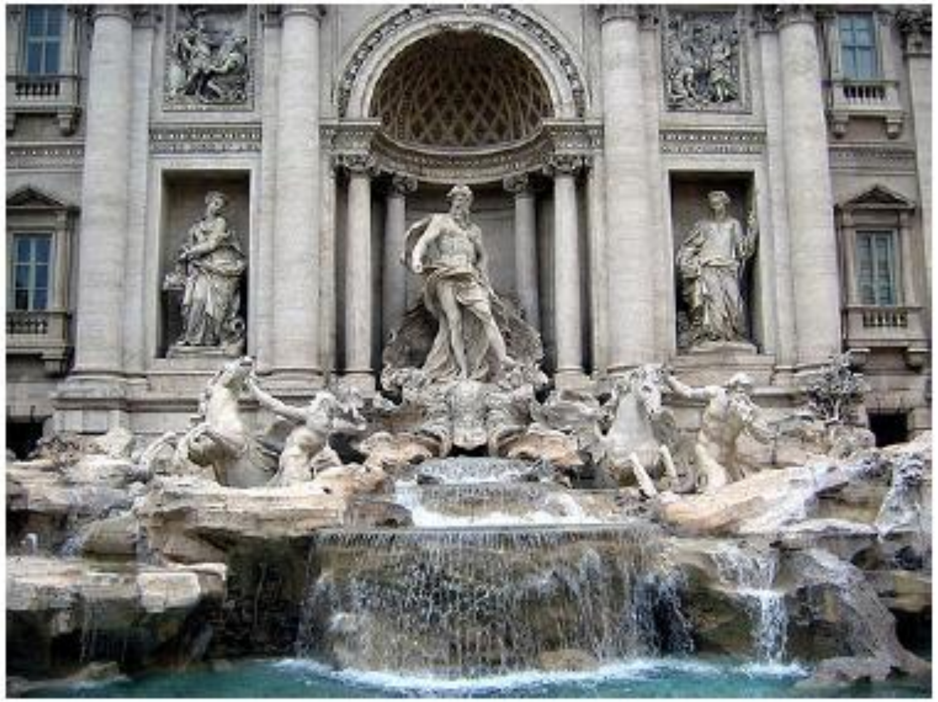
Feature Extraction

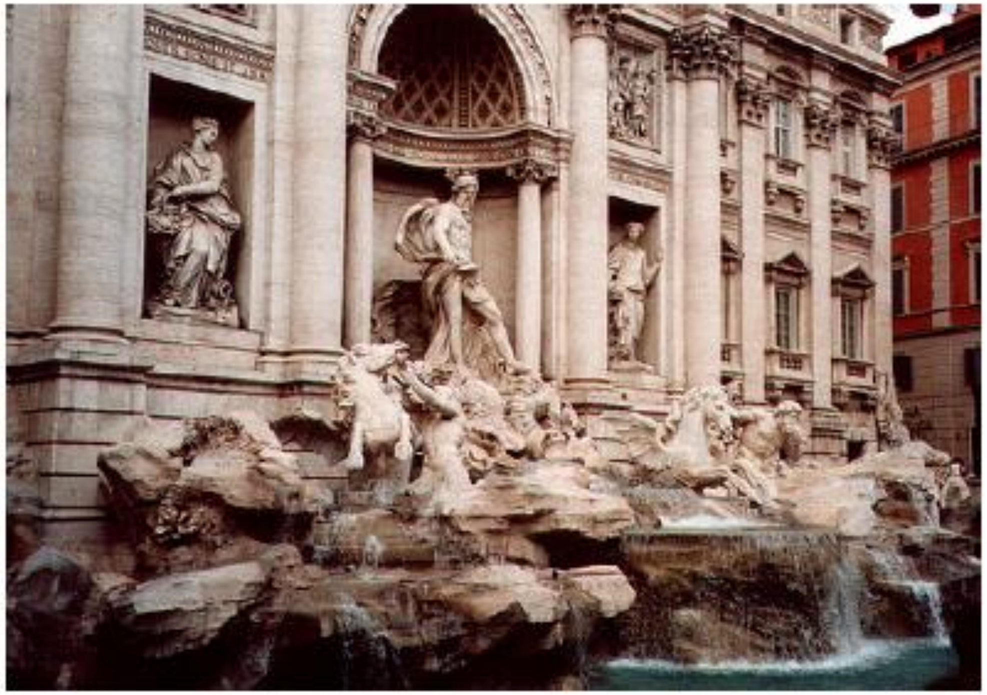
FCV

a challenging problem

Harder Case



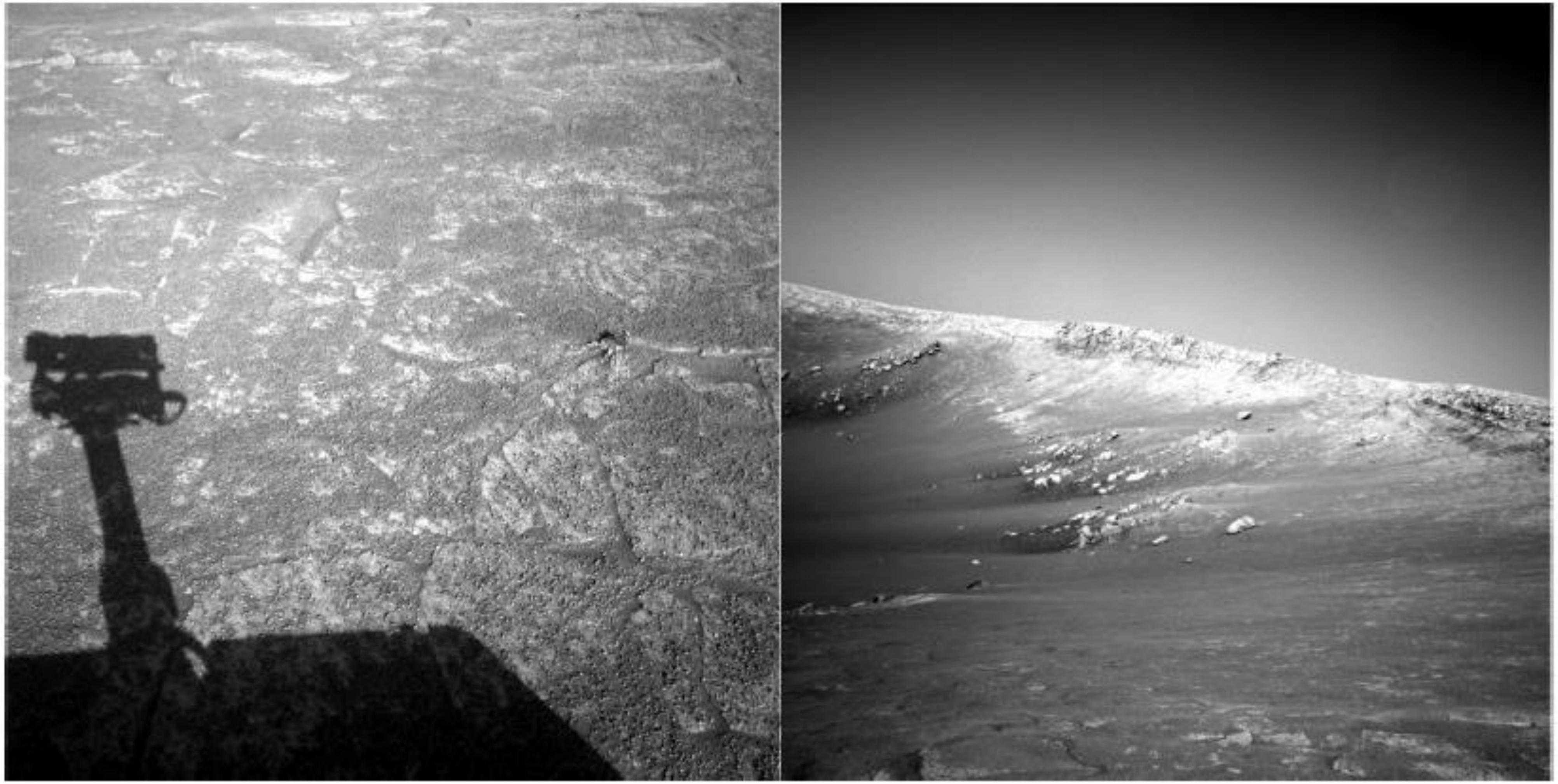
by Diva Sian



by scgbt

Slide credit: Steve Seitz

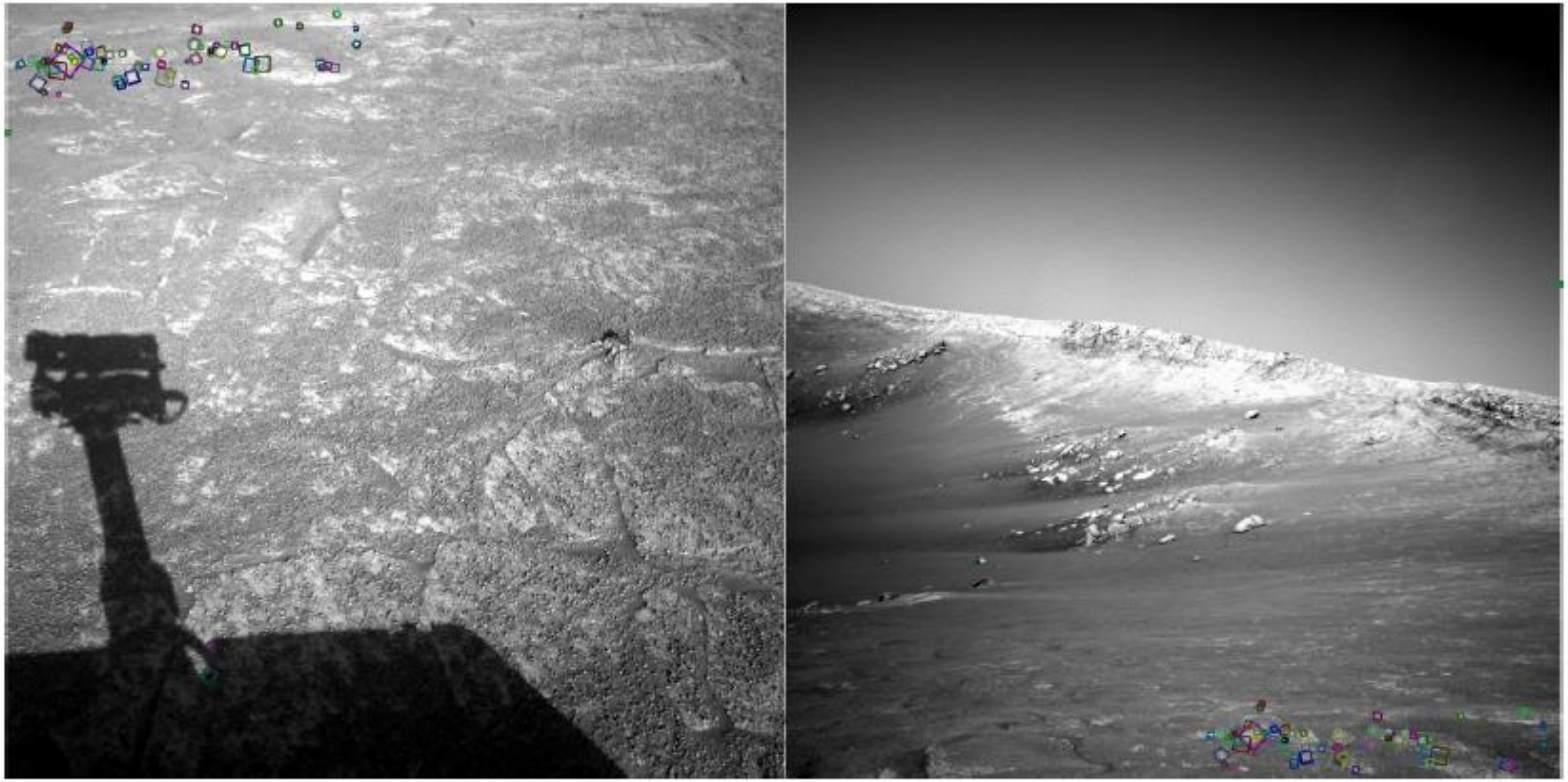
Harder Still?



NASA Mars Rover images

Slide credit: Steve Seitz

Answer Below (Look for tiny colored squares)

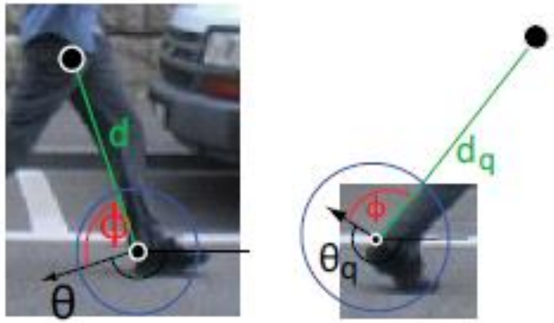


NASA Mars Rover images with SIFT feature matches
(Figure by Noah Snavely)

Slide credit: Steve Seitz

Motivation for using local features

- Global representations have major limitations
- Instead, describe and match only local regions
- Increased robustness to
  - Occlusions
  - Articulation
  - Intra-category variations



Objects and Interest Points (IPs)

1. **Feature detection**

   Extract interest points
   (unique image regions)

2. **Feature description**

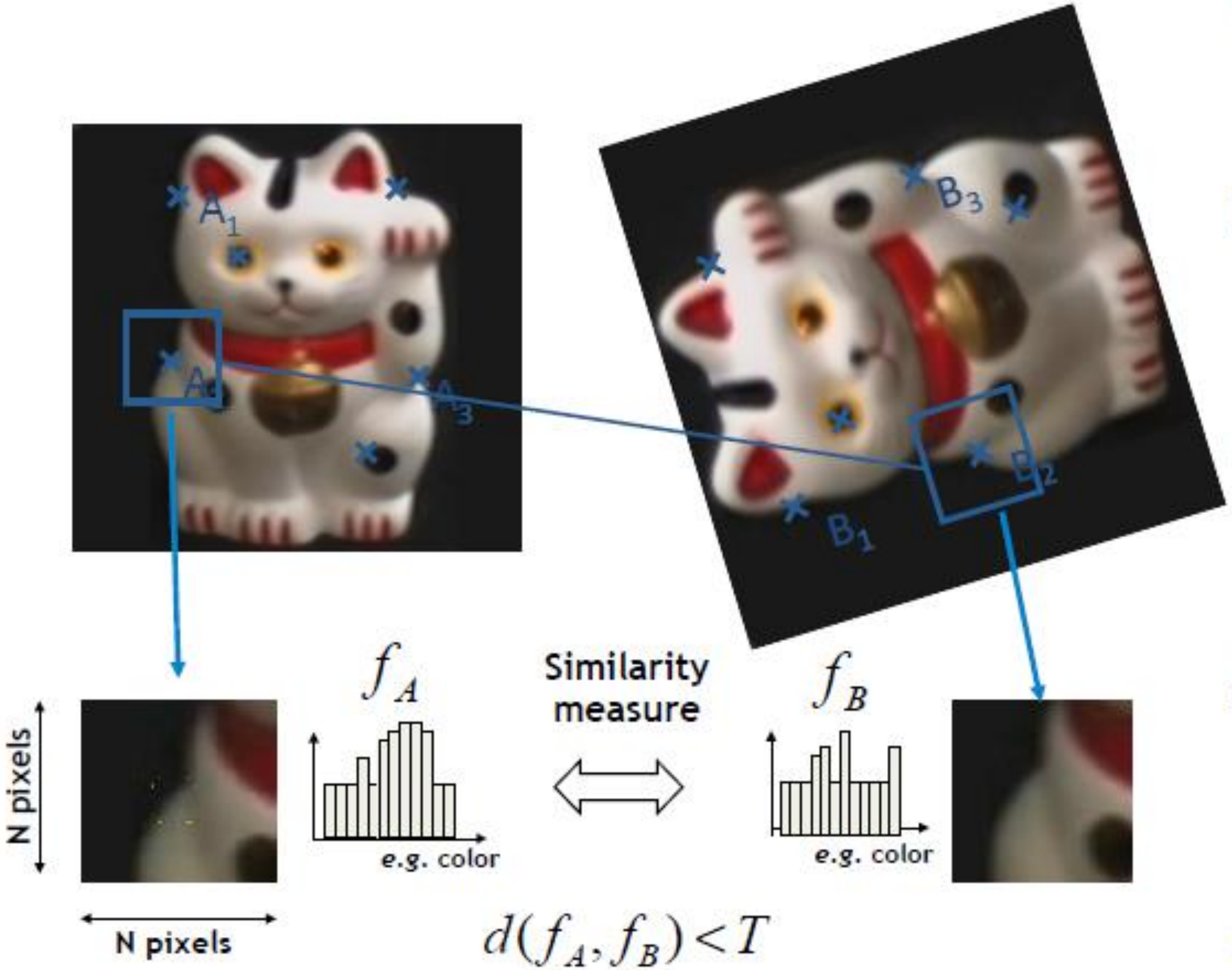
   Calculate local
   (invariant) descriptors

