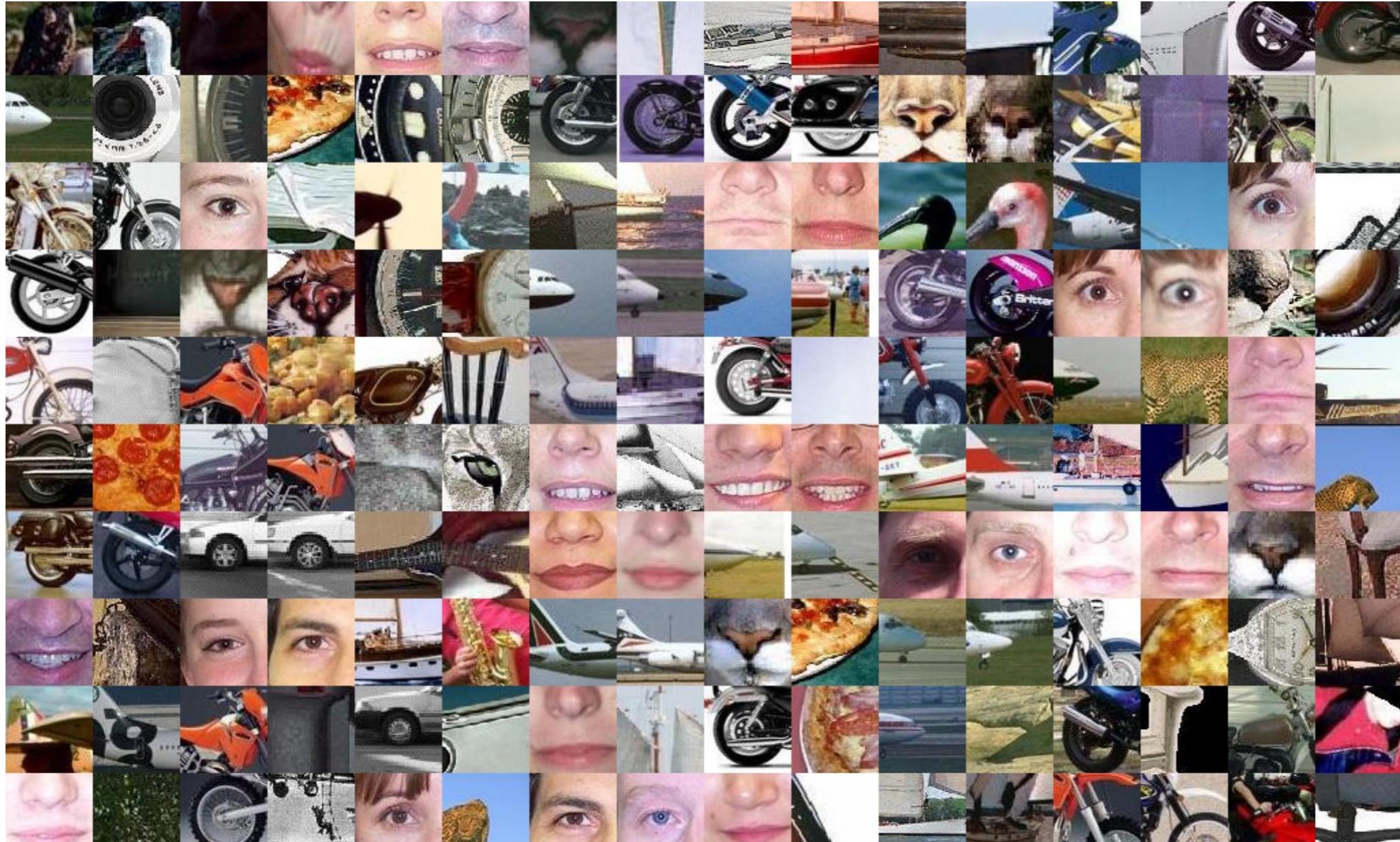


Feature detectors and descriptors



Guest Lecturers



Adithya Pediredla

Lecture 7: 2D Transformations

Lecture 8: Image Homographies



Mark Sheinin

Lecture 9: Geometric
Camera Models

Overview of today's lecture

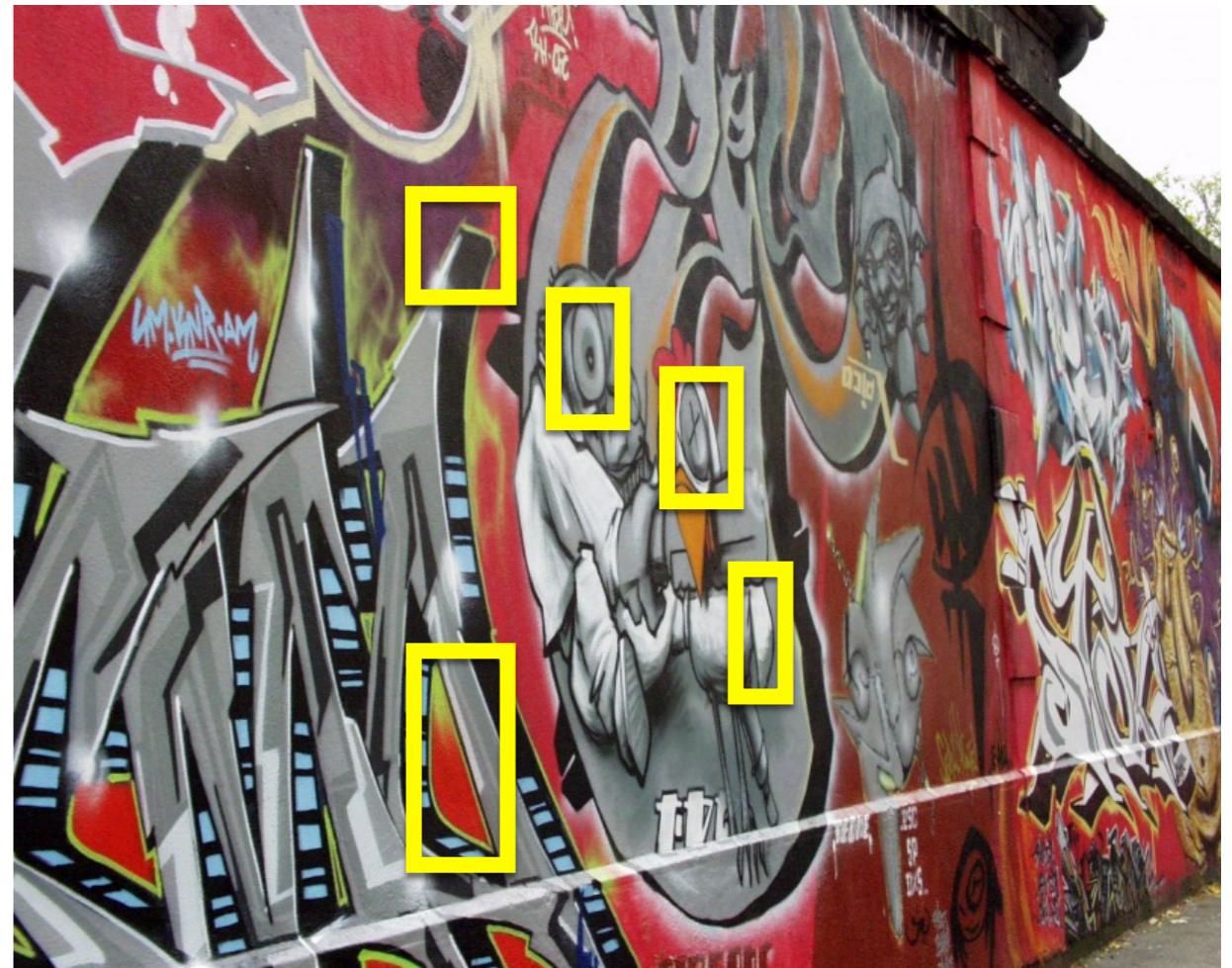
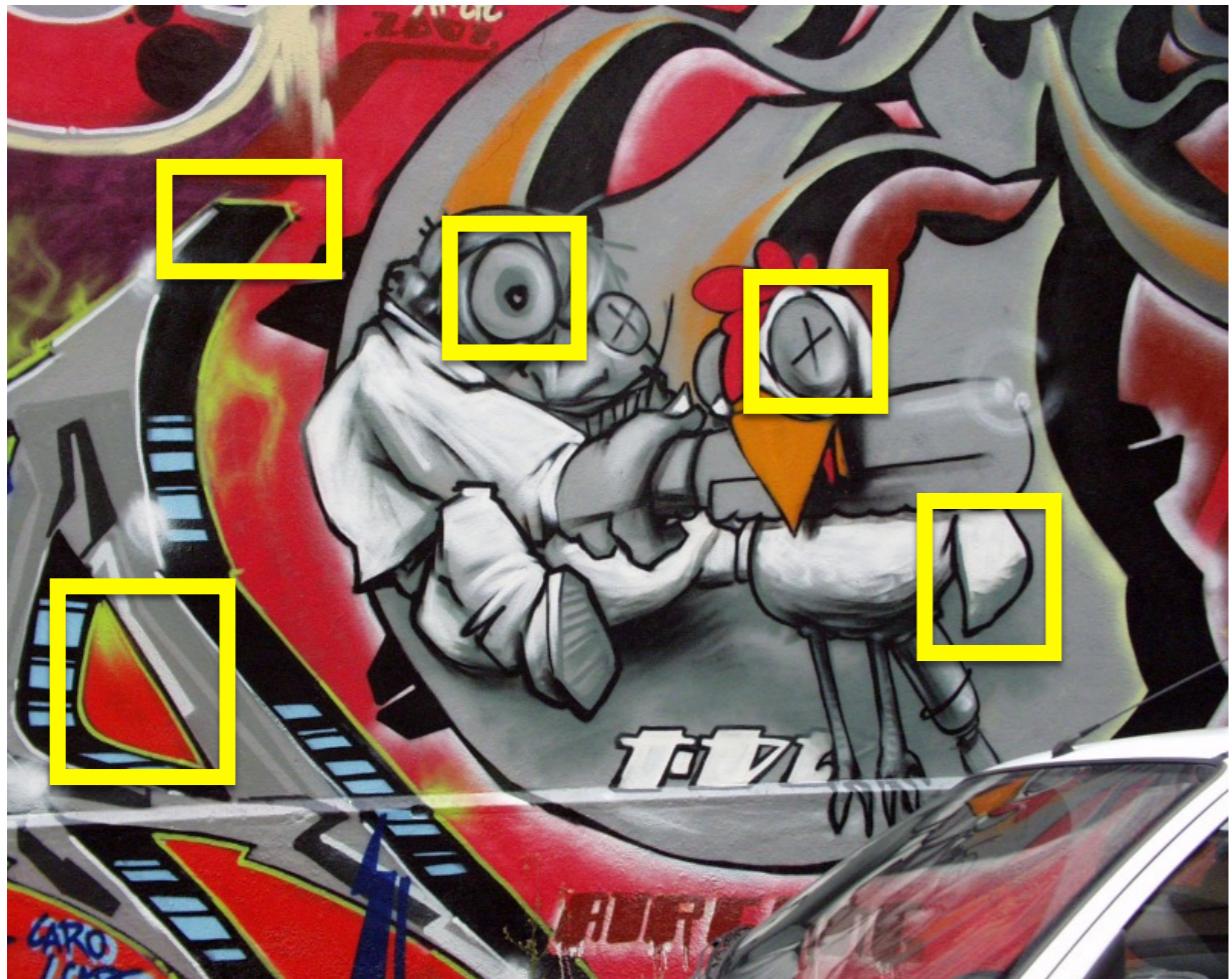
- Why do we need feature descriptors?
- Designing feature descriptors.
- MOPS descriptor.
- GIST descriptor.

Slide credits

Most of these slides were adapted from:

- Kris Kitani (16-385, Spring 2017).

Why do we need feature
descriptors?

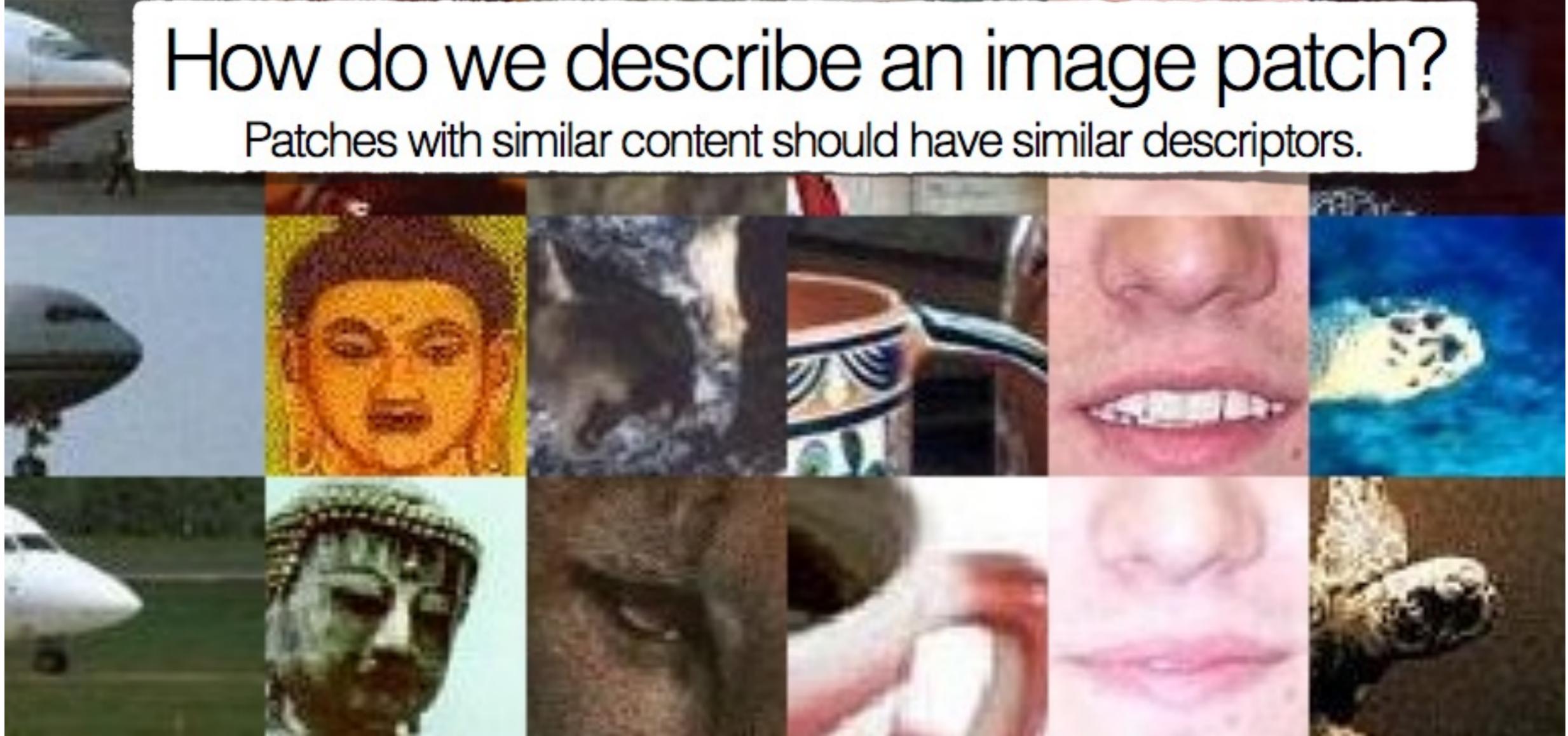


*If we know where the good features are,
how do we match them?*



How do we describe an image patch?

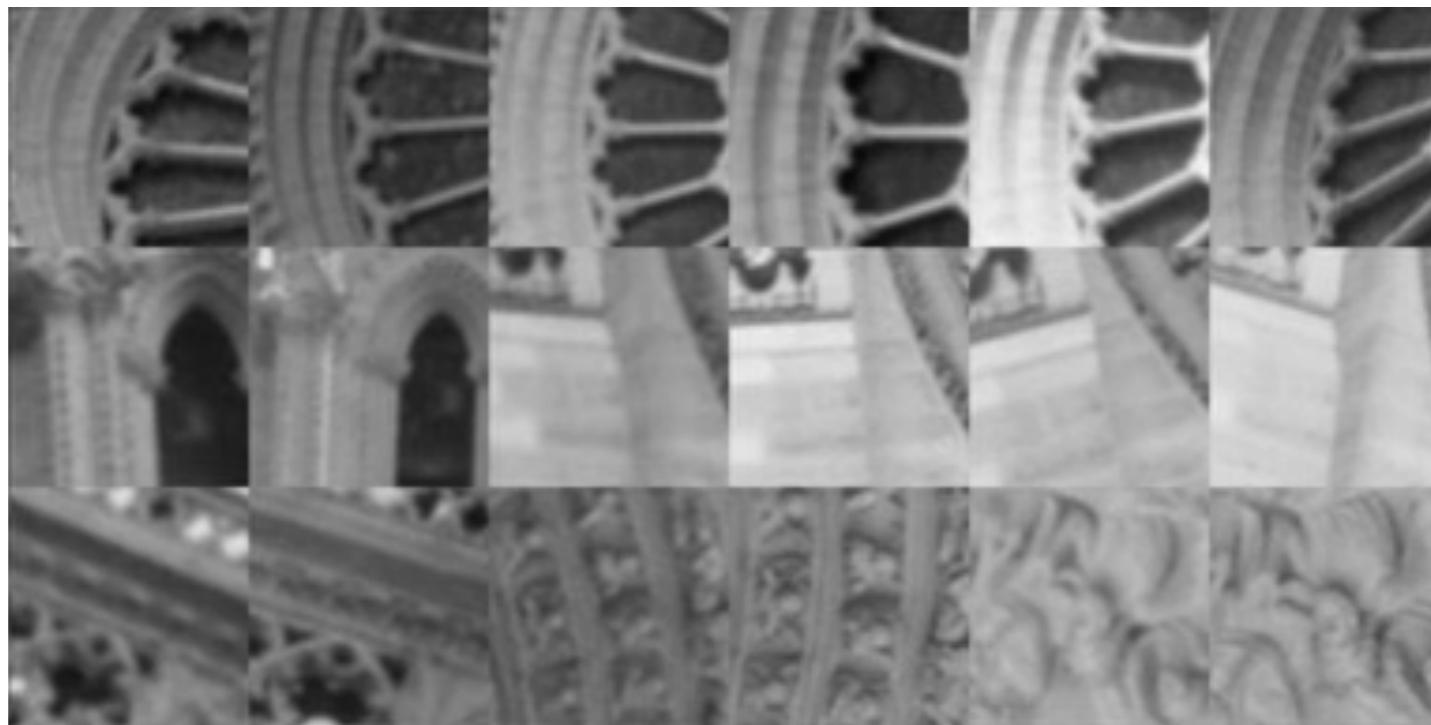
Patches with similar content should have similar descriptors.



