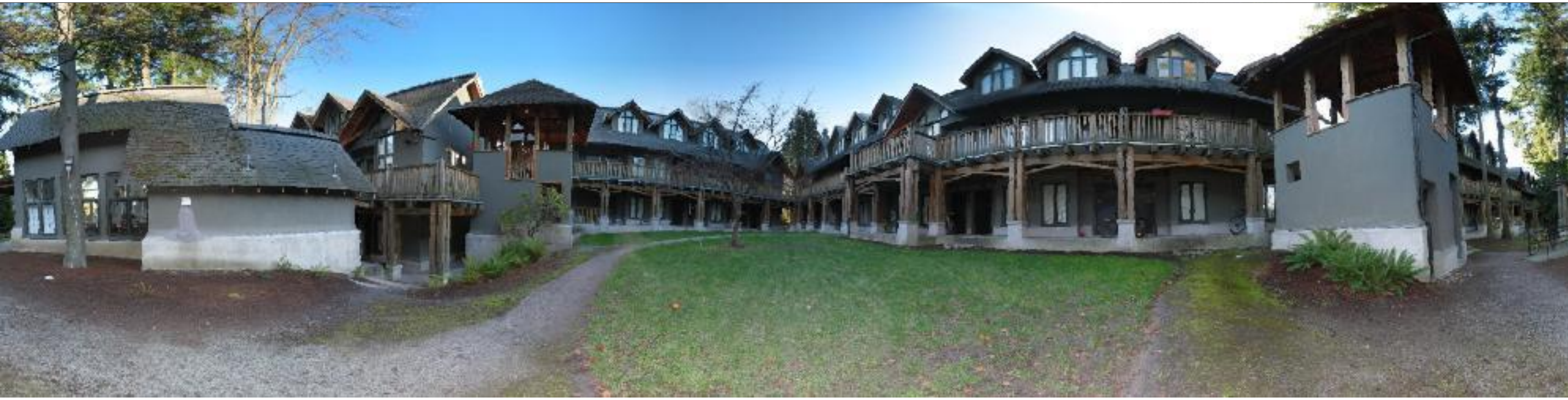
Local Invariant Features

This is a compilation of slides by:

Darya Frolova, Denis Simakov,The Weizmann Institute of Science
Jiri Matas, Martin Urban Center for Machine Percpetion Prague
Matthew Brown,David Lowe, University of British Columbia

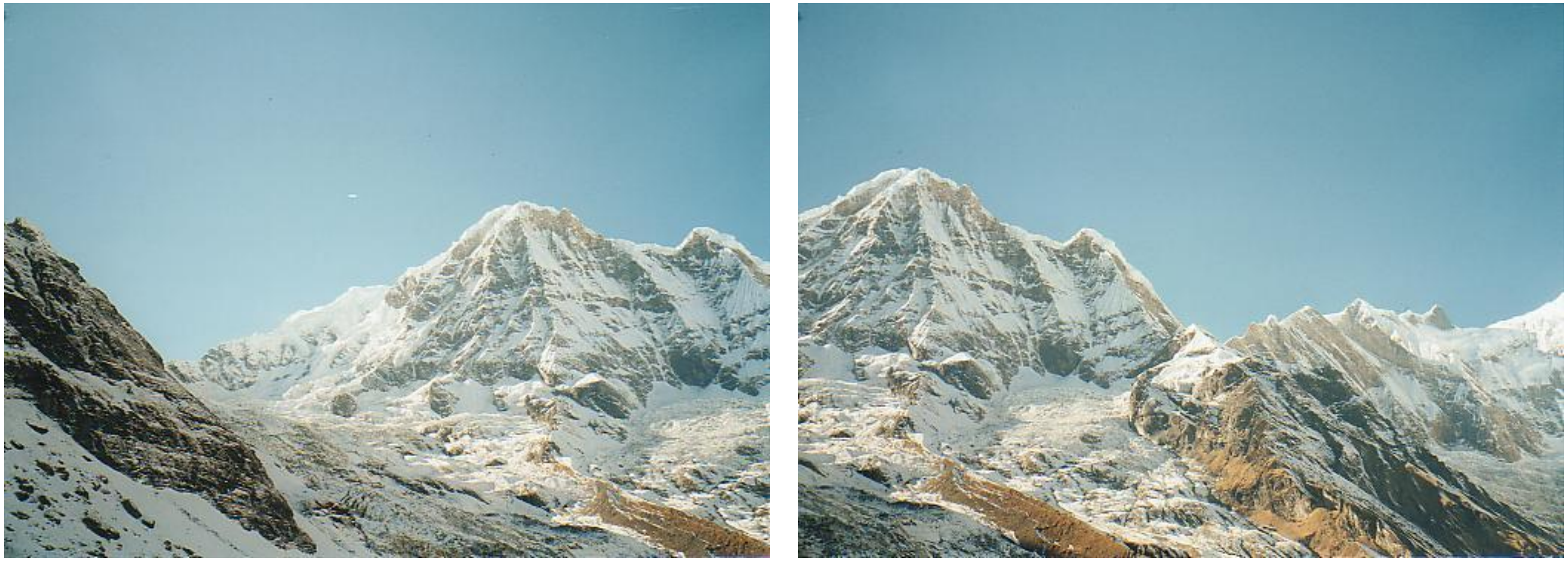
Building a Panorama



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

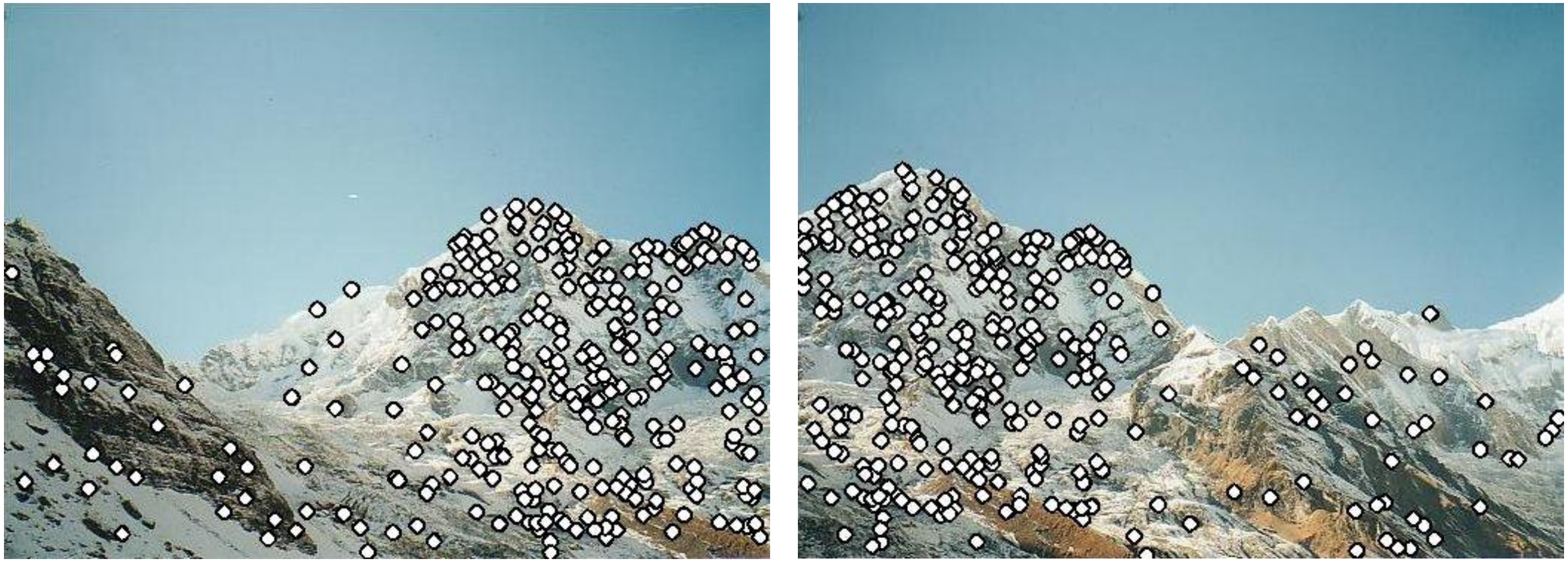
How do we build panorama?

- We need to match (align) images



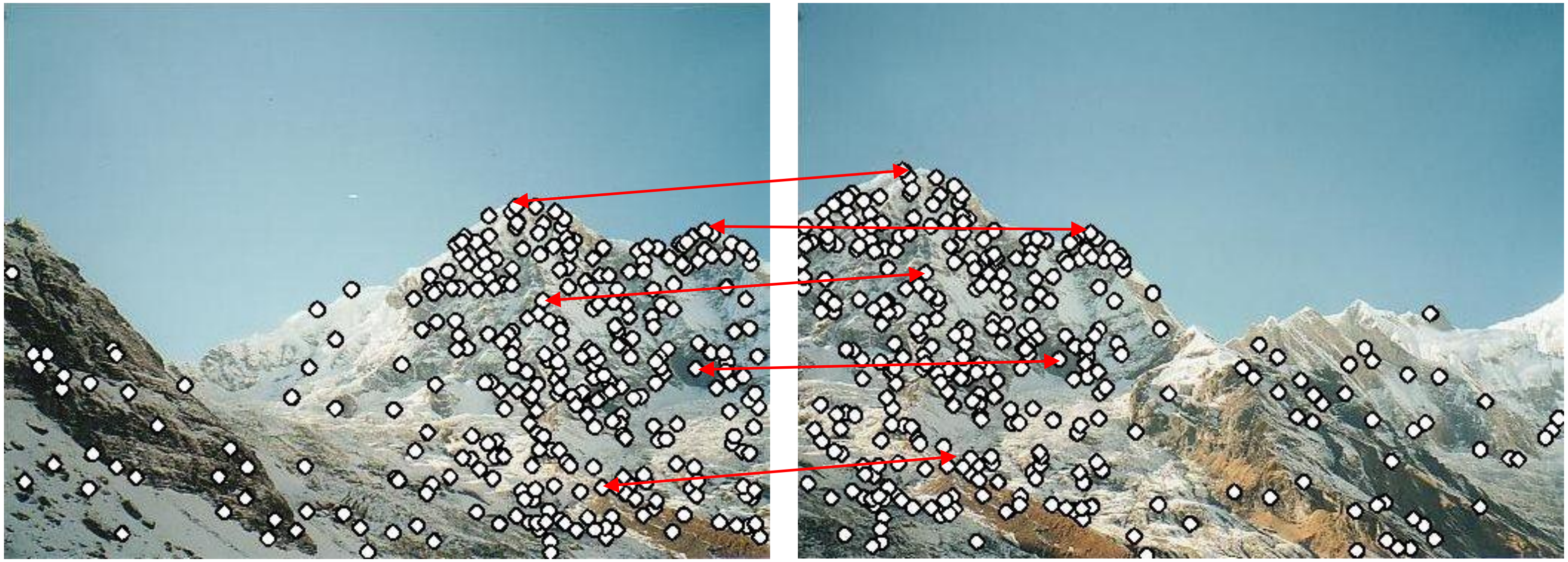
Matching with Features

- Detect feature points in both images



Matching with Features

- Detect feature points in both images
- Find corresponding pairs

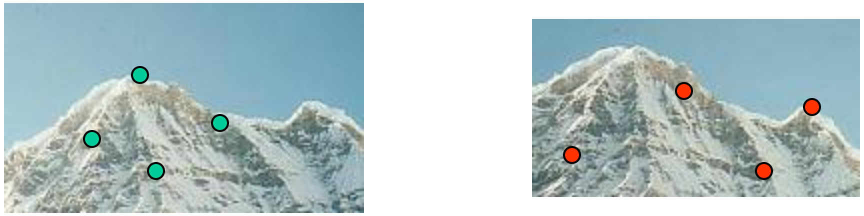


Matching with Features

- 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

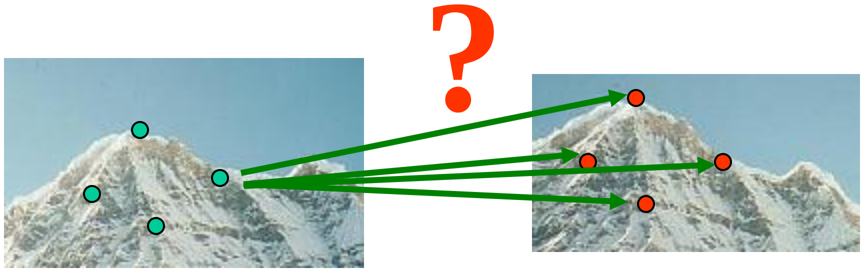


no chance to match!

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

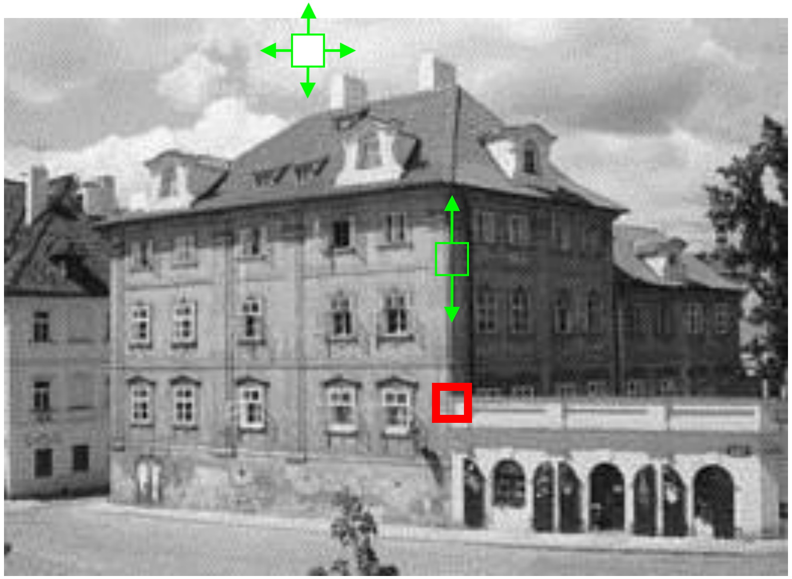
More motivation…

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

Selecting Good Features

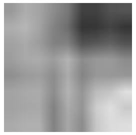
- What’s a “good feature”?
  - Satisfies brightness constancy
  - Has sufficient texture variation
  - Does not have too much texture variation
  - Corresponds to a “real” surface patch
  - Does not deform too much over time

Corner Detection: Introduction



undistinguished patches:

distinguished patches:

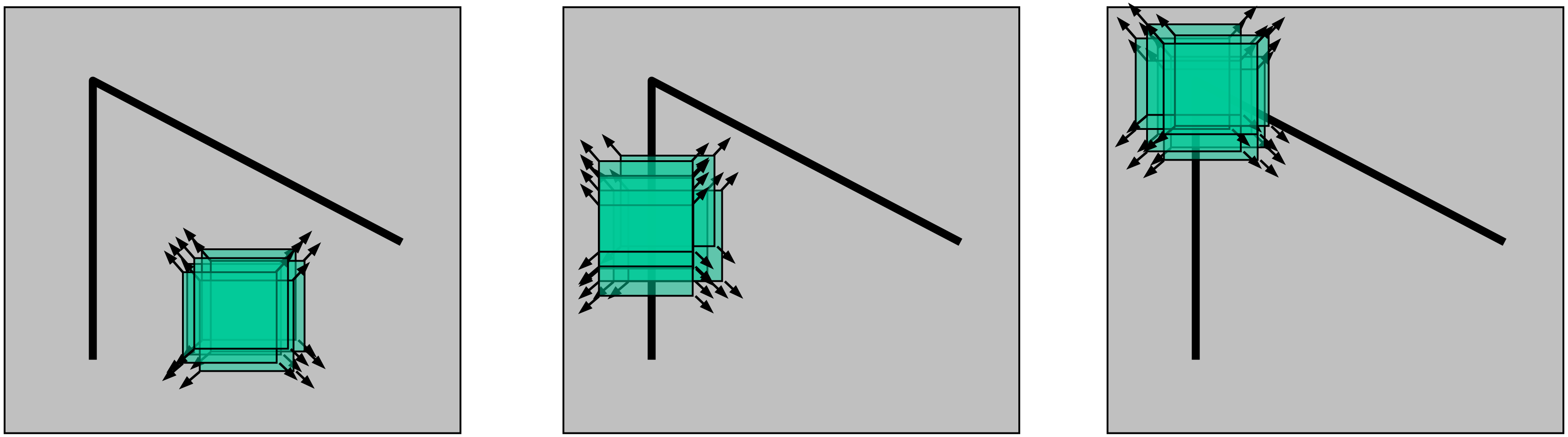


“Corner” (“interest point”) detector detects points with distinguished neighbourhood(*) well suited for matching verification.

Detectors

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

Harris Detector: Basic Idea



"flat" region:
no change in
all directions

"edge":
no change along
the edge direction

"corner":
significant change
in all directions

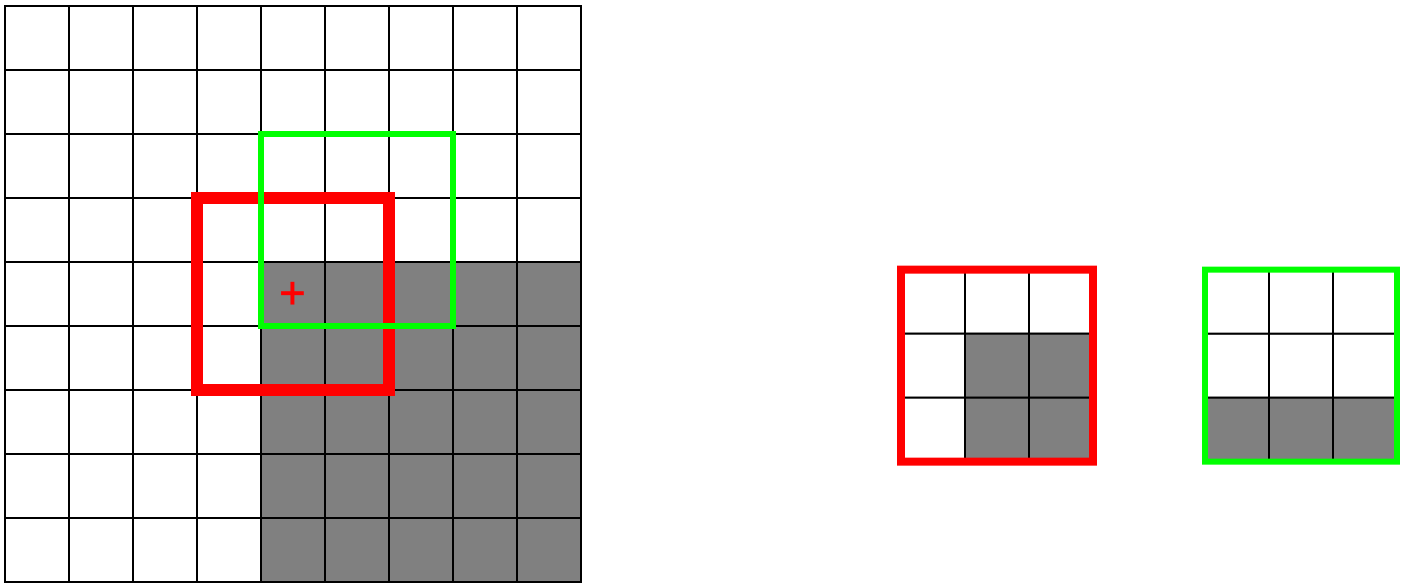
- We should easily recognize the point by looking through a small window
- Shifting a window in *any direction* should give *a large change* in intensity

Harris detector

Based on the idea of auto-correlation

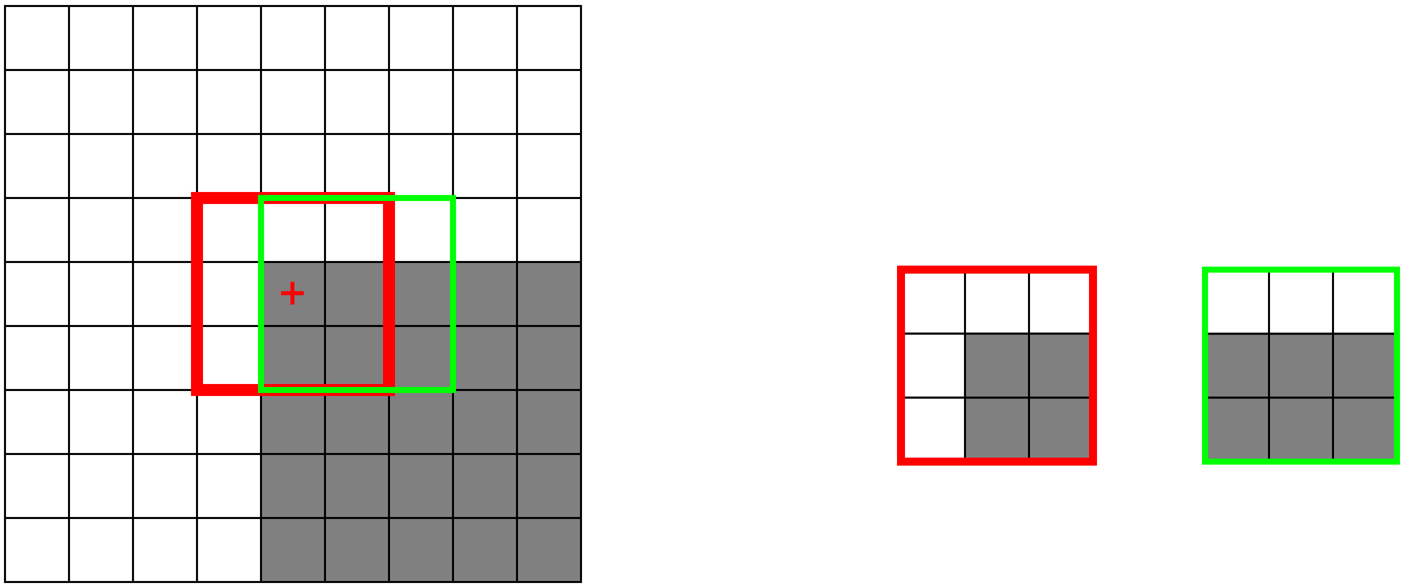
Important difference in all directions => interest point

Corner Detection: Introduction



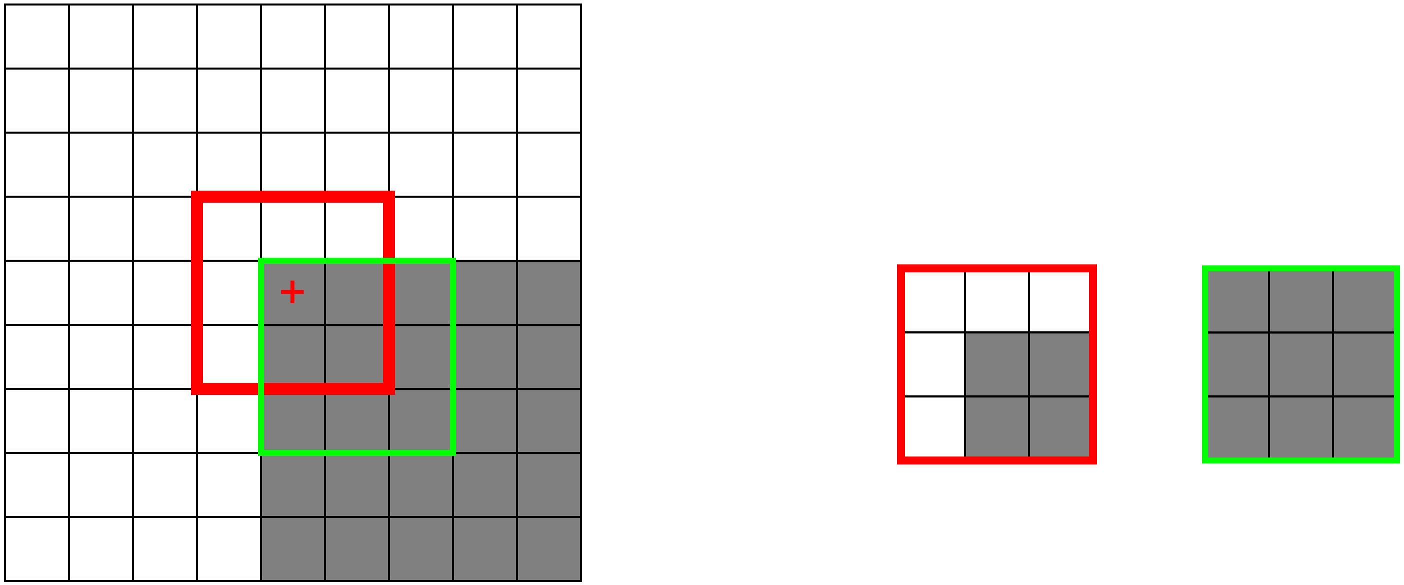
Demo of a point + with well distinguished neighbourhood.

