Local invariant feature detection

Tues Feb 27

Kristen Grauman

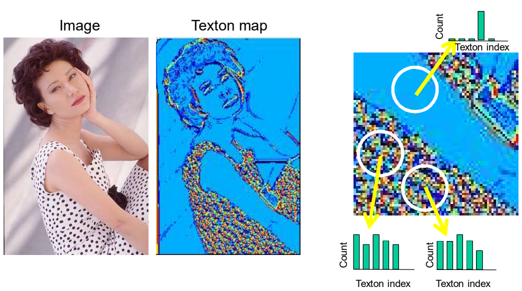
UT Austin

Survey feedback

- **Generally like**
  - Assignments
  - Topics
  - Lecture engaging, like examples, interactive nature
- Lecture can be fast
  - Would like discussion section, more review
  - Careful about tangential questions
  - Questions are on slides but answers not written there too
  - Would like to videotape lectures for review later
- Content:
  - Programming (would like more) vs. math (difficult)
- Grading: make sure fair partial credit
- Book can be difficult to follow
- Website:
  - Add direct link to current lecture (we have this)
  - Add TA emails (now added)
- My office hours
  - Schedule with me if you can't make standard window

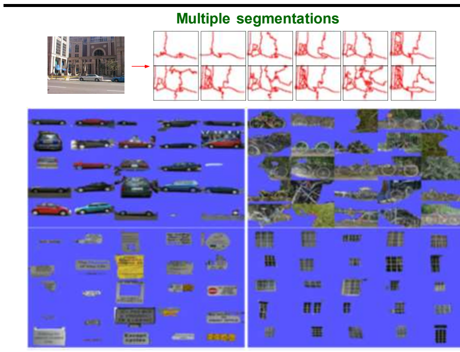
Review: Segmentation with texture

- Find "textons" by **clustering** vectors of filter bank outputs
- Describe texture in a window based on *texton histogram*



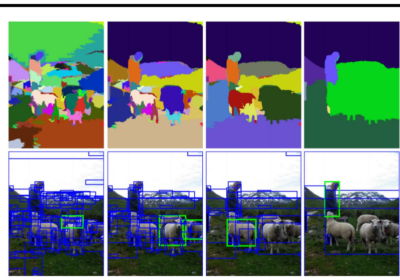
Malik, Belongie, Leung and Shi. IJCV 2001. Adapted from Lana Lazebnik

Segments as primitives for recognition



B. Russell et al., "Using Multiple Segmentations to Discover Objects and their Extent in Image Collections," CVPR 2006 Slide credit: Lana Lazebnik

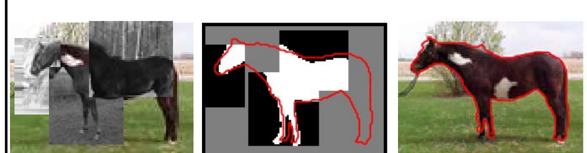
Category-agnostic object "proposals"



Selective search for object recognition. Uijlings et al. IJCV 2013.

Slide credit: Kristen Grauman

Top-down segmentation



E. Borenstein and S. Ullman, "Class-specific, top-down segmentation," ECCV 2002

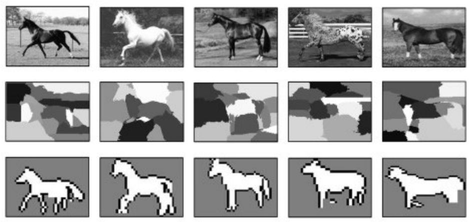
A. Levin and Y. Weiss, "Learning to Combine Bottom-Up and Top-Down Segmentation," ECCV 2006.

