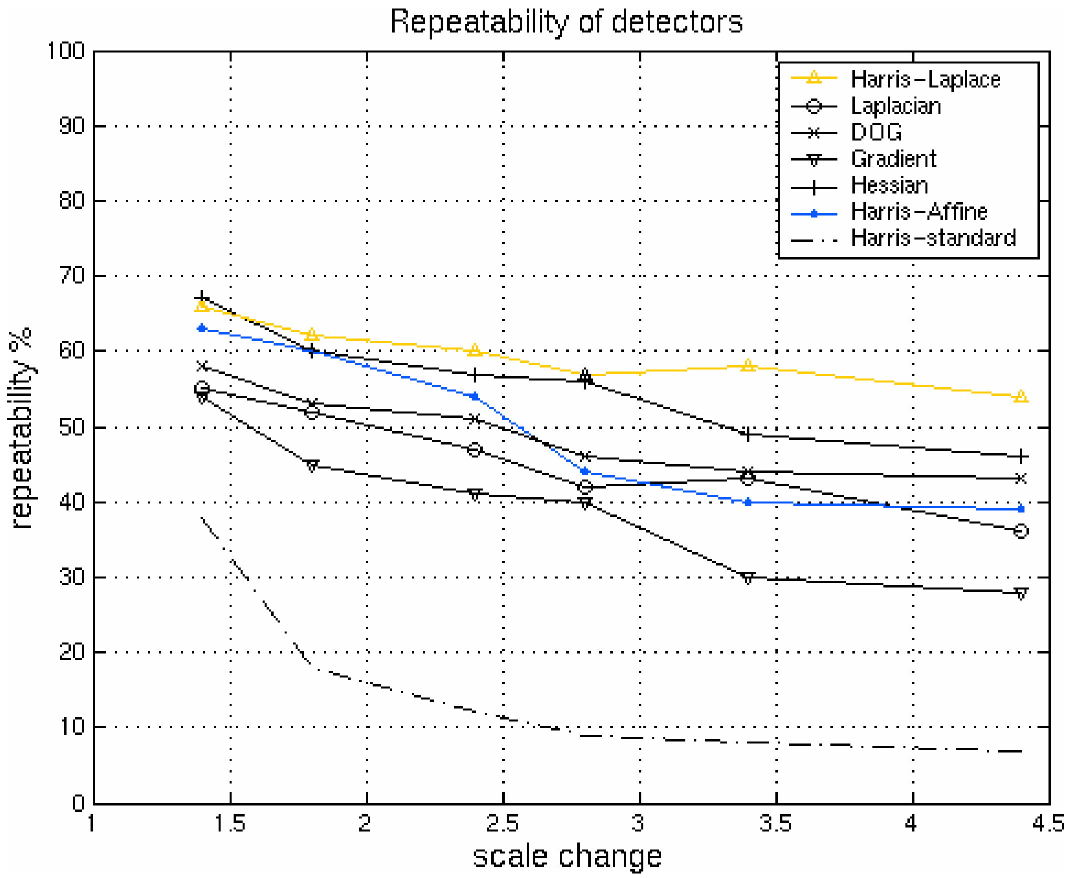






Evaluation of scale invariant detectors

repeatability – scale changes



Evaluation of affine invariant detectors

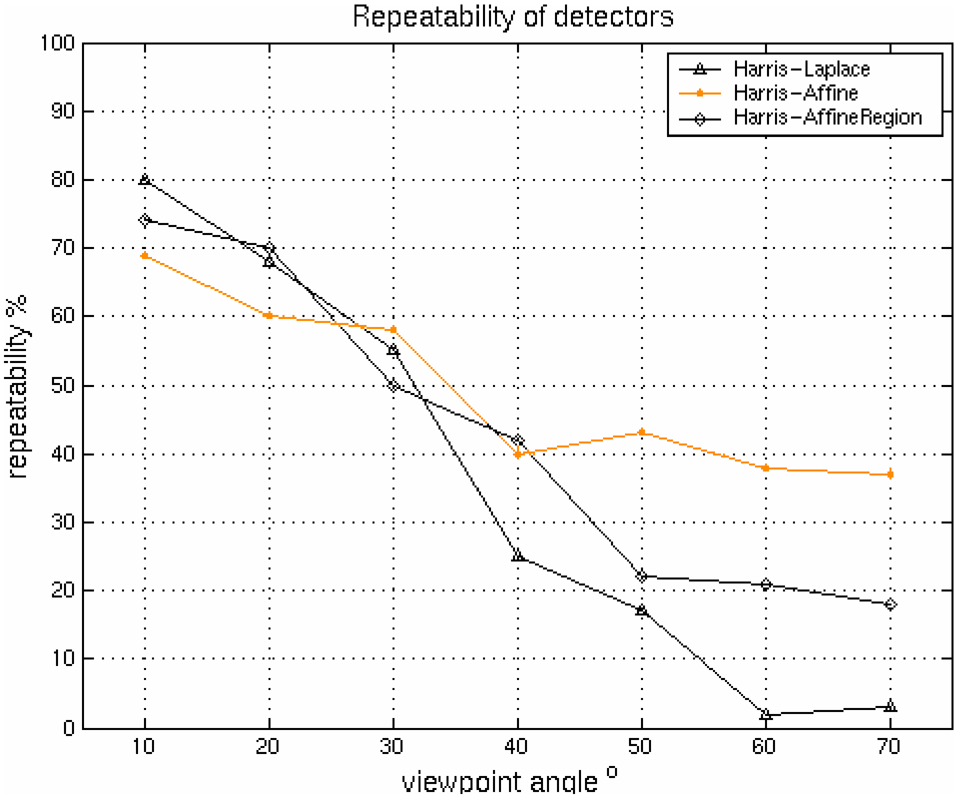
repeatability – perspective transformation