3. **Feature matching / feature tracking**

   Find correspondences

4. **Find similar image regions/objects**

General Approach



Similarity measure

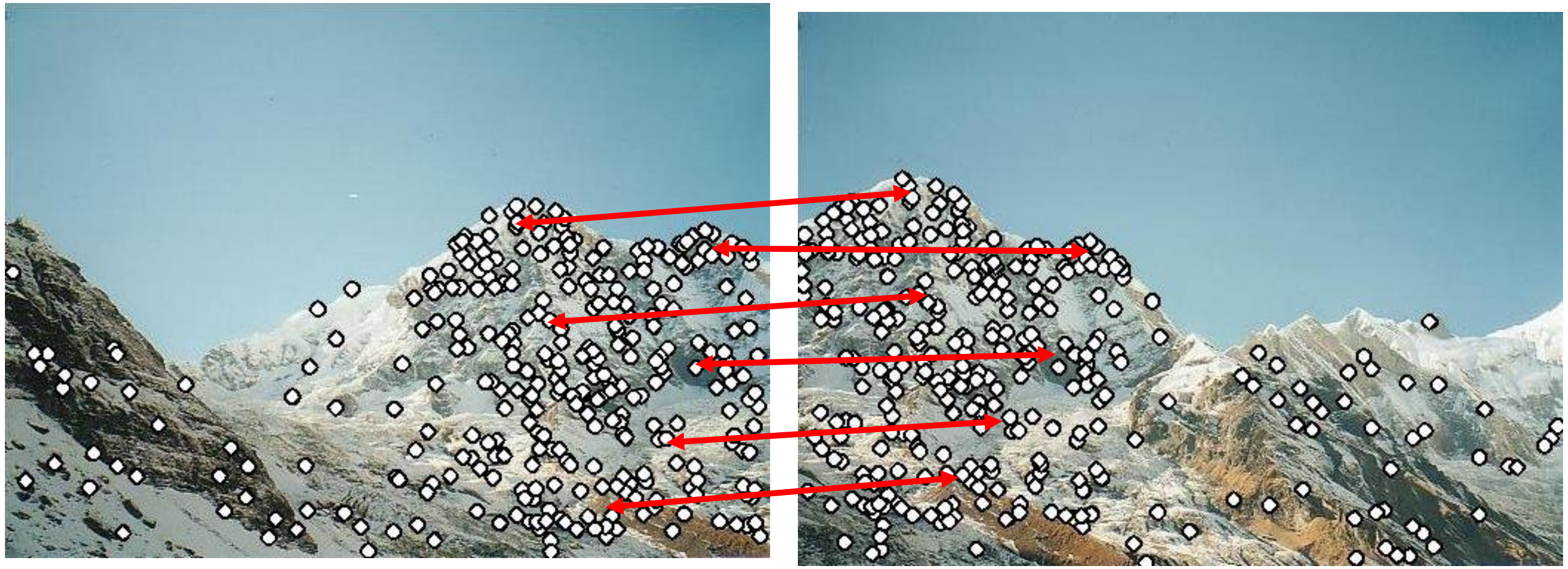
$$d(f_A, f_B) < T$$

1. Find a set of distinctive keypoints
2. Define a region around each keypoint
3. Extract and normalize the region content
4. Compute a local descriptor from the normalized region
5. Match local descriptors

Slide credit: Bastian Leibe

Example

- Motivation: Image stitching
    - We have two images – how do we combine them?

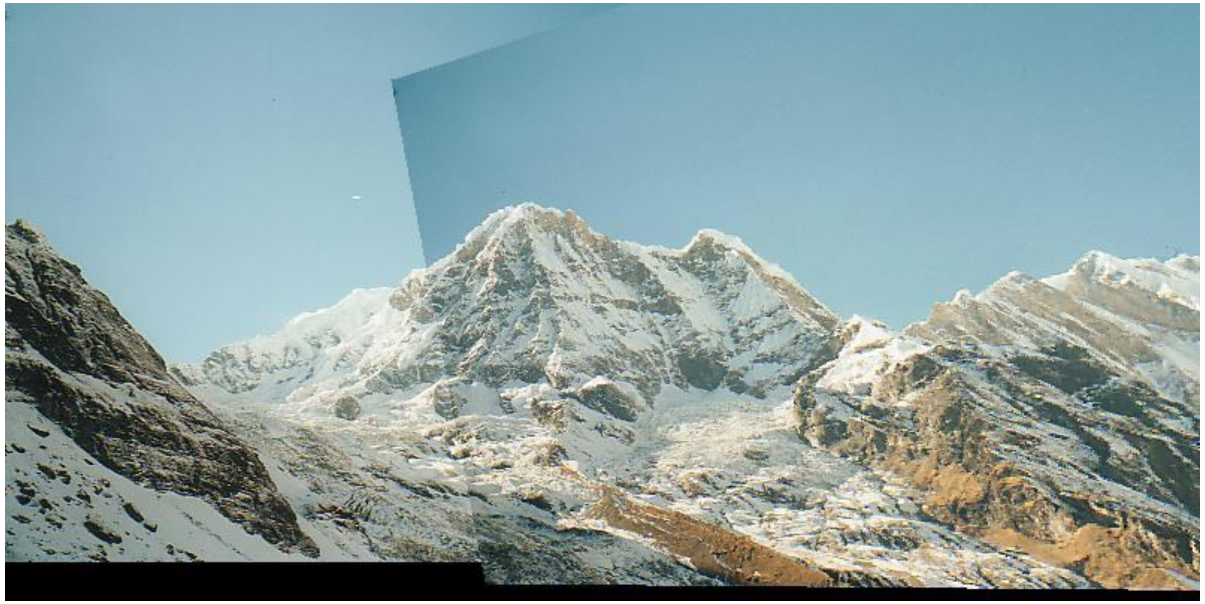


Step 1: extract features

Step 2: match features

Example

- Motivation: Image stitching
  - We have two images – how do we combine them?



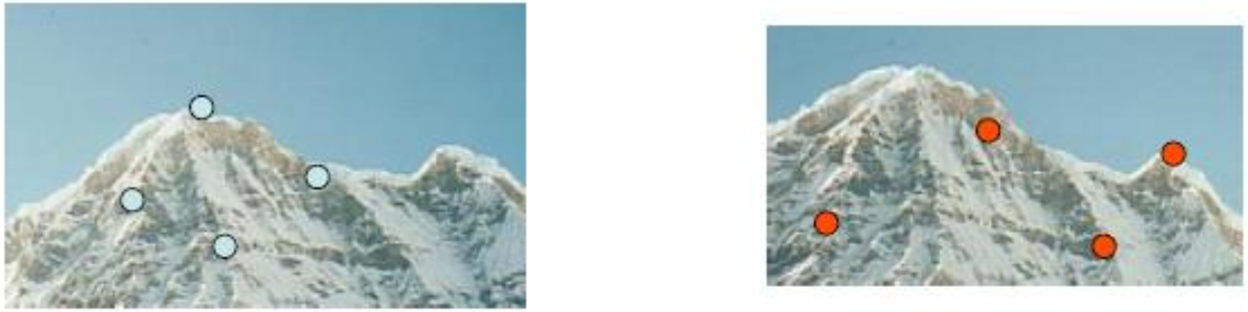
Step 1: extract features

Step 2: match features

Step 3: align images

Common Requirements

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

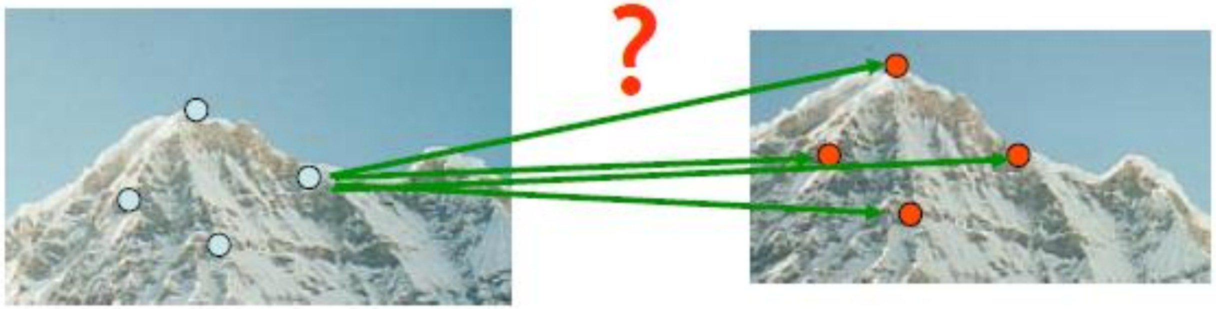


No chance to match!

We need a repeatable detector!

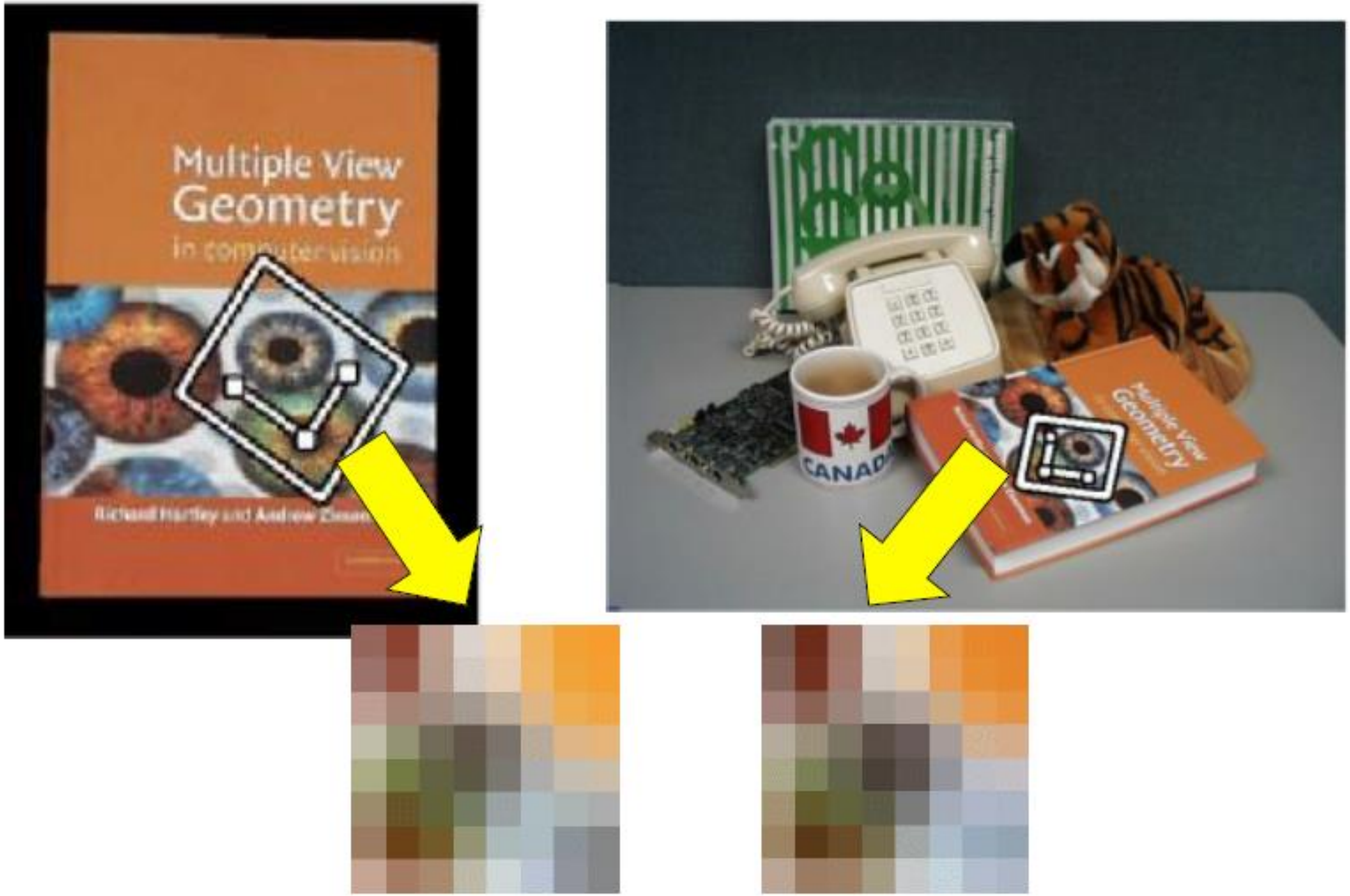
Common Requirements

- Problem 1:
  - Detect the same point *independently* in both images
- Problem 2:
  - For each point correctly recognize the corresponding one

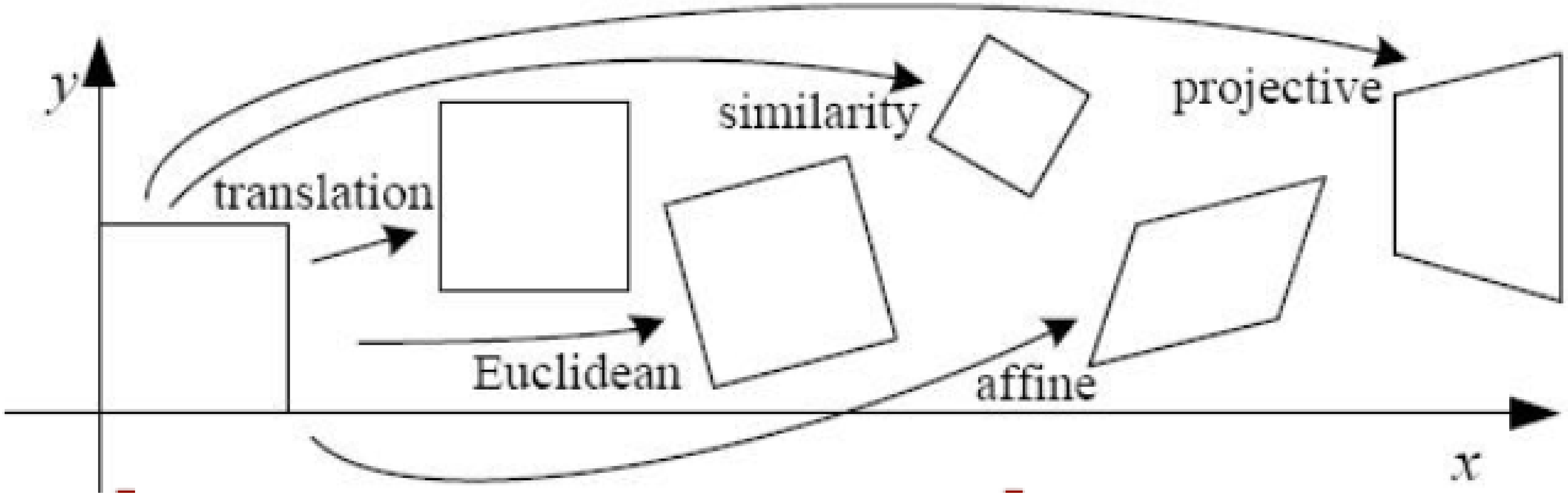


We need a reliable and distinctive descriptor!

