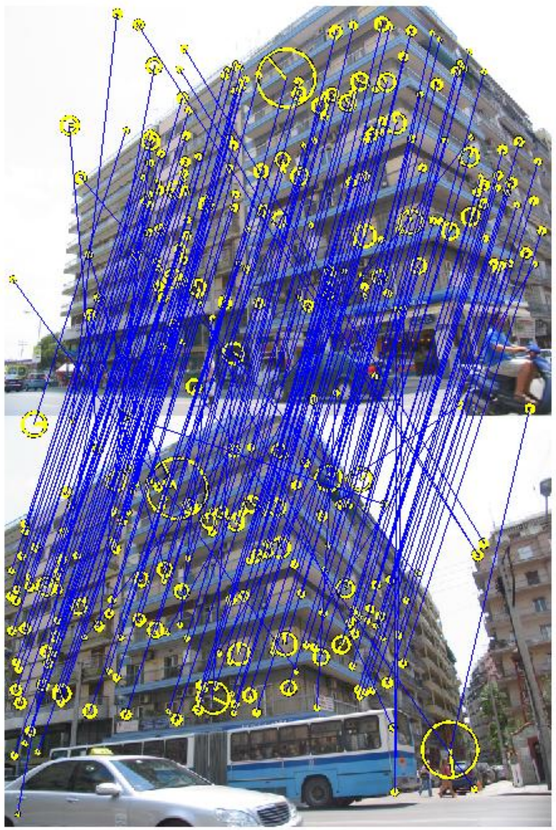
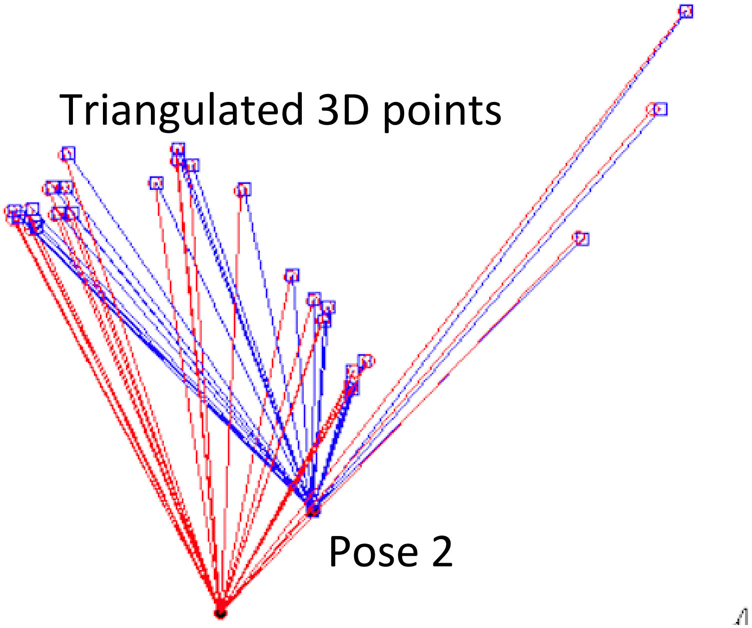
Perception: Visual Features

Kostas Daniilidis

What do we need features for?

- For finding points so that we solve localization and reconstruction





What do we need features for?

- For place recognition and other retrieval tasks

Query image

Five most similar image from a database

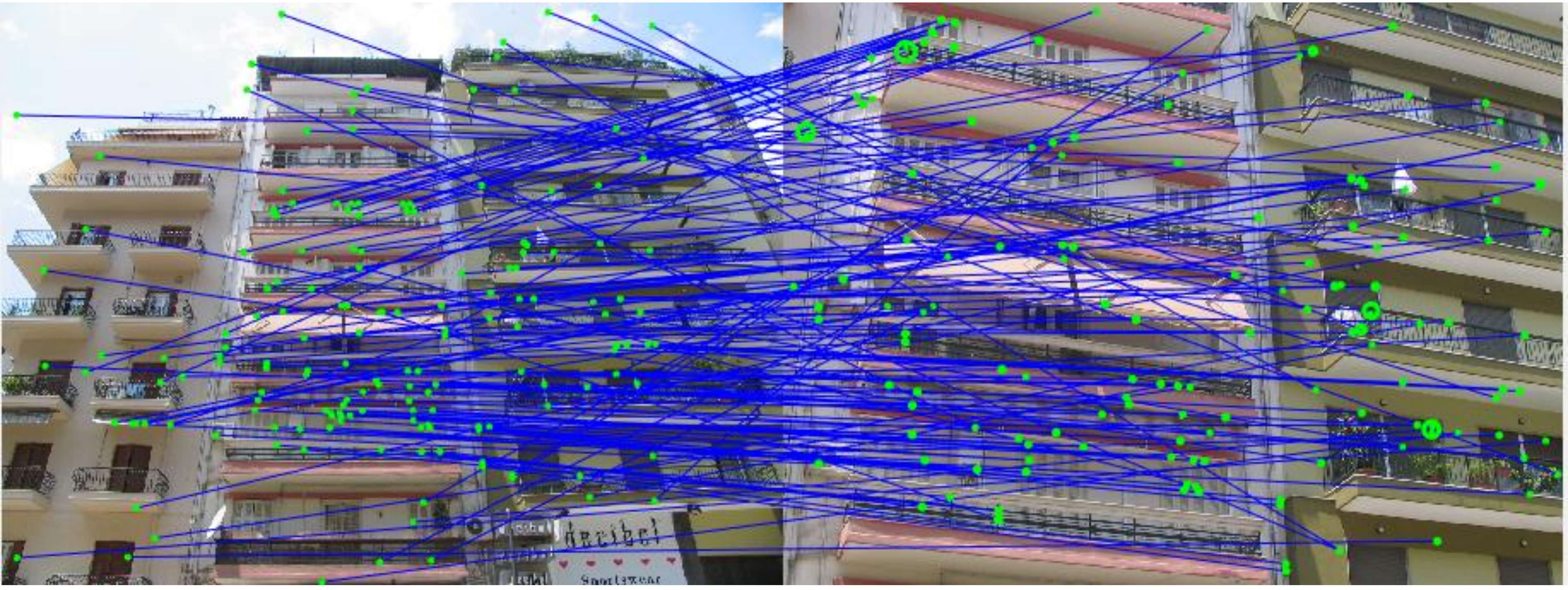
Where am I if I see this image

What we want from the features?

1. Detection repeatability:
   – When they are detected in one image to be detected in another one from the same scene even if image differs in scale and orientation

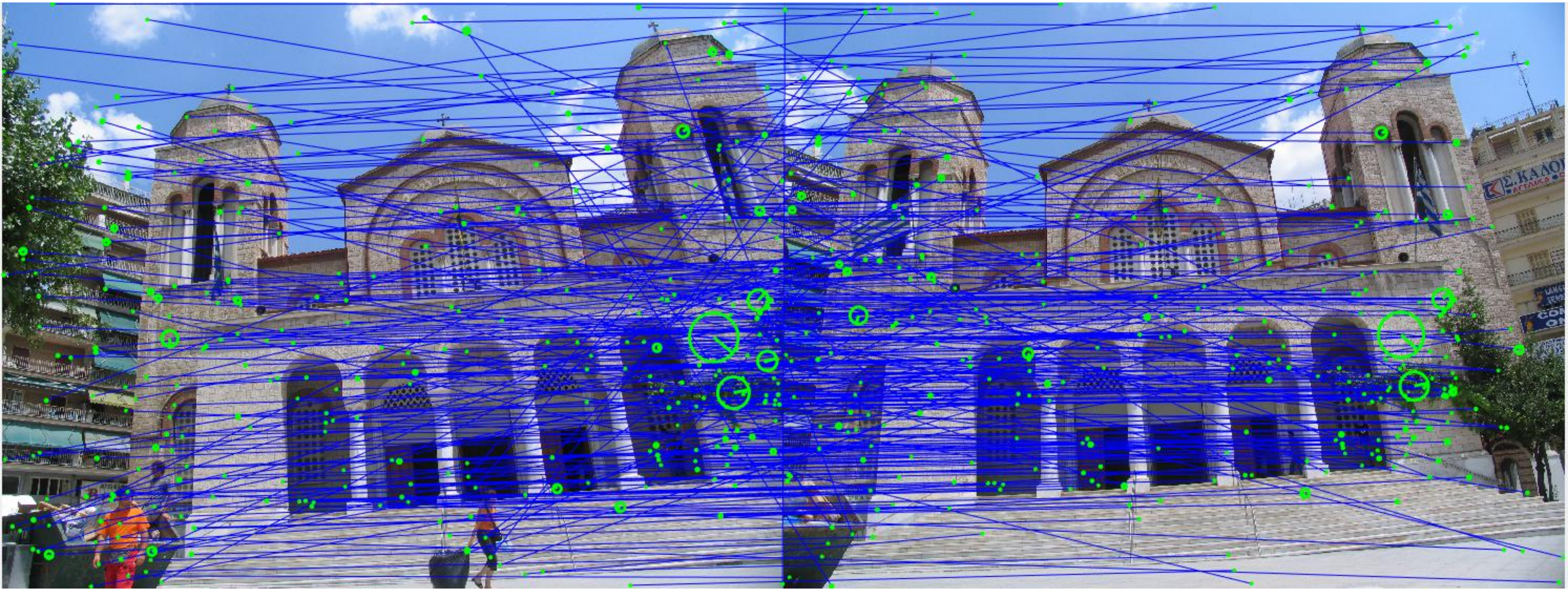
What we want from the features?

Features should be detected again even under severe scale changes. We call this property detection invariance.



What else do we want from features?

- We should be able to match features using a descriptor of the neighborhood.
- This descriptor should not change significantly under viewpoint changes like scale and rotation.
- We call this property descriptor invariance.



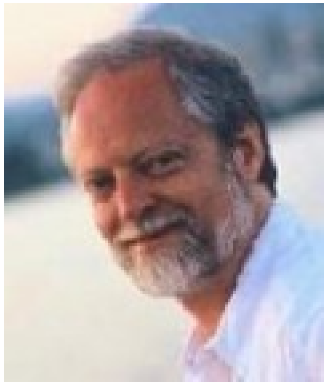
Invariant detection and description

Probably the most challenging of all properties is the scale invariance!


