Juergen Gall

Image features and descriptors
MA-INF 2201 - Computer Vision
WS24/25

Motivation for feature-based alignment: Image mosaics

UNIVERSITÄT BONN

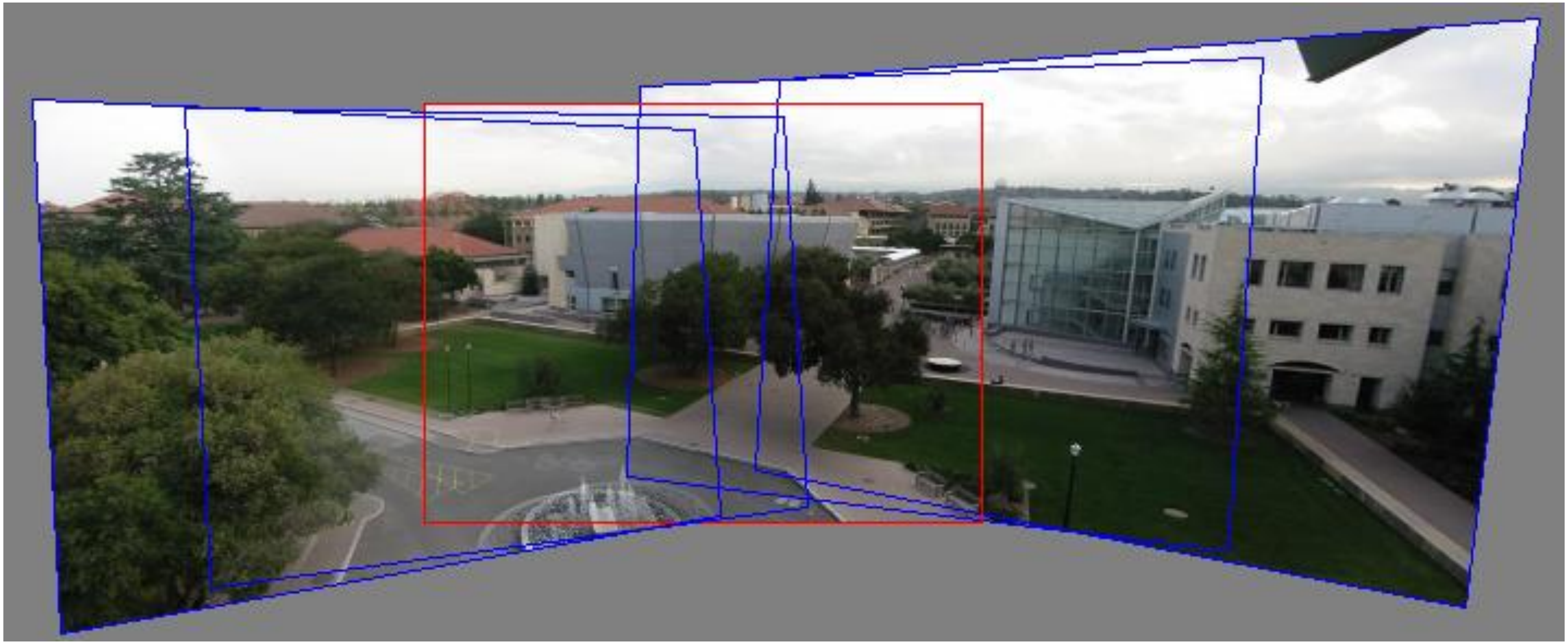
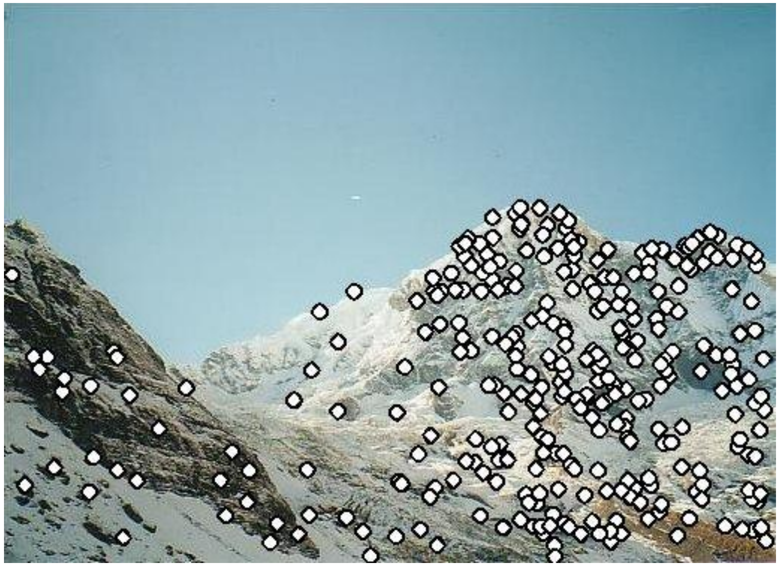


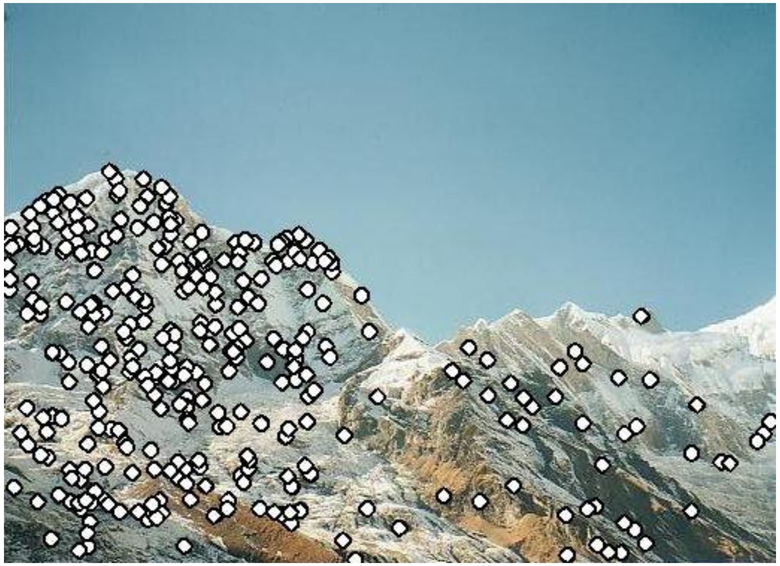
Image from http://graphics.cs.cmu.edu/courses/15-463/2010_fall/

Feature-based alignment

Source: L. Lazebnik

Feature-based alignment





- Extract features

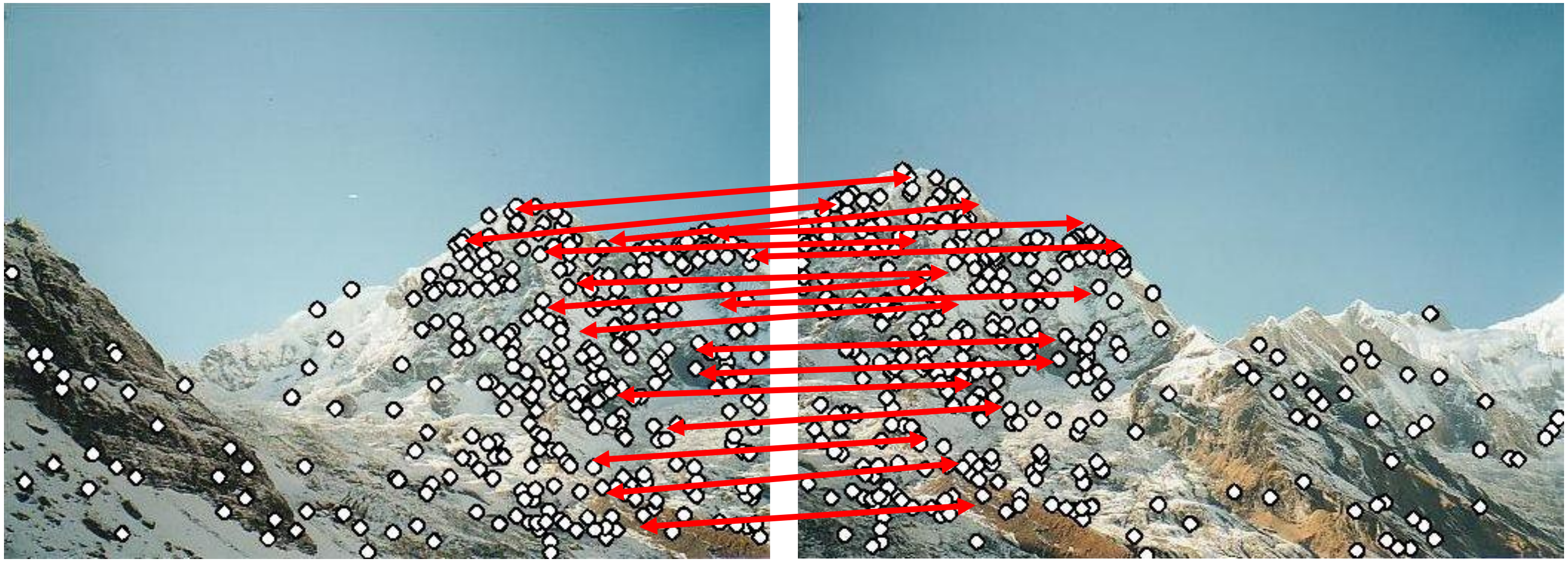
Source: L. Lazebnik

Feature-based alignment

- Extract features
- Compute matches

Source: L. Lazebnik

Feature-based alignment



- Extract features
- Compute matches
- Filter matches

Source: L. Lazebnik

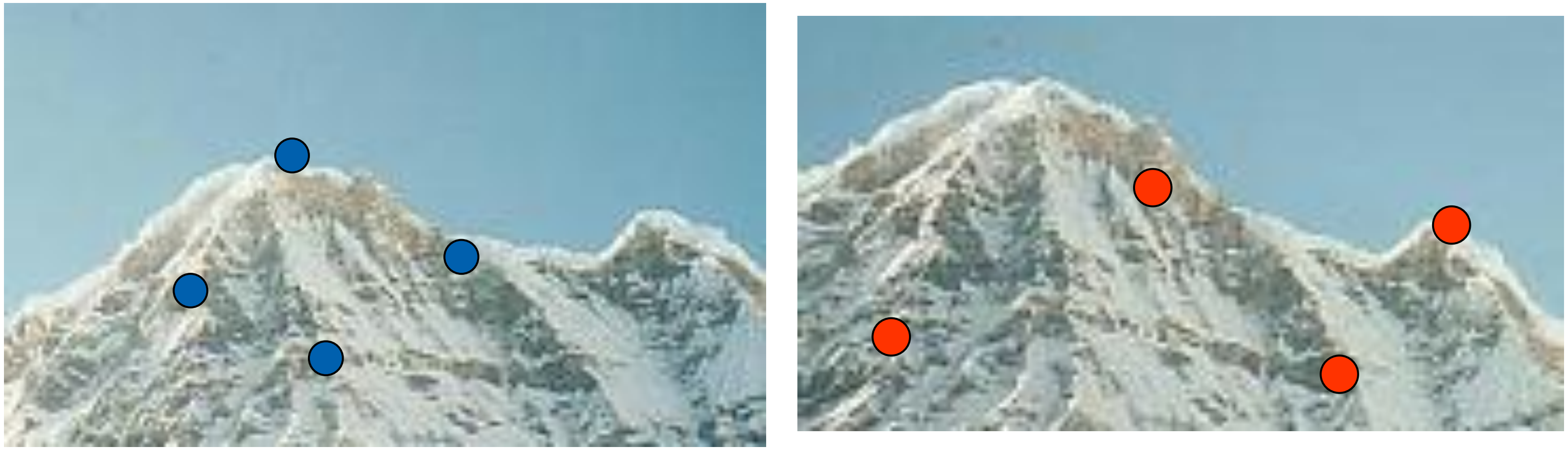
Feature-based alignment

- Extract features
- Compute matches
- Filter matches
- Compute transformation

Source: L. Lazebnik

Local features: desired properties

- Repeatability:
  - The same feature can be found in several images despite geometric and photometric transformations



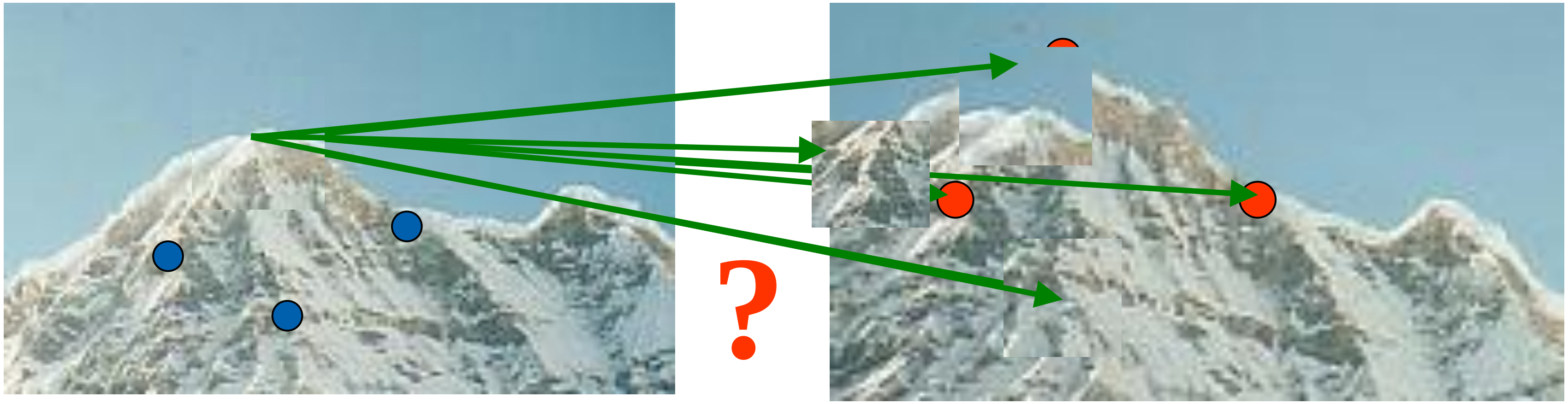
No chance to find true matches!

We want to detect (at least some of) the same points in both images.

Kristen Grauman

Local features: desired properties

- Repeatability:
  - The same feature can be found in several images despite geometric and photometric transformations
- Saliency:
  - Each feature has a distinctive description



Invariance to geometric and photometric differences between the two views.

Kristen Grauman

Local features: desired properties

- Repeatability:
  - The same feature can be found in several images despite geometric and photometric transformations
- Saliency:
  - Each feature has a distinctive description
- Locality:
  - A feature occupies a relatively small area of the image; robust to clutter and occlusion
- Sometimes compactness and efficiency:
  - Low dimensionality or binary

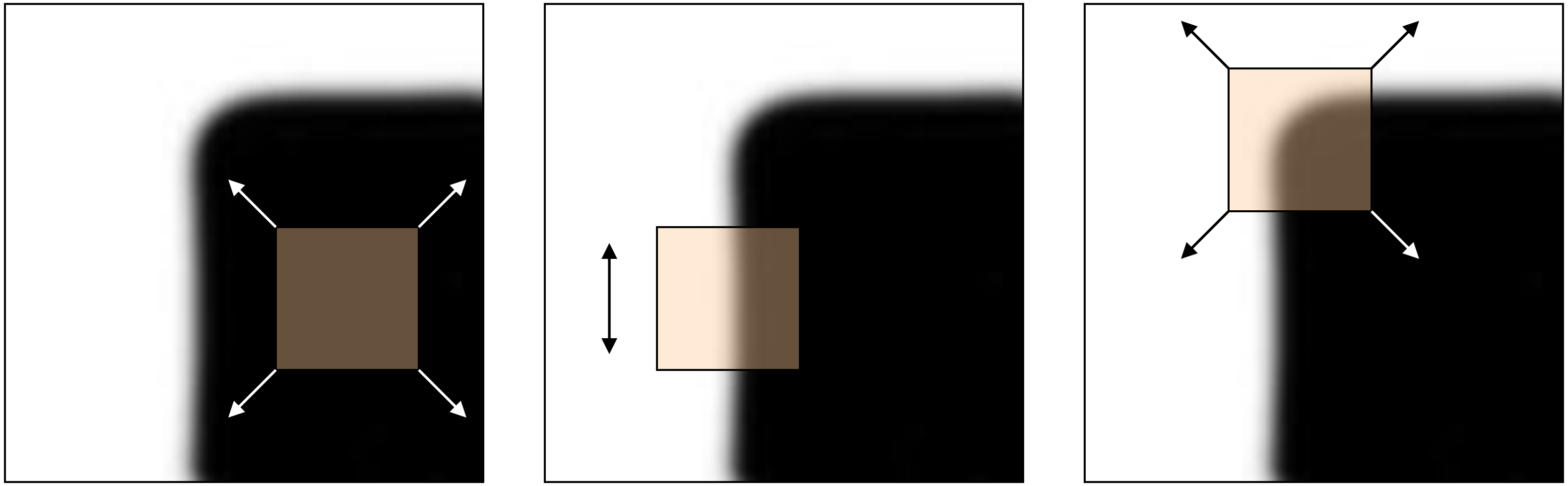
Kristen Grauman

What points would you choose?

Kristen Grauman

Distinctive interest points

- 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



"flat" region:
no change in all directions

"edge":
no change along the edge direction

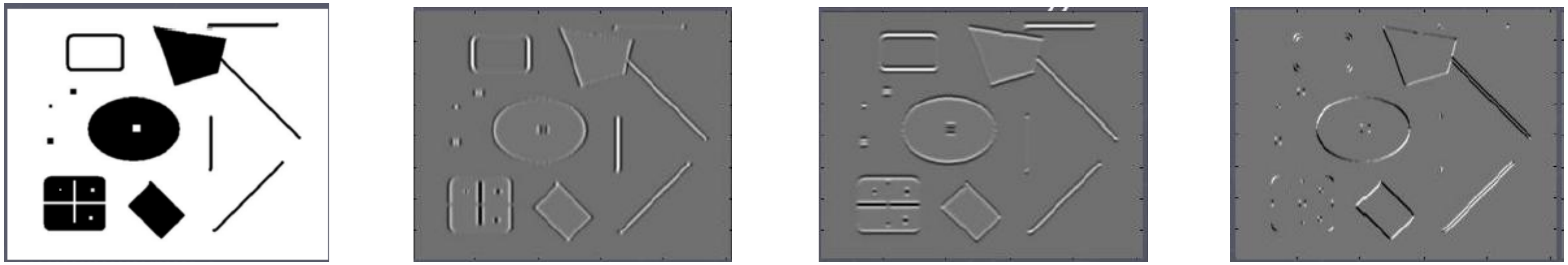
"corner":
significant change in all directions

Slide credit: Alyosha Efros, Darya Frolova, Denis Simakov

Corners as distinctive interest points

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

2 x 2 matrix of image derivatives (averaged in neighborhood of a point).