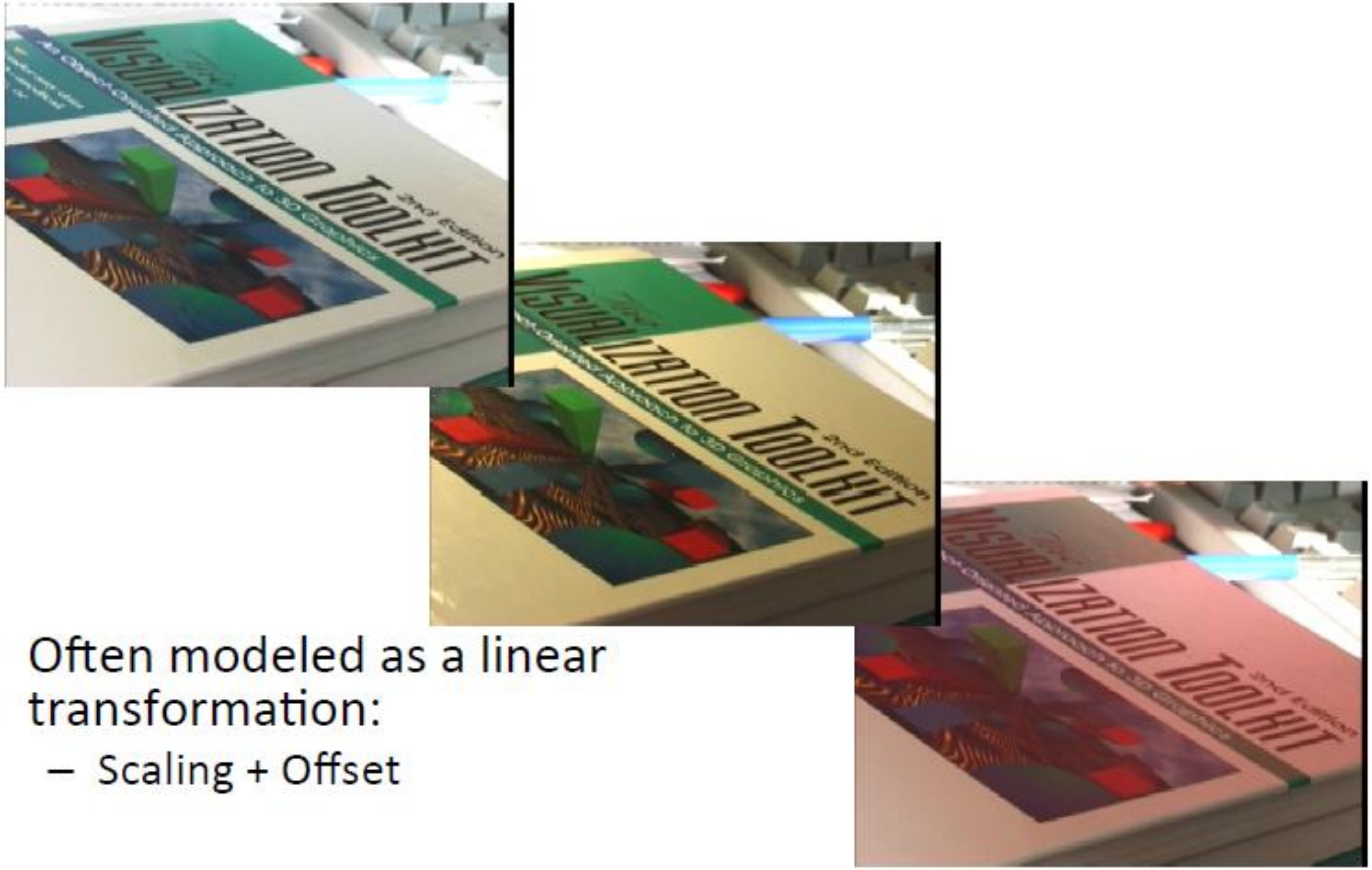
Invariance: Geometric Transformations



Levels of Geometric Invariance



Invariance: Photometric Transformation



- Often modeled as a linear transformation:
  - Scaling + Offset

Requirements

- Region extraction needs to be repeatable and accurate
  - Invariant to translation, rotation, scale changes
  - Robust or covariant to out-of-plane (?affine) transformations
  - Robust to lighting variations, noise, blur, quantization
- Locality: Features are local, therefore robust to occlusion and clutter.
- Quantity: We need a sufficient number of regions to cover the object.
- Distinctivenes : The regions should contain “interesting” structure.
- Efficiency: Close to real-time performance.

Interest Point Detection

Many different approaches

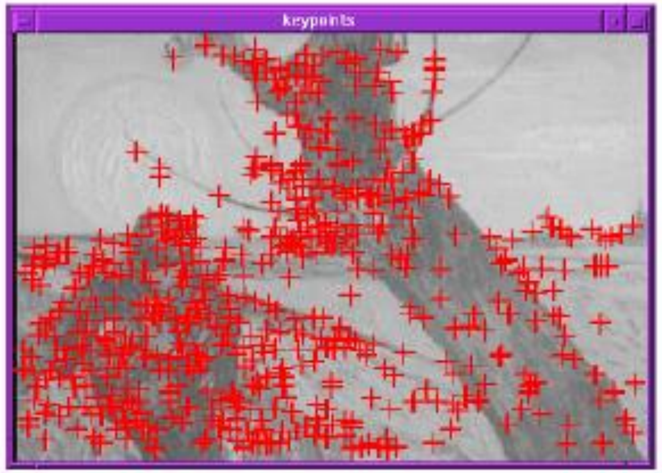
- Corner detector: Harris (1988), Hessian
- Multi-scale corner detector with scale selection
  - Scale invariant Harris and Hessian corners
  - Difference of Gaussian (DoG) (Lowe 2004)
- Affine covariant Regions
  - Harris-Affine (Mikolajczyk, Schmid ‘02, Schaffalitzky, Zisserman ’02)
  - Hessian-Affine (Mikolajczyk and Schmid ’02)
  - Maximally stable extremal regions (MSER) (Matas et al. ’02)
  - Intensity based regions (IBR) (Tuytelaars and Van Gool ’00)
  - Edge based regions (EBR) (Tuytelaars and Van Gool ’00)
  - Entropy-based regions (salient regions) (Kadir et al. ’04)
  - Features from accelerated segment test (FAST) (Rosten et al. '05)

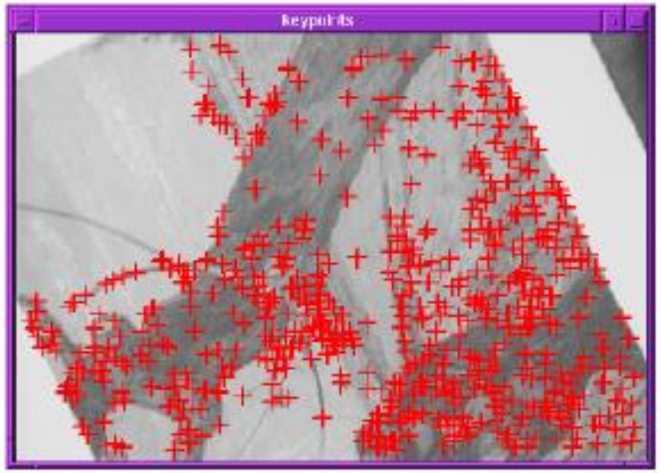
Credit: Markus Vincze, Technische Universität Wien

Applications

- Feature points are used for:
  - Image alignment
  - 3D reconstruction
  - Motion tracking
  - Robot navigation
  - Indexing and database retrieval
  - Object recognition

Finding Corners



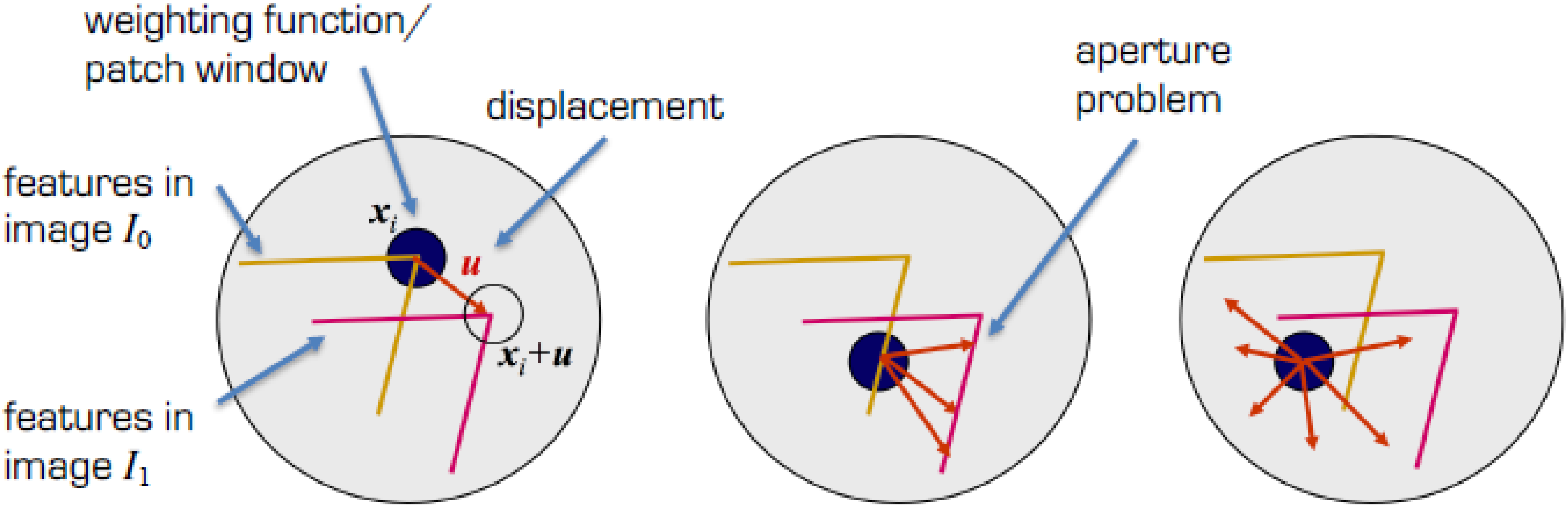


- Key property:
  - In the region around a corner, image gradient has two or more dominant directions
- Corners are *repeatable* and *distinctive*

C.Harris and M.Stephens. "A Combined Corner and Edge Detector.“
Proceedings of the 4th Alvey Vision Conference, 1988.

Interest Point Detection

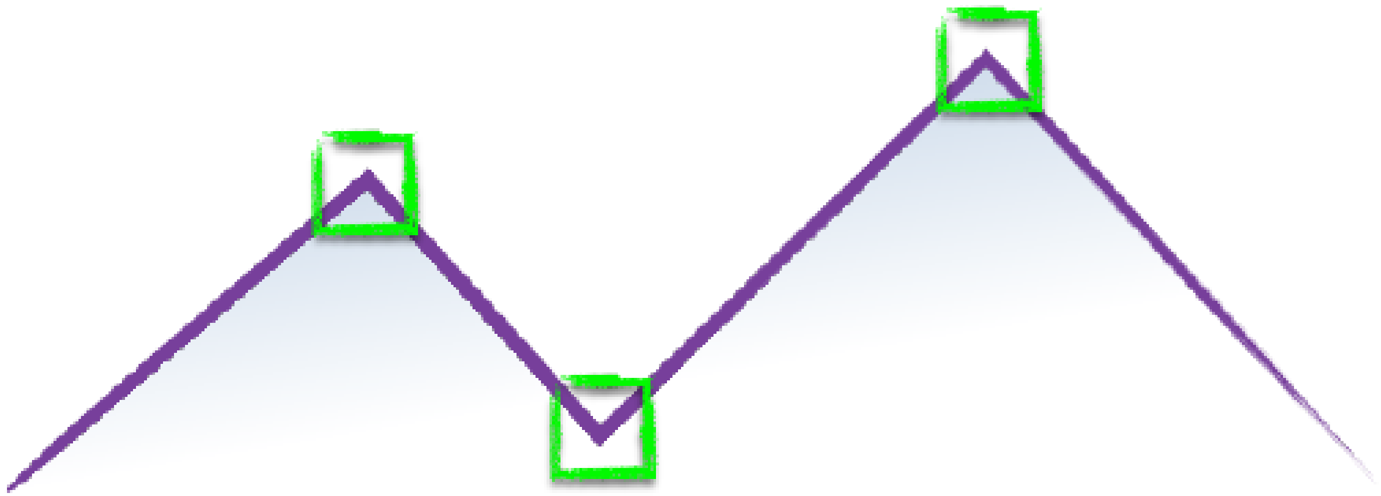
- Textureless patches are almost impossible to localize
- Patches with high contrast (gradient) are easier to localize
- Straight-line segments suffer from the aperture problem



Corners ☺ Edges ☹ Textureless ☹☹

How do you find a corner?