Tony Lindeberg

Feature detection with automatic scale selection

Authors	Tony Lindeberg
Publication date	1998/11/1
Journal	International Journal of Computer Vision
Volume	30
Issue	2


David Lowe

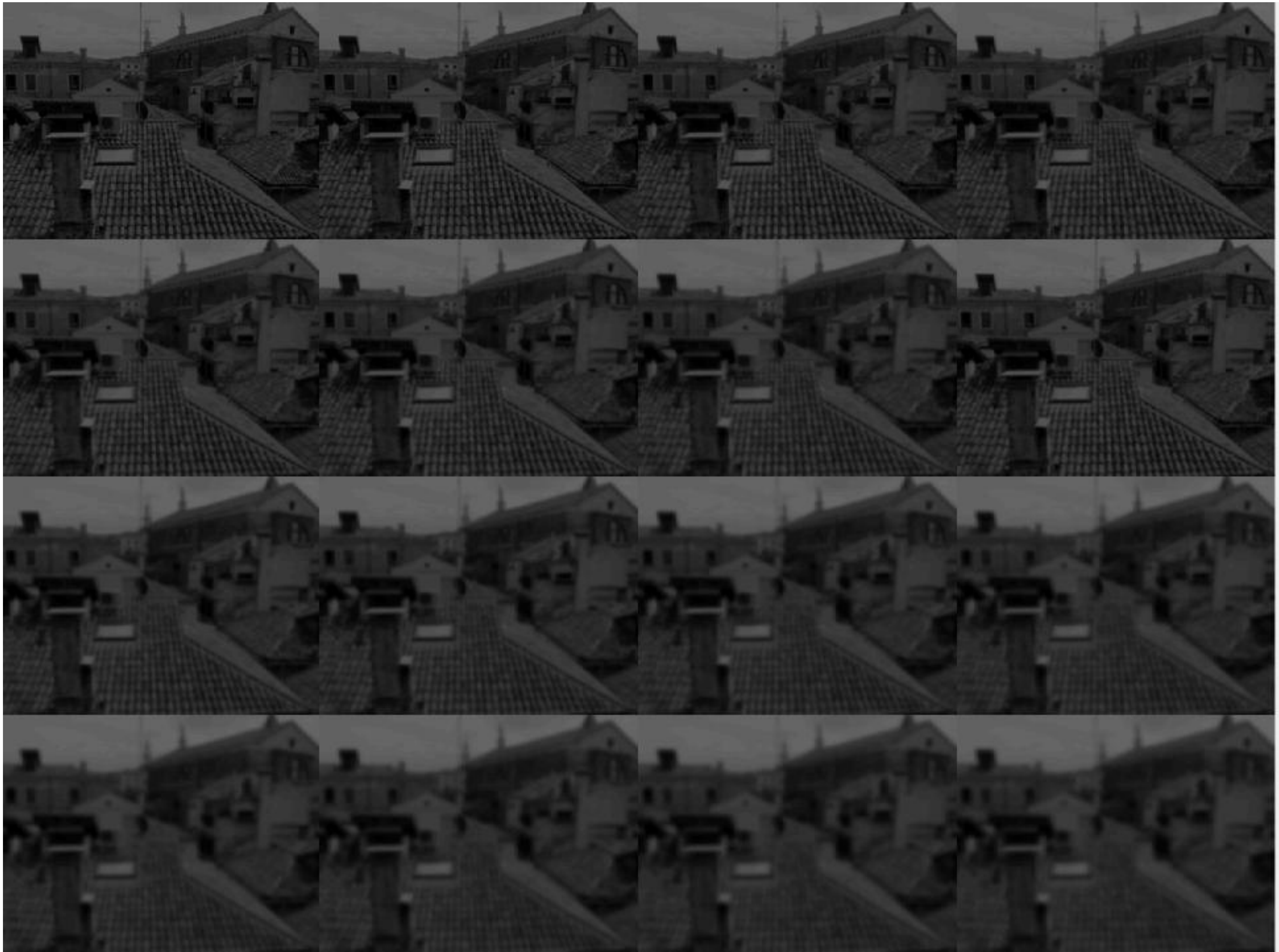
Distinctive image features from scale-invariant keypoints

Authors	David G Lowe
Publication date	2004/11/1
Journal	International journal of computer vision
Volume	60
Issue	2

SIFT: Scale Invariant Feature Transform

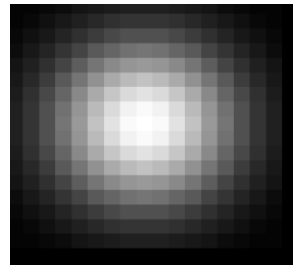
Title 1–20	Cited by	Year
Distinctive image features from scale-invariant keypoints DG Lowe International journal of computer vision 60 (2), 91-110	34581	2004
Object recognition from local scale-invariant features DG Lowe International Conference on Computer Vision, 1999, 1150-1157	10955	1999

The notion of scale space



How is scale space built?

$*$

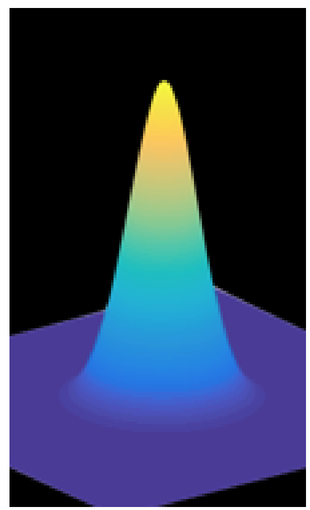


2D Gaussian

$=$

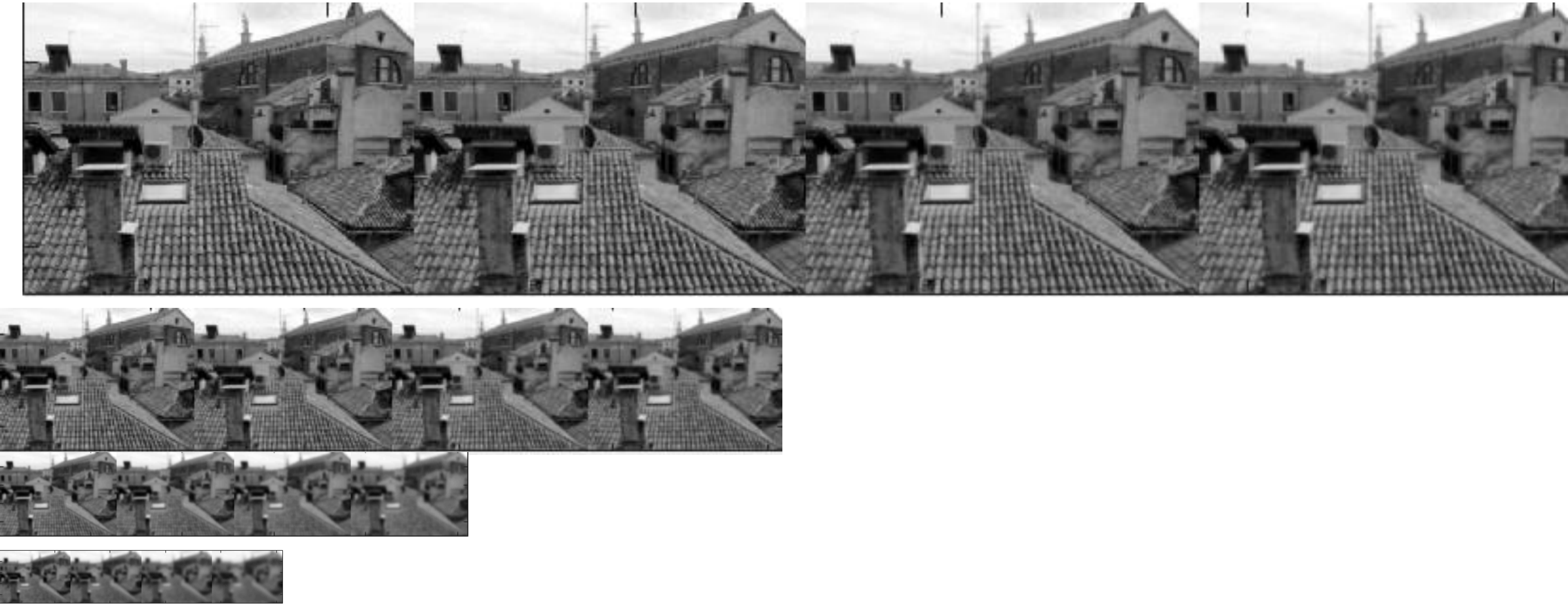
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

$*$ means convolution



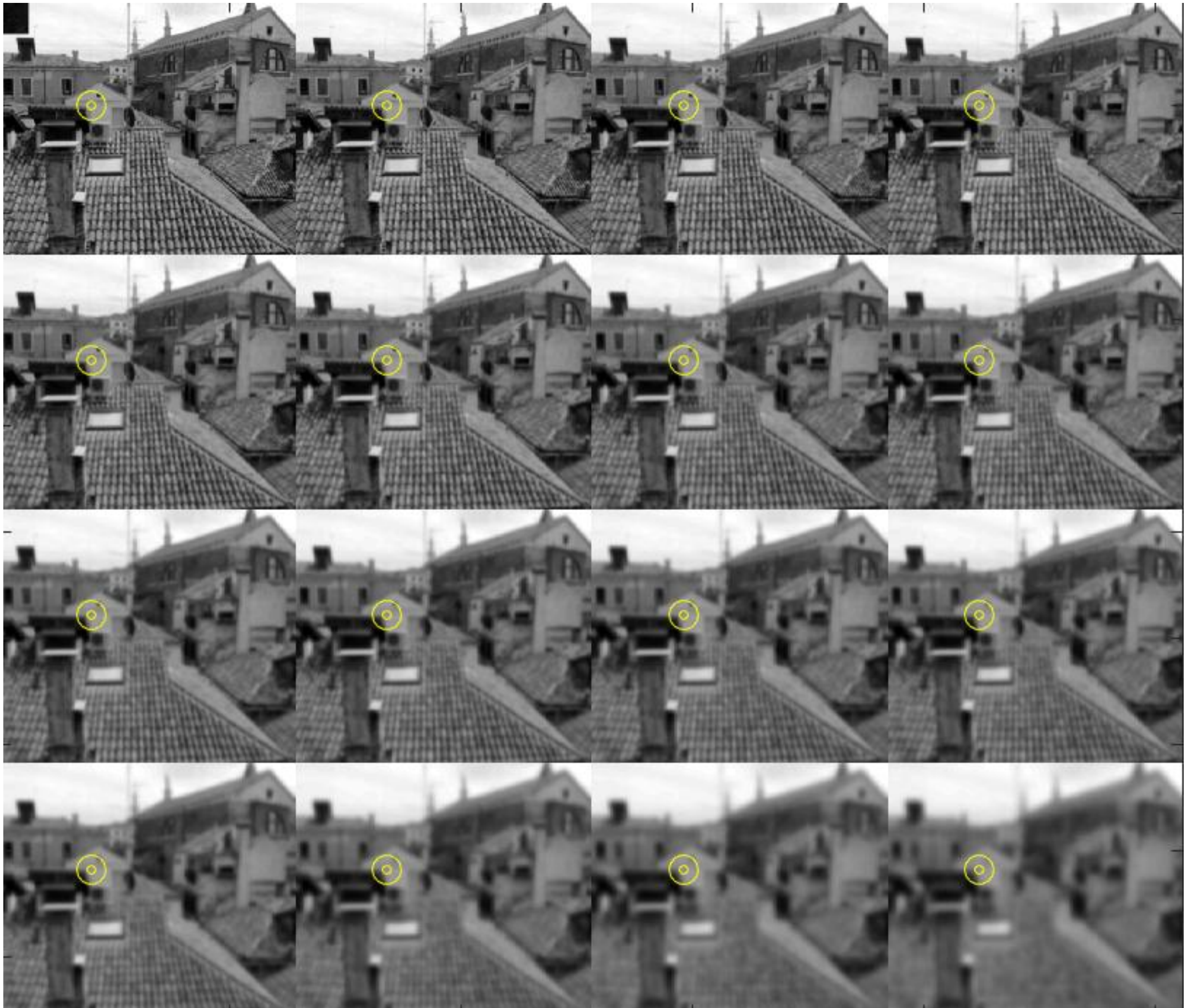
$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

The same scale but subsampled

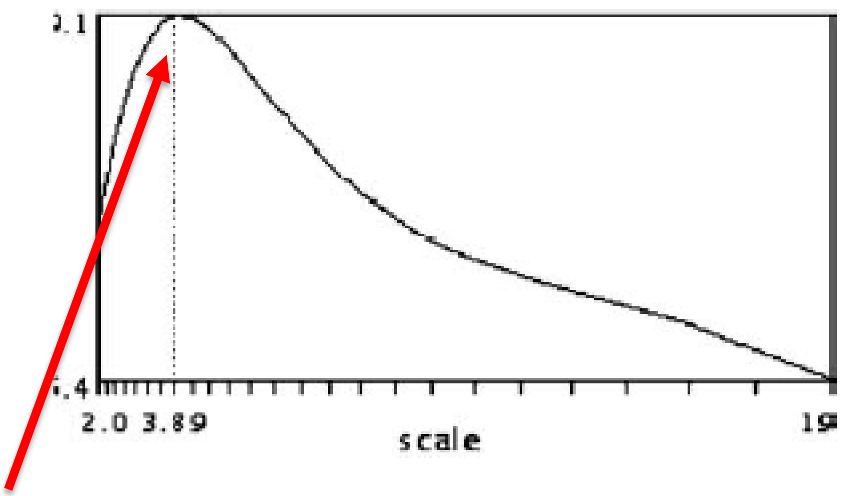


We subsampled it every time that the sigma of the Gaussian was doubled!