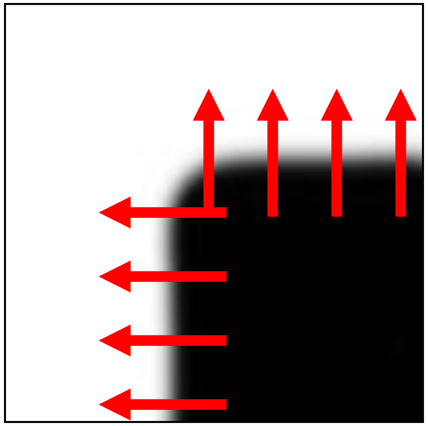


Notation: $I_x \Leftrightarrow \dfrac{\partial I}{\partial x}$ $\quad I_y \Leftrightarrow \dfrac{\partial I}{\partial y}$ $\quad I_x I_y \Leftrightarrow \dfrac{\partial I}{\partial x}\dfrac{\partial I}{\partial y}$

Kristen Grauman

What does this matrix reveal?

First, consider an axis-aligned corner:



Kristen Grauman

What does this matrix reveal?

First, consider an axis-aligned corner:

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

This means dominant gradient directions align with x or y axis

Look for locations where **both** λ's are large.

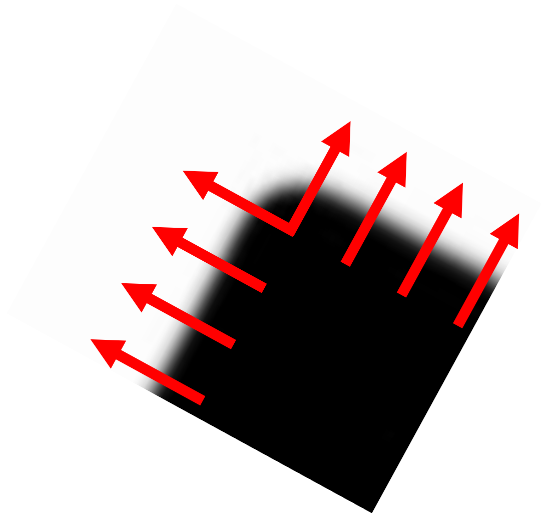
If either λ is close to 0, then this is **not** corner-like.

What if we have a corner that is not aligned with the image axes?

Kristen Grauman

What does this matrix reveal?

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

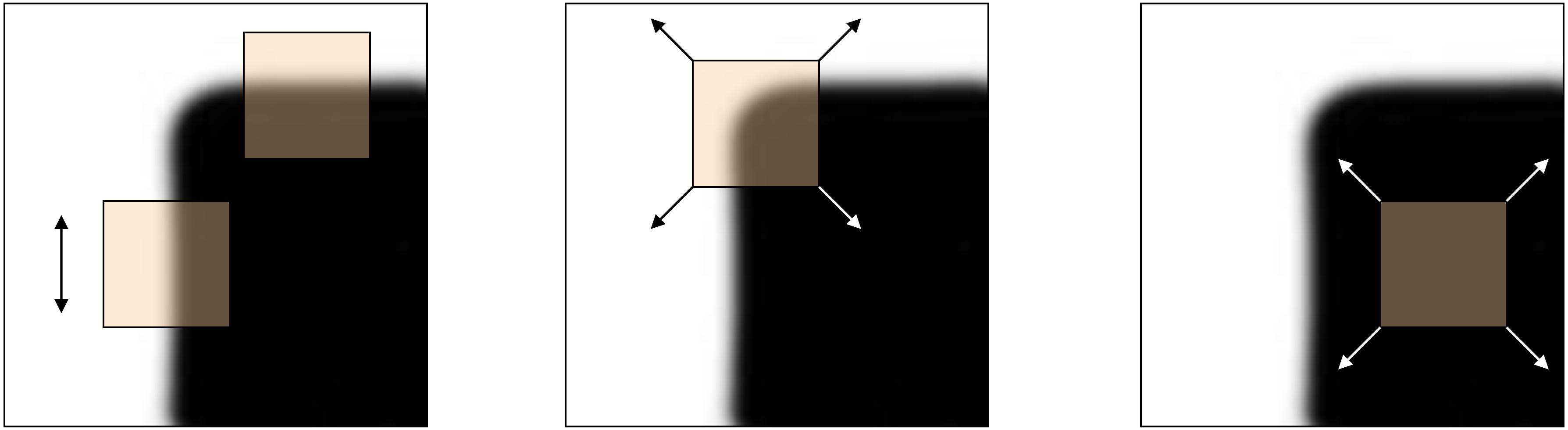


$$Mx_i = \lambda_i x_i$$

The *eigenvalues* of M reveal the amount of intensity change in the two principal orthogonal gradient directions in the window.

Kristen Grauman

Corner response function



"edge":

$\lambda_1 >> \lambda_2$

$\lambda_2 >> \lambda_1$

"corner":

λ_1 and λ_2 are large,
$\lambda_1 \sim \lambda_2$;

"flat" region

λ_1 and λ_2 are small;

$$f = |M| - k\,(\mathrm{Tr}(M))^2 = \lambda_1\lambda_2 - k\,(\lambda_1 + \lambda_2)^2$$

$$k \approx 0.04$$

Harris corner detector

1) Compute *M* matrix for each image window to get their *cornerness* scores.
2) Find points whose surrounding window gave large corner response (*f*> threshold)
3) Take the points of local maxima, i.e., perform non-maximum suppression

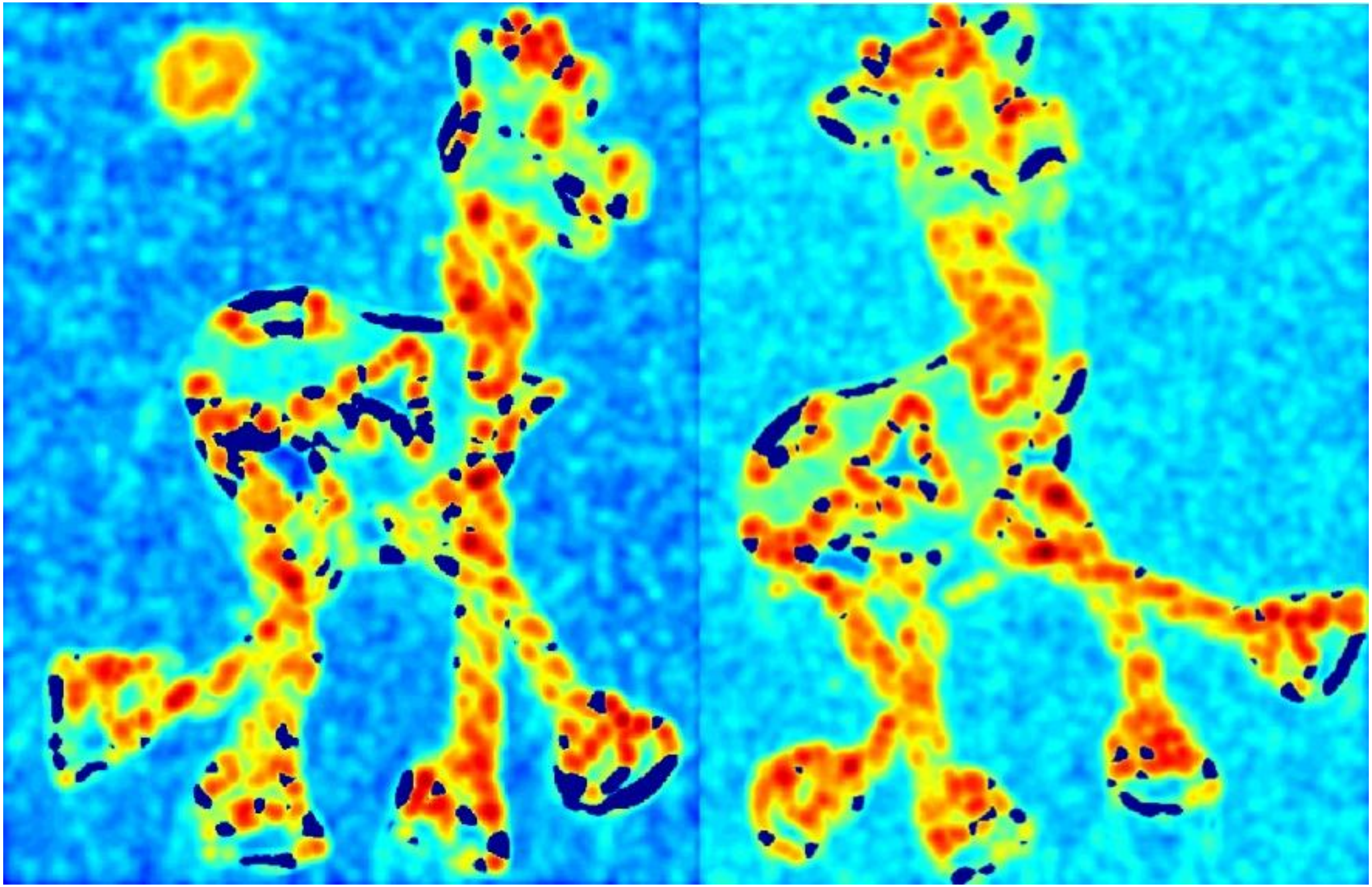
Kristen Grauman

Harris Detector

Kristen Grauman

Harris Detector

Compute corner response f

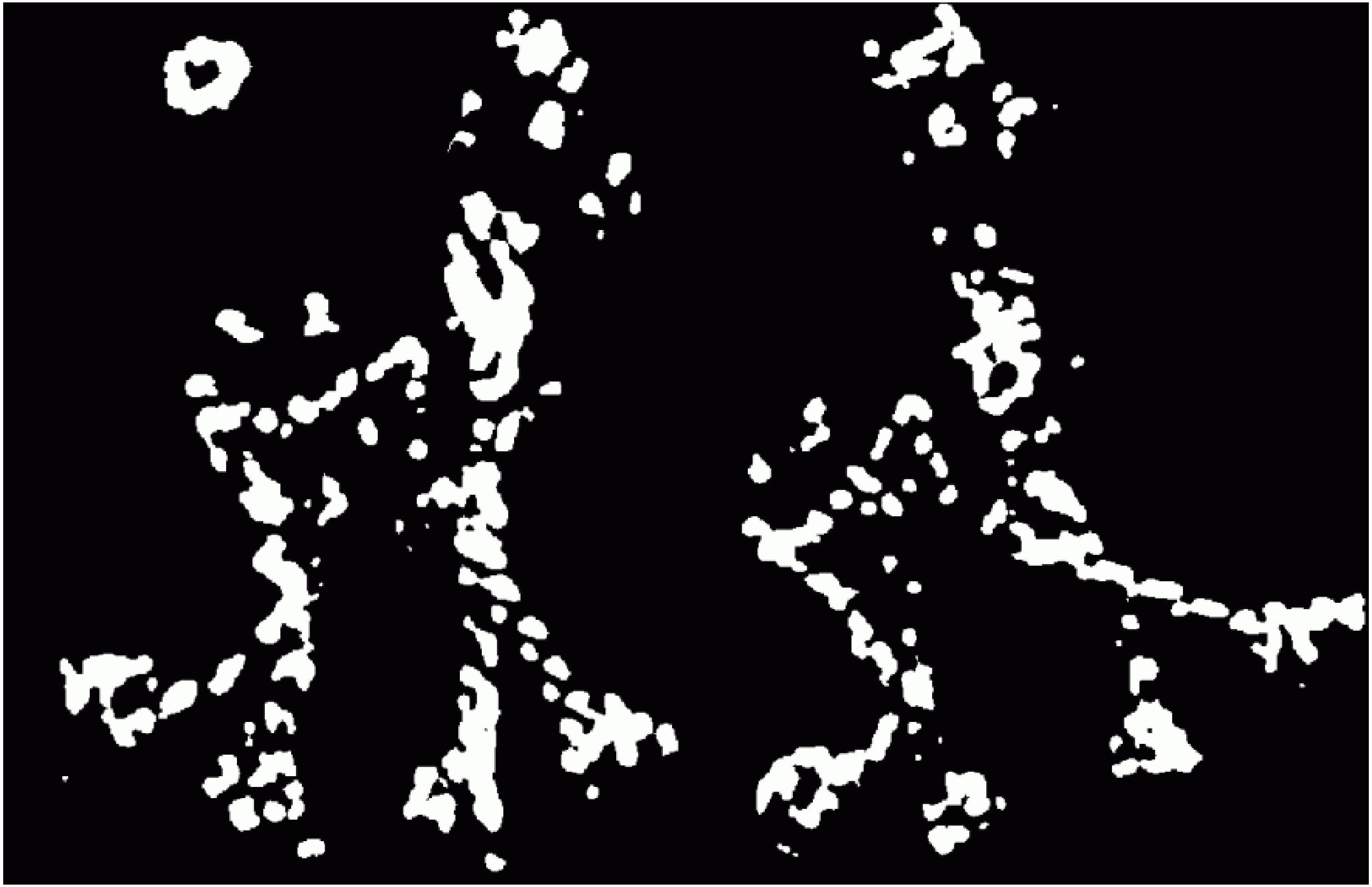


Kristen Grauman

Harris Detector

UNIVERSITÄT BONN

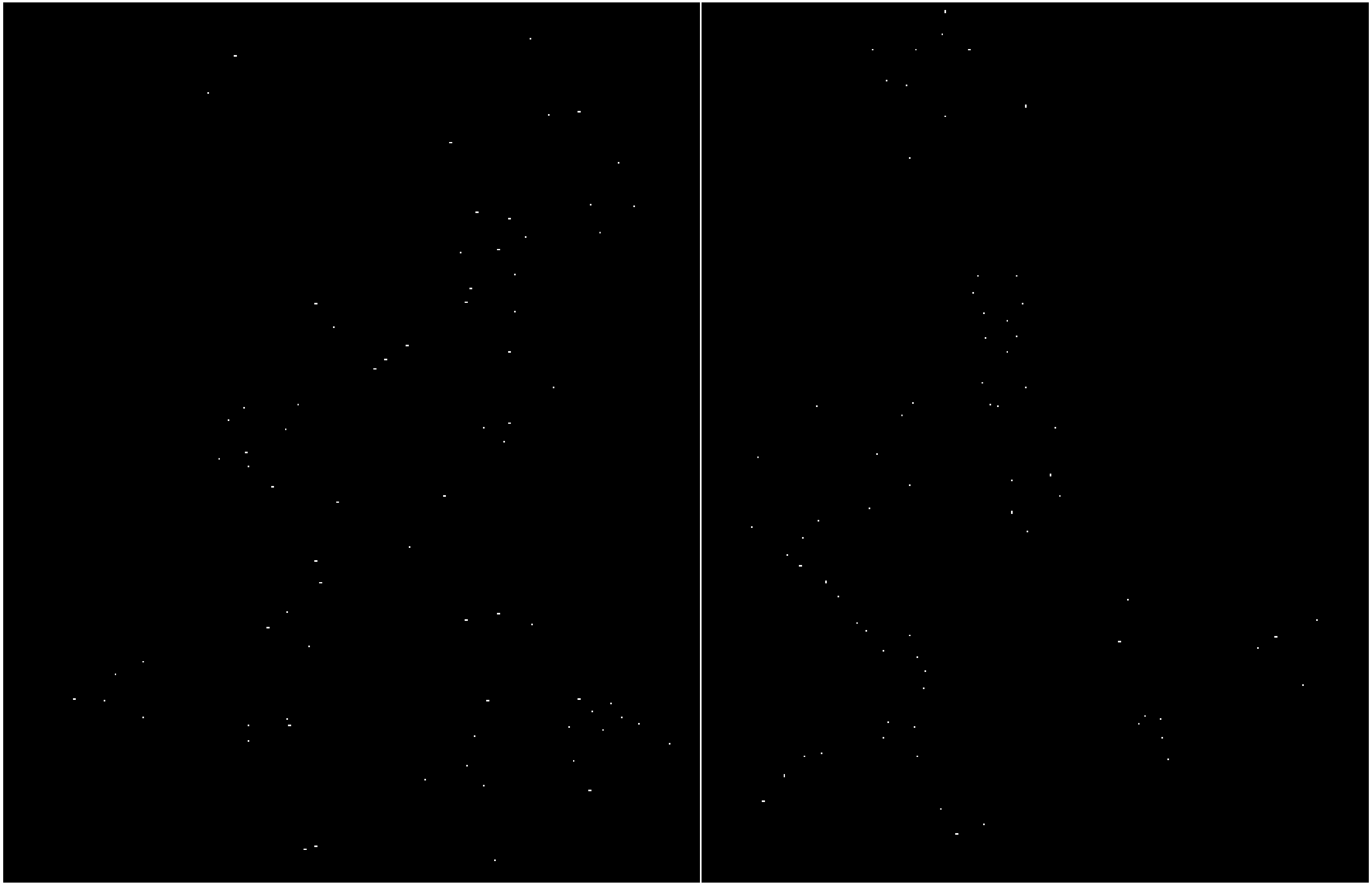
Find points with large corner response: f > threshold