Slide credit: Lana Lazebnik

Top-down segmentation

Normalized cuts

Top-down segmentation

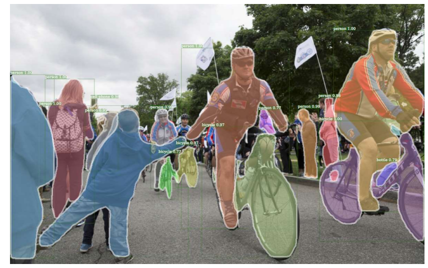


E. Borenstein and S. Ullman, "Class-specific, top-down segmentation," ECCV 2002

A. Levin and Y. Weiss, "Learning to Combine Bottom-Up and Top-Down Segmentation," ECCV 2006.

Slide credit: Lana Lazebnik

Joint segmentation and recognition

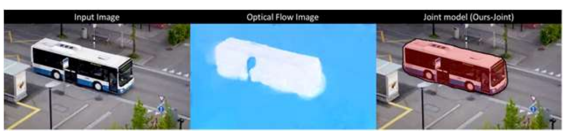


Mask R-CNN, K. He et al., ICCV 2017

Video object segmentation

Goal: Extract all foreground objects

- ✓ even those unseen during training
- ✓ without manual intervention.



`http://vision.cs.utexas.edu/projects/fusionseg/`

*S. Jain et al., **FusionSeg:** Learning to combine motion and appearance for fully automatic segmentation of generic objects in videos, CVPR 2017*

Slide credit: Kristen Grauman

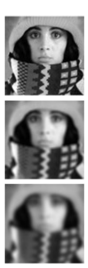
Interactive image and video segmentation

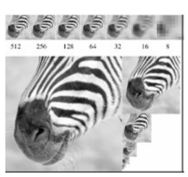
Results achieved with average of 2 user clicks

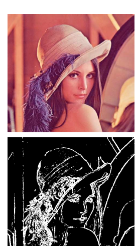
[Jain & Grauman, HCOMP 2016] Click Carving
https://github.com/suyogduttjain/click_carving

Slide credit: Kristen Grauman

Previously: Features and filters



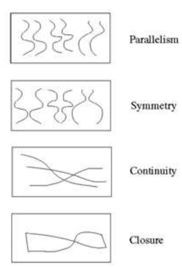


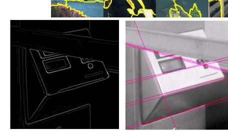


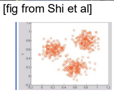
Transforming and describing images; textures, colors, edges

Slide credit: Kristen Grauman

Previously: Grouping & fitting





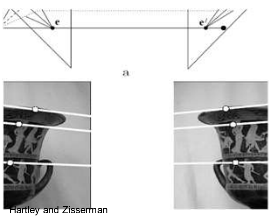


[fig from Shi et al]

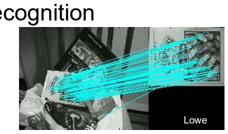
Clustering, segmentation, fitting; what parts belong together?

Slide credit: Kristen Grauman

Now: Multiple views

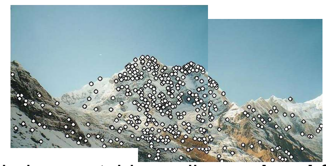


Matching, invariant features, stereo vision, instance recognition

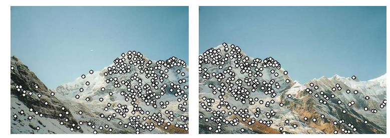


Slide credit: Kristen Grauman

Important tool for multiple views: Local features



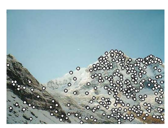
Multi-view matching relies on **local feature** correspondences.



How to detect *which local features* to match?

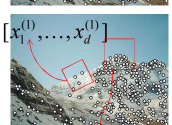
Local features: main components

1) Detection: Identify the interest points



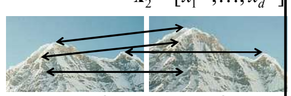
2) Description: Extract vector feature descriptor surrounding each interest point.

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



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

3) Matching: Determine correspondence between descriptors in two views



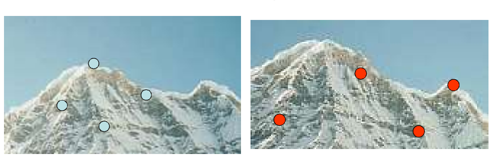
Slide credit: Kristen Grauman

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
- Compactness and efficiency
  - Many fewer features than image pixels
- Locality
  - A feature occupies a relatively small area of the image; robust to clutter and occlusion

Goal: interest operator repeatability

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

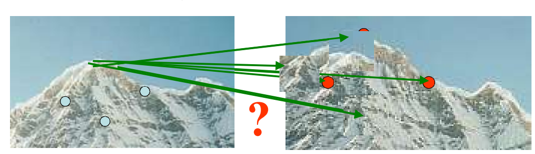


No chance to find true matches!