Now look at the same pixel across scale

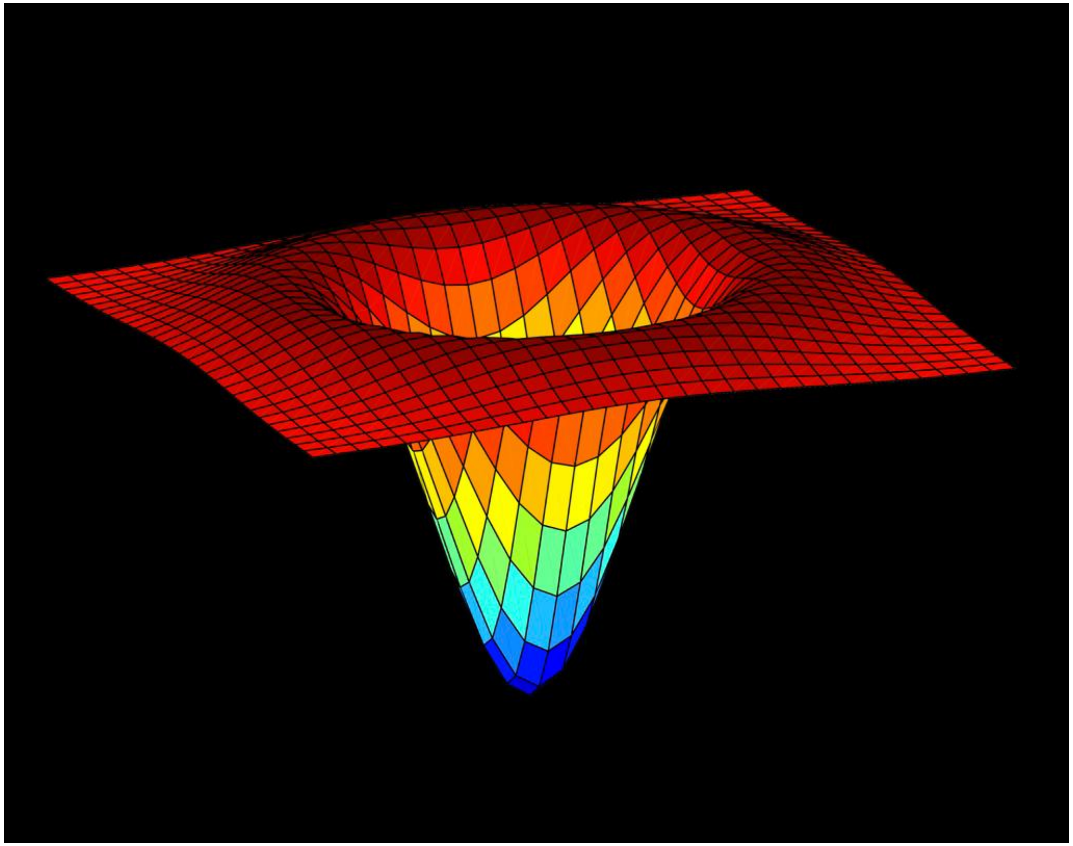


Scale selection



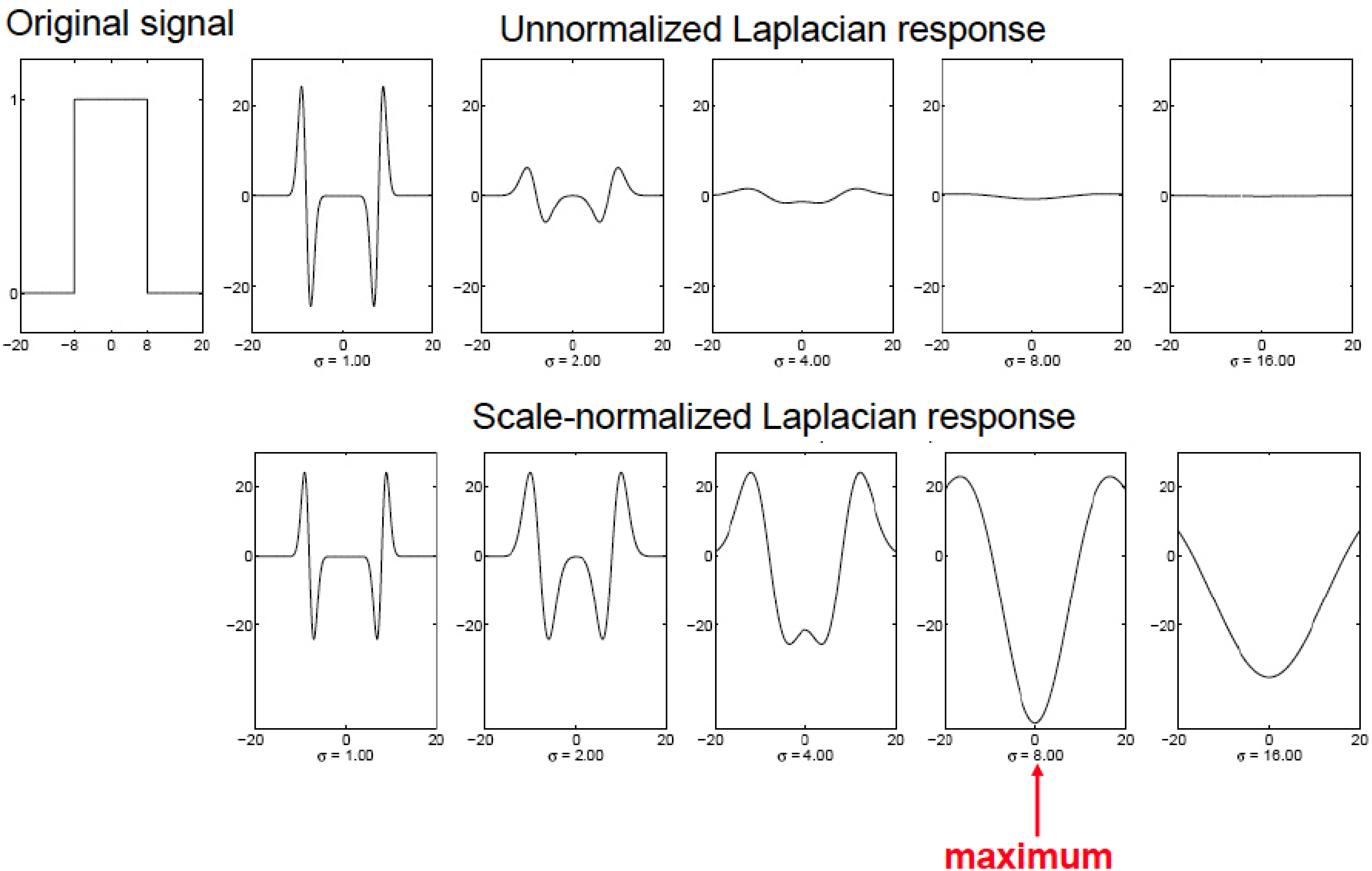
- The maximum across scale is the intrinsic scale of the image structure
- if the smoothed value is scale normalized.
- It turns out that only the derivatives of the Gaussian responses can be normalized.

We choose the 2nd derivative (trace of Hessian) , called Laplacian of Gaussian (LoG)



Which has the nice property that it can be approximated as the difference of two Gaussians and can detect blob like features!

Why do we need normalization?

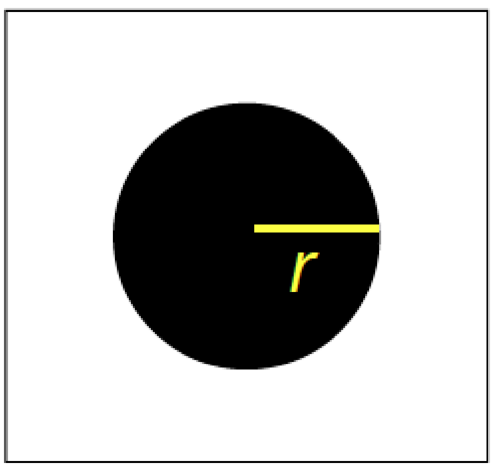


Scale selection

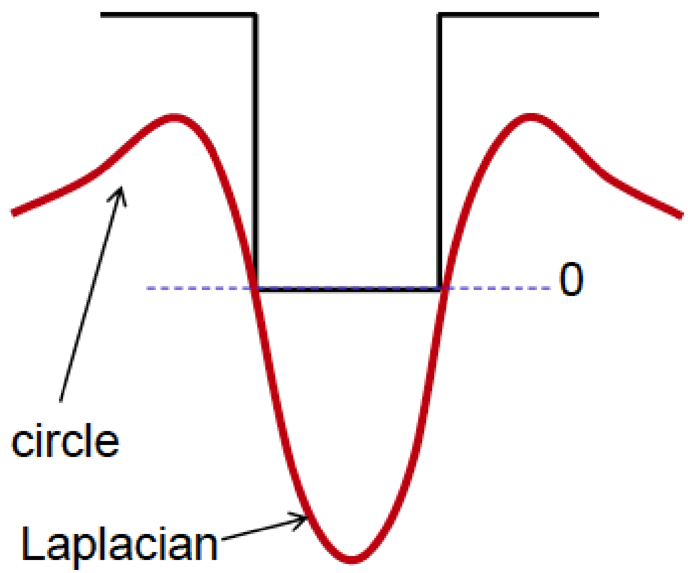
- At what scale does the Laplacian achieve a maximum response to a binary circle of radius r?
- To get maximum response, the zeros of the Laplacian have to be aligned with the circle
- The Laplacian is given by (up to scale):

$$(x^2 + y^2 - 2\sigma^2)\, e^{-(x^2+y^2)/2\sigma^2}$$

- Therefore, the maximum response occurs at $\sigma = r/\sqrt{2}$.

