

Feature Detection

CSE 5005: Computer vision & Image
Processing

a challenging problem



Harder Case

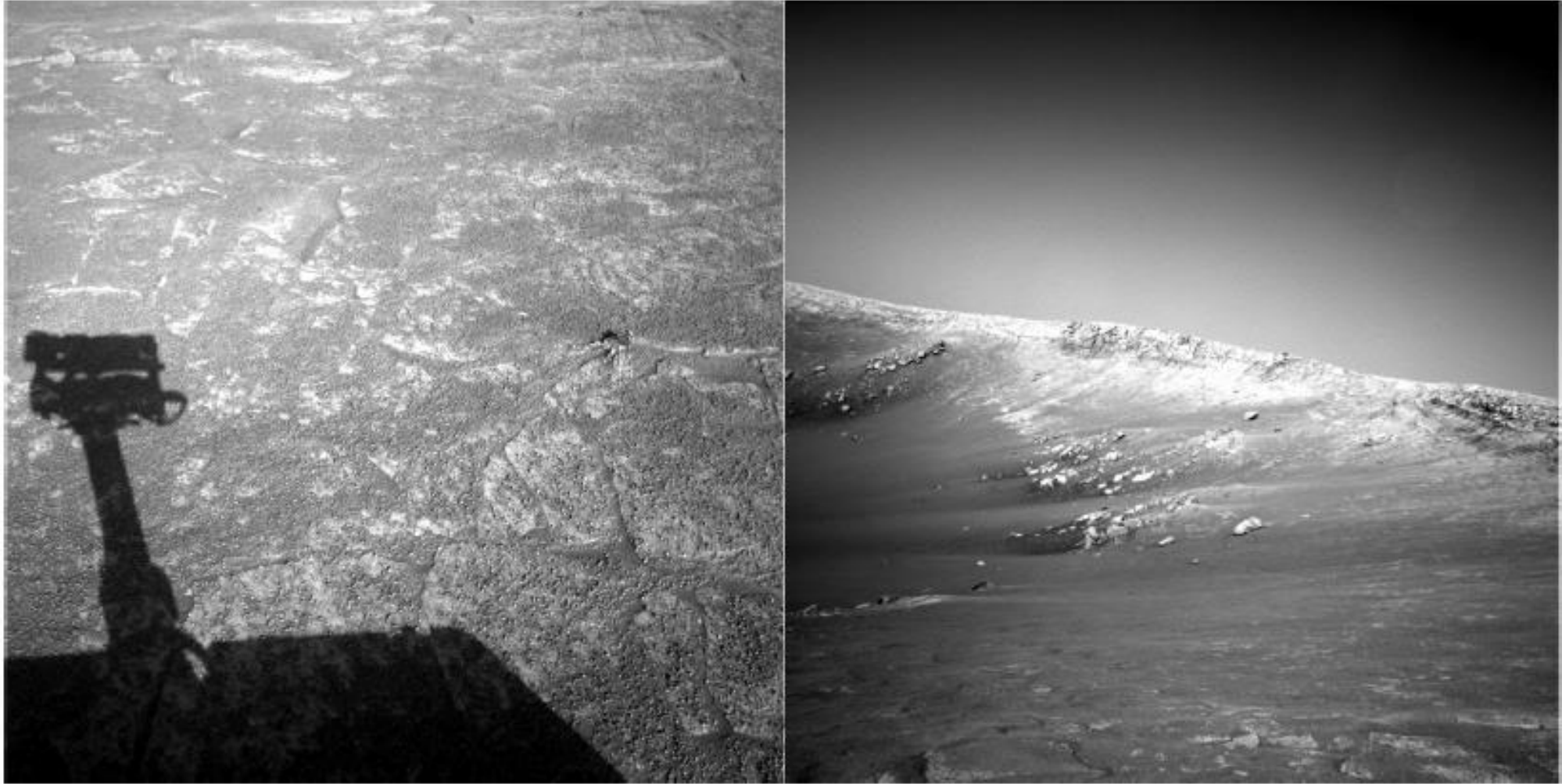


by [Diva Sian](#)



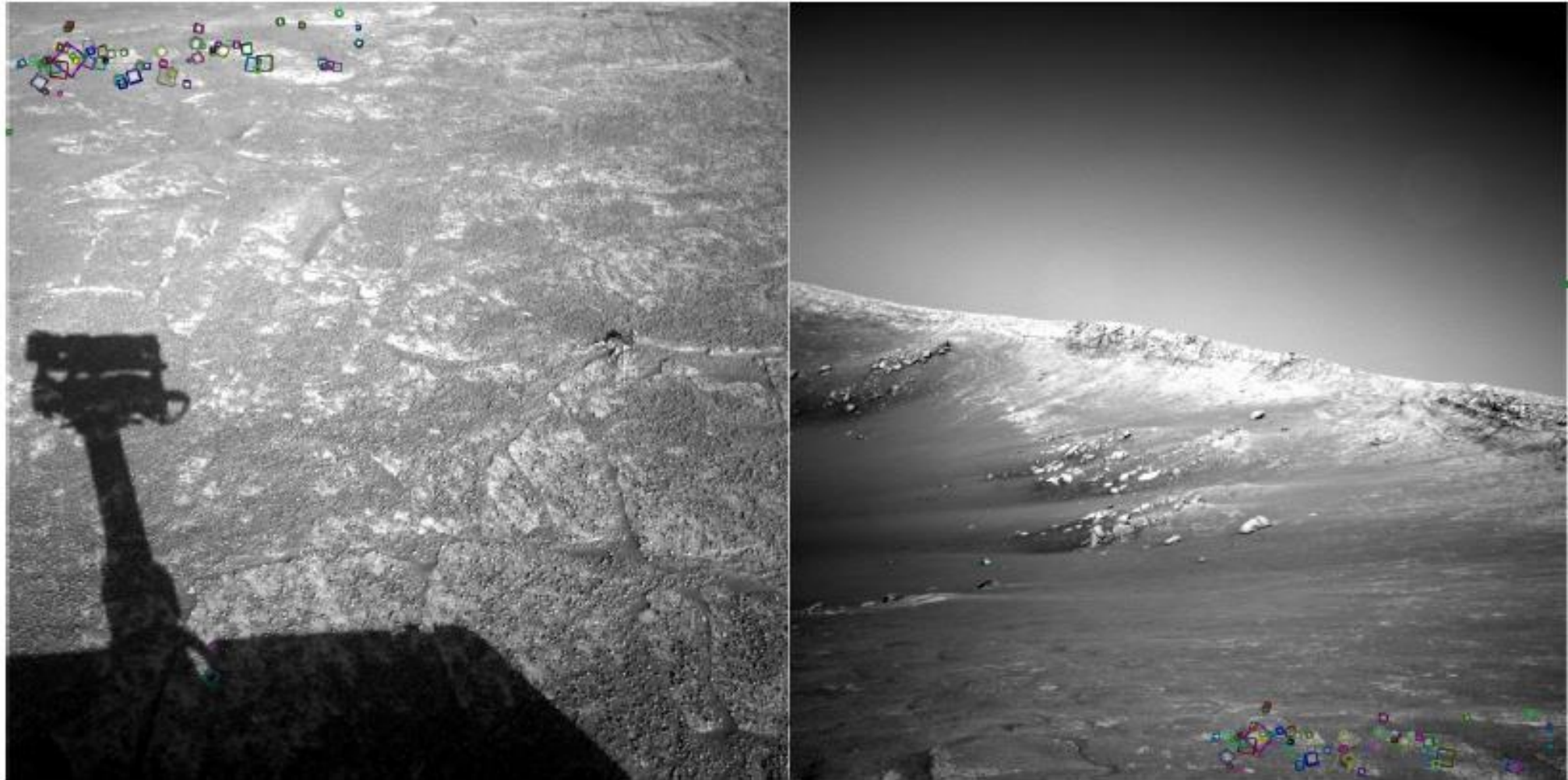
by [scgbt](#)

Harder Still?