Corner Detection: Introduction



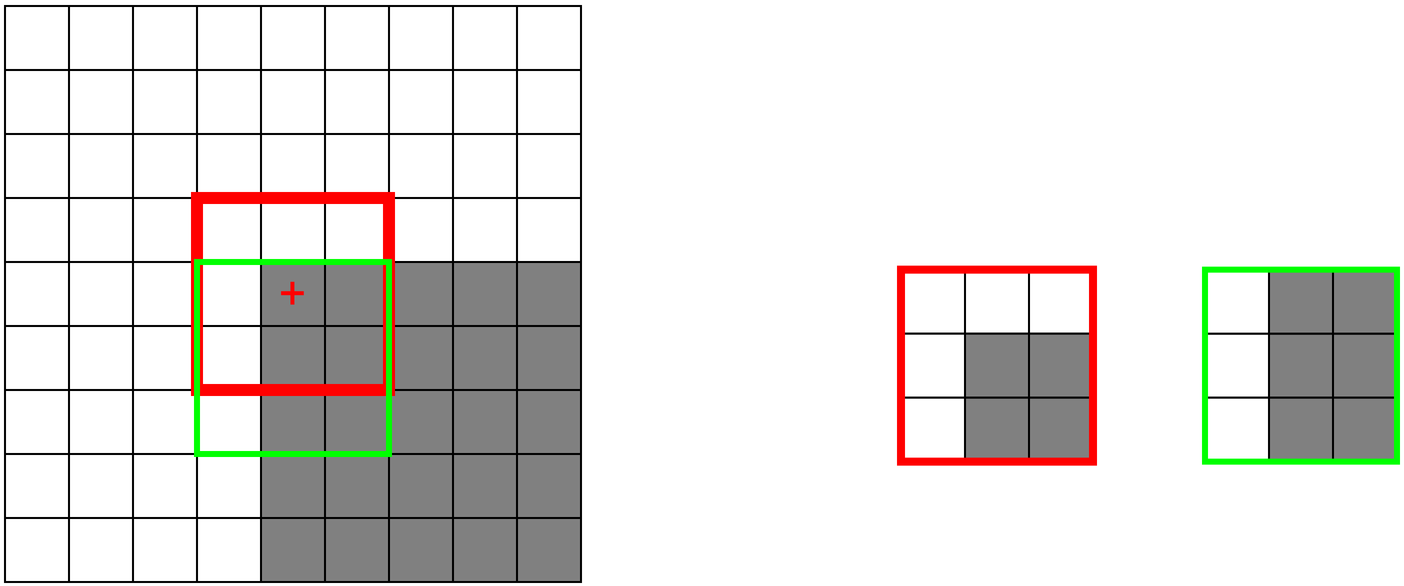
Demo of a point + with well distinguished neighbourhood.

Corner Detection: Introduction



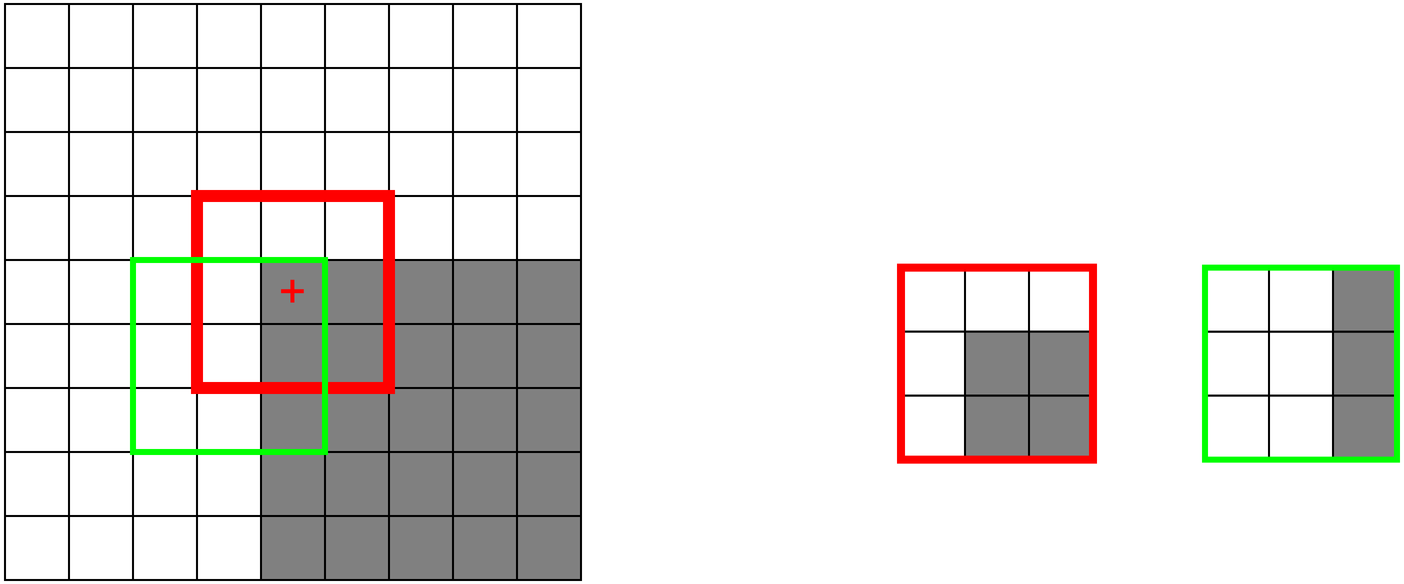
Demo of a point + with well distinguished neighbourhood.

Corner Detection: Introduction



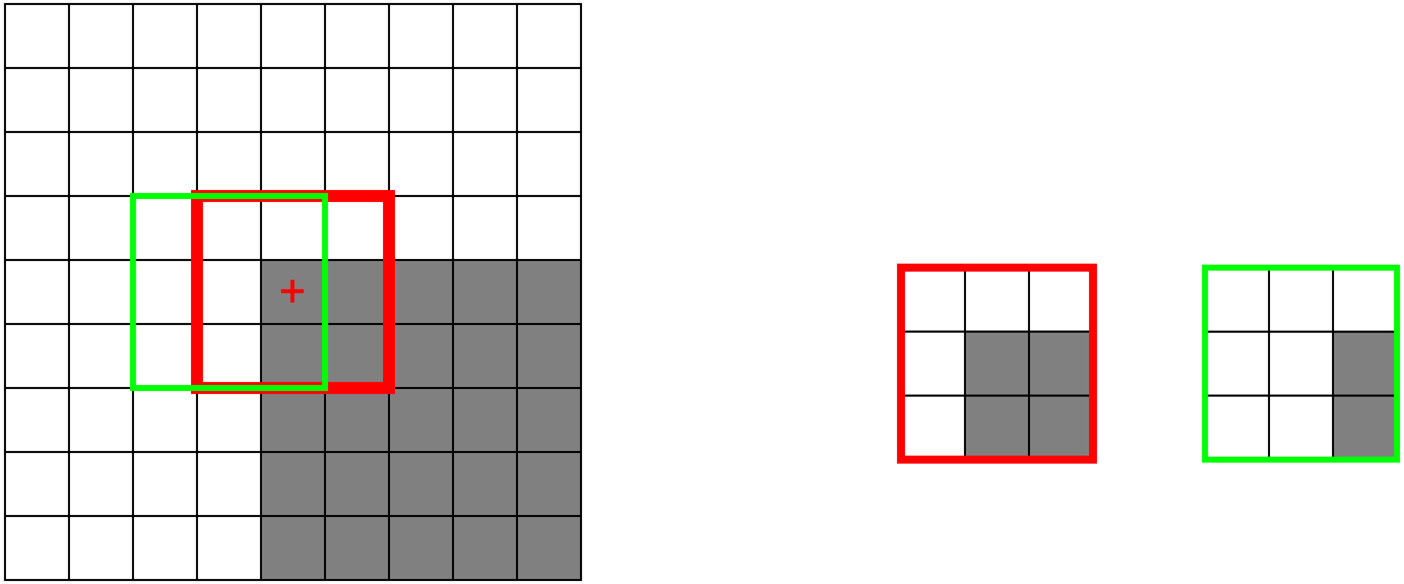
Demo of a point + with well distinguished neighbourhood.

Corner Detection: Introduction



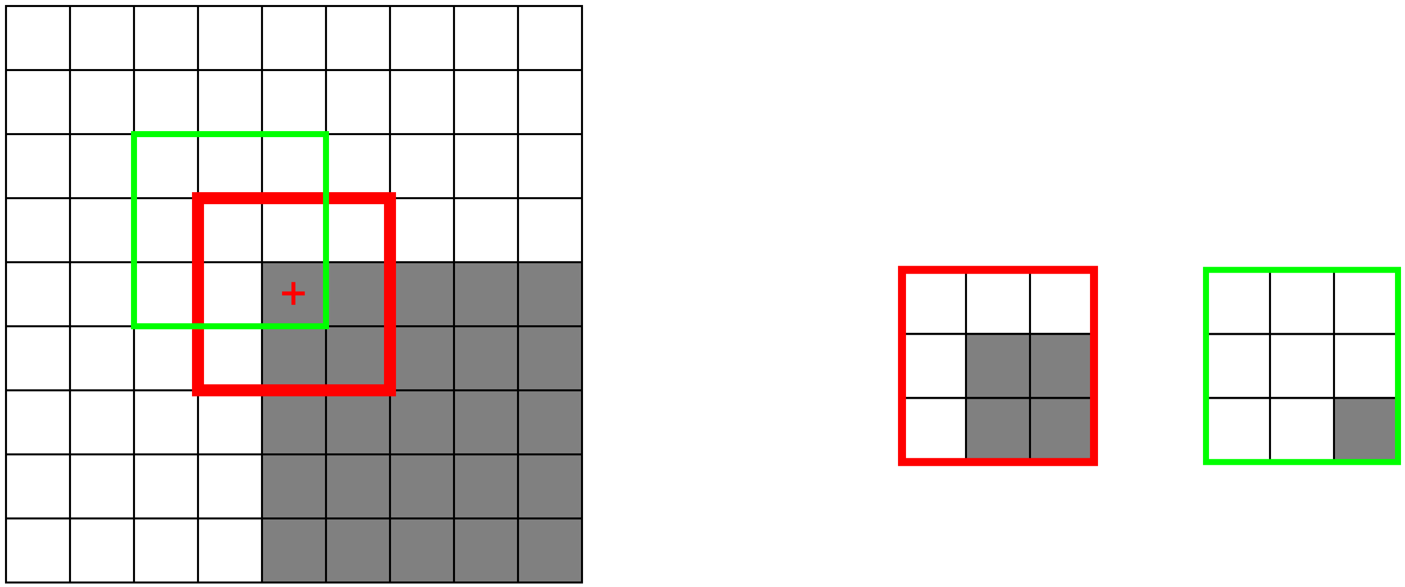
Demo of a point + with well distinguished neighbourhood.

Corner Detection: Introduction



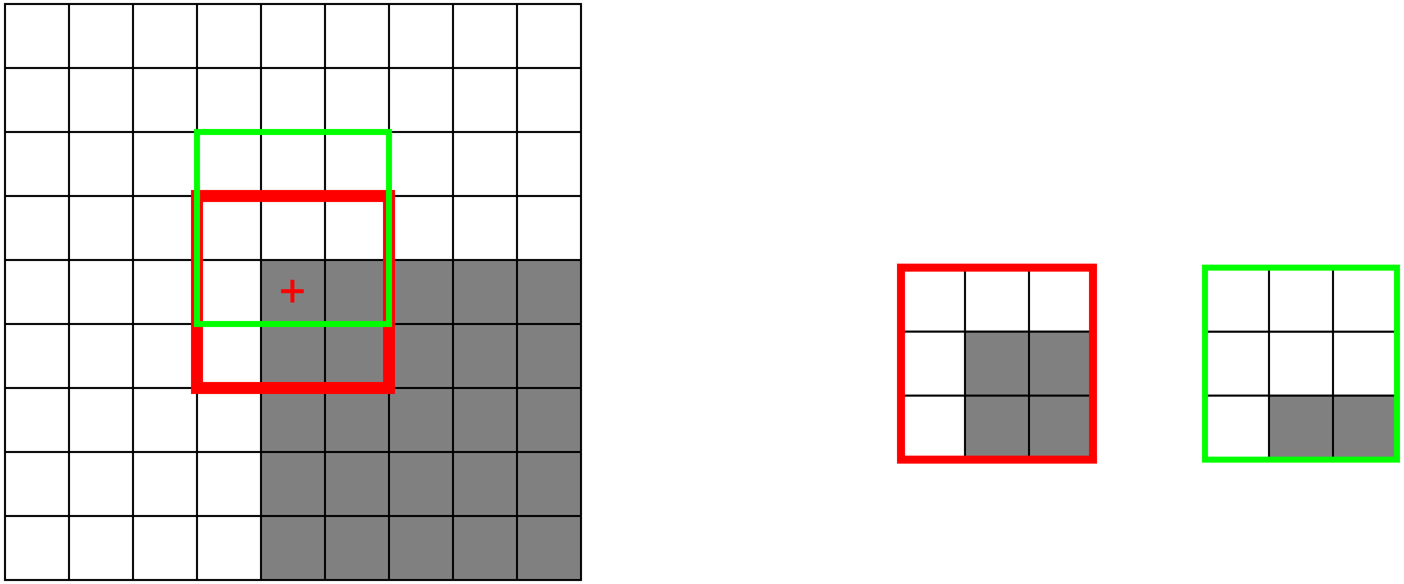
Demo of a point + with well distinguished neighbourhood.

Corner Detection: Introduction



Demo of a point + with well distinguished neighbourhood.

Corner Detection: Introduction



Demo of a point + with well distinguished neighbourhood.

Harris detection

- Auto-correlation matrix
  - captures the structure of the local neighborhood
  - measure based on eigenvalues of this matrix
    - 2 strong eigenvalues => interest point
    - 1 strong eigenvalue => contour
    - 0 eigenvalue => uniform region

- Interest point detection
  - threshold on the eigenvalues
  - local maximum for localization

Harris detector

Auto-correlation function for a point (x,y) and a shift $(\Delta x,\Delta y)$

$$f(x,y)=\sum_{(x_k,y_k)\in W}(I(x_k,y_k)-I(x_k+\Delta x,y_k+\Delta y))^2$$

Discrete shifts can be avoided with the auto-correlation matrix

$$\text{with}\quad I(x_k+\Delta x,y_k+\Delta y)=I(x_k,y_k)+(I_x(x_k,y_k)\quad I_y(x_k,y_k))\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}$$

$$f(x,y)=\sum_{(x_k,y_k)\in W}\left(\big(I_x(x_k,y_k)\quad I_y(x_k,y_k)\big)\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}\right)^2$$

Harris detector

$$= \begin{pmatrix}\Delta x & \Delta y\end{pmatrix} \begin{bmatrix} \sum\limits_{(x_k,y_k)\in W} (I_x(x_k,y_k))^2 & \sum\limits_{(x_k,y_k)\in W} I_x(x_k,y_k) I_y(x_k,y_k) \\ \sum\limits_{(x_k,y_k)\in W} I_x(x_k,y_k) I_y(x_k,y_k) & \sum\limits_{(x_k,y_k)\in W} (I_y(x_k,y_k))^2 \end{bmatrix} \begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix}$$

Auto-correlation matrix

Harris Detector: Mathematics

Window-averaged change of intensity for the shift [*u*,*v*]:

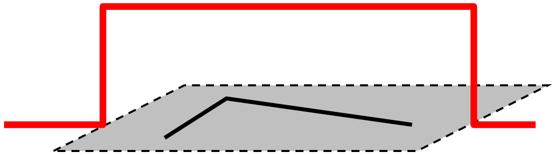
$$E(u,v)=\sum_{x,y} w(x,y)\left[I(x+u,y+v)-I(x,y)\right]^2$$

Window function

Shifted intensity

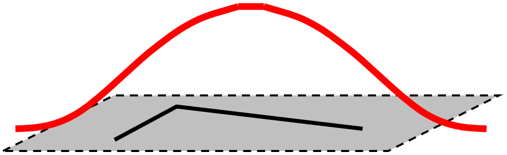
Intensity

Window function $w(x,y)$ =



1 in window, 0 outside

or



Gaussian

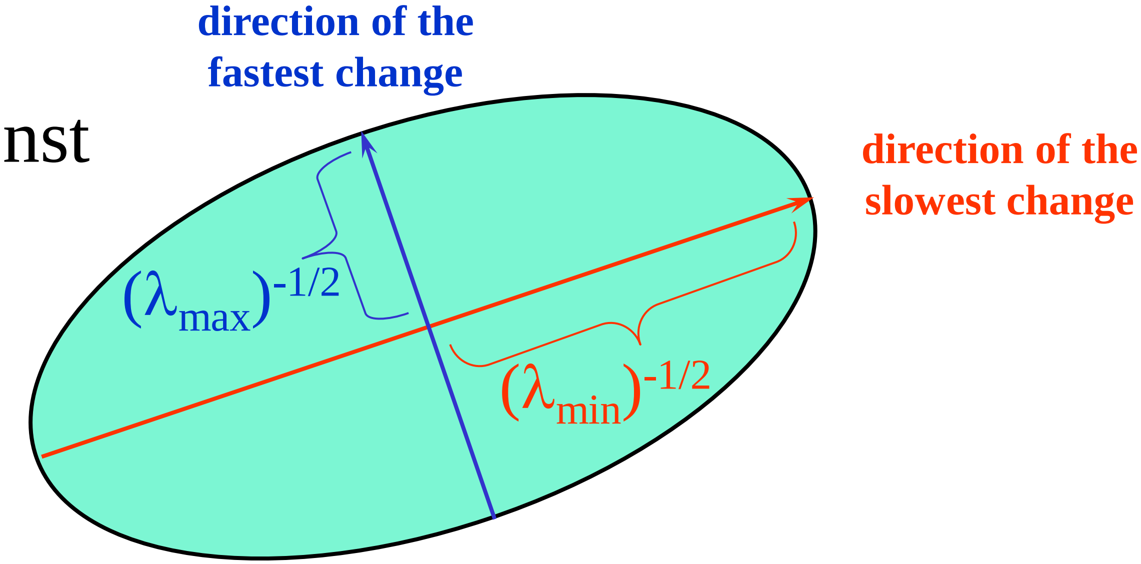
Harris Detector: Mathematics

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


direction of the fastest change
direction of the slowest change
$(\lambda_{\max})^{-1/2}$
$(\lambda_{\min})^{-1/2}$

Harris Detector: Mathematics

Expanding E(u,v) in a 2nd order Taylor series expansion, we have,for small shifts $[u,v]$, a *bilinear* approximation:

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

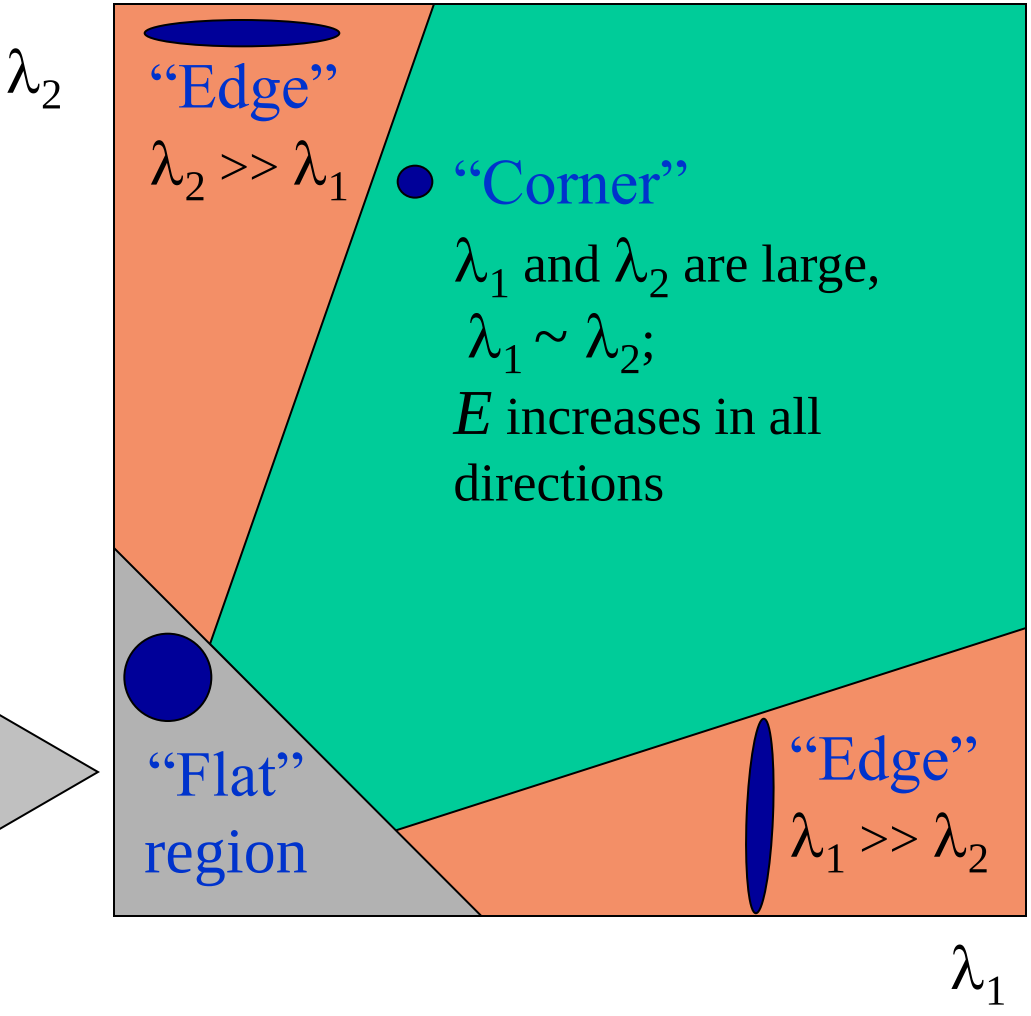
where M is a 2×2 matrix computed from image derivatives:

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

Harris Detector: Mathematics

Classification of image points using eigenvalues of *M*:

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



“Edge”

$\lambda_2 >> \lambda_1$

“Corner”

λ_1 and λ_2 are large, $\lambda_1 \sim \lambda_2$; *E* increases in all directions

“Flat” region

“Edge”

$\lambda_1 >> \lambda_2$

Harris Detector: Mathematics

Measure of corner response:

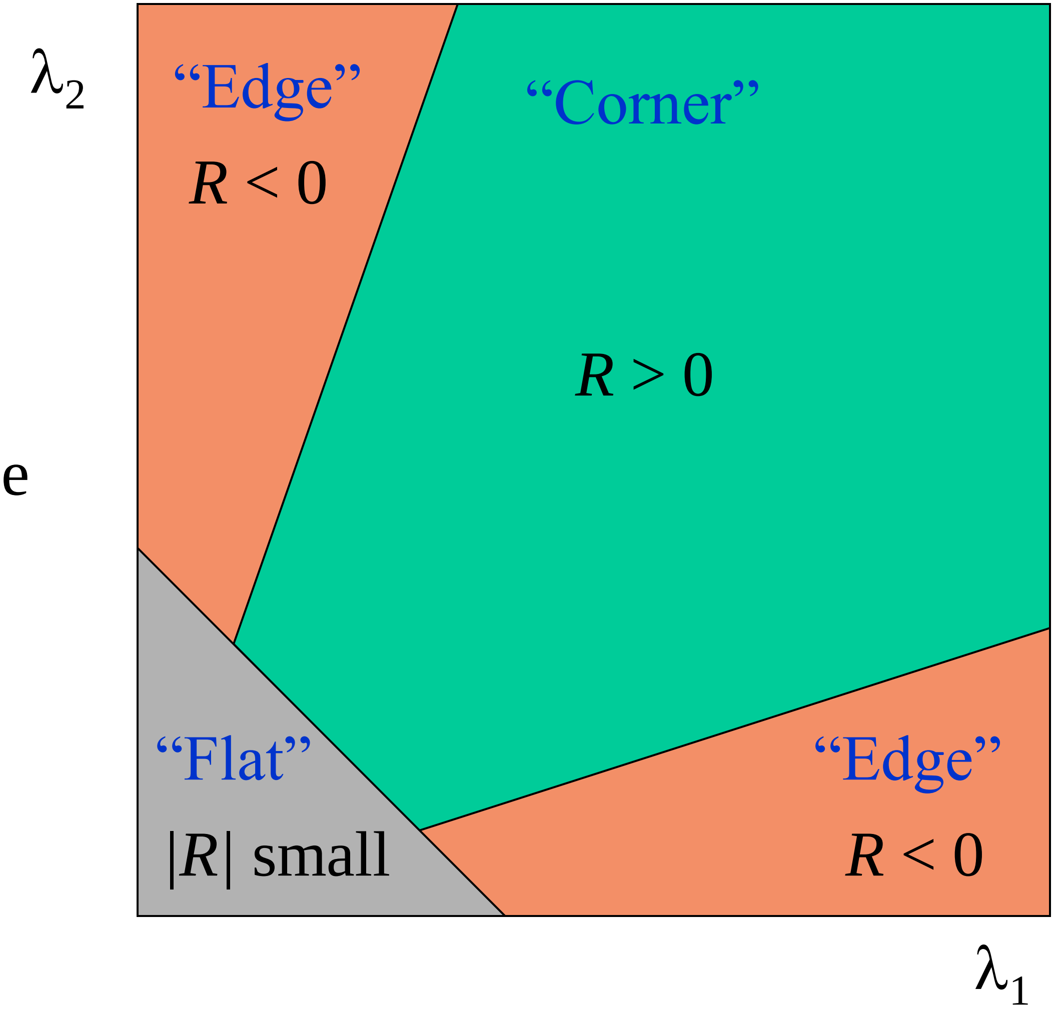
$$R = \det M - k\left(\operatorname{trace} M\right)^2$$

$$\begin{aligned}\det M &= \lambda_1 \lambda_2 \\ \operatorname{trace} M &= \lambda_1 + \lambda_2\end{aligned}$$

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

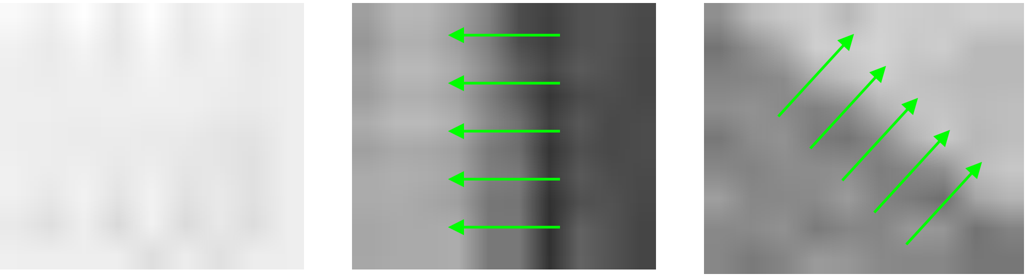
Harris Detector: Mathematics

- *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



Corner Detection: Basic principle

undistinguished patches:



distinguished patches:

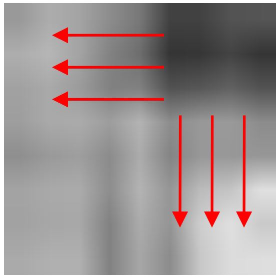


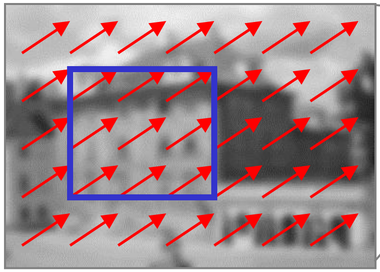
Image gradients $\nabla I(x,y)$ of undist. patches are (0,0) or have only one principle component.

Image gradients $\nabla I(x,y)$ of dist. patches have two principle components.

$$\Rightarrow \quad \mathrm{rank}\left(\sum \nabla I(x,y) * \nabla I(x,y)^{\top} \right) = 2$$

Algorithm (R. Harris, 1988)

1. filter the image by gaussian (2x 1D convolution), sigma_d



2. compute the intensity gradients $\nabla I(x,y)$, (2x 1D conv.)

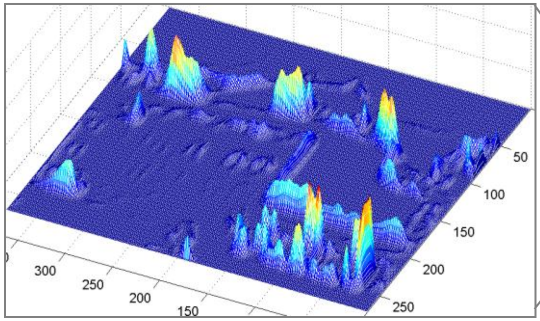
3. for each pixel and given neighbourhood, sigma_i:

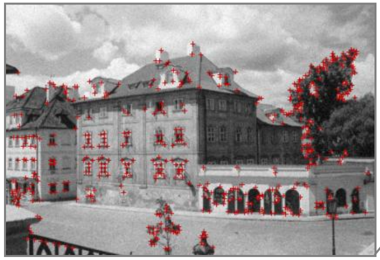
- compute auto-correlation matrix

$$\mathbf{A} = \sum \nabla I(x,y) * \nabla I(x,y)^{T}$$

- and evaluate the response function $R(\mathbf{A})$:

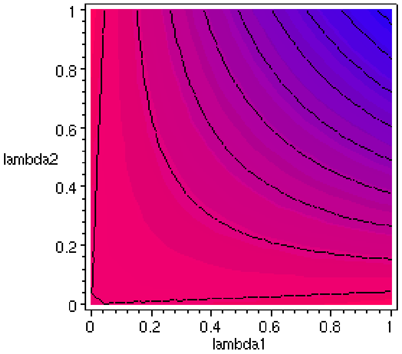
$R(\mathbf{A}) \gg 0$ for rank($\mathbf{A}$)=2, $R(\mathbf{A}) \rightarrow 0$ for rank($\mathbf{A}$)<2

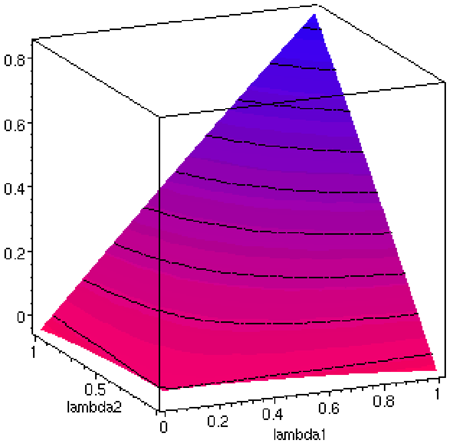




4. choose the best candidates (non-max suppression and thresholding)

Corner Detection: Algorithm (R. Harris, 1988)





Harris response function $R(\mathbf{A})$:

$$R(\mathbf{A}) = \det(\mathbf{A}) - k*\mathrm{trace}^{2}(\mathbf{A}) \ ,$$

[lamda1,lambda2] = eig(**A**)

Corner Detection: Algorithm (R. Harris, 1988)

Algorithm properties:

+ “invariant” to 2D image shift and rotation

+ invariant to shift in illumination

+ “invariant” to small view point changes

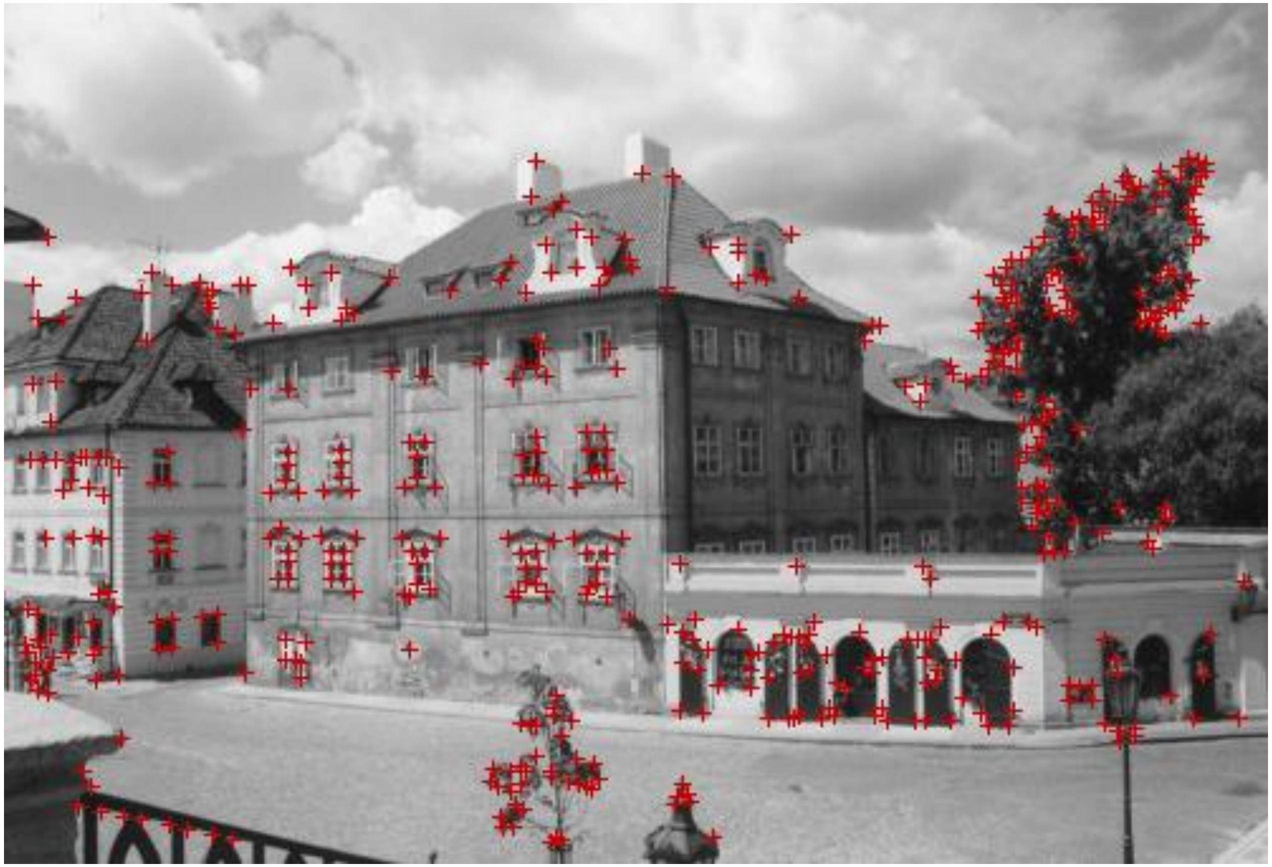
+ low numerical complexity

- not invariant to larger scale changes

- not completely invariant to high contrast changes

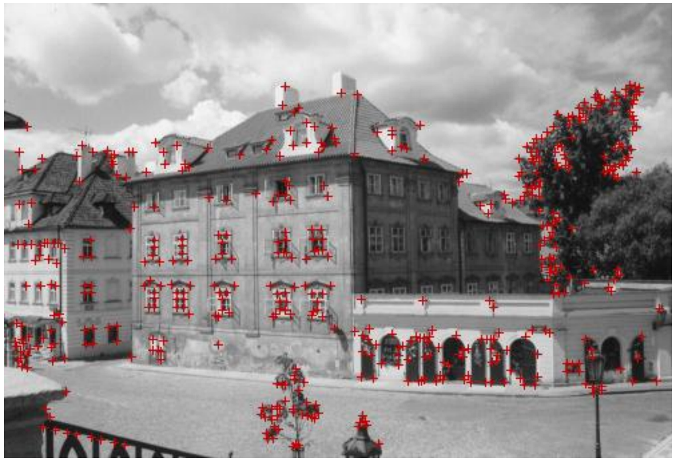
- not invariant to bigger view point changes

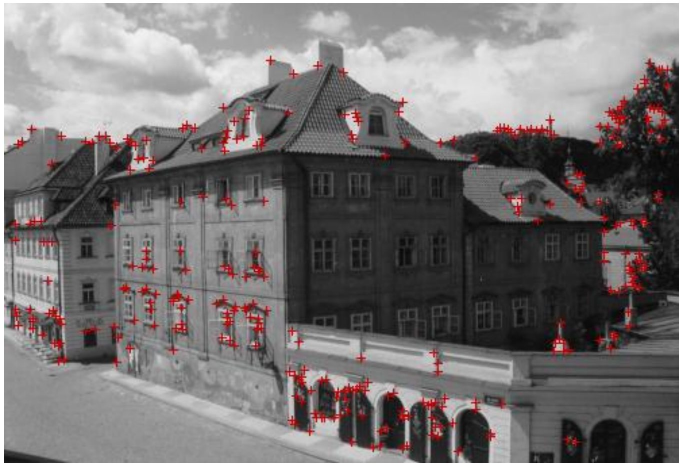
Corner Detection: Introduction



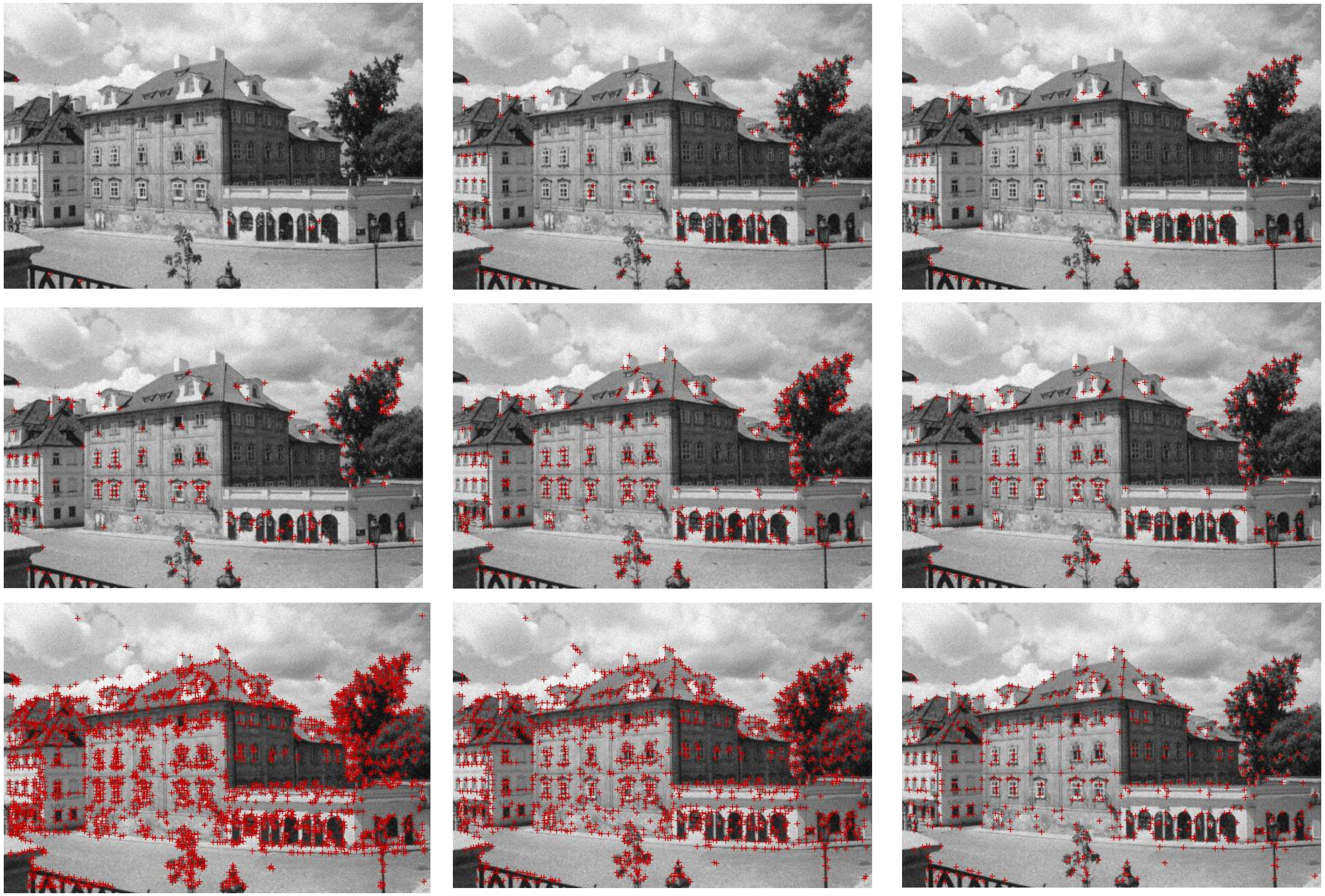
Example of detected points

Corner Detection: Algorithm (R. Harris, 1988)

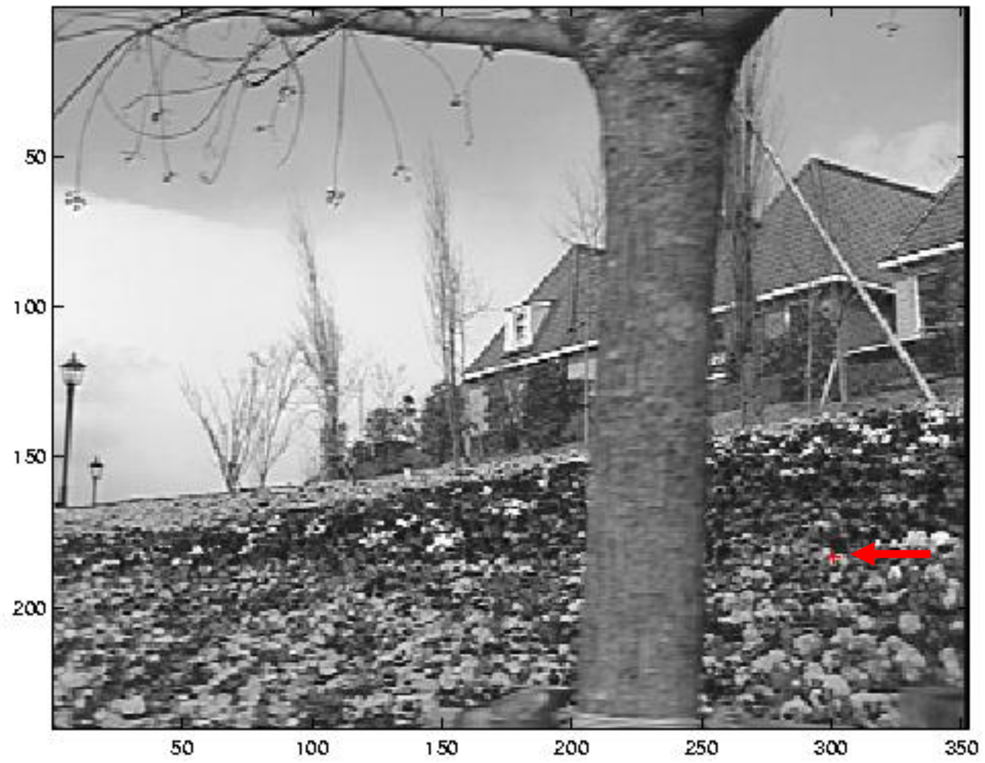


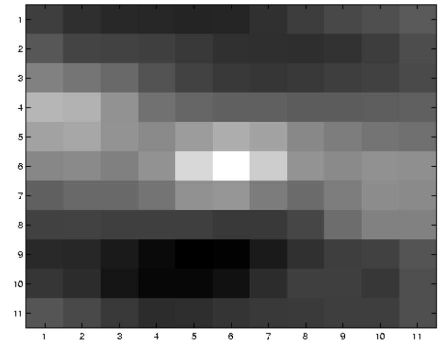


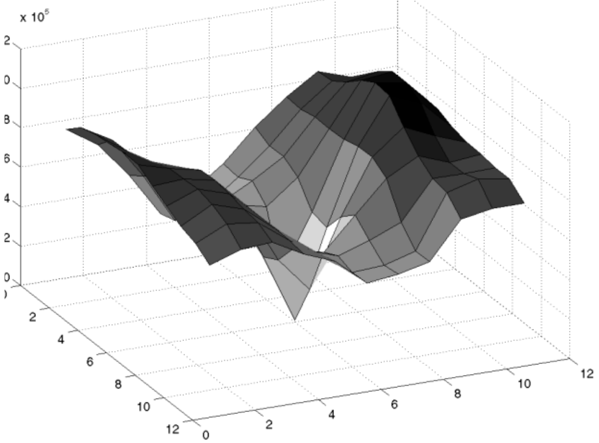
Corner Detection: Harris points versus sigma_d and sigma_i