0

40

60

70



Quantitative evaluation of descriptors

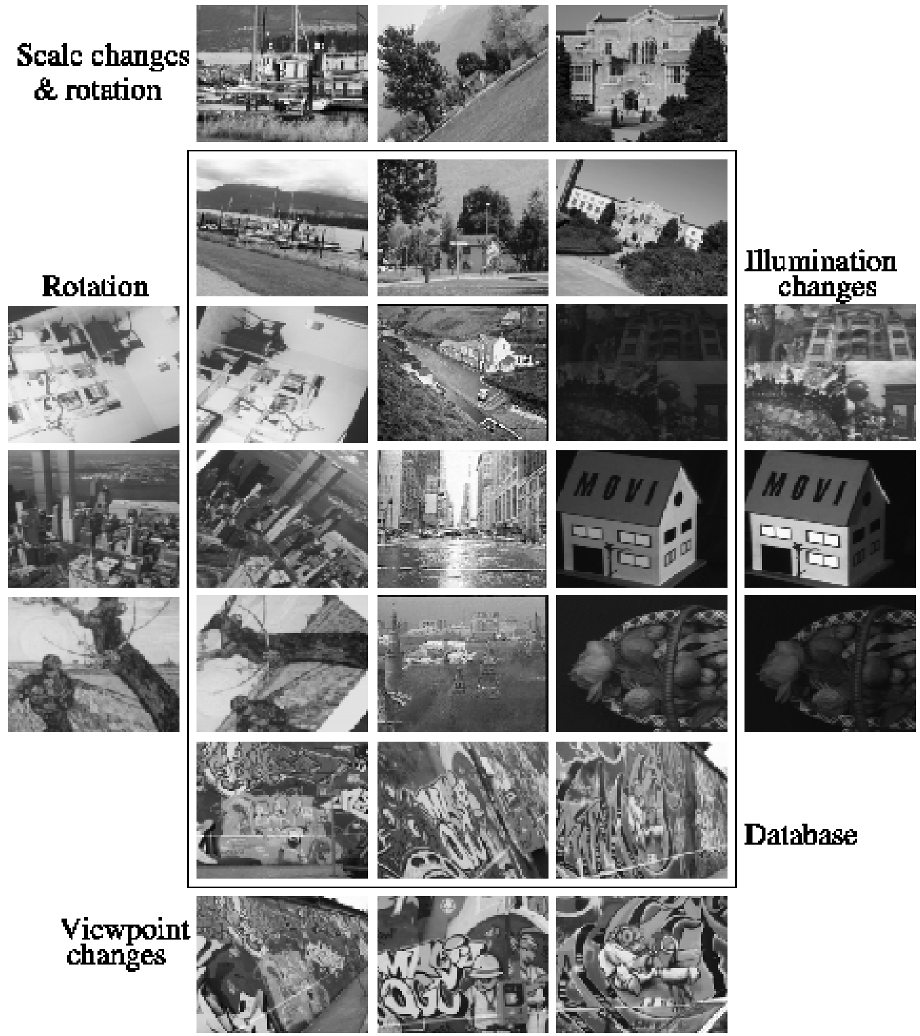
- Evaluation of different local features
  - SIFT, steerable filters, differential invariants, moment invariants, cross-correlation

- Measure : distinctiveness
  - receiver operating characteristics of
    detection rate with respect to false positives

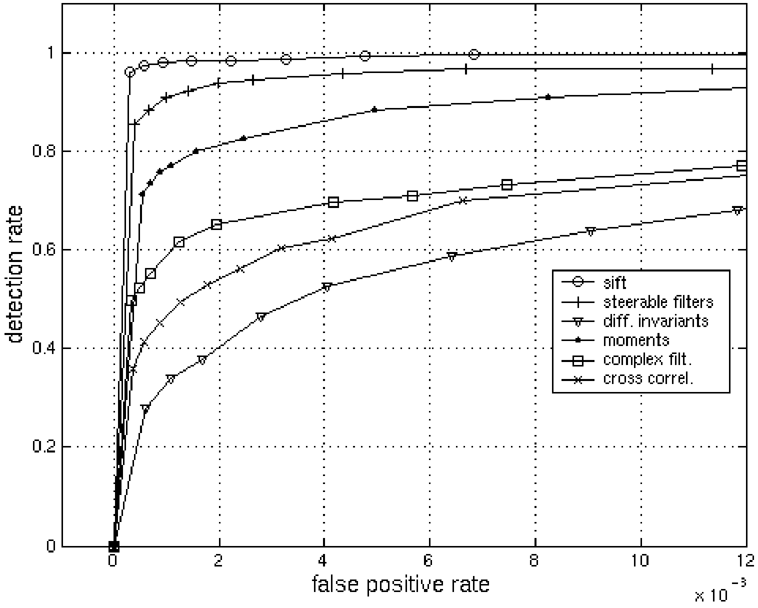
  - detection rate = correct matches / possible matches
  - false positives = false matches / (database points * query points)