NASA Mars Rover images

Answer Below (Look for tiny colored squares)



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

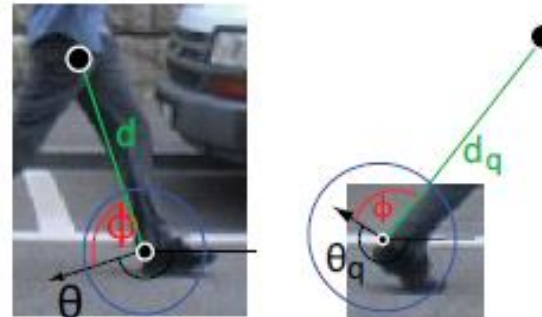
Motivation for using local features

- Global representations have major limitations
- Instead, describe and match only local regions
- Increased robustness to

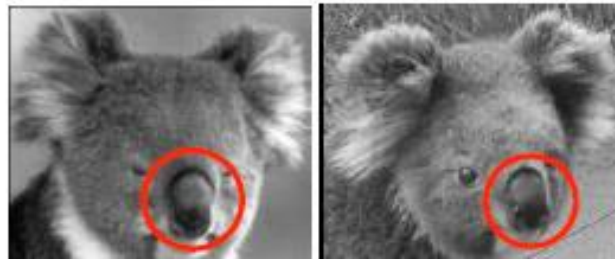
- Occlusions



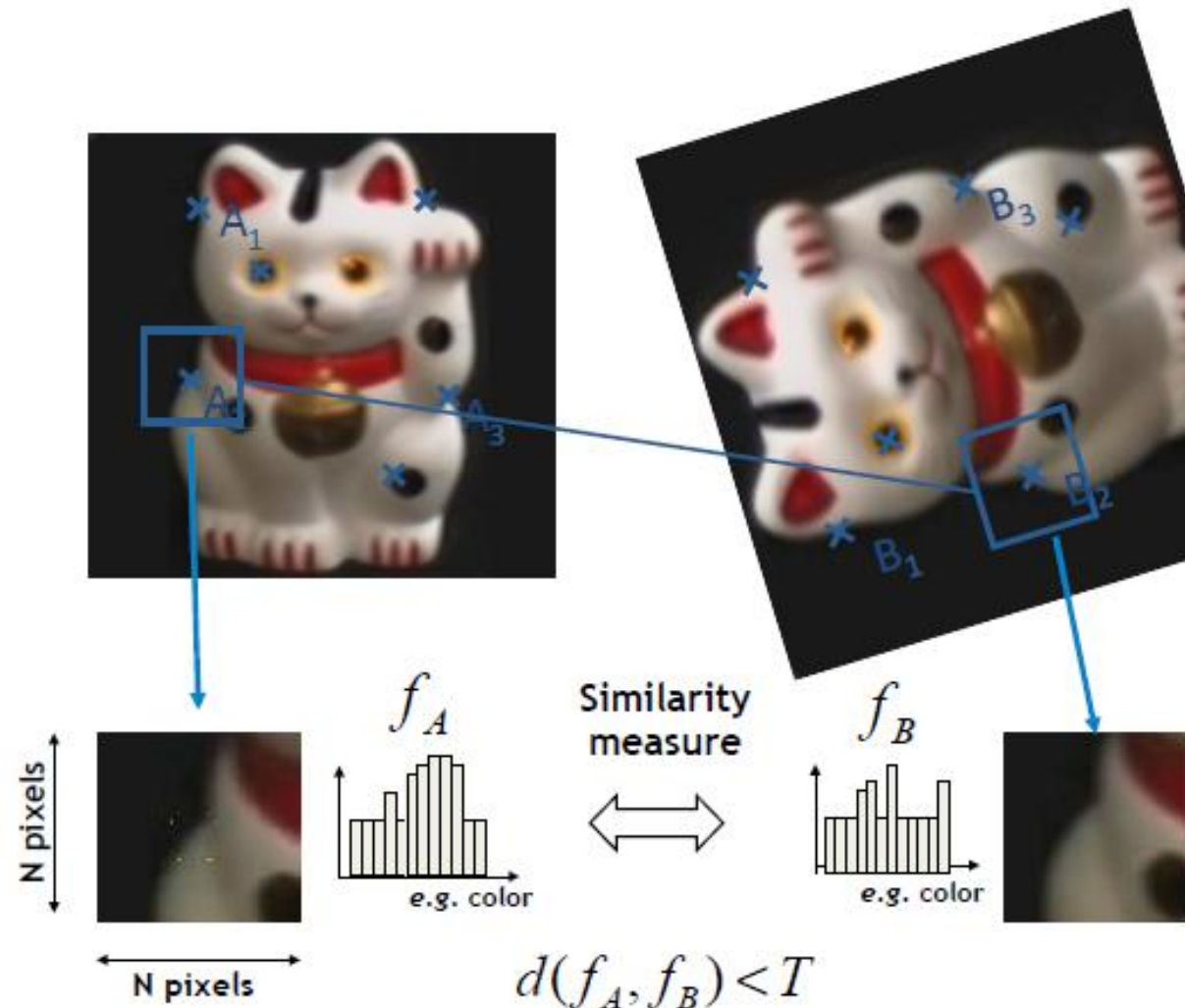
- Articulation



- Intra-category variations



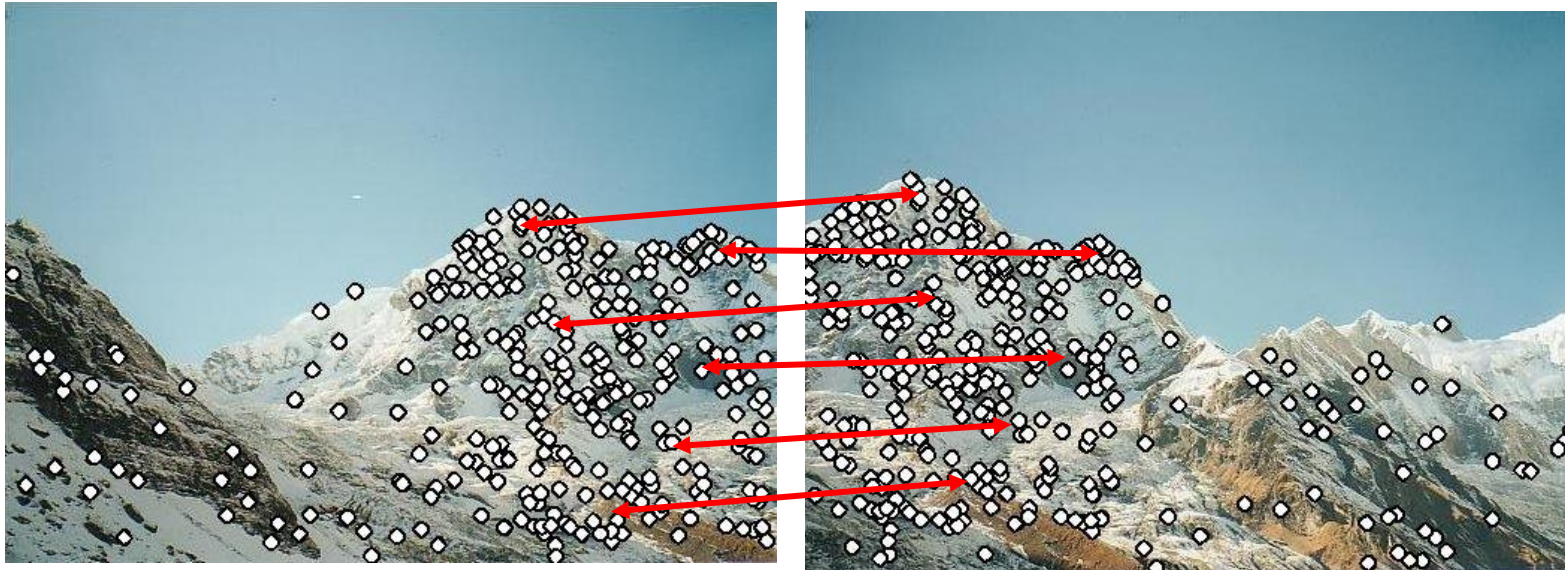
General Approach



1. Find a set of distinctive key-points
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

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

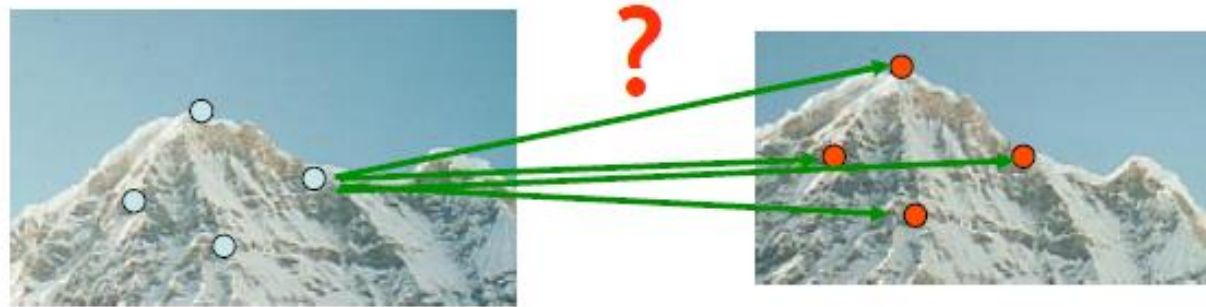


No chance to match!

We need a repeatable detector!