Designing feature
descriptors



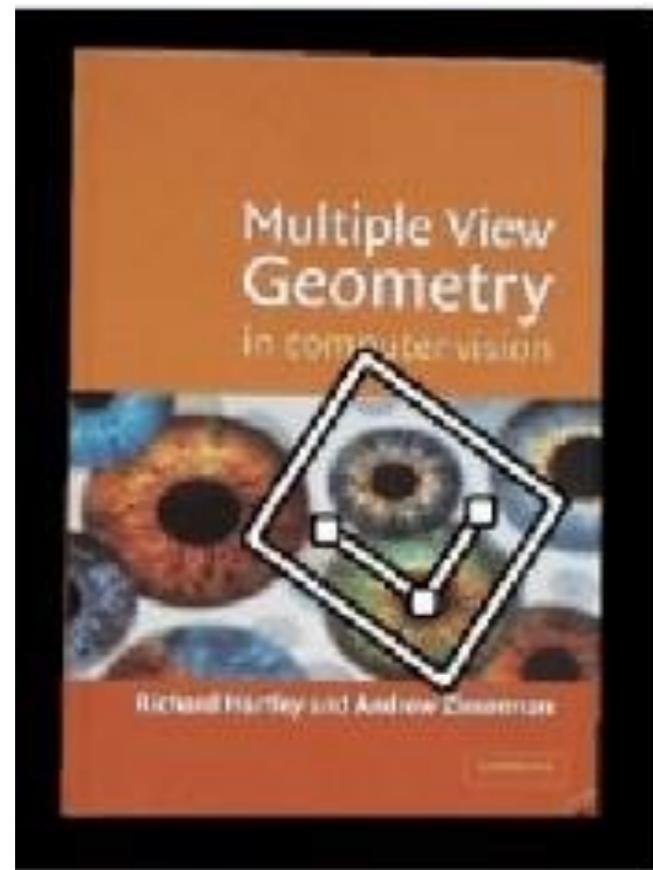
What is the best descriptor for an image feature?



Photometric transformations



Geometric transformations



objects will appear at different scales,
translation and rotation

Image patch

Just use the pixel values of the patch!



Perfectly fine if geometry and appearance is unchanged
(a.k.a. template matching)

Image patch

Just use the pixel values of the patch!



Perfectly fine if geometry and appearance is unchanged
(a.k.a. template matching)

What are the problems?

Image patch

Just use the pixel values of the patch!



Perfectly fine if geometry and appearance is unchanged
(a.k.a. template matching)

What are the problems?

How can you be less sensitive to absolute intensity values?

Image gradients

Use pixel differences

1	2	3
4	5	6
7	8	9



$$(\quad - \quad + \quad + \quad - \quad - \quad + \quad)$$

vector of x derivatives

'binary descriptor'

Feature is invariant to absolute intensity values

What are the problems?

Image gradients

Use pixel differences

1	2	3
4	5	6
7	8	9



$$(\quad - \quad + \quad + \quad - \quad - \quad + \quad)$$

vector of x derivatives

'binary descriptor'

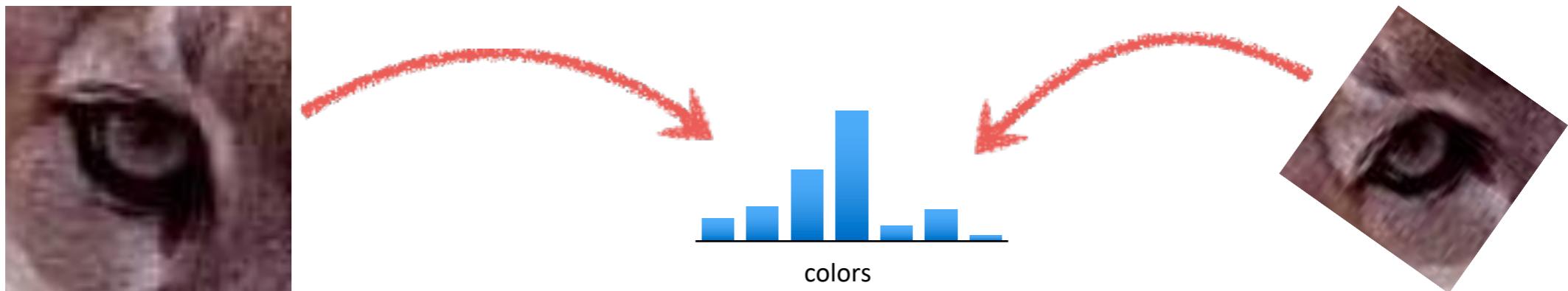
Feature is invariant to absolute intensity values

What are the problems?

How can you be less sensitive to deformations?

Color histogram

Count the colors in the image using a histogram

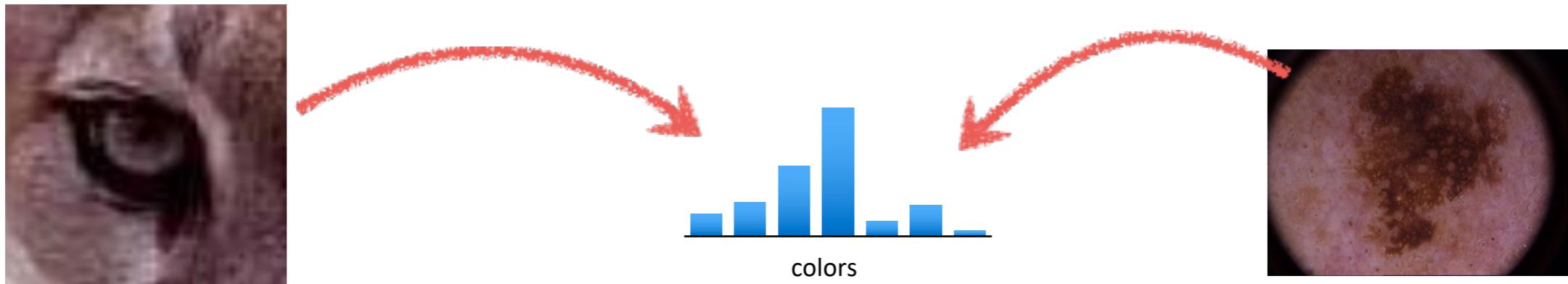


Invariant to changes in scale and rotation

What are the problems?

Color histogram

Count the colors in the image using a histogram

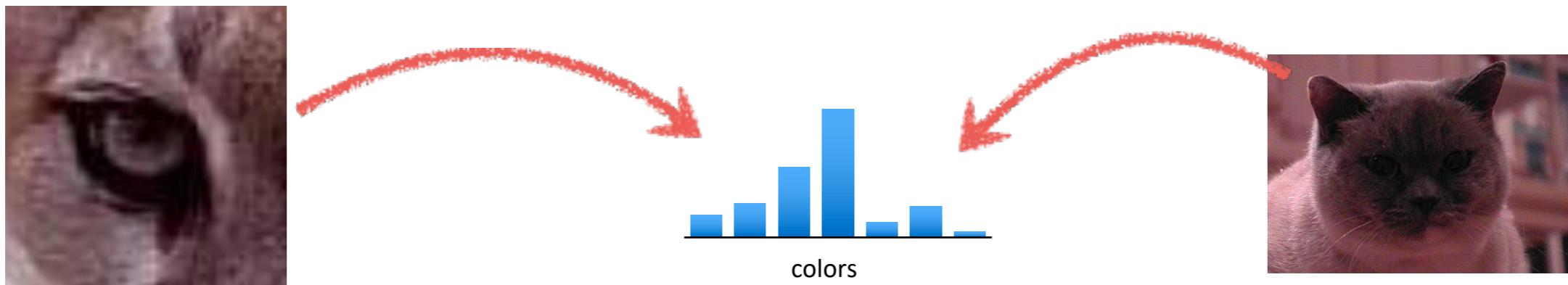


Invariant to changes in scale and rotation

What are the problems?

Color histogram

Count the colors in the image using a histogram



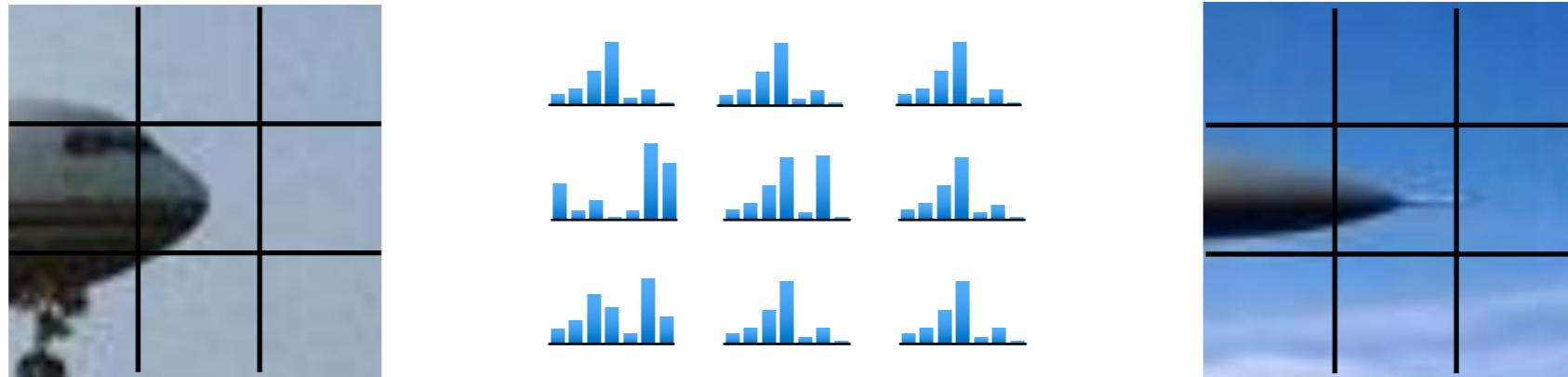
Invariant to changes in scale and rotation

What are the problems?

How can you be more sensitive to spatial layout?

Spatial histograms

Compute histograms over spatial ‘cells’

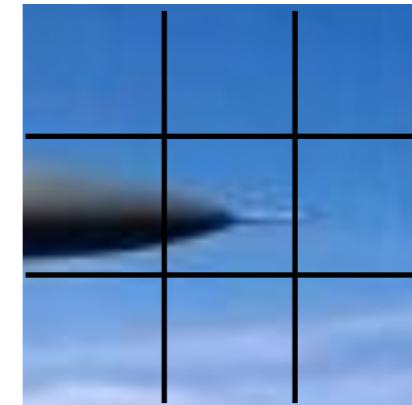
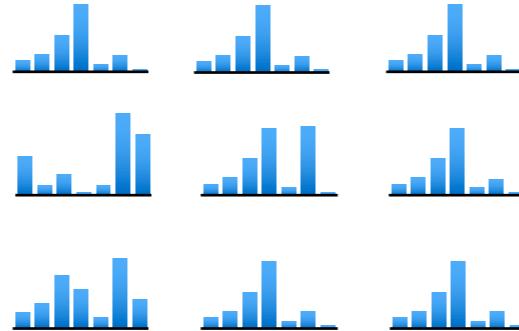


Retains rough spatial layout
Some invariance to deformations

What are the problems?

Spatial histograms

Compute histograms over spatial ‘cells’



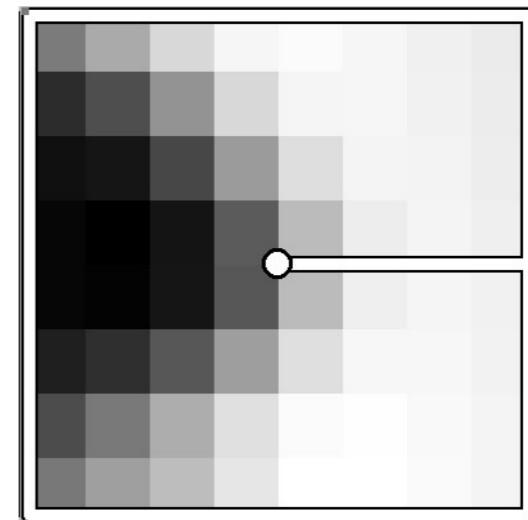
Retains rough spatial layout
Some invariance to deformations

What are the problems?

How can you be completely invariant to rotation?

Orientation normalization

Use the dominant image gradient direction to normalize the orientation of the patch

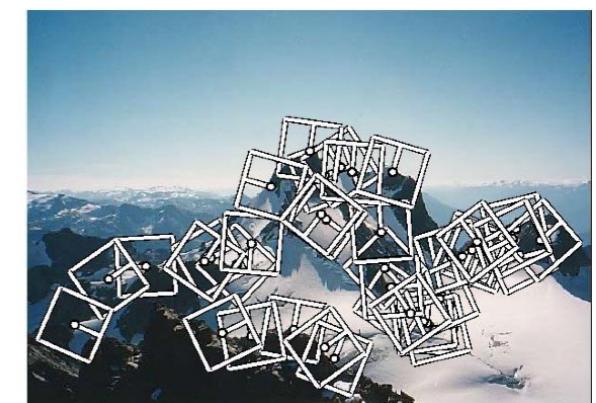
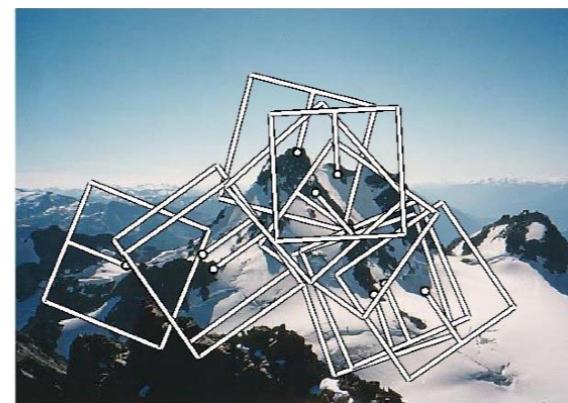
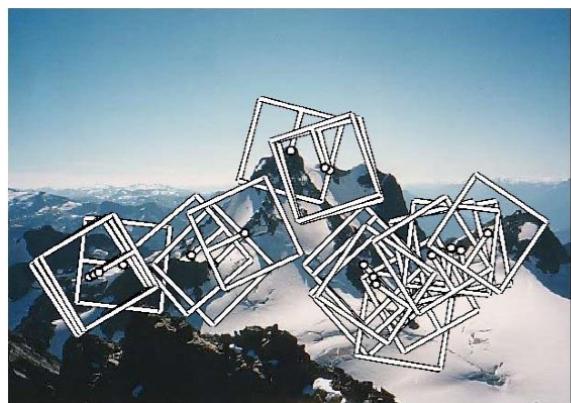


save the orientation angle θ along with (x, y, s)

MOPS descriptor

Multi-Scale Oriented Patches (MOPS)

Multi-Image Matching using Multi-Scale Oriented Patches. M. Brown, R. Szeliski and S. Winder.
International Conference on Computer Vision and Pattern Recognition (CVPR2005). pages 510-517



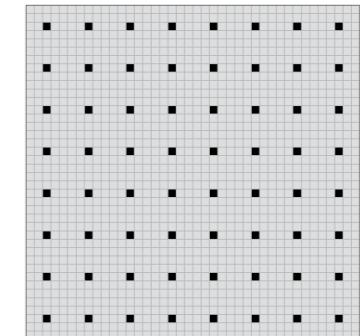
Multi-Scale Oriented Patches (MOPS)

Multi-Image Matching using Multi-Scale Oriented Patches. M. Brown, R. Szeliski and S. Winder.
International Conference on Computer Vision and Pattern Recognition (CVPR2005). pages 510-517

Given a feature (x, y, s, θ)

Get 40 x 40 image patch, subsample
every 5th pixel

(*what's the purpose of this step?*)



Subtract the mean, divide by standard
deviation

(*what's the purpose of this step?*)

Haar Wavelet Transform

(*what's the purpose of this step?*)

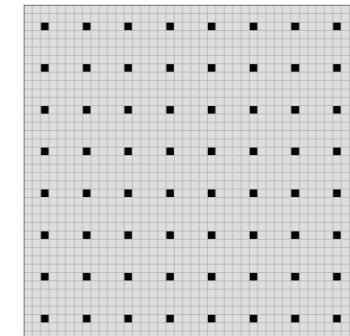
Multi-Scale Oriented Patches (MOPS)

Multi-Image Matching using Multi-Scale Oriented Patches. M. Brown, R. Szeliski and S. Winder.
International Conference on Computer Vision and Pattern Recognition (CVPR2005). pages 510-517

Given a feature (x, y, s, θ)

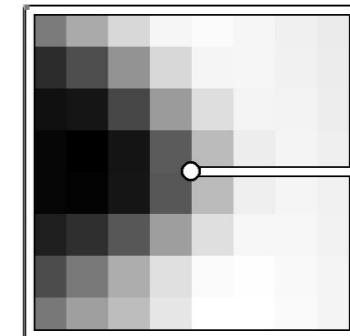
Get 40×40 image patch, subsample
every 5th pixel

(low frequency filtering, absorbs localization errors)



Subtract the mean, divide by standard
deviation

(*what's the purpose of this step?*)



Haar Wavelet Transform

(*what's the purpose of this step?*)

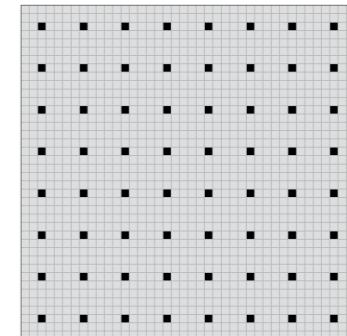
Multi-Scale Oriented Patches (MOPS)

Multi-Image Matching using Multi-Scale Oriented Patches. M. Brown, R. Szeliski and S. Winder.
International Conference on Computer Vision and Pattern Recognition (CVPR2005). pages 510-517

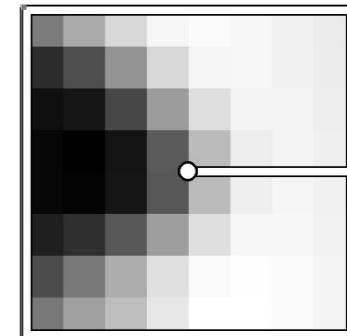
Given a feature (x, y, s, θ)

Get 40×40 image patch, subsample
every 5th pixel

(low frequency filtering, absorbs localization errors)



Subtract the mean, divide by standard
deviation
(removes bias and gain)



Haar Wavelet Transform
(what's the purpose of this step?)