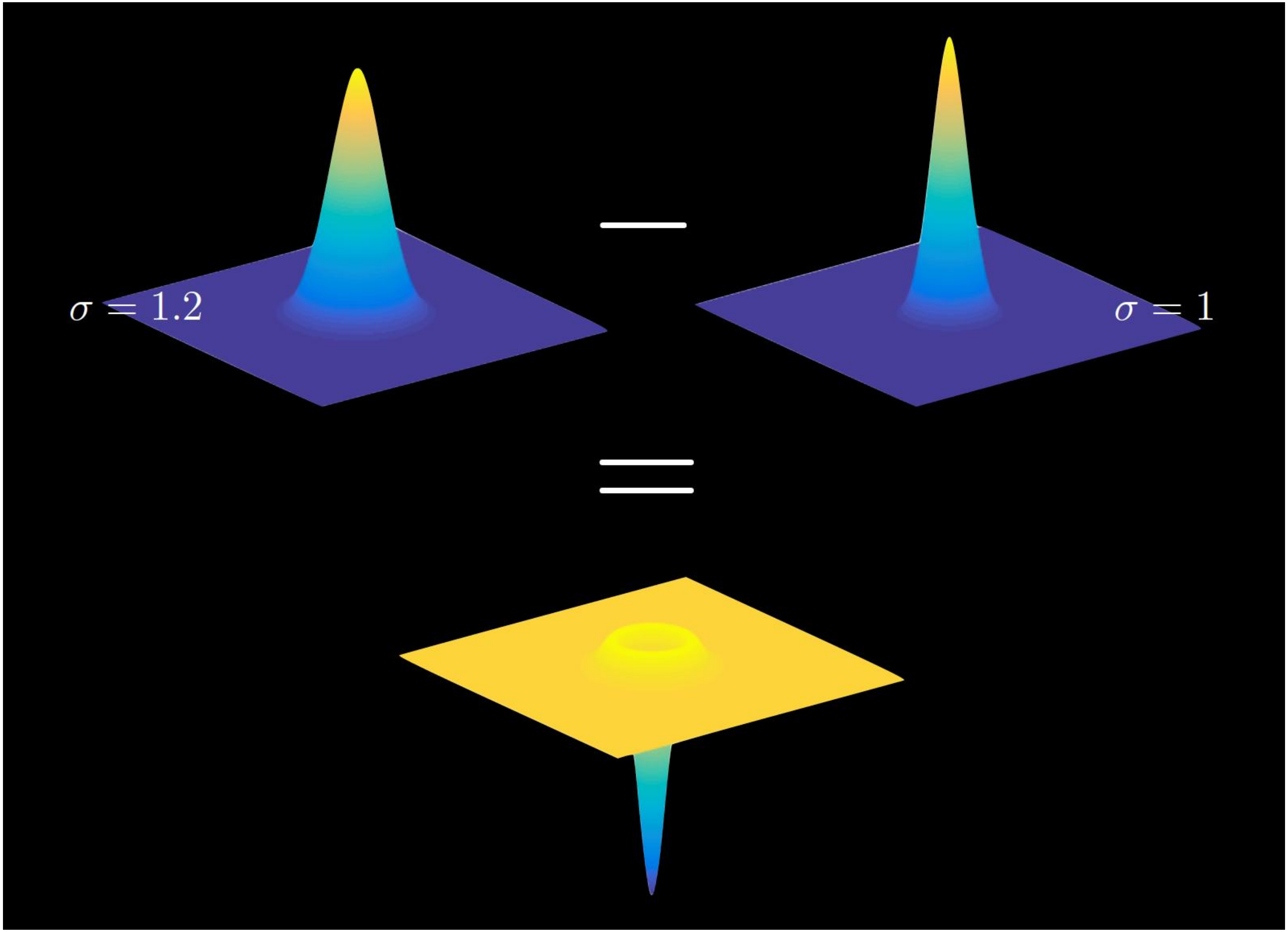


image

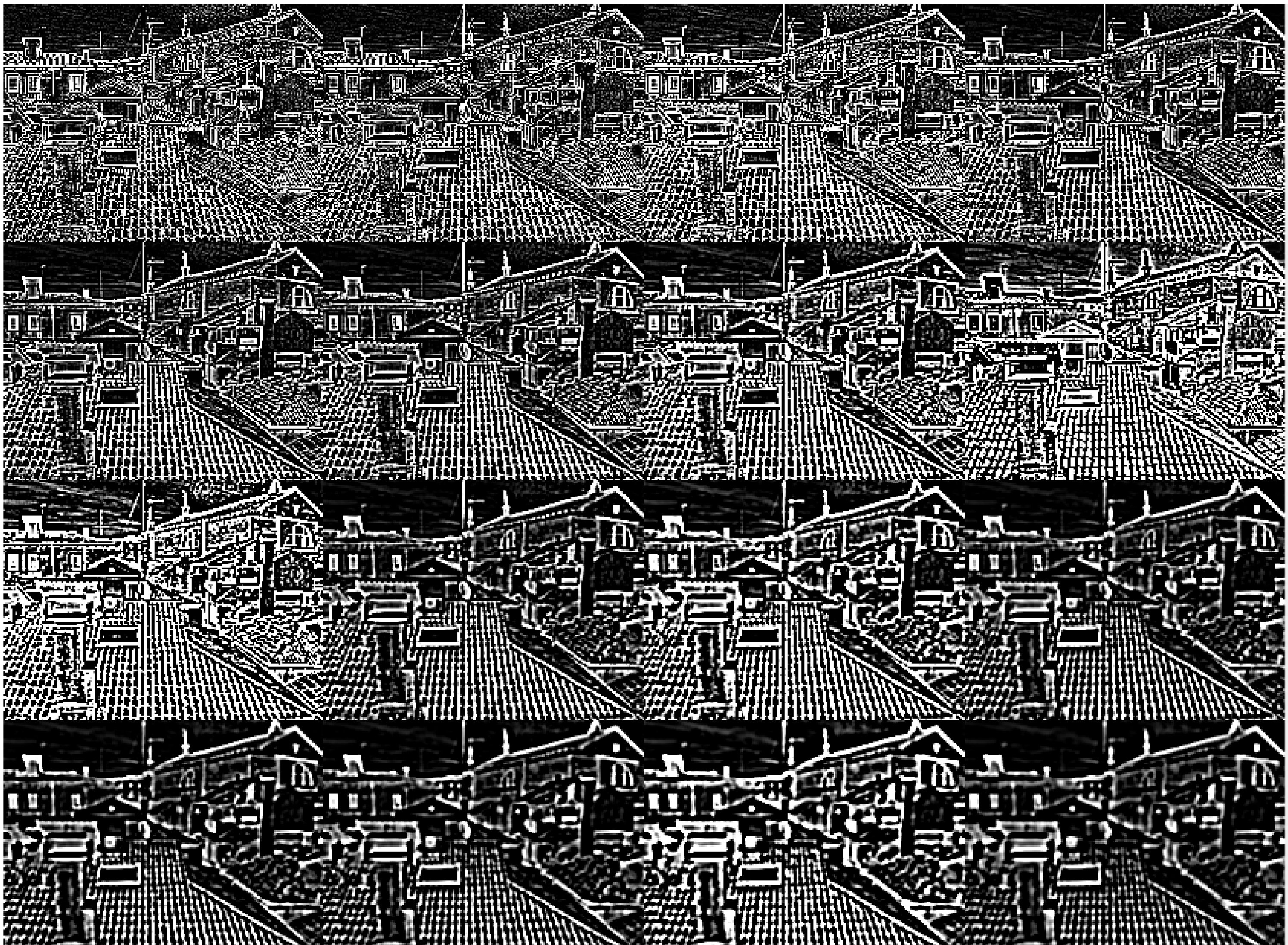


Slide from S. Lazebnik

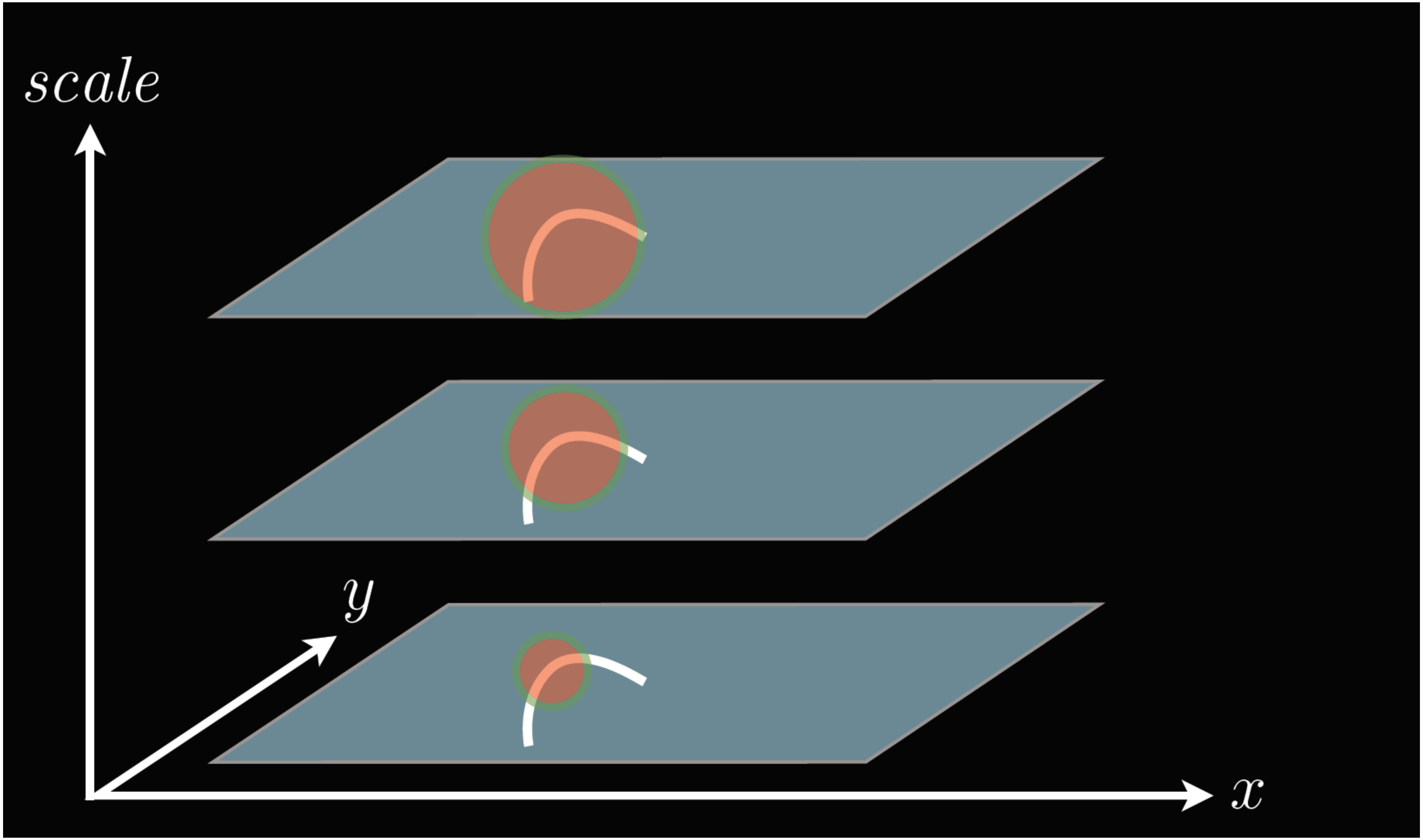
Difference of Gaussians (DoG)