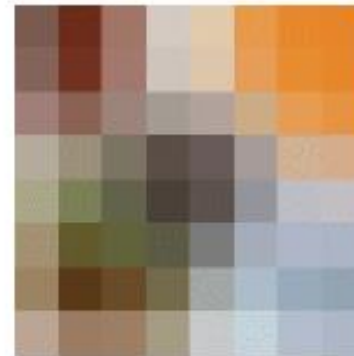
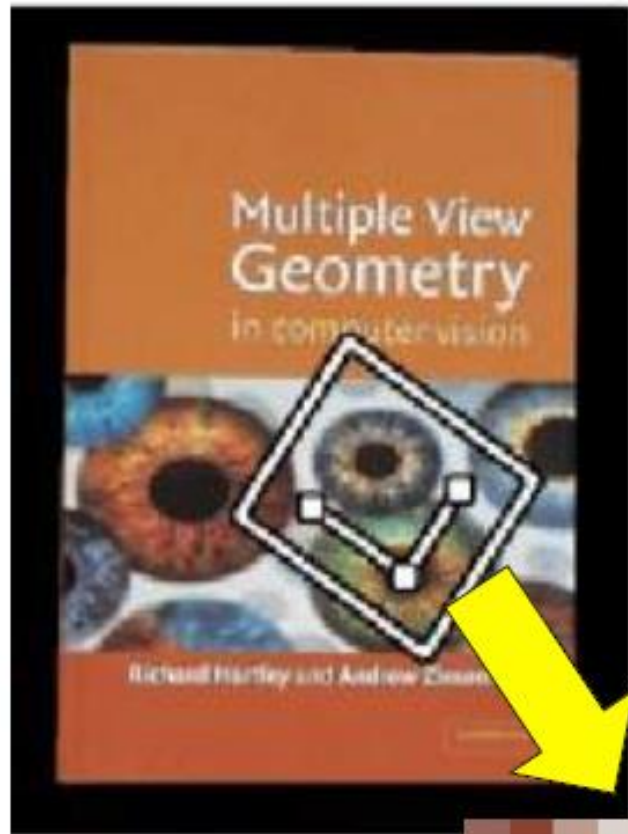
Common Requirements

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

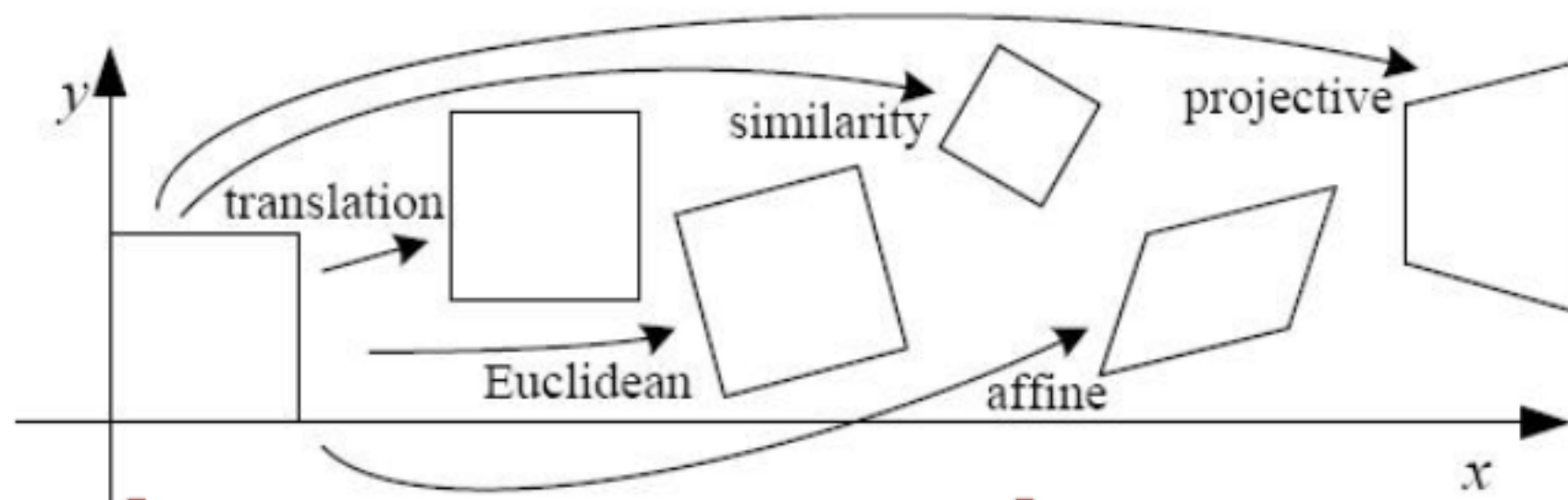


We need a reliable and distinctive descriptor!

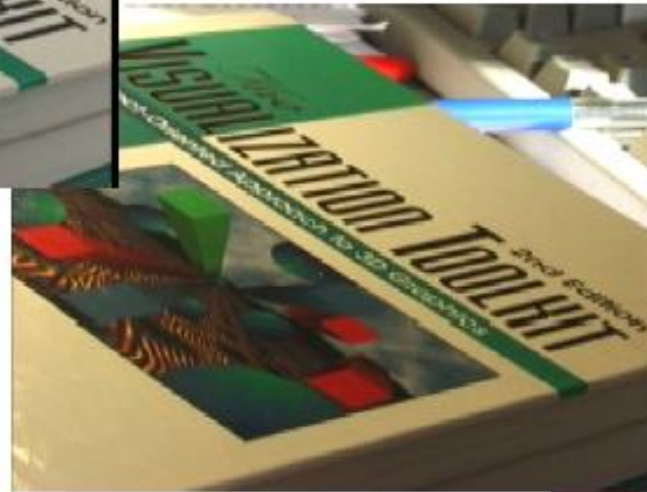
Invariance: Geometric Transformations



Levels of Geometric Invariance



Invariance: Photometric Transformation



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

Requirements

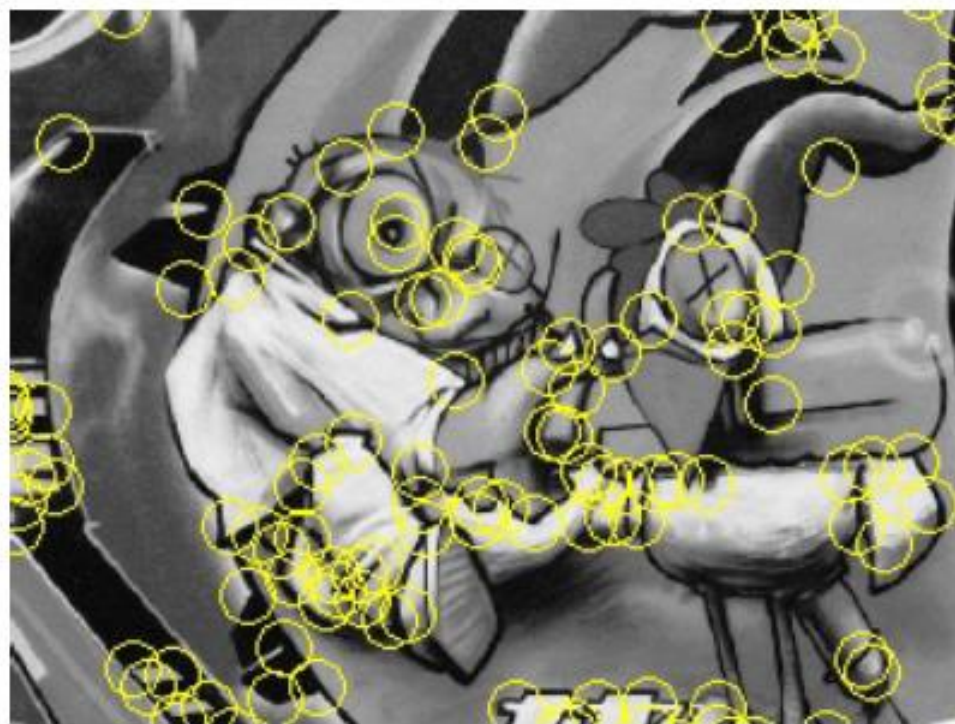
- Region extraction needs to be **repeatable** and **accurate**
 - **Invariant** to translation, rotation, scale changes
 - **Robust** or **covariant** to out-of-plane (\square affine) transformations
 - **Robust** to lighting variations, noise, blur, quantization
- **Locality**: Features are local,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

Applications

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

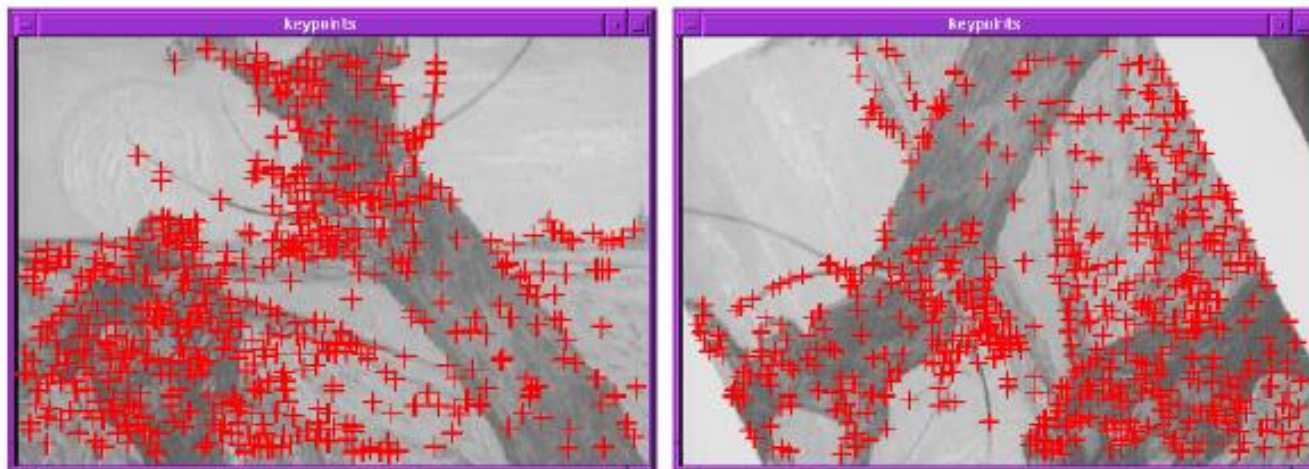


Keypoint Localization



- Goals:
 - Repeatable detection
 - Precise localization
 - Interesting content
- [?]** *Look for two-dimensional signal changes*