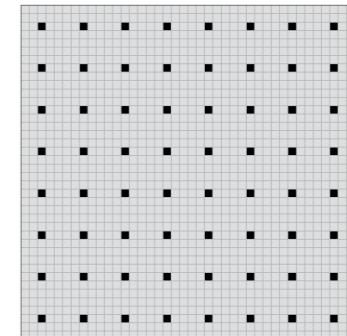
Multi-Scale Oriented Patches (MOPS)

Multi-Image Matching using Multi-Scale Oriented Patches. M. Brown, R. Szeliski and S. Winder.
International Conference on Computer Vision and Pattern Recognition (CVPR2005). pages 510-517

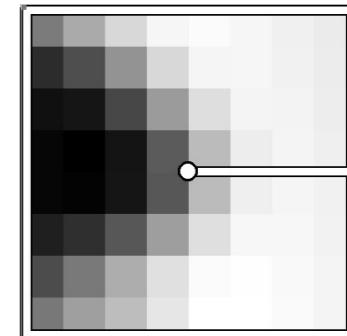
Given a feature (x, y, s, θ)

Get 40 x 40 image patch, subsample
every 5th pixel

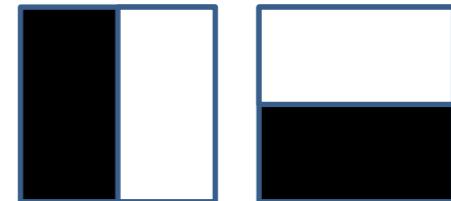
(low frequency filtering, absorbs localization errors)



Subtract the mean, divide by standard
deviation
(removes bias and gain)



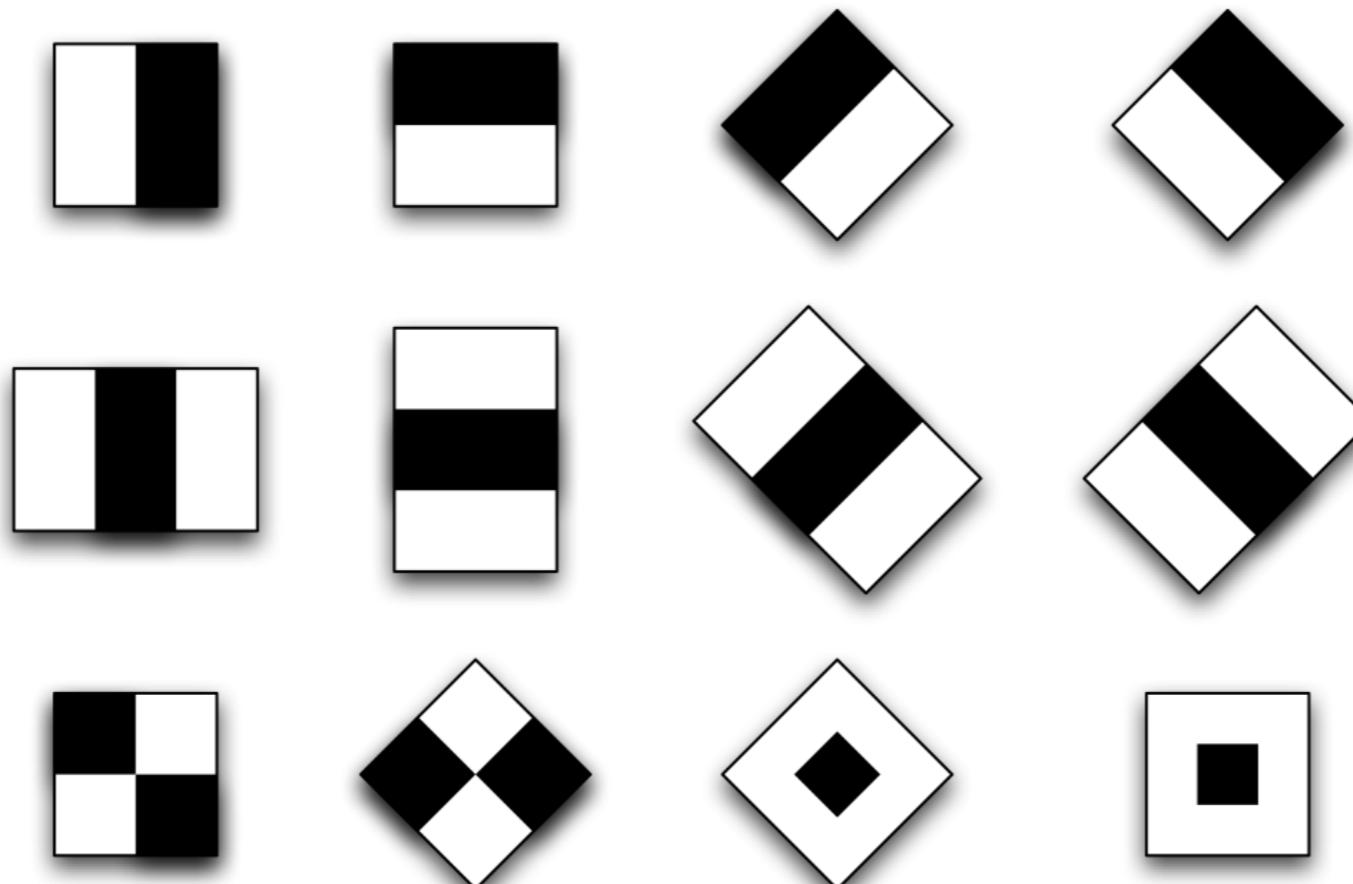
Haar Wavelet Transform
(low frequency projection)



Haar Wavelets

(actually, Haar-like features)

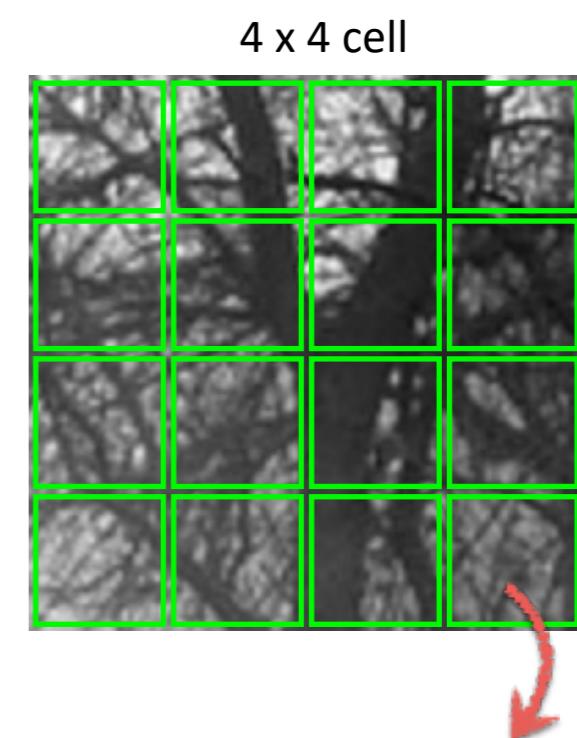
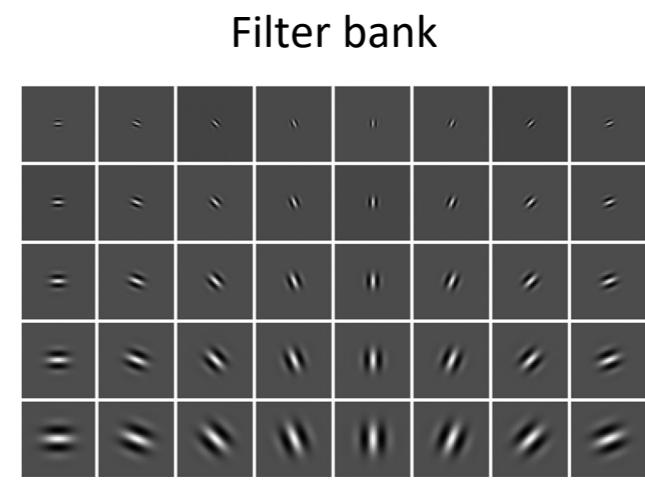
Use responses of a bank of filters as a descriptor



GIST descriptor

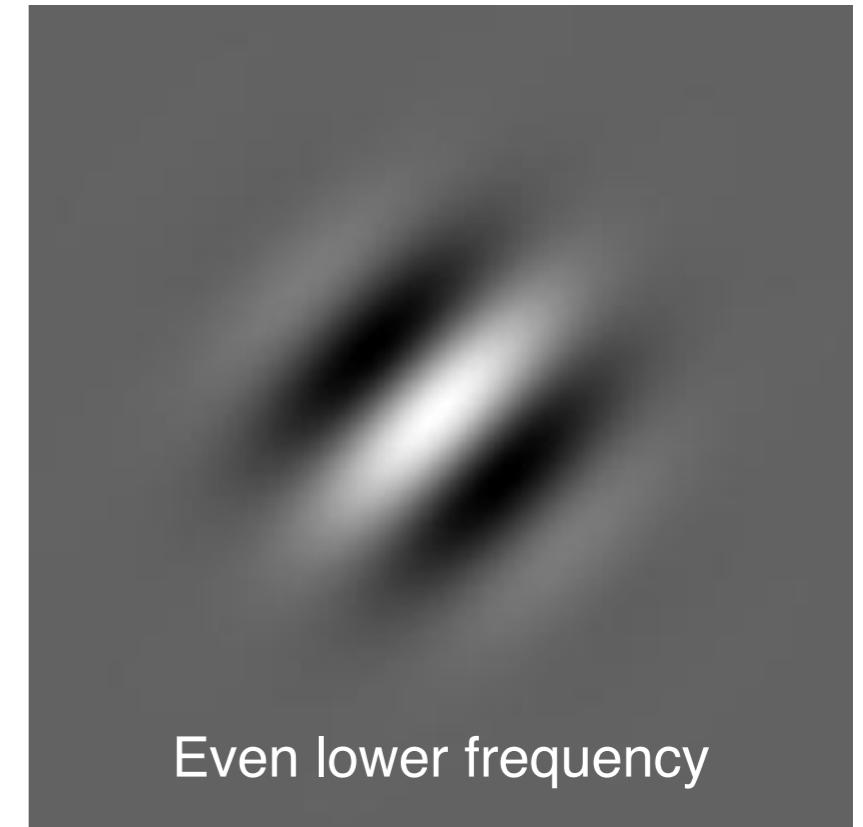
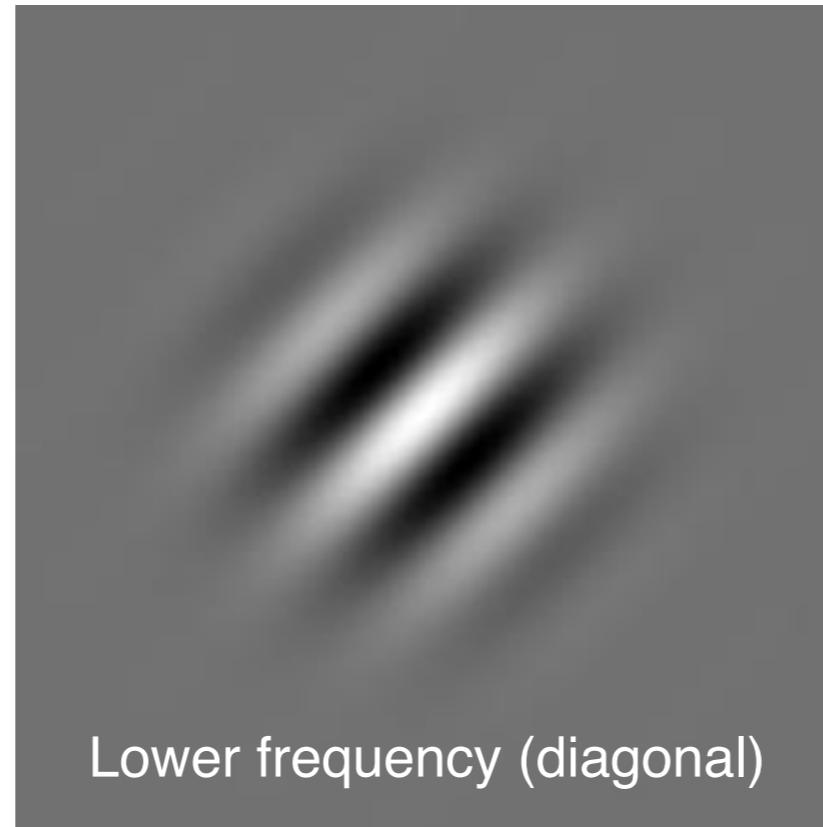
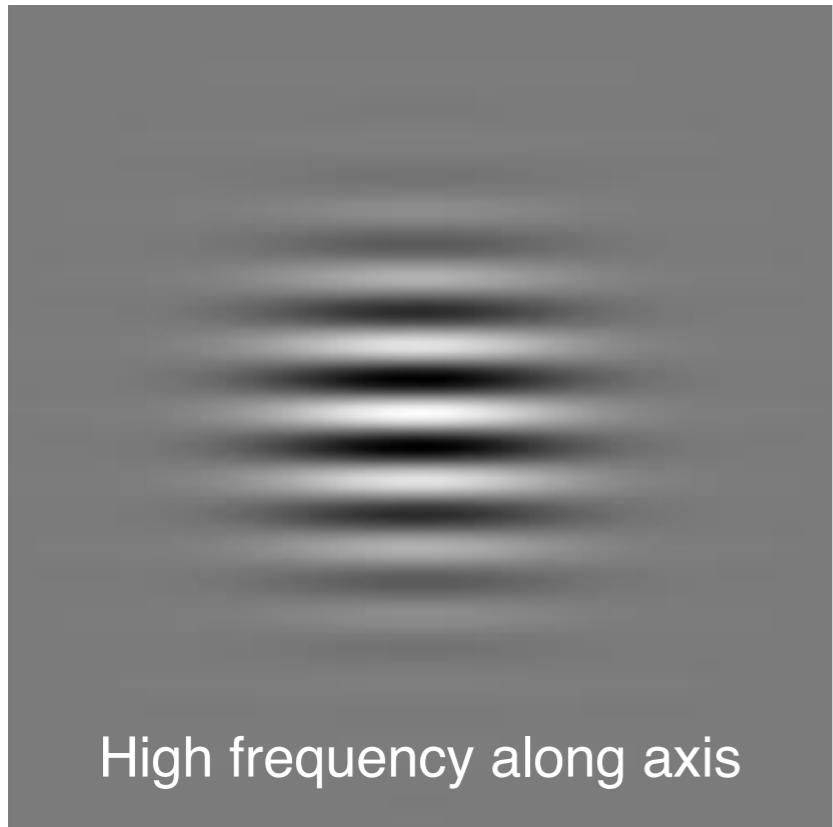
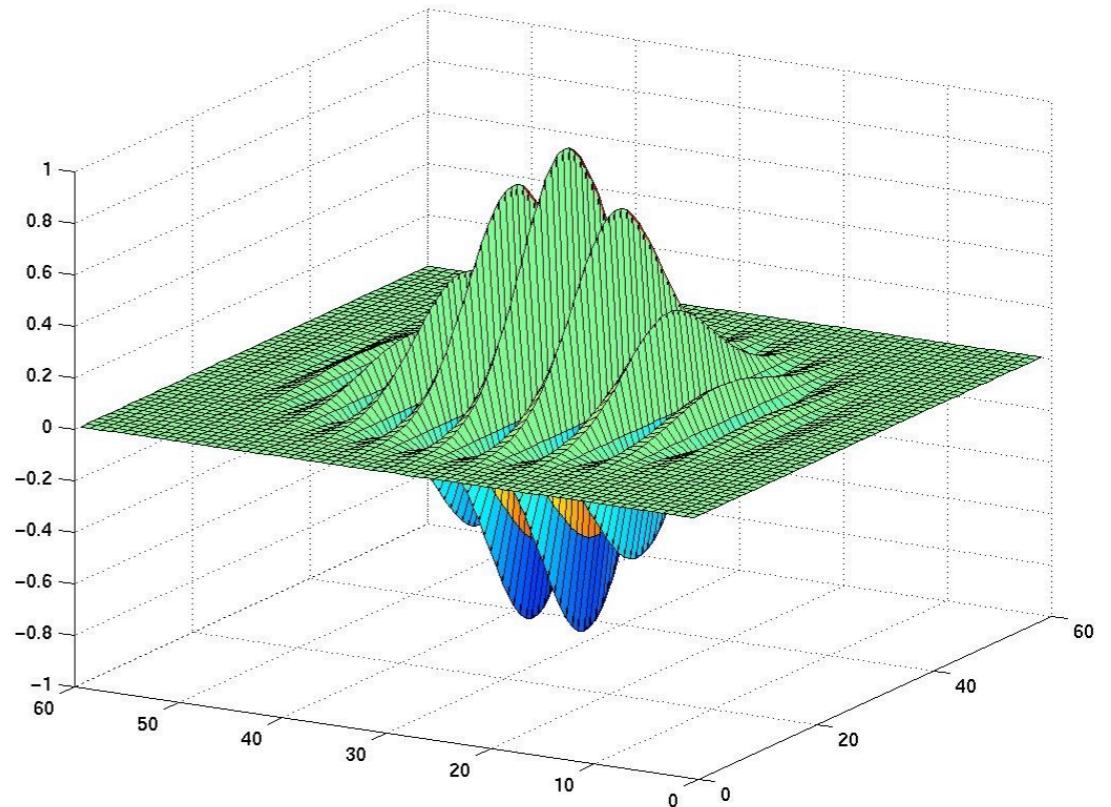
GIST