Selecting Good Features

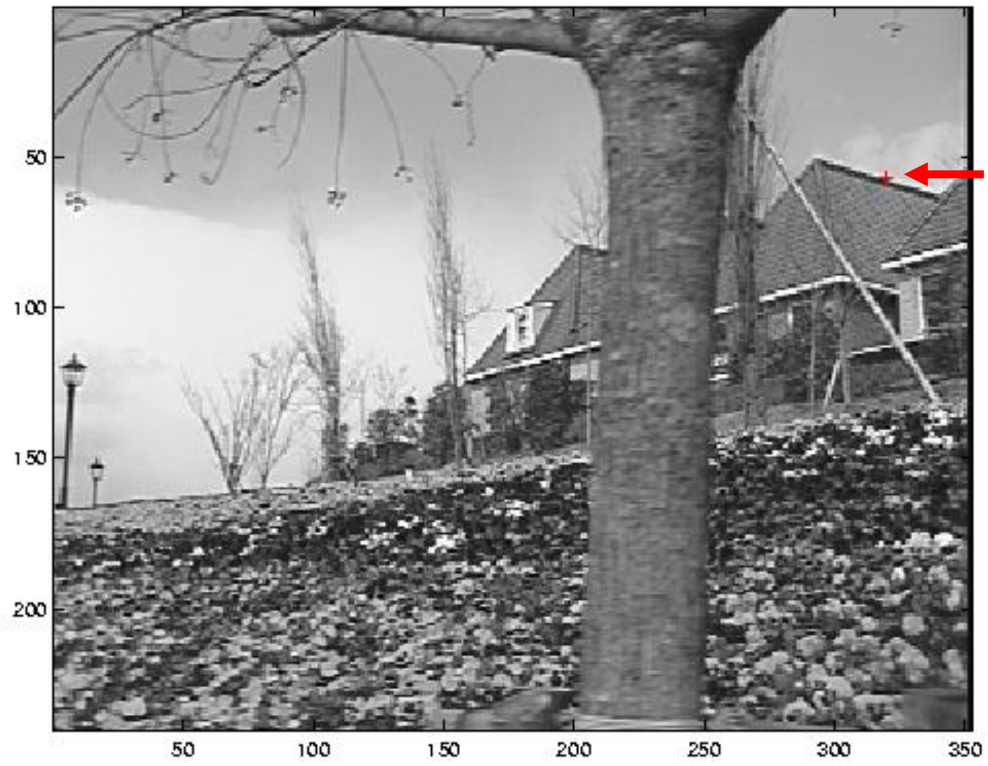


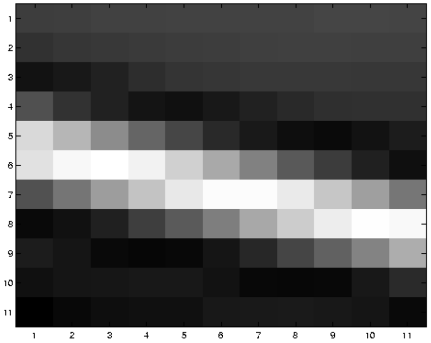


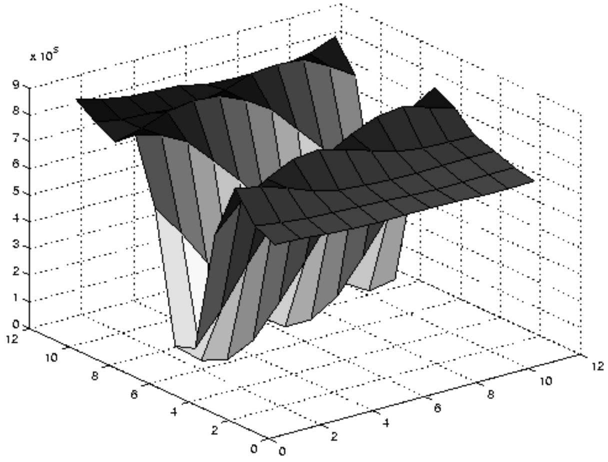


λ_1 and λ_2 are large

Selecting Good Features

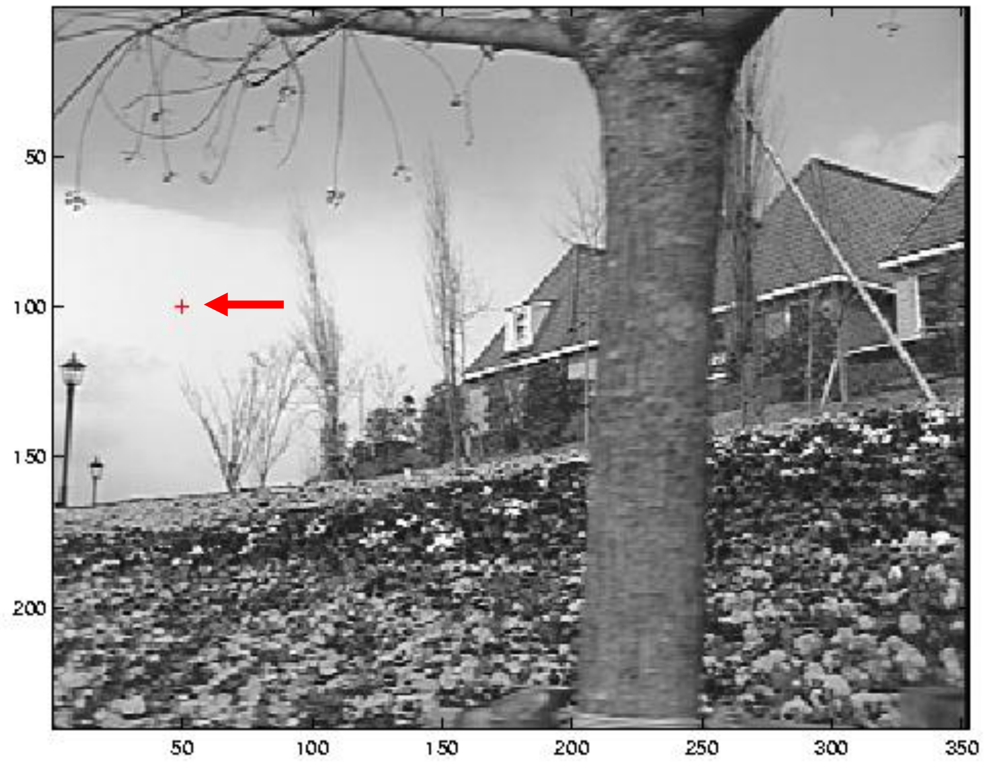


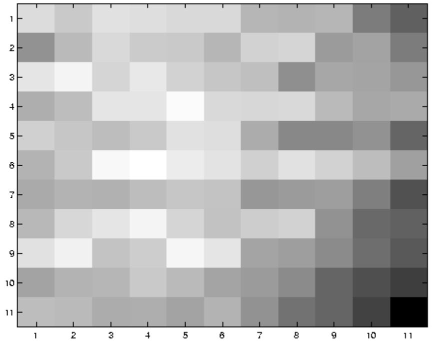


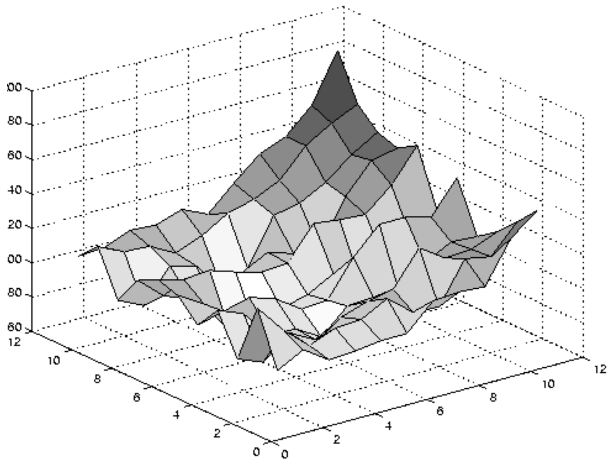


large λ_1, small λ_2

Selecting Good Features





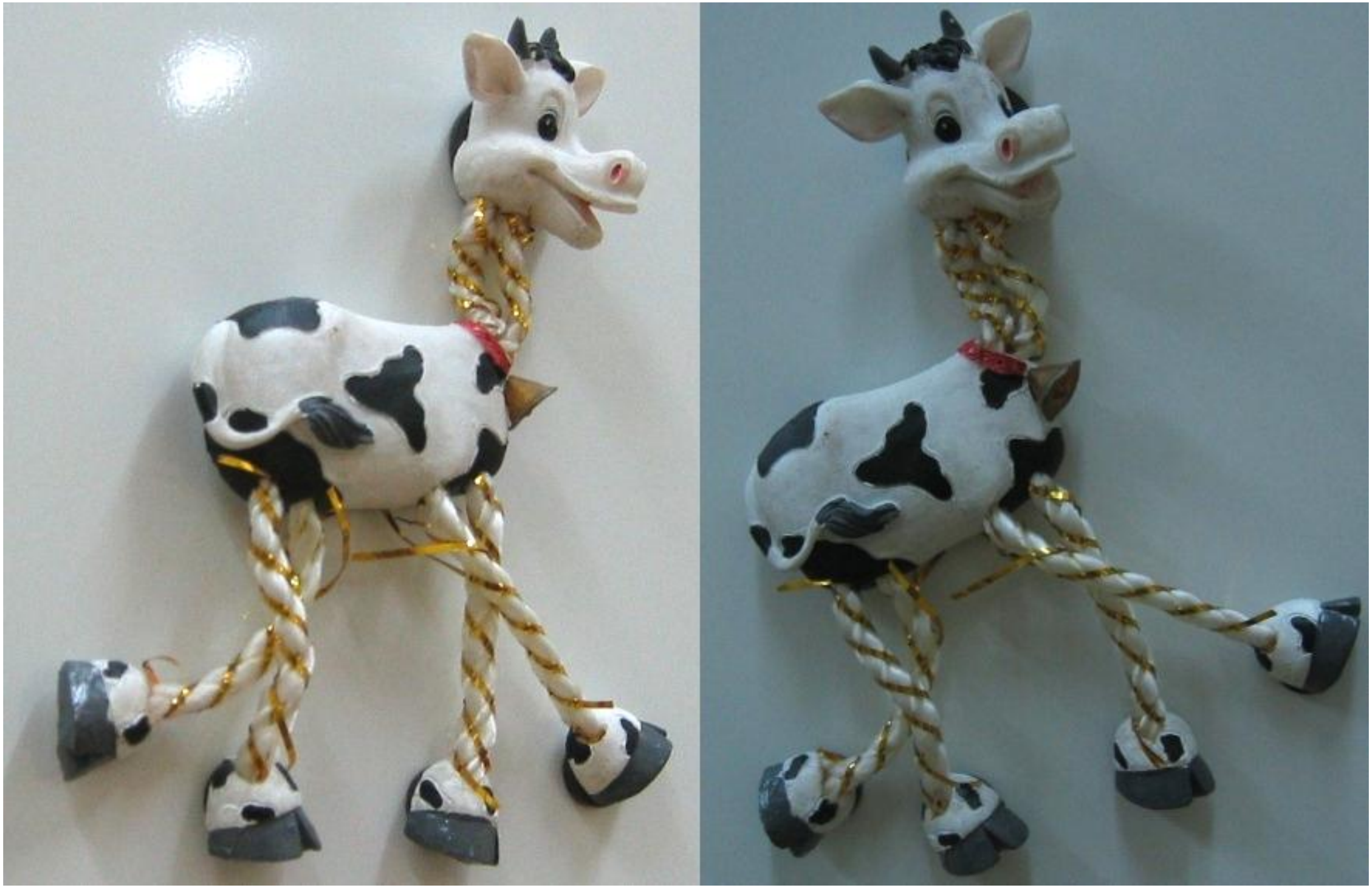


small λ_1, small λ_2

Harris Detector

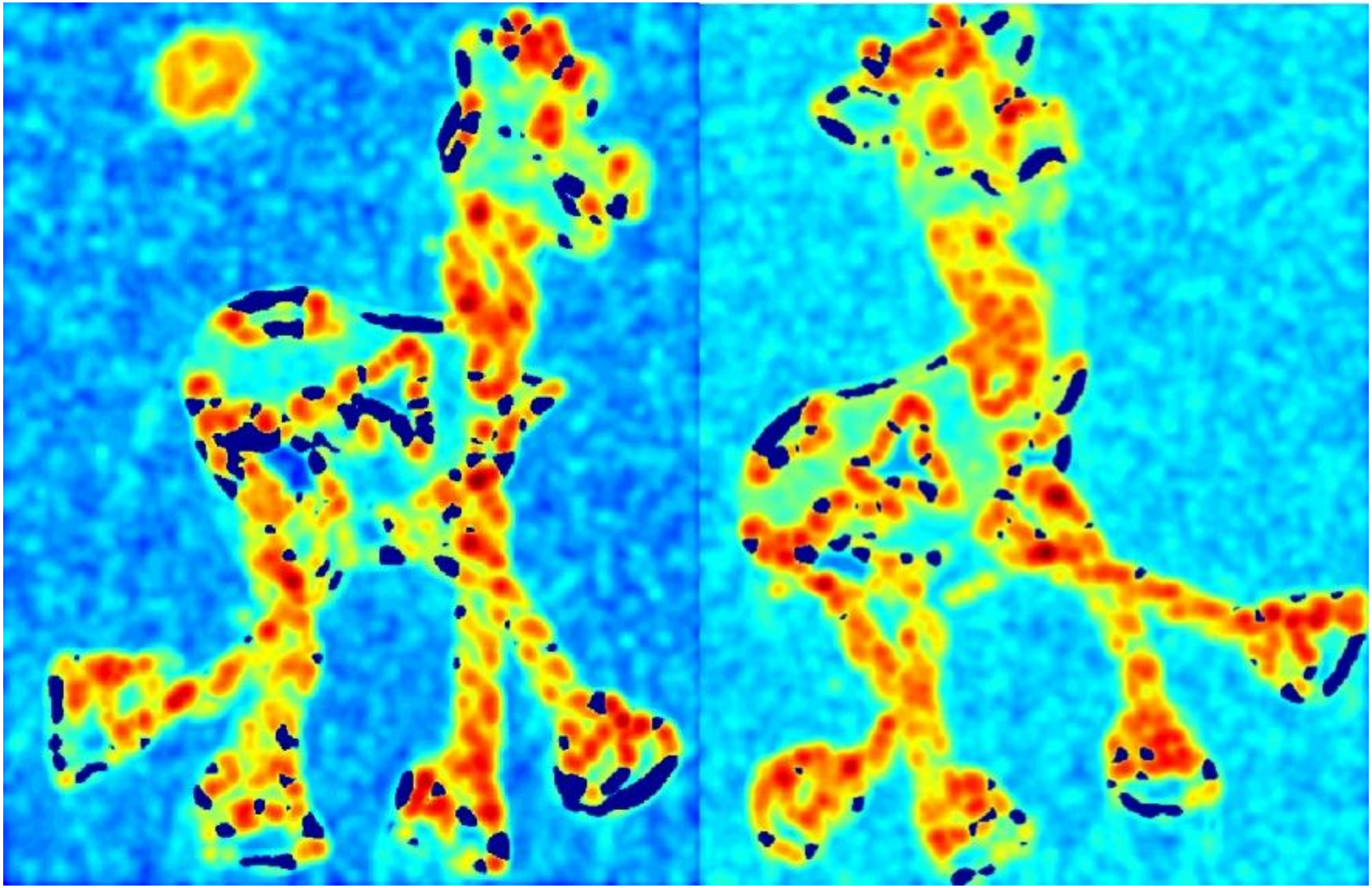
- The Algorithm:
  - Find points with large corner response function *R* (*R* > threshold)
  - Take the points of local maxima of *R*