- Yet we have to be able to run the detection procedure *independently* per image.

Goal: descriptor distinctiveness

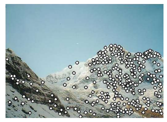
- We want to be able to reliably determine which point goes with which.



- Must provide some invariance to geometric and photometric differences between the two views.

Local features: main components

1) Detection: Identify the interest points



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

3) Matching: Determine correspondence between descriptors in two views

Slide credit: Kristen Grauman

- What points would you choose?

Slide credit: Kristen Grauman

Detecting corners

Slide credit: Kristen Grauman

Detecting corners

Compute "cornerness" response at every pixel.

Slide credit: Kristen Grauman

Detecting corners

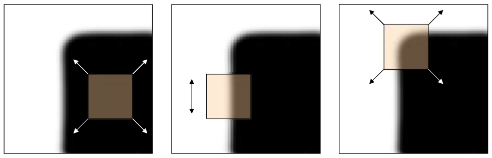
Slide credit: Kristen Grauman

Detecting local invariant features

- Detection of interest points
  - Harris corner detection
  - Scale invariant blob detection: LoG
- (Next time: description of local patches)

Corners as 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

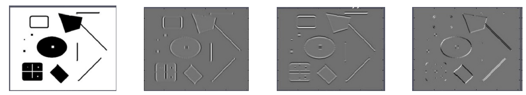
"corner": significant change in all directions

Slide credit: Alyosha Efros, Darya Frolova, Denis Simakov

Corners as distinctive interest points

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

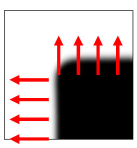
2 x 2 matrix of image derivatives (averaged in neighborhood of a point).



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

What does this matrix reveal?

First, consider an axis-aligned corner:



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

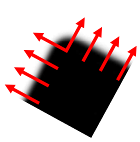
Look for locations where **both** λ's are large.

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?

What does this matrix reveal?

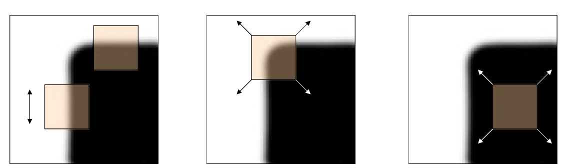
Since *M* is symmetric, we have $M = X \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} X^T$



$$Mx_i = \lambda_i x_i$$

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

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;

Cornerness score (other variants possible)