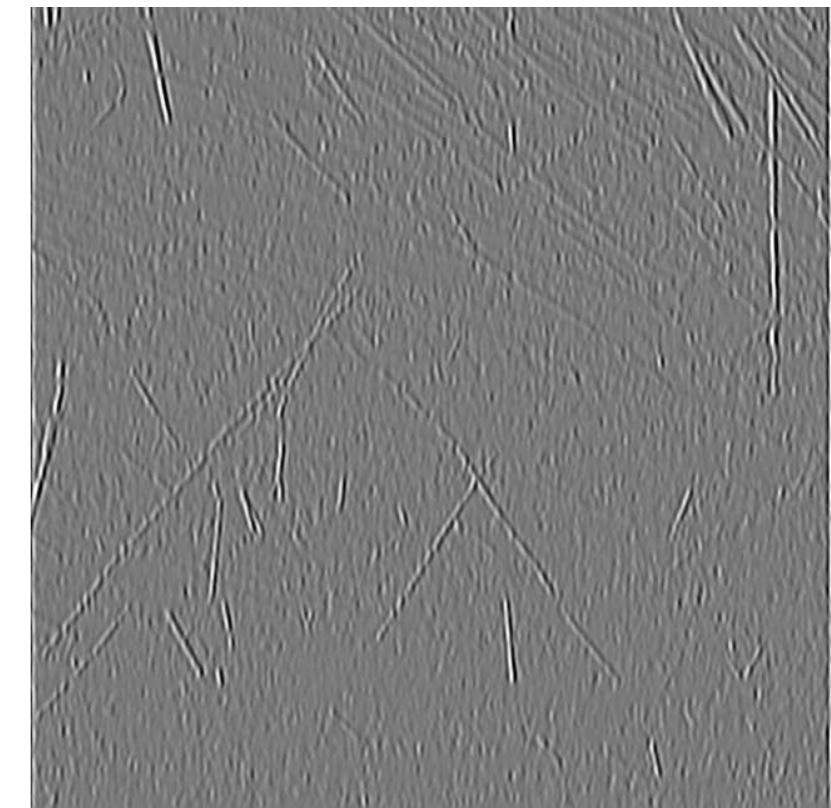
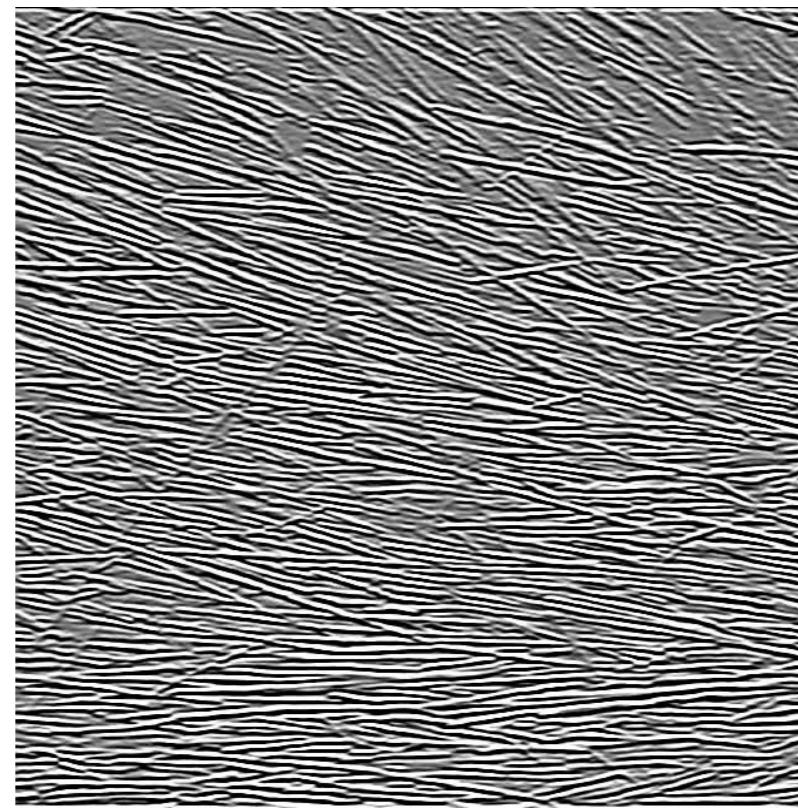
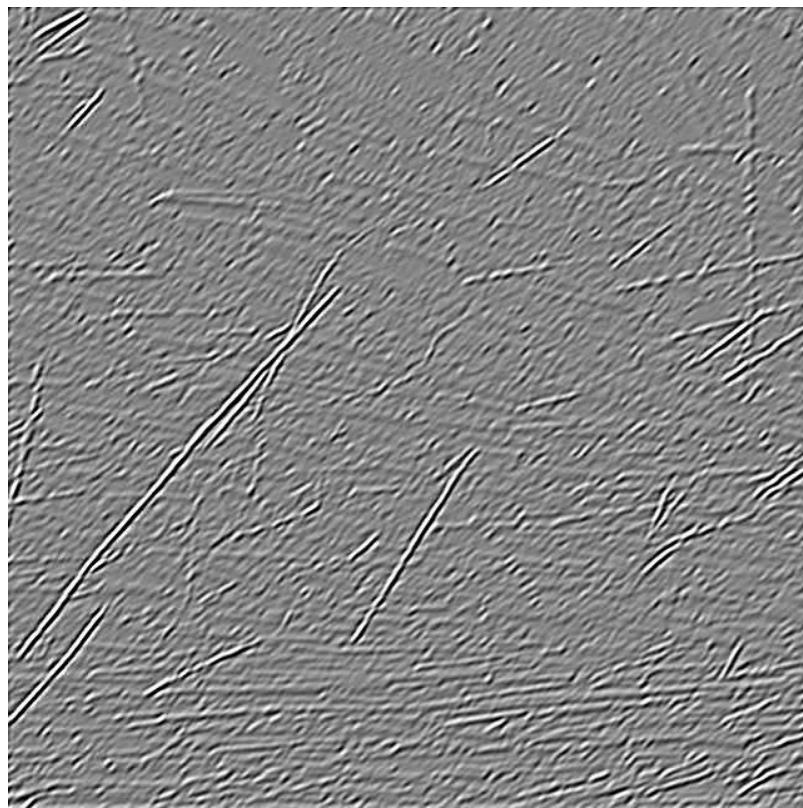
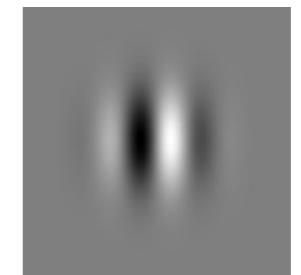
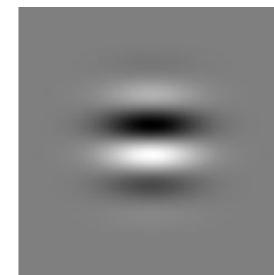
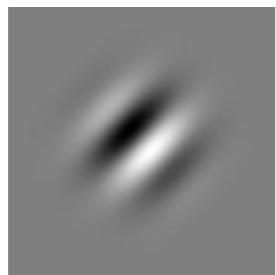
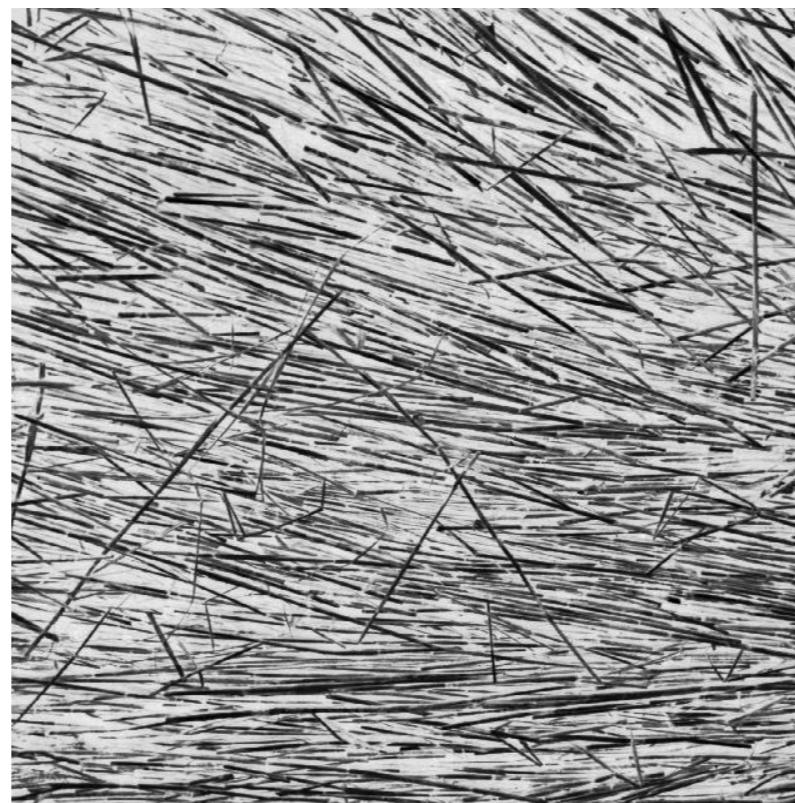
1. Compute filter responses (filter bank of Gabor filters)
2. Divide image patch into 4×4 cells
3. Compute filter response averages for each cell
4. Size of descriptor is $4 \times 4 \times N$, where N is the size of the filter bank

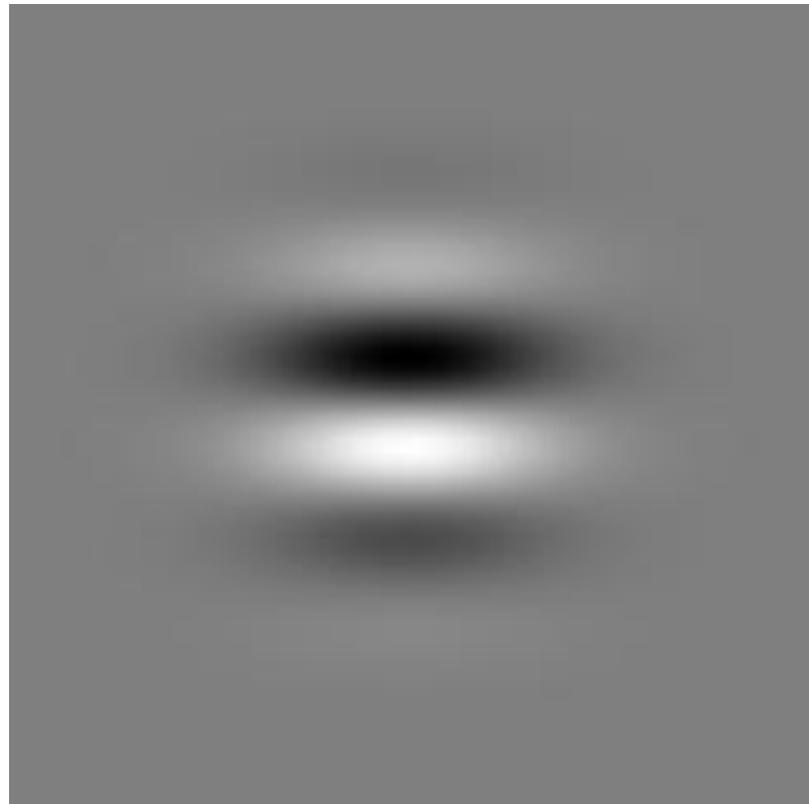


2D Gabor Filters

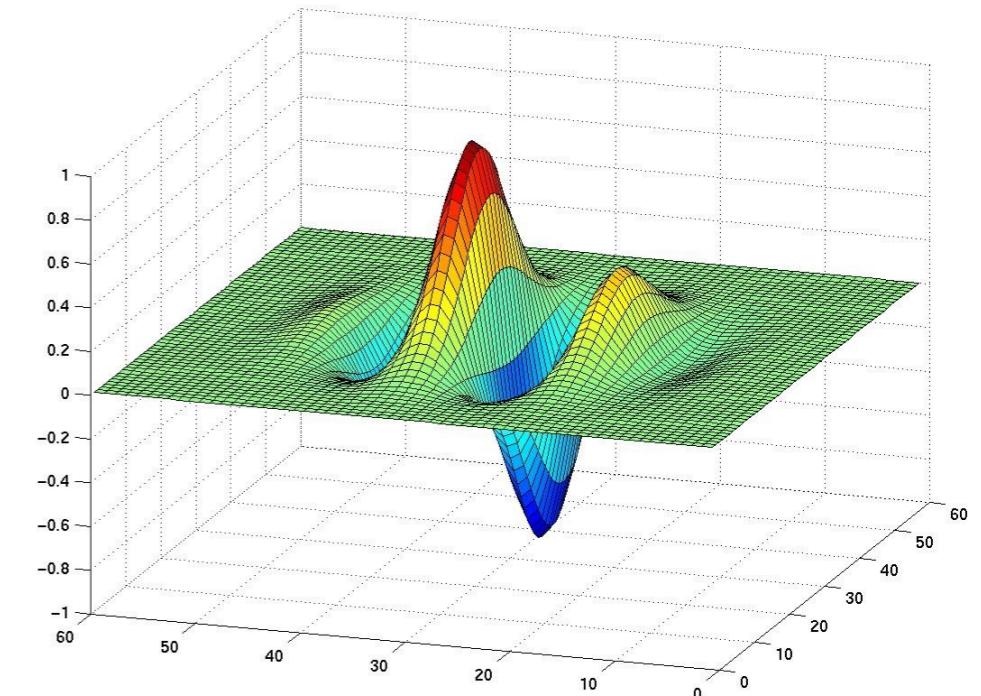
$$e^{-\frac{x^2+y^2}{2\sigma^2}} \cos(2\pi(k_x x + k_y y))$$



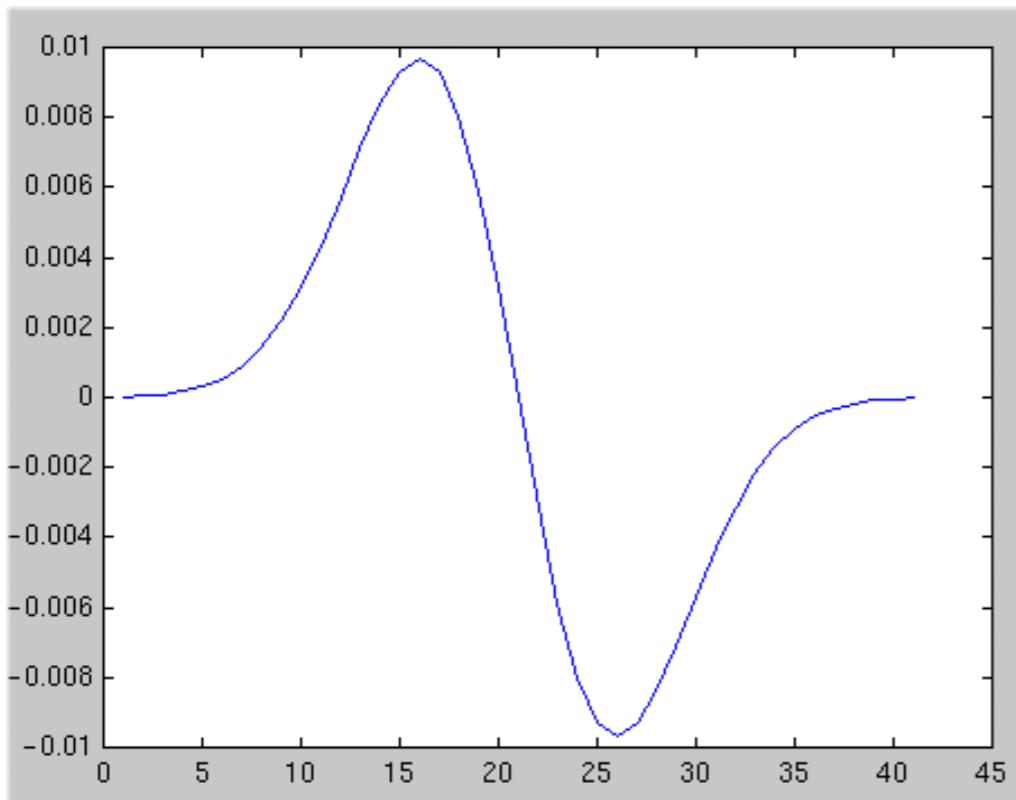




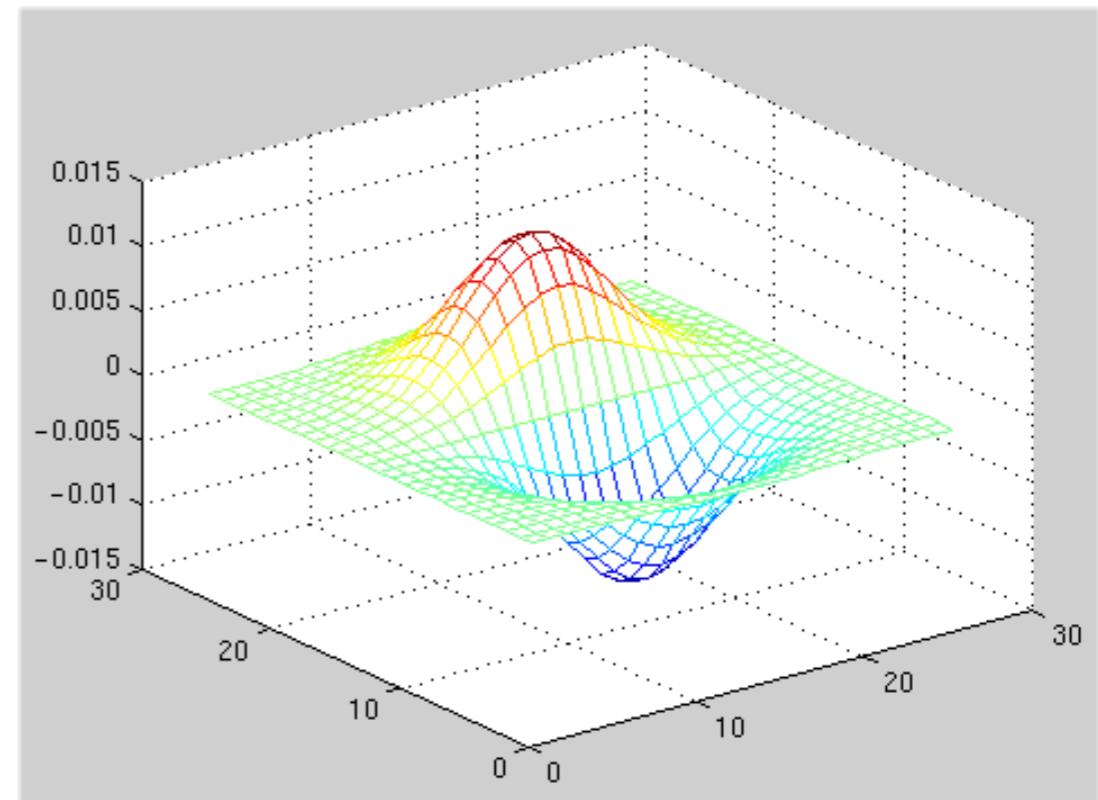
Odd
Gabor
filter

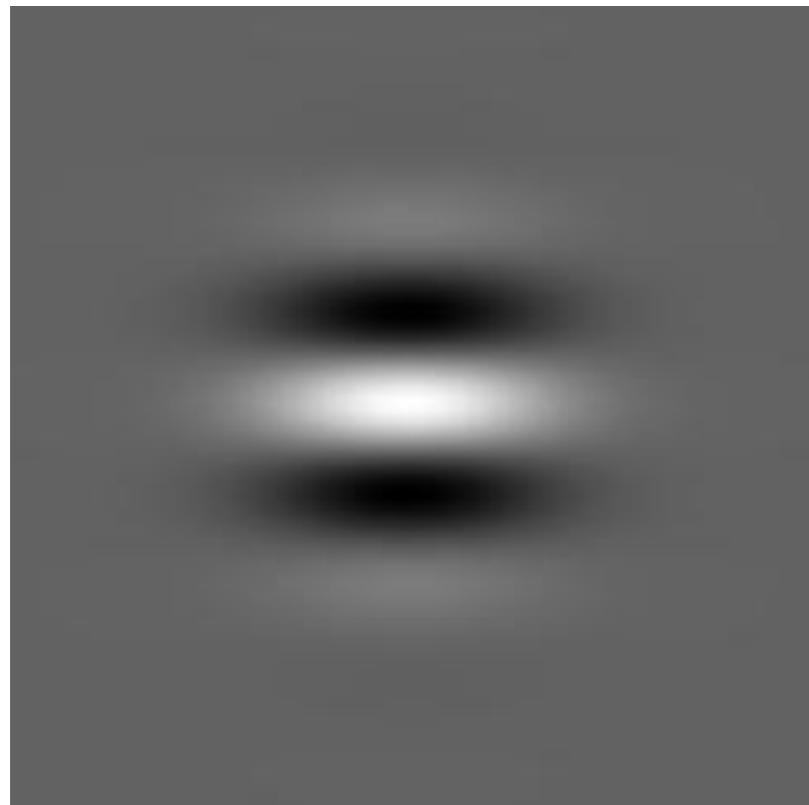


... looks a lot like...