Finding Corners

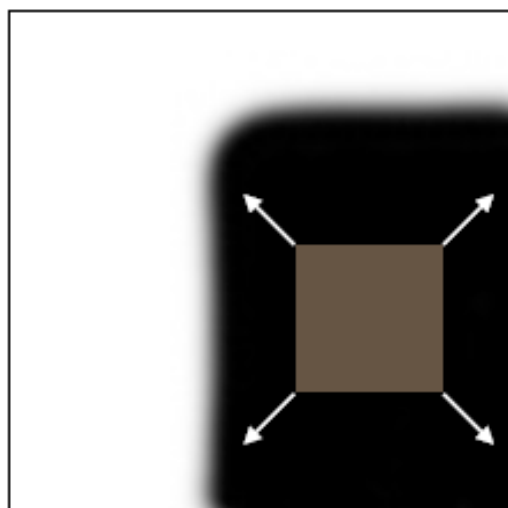


- Key property:
 - In the region around a corner, image gradient has two or more dominant directions
- Corners are *repeatable* and *distinctive*

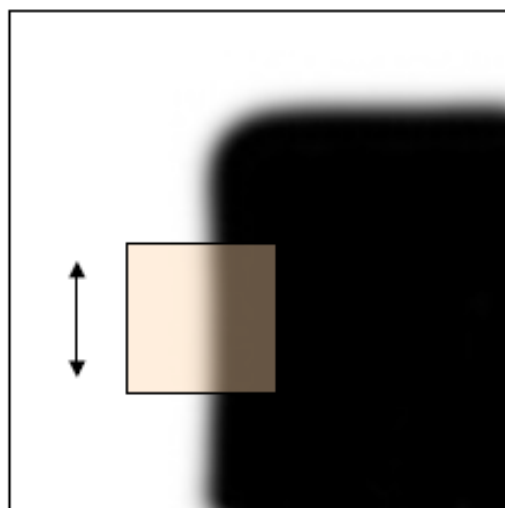
C.Harris and M.Stephens. ["A Combined Corner and Edge Detector."](#)
Proceedings of the 4th Alvey Vision Conference, 1988.

Corners as Distinctive Interest Points

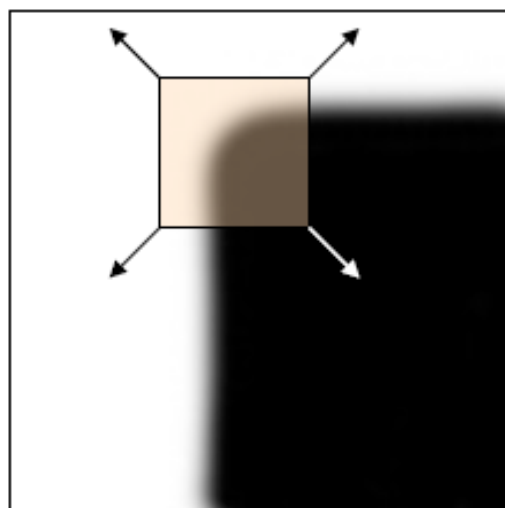
- Design criteria
 - We should easily recognize the point by looking through a small window (*locality*)
 - Shifting the window in *any direction* should give a *large change* in intensity (*good localization*)



“flat” region:
no change in all
directions

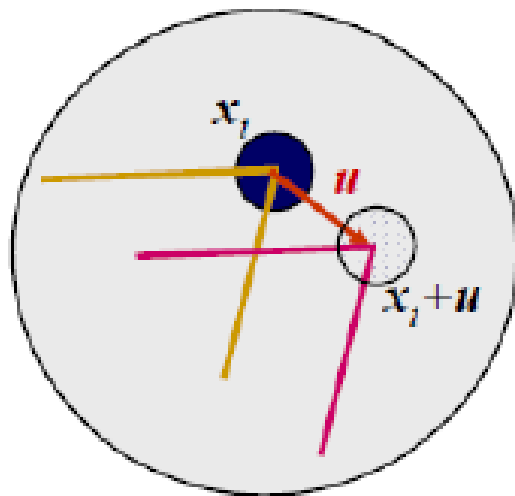


“edge”:
no change along
the edge direction

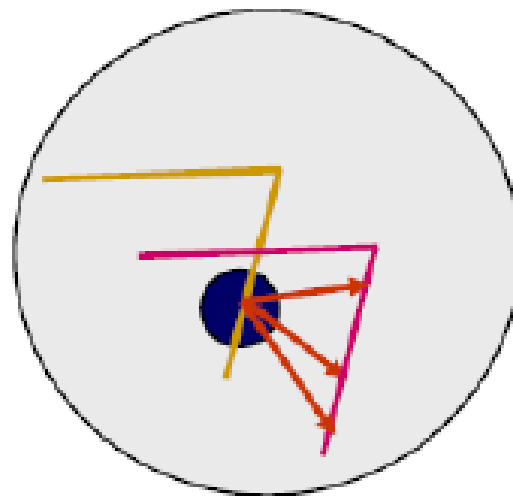


“corner”:
significant change
in all directions

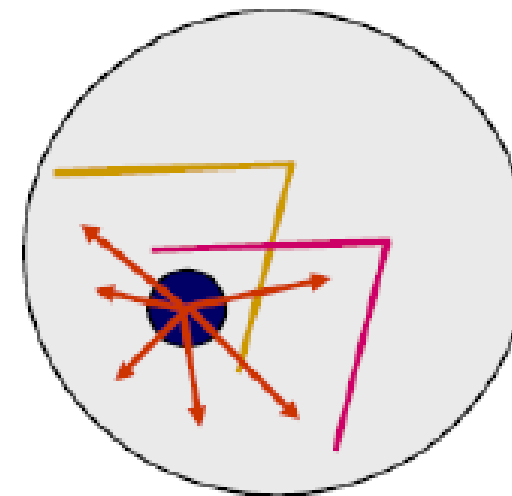
Aperture problem



(a)