Harris Detector: Workflow



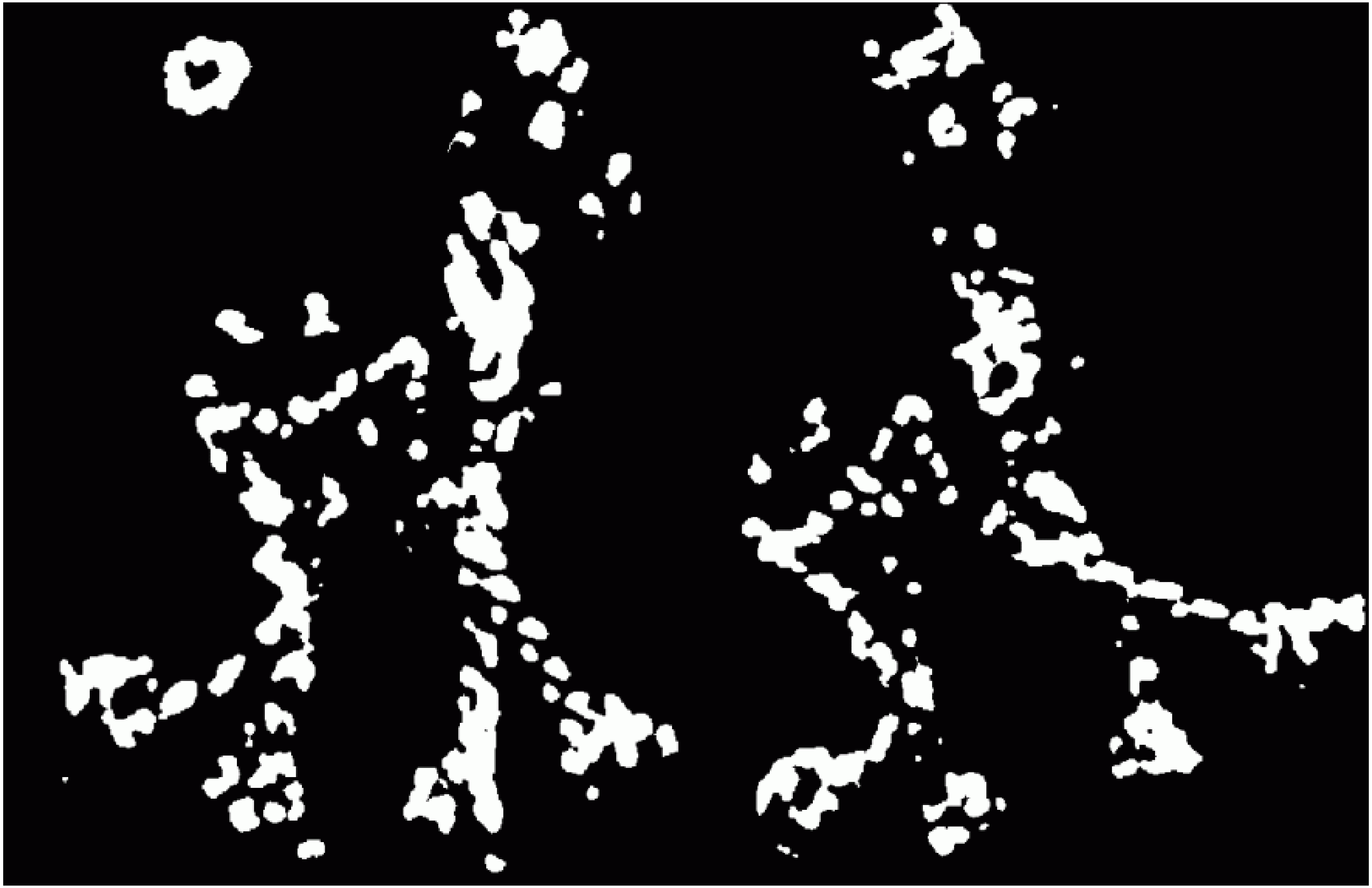
Harris Detector: Workflow

Compute corner response R



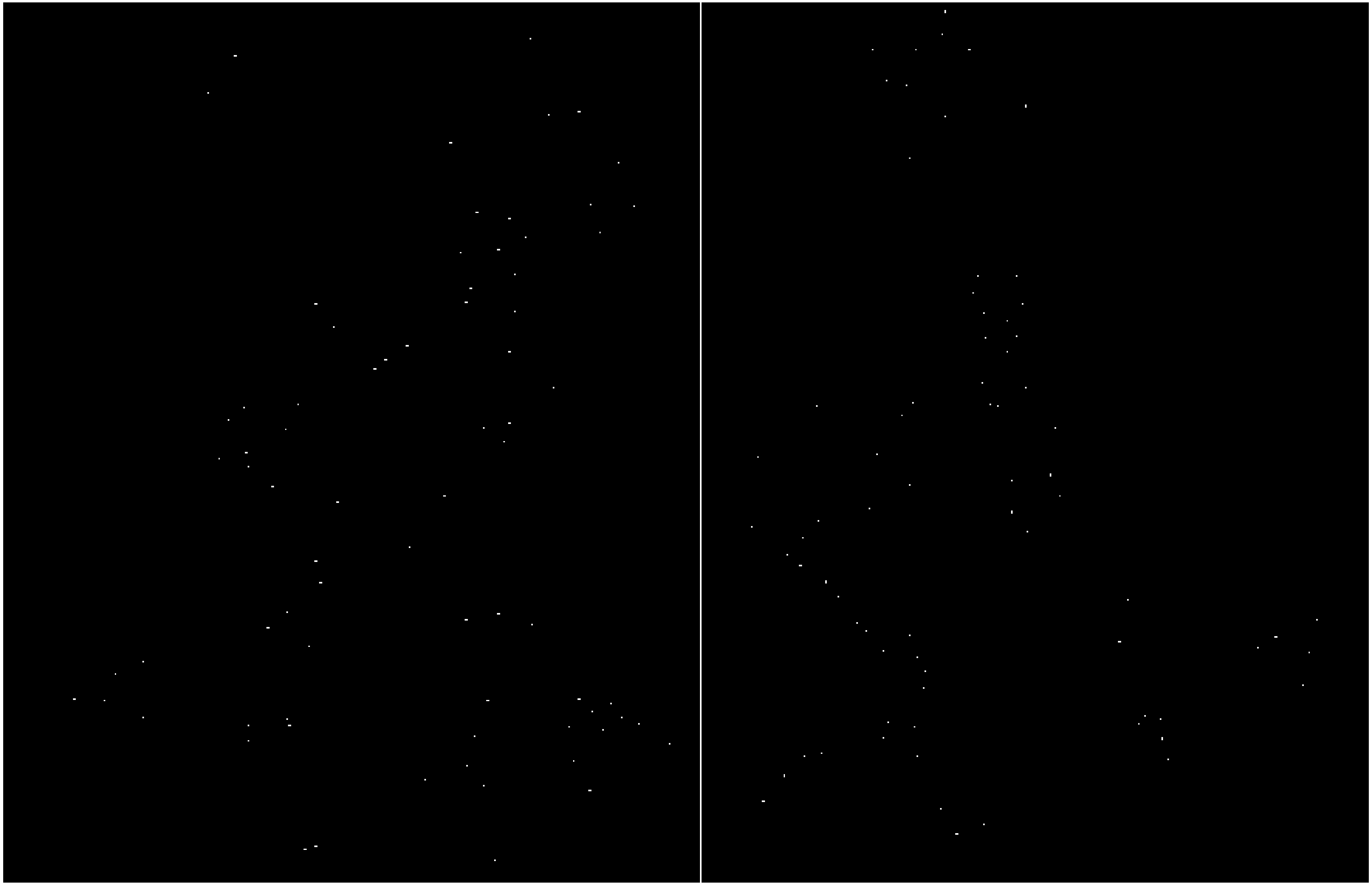
Harris Detector: Workflow

Find points with large corner response: R>threshold

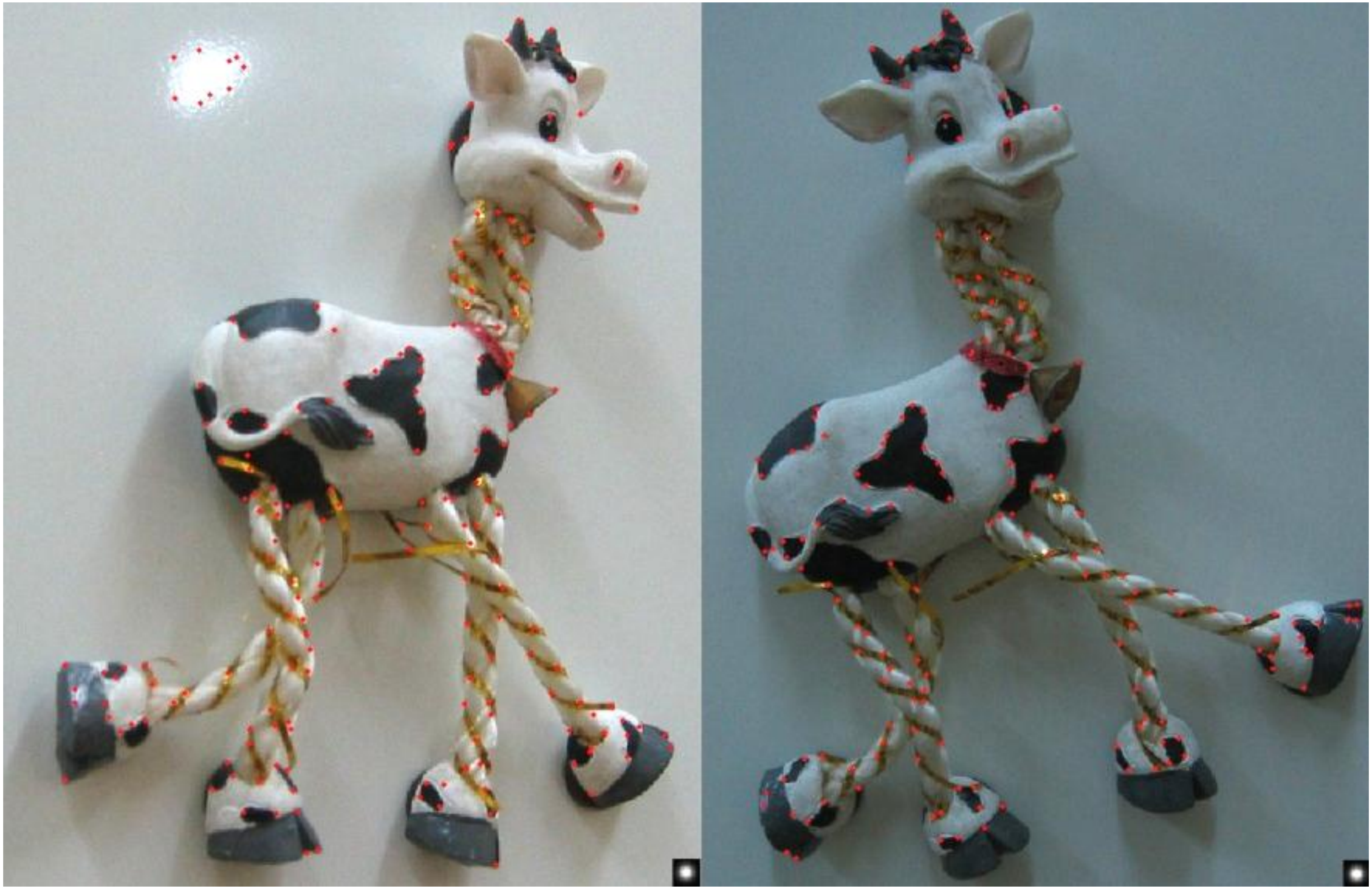


Harris Detector: Workflow

Take only the points of local maxima of R



Harris Detector: Workflow



Harris Detector: 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

Corner Detection: Application

Algorithm:

1. Corner detection

2. Tentative correspondences
   - by comparing similarity of the corner neighb. in the searching window (e.g. cross-correlation)

3. Camera motion geometry estimation (e.g. by RANSAC)
   - finds the motion geometry and consistent correspondences

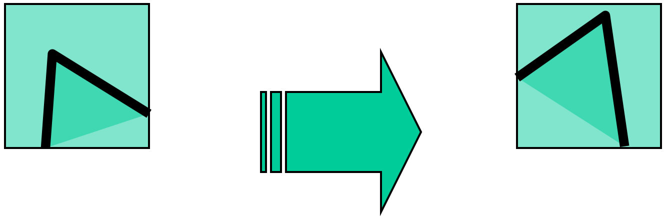
4. 3D reconstruction
   - triangulation, bundle adjustment

Contents

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

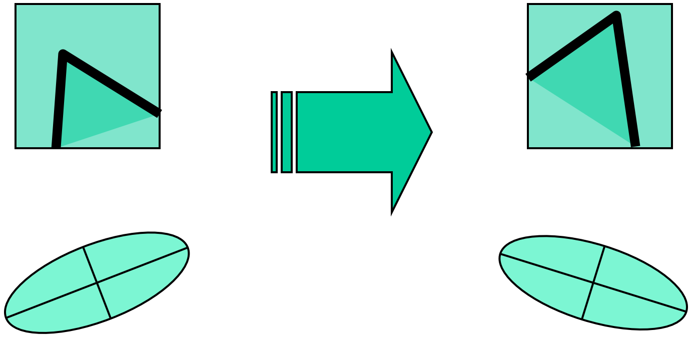
Harris Detector: Some Properties

- Rotation invariance?



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

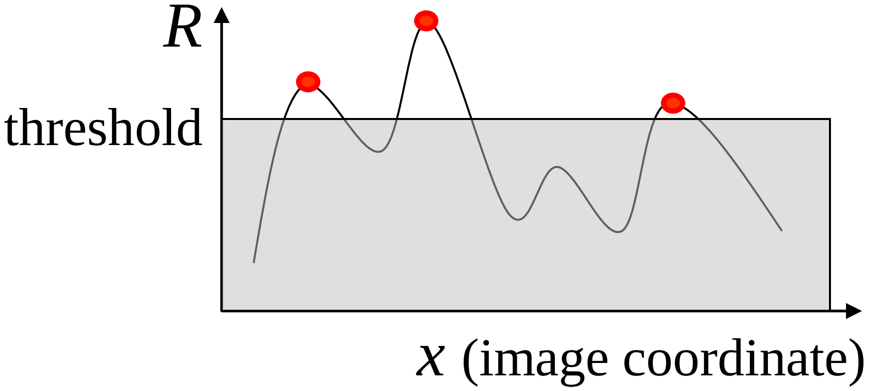
- Invariance to image intensity change?

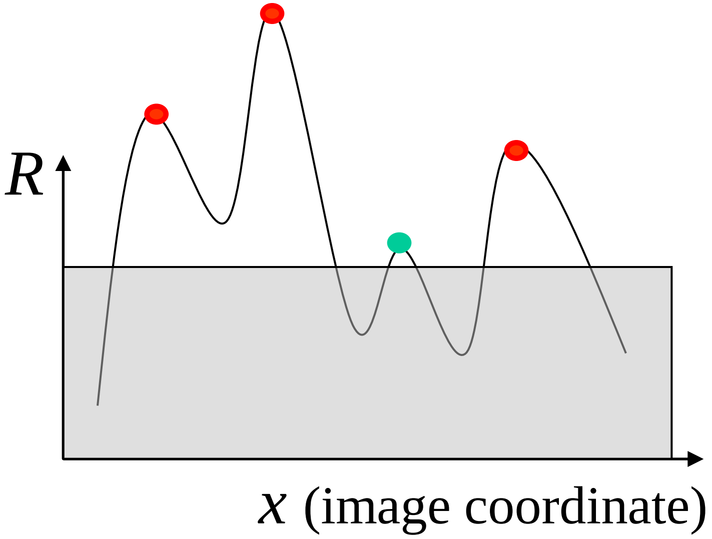
Harris Detector: Some Properties

- Partial invariance to additive and multiplicative intensity changes

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

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



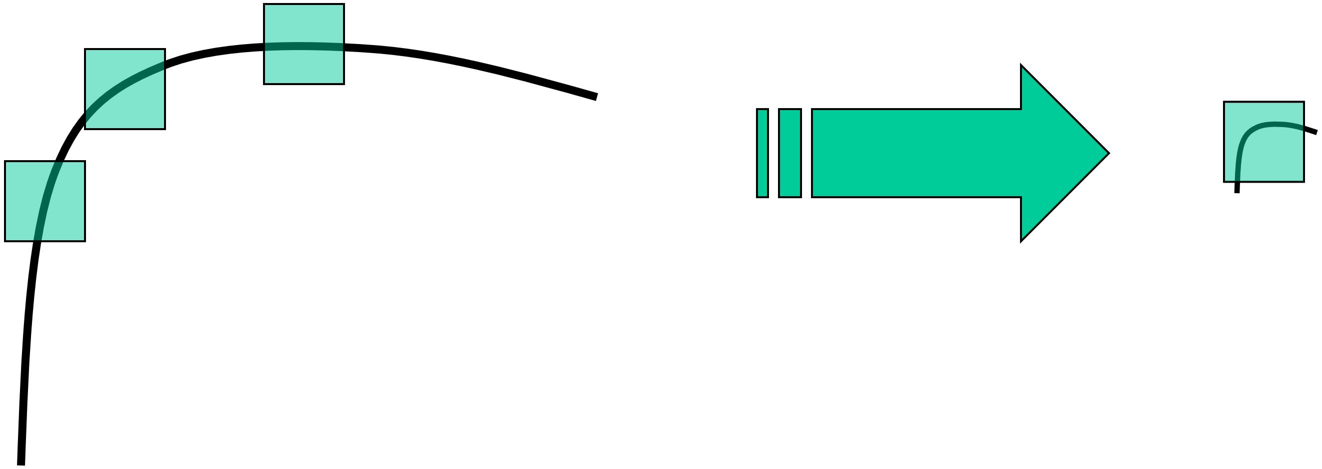


Harris Detector: Some Properties

- Invariant to image scale?

Harris Detector: Some Properties

- Not invariant to *image scale*!



All points will be
classified as edges

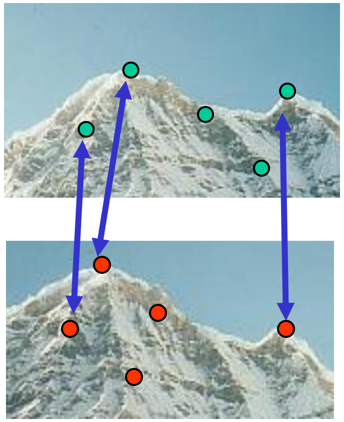
Corner !

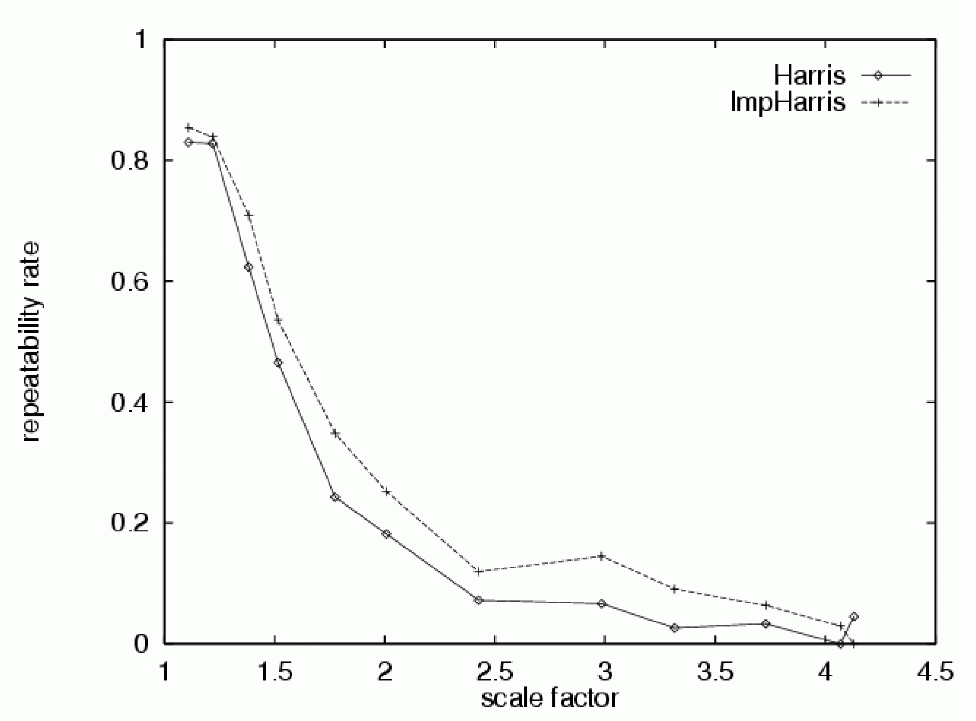
Harris Detector: Some Properties

- 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

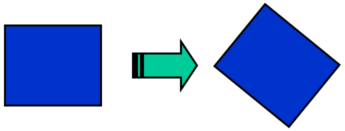
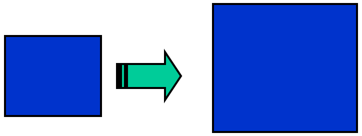
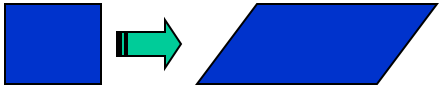
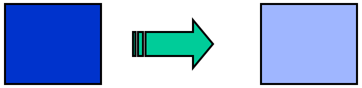
Contents

- 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

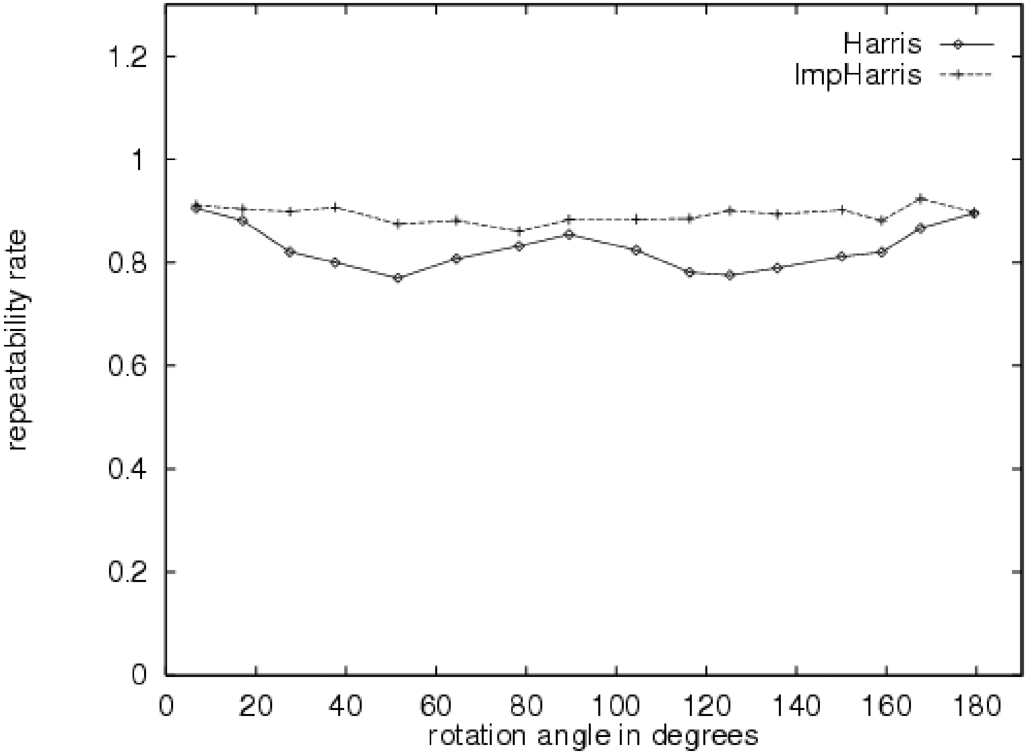
- Geometry
  - Rotation 
  - Similarity (rotation + uniform scale) 
  - Affine (scale dependent on direction)  valid for: orthographic camera, locally planar object
- Photometry
  - Affine intensity change ($I \rightarrow a\, I + b$) 

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



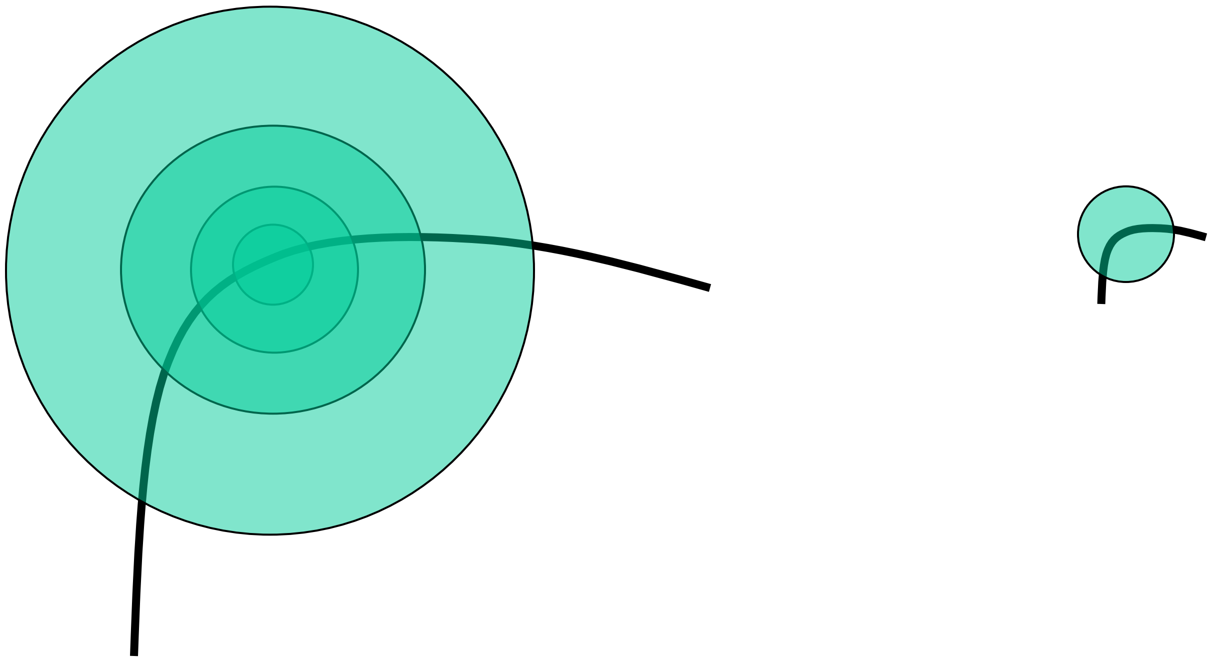
C.Schmid et.al. “Evaluation of Interest Point Detectors”. IJCV 2000

Contents

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

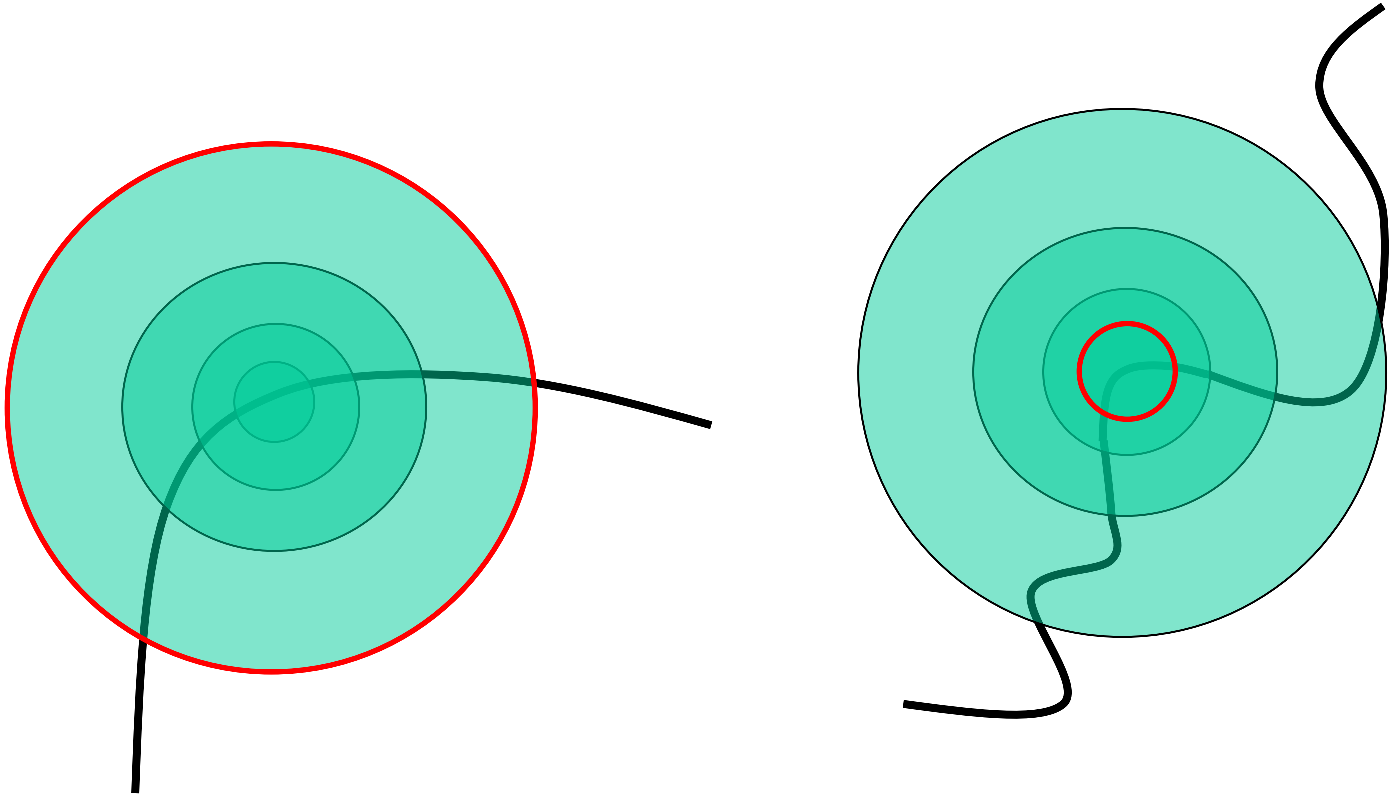
Scale Invariant Detection

- 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?