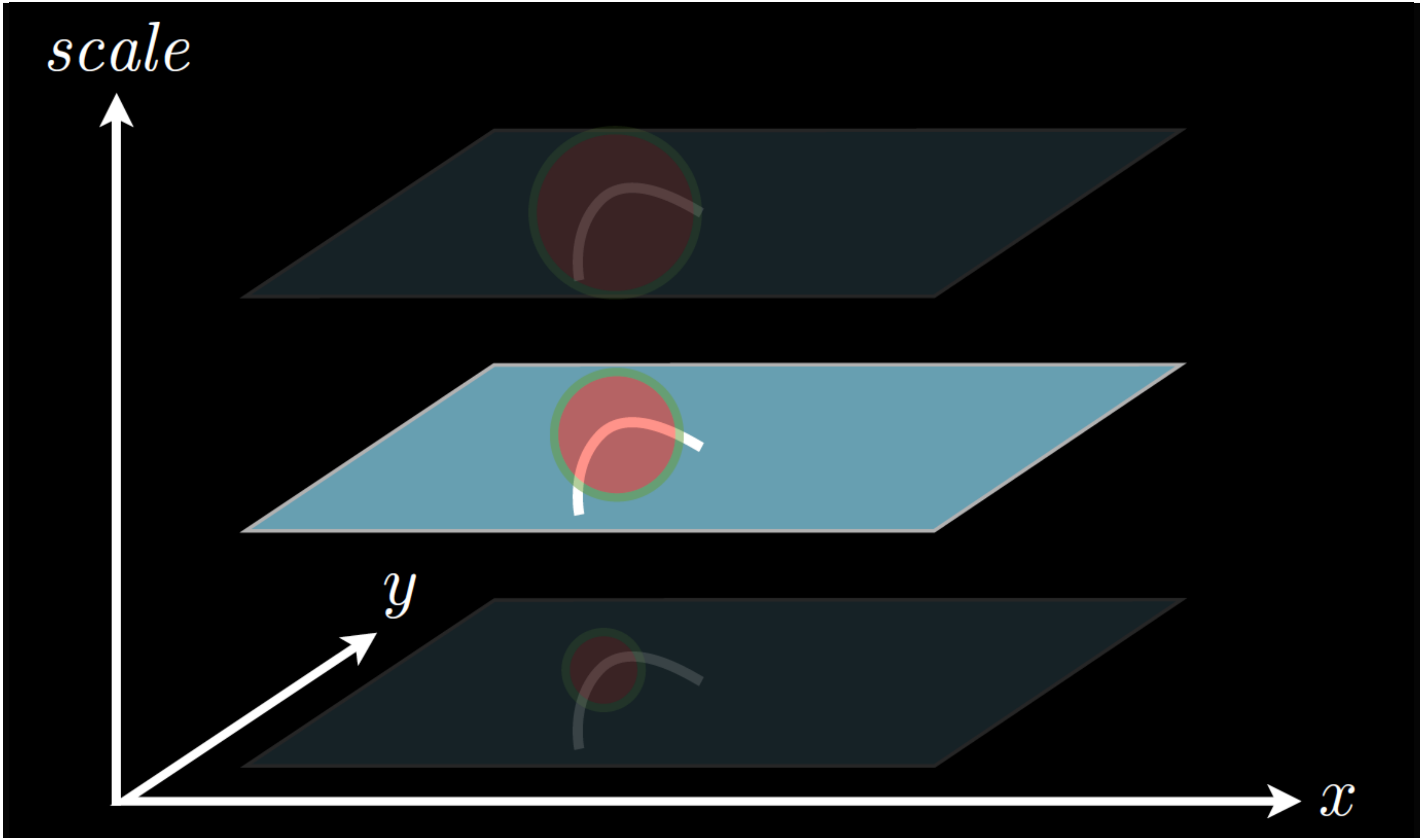
Laplacian Scale Space



We convolve DoG across space at different scales

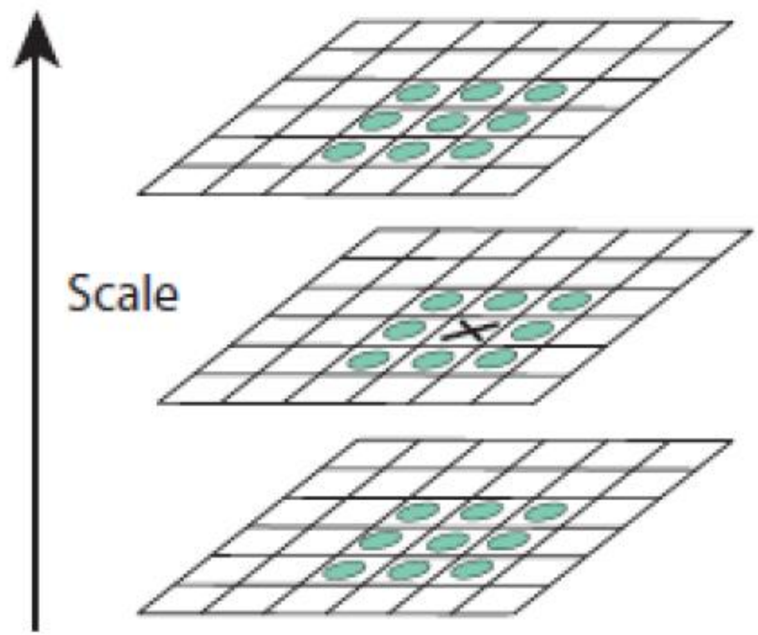


We convolve DoG across space at different scales and detect maximum



So, where is a SIFT keypoint?

- Definition of a keypoint: Maximum in the 3x3x3 (x,y,σ) region of the point.



Vedaldi's vlfeat

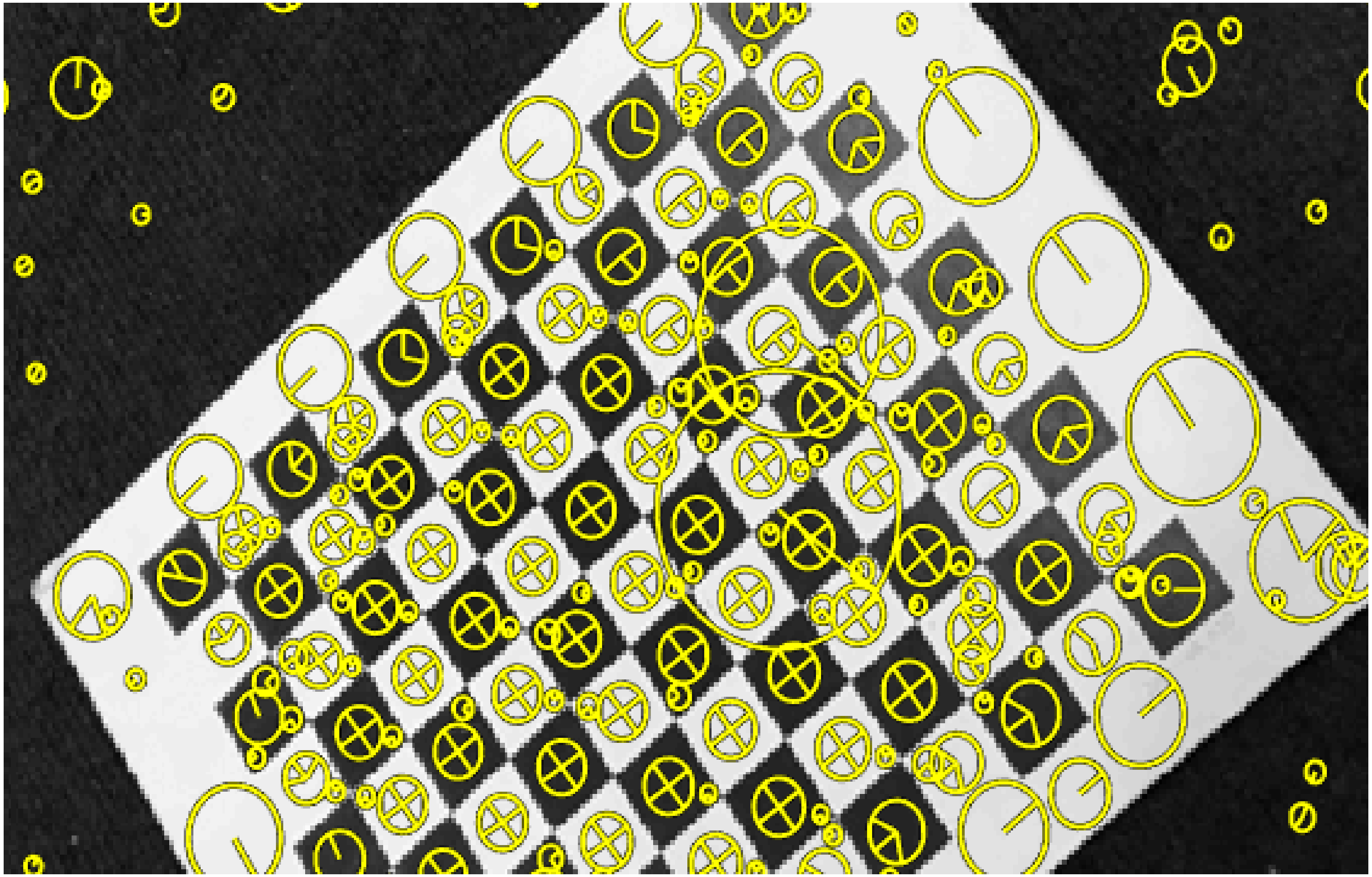
VLFeat.org

Home **Download** **Tutorials** **Applications** **Documentation**

Andrea Vedaldi, Ph.D. vedaldi@robots.ox.ac.uk
Associate Professor in Engineering Science
Information Engineering Building 30.05, Parks Road, Oxford, OX1 3PJ
Visual Geometry Group (directions)
Tel. +44 1865 273 127