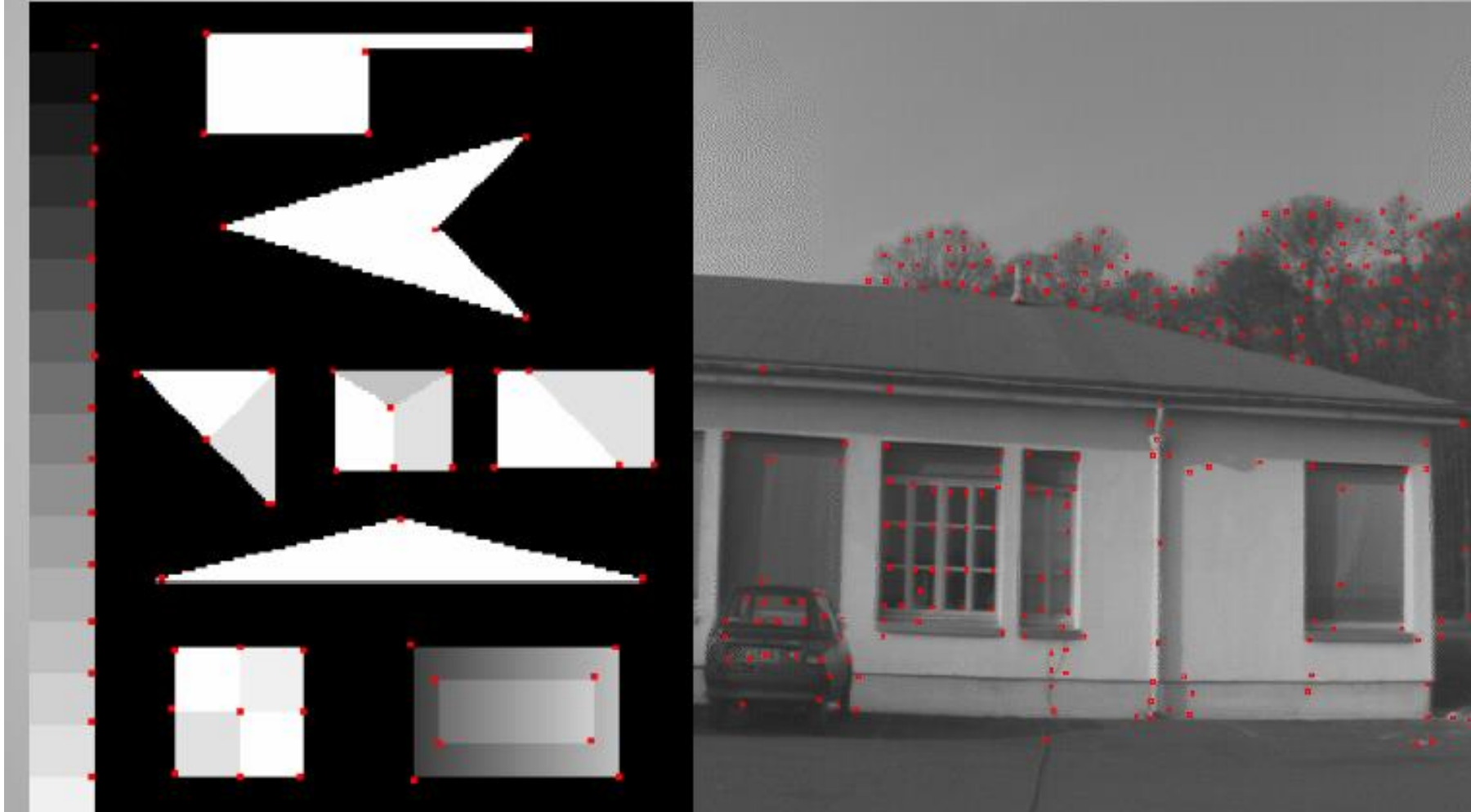


(b)



(c)

What is an interest point?



Harris Detector Formulation

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

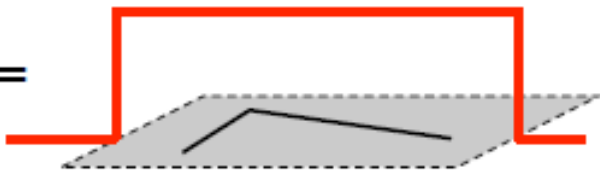
$$E(u,v) = \sum_{x,y} w(x,y) \left[I(x+u, y+v) - I(x,y) \right]^2$$

Window
function

Shifted
intensity

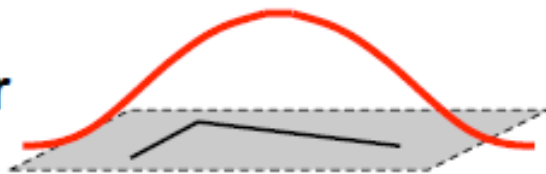
Intensity

Window function $w(x,y) =$




1 in window, 0 outside

or



Gaussian

Correlation


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

f = Image

h = Kernel

$$\begin{array}{|c|c|c|} \hline f_1 & f_2 & f_3 \\ \hline f_4 & f_5 & f_6 \\ \hline f_7 & f_8 & f_9 \\ \hline \end{array} \otimes \begin{array}{|c|c|c|} \hline h_1 & h_2 & h_3 \\ \hline h_4 & h_5 & h_6 \\ \hline h_7 & h_8 & h_9 \\ \hline \end{array} \rightarrow \begin{aligned} f * h &= f_1h_1 + f_2h_2 + f_3h_3 \\ &+ f_4h_4 + f_5h_5 + f_6h_6 \\ &+ f_7h_7 + f_8h_8 + f_9h_9 \end{aligned}$$

Correlation vs SSD

minimize $SSD = \sum_k \sum_l (f(k,l) - h(k,l))^2$ Sum of Squares Difference

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

$SSD = \sum_k \sum_l (-2h(k,l)f(k,l))$ These terms do not depend on correlation

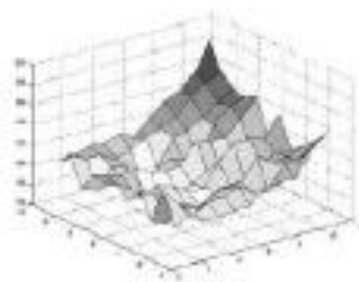
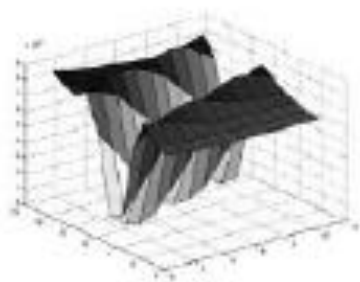
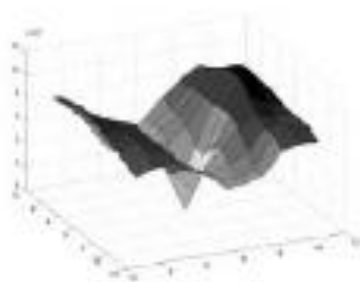
maximize $SSD = \sum_k \sum_l (2h(k,l)f(k,l))$

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

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



(a)



Harris Detector Formulation

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

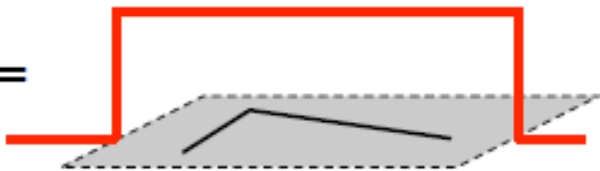
$$E(u,v) = \sum_{x,y} w(x,y) \left[I(x+u, y+v) - I(x,y) \right]^2$$

Window
function

Shifted
intensity

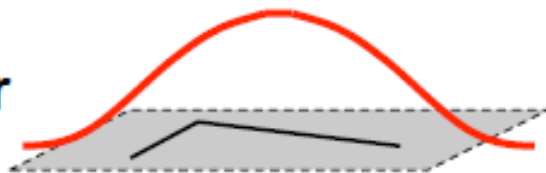
Intensity

Window function $w(x,y) =$



1 in window, 0 outside

or



Gaussian

Taylor series

$\mathcal{A}(x)$ Can be represented at point a in terms of its derivatives

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

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

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

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

Taylor Series

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

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

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

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

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

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

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:

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

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

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

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

Harris Detector Formulation

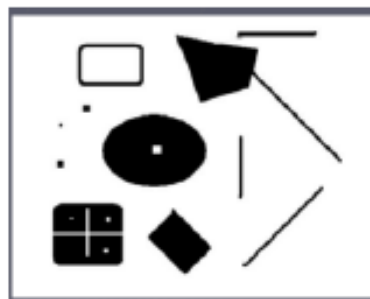


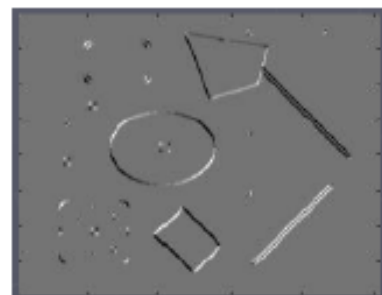
Image I



I_x



I_y



$I_x I_y$

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

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

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

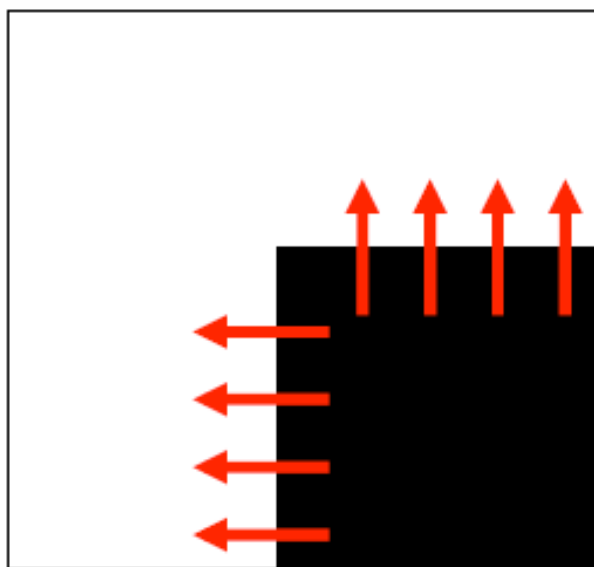
← Gradient with respect to x , times gradient with respect to y

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

What Does This Matrix Reveal?

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

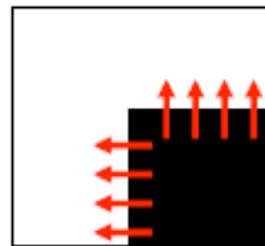
$$M = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



What Does This Matrix Reveal?