$$f = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$$

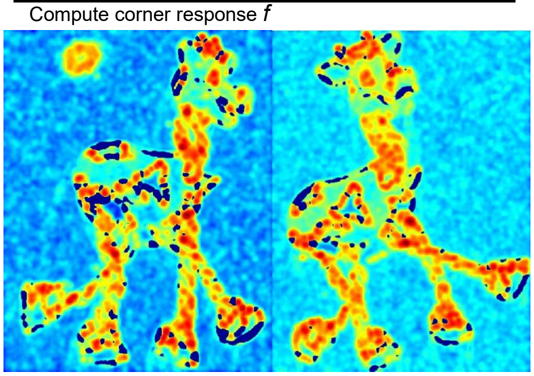
Harris corner detector

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

Harris Detector: Steps

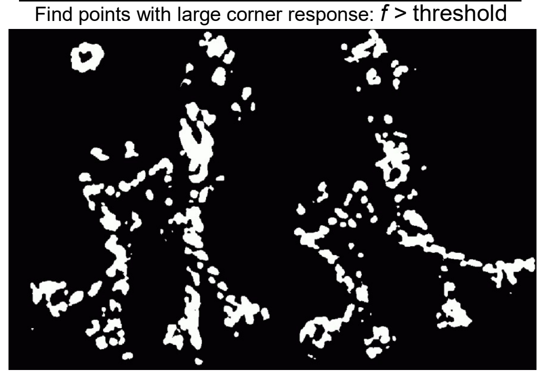
Harris Detector: Steps

Compute corner response *f*



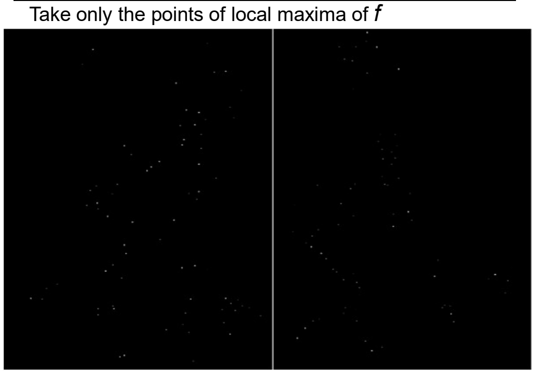
Harris Detector: Steps

Find points with large corner response: *f* > threshold

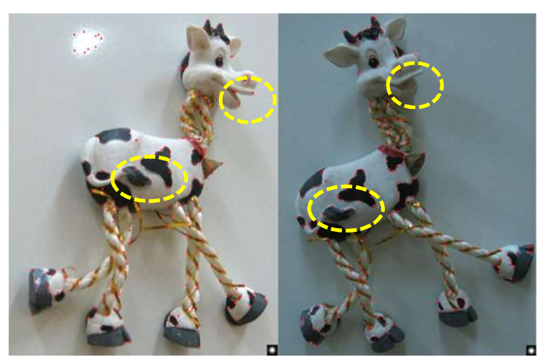


Harris Detector: Steps

Take only the points of local maxima of *f*



Harris Detector: Steps



Properties of the Harris corner detector

Rotation invariant? Yes

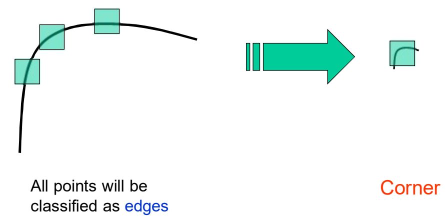
$$M = X \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} X^T$$

Scale invariant?

Properties of the Harris corner detector

Rotation invariant? Yes

Scale invariant? No



Scale invariant interest points

How can we independently select interest points in each image, such that the detections are repeatable across different scales?

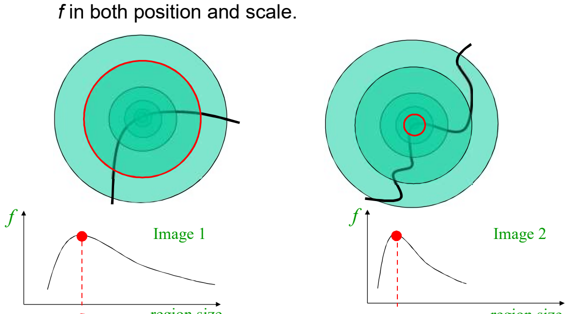
Automatic Scale Selection

How to find corresponding patch sizes, with only one image in hand?

K. Grauman, B. Leibe

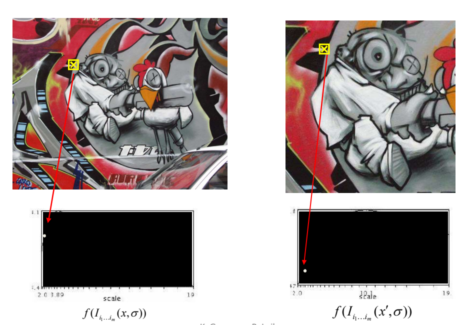
Automatic scale selection

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



Automatic Scale Selection

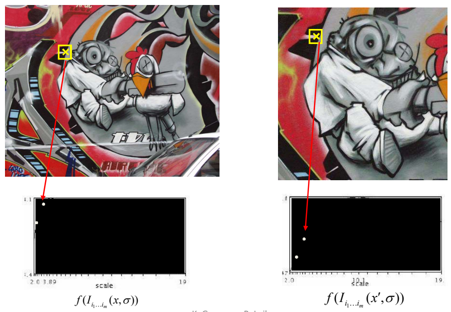
- Function responses for increasing scale (scale signature)