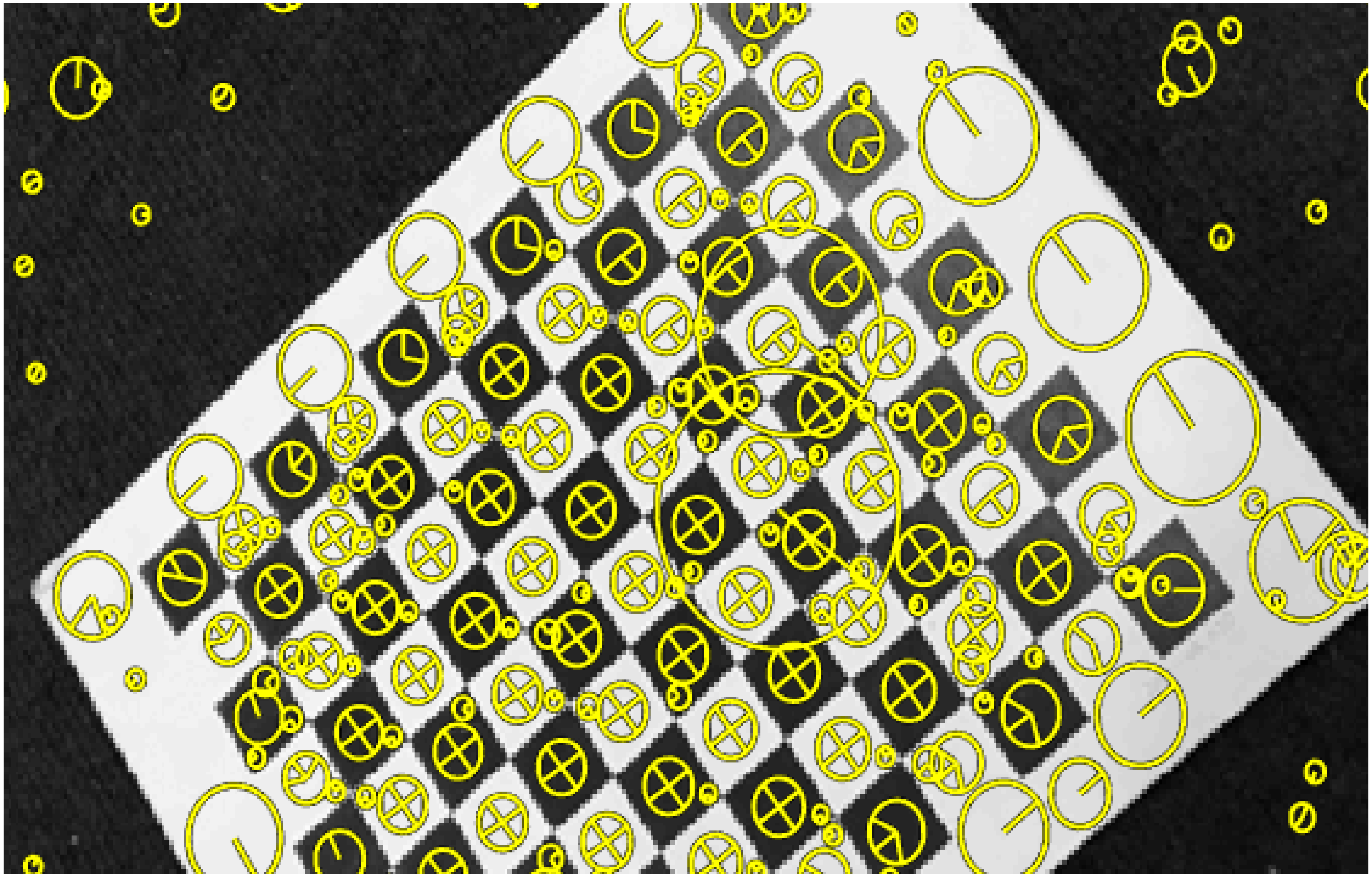
Résumé Google Scholar

Selected σ is visualized with a circle



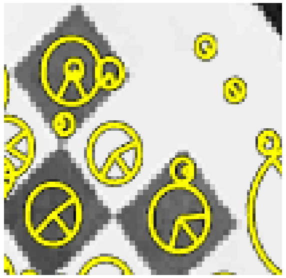
Denoting the support region of the feature

Detector is rotation invariant



Because Laplacian is isotropic and a maximum in (x,y,σ) is invariant to rotations.

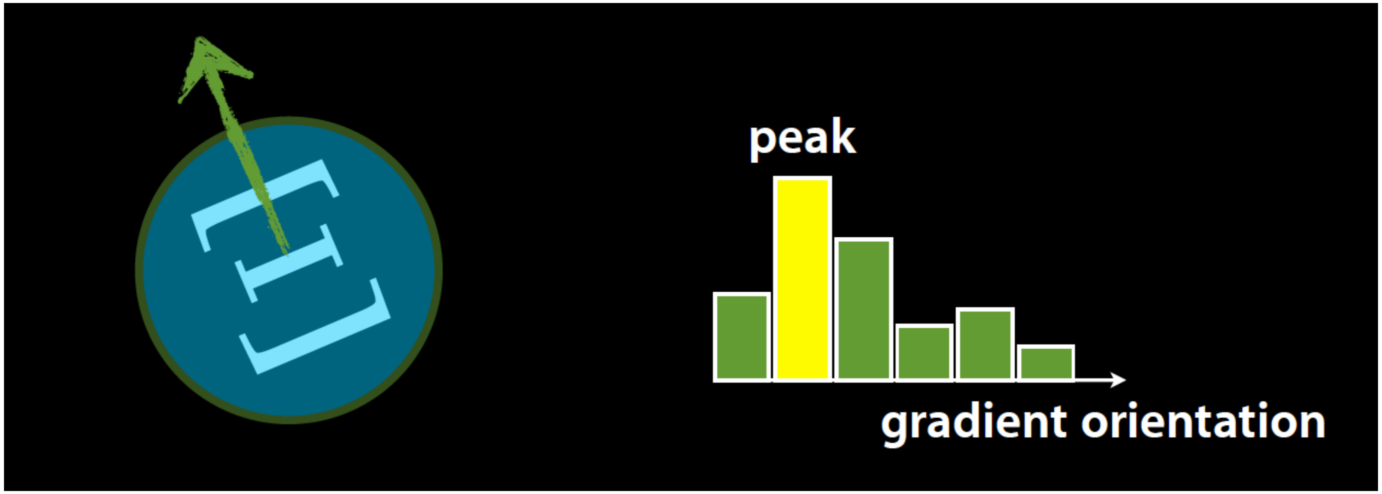
Descriptor invariance



- Since the intrinsic scale is detected (circle size) all circles will be normalized to a 16x16 region.

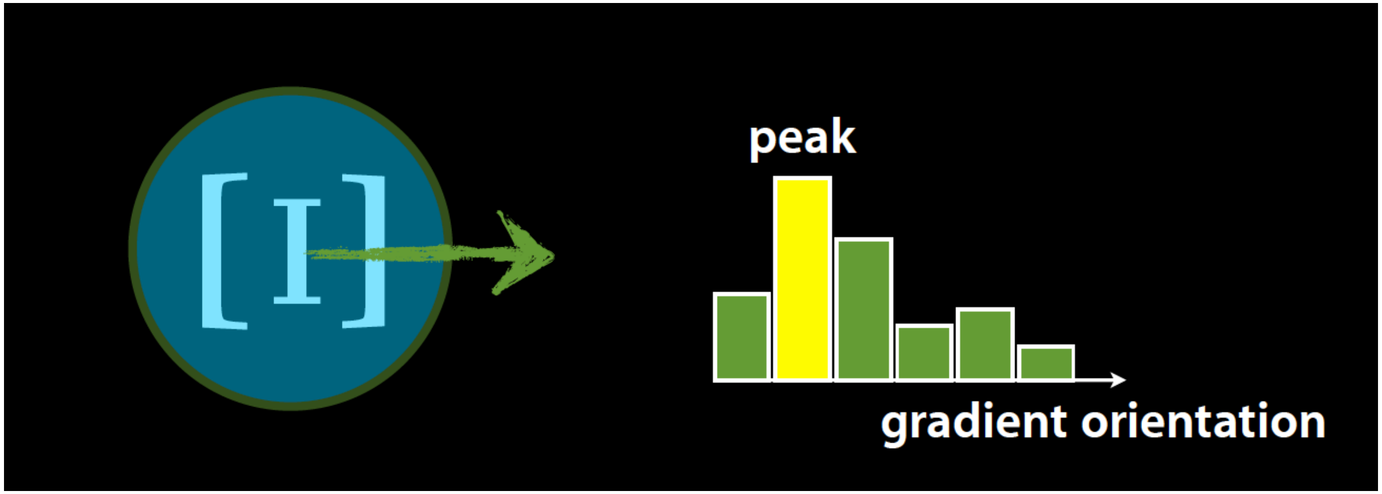
The descriptor should be also rotation invariant

- 1st Step: Find dominant orientation for the patch



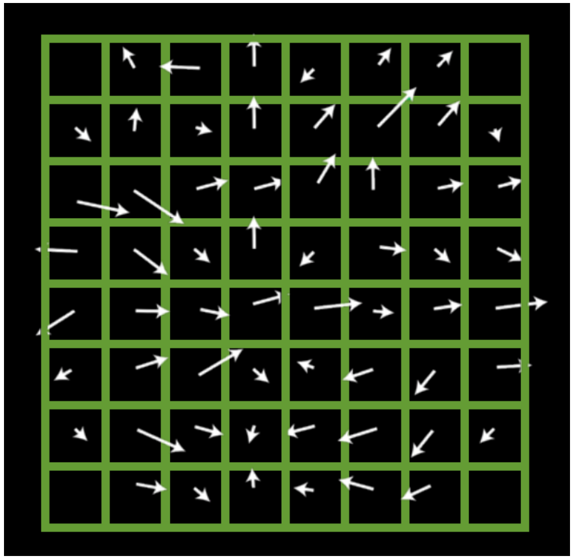
The descriptor should be rotation invariant

- 1st Step: Find dominant orientation for the patch
- 2nd Step: Rotate patch to point along x-axis