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

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

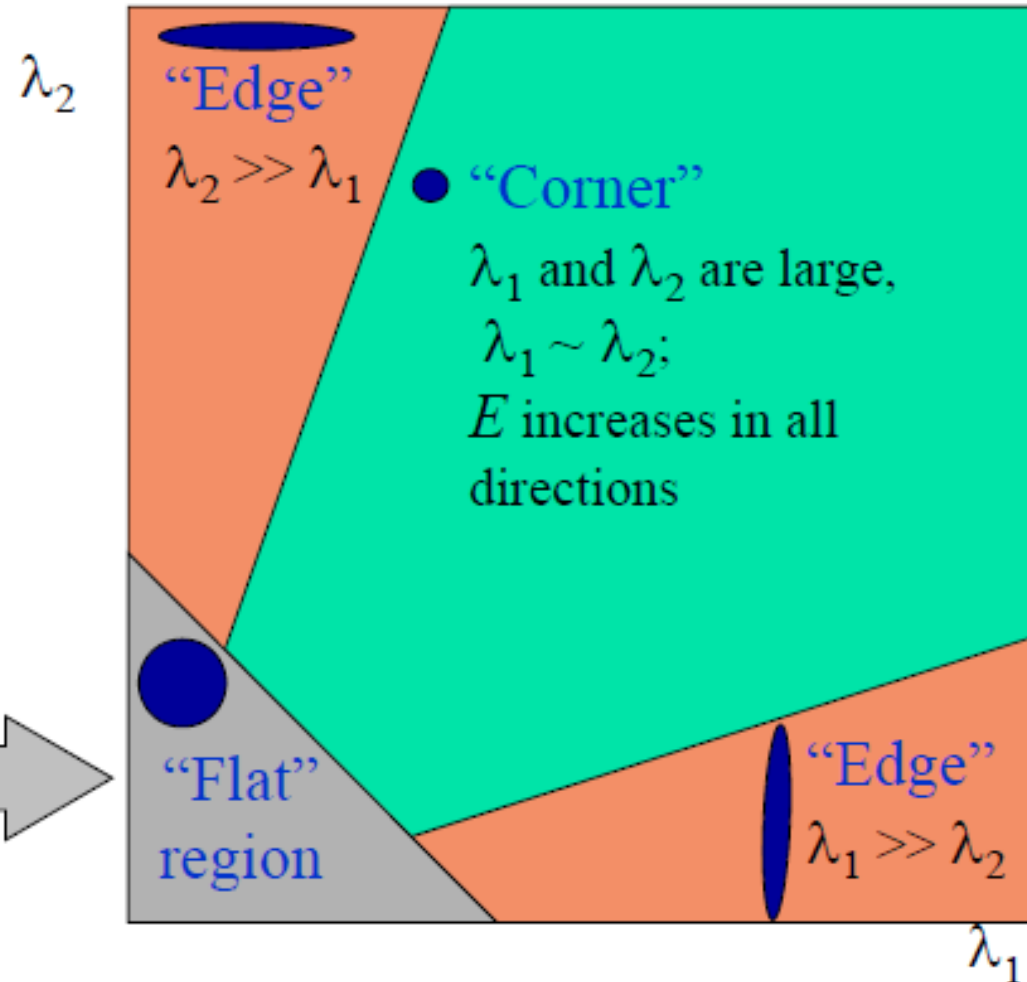


- This means:
 - Dominant gradient directions align with x or y axis
 - If either λ is close to 0, then this is not a corner, so look for locations where both are large.
- What if we have a corner that is not aligned with the image axes?

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

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

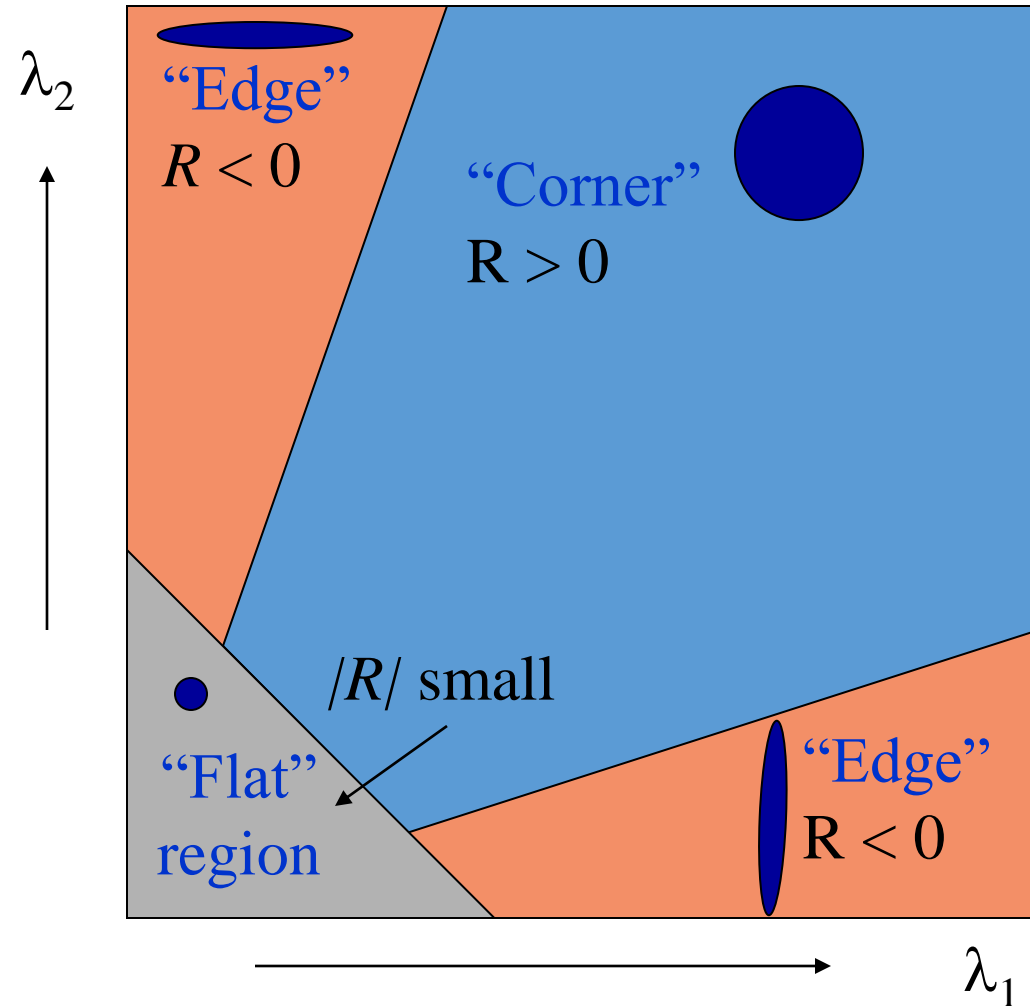
α : constant (0.04 to 0.06)

Determinant ($\det(A)$):

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

Trace ($\text{trace}(A)$):

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



Window Function $w(x,y)$

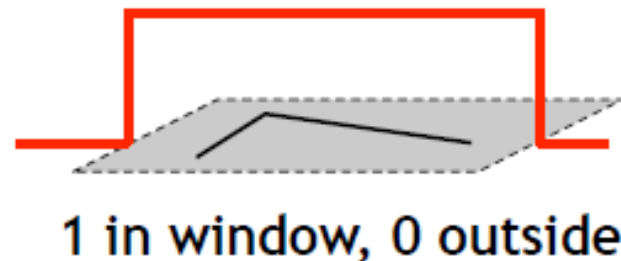
$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- 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

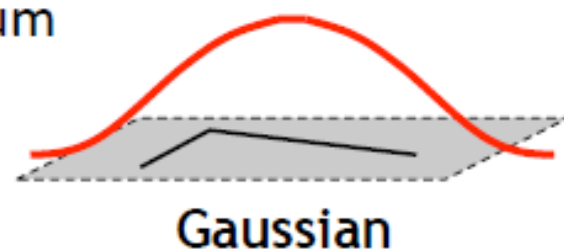


- Option 2: Smooth with Gaussian

- Gaussian already performs weighted sum

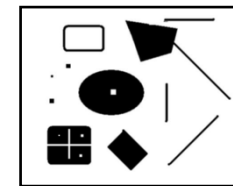
$$M = g(\sigma) * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- Result is rotation invariant

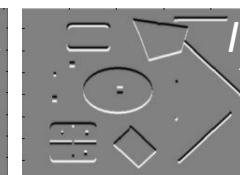


Summary Harris Detector [Harris88]

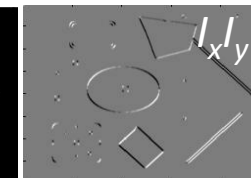
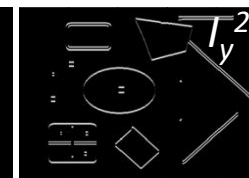
$$M(\sigma_I, \sigma_D) = g(\sigma_I) * \begin{bmatrix} I_x^2(\sigma_D) & I_x I_y(\sigma_D) \\ I_x I_y(\sigma_D) & I_y^2(\sigma_D) \end{bmatrix}$$



1. Image derivatives
(optionally, blur first)



2. Square of derivatives



$$\det M = \lambda_1 \lambda_2$$

$$\text{trace } M = \lambda_1 + \lambda_2$$

3. Gaussian filter $g(\sigma_I)$



4. Cornerness function – both eigenvalues are strong

$$\begin{aligned} har &= \det[M(\sigma_I, \sigma_D)] - \alpha [\text{trace}(M(\sigma_I, \sigma_D))]^2 \\ &= g(I_x^2)g(I_y^2) - [g(I_x I_y)]^2 - \alpha [g(I_x^2) + g(I_y^2)]^2 \end{aligned}$$