[Moravec 1980]

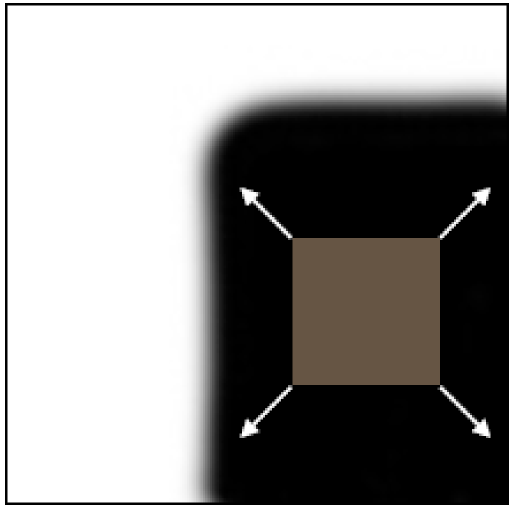


Easily recognized by looking through a small window

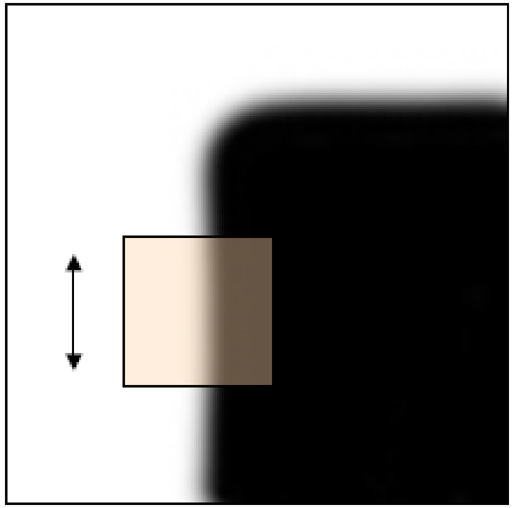
Shifting (displacing) the window should give large change in intensity

Corners as Distinctive Interest Points

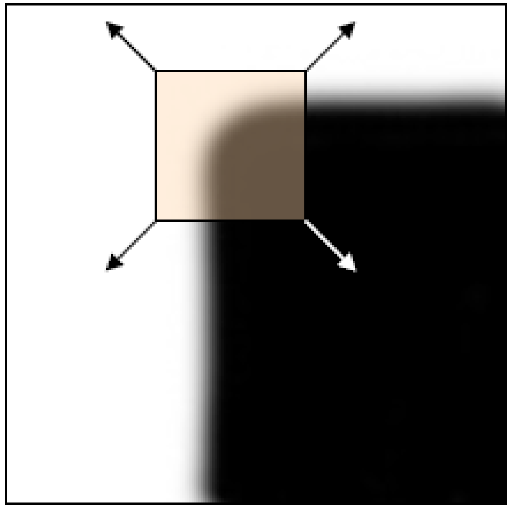
- Design criteria
  - We should easily recognize the point by looking through a small window (*locality*)
  - Shifting the window in *any direction* should give *a large change* in intensity (*good localization*)



“flat” region:
no change in all directions

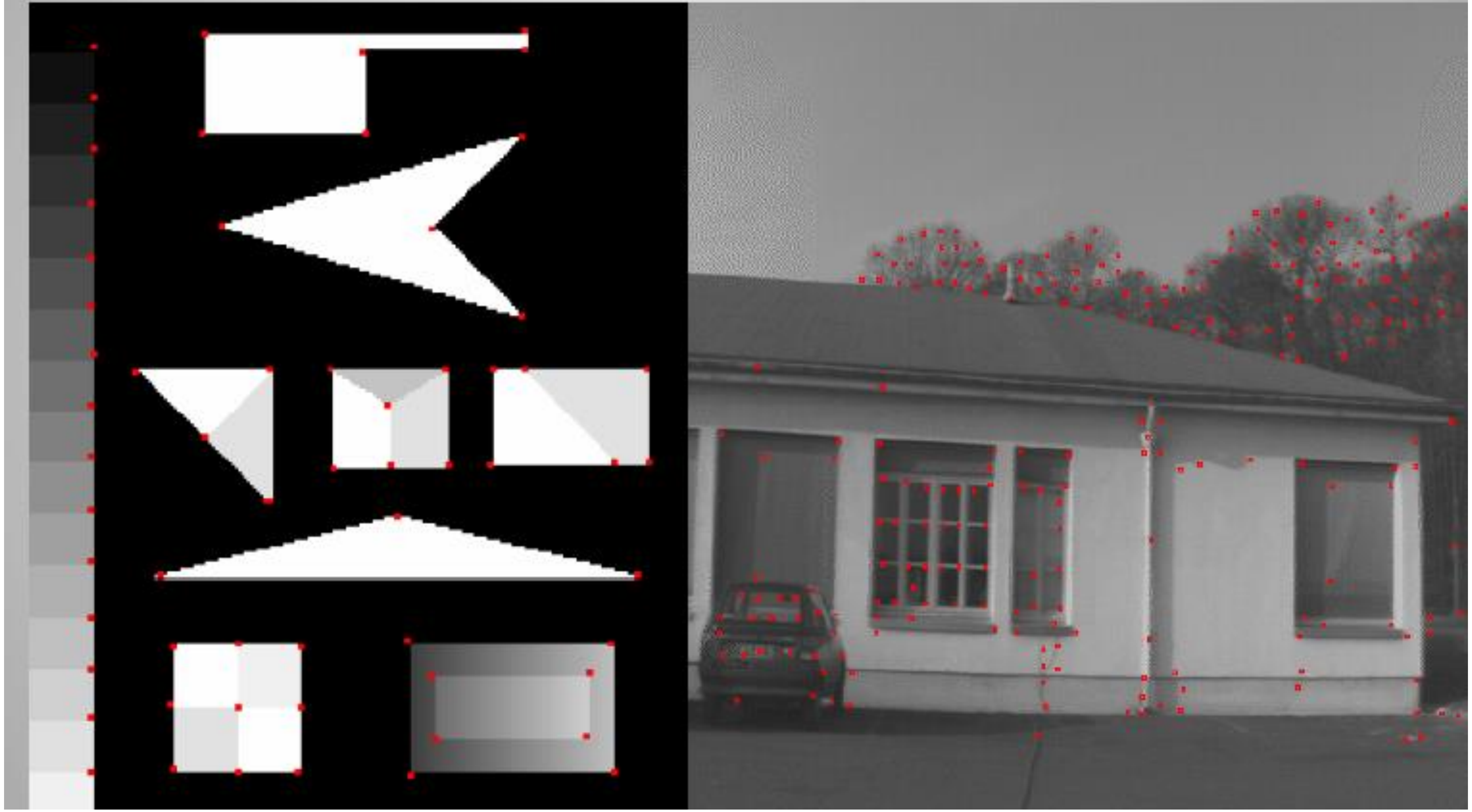


“edge”:
no change along the edge direction



“corner”:
significant change in all directions

What is an interest point?

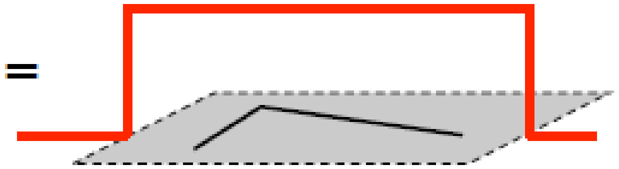


Harris Detector Formulation

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

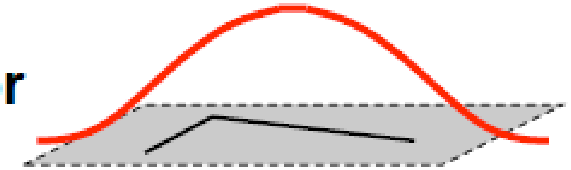
$$E(u,v)=\sum_{x,y}\underbrace{w(x,y)}_{\text{Window function}}\Big[\underbrace{I(x+u,y+v)}_{\text{Shifted intensity}}-\underbrace{I(x,y)}_{\text{Intensity}}\Big]^2$$

Window function $w(x,y)$ =



1 in window, 0 outside

or



Gaussian

Correlation

$$f \otimes h = \sum_{k} \sum_{l} f(k,l) h(k,l)$$

f = Image

h = Kernel

f

f_1	f_2	f_3
f_4	f_5	f_6
f_7	f_8	f_9

$\otimes$

h

h_1	h_2	h_3
h_4	h_5	h_6
h_7	h_8	h_9

$$\rightarrow \quad \begin{aligned} f * h = {} & f_1 h_1 + f_2 h_2 + f_3 h_3 \\ & + f_4 h_4 + f_5 h_5 + f_6 h_6 \\ & + f_7 h_7 + f_8 h_8 + f_9 h_9 \end{aligned}$$

Correlation vs SSD

$$SSD = \sum_k \sum_l \left(f(k,l) - h(k,l)\right)^2$$

Sum of Squares Difference

minimize

$$SSD = \sum_k \sum_l \left(\cancel{f}(k,l)^2 - 2h(k,l)f(k,l) + h(\cancel{k},l)^2\right)$$

minimize

$$SSD = \sum_k \sum_l \left(-2h(k,l)f(k,l)\right)$$

These terms do not depend on correlation

$$SSD = \sum_k \sum_l \left(\cancel{2}h(k,l)f(k,l)\right)$$

maximize

$$Correlation = \sum_k \sum_l \left(h(k,l)f(k,l)\right)$$

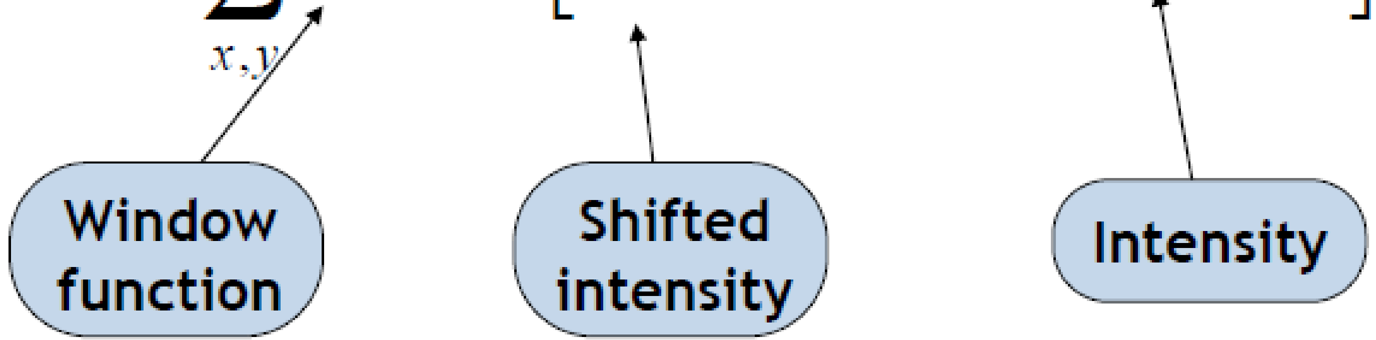
maximize

$$f \otimes f = \sum_k \sum_l f(k,l)f(k,l)$$

Harris Detector Formulation

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

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

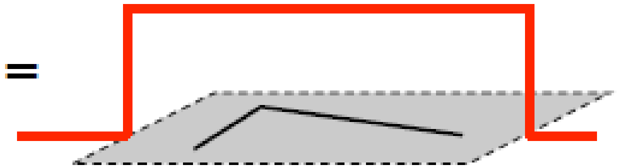


Window function

Shifted intensity

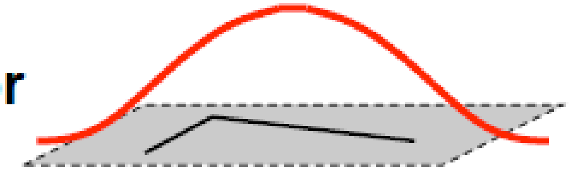
Intensity

Window function $w(x,y)$ =



1 in window, 0 outside

or



Gaussian

Taylor series

$f(x)$ Can be represented at point a in terms of its derivatives

$$f(x) = f(a) + (x-a)f_x + \frac{(x-a)^2}{2!} f_{xx} + \frac{(x-a)^3}{3!} f_{xxx} + \dots$$

Approximating the displaced image as follows

Express $I(x+u, y+v)$ at (x, y):

$$I(x+u, y+v) = I(x, y) + I_x(x+u-x) + I_y(y+v-y)$$

$$I(x+u, y+v) = I(x, y) + I_x u + I_y v$$

Mathematics of Harris detector

$$E(u,v)=\sum_{x,y}[\underbrace{I(x+u,y+v)}_{\text{shifted intensity}}-\underbrace{I(x,y)}_{\text{intensity}}]^2$$

$$E(u,v)=\sum_{x,y}[\underbrace{\cancel{I(x,y)}+uI_x+vI_y}_{\text{shifted intensity}}-\underbrace{\cancel{I(x,y)}}_{\text{intensity}}]^2$$

Taylor Series

$$E(u,v)=\sum_{x,y}[uI_x+vI_y]^2$$

$$E(u,v)=\sum_{x,y}\left[\begin{pmatrix}u & v\end{pmatrix}\begin{pmatrix}I_x\\ I_y\end{pmatrix}\right]^2$$

$$E(u,v)=\sum_{x,y}\begin{pmatrix}u & v\end{pmatrix}\begin{pmatrix}I_x\\ I_y\end{pmatrix}\begin{pmatrix}I_x & I_y\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}$$

$$E(u,v)=\begin{pmatrix}u & v\end{pmatrix}\left[\sum_{x,y}\begin{pmatrix}I_x\\ I_y\end{pmatrix}\begin{pmatrix}I_x & I_y\end{pmatrix}\right]\begin{pmatrix}u\\ v\end{pmatrix}$$

$$M=\sum_{x,y}\begin{bmatrix}I_xI_x & I_xI_y\\ I_xI_y & I_yI_y\end{bmatrix}$$

$$E(u,v)=\begin{pmatrix}u & v\end{pmatrix}M\begin{pmatrix}u\\ v\end{pmatrix}$$

Harris Detector Formulation

- This measure of change can be approximated by:

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

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

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

Gradient with respect to x, times gradient with respect to y

Sum over image region – the area we are checking for corner

$$M = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \; I_y]$$

Harris Detector Formulation

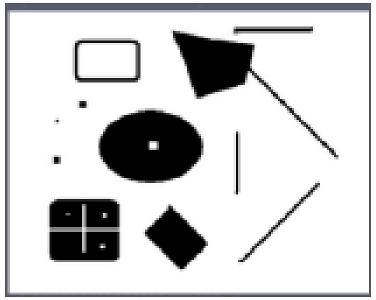
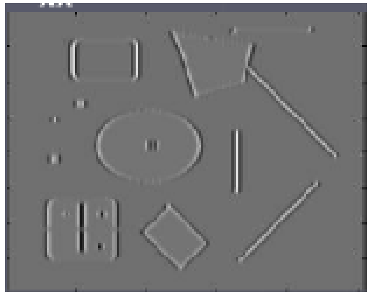
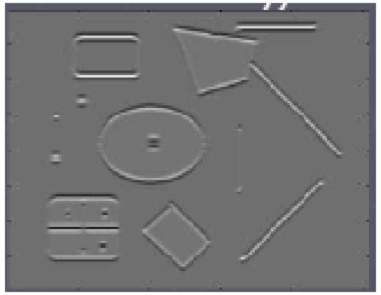