K. Grauman, B. Leibe

Automatic Scale Selection

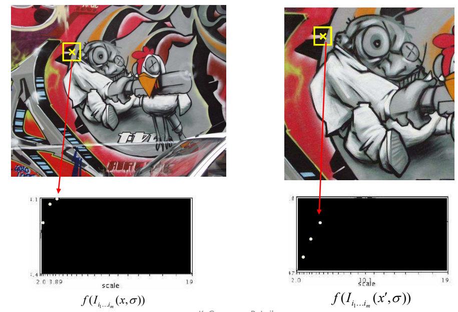
- Function responses for increasing scale (scale signature)



K. Grauman, B. Leibe

Automatic Scale Selection

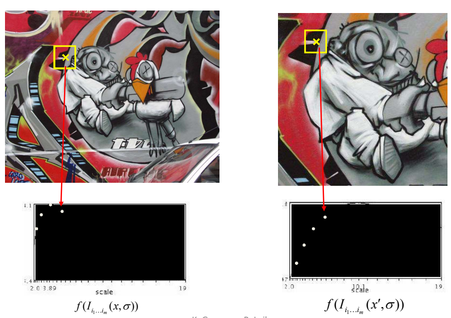
- Function responses for increasing scale (scale signature)



K. Grauman, B. Leibe

Automatic Scale Selection

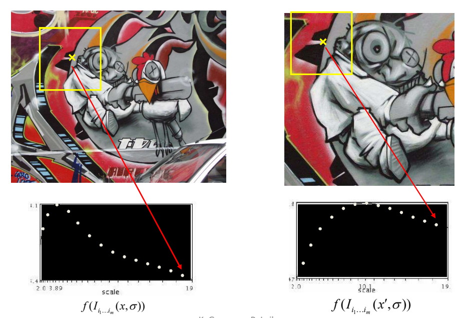
- Function responses for increasing scale (scale signature)



K. Grauman, B. Leibe

Automatic Scale Selection

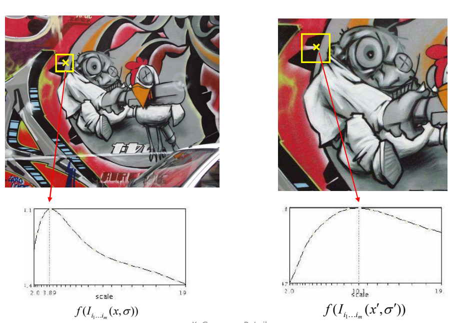
- Function responses for increasing scale (scale signature)



K. Grauman, B. Leibe

Automatic Scale Selection

- Function responses for increasing scale (scale signature)

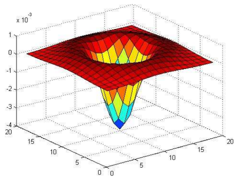


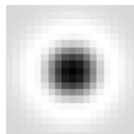
K. Grauman, B. Leibe

What can be the "signature" function?

Blob detection in 2D

Laplacian of Gaussian: Circularly symmetric operator for blob detection in 2D

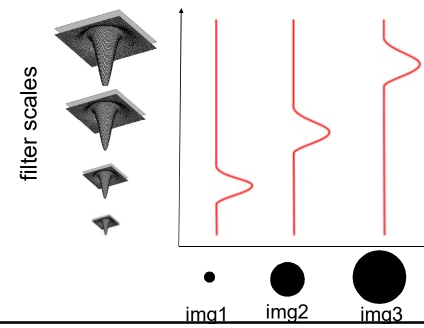




$$\nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$$

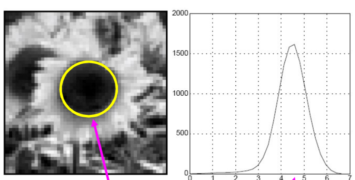
Blob detection in 2D: scale selection

Laplacian-of-Gaussian = “blob” detector $\nabla^2 g = \frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 g}{\partial y^2}$