Kristen Grauman

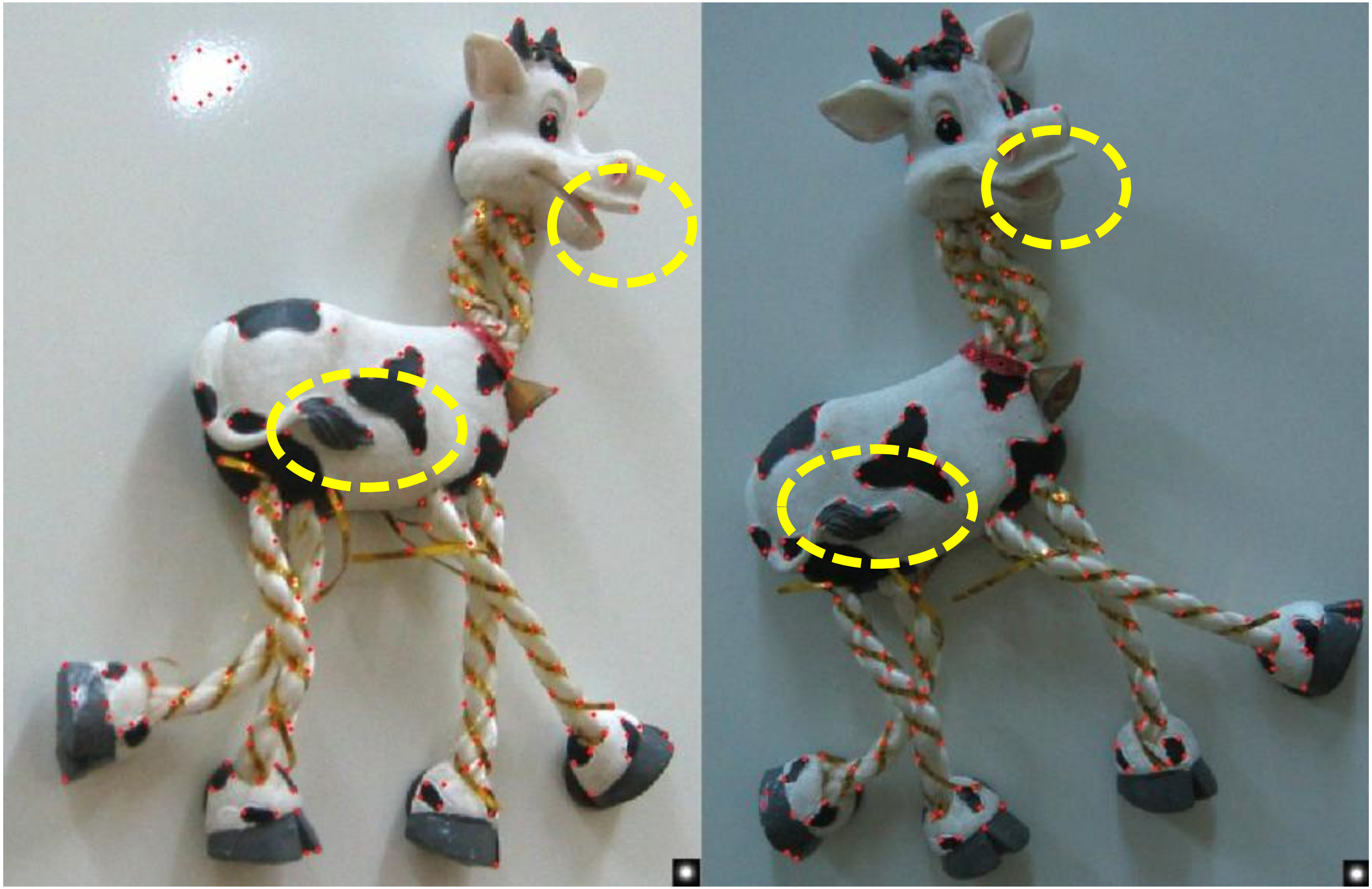
Harris Detector

Take only the points of local maxima of *f*



Kristen Grauman

Harris Detector



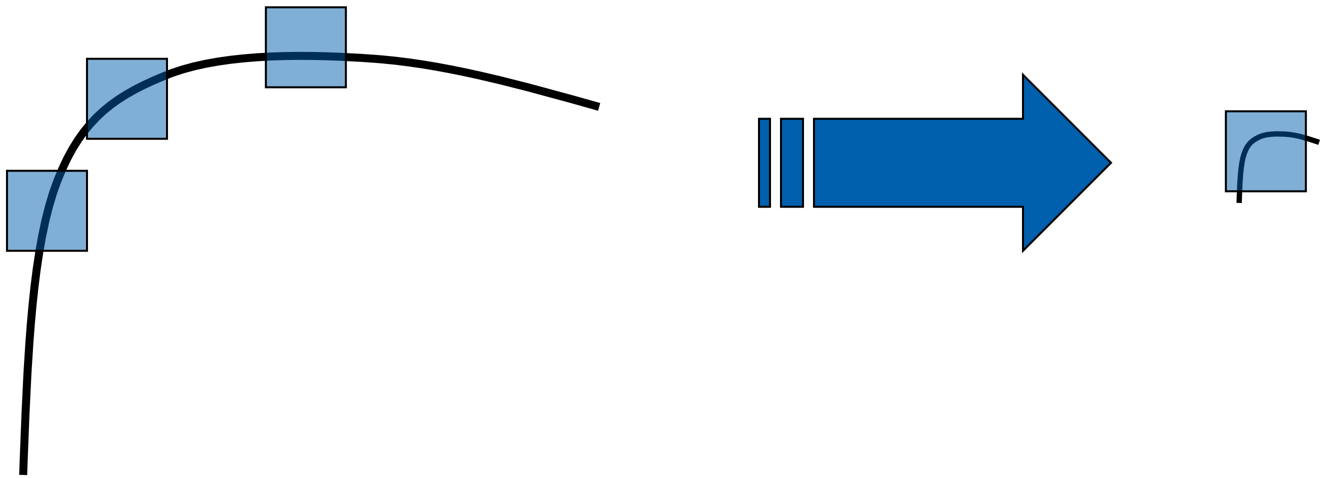
Kristen Grauman

Properties of the Harris corner detector

- Rotation invariant? Yes

$$M = X \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} X^T$$

- Scale invariant? No



All points will be classified as edges

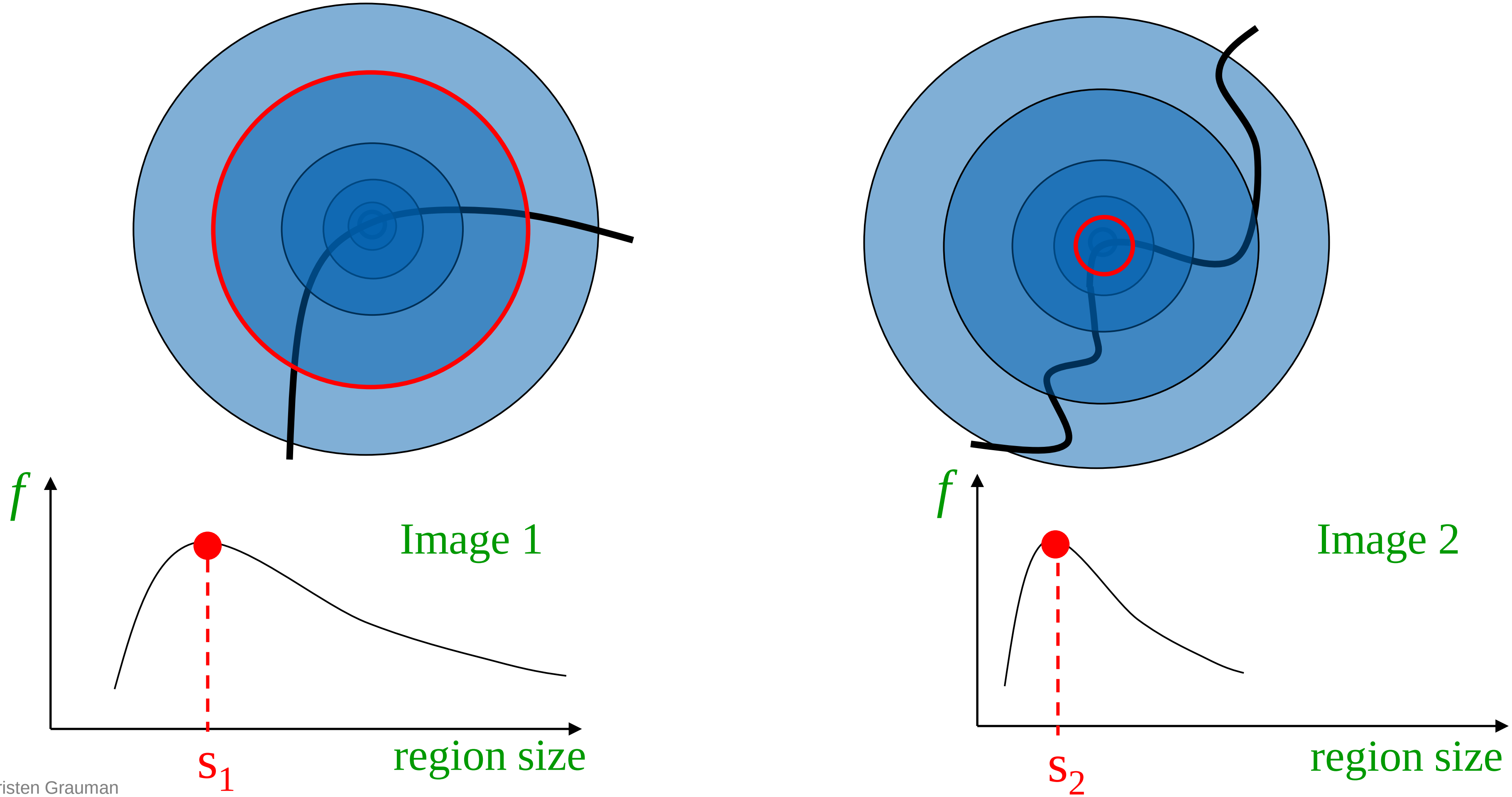
Corner !

Kristen Grauman

Automatic scale selection

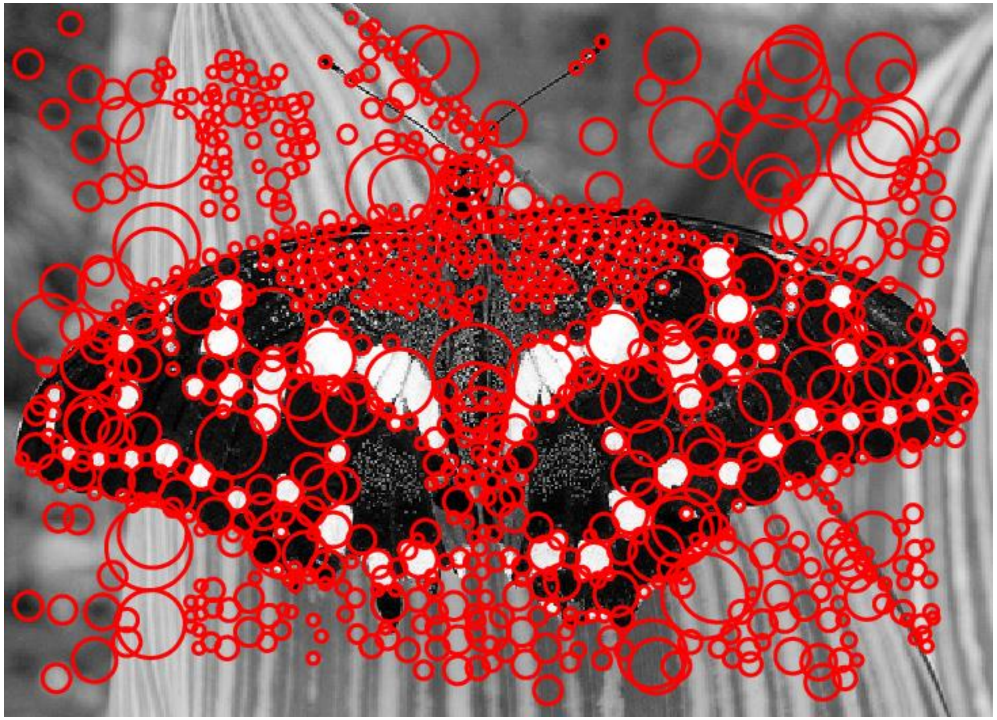
Intuition:

- Find scale that gives local maxima of some function *f* in both position and scale.



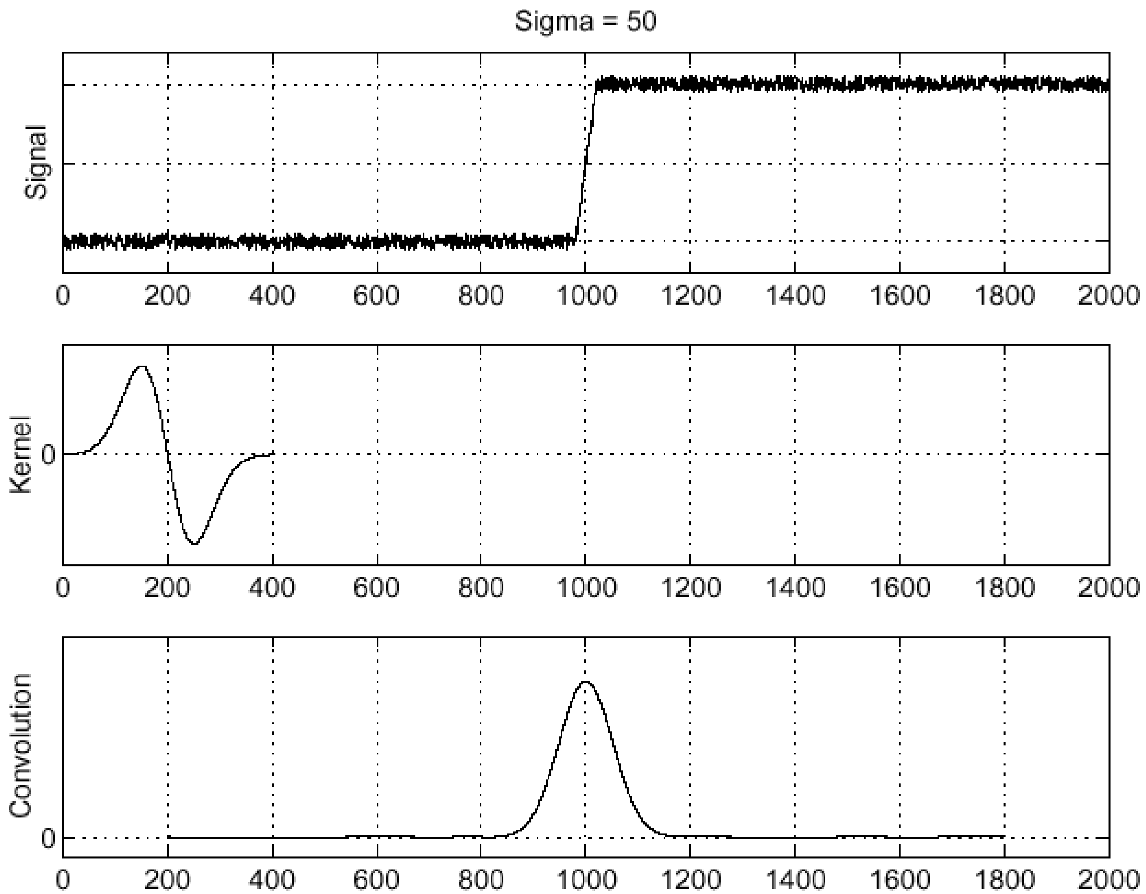
Kristen Grauman

Scale invariant blob detection



Kristen Grauman

Recall: Edge detection



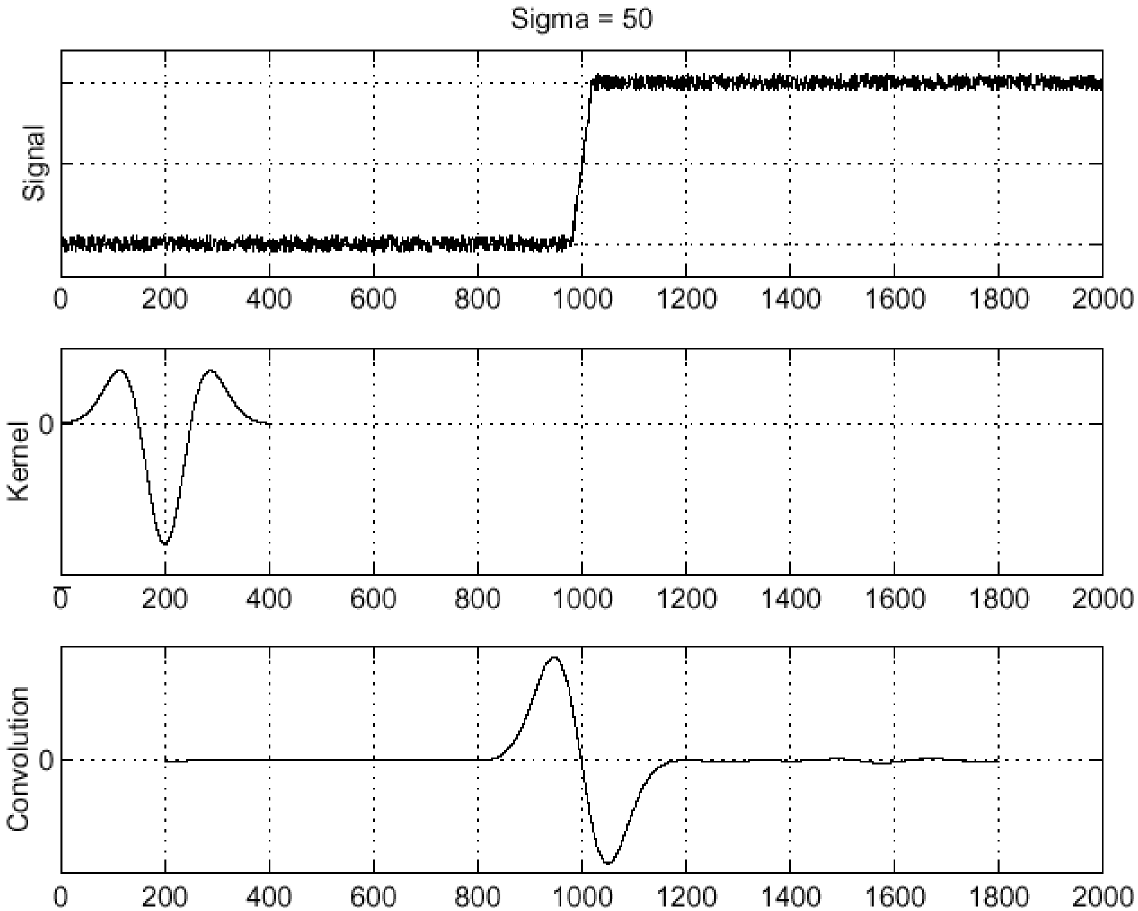
f Edge

$\frac{d}{dx} g$ Derivative of Gaussian

$f * \frac{d}{dx} g$ Edge = maximum of derivative

Source: S. Seitz

Recall: Edge detection



f — Edge

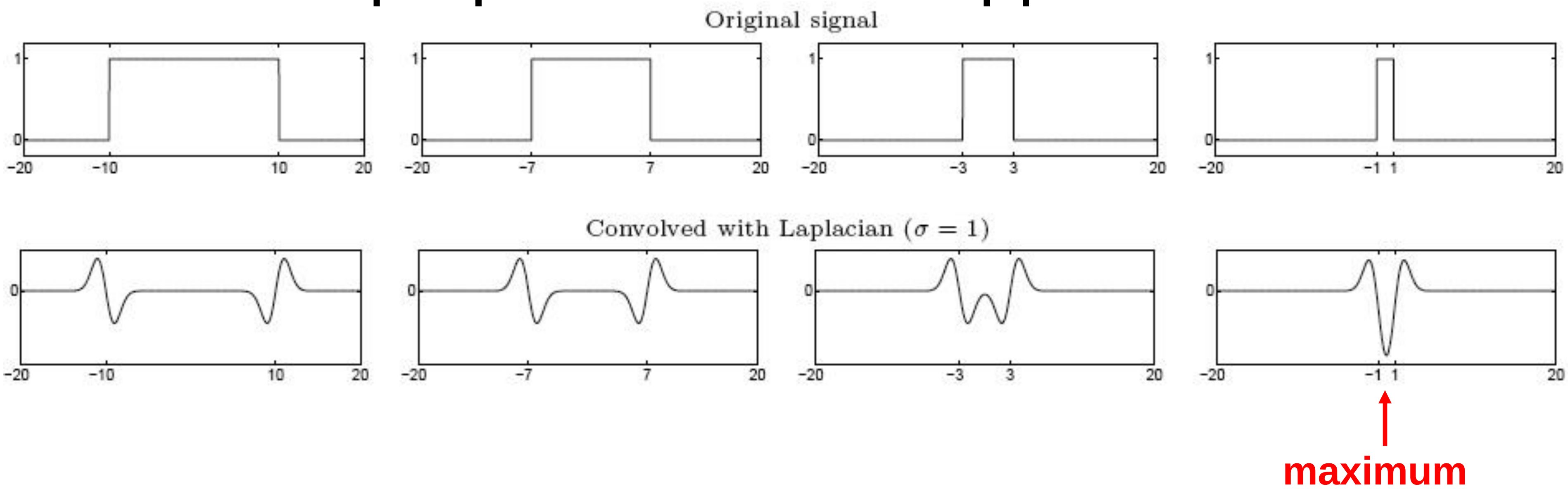
$\frac{d^2}{dx^2} g$ — Second derivative of Gaussian (Laplacian)