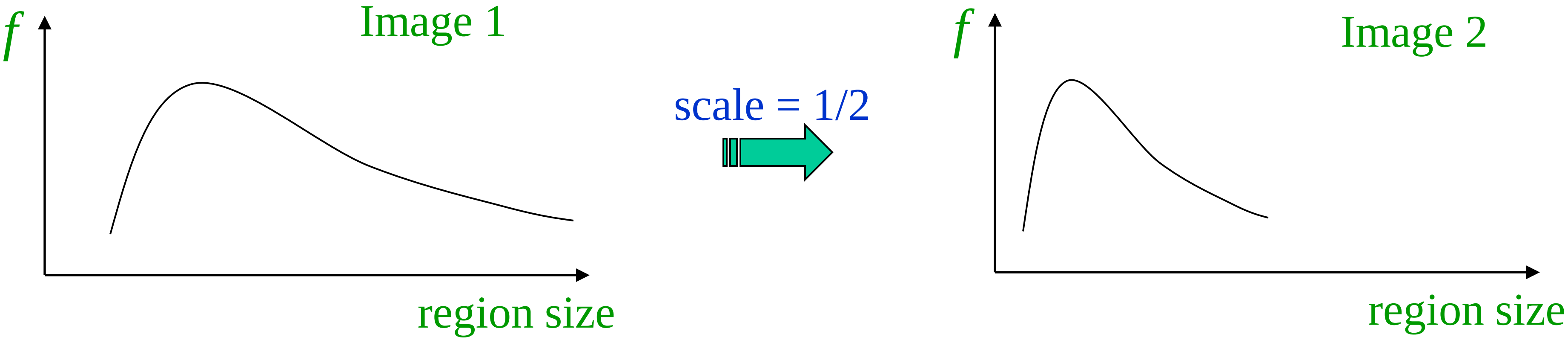


Scale Invariant Detection

- 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)



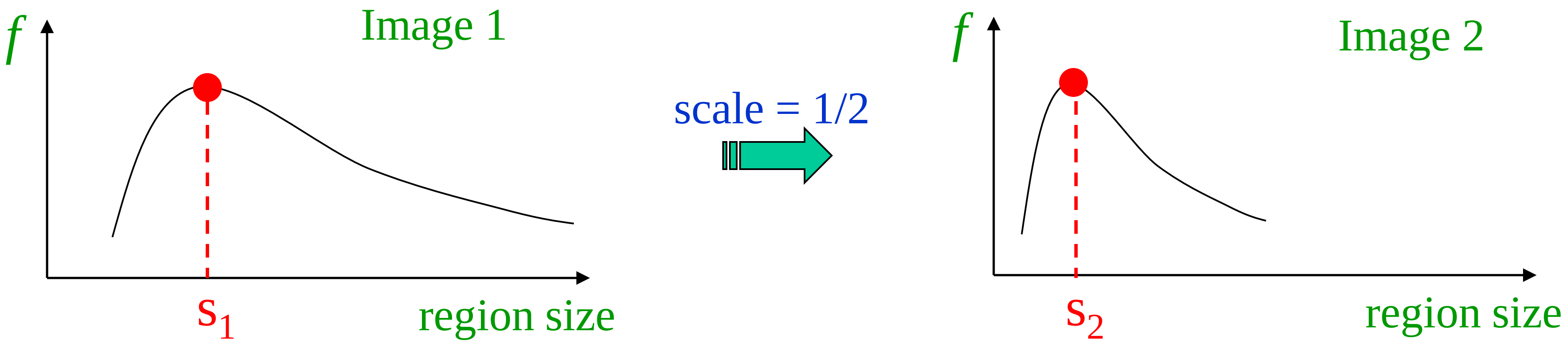
Scale Invariant Detection

- Common approach:

Take a local maximum of this function

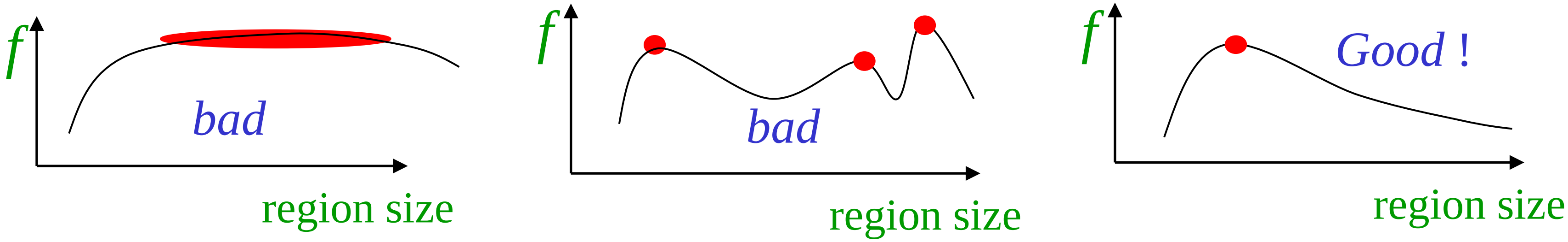
Observation: region size, for which the maximum is achieved, should be *invariant* to image scale.

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:

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

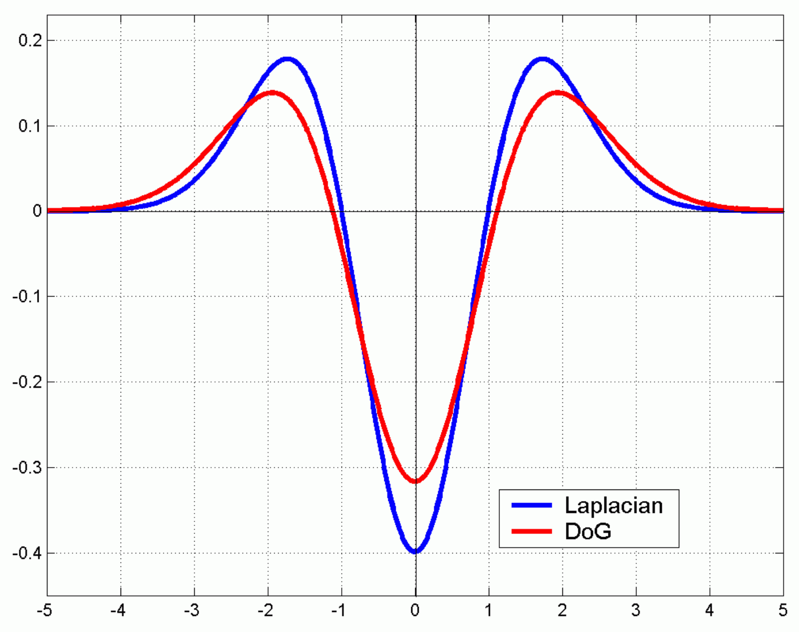
(Laplacian)

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

(Difference of Gaussians)

where Gaussian

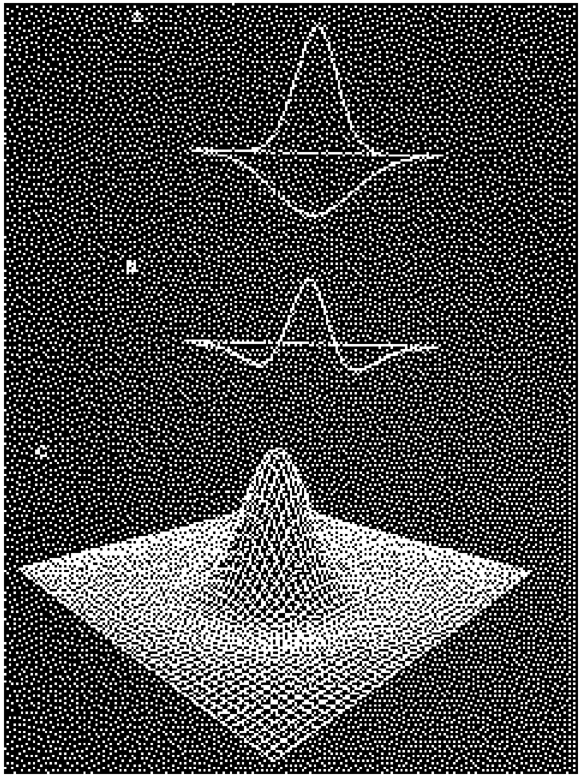
$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



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

Scale Invariant Detection

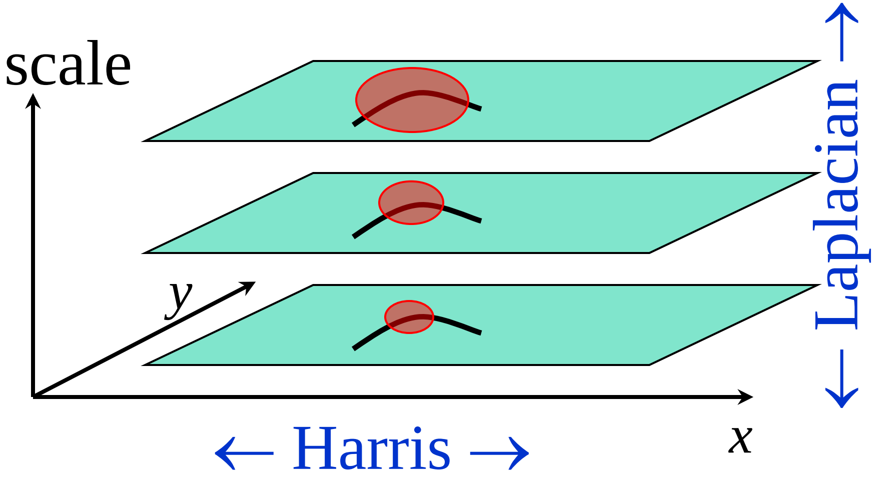
- Compare to human vision: eye’s response



Scale Invariant Detectors

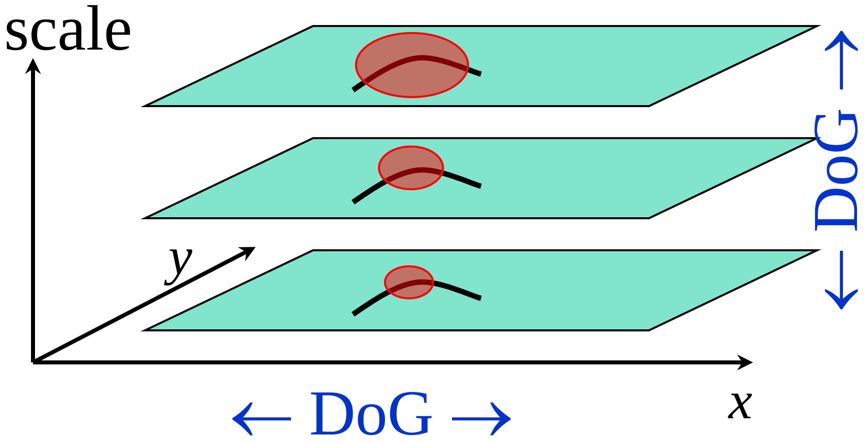
- Harris-Laplacian[1]

  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

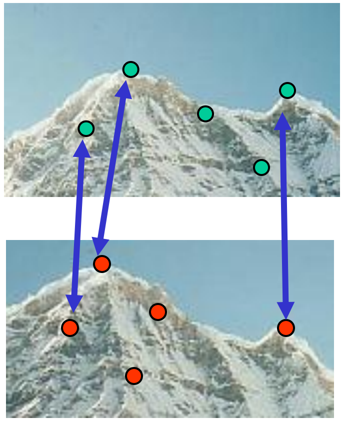
[2] D.Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. Accepted to IJCV 2004

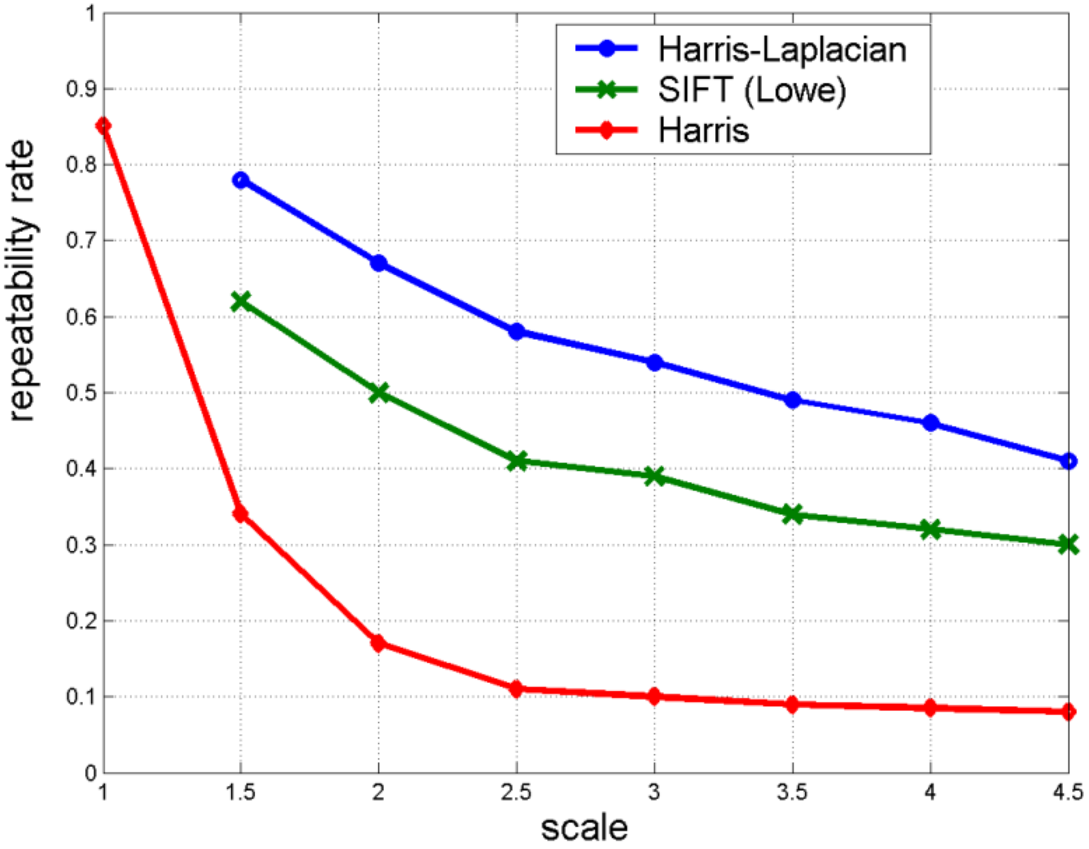
Scale Invariant Detectors

- 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:

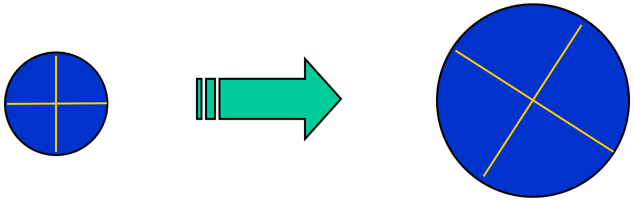
1. Harris-Laplacian [Mikolajczyk, Schmid]: maximize Laplacian over scale, Harris' measure of corner response over the image
2. SIFT [Lowe]: maximize Difference of Gaussians over scale and space

Contents

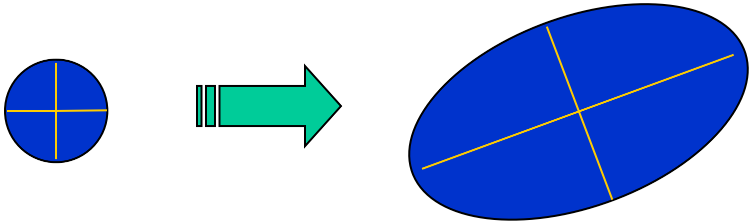
- Harris Corner Detector
  - Description
  - Analysis
- Detectors
  - Rotation invariant
  - Scale invariant
  - Affine invariant
- Descriptors
  - Rotation invariant
  - Scale invariant
  - Affine invariant

Affine Invariant Detection

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

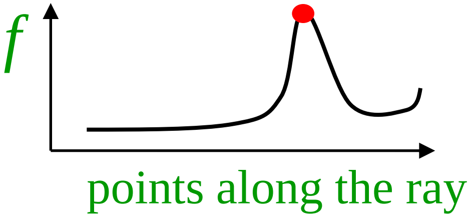


- Now we go on to:
  Affine transform (rotation + non-uniform scale)



Affine Invariant Detection

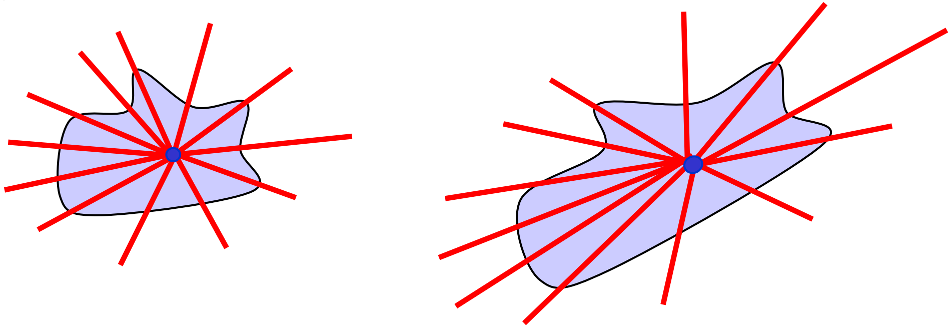
- Take a local intensity extremum as initial point
- Go along every ray starting from this point and stop when extremum of function f is reached



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

- 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.

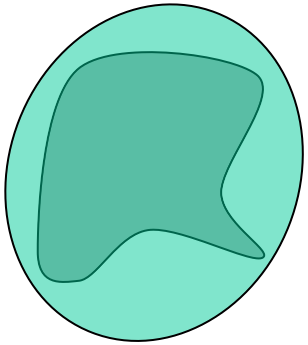
Affine Invariant Detection

- 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

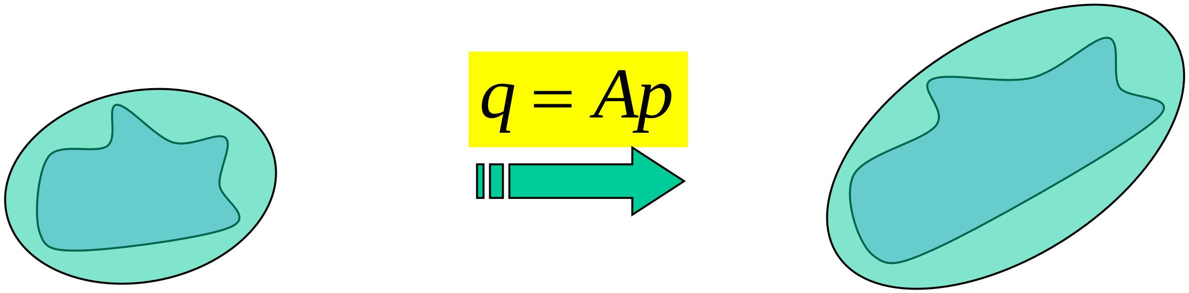
Taking f to be the characteristic function of a region (1 inside, 0 outside), moments of orders up to 2 allow to approximate the region by an ellipse



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:



$$q = Ap$$

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

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

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

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

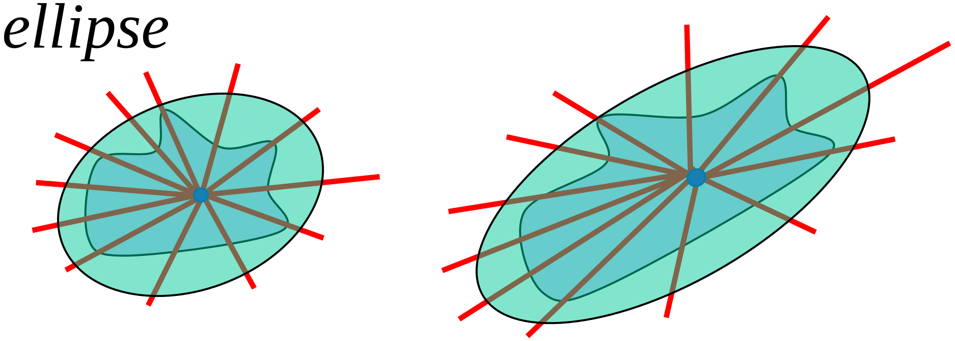
$$\Sigma_2 = \left\langle qq^T \right\rangle_{\text{region 2}}$$

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

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.]