Blob detection in 2D

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



characteristic scale

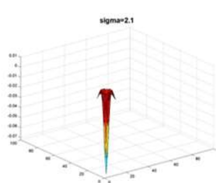
Slide credit: Lana Lazebnik

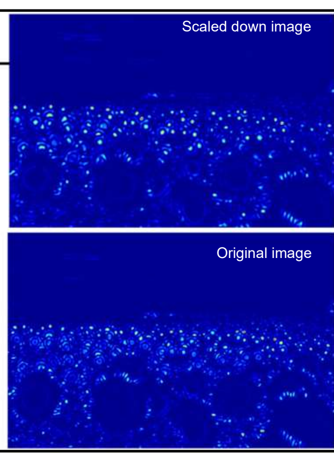
Example

Original image at ¾ the size

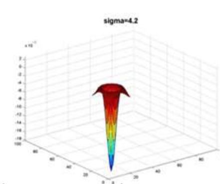
Slide credit: Kristen Grauman

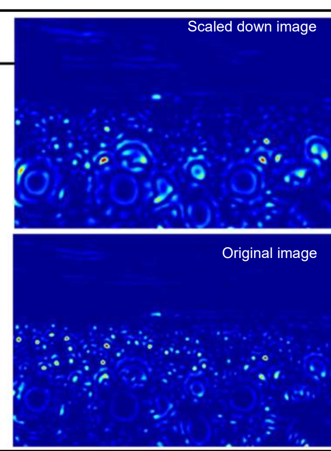
Original image at ¾ the size



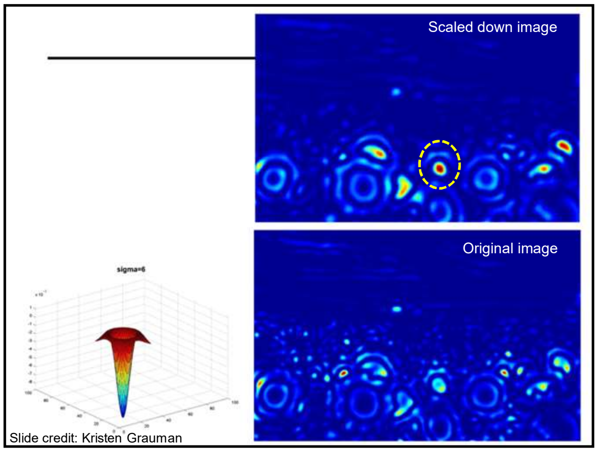


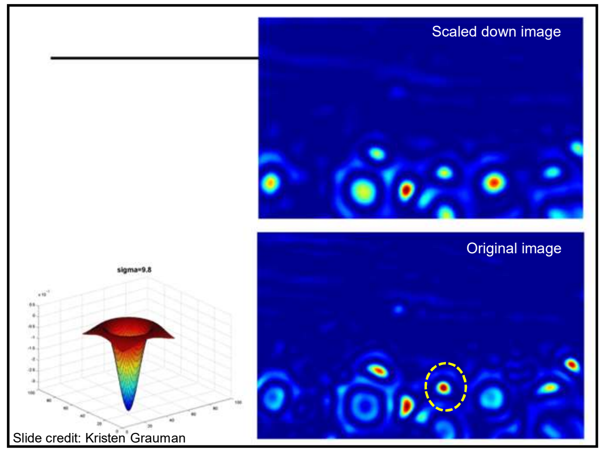
Slide credit: Kristen Grauman

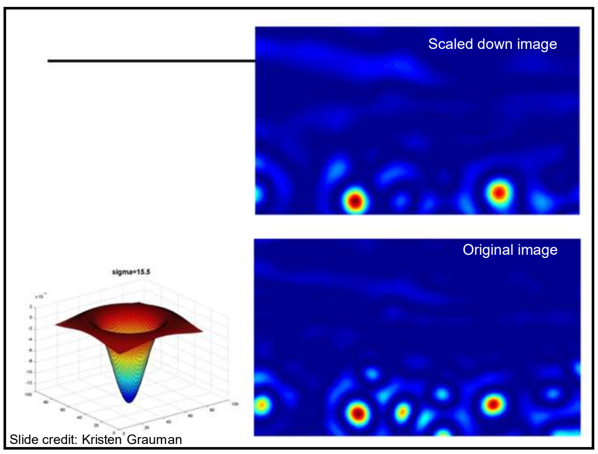


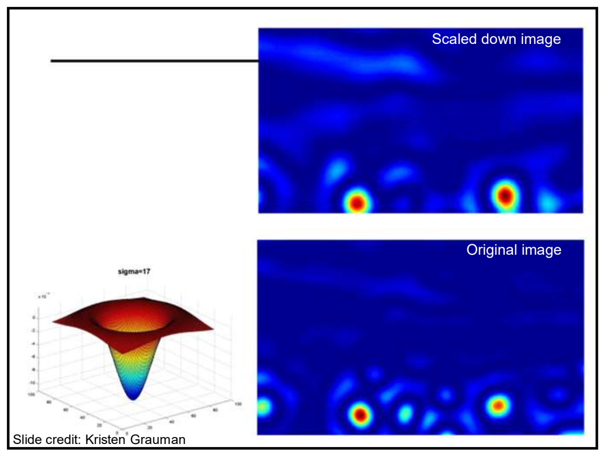


Slide credit: Kristen Grauman



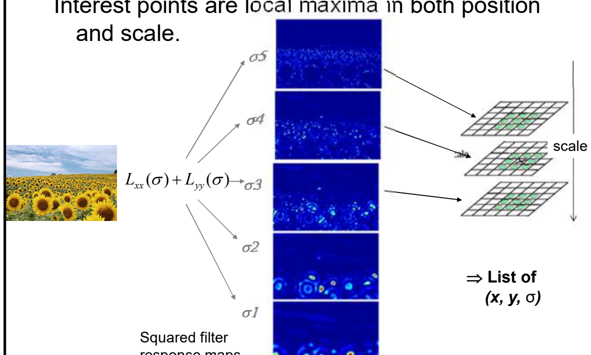






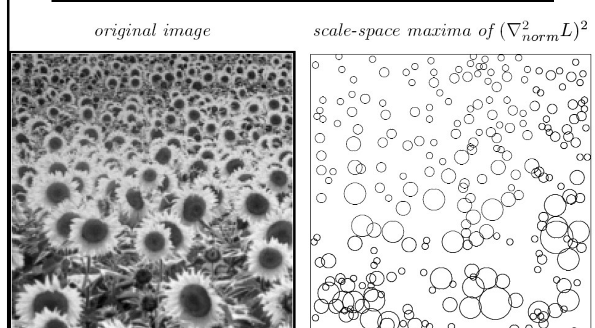
Scale invariant interest points

Interest points are local maxima in both position and scale.



Slide credit: Kristen Grauman

Scale-space blob detector: Example



T. Lindeberg. Feature detection with automatic scale selection. IJCV 1998.

Scale-space blob detector: Example

Image credit: Lana Lazebnik

Technical detail

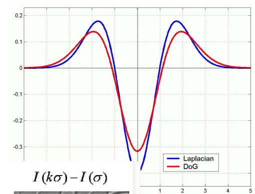
We can approximate the Laplacian with a difference of Gaussians; more efficient to implement.

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

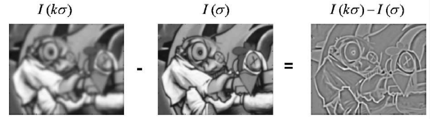
(Laplacian)

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

(Difference of Gaussians)



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



Summary

- Desirable properties for local features for correspondence
- Basic matching pipeline
- Interest point detection
  - Harris corner detector
  - Laplacian of Gaussian, automatic scale selection

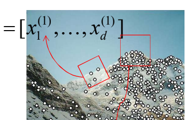
Local features: main components

1) Detection: Identify the interest points

NEXT TIME

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

Slide credit: Kristen Grauman