$f * \frac{d^2}{dx^2} g$ — Edge = zero crossing of second derivative

Source: S. Seitz

From edges to blobs

- Edge = ripple
- Blob = superposition of two ripples

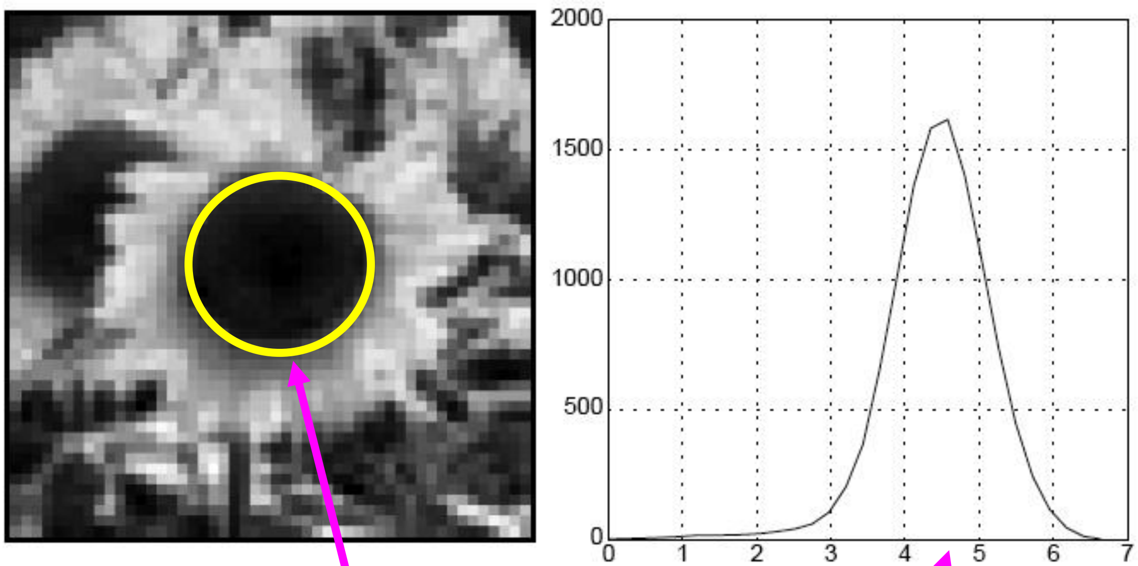


Spatial selection: the **magnitude** of the Laplacian response will achieve a maximum at the center of the blob, provided the scale of the Laplacian is "matched" to the scale of the blob

Lana Lazebnik

Blob detection in 2D

We define the *characteristic scale* as the scale that produces peak of Laplacian response