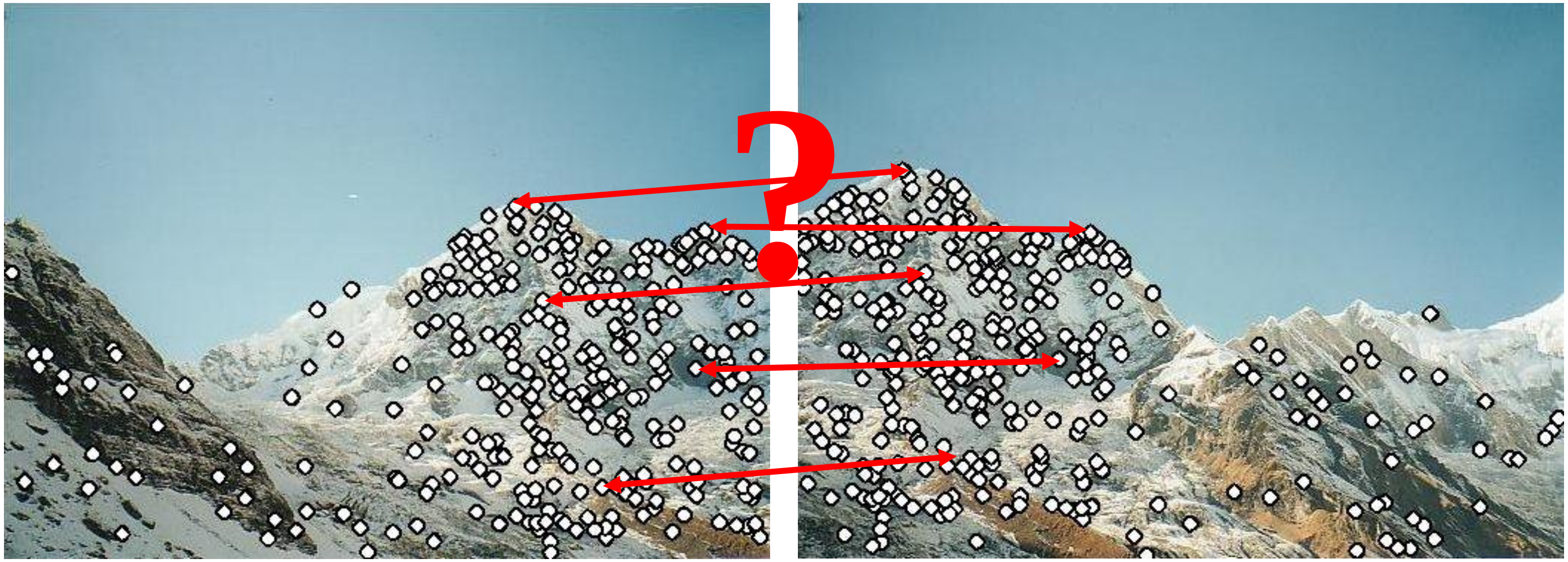
Contents

- 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

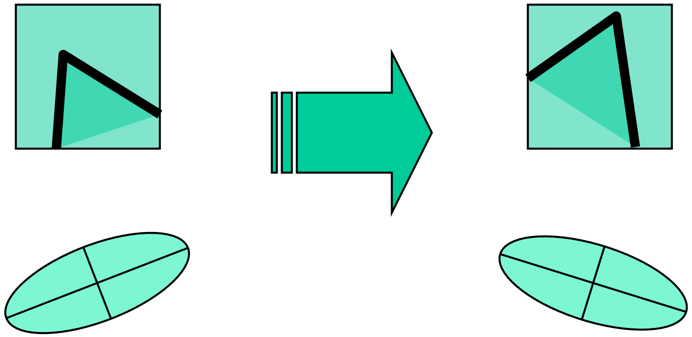
Contents

- 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:

$$|m_{kl}|$$

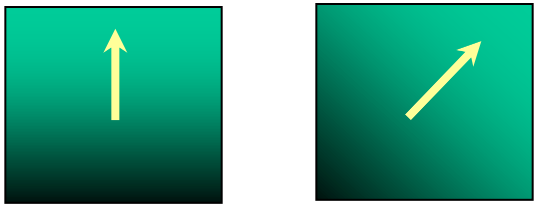
Matching is done by comparing vectors $[|m_{kl}|]_{k,l}$

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

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

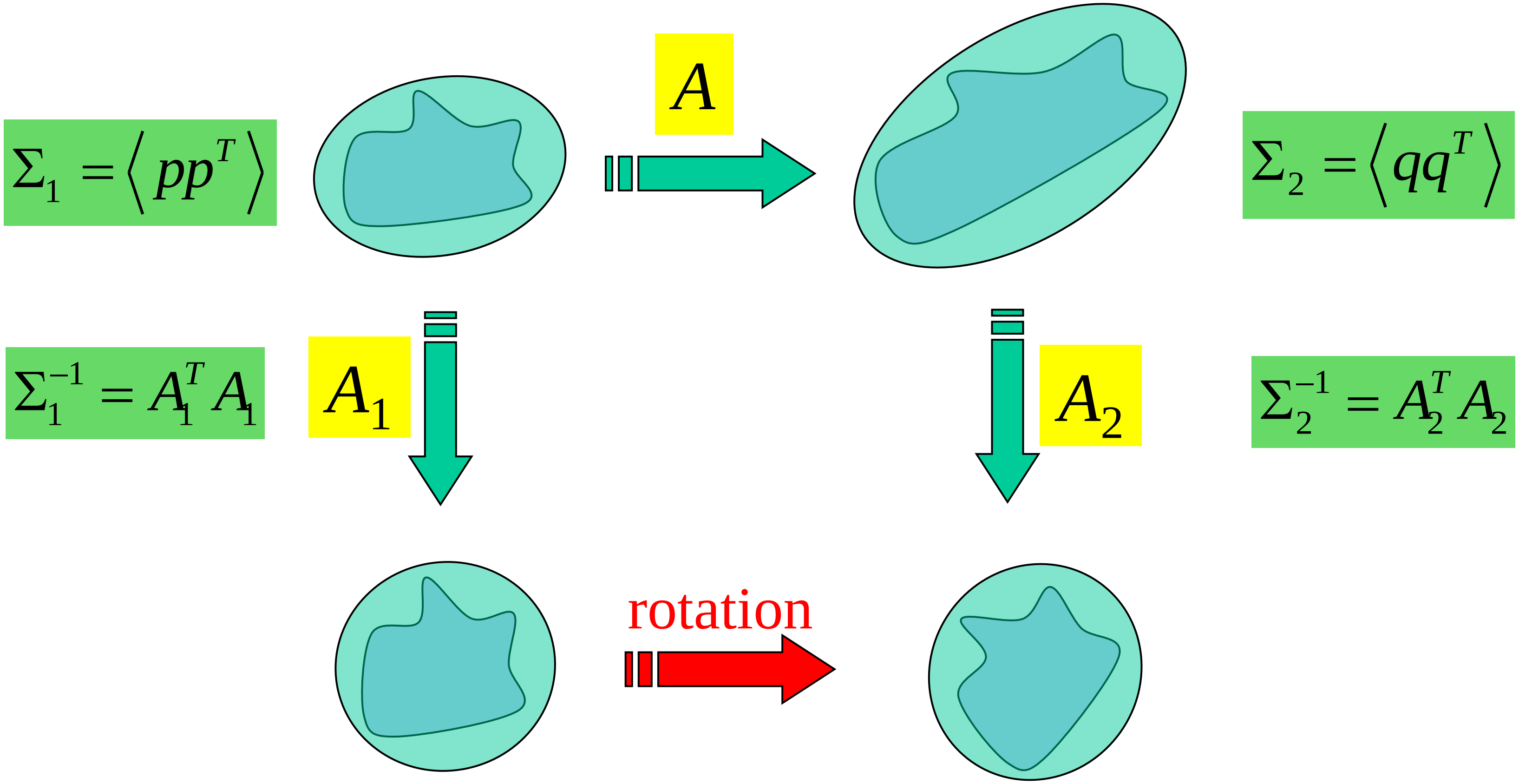
Different combinations of these moments are fully affine invariant

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



$\Sigma_1 = \langle pp^T \rangle$

A

$\Sigma_2 = \langle qq^T \rangle$

$\Sigma_1^{-1} = A_1^T A_1$

A_1

A_2

$\Sigma_2^{-1} = A_2^T A_2$

rotation

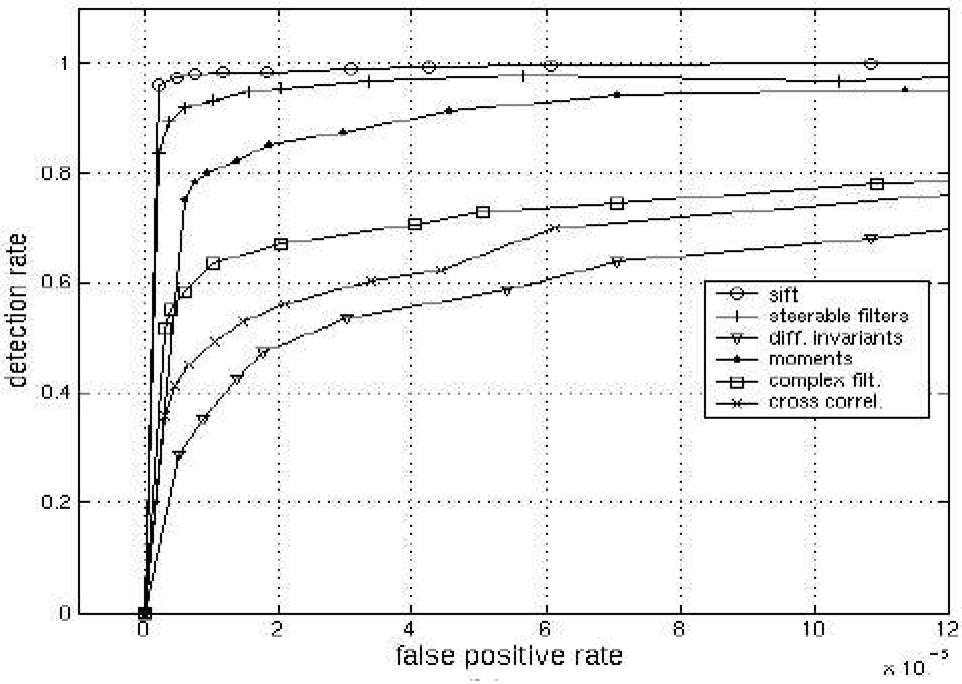
- Compute rotational invariant descriptor in this normalized frame

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

CVPR 2003 Tutorial

Recognition and Matching Based on Local Invariant Features

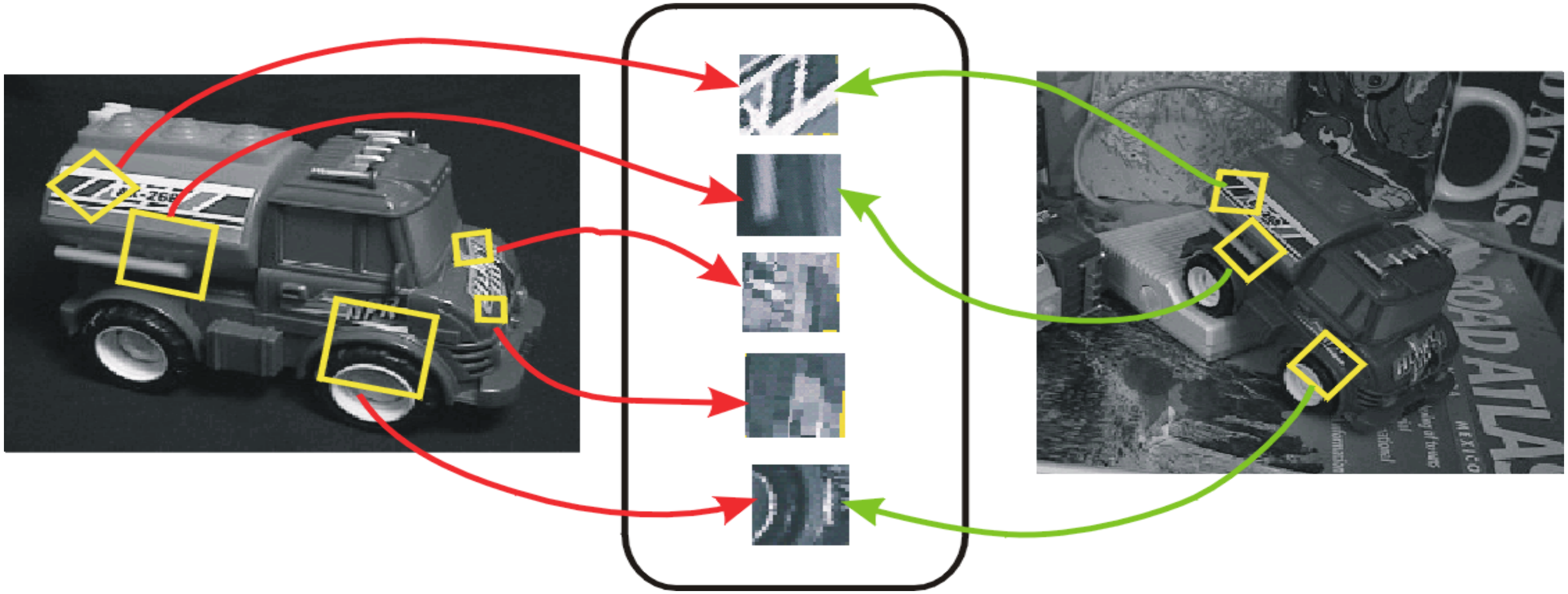
David Lowe

Computer Science Department

University of British Columbia

Invariant Local Features

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



SIFT Features

Advantages of invariant local features

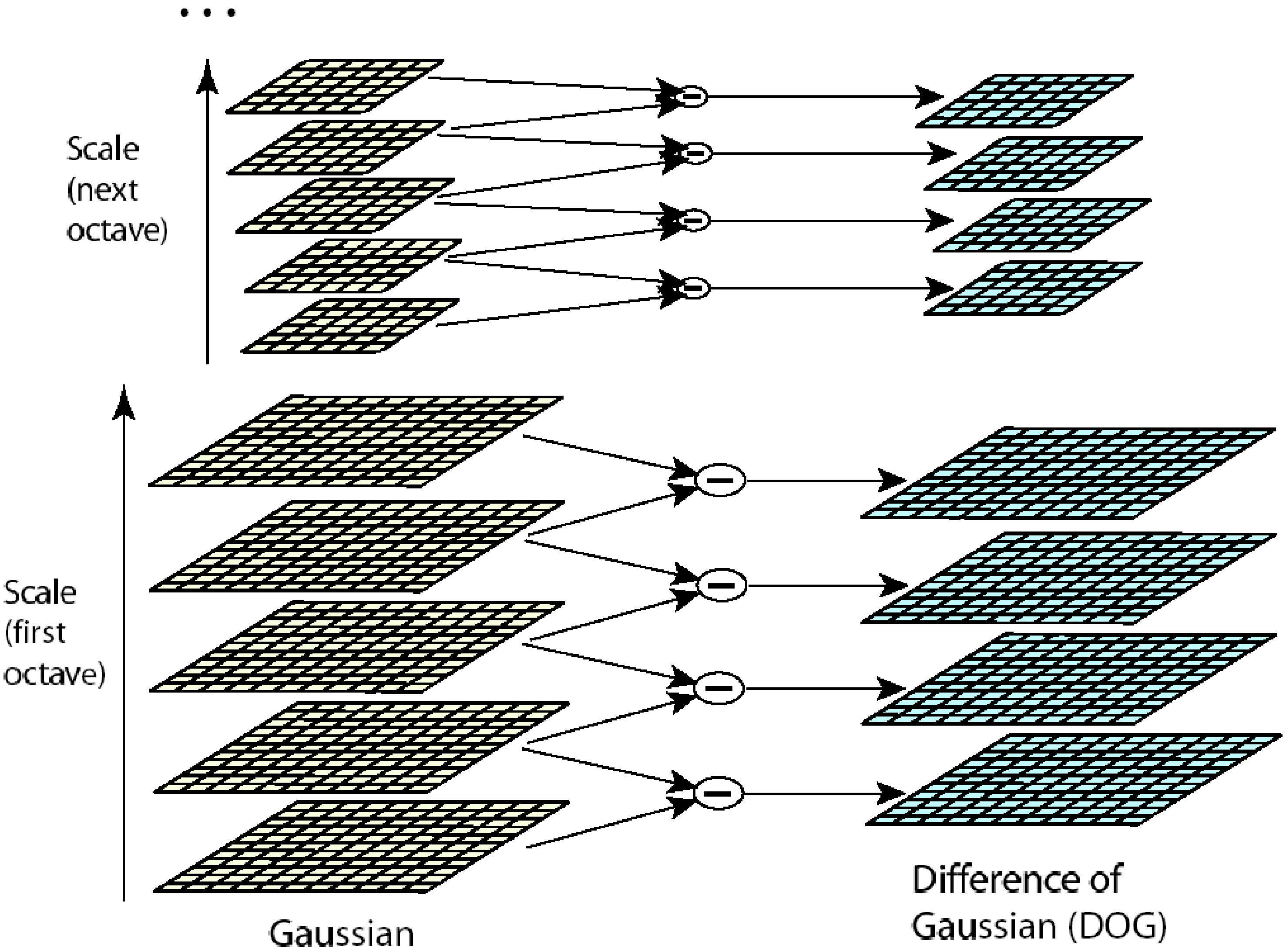
- **Locality:** features are local, so robust to occlusion and clutter (no prior segmentation)
- **Distinctiveness:** individual features can be matched to a large database of objects
- **Quantity:** many features can be generated for even small objects
- **Efficiency:** close to real-time performance
- **Extensibility:** can easily be extended to wide range of differing feature types, with each adding robustness

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)
- An efficient choice is to detect peaks in the difference of Gaussian pyramid (Burt & Adelson, 1983; Crowley & Parker, 1984 – 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 space processed one octave at a time



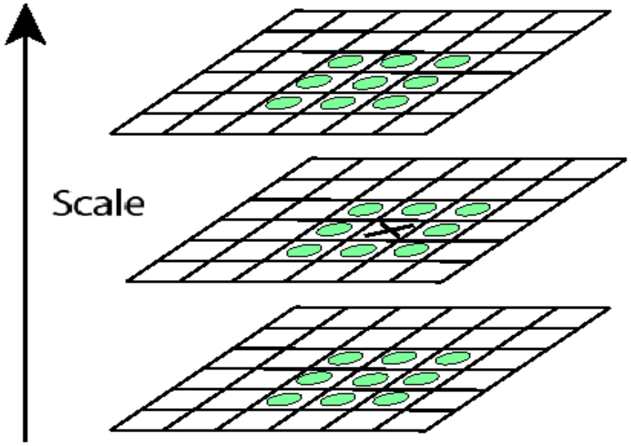
Key point localization

- 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:

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

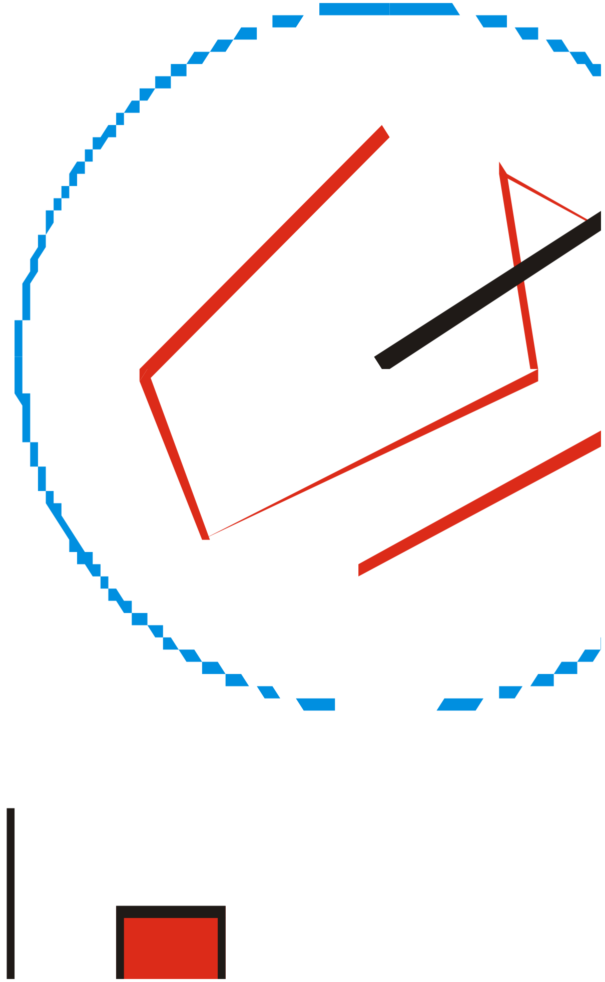
- Offset of extremum (use finite differences for derivatives):