5. Non-maxima suppression

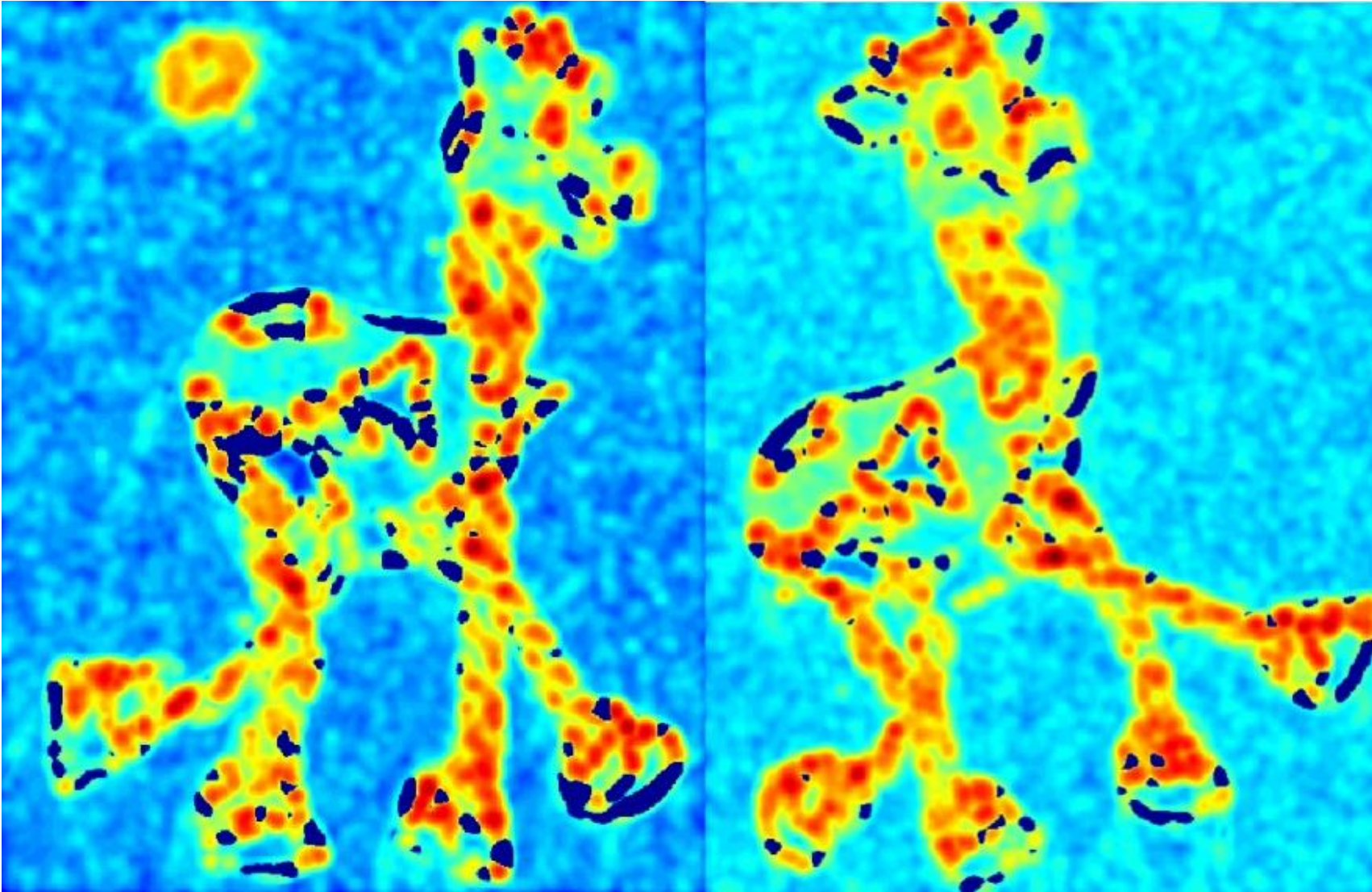


Harris Detector: Steps



Harris Detector: Steps

Compute corner response R



Harris Detector: Steps

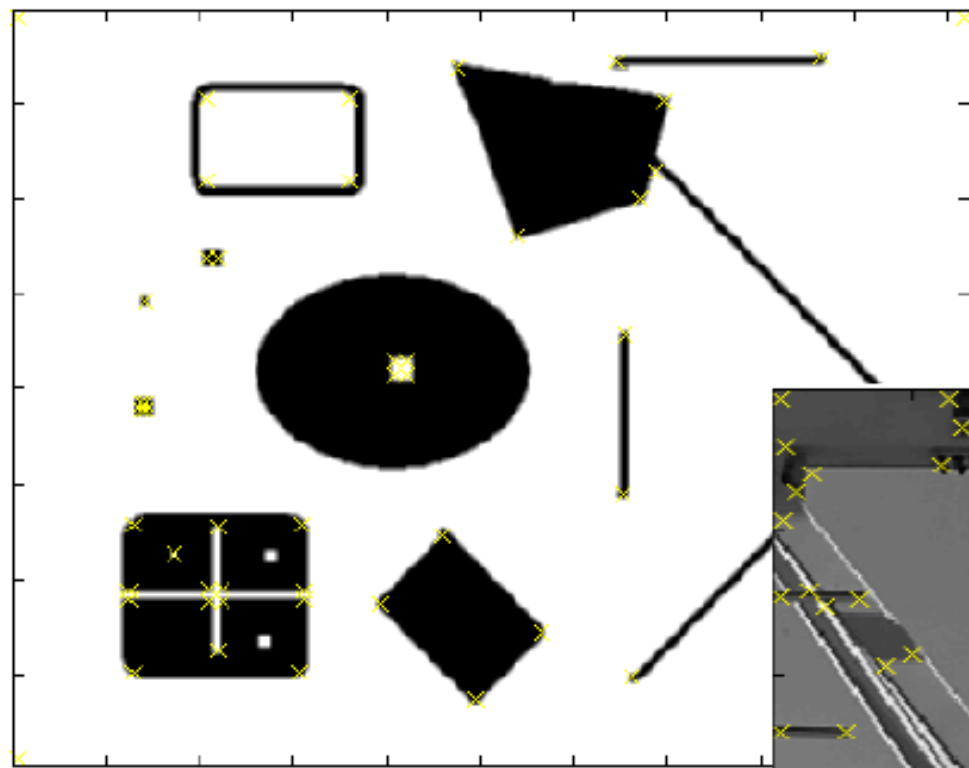
Take only the points of local maxima of R