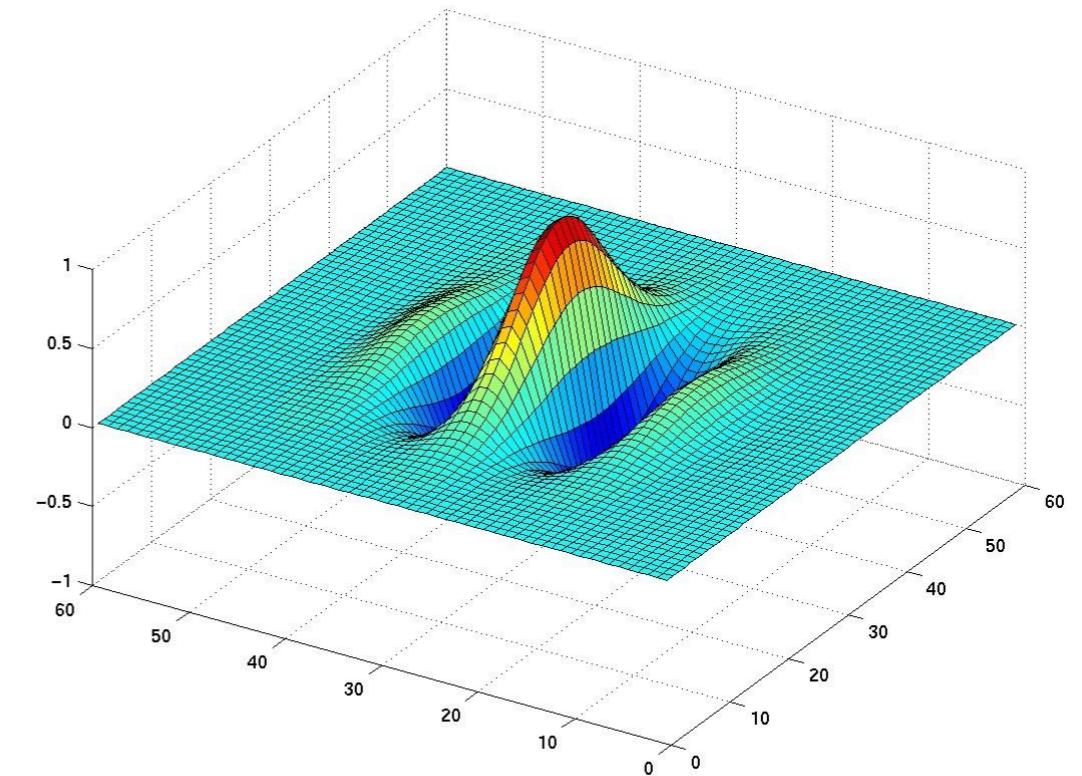


Gaussian
Derivative

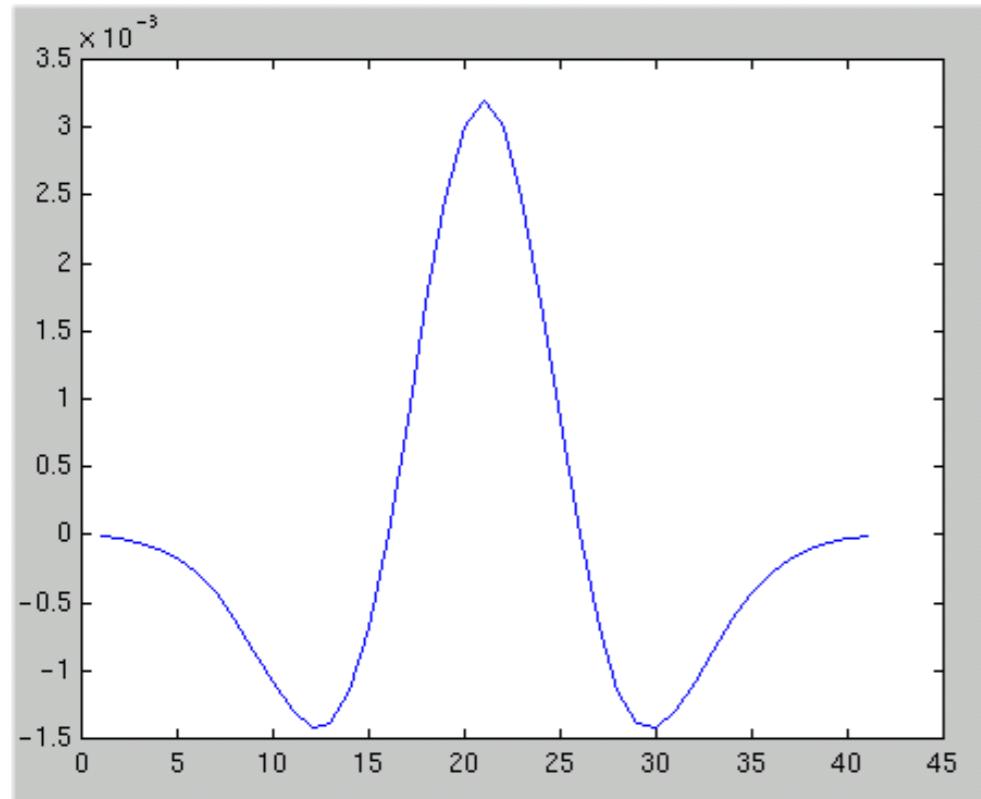




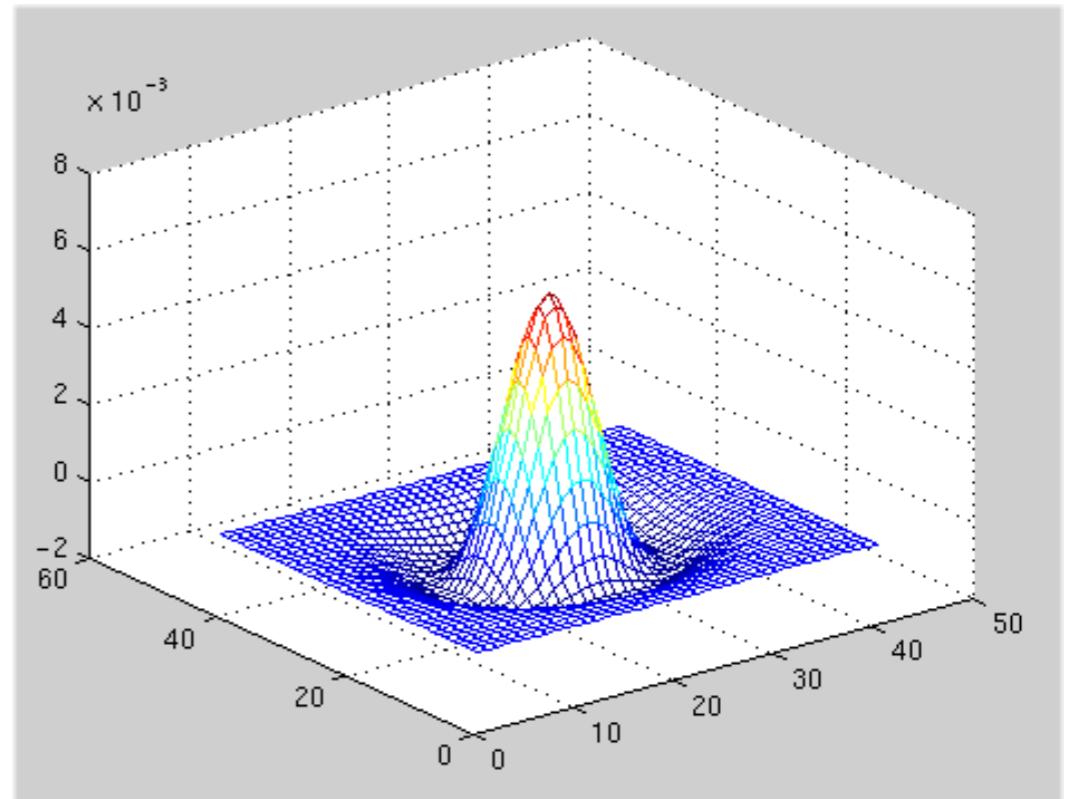
Even
Gabor
filter



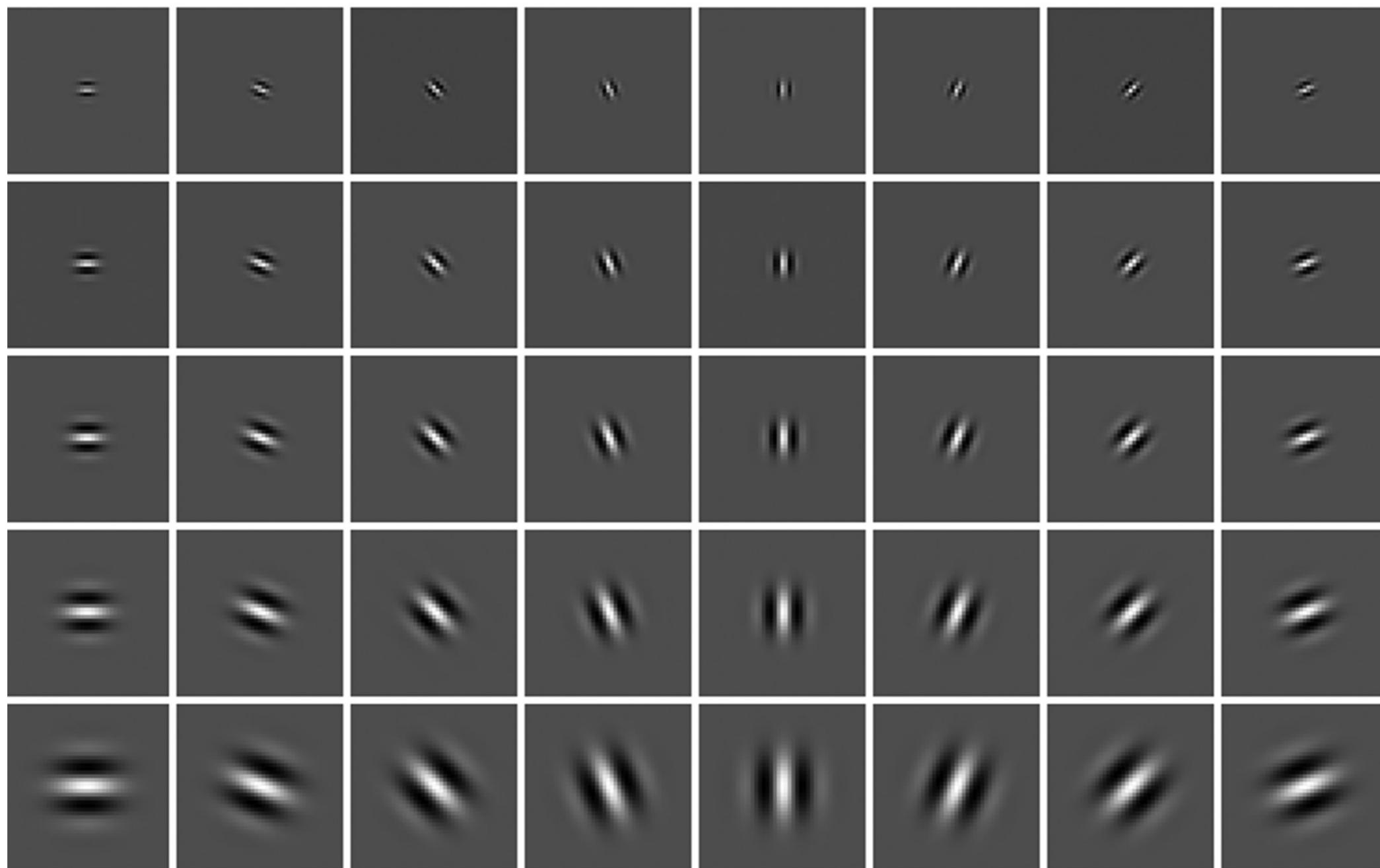
... looks a lot like...



Laplacian

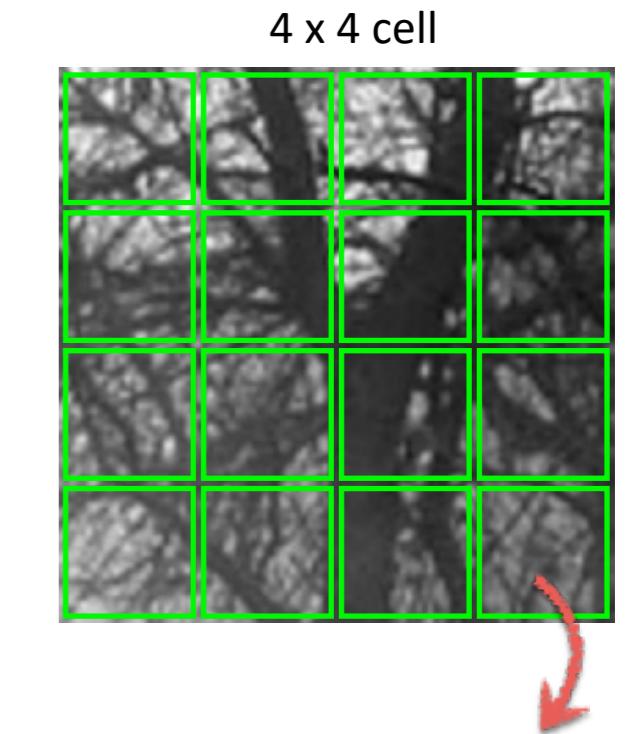
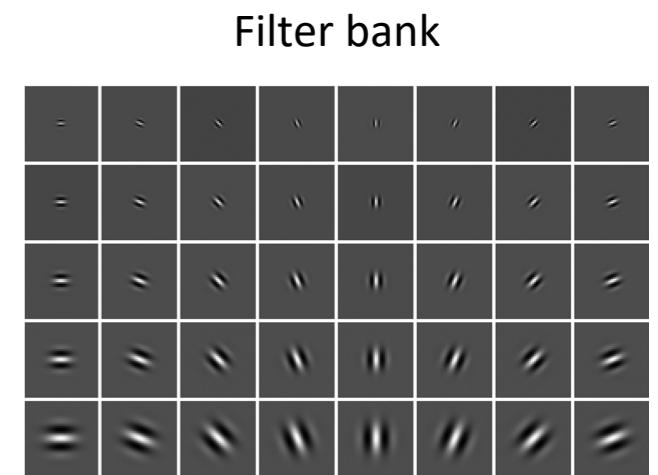


Directional edge detectors



GIST

1. Compute filter responses (filter bank of Gabor filters)
2. Divide image patch into 4×4 cells
3. Compute filter response averages for each cell
4. Size of descriptor is $4 \times 4 \times N$, where N is the size of the filter bank

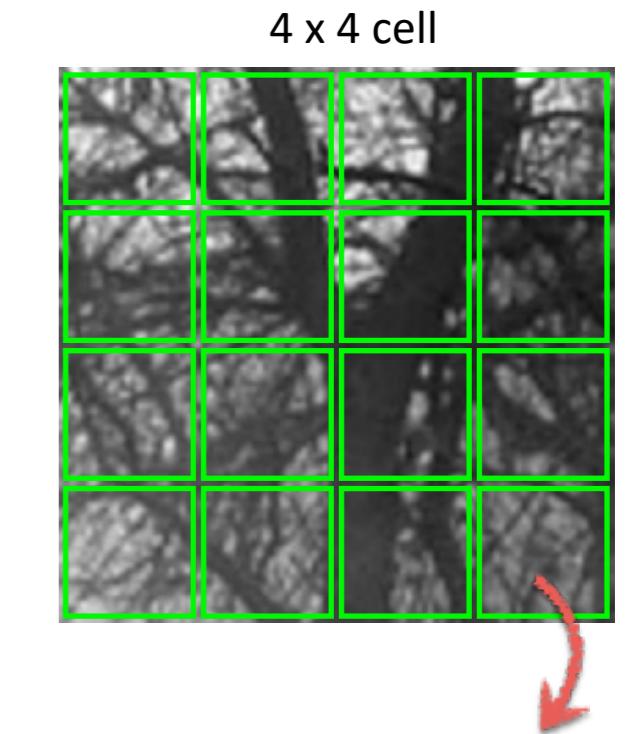
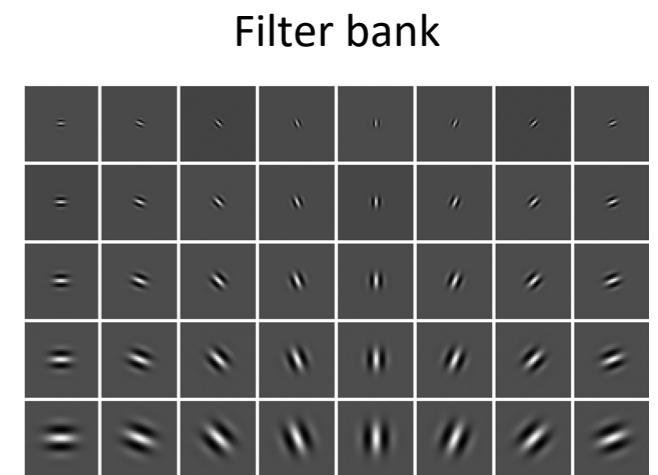


What is the GIST descriptor encoding?



GIST

1. Compute filter responses (filter bank of Gabor filters)
2. Divide image patch into 4×4 cells
3. Compute filter response averages for each cell
4. Size of descriptor is $4 \times 4 \times N$, where N is the size of the filter bank



What is the GIST descriptor encoding?

Rough spatial distribution of image gradients

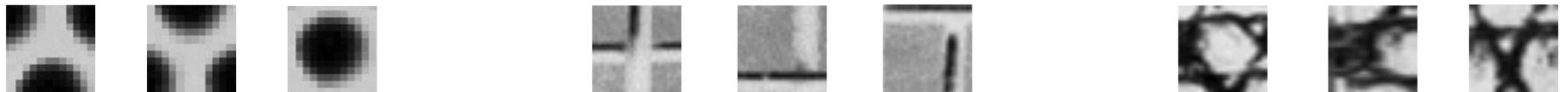


Histogram of Textons descriptor

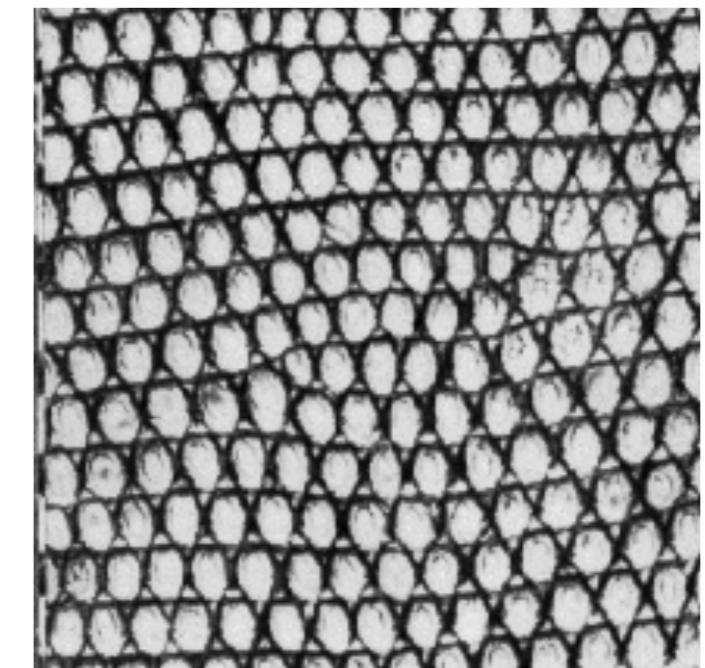
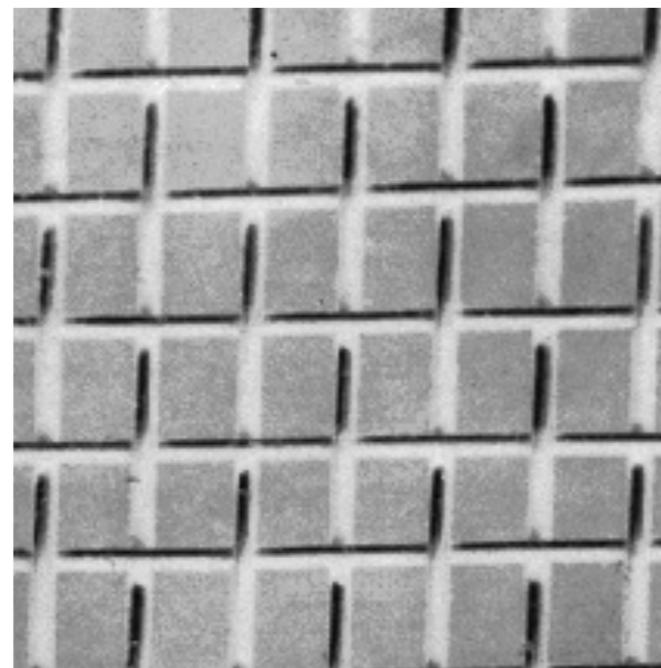
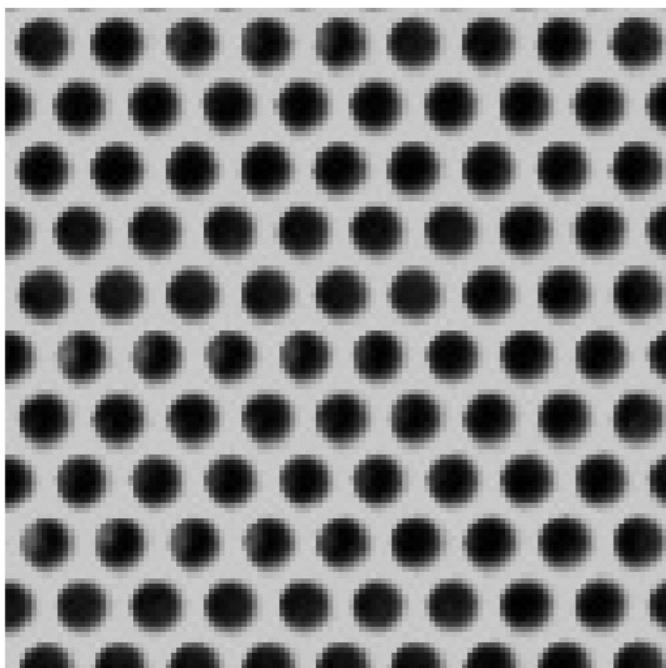
Textons