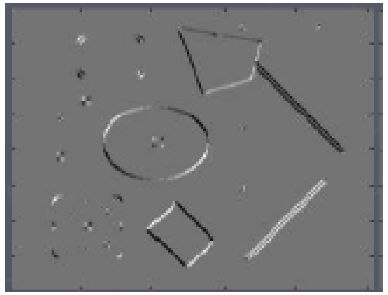
Image I



I_x



I_y



$I_x I_y$

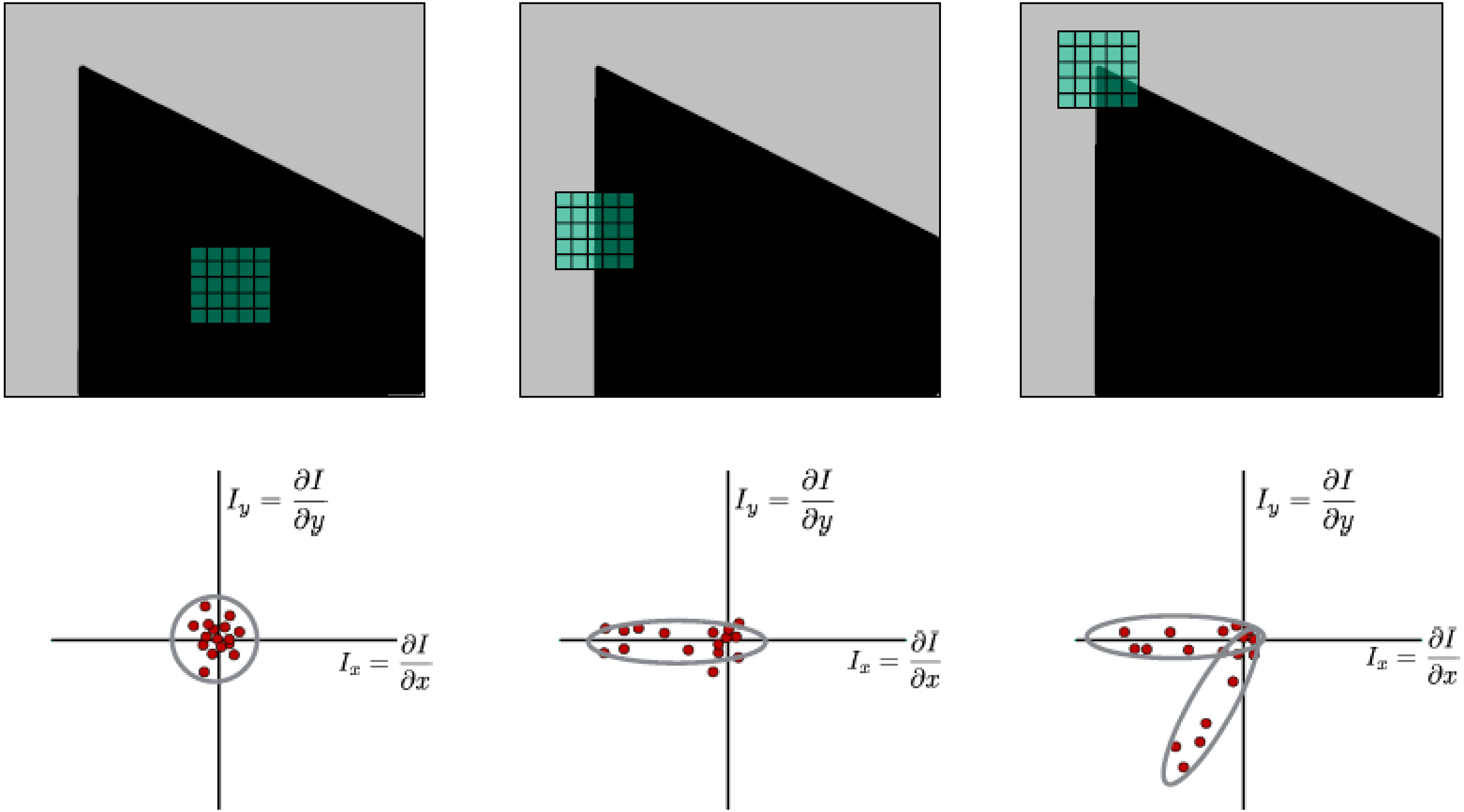
where M is a 2$\boxed{?}$2 matrix computed from image derivatives:

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

Gradient with respect to x, times gradient with respect to y

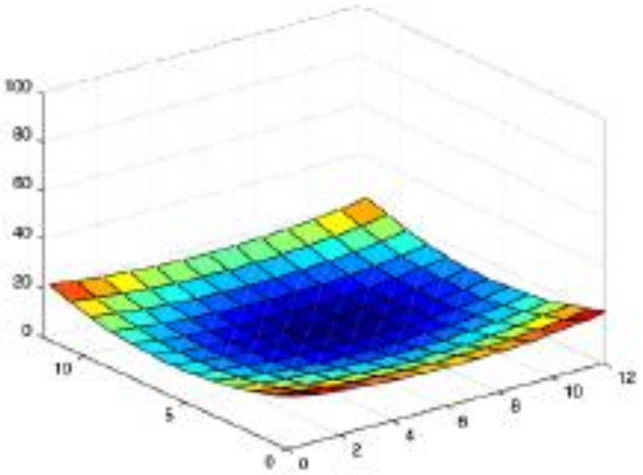
Sum over image region – the area we are checking for corner

$$M = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \ I_y]$$

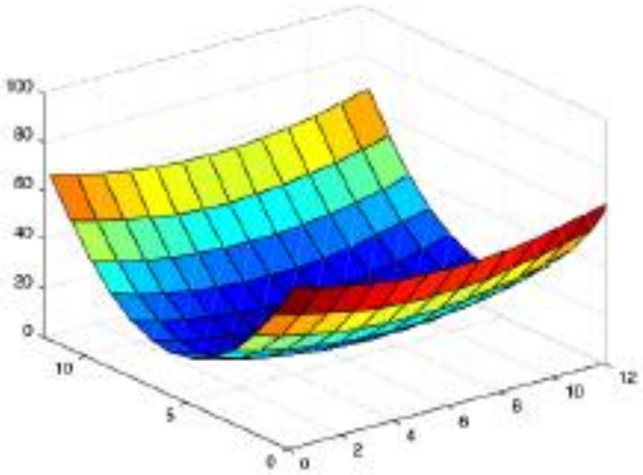

$I_y = \frac{\partial I}{\partial y}$
$I_x = \frac{\partial I}{\partial x}$
$I_y = \frac{\partial I}{\partial y}$
$I_x = \frac{\partial I}{\partial x}$
$I_y = \frac{\partial I}{\partial y}$
$I_x = \frac{\partial I}{\partial x}$

Distribution reveals edge orientation and magnitude

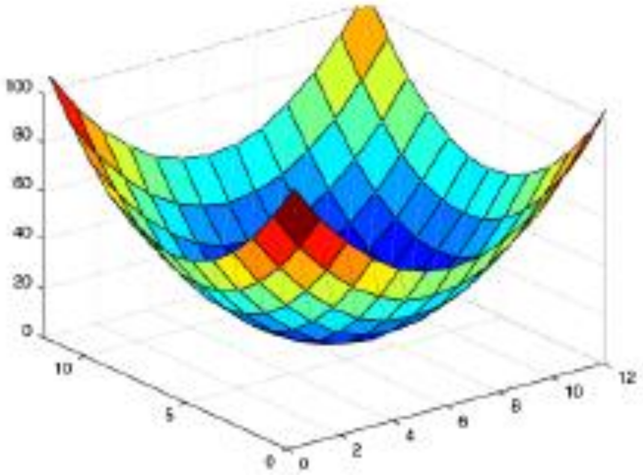
Which error surface indicates a good image feature?



flat

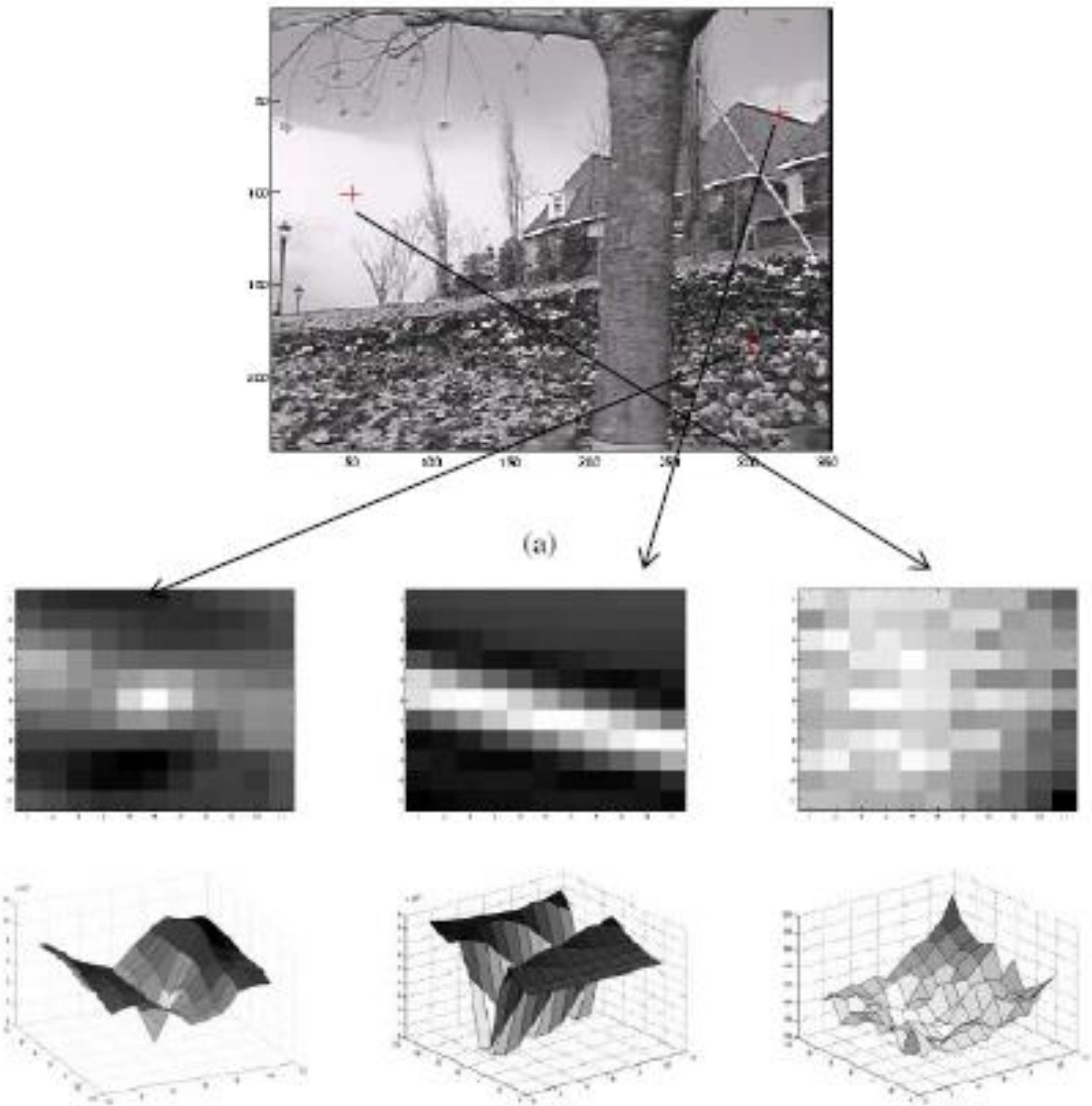


edge
'line'



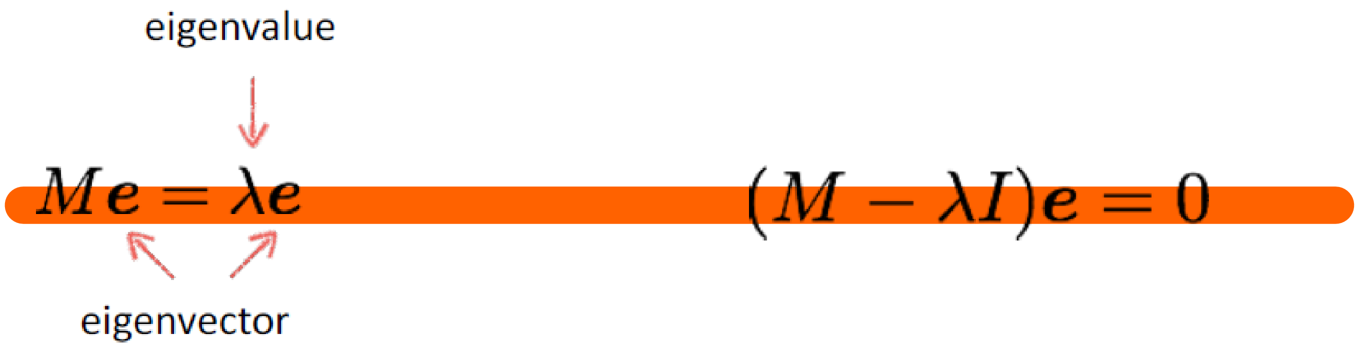
corner
'dot'

What kind of image patch do these surfaces represent?



(a)

Compute eigenvalues and eigenvectors


eigenvalue
$M\boldsymbol{e} = \lambda\boldsymbol{e}$
eigenvector
$(M - \lambda I)\boldsymbol{e} = 0$

1. Compute the determinant of (returns a polynomial) $M - \lambda I$

2. Find the roots of polynomial (returns eigenvalues) $\det(M - \lambda I) = 0$

3. For each eigenvalue, solve (returns eigenvectors) $(M - \lambda I)\boldsymbol{e} = 0$

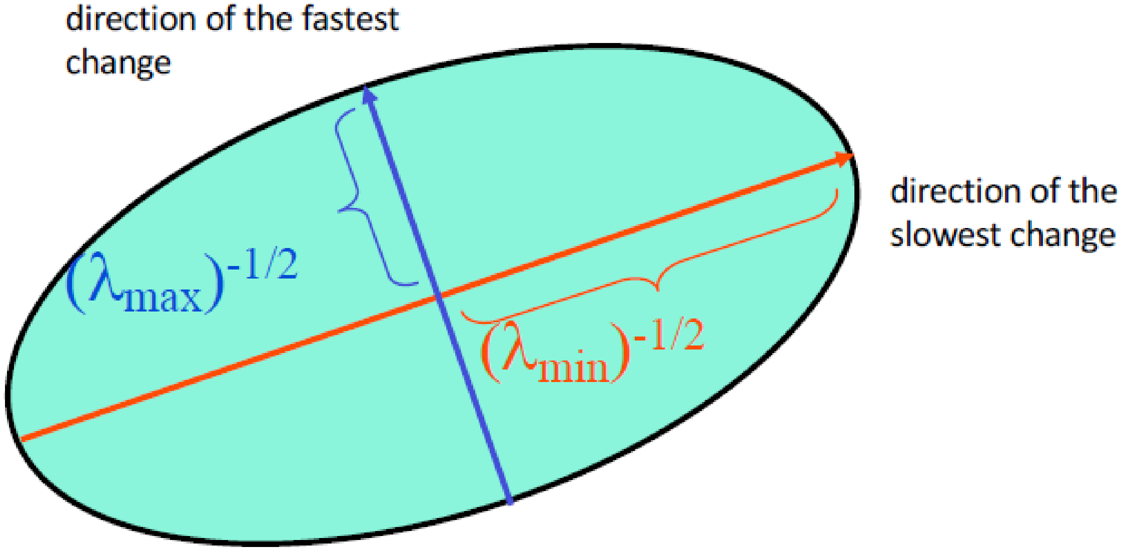
Credit: Kris Kitani, Carnegie Mellon University

General Case

- Since M is symmetric, we have $M = R^{-1}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} R$

  (Eigenvalue decomposition)

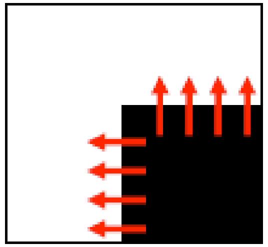
- We can visualize M as an ellipse with axis lengths determined by the eigenvalues and orientation determined by R


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

What Does This Matrix Reveal?