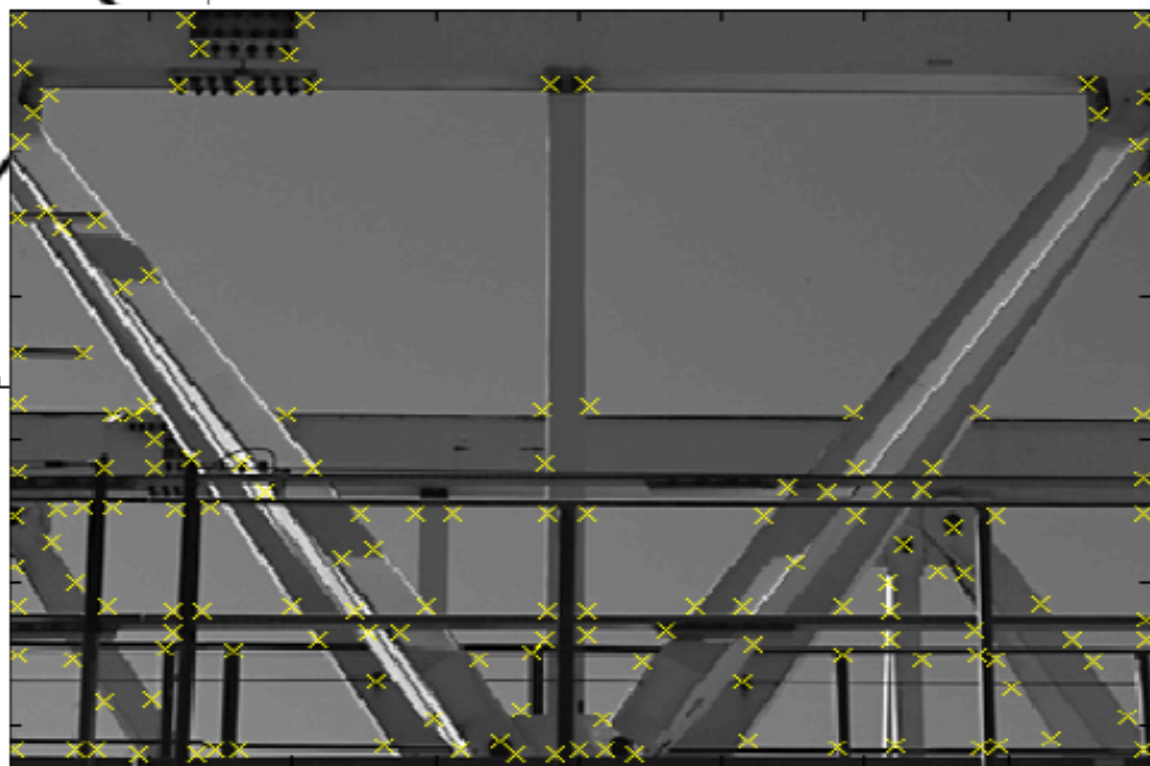
Harris Detector: Steps



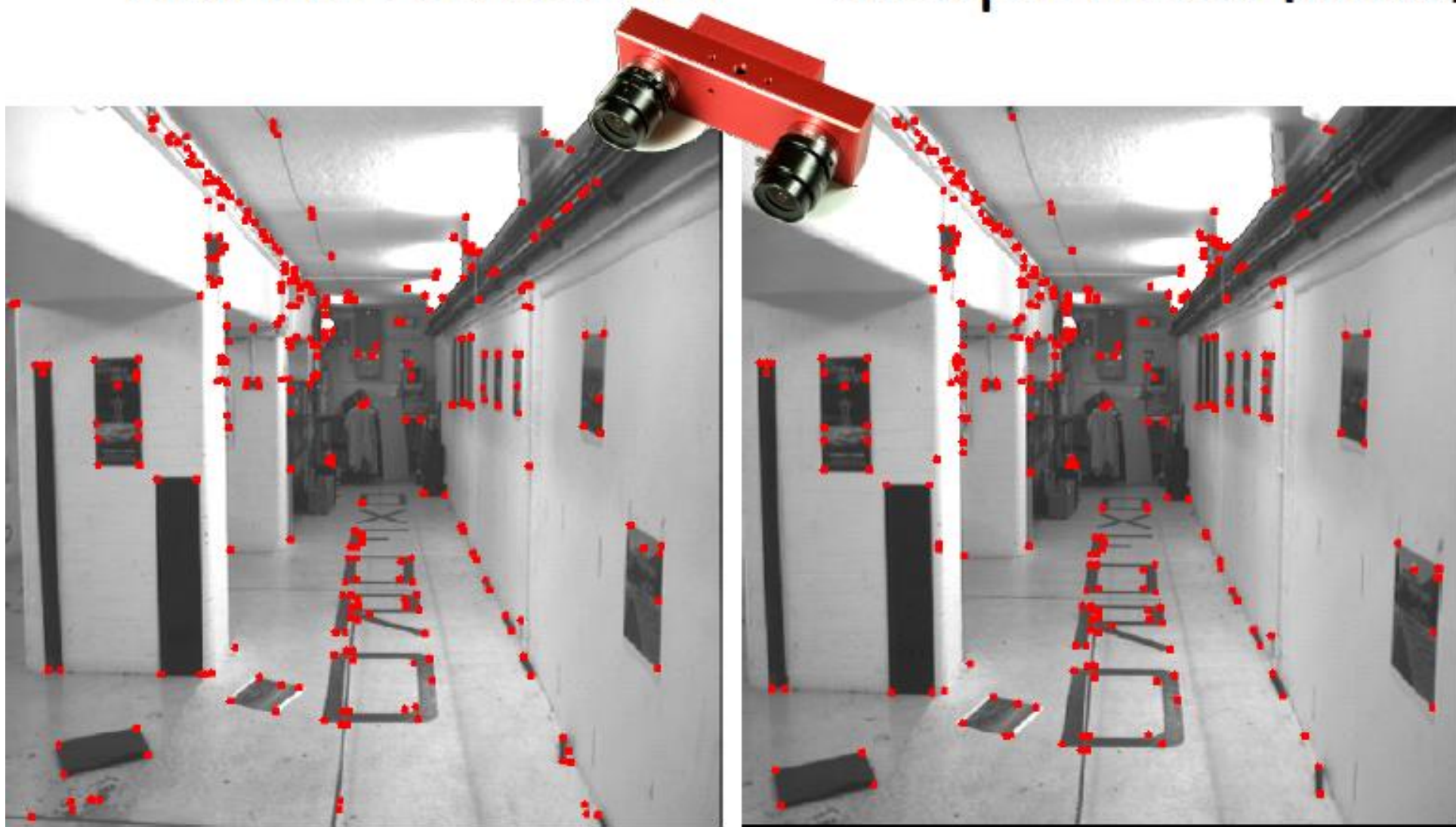
Harris Detector – Responses [Harris88]



Effect: A very precise corner detector.



Harris Detector – Responses [Harris88]



- Results are well suited for finding stereo correspondences

Hessian Matrix Detector

- For an Image the Hessian matrix can be expressed as

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

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



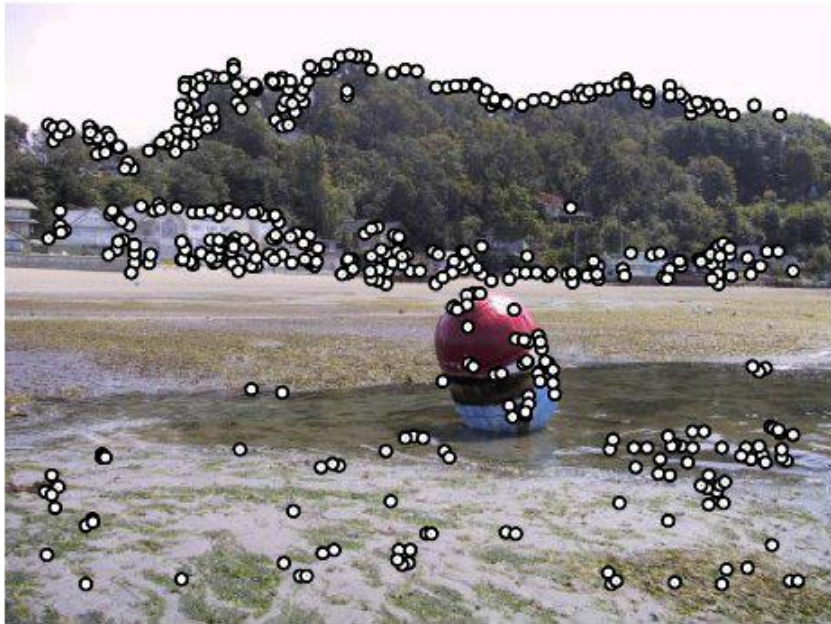
Result using vlfeat open library

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 .



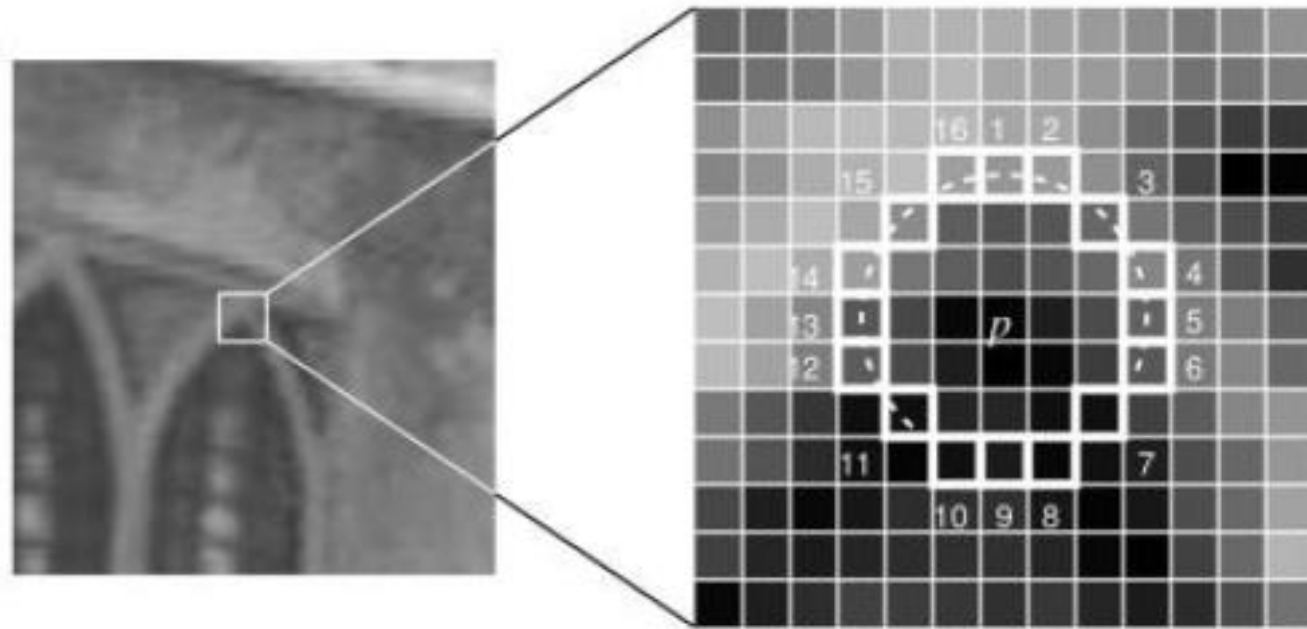
Strongest 500



ANMS 500, $r = 16$

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.