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



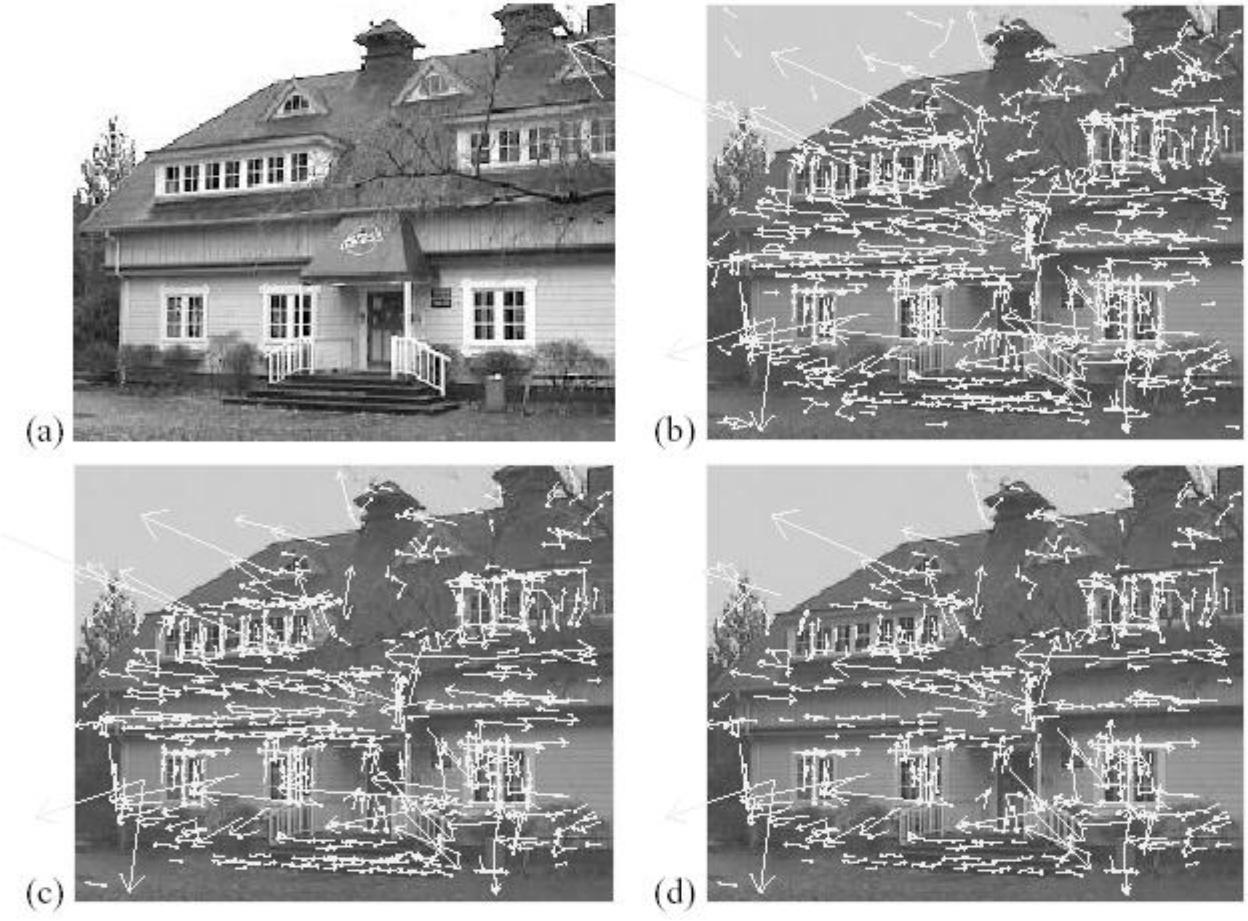
Select canonical orientation

- 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)



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 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

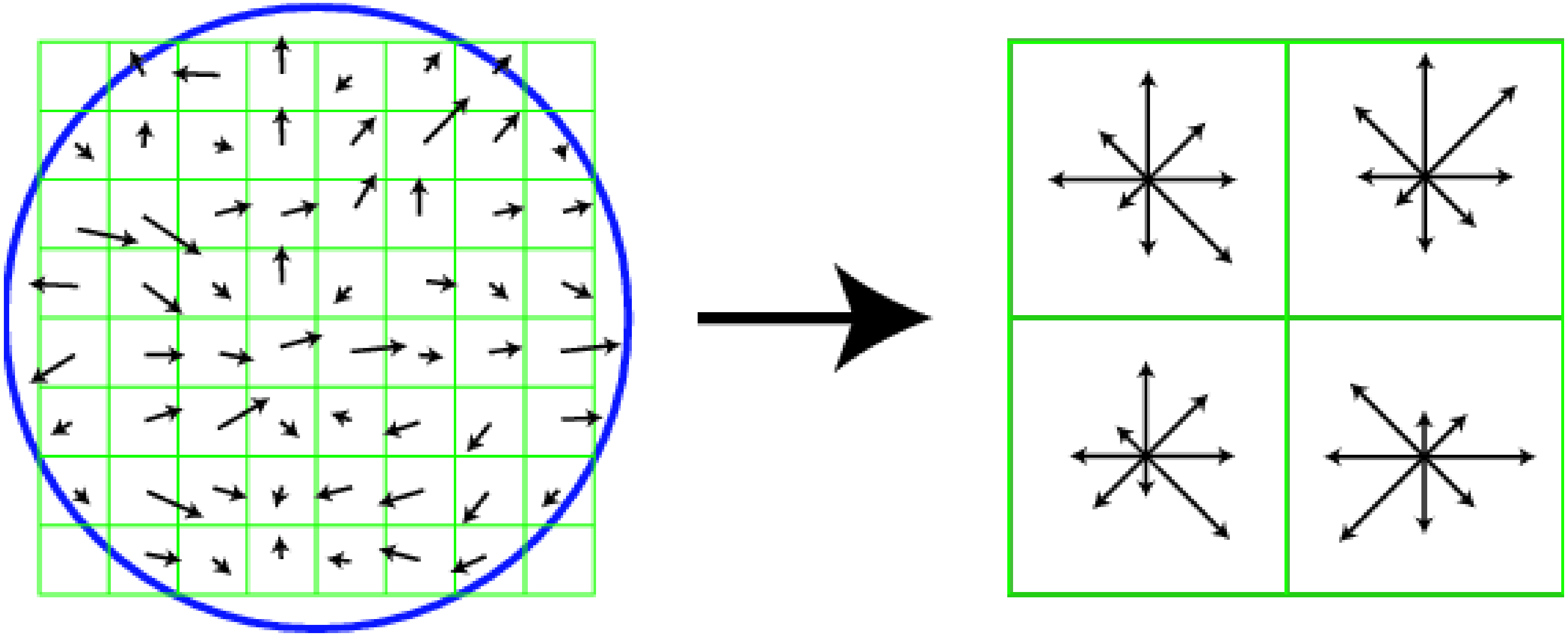
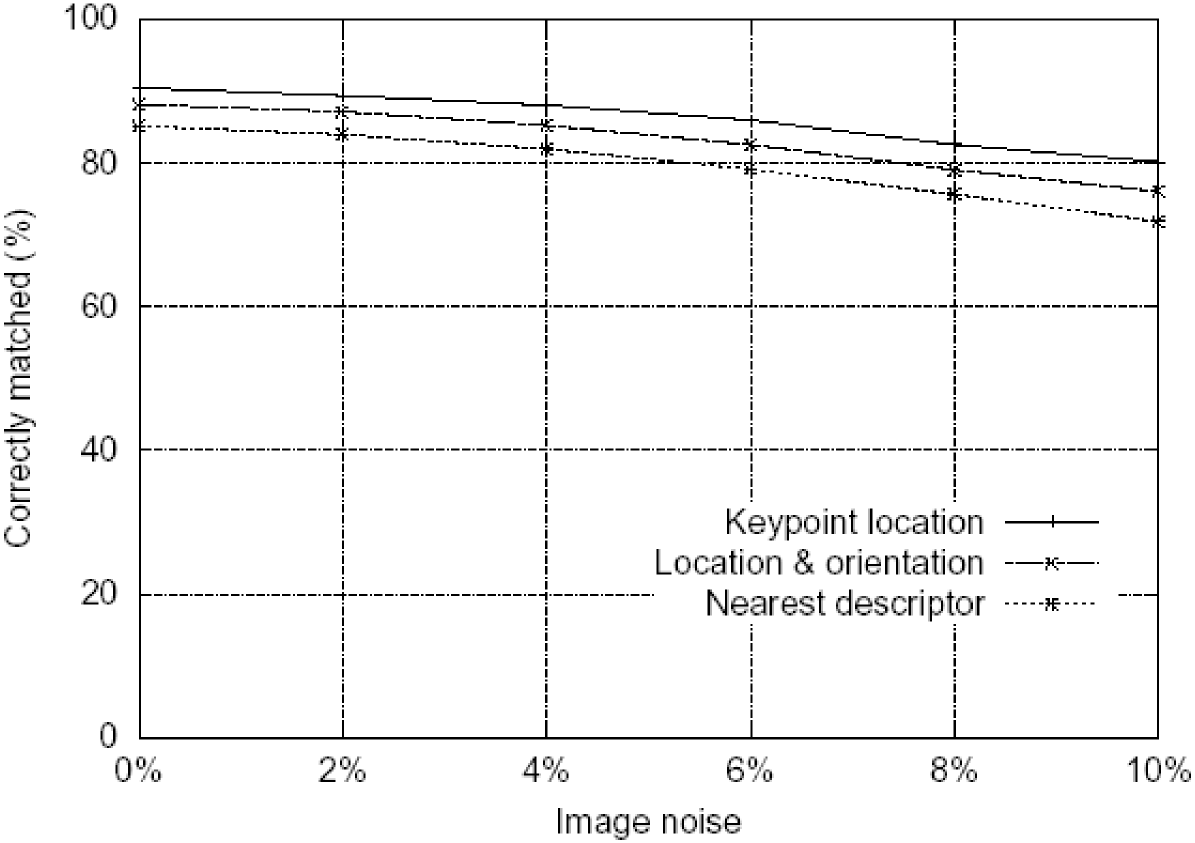


Image gradients

Keypoint descriptor

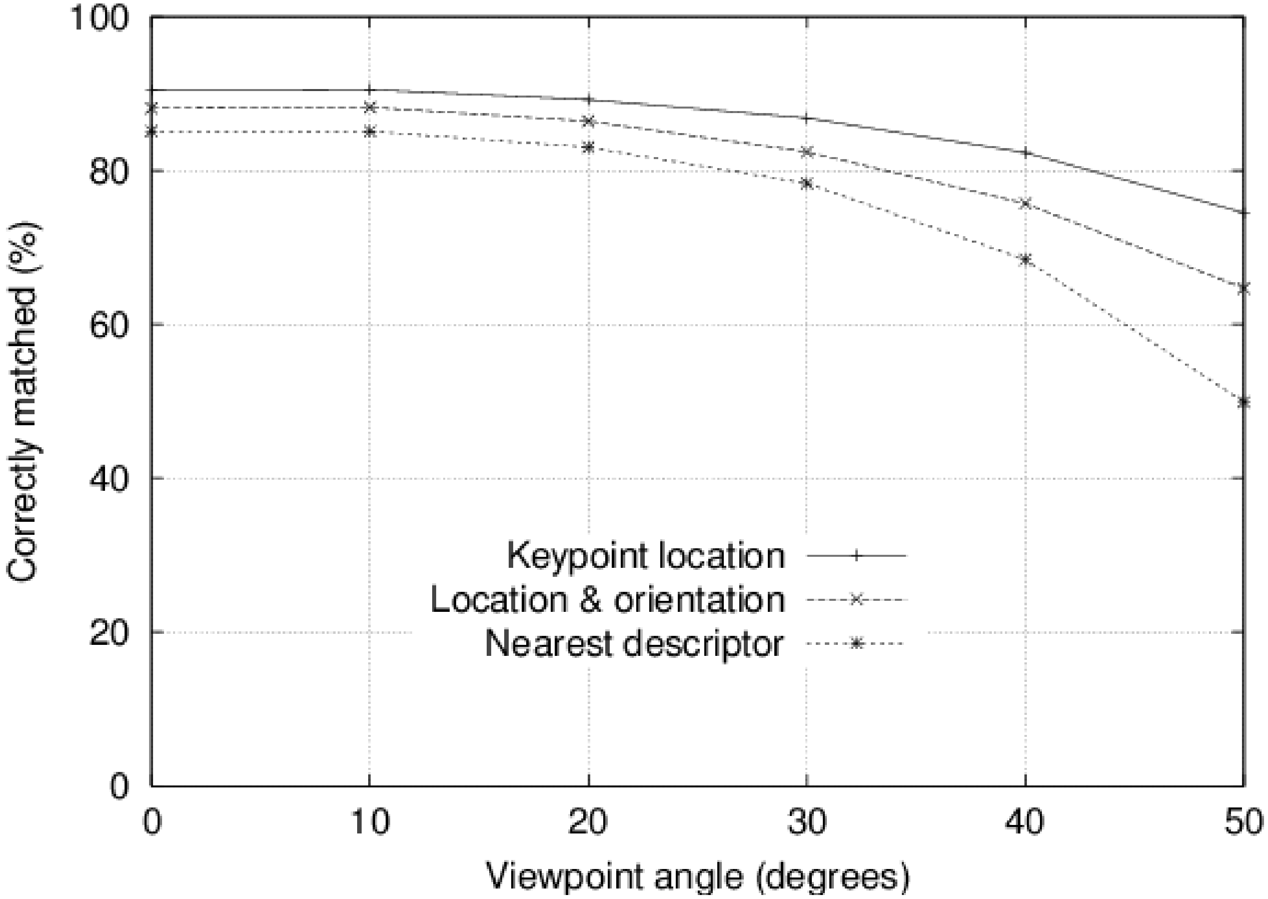
Feature stability to noise

- Match features after random change in image scale & orientation, with differing levels of image noise
- Find nearest neighbor in database of 30,000 features



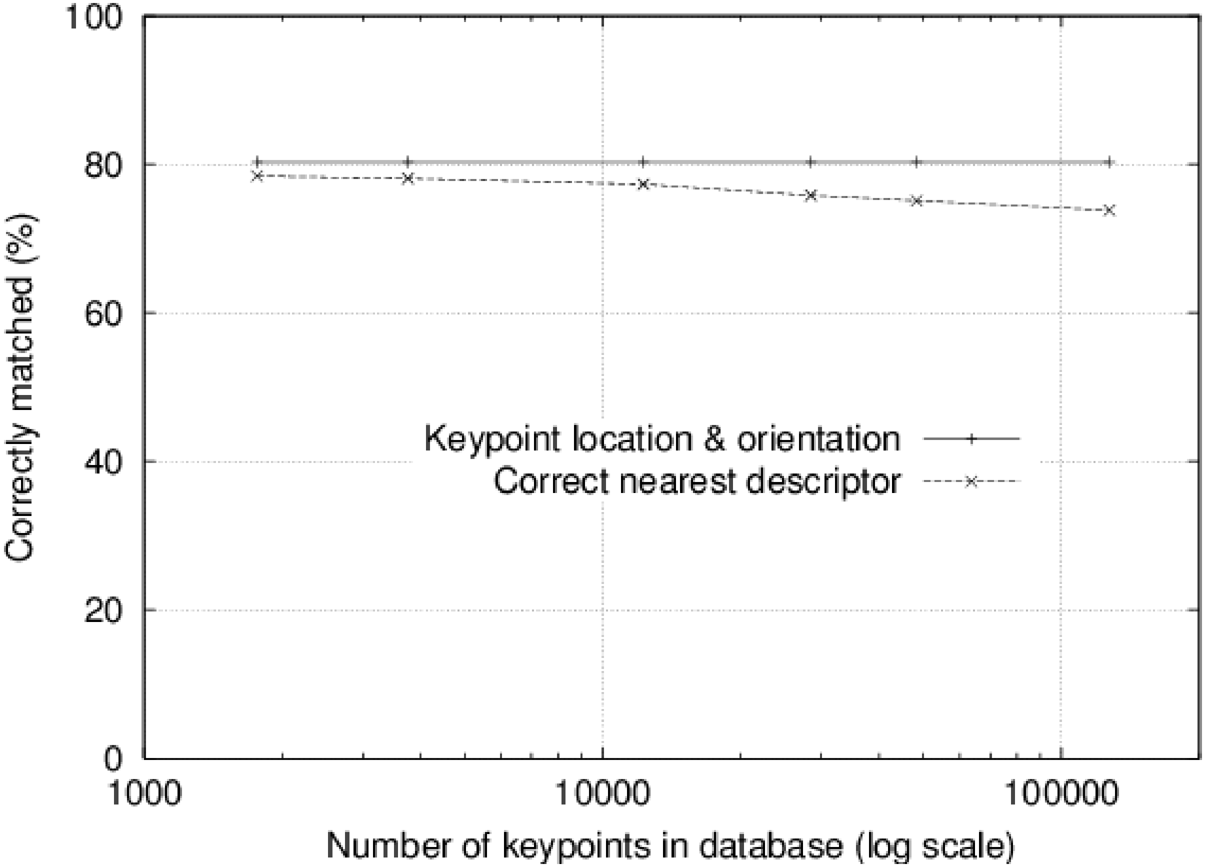
Feature stability to affine change

- Match features after random change in image scale & orientation, with 2% image noise, and affine distortion
- Find nearest neighbor in database of 30,000 features



Distinctiveness of features

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



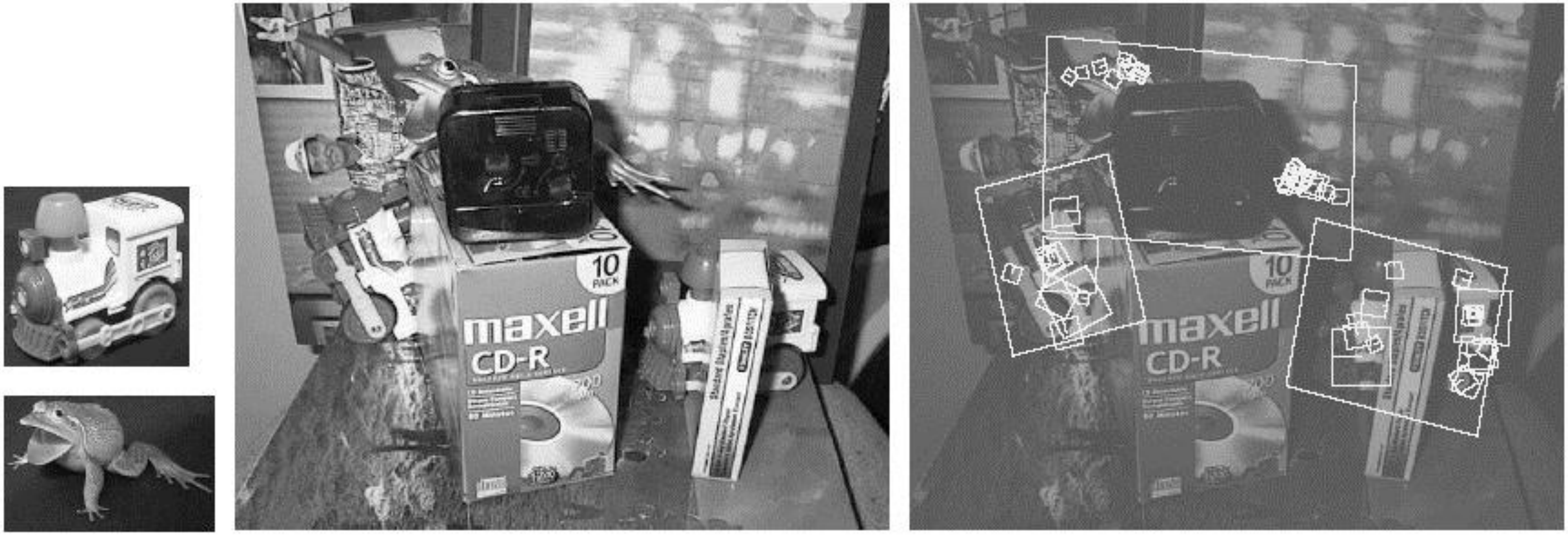


Figure 12: The training images for two objects are shown on the left. These can be recognized in a cluttered image with extensive occlusion, shown in the middle. The results of recognition are shown on the right. A parallelogram is drawn around each recognized object showing the boundaries of the original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

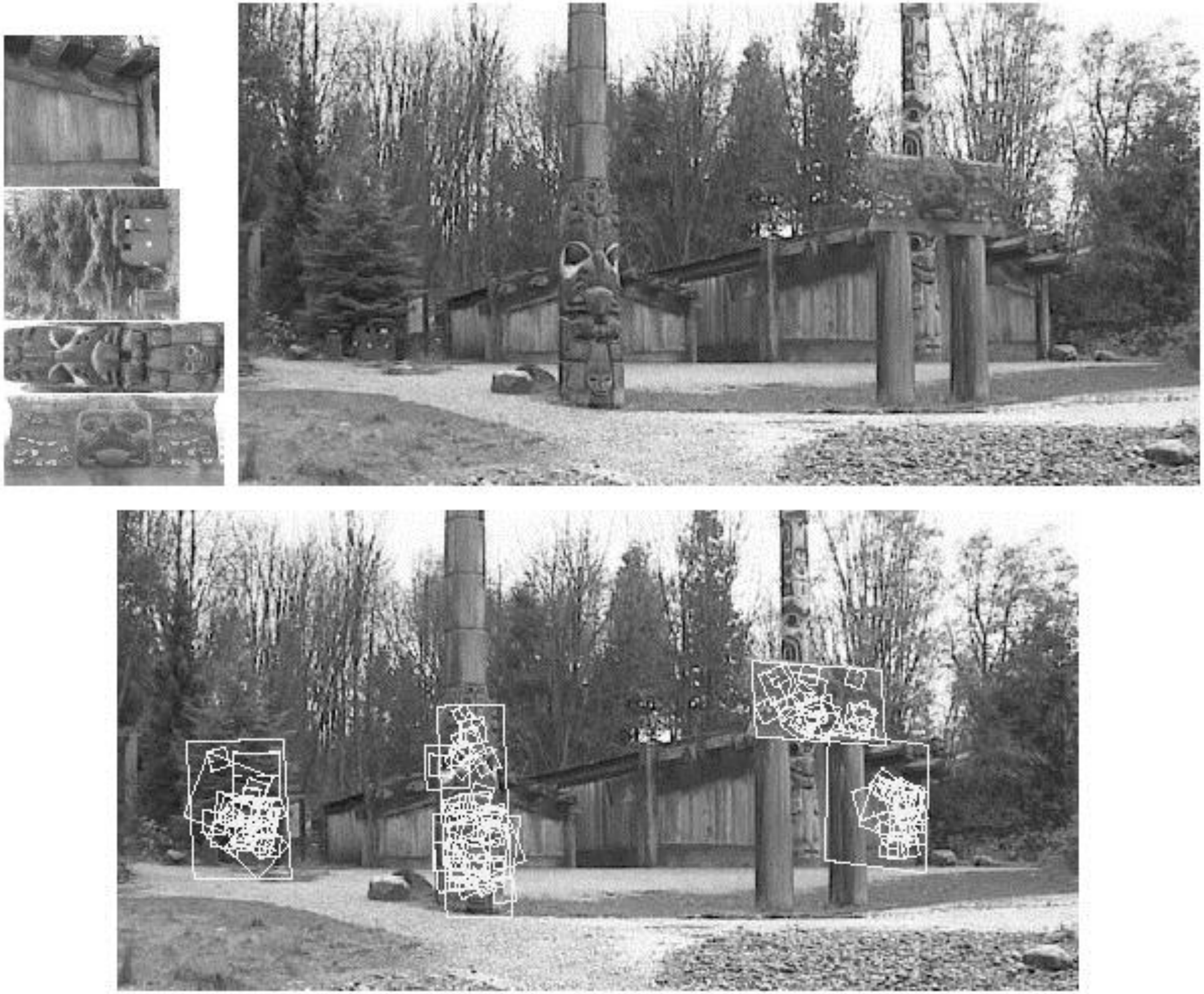


Figure 13: This example shows location recognition within a complex scene. The training images for locations are shown at the upper left and the 640x315 pixel test image taken from a different viewpoint is on the upper right. The recognized regions are shown on the lower image, with keypoints shown as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

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

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)