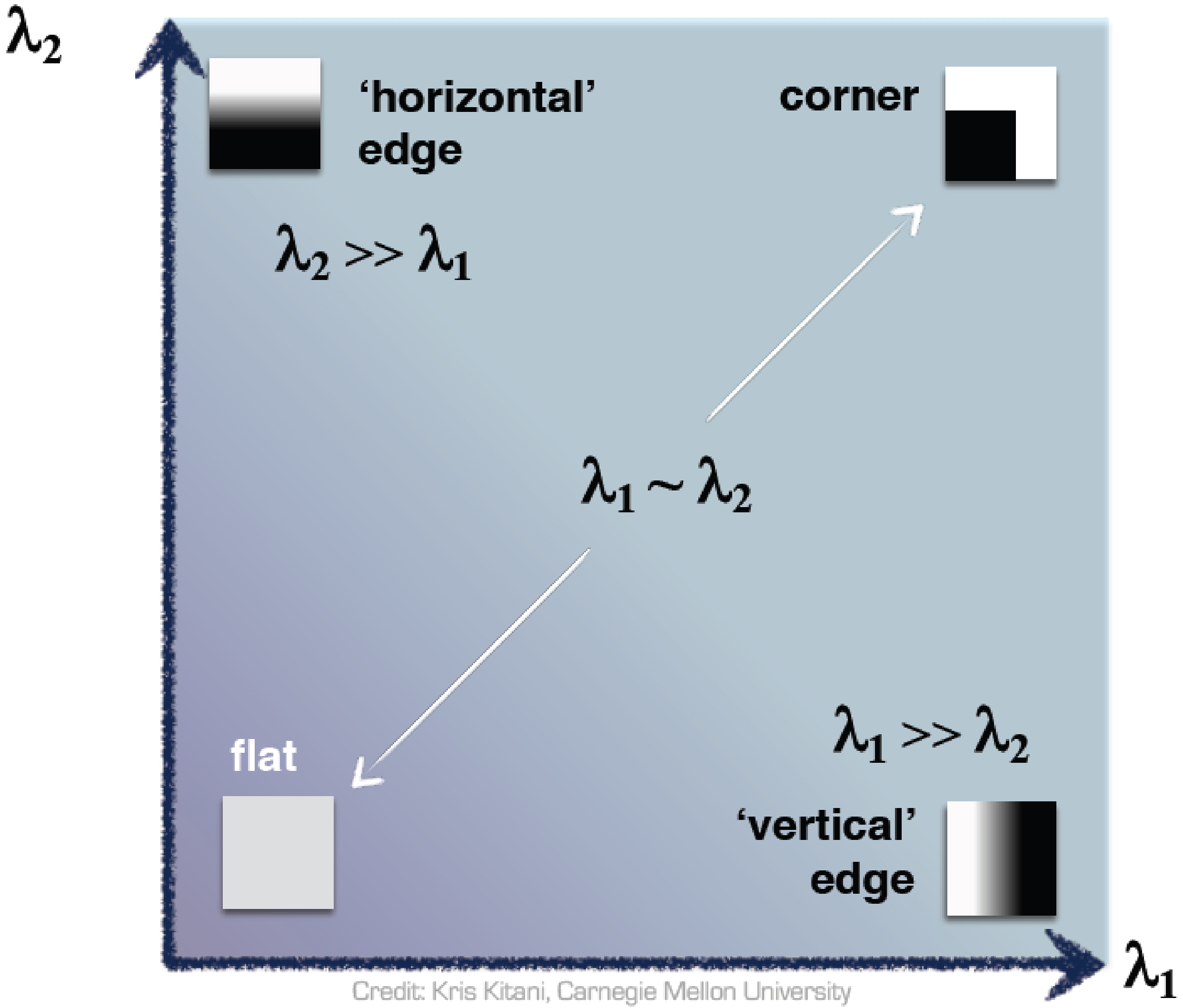
- First, let’s consider an axis-aligned corner:

$$M = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



- This means:
  - Dominant gradient directions align with x or y axis
  - If either λ is close to 0, then this is not a corner, so look for locations where both are large.

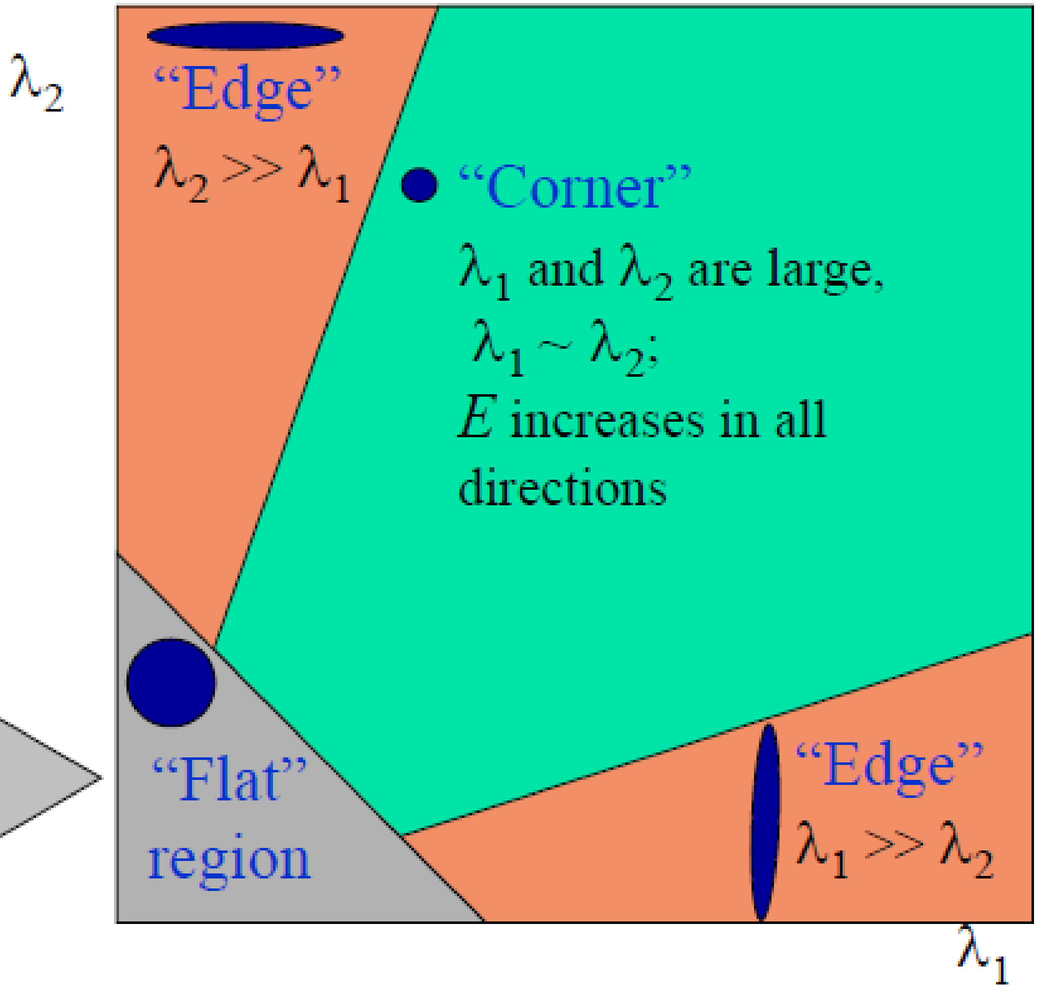
Interpreting eigenvalues



Credit: Kris Kitani, Carnegie Mellon University

Interpreting the Eigenvalues

Classification of image points using eigenvalues of *M*:



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

Harris Corner response function

$$R = \lambda_1\lambda_2 - \alpha(\lambda_1 + \lambda_2)^2 = \det(M) - \alpha\, \mathrm{trace}(M)^2$$

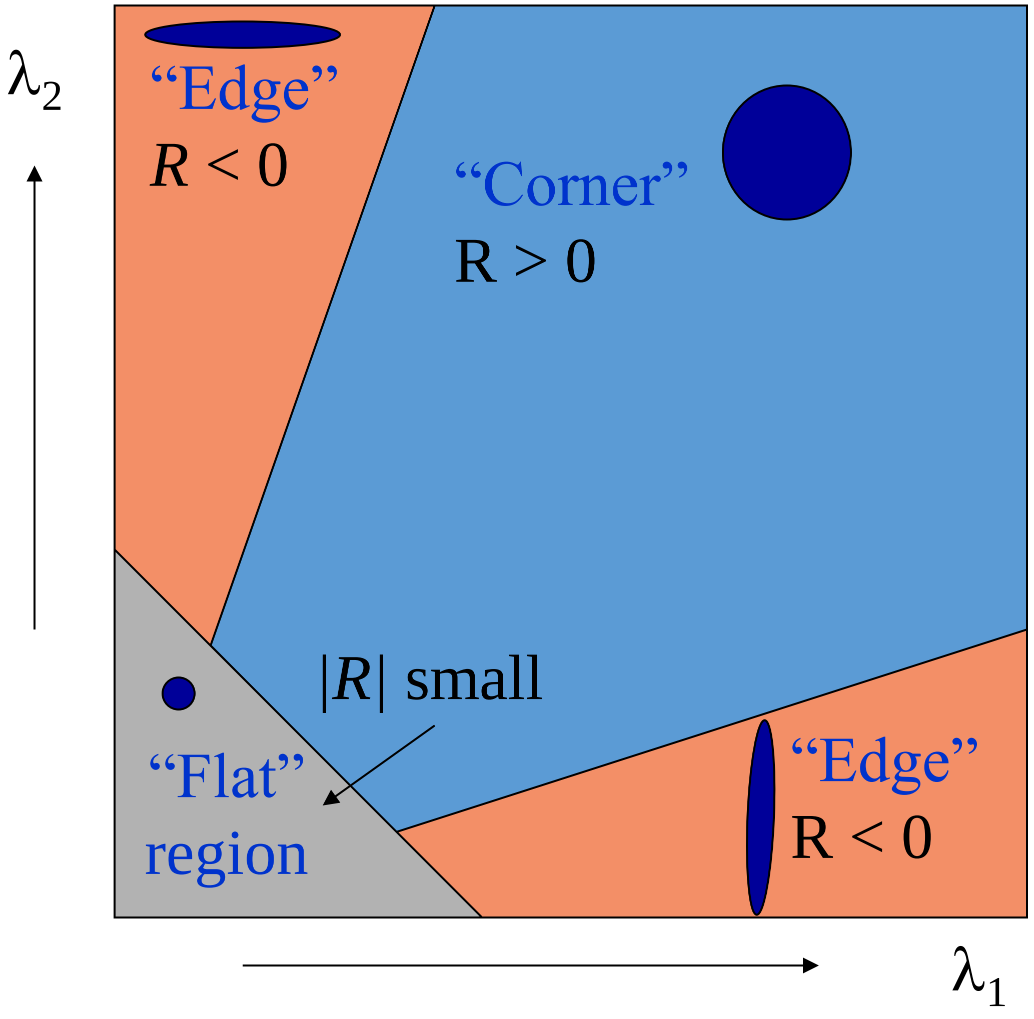
α: constant (0.04 to 0.06)

Determinant (det(A)):

$$|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Trace (trace(A)):

$$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$$



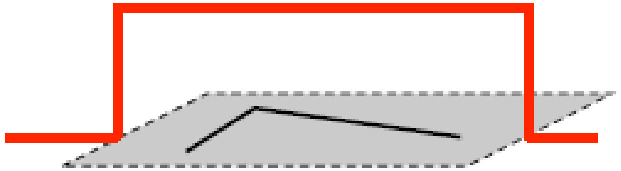
Window Function $w(x,y)$

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

- Option 1: uniform window
  - Sum over square window

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

  - Problem: not rotation invariant

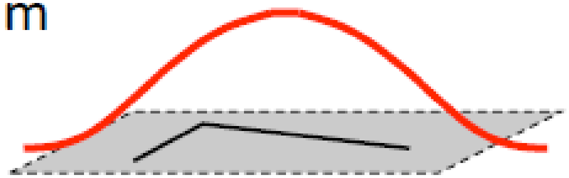


1 in window, 0 outside

- Option 2: Smooth with Gaussian
  - Gaussian already performs weighted sum

$$M = g(\sigma) * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

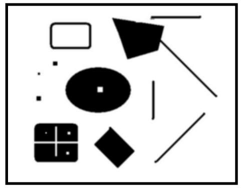
  - Result is rotation invariant



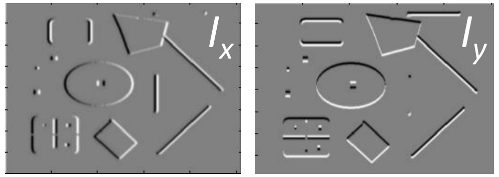
Gaussian

Summary Harris Detector [Harris88]

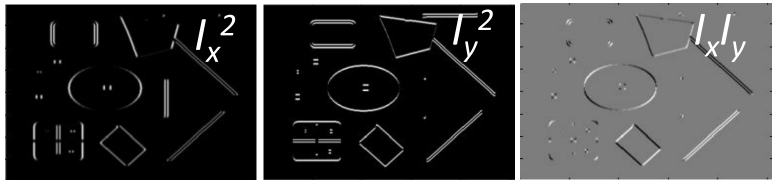
$$M(\sigma_I,\sigma_D) = g(\sigma_I) * \begin{bmatrix} I_x^2(\sigma_D) & I_xI_y(\sigma_D) \\ I_xI_y(\sigma_D) & I_y^2(\sigma_D) \end{bmatrix}$$