Julesz. Textons, the elements of texture perception, and their interactions. Nature 1981

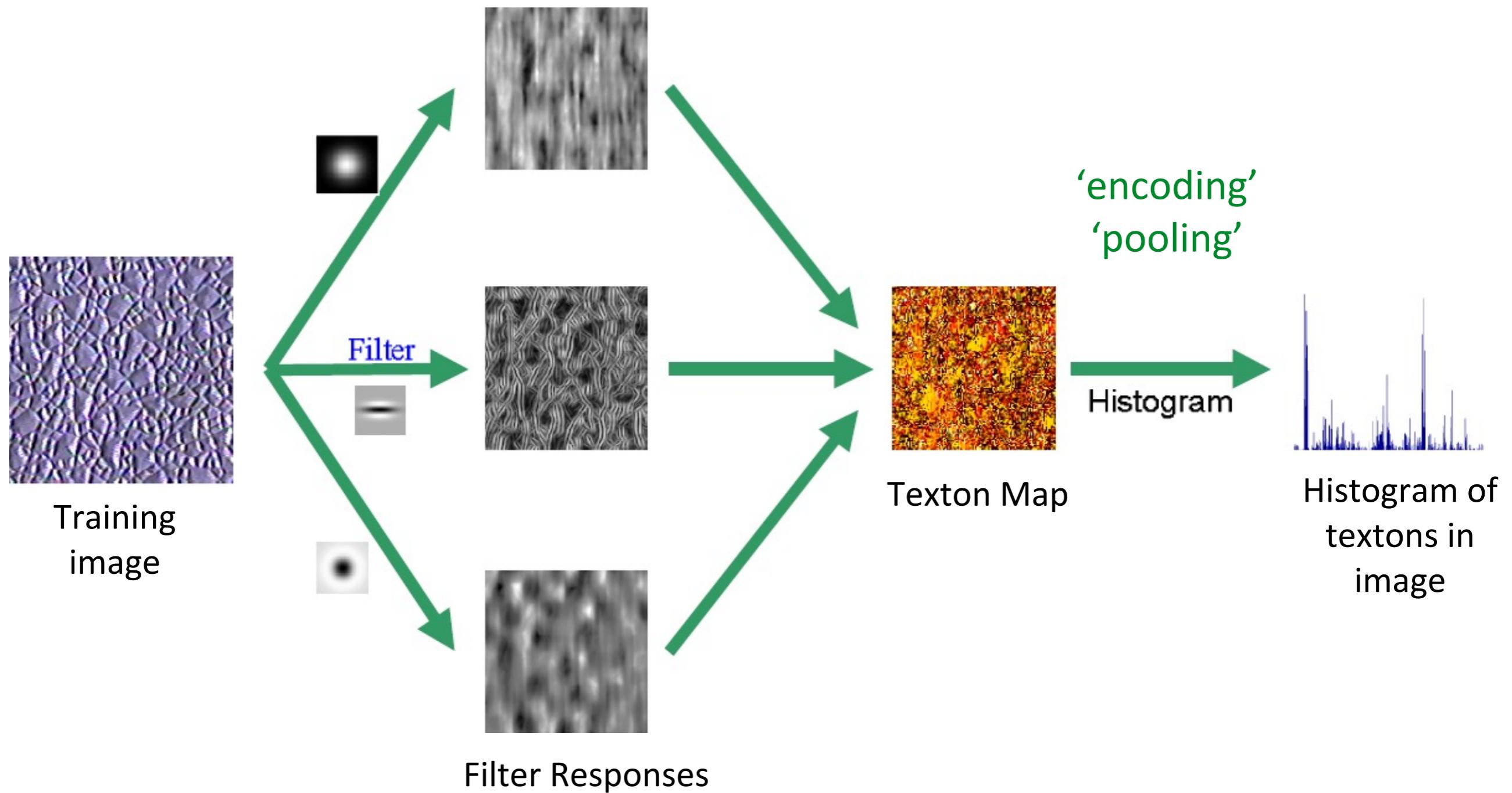
Texture is characterized by the repetition of basic elements or ***textons***



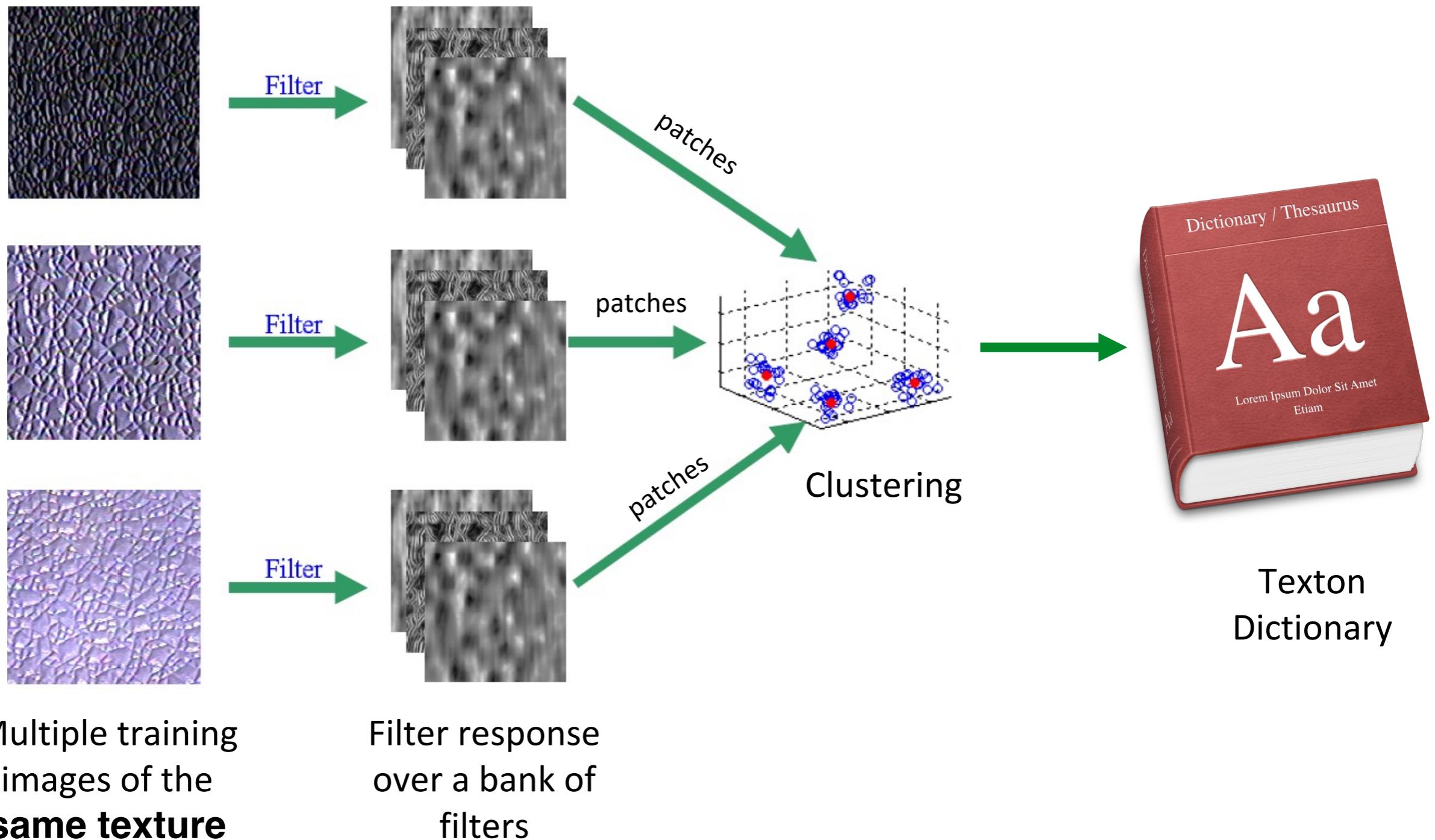
For stochastic textures, it is the identity of the ***textons***, not their spatial arrangement, that matters



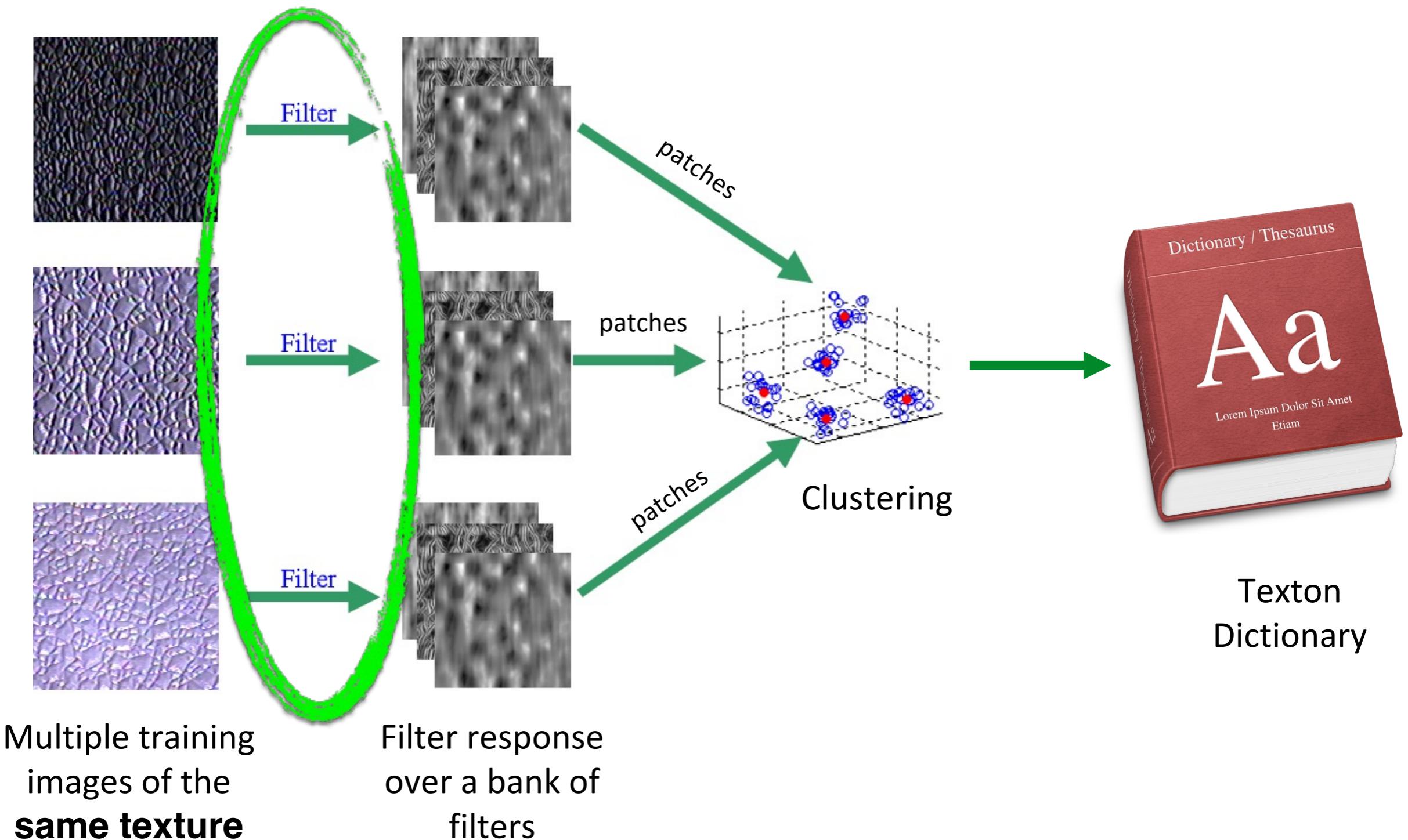
Histogram of Textons descriptor



Learning Textons from data

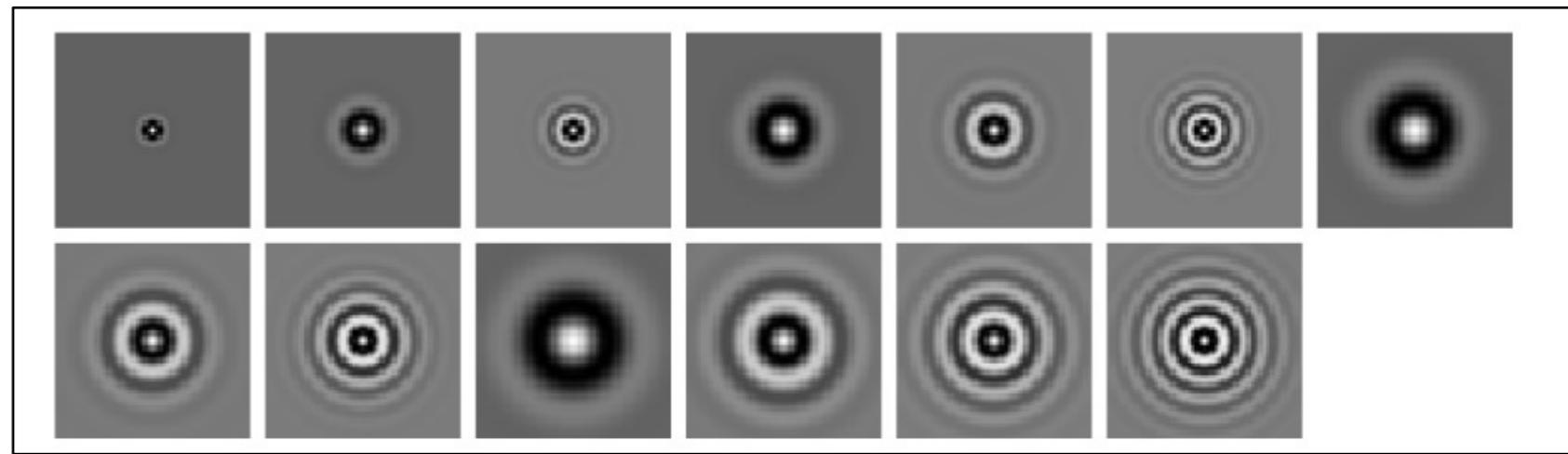


Learning Textons from data



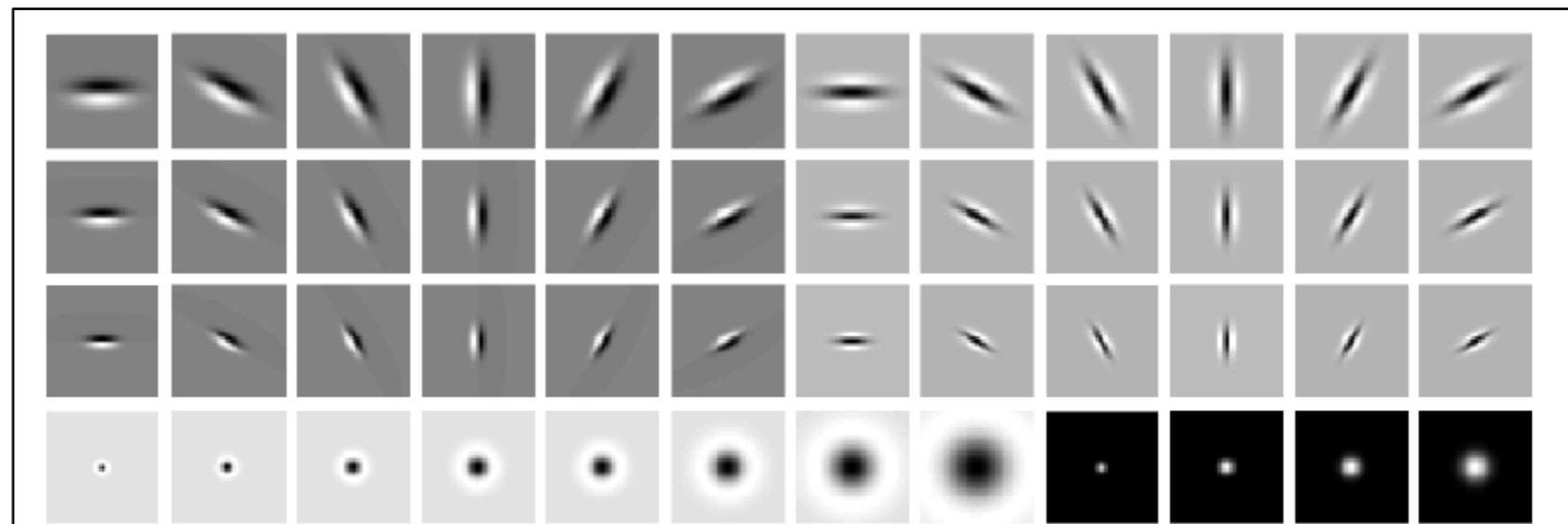
Example of Filter Banks

Isotropic Gabor

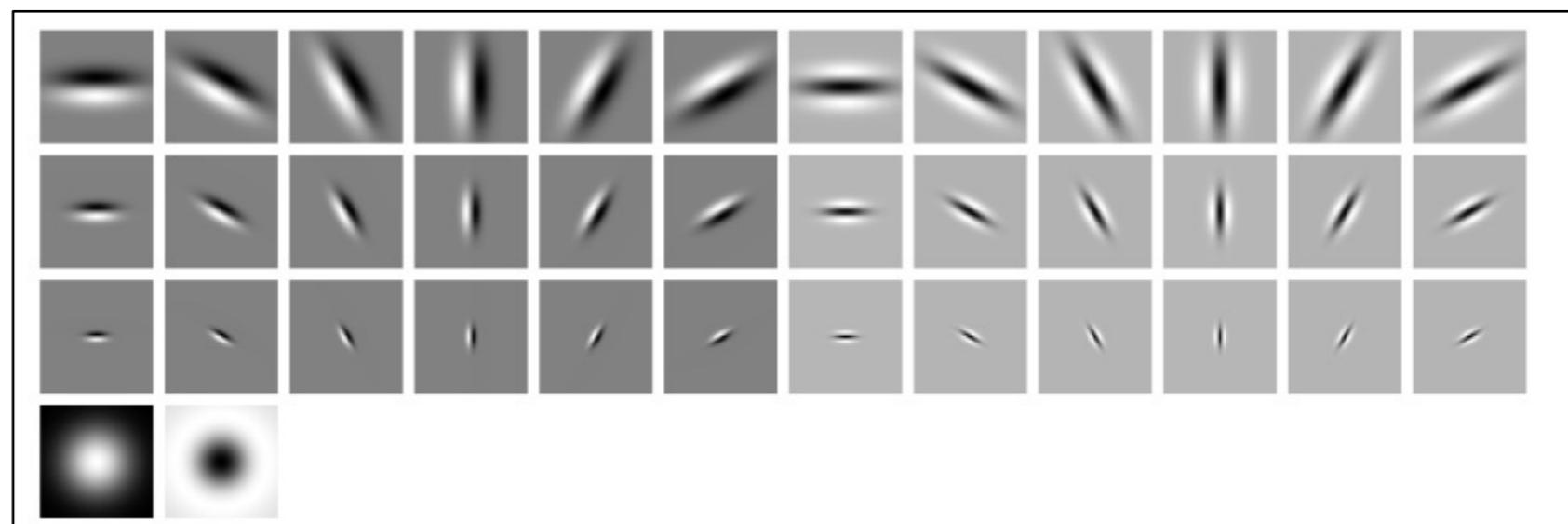


'S'