characteristic scale

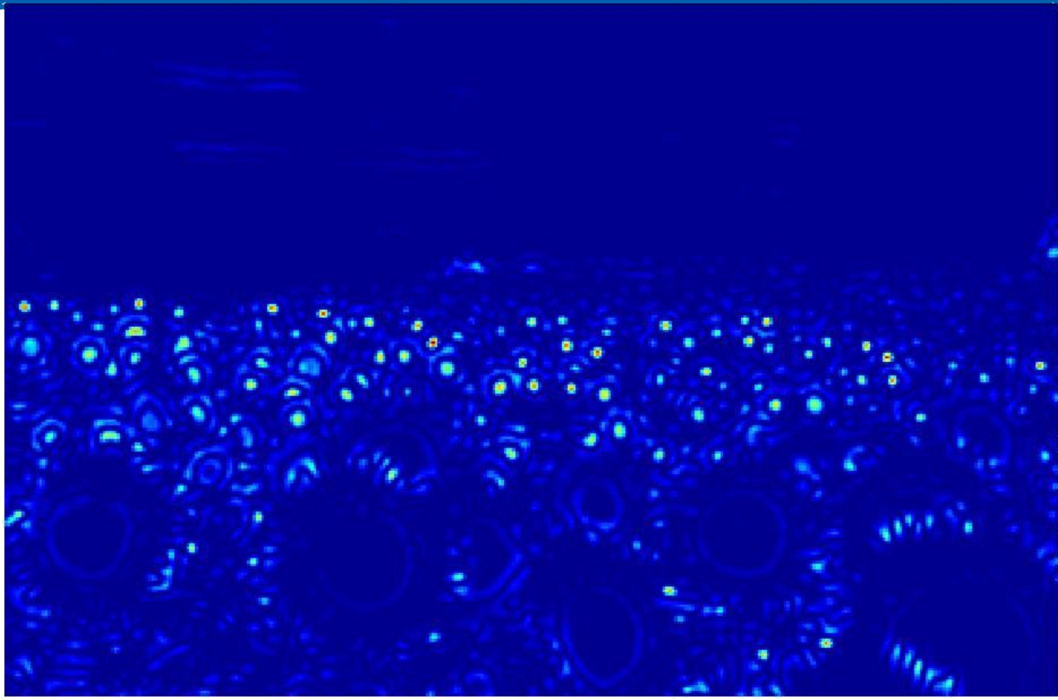
Lana Lazebnik

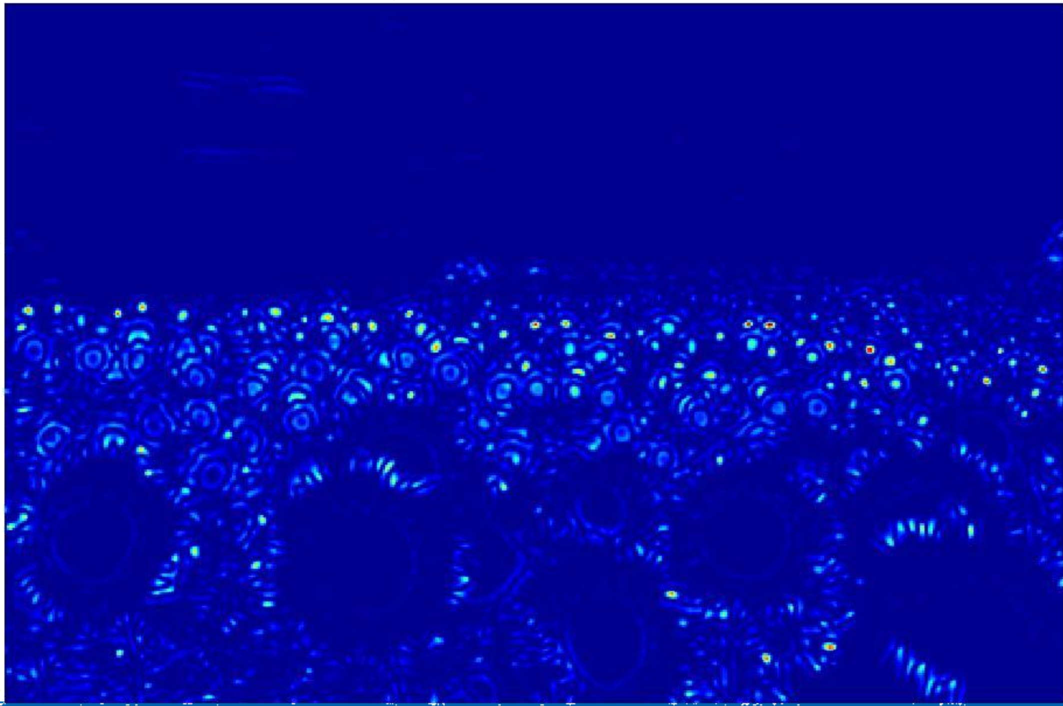
Example

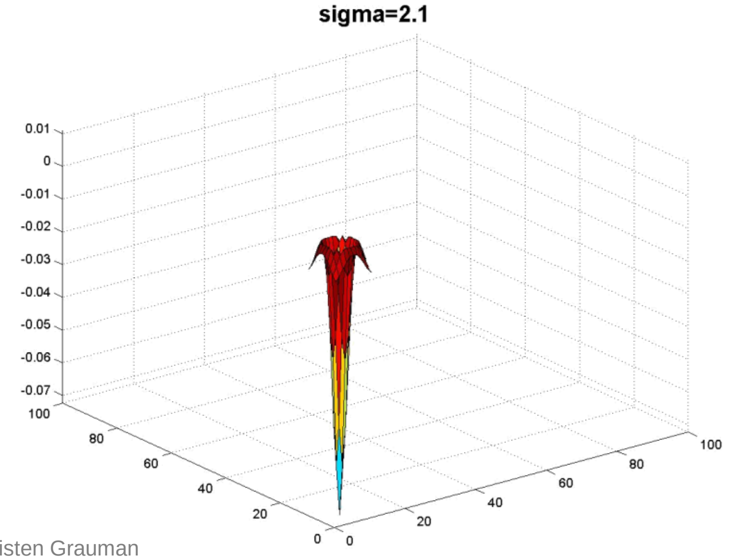
Original image
at ¾ the size

Kristen Grauman

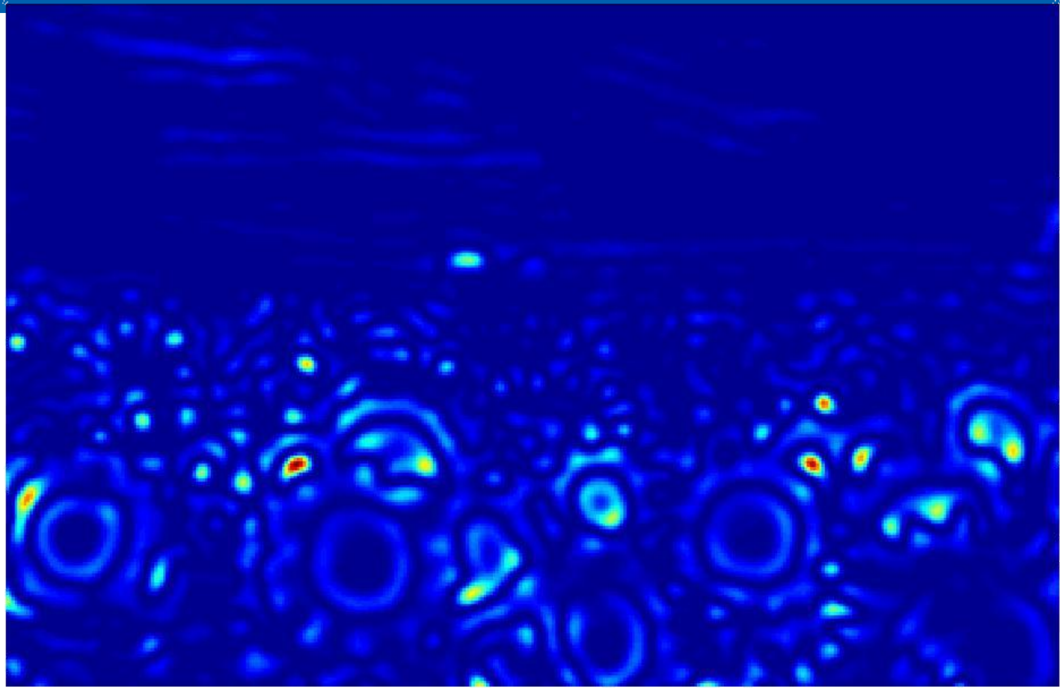
Original image
at ¾ the size

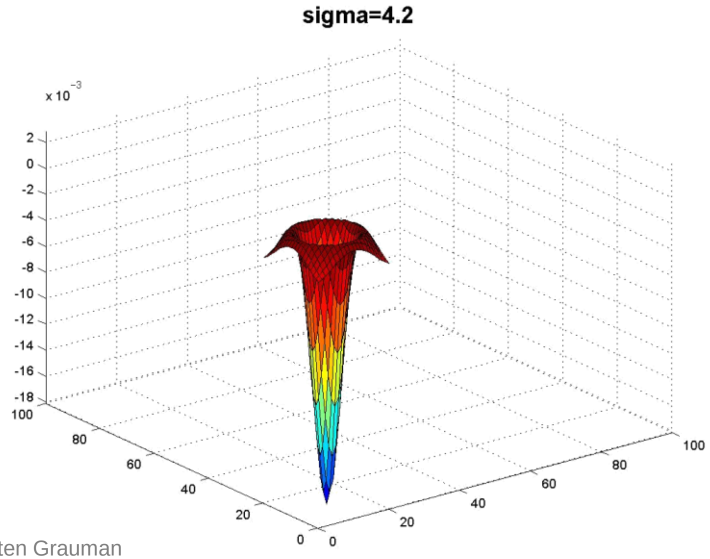


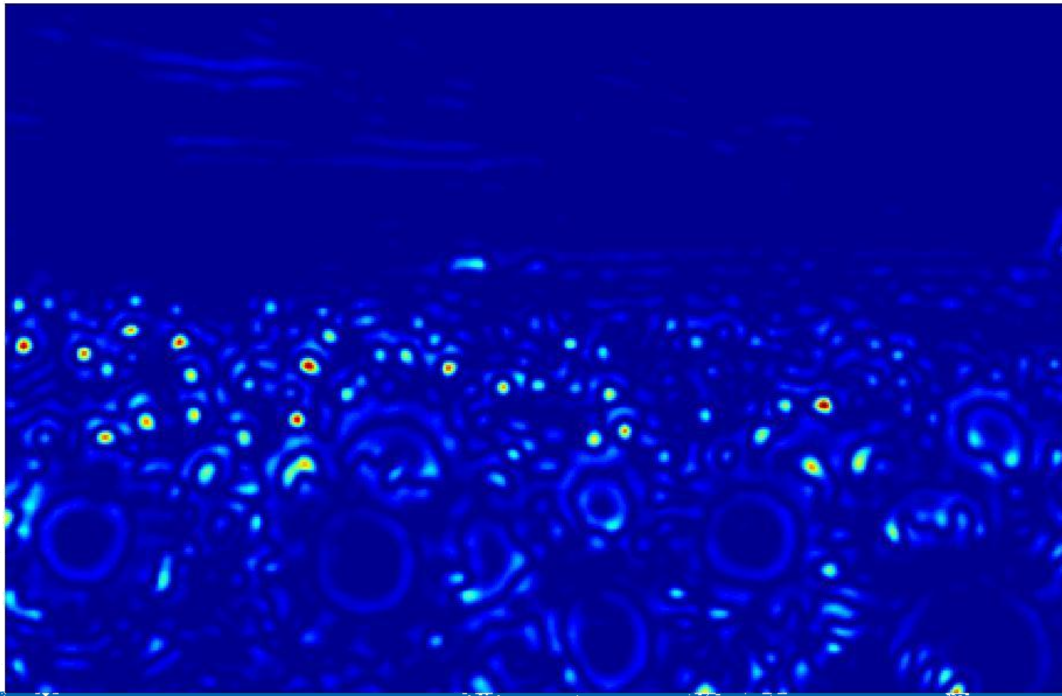




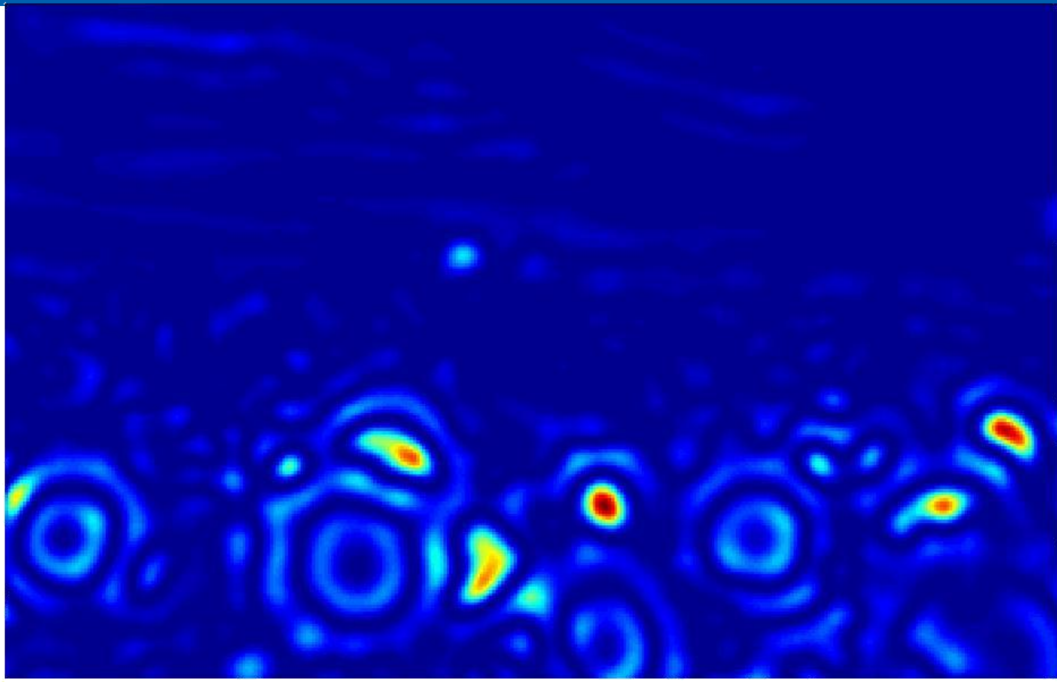
Kristen Grauman

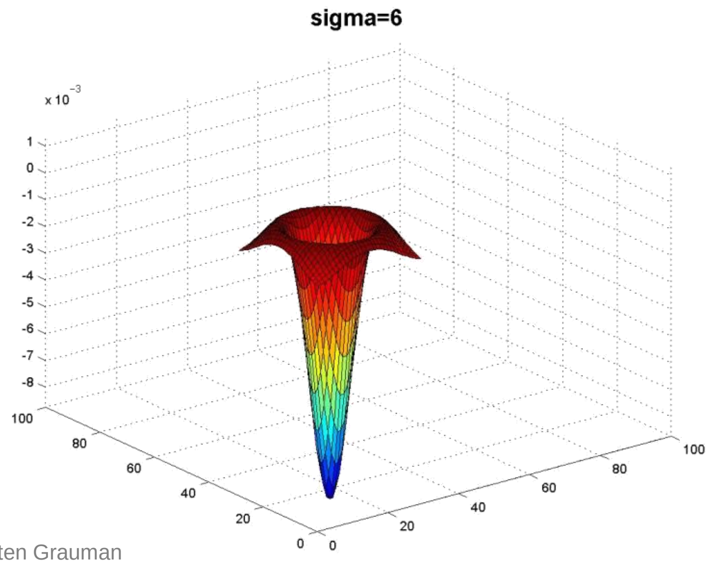


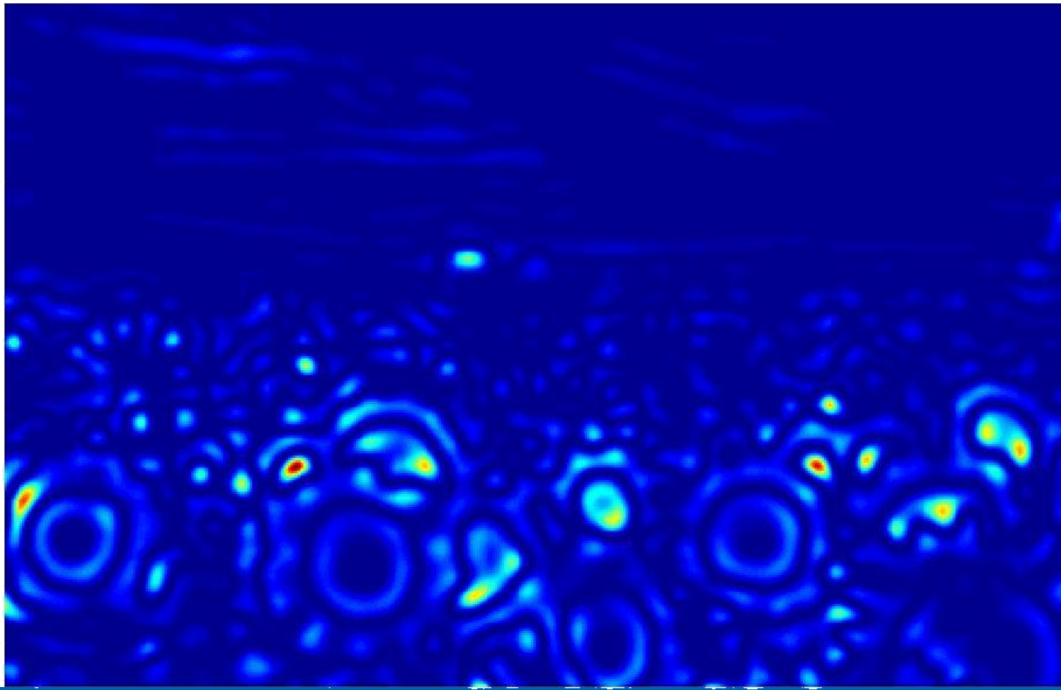




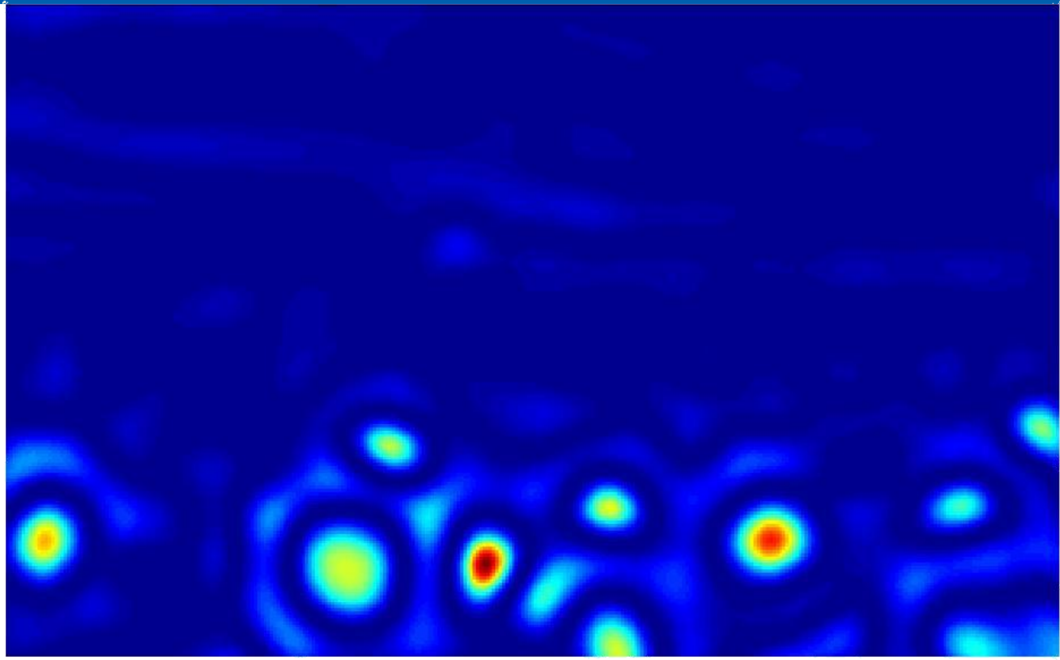
Kristen Grauman

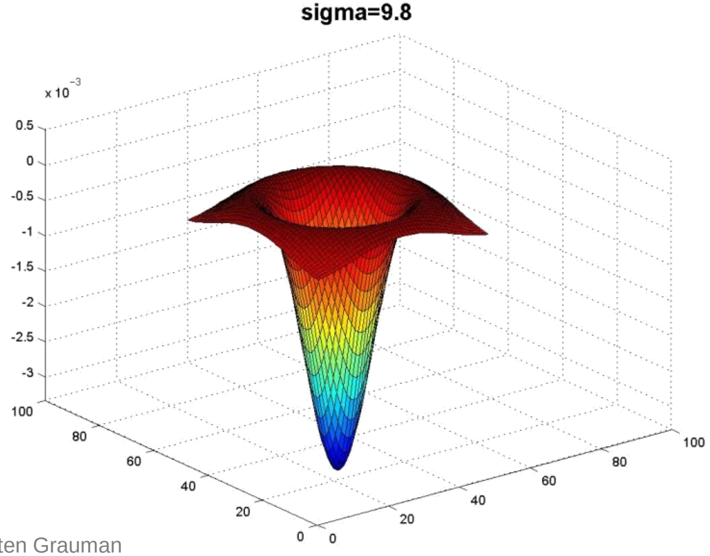


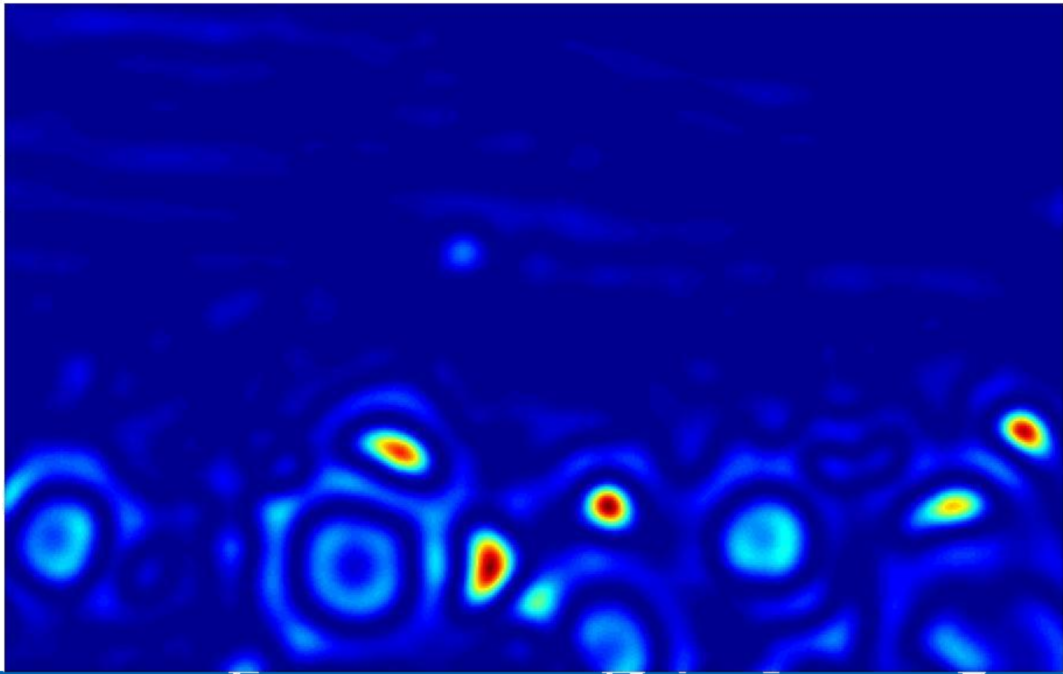




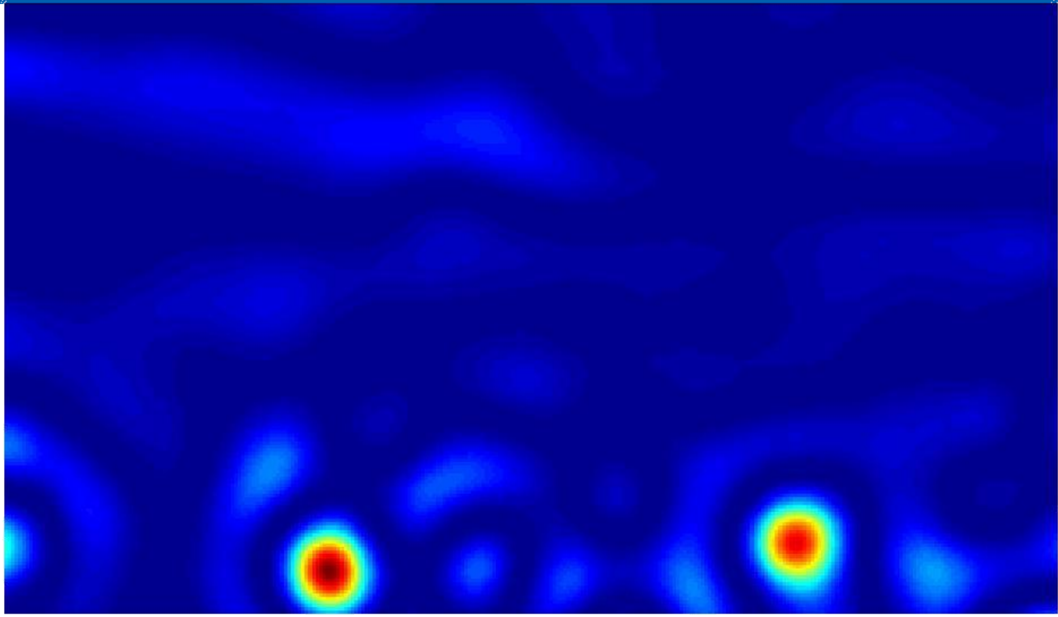
Kristen Grauman

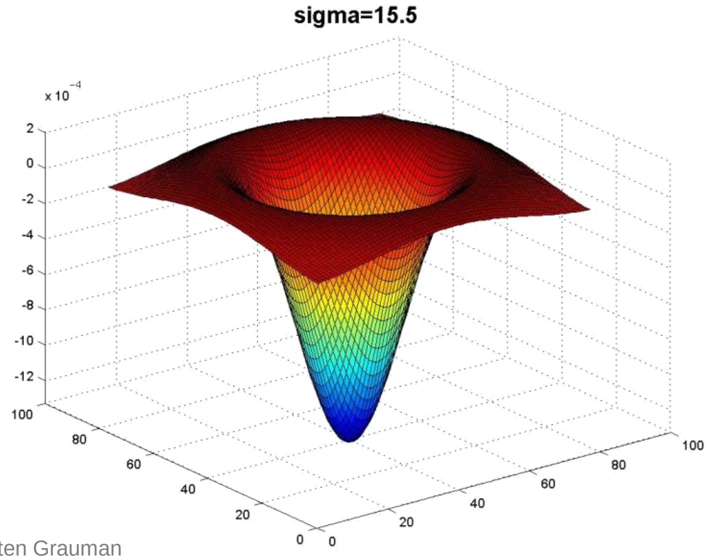


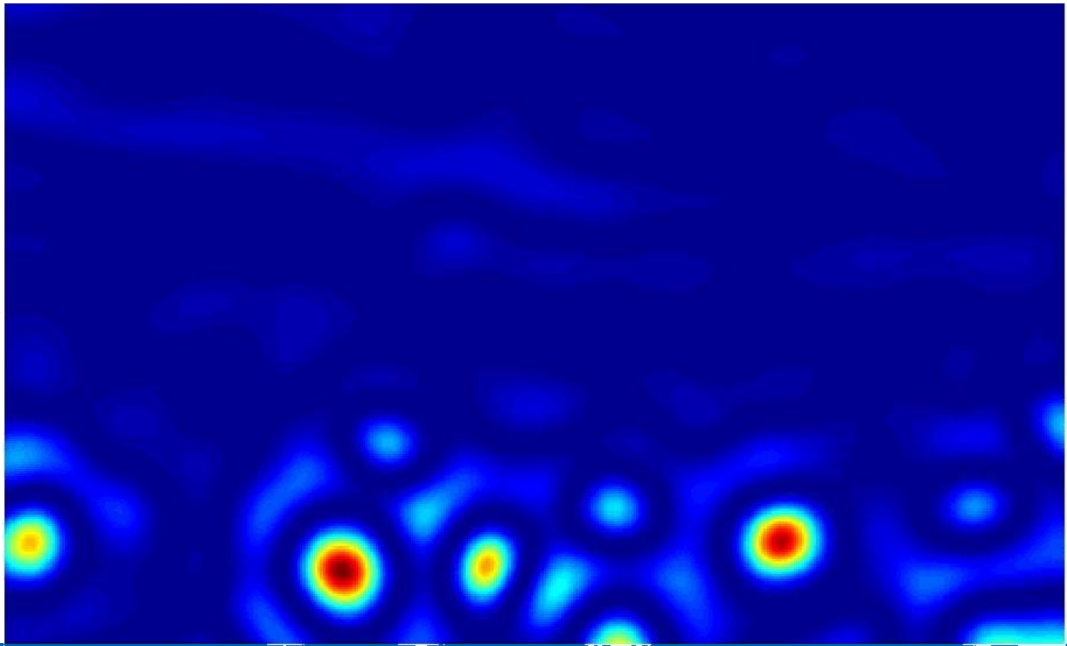




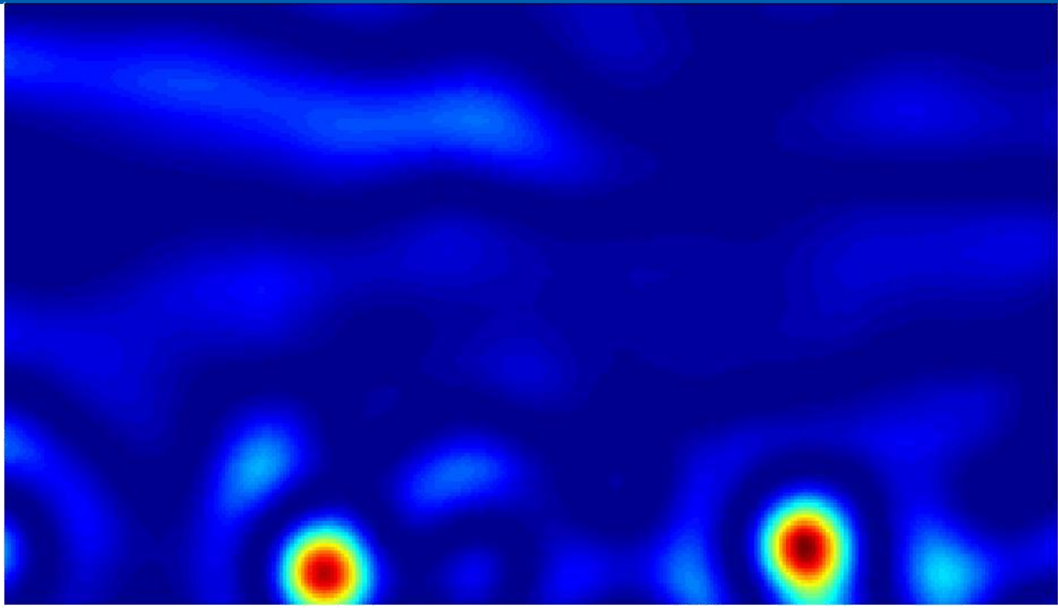
Kristen Grauman

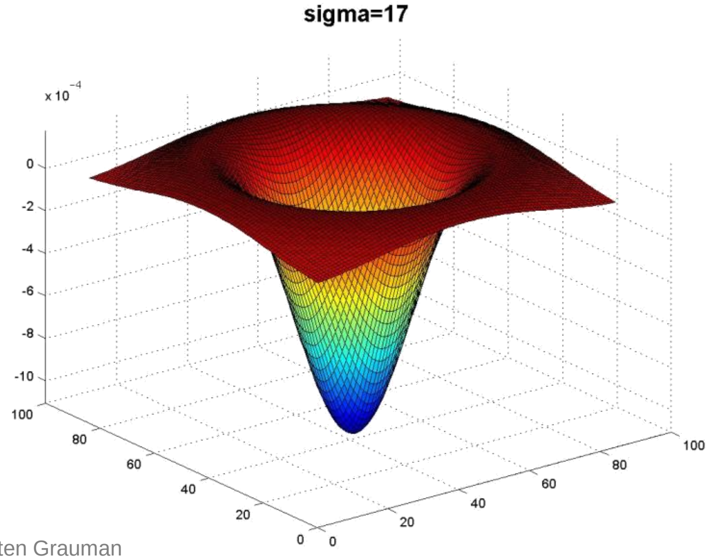


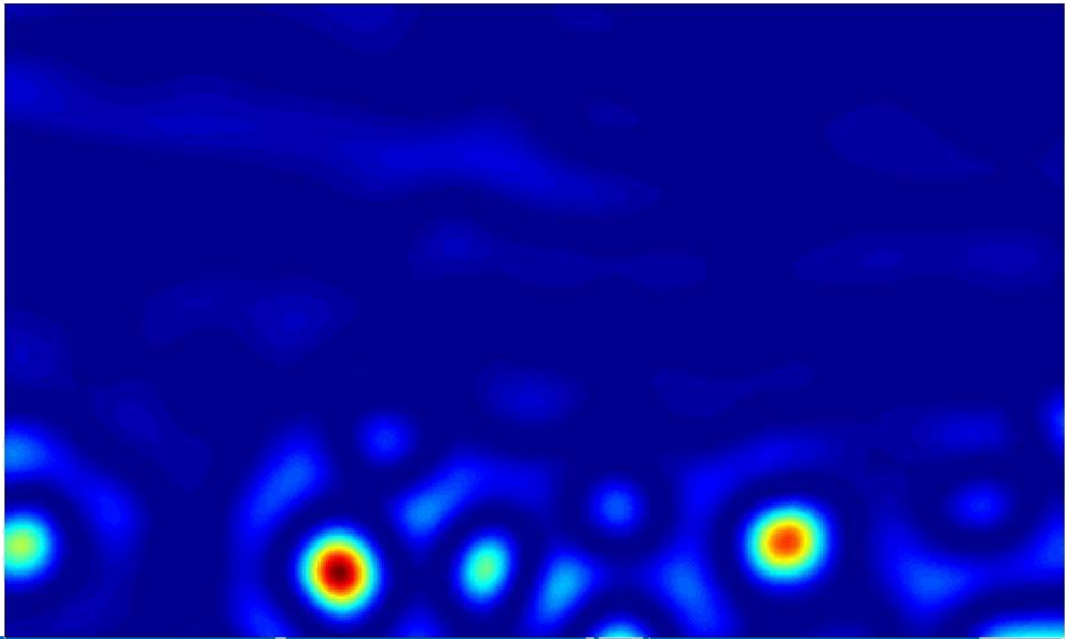




Kristen Grauman



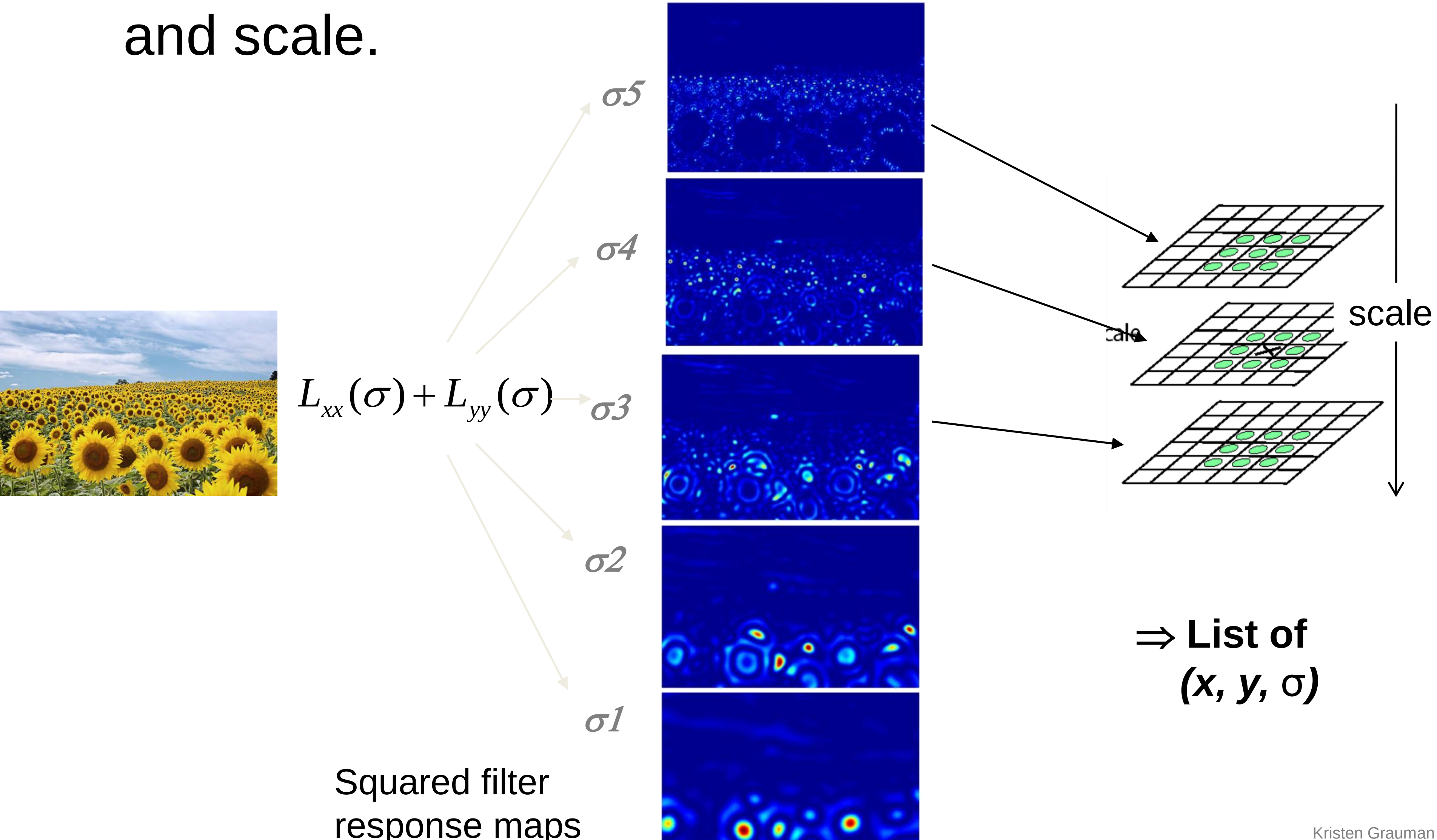




Kristen Grauman

Scale invariant interest points

Interest points are local maxima in both position and scale.



Squared filter response maps

⇒ **List of (*x*, *y*, σ)**

Kristen Grauman

Scale-space blob detector

Lana Lazebnik

Technical detail

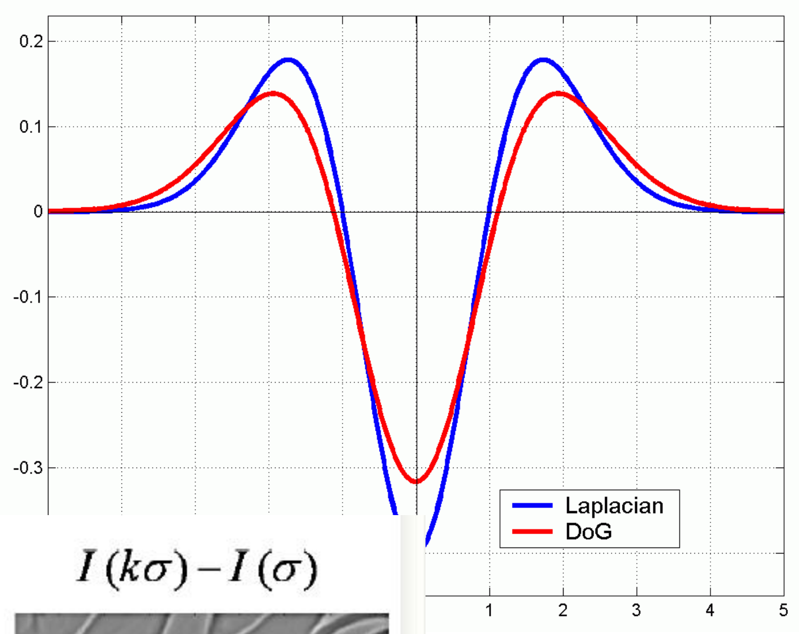
We can approximate the Laplacian with a difference of Gaussians; more efficient to implement.

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

(Laplacian)

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