1. Image derivatives (optionally, blur first)

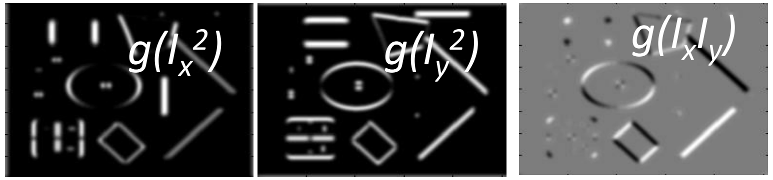


2. Square of derivatives



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

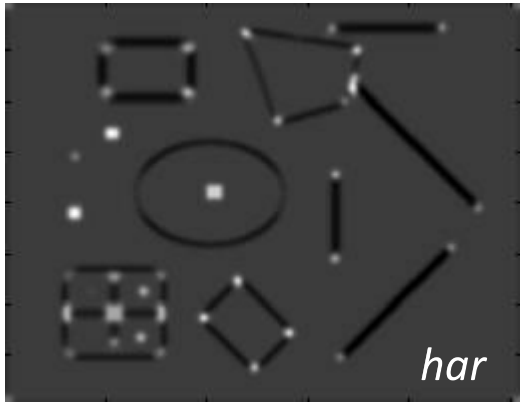
3. Gaussian filter $g(\sigma_I)$



4. Cornerness function – both eigenvalues are strong

$$har = \det[M(\sigma_I,\sigma_D)] - \alpha[\operatorname{trace}(M(\sigma_I,\sigma_D))^2]$$

$$= g(I_x^2)g(I_y^2) - [g(I_xI_y)]^2 - \alpha[g(I_x^2) + g(I_y^2)]^2$$



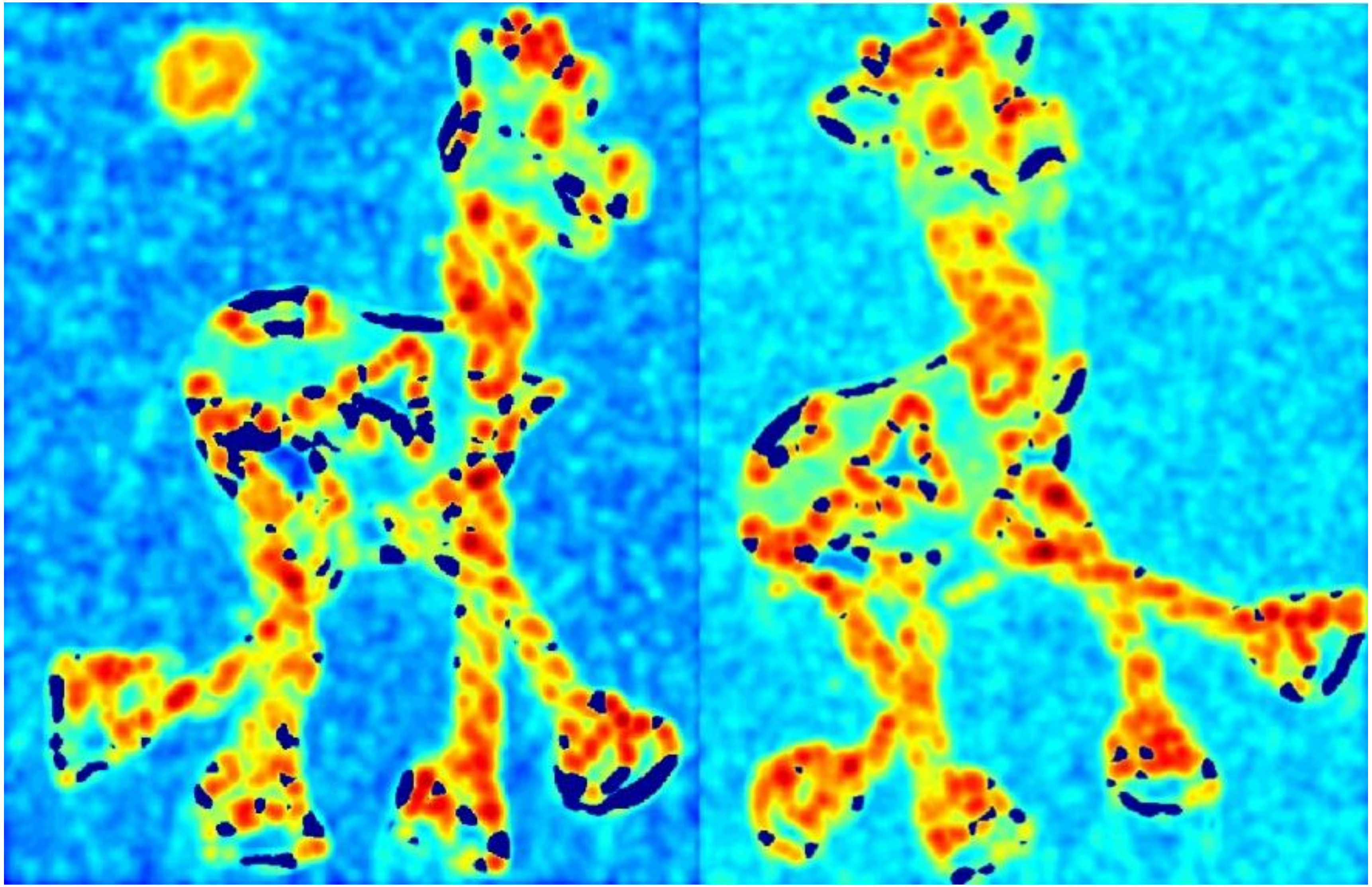
5. Non-maxima suppression

James Hays

Harris Detector: Steps

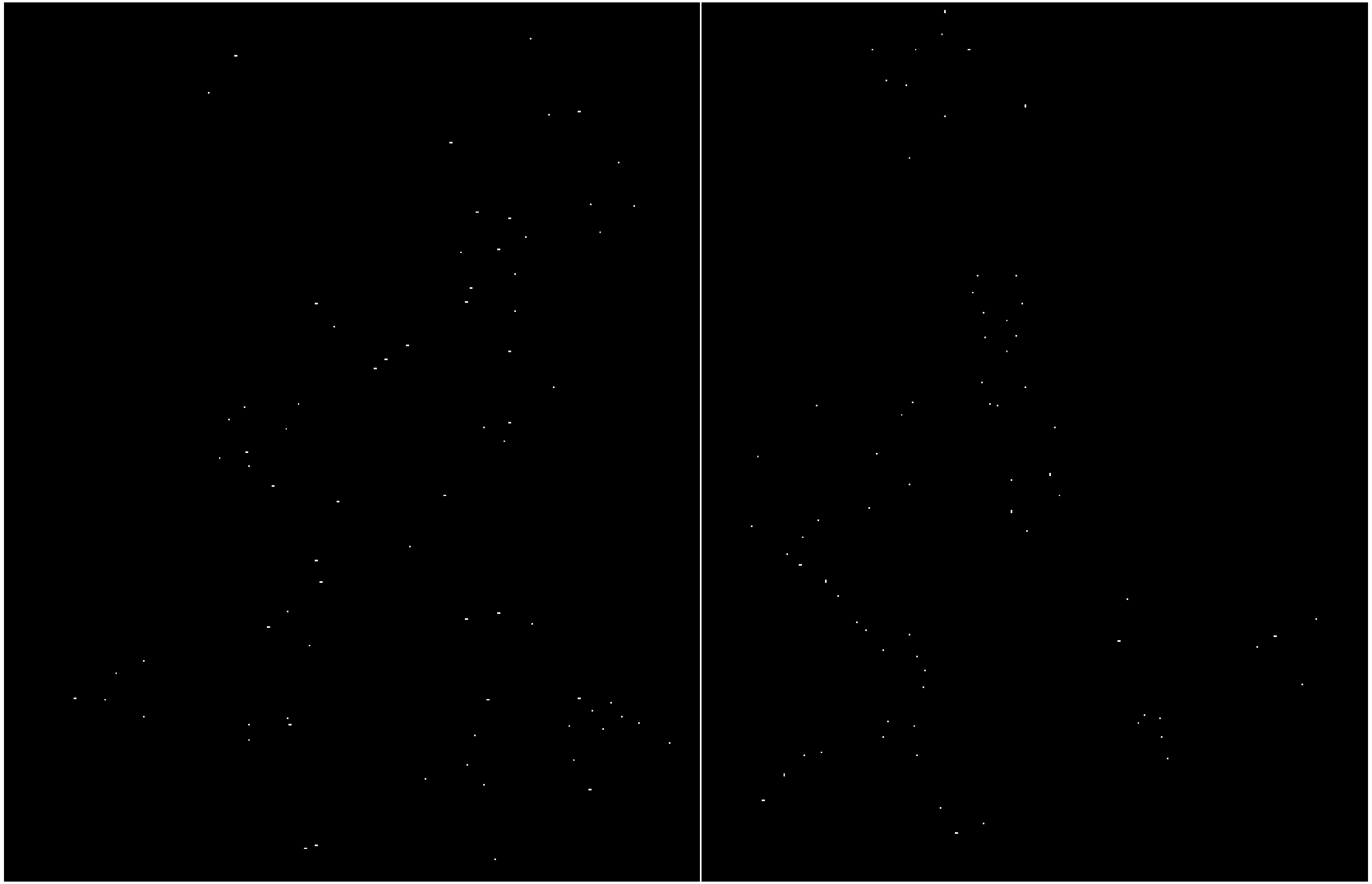
Harris Detector: Steps

Compute corner response R

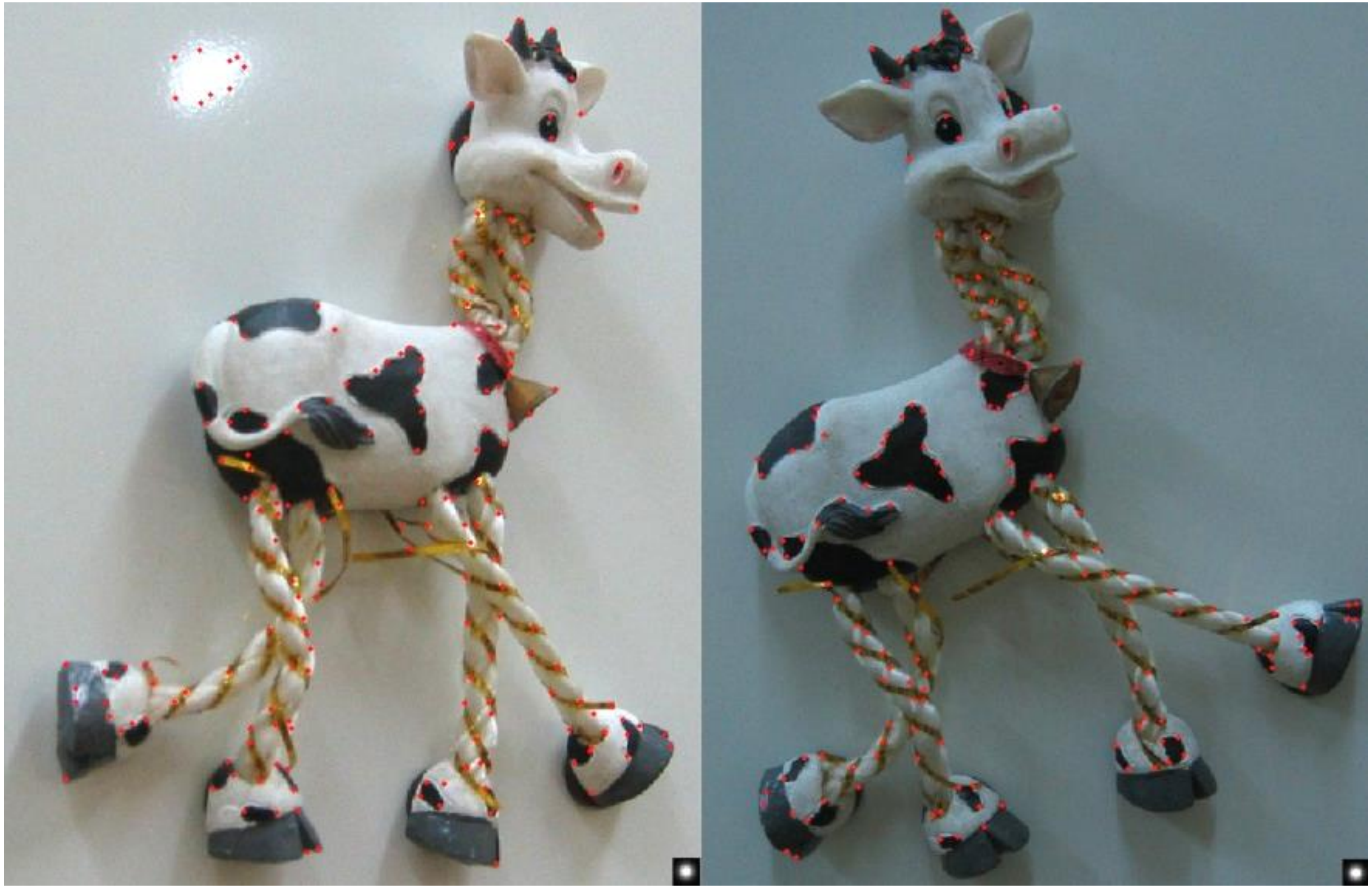


Harris Detector: Steps

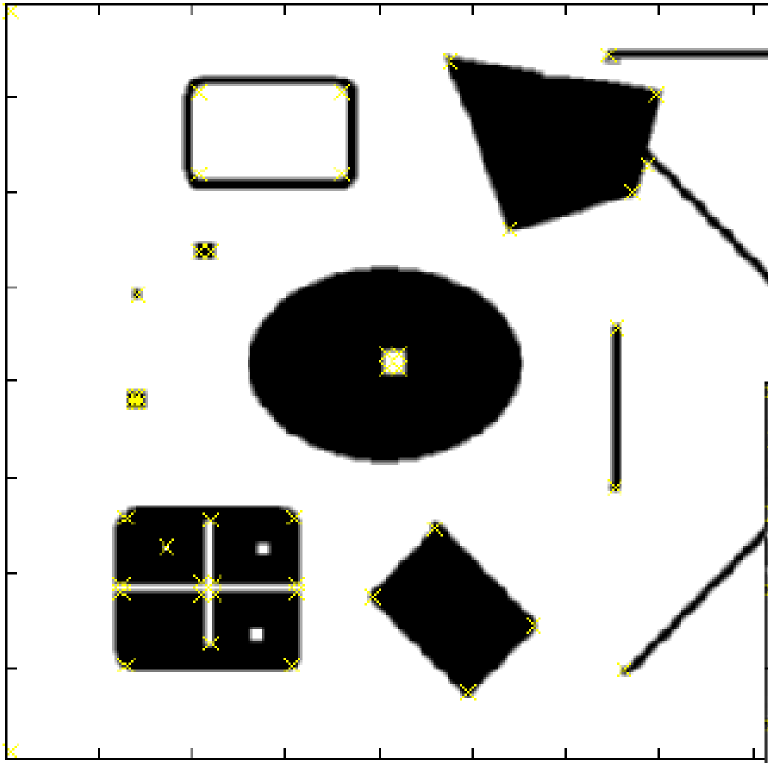
Take only the points of local maxima of *R*