To extract a feature descriptor from a cell

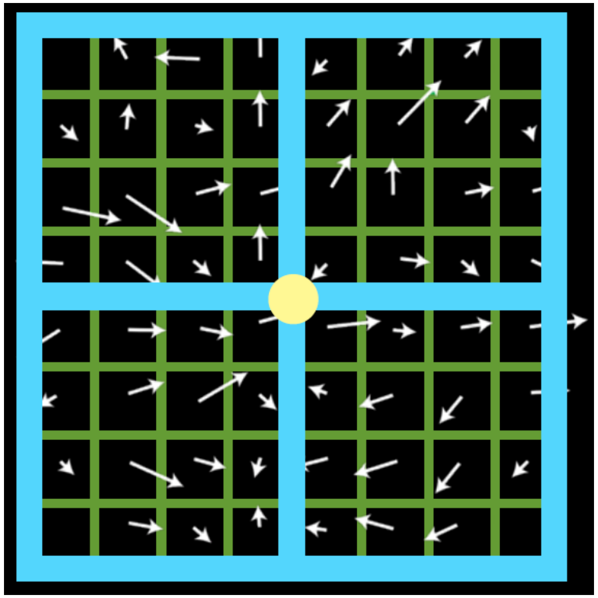
- Compute Image Gradients



To extract a feature descriptor from a cell

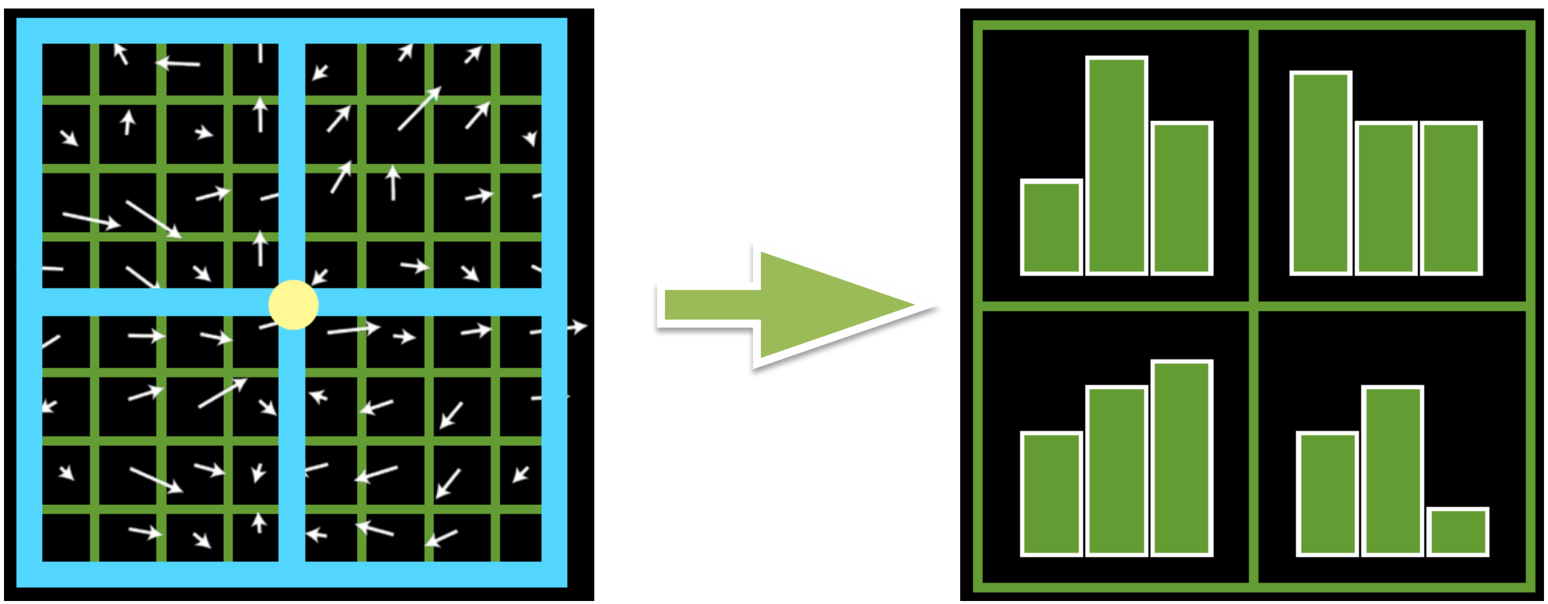
- Compute Image Gradients

- Accumulate gradients along cells

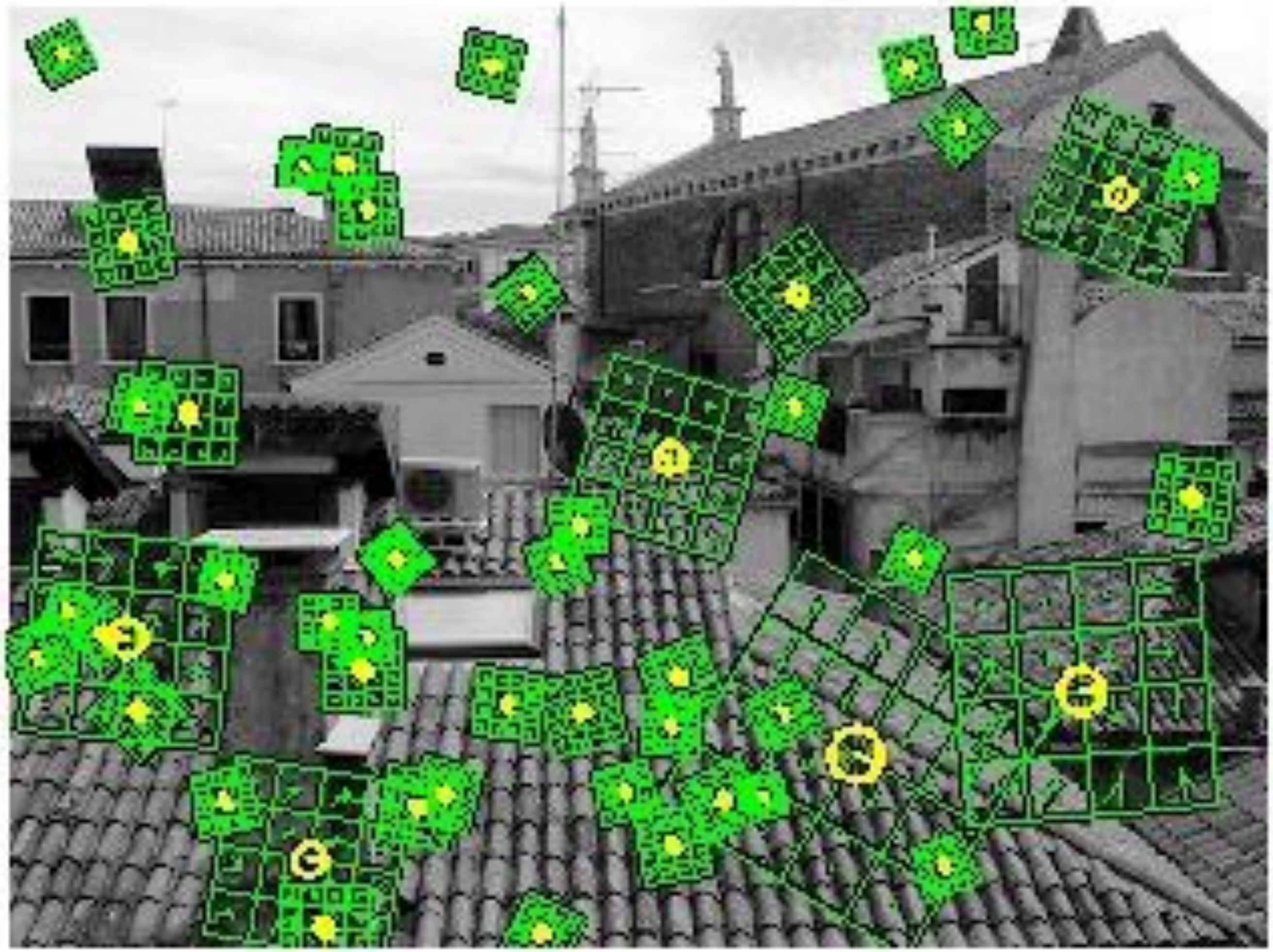


To extract a feature descriptor from a cell

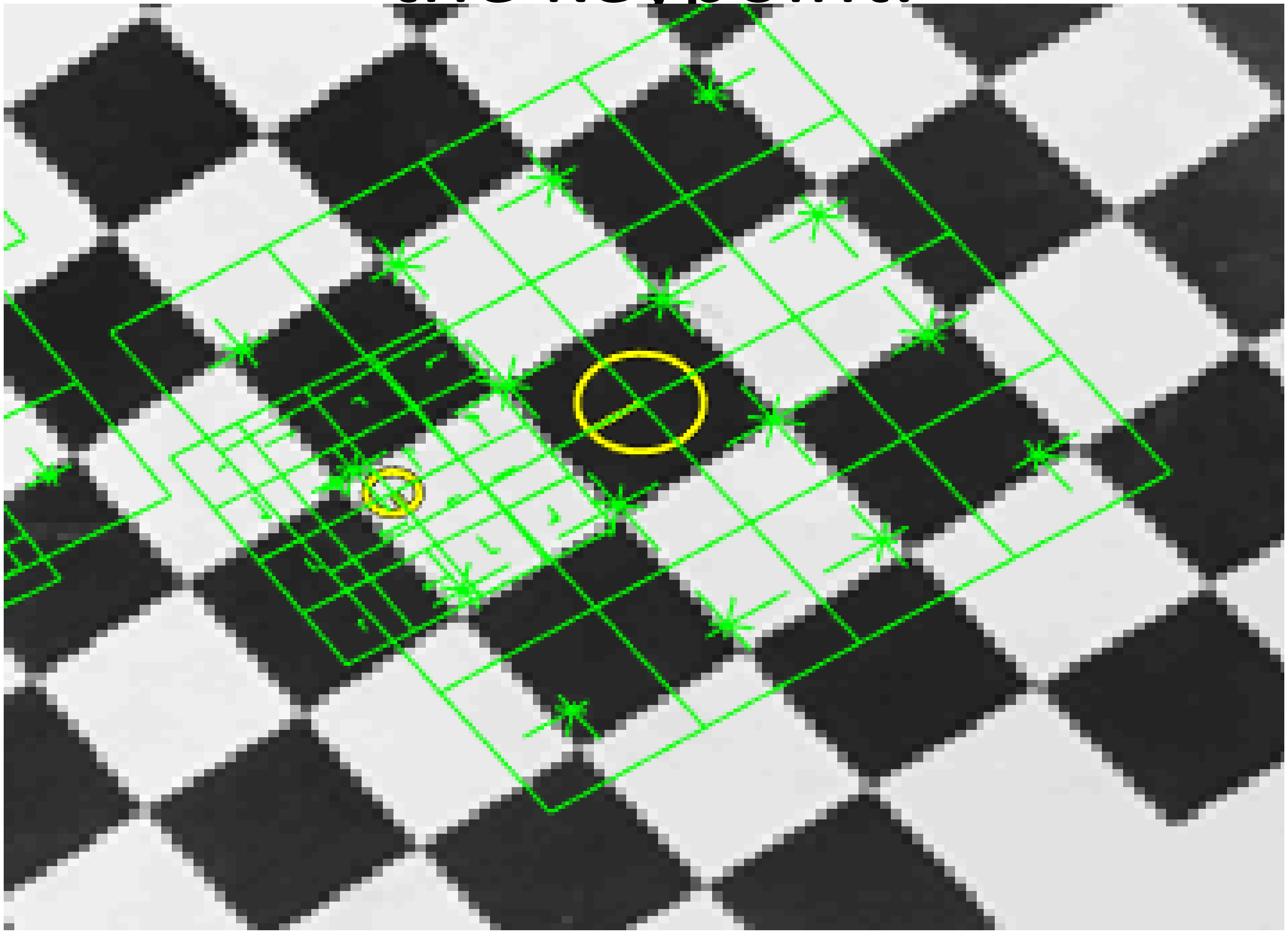
- Compute Image Gradients
- Accumulate gradients along cells
- Form image descriptor



As a matter of fact it is a 4x4 grid of histograms at each keypoint

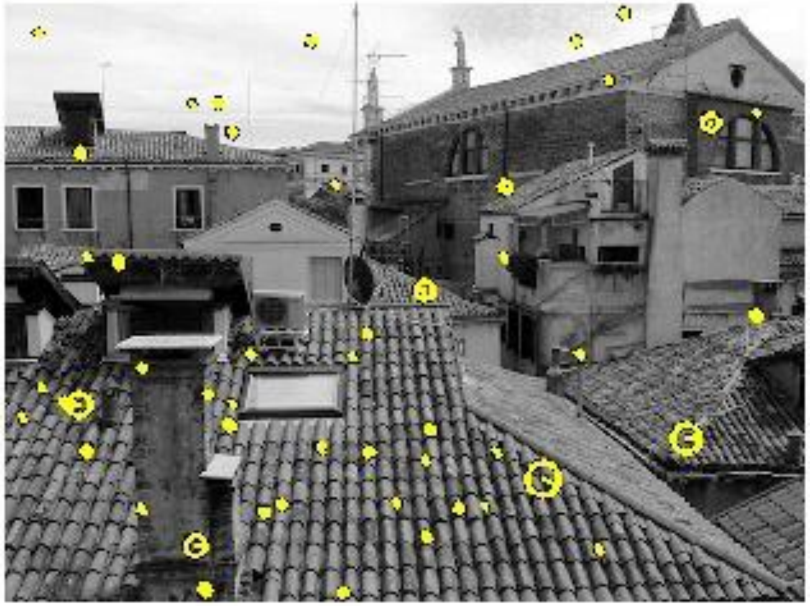


The descriptor is an 128x1 vector which together with σ,θ characterize the keypoint.