(Difference of Gaussians)



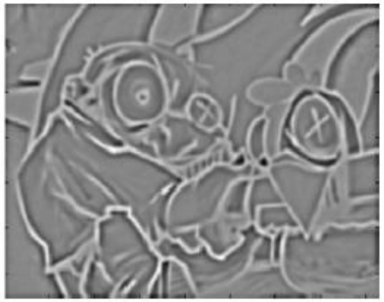
$I(k\sigma)$

-

$I(\sigma)$

=

$I(k\sigma) - I(\sigma)$

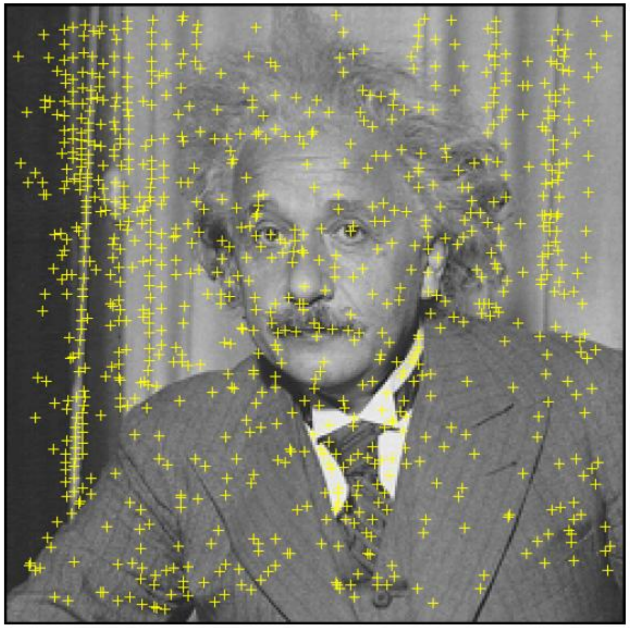


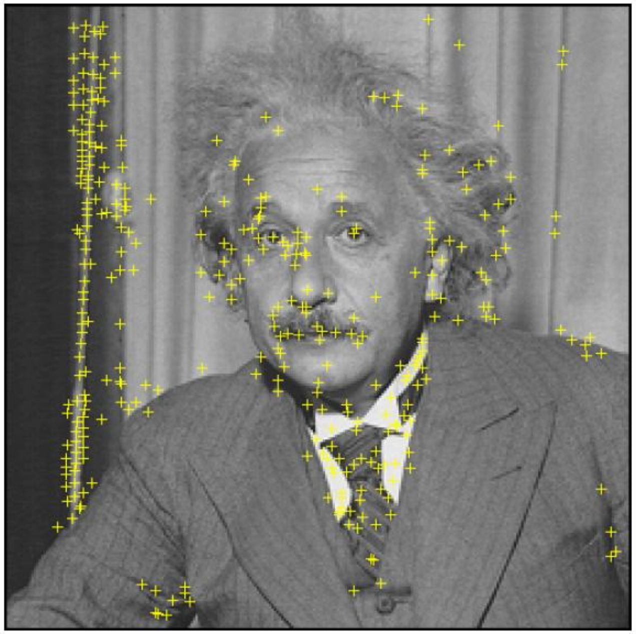
Kristen Grauman

Technical detail

- Reject points with low contrast
- Reject edge points using Hessian H (like corner detector)

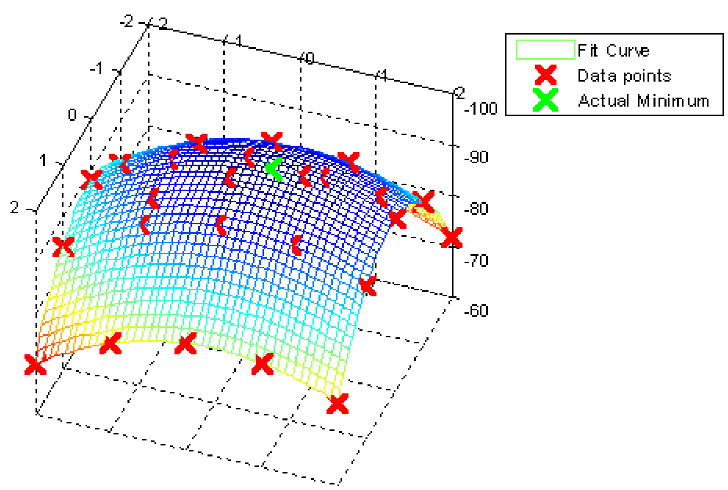
$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$

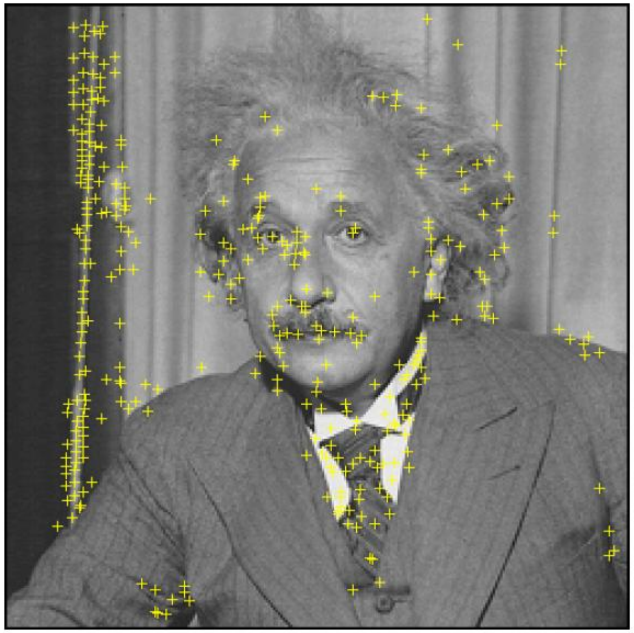




Technical detail

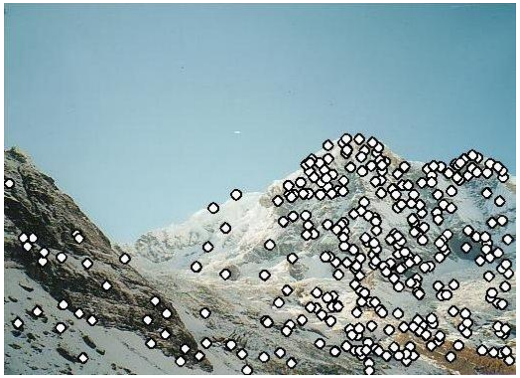
Accurate keypoint localization by fitting quadratic function for interpolation





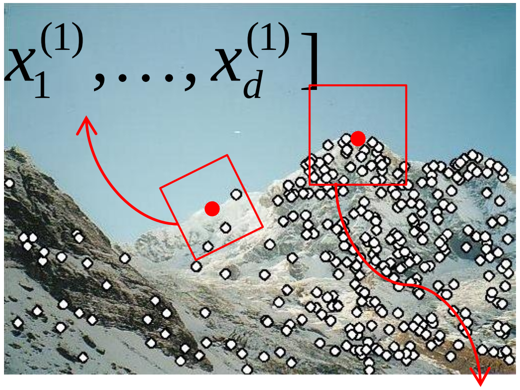
Local features and matching

1) Detection: Identify the interest points



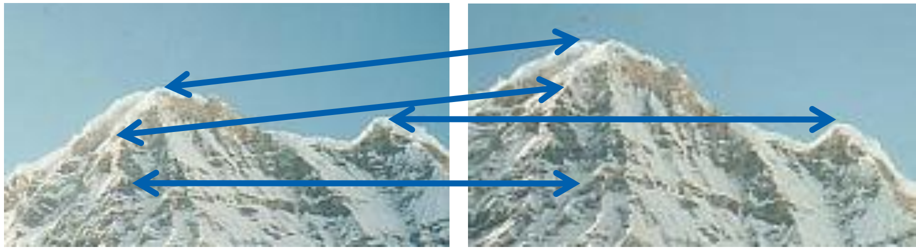
2) Description: Extract vector feature descriptor surrounding each interest point.

$$\mathbf{x}_1 = [x_1^{(1)}, \ldots, x_d^{(1)}]$$



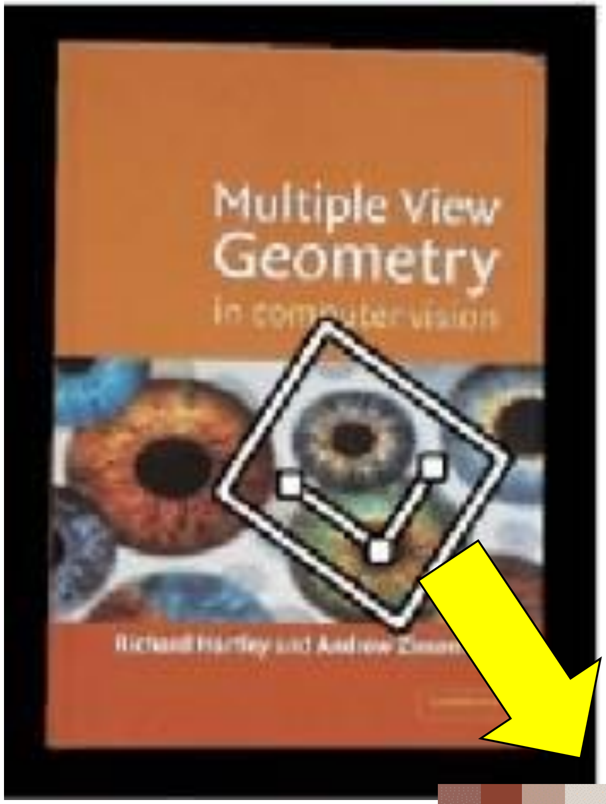
$$\mathbf{x}_2 = [x_1^{(2)}, \ldots, x_d^{(2)}]$$

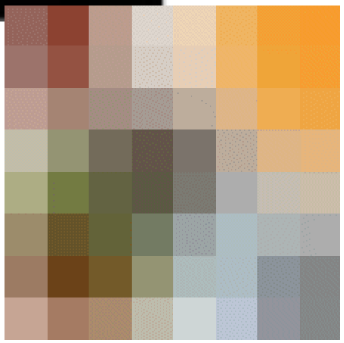
3) Matching: Determine correspondence between descriptors in two views

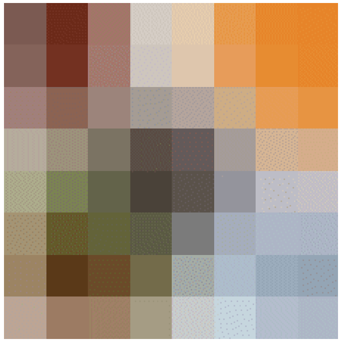


Kristen Grauman

Geometric transformations







e.g. scale,
translation,
rotation

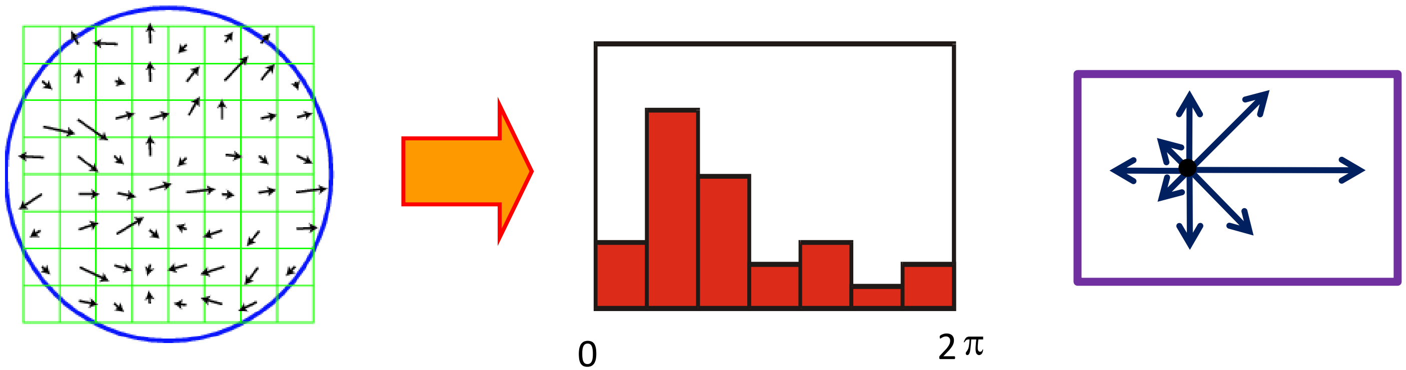
Kristen Grauman

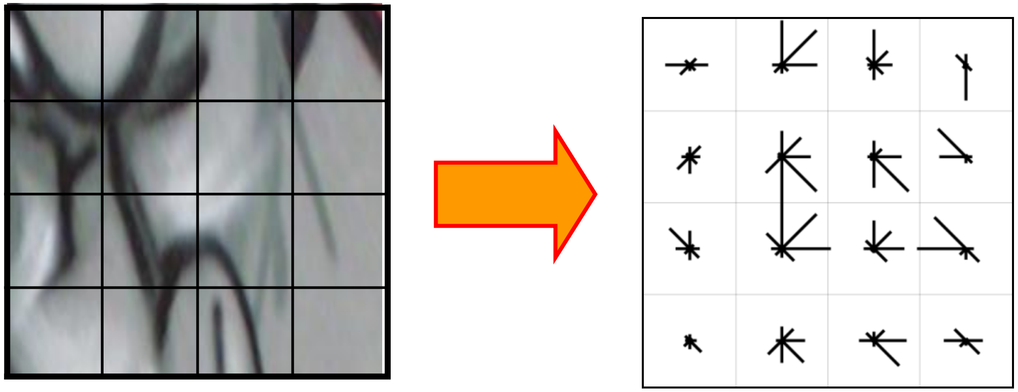
Photometric transformations

T. Tuytelaars

SIFT descriptor [Lowe 2004]

- Use histograms to bin pixels within sub-patches according to their orientation.