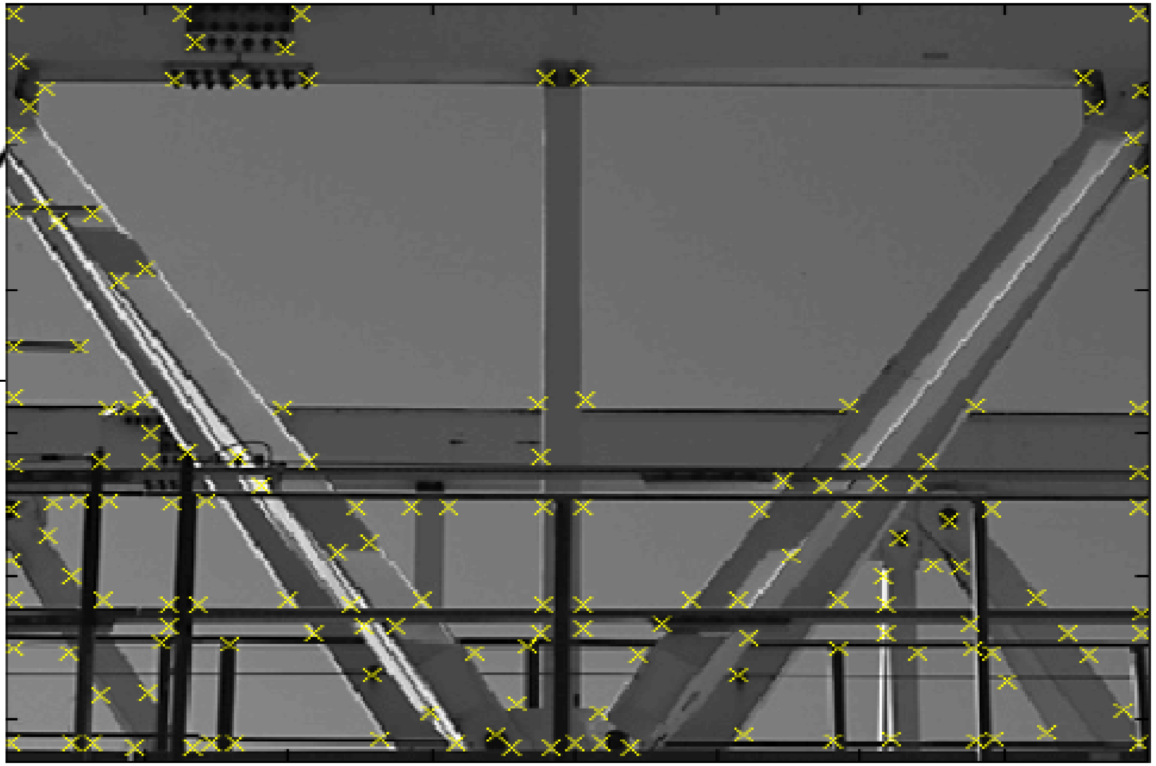


Harris Detector: Steps



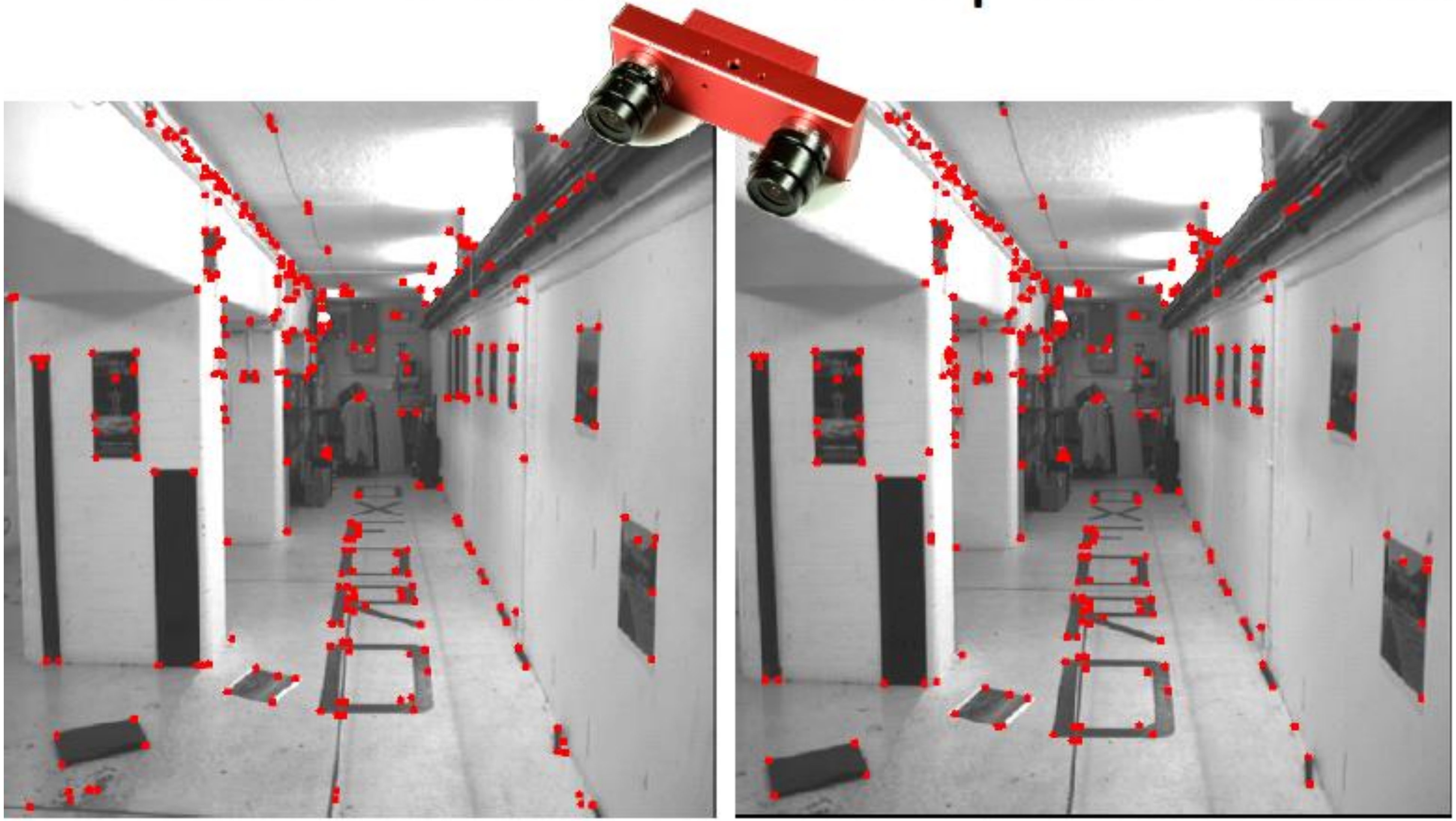
Harris Detector – Responses [Harris88]





Effect: A very precise corner detector.

Harris Detector – Responses [Harris88]



- Results are well suited for finding stereo correspondences

Reading

R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2010.

Section 4.1 Points and Patches

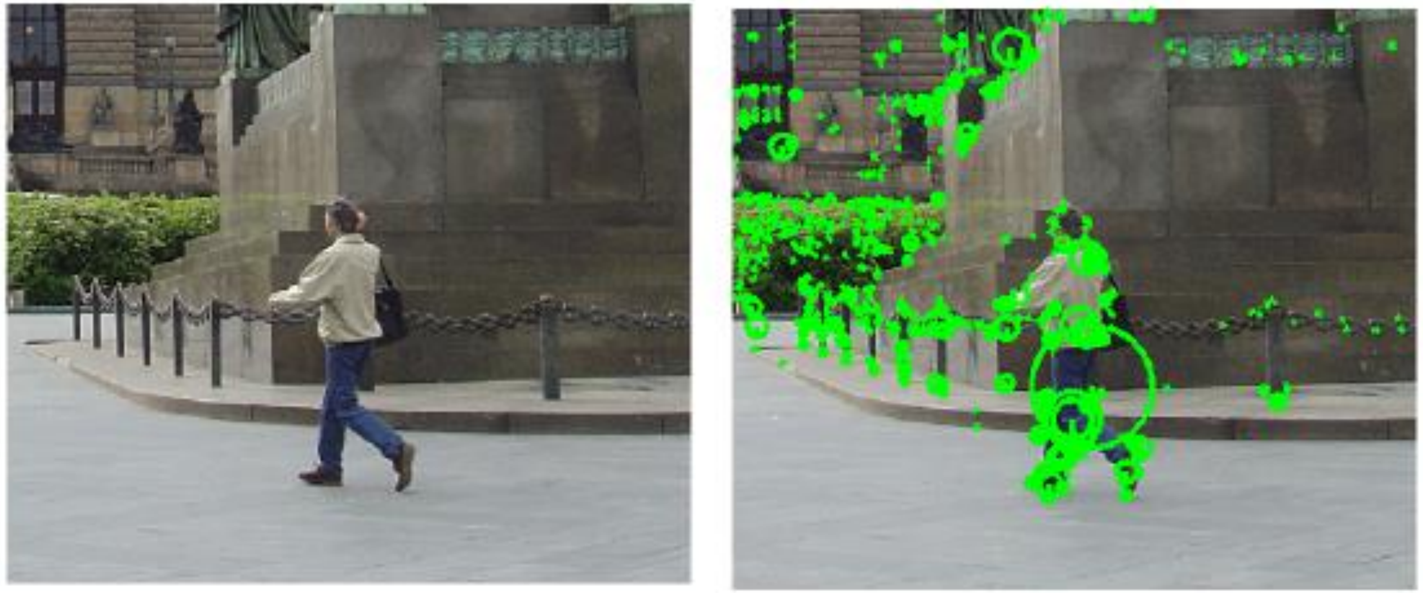
Section 4.1.1 Feature Detectors

Hessian Matrix Detector

- For an Image the Hessian matrix can be expressed as

$$H = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix}$$

$$D = Det(H) = I_{xx} I_{yy} - I_{xy}^2$$



Result using vlfeat open library

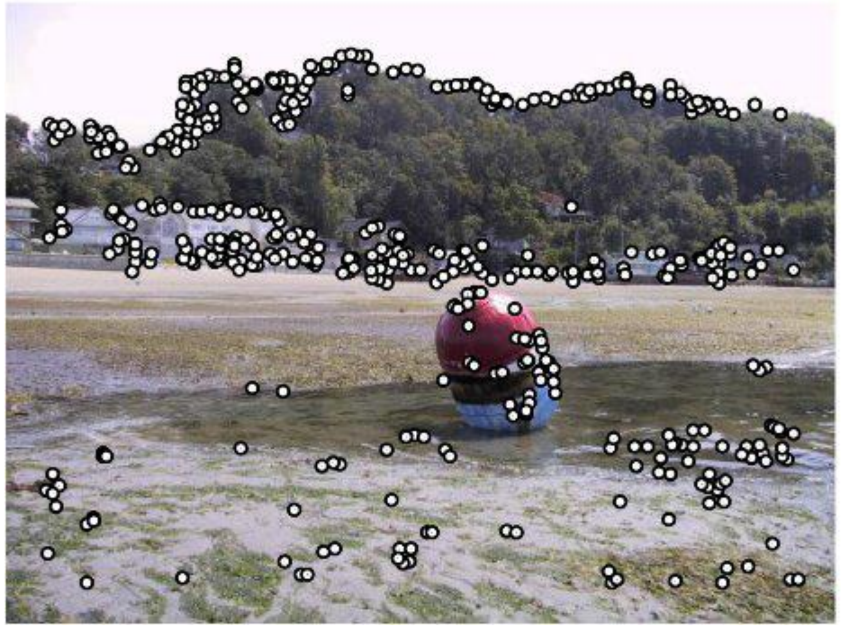
Measuring repeatability

- **Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172.**

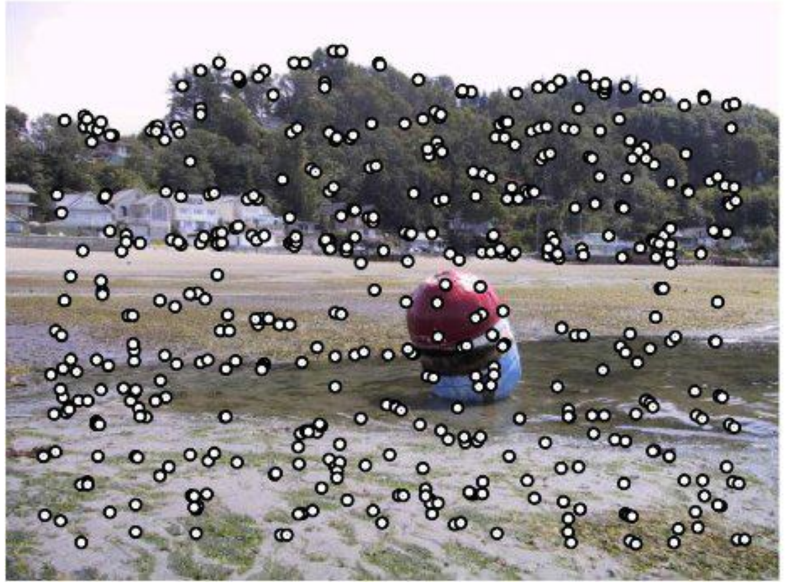
- Frequency with which interest points are detected in one image are to be within ϵ pixels in transformed image
- Measured information content at each detected point
- Reported Harris method works best

Adaptive non-maxima suppression (ANMS)

- Local maxima leads to uneven distribution of feature points.
- Look for local maxima with a response significantly (10%) greater than that of its neighbours within a radius *r.*



Strongest 500

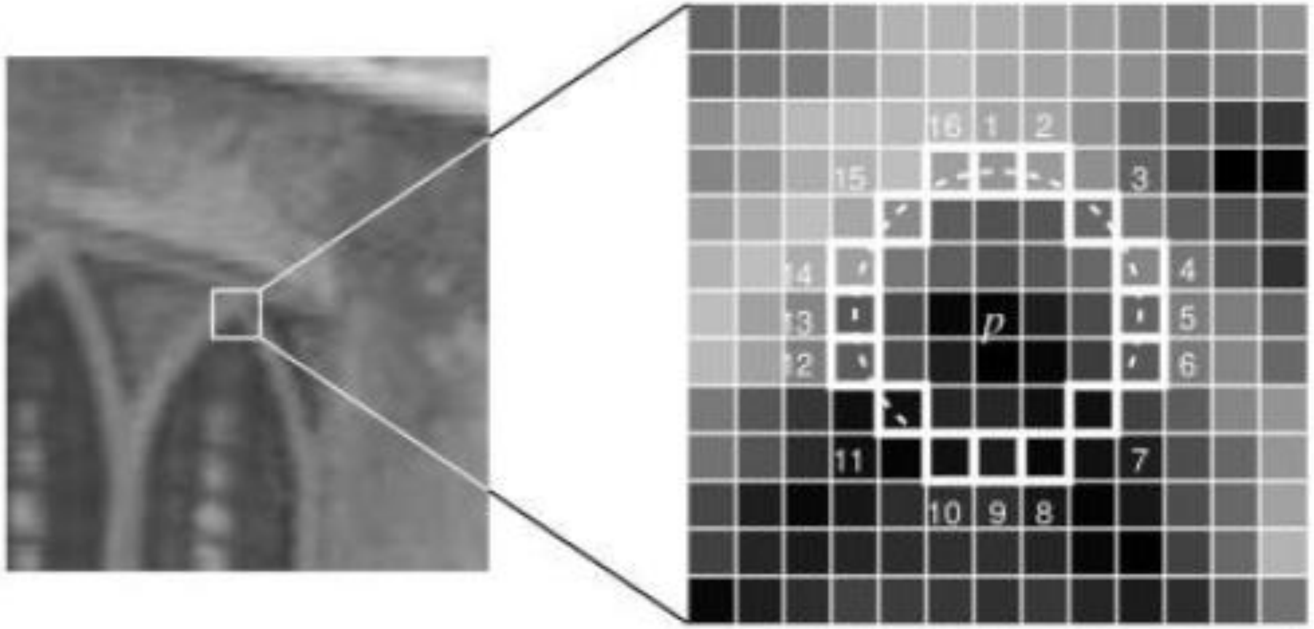


ANMS 500, $r = 16$

M. Brown ; R. Szeliski ; S. Winder, Multi-image matching using multi-scale oriented patches, CVPR 2005

Feature from Accelerated Segment Test (FAST)

Pixel ***p*** is considered as a corner if there are n contiguous pixels in the circle out of 16 pixels, which are all brighter than $I_p + t$, or all darker than $I_p - t$.



Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, pp. 105–119, Jan. 2010.