Gaussian derivatives at different scales and orientations

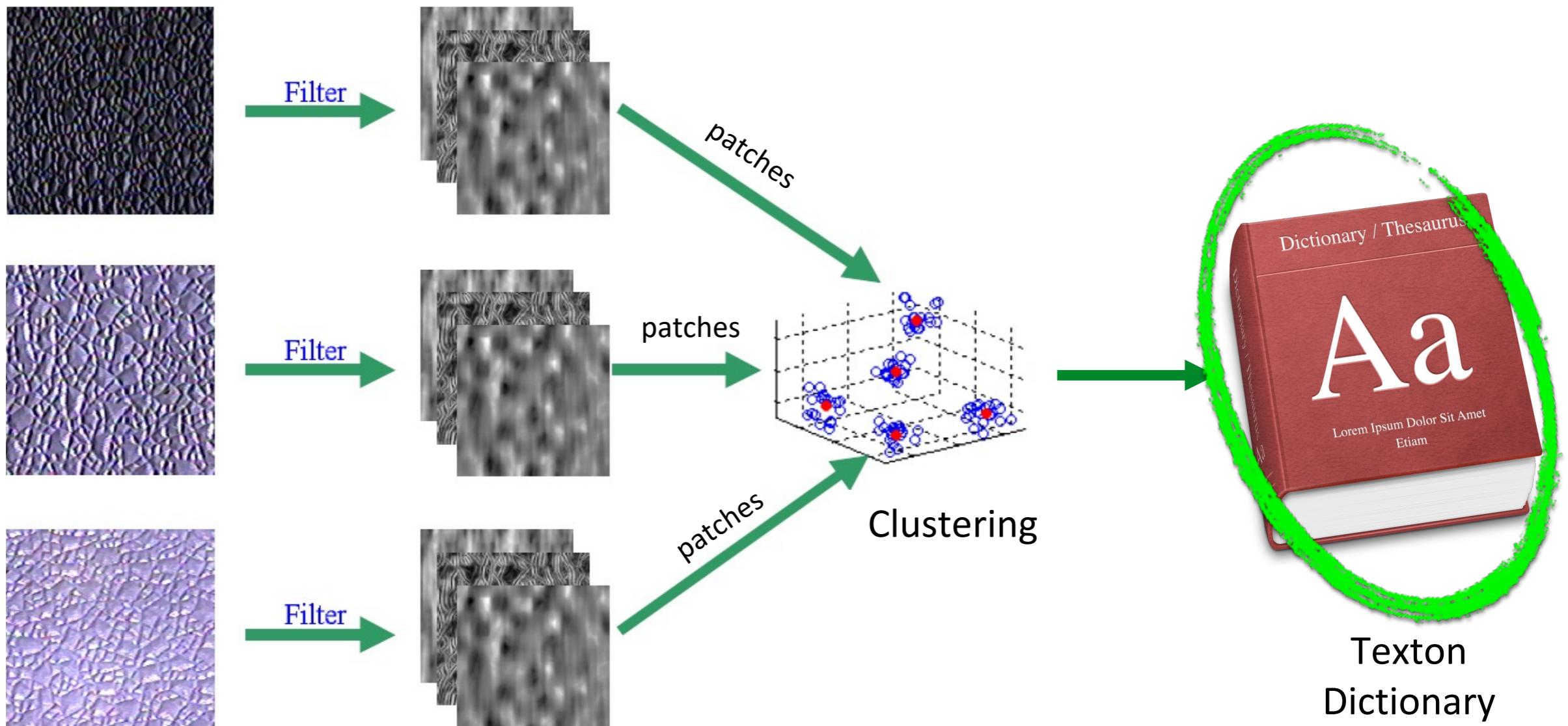


'LM'



'MR8'

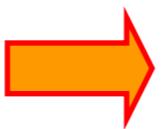
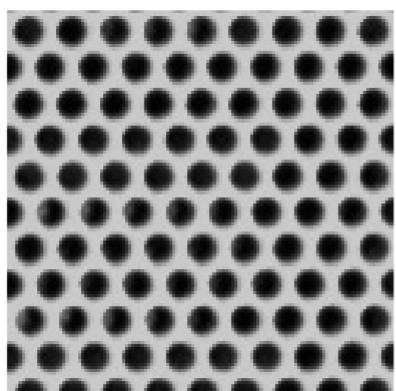
Learning Textons from data



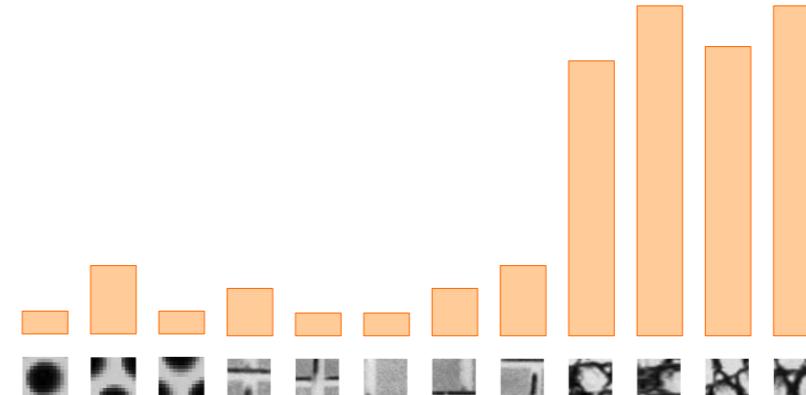
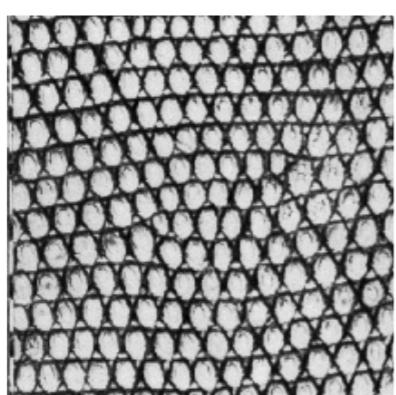
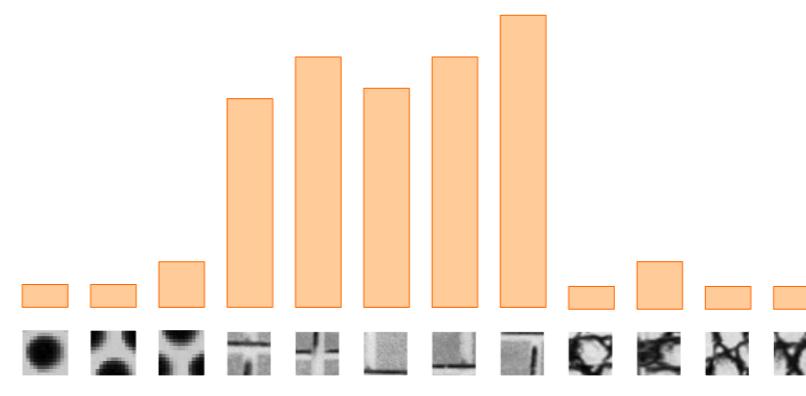
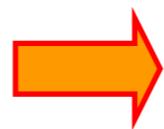
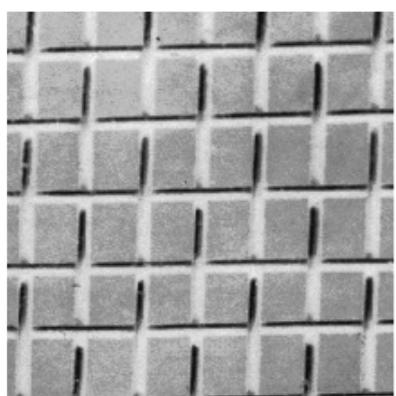
Multiple training
images of the same
texture

Filter response
over a bank of
filters

We will learn more about clustering
later in class (Bag of Words lecture).



Universal texton dictionary



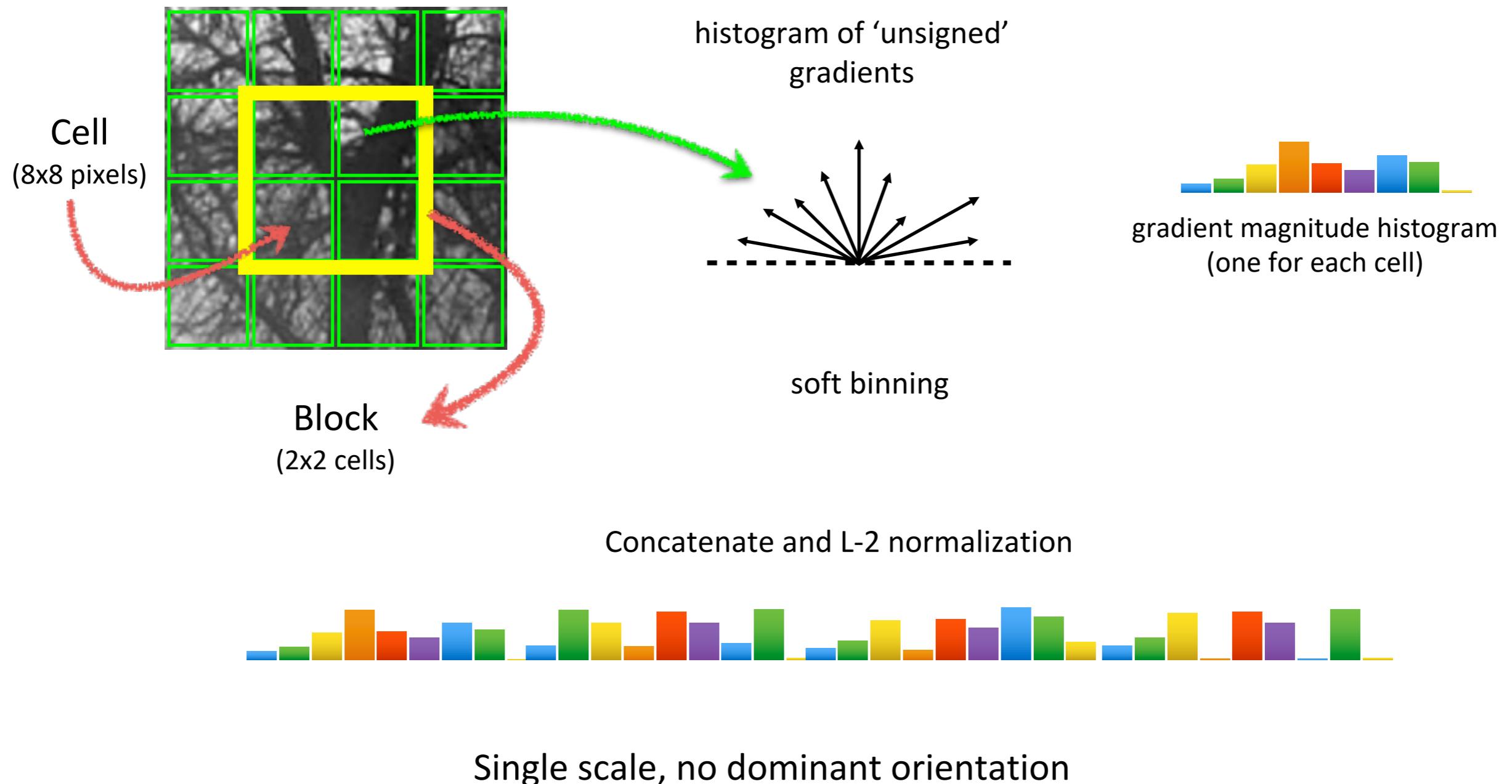
Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001;
Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

HOG descriptor

HOG



Dalal, Triggs. **Histograms of Oriented Gradients** for Human Detection. CVPR, 2005

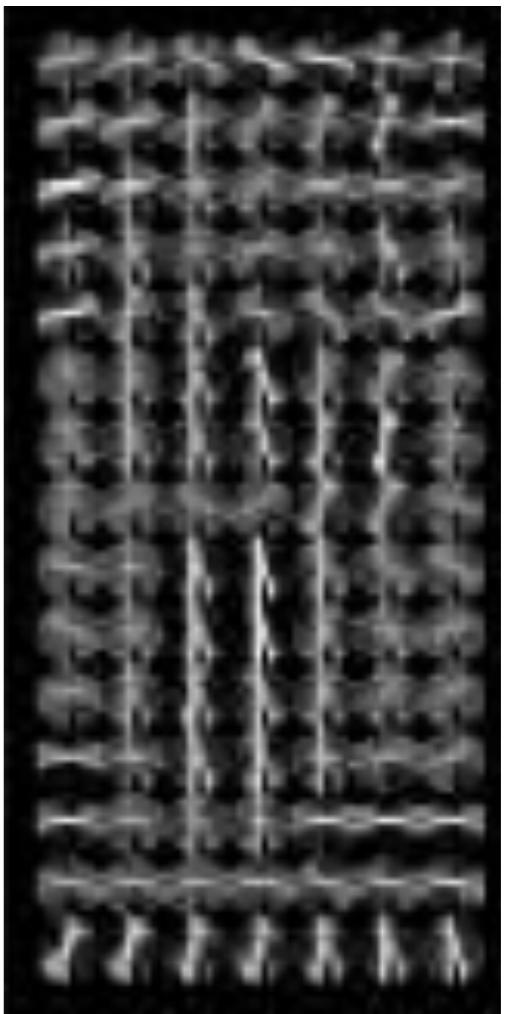


Pedestrian detection

128 pixels
16 cells
15 blocks

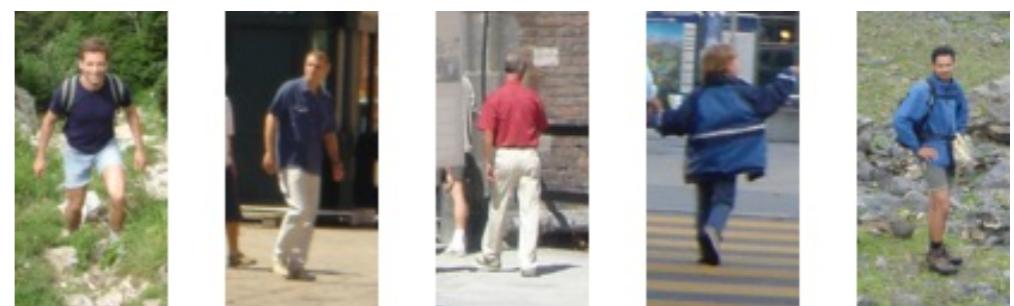


$$15 \times 7 \times 4 \times 9 = 3780$$



64 pixels
8 cells
7 blocks

Redundant representation due to overlapping blocks
How many times is each inner cell encoded?



SIFT



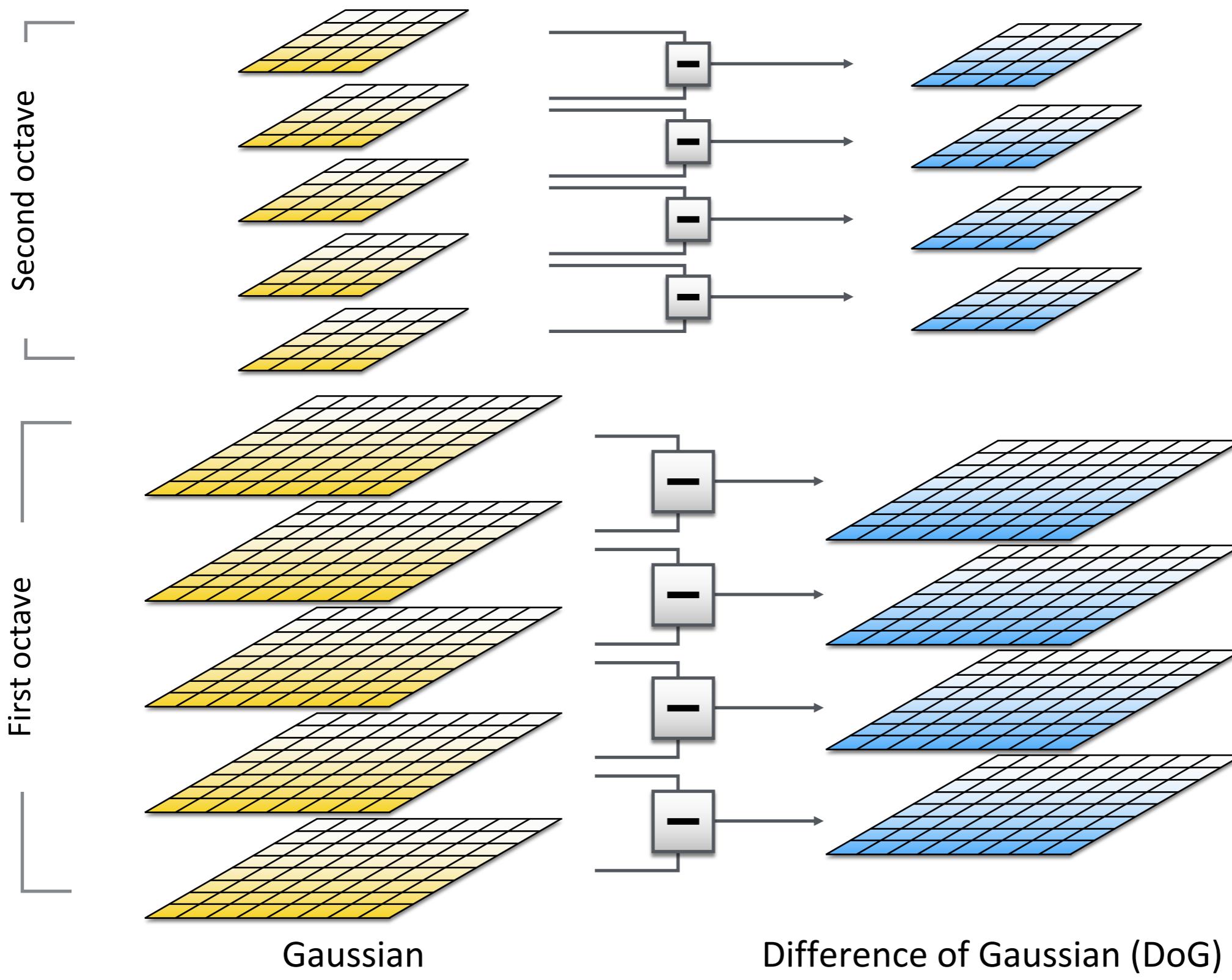
SIFT

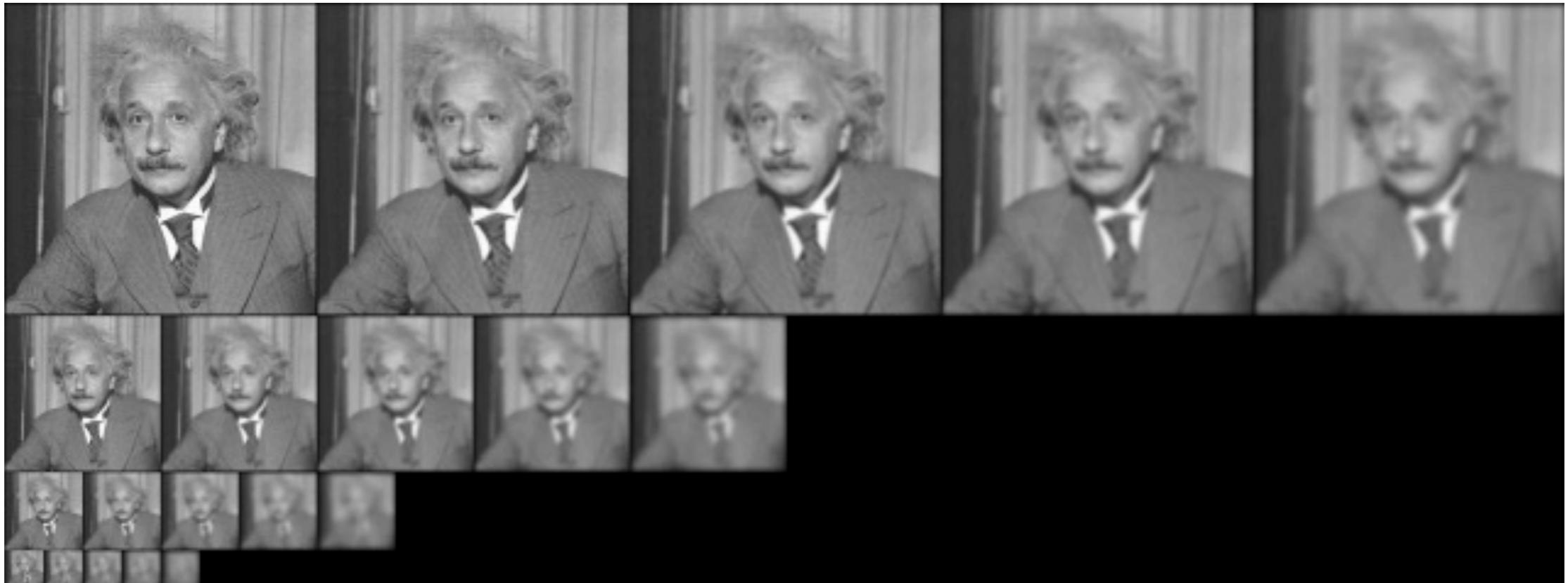
(Scale Invariant Feature Transform)