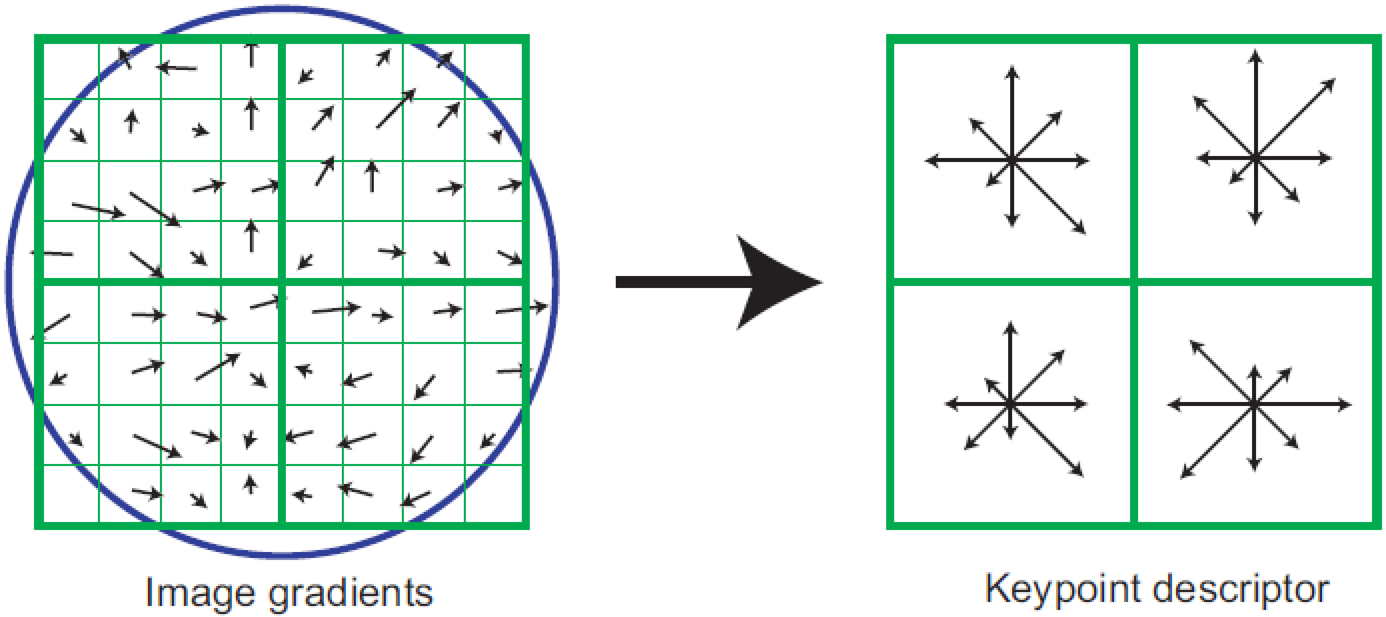




Kristen Grauman

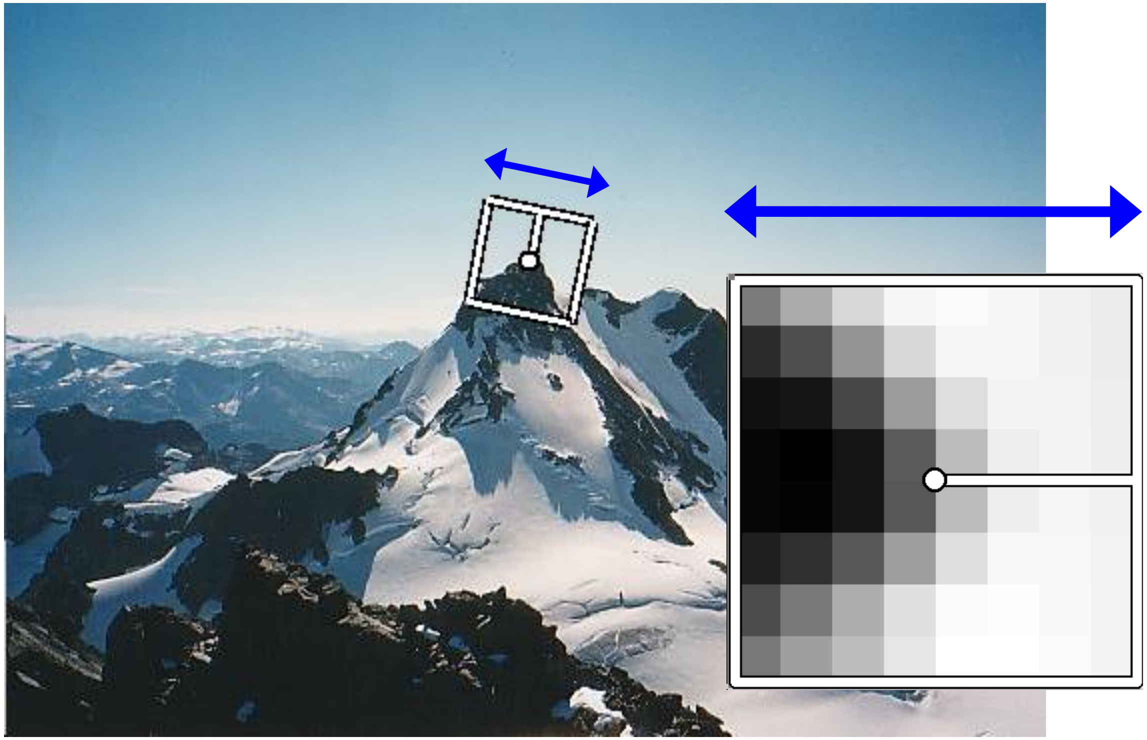
SIFT descriptor [Lowe 2004]

- 128 D vector (4x4x8)



- Normalize to 1 (robust to illumination changes)
- Entries above 0.2 are set to 0.2 (reduce impact of very strong gradients e.g. shadow)
- Normalize to 1

Making descriptor rotation invariant

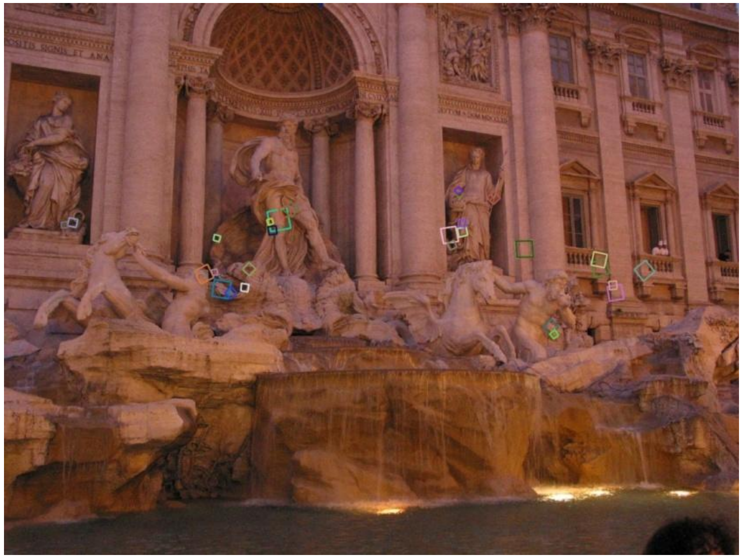


- Rotate patch according to its dominant gradient orientation
- This puts the patches into a canonical orientation.

Matthew Brown

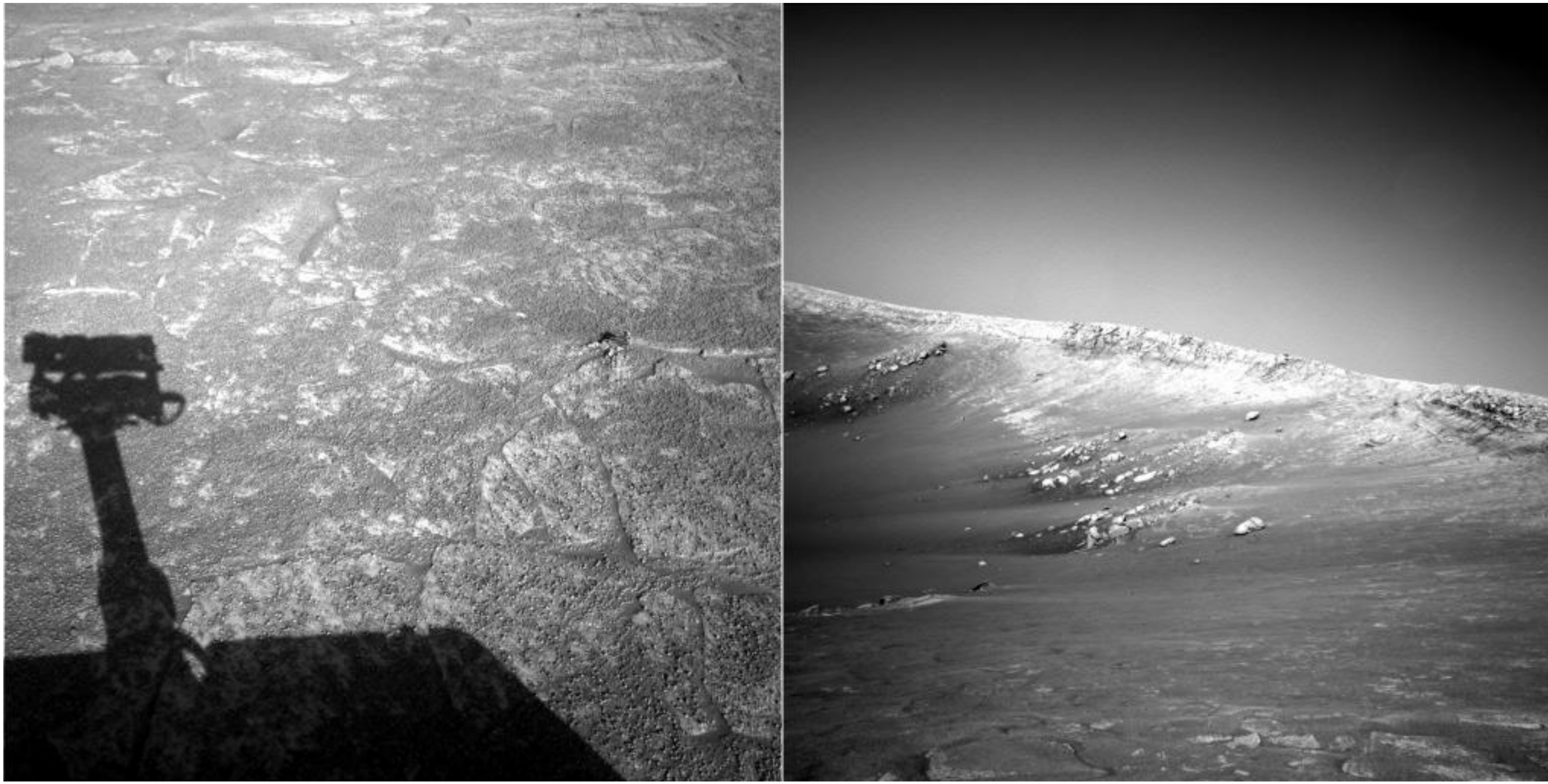
SIFT descriptor [Lowe 2004]

- Extraordinarily robust matching technique
  - Can handle changes in viewpoint
    - Up to about 60 degree out of plane rotation
  - Can handle significant changes in illumination
    - Sometimes even day vs. night (below)
  - Fast and efficient



Steve Seitz

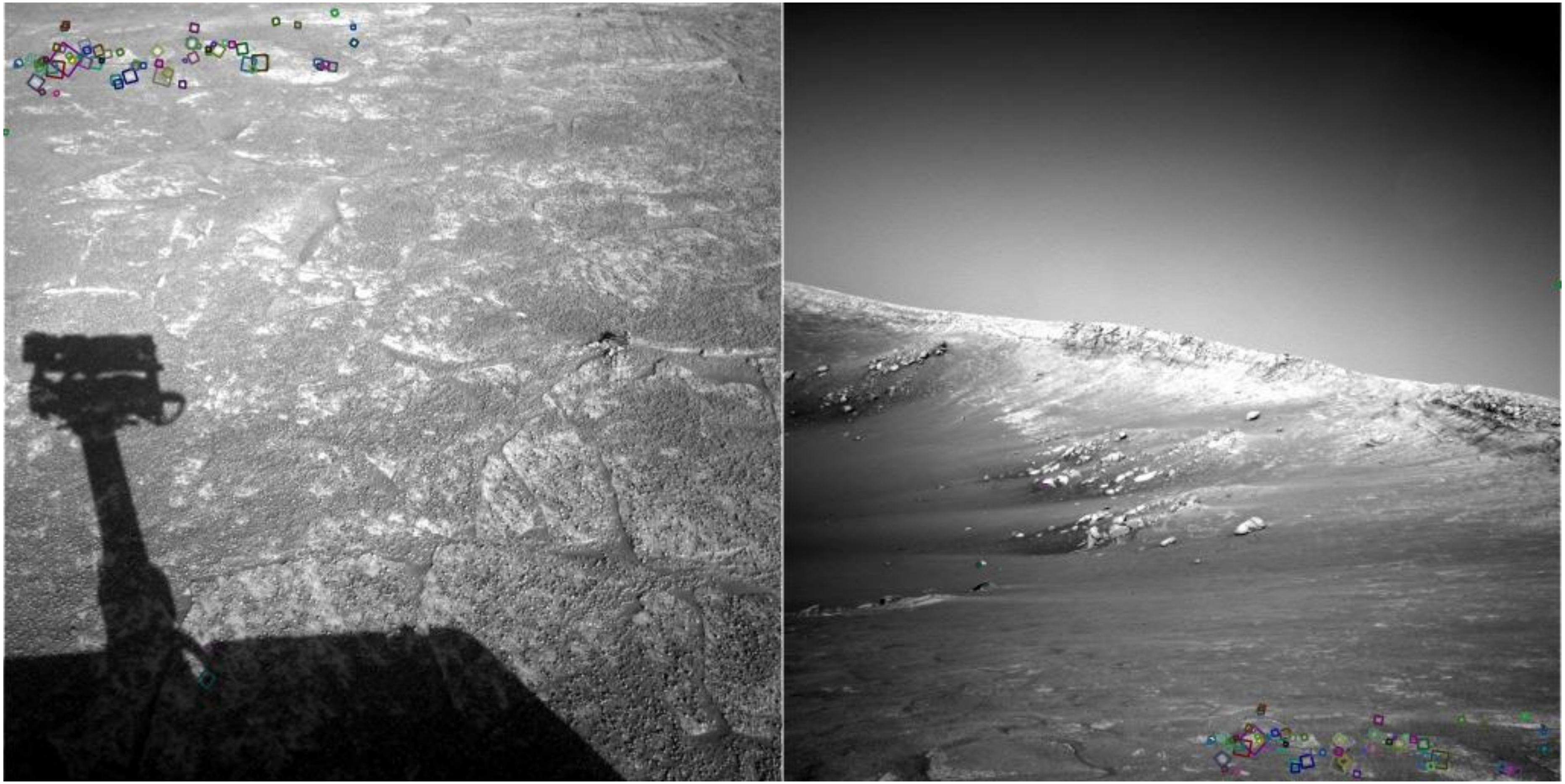
Example



NASA Mars Rover images

Kristen Grauman

Example



NASA Mars Rover images
with SIFT feature matches

Noah Snavely

SIFT properties

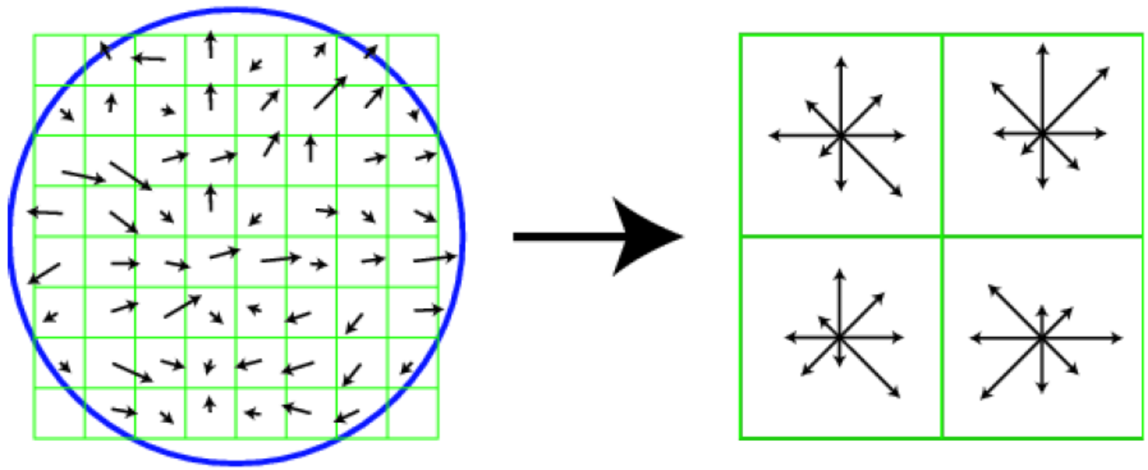
- Invariant to
  - Scale
  - Rotation

- Partially invariant to
  - Illumination changes
  - Camera viewpoint
  - Occlusion, clutter

Kristen Grauman

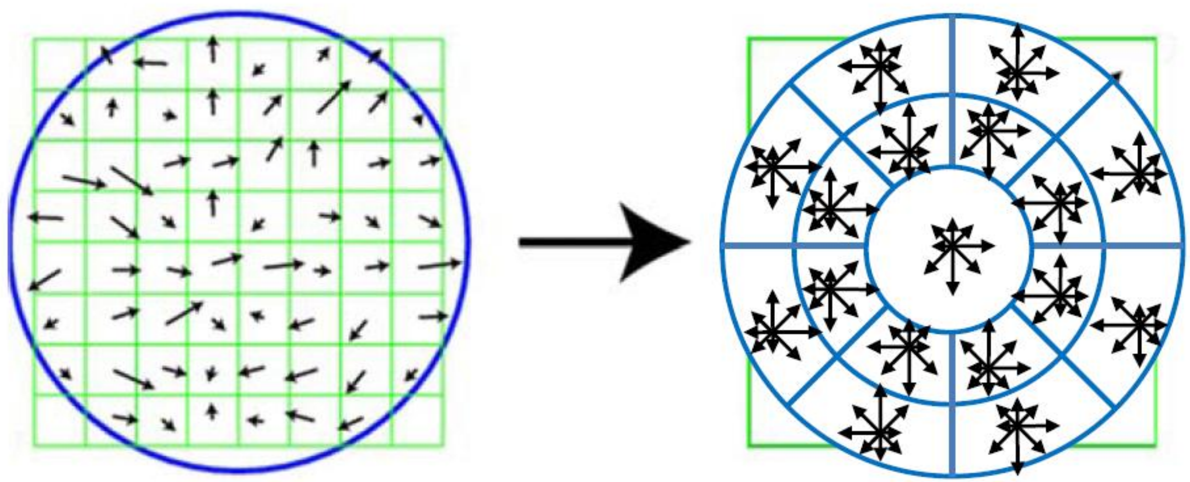
Other feature descriptors

SIFT:



(a) image gradients (b) keypoint descriptor

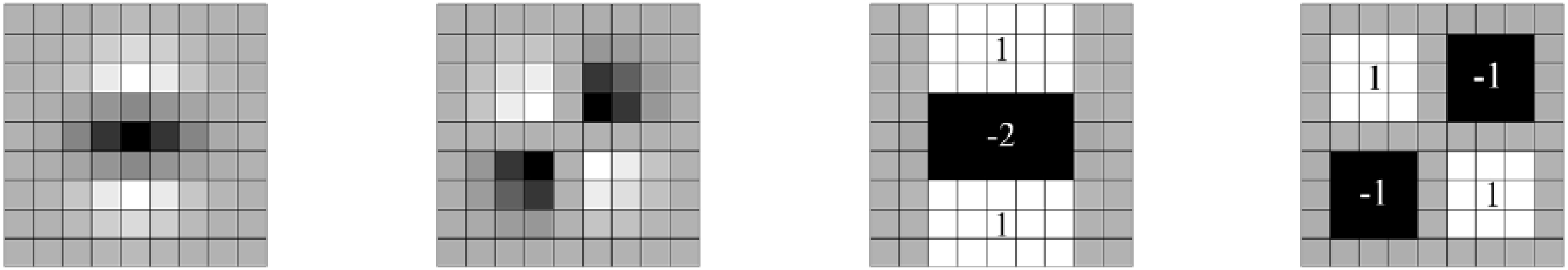
GLOH:



(a) image gradients (b) keypoint descriptor

Other feature descriptors

SURF: Approximate second order derivatives of Gaussian



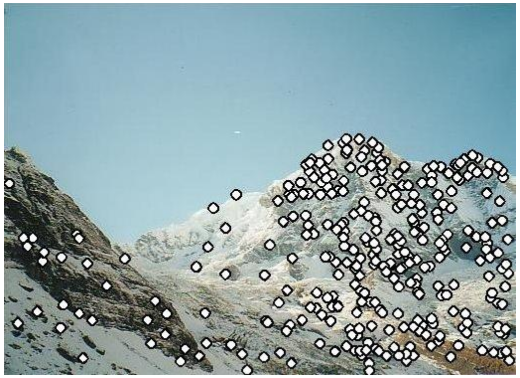
Fast to compute with integral images

Feature descriptors

- Countless number of SIFT variations
- SIFT sometimes too slow, many approximations for real-time applications
- Color versions exist
- Most popular SIFT, SURF

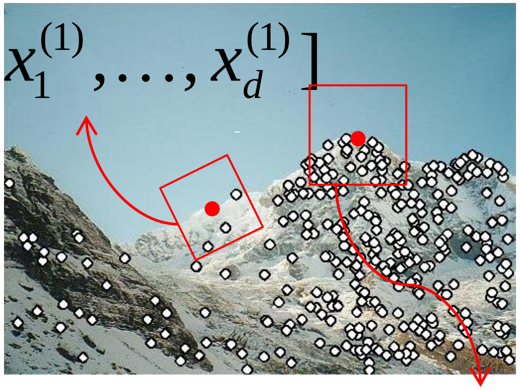
Local features and matching

1) Detection: Identify the interest points



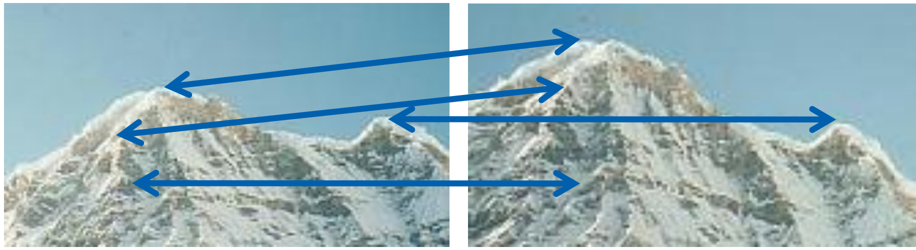
2) Description: Extract vector feature descriptor surrounding each interest point.

$$\mathbf{x}_1 = [x_1^{(1)}, \ldots, x_d^{(1)}]$$



$$\mathbf{x}_2 = [x_1^{(2)}, \ldots, x_d^{(2)}]$$

3) Matching: Determine correspondence between descriptors in two views



Kristen Grauman

Matching local features

Kristen Grauman

Matching local features

UNIVERSITÄT BONN

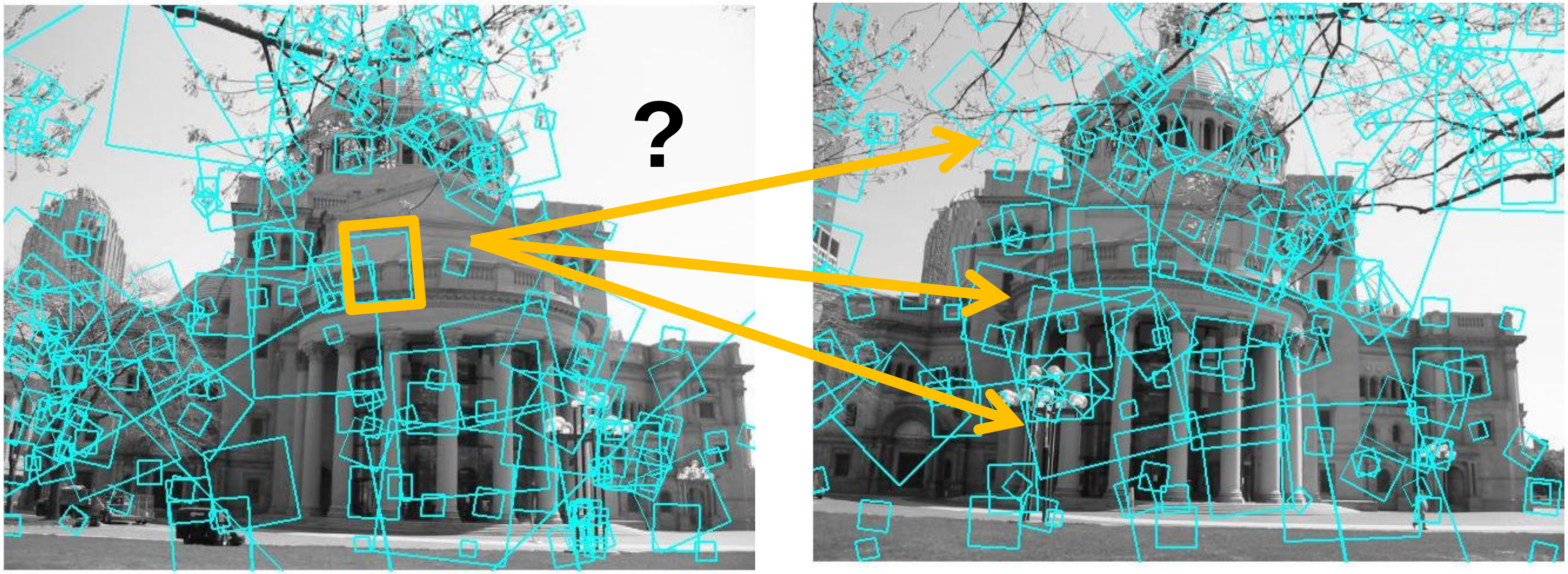


Image 1

Image 2

To generate **candidate matches**, find patches that have the most similar descriptor (e.g., Euclidean distance)

Simplest approach: compare them all, take the closest (or closest k, or within a thresholded distance)

Kristen Grauman

Matching local features

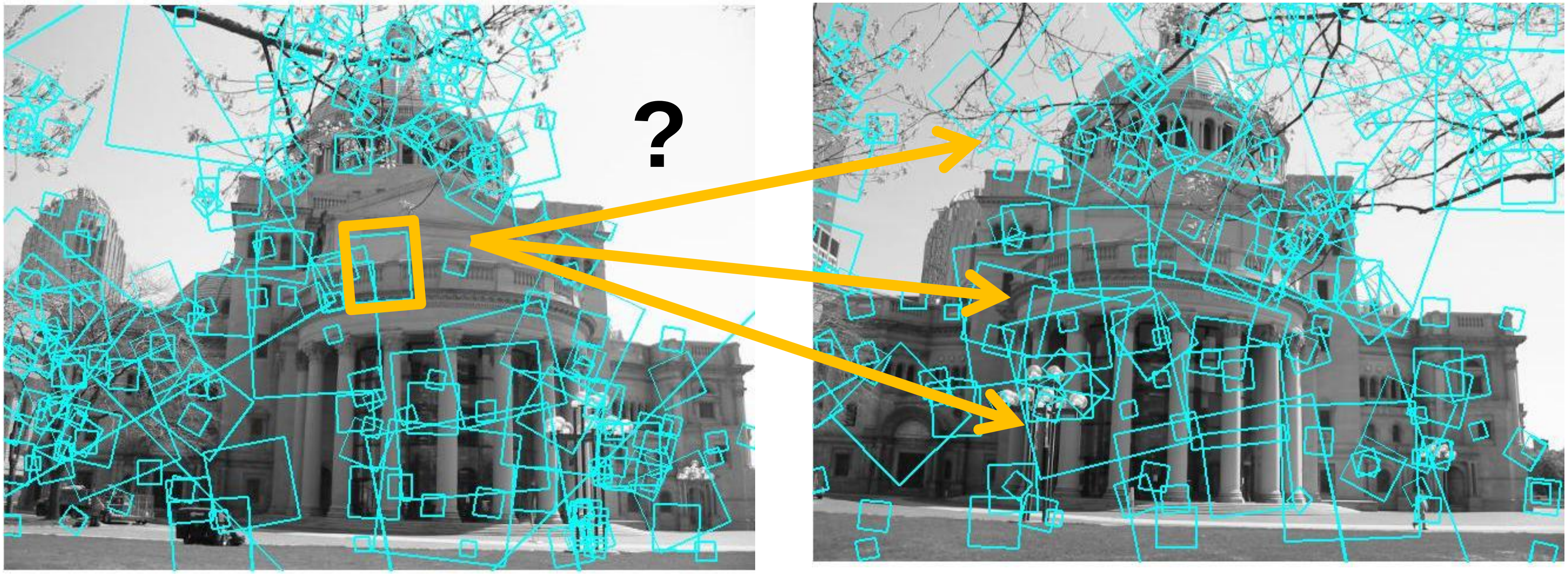


Image 1 Image 2

To add robustness to matching, can consider **ratio** : distance to best match / distance to second best match

If low, first match looks good.

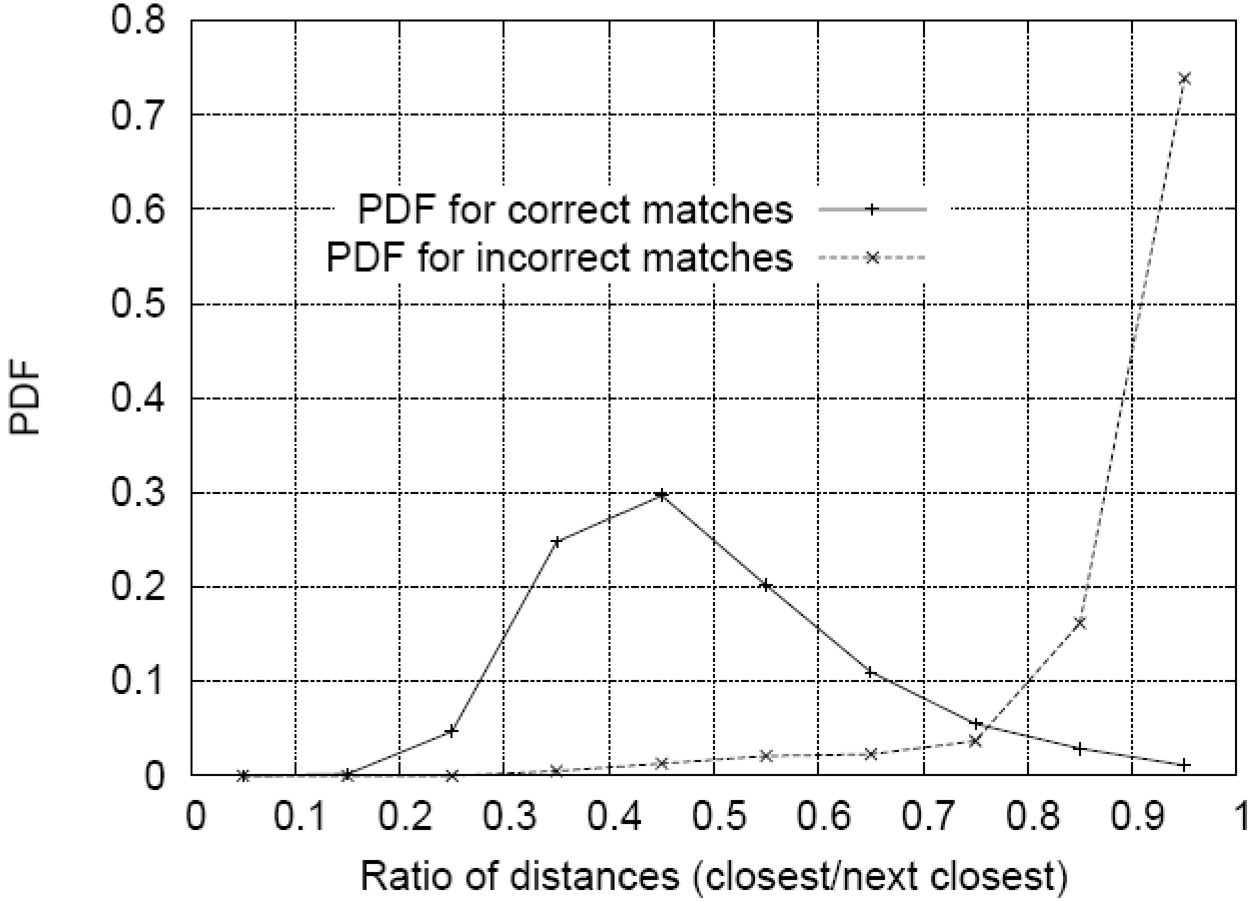
If high and close to one, could be ambiguous match.

Kristen Grauman

Matching SIFT Descriptors

UNIVERSITÄT BONN

- Nearest neighbor (Euclidean distance)
- Threshold ratio of nearest to 2nd nearest descriptor



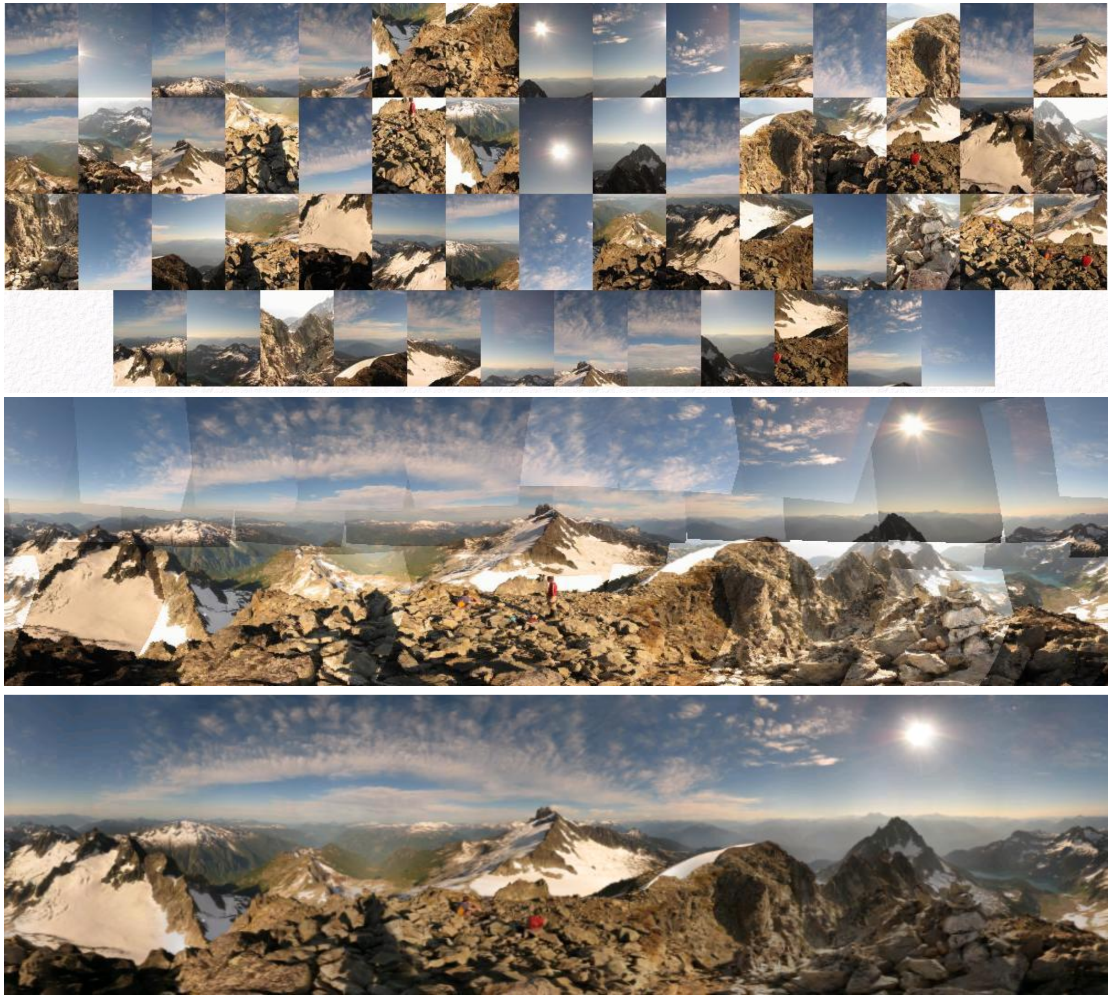
Lowe IJCV 2004

Applications of local invariant features

UNIVERSITÄT BONN

- Wide baseline stereo
- Motion tracking
- Panoramas
- Mobile robot navigation
- 3D reconstruction
- Recognition
- …

Automatic mosaicing



http://www.cs.ubc.ca/~mbrown/autostitch/autostitch.html

Wide baseline stereo

T. Tuytelaars