[A performance evaluation of local descriptors, Mikolajczyk & Schmid, CVPR’03]

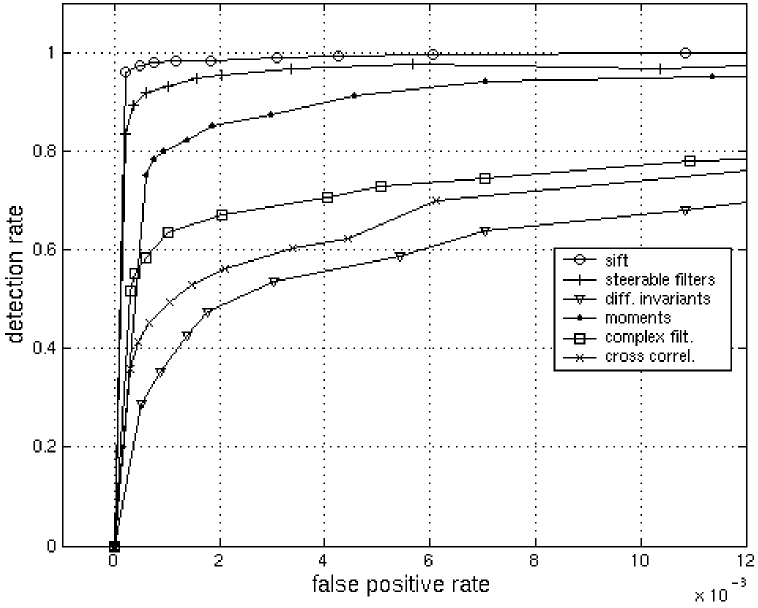
Experimental evaluation



Scale change (factor 2.5)

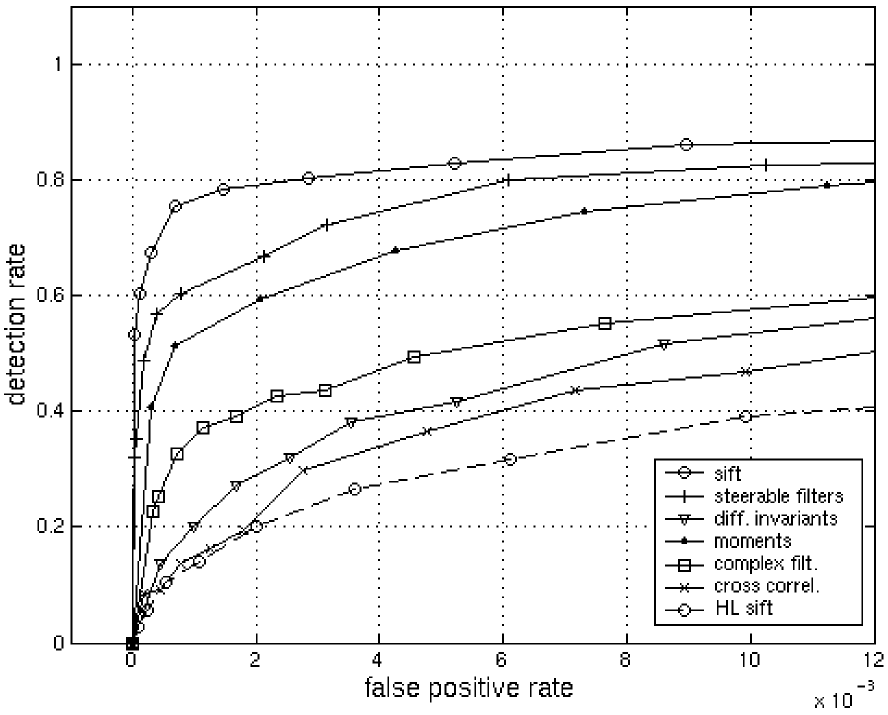


Harris-Laplace



DoG

Viewpoint change (60 degrees)



Harris-Affine (Harris-Laplace)

Descriptors - conclusion

- SIFT + steerable perform best
- Performance of the descriptor independent of the detector
- Errors due to imprecision in region estimation, localization

shape context slides

- Slides from Jitendra Malik, U.C. Berkeley

Shape context application: CAPTCHA