SIFT describes both a **detector** and **descriptor**

1. Multi-scale extrema detection
2. Keypoint localization
3. Orientation assignment
4. Keypoint descriptor

1. Multi-scale extrema detection



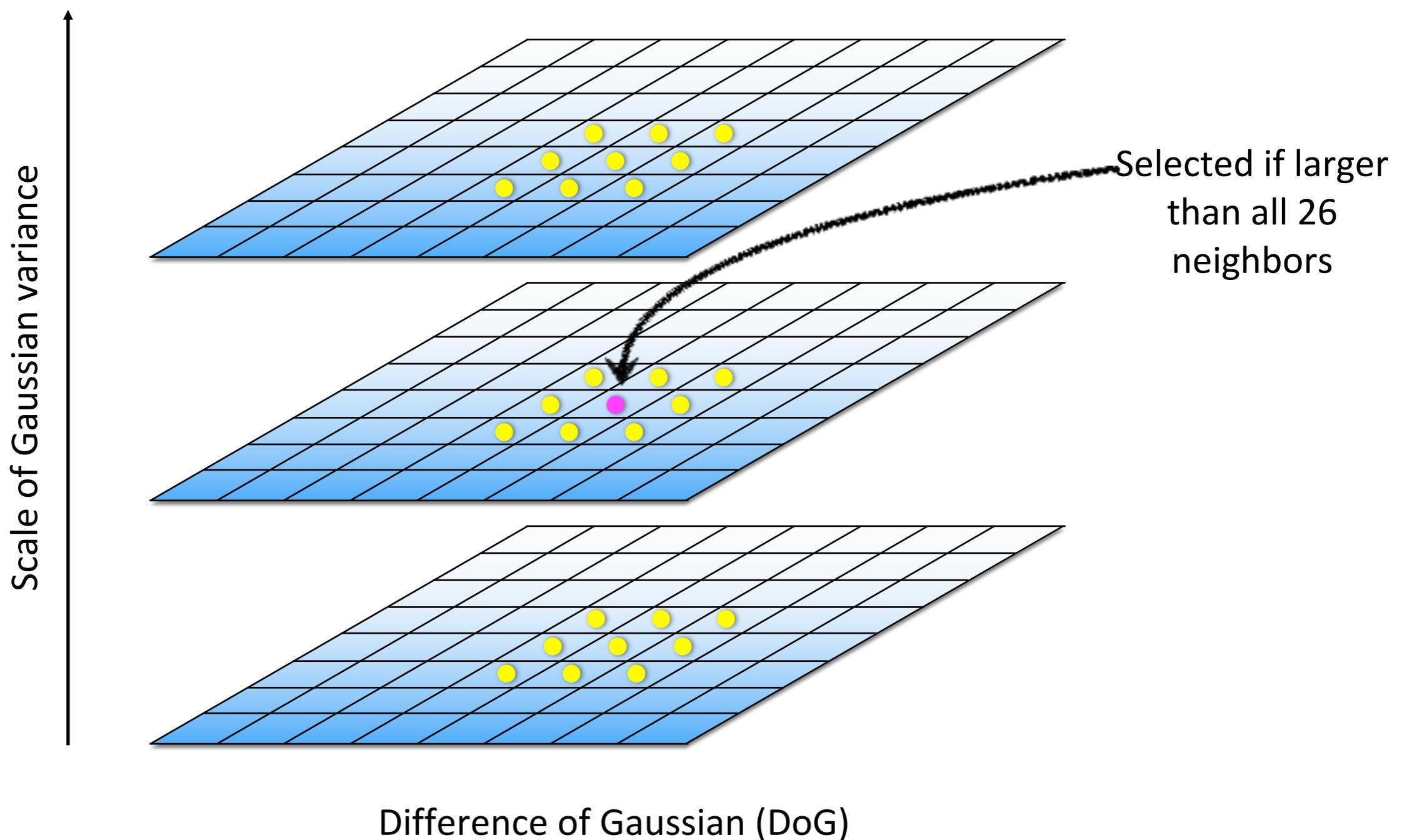


Gaussian



Laplacian

Scale-space extrema



2. Keypoint localization

2nd order Taylor series approximation of DoG scale-space

$$f(\mathbf{x}) = f + \frac{\partial f}{\partial \mathbf{x}}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 f}{\partial \mathbf{x}^2} \mathbf{x}$$

$$\mathbf{x} = \{x, y, \sigma\}$$

Take the derivative and solve for extrema

$$\mathbf{x}_m = - \frac{\partial^2 f}{\partial \mathbf{x}^2}^{-1} \frac{\partial f}{\partial \mathbf{x}}$$

Additional tests to retain only strong features

3. Orientation assignment

For a keypoint, \mathbf{L} is the **Gaussian-smoothed** image
with the closest scale,

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2}$$

x-derivative y-derivative

$$\theta(x, y) = \tan^{-1}((L(x, y + 1) - L(x, y - 1)) / (L(x + 1, y) - L(x - 1, y)))$$

Detection process returns

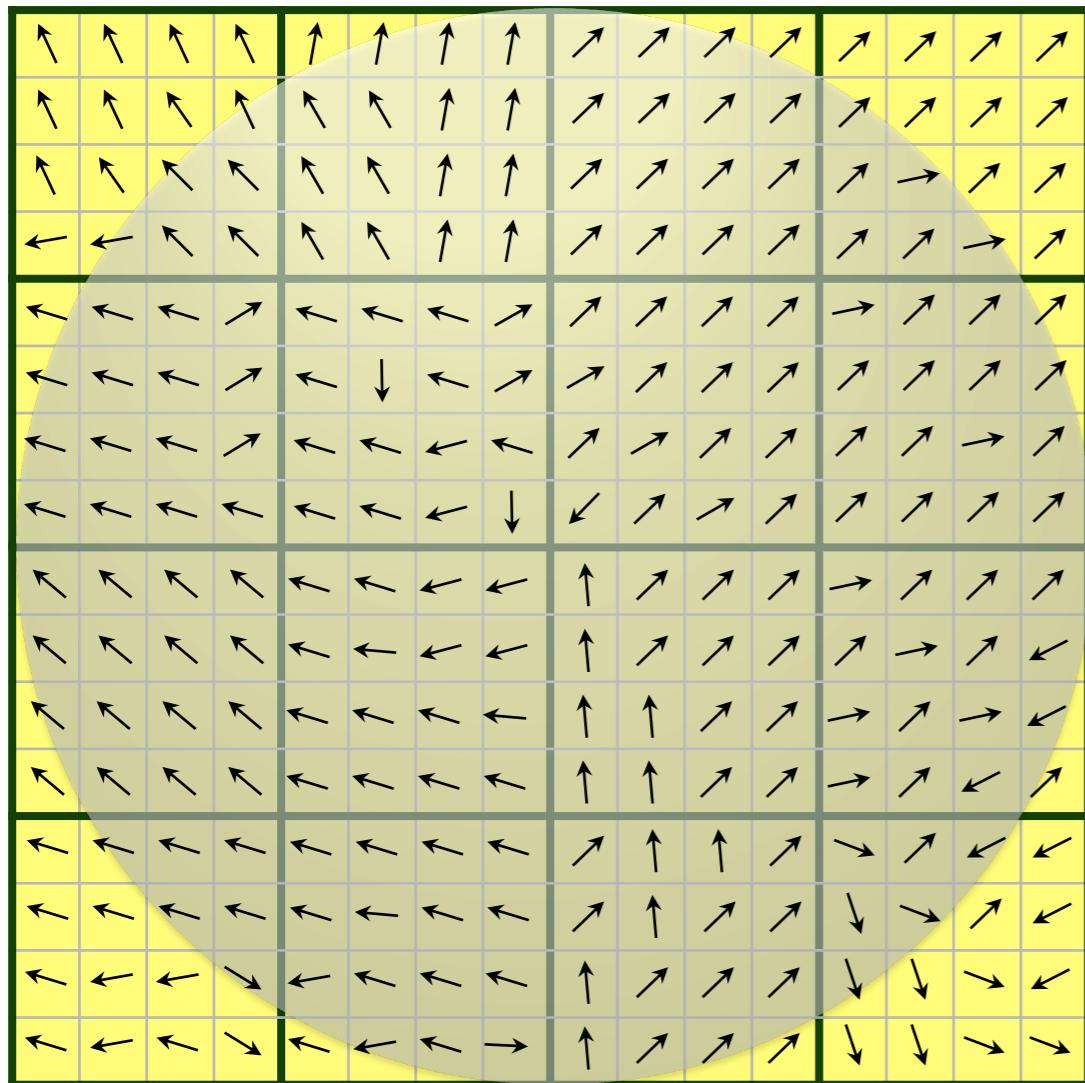
$$\{x, y, \sigma, \theta\}$$

location scale orientation

4. Keypoint descriptor

Image Gradients

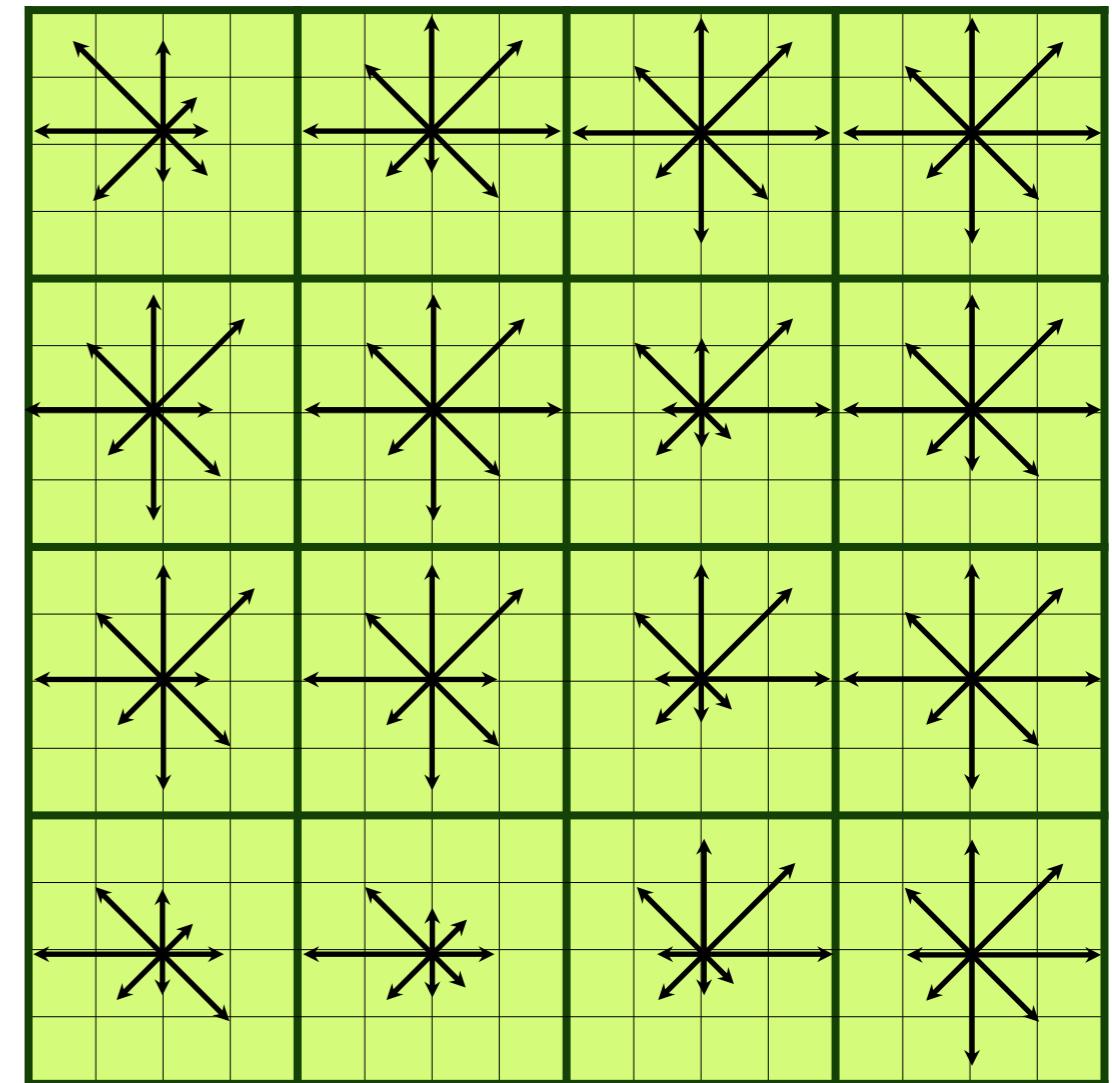
(4 x 4 pixel per cell, 4 x 4 cells)



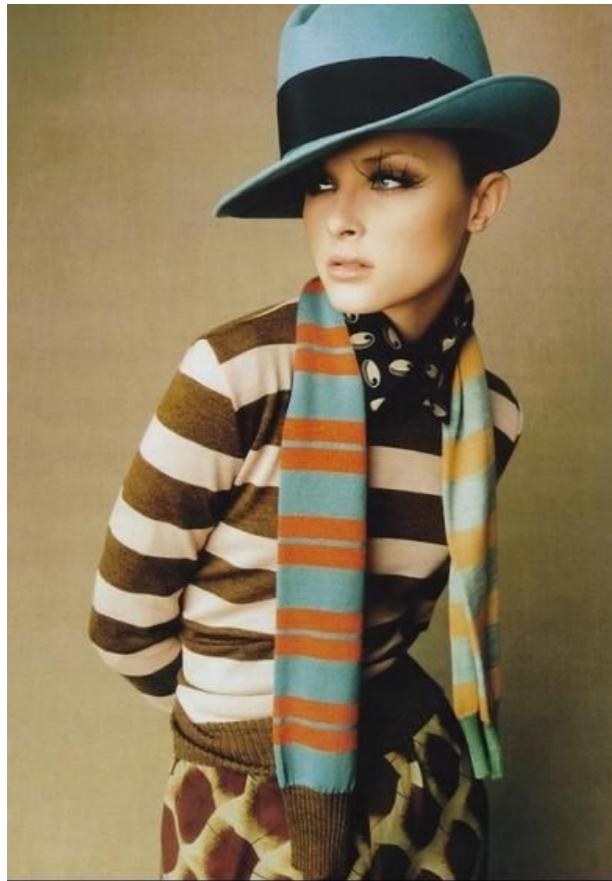
Gaussian weighting
(sigma = half width)

SIFT descriptor

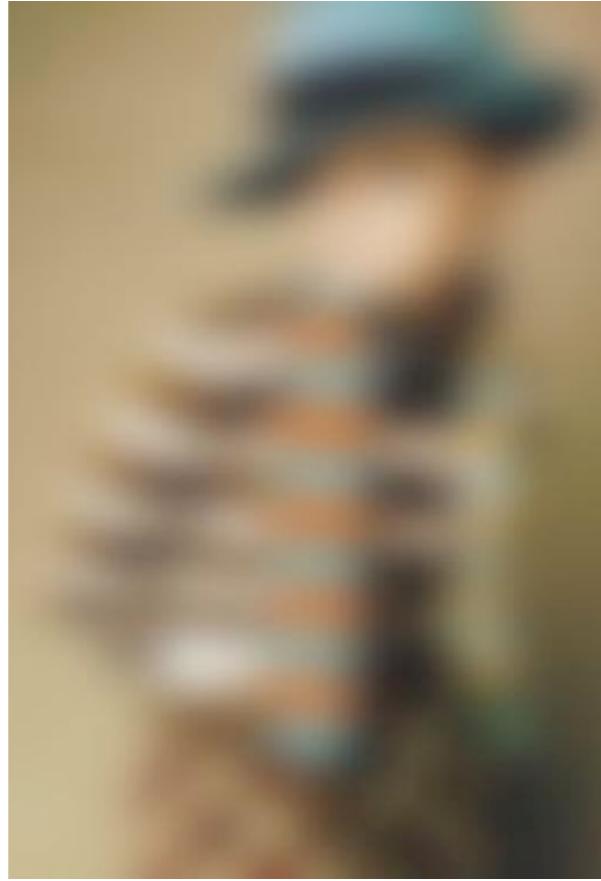
(16 cells x 8 directions = 128 dims)



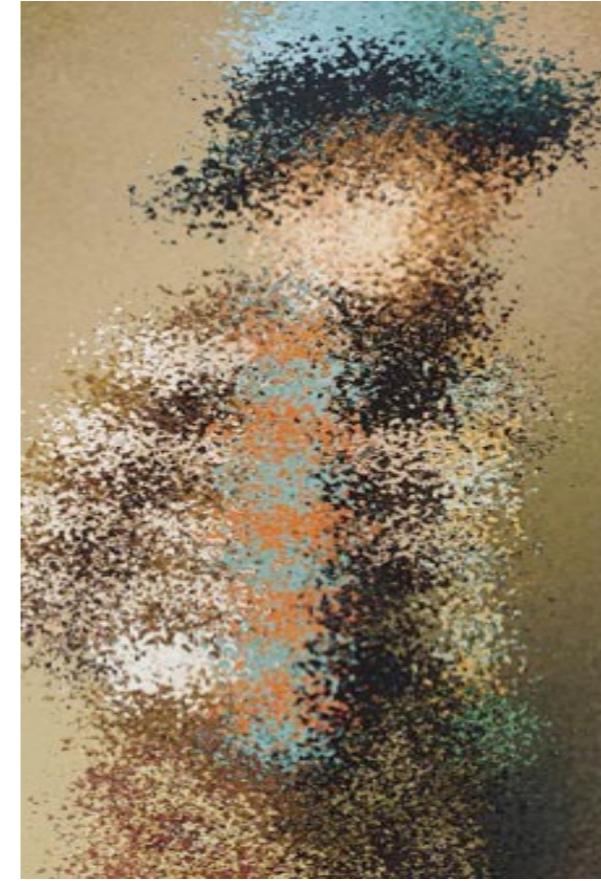
Discriminative power



Raw pixels



Sampled



Locally orderless



Global histogram

Generalization power