Example of SIFT detections and feature extraction

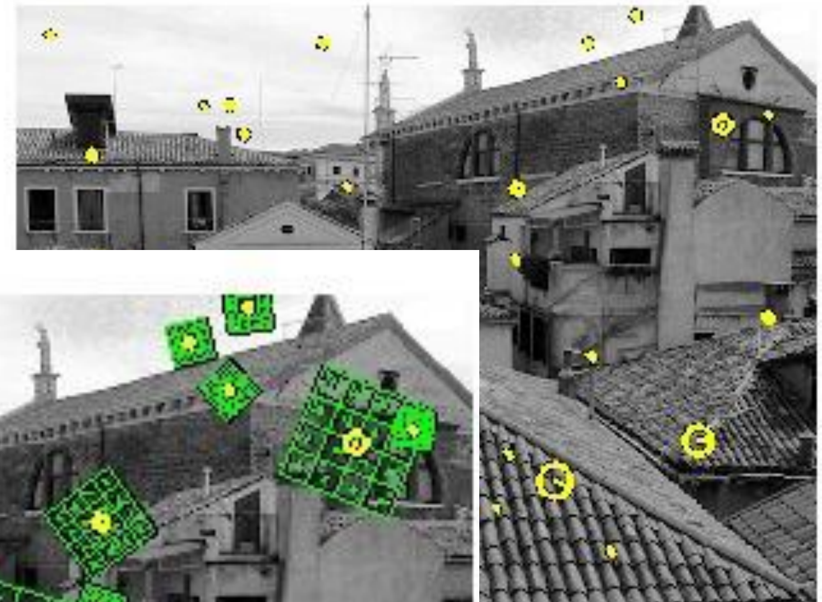
Input Image



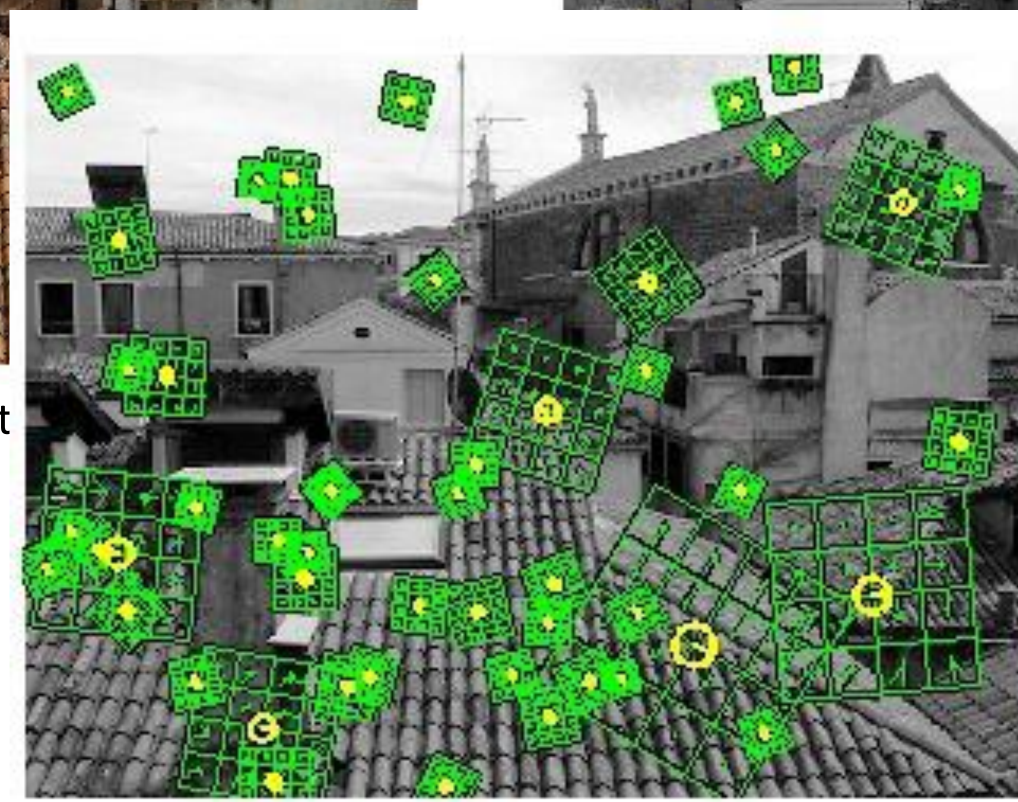
Example Detections

Example of SIFT detections and feature extraction

Input



ections



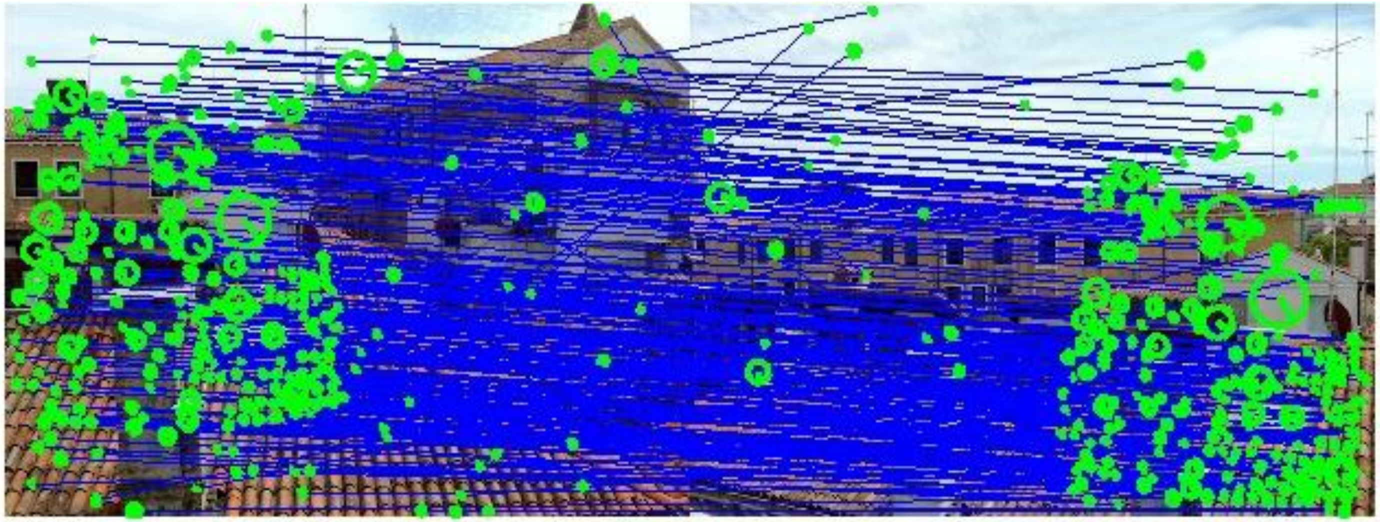
Extracted Feature Descriptors

Using SIFT for image matching

Original Image Pair

Using SIFT for image matching

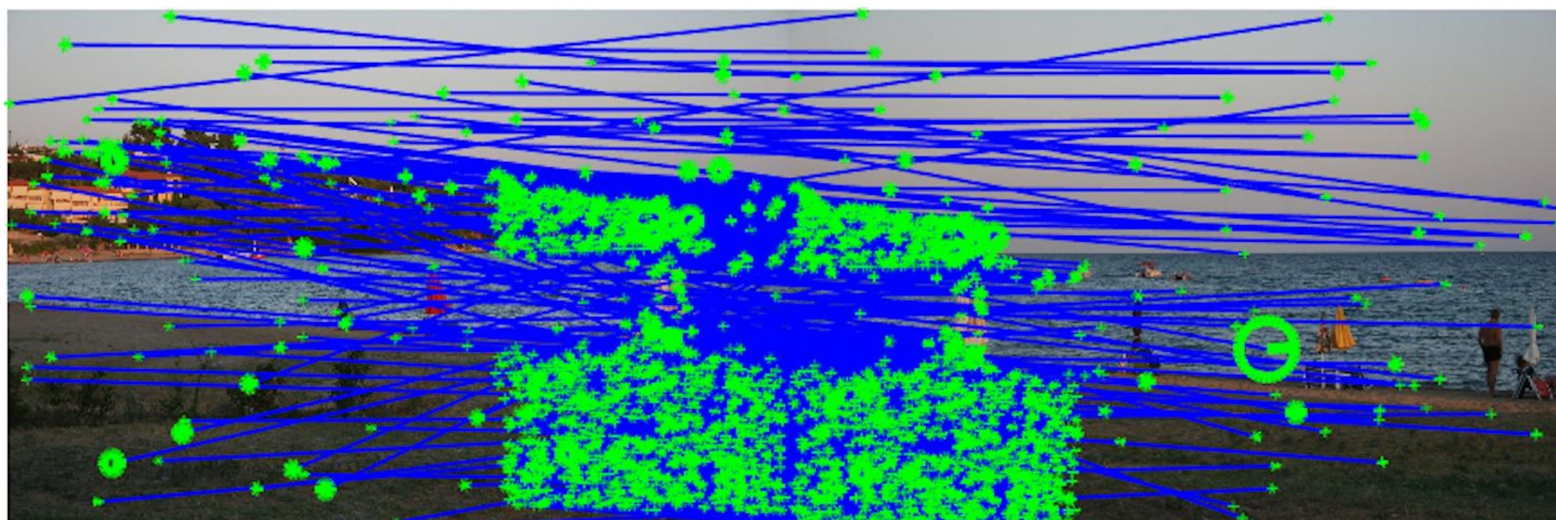
Original Image Pair



Matched features

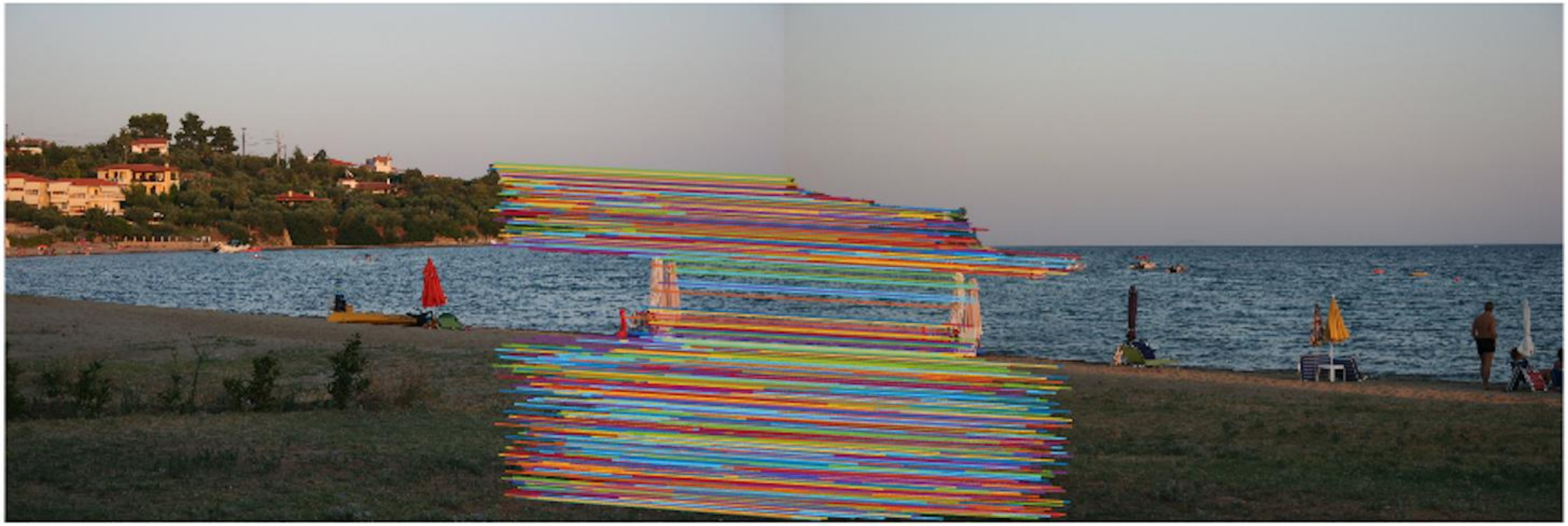
Create Image Mosaic

1. Get an image pair



2. Establish correspondences between matching features

Create Image Mosaic



3. Keep only consistent matches (inliers)

4. Compute homography and warp 2^{nd} image

Create Image Mosaic

5. Repeat to extend the mosaic

Find Location

Query Image

We want to find a match in a dataset of given images

Find Location

Query Image

Find Location

Query Image

Good Matches

Medium Matches

Bad Matches

Find Location

Good Match

Medium Match

Bad Match

SIFT Features

- SIFT detector can automatically
  - Select scale
  - Compute dominant rotation
- SIFT descriptor
  - Is a grid of histogram of gradient orientations
  - On a region normalized with respect to scale and rotation