CS 6550
Unit 2 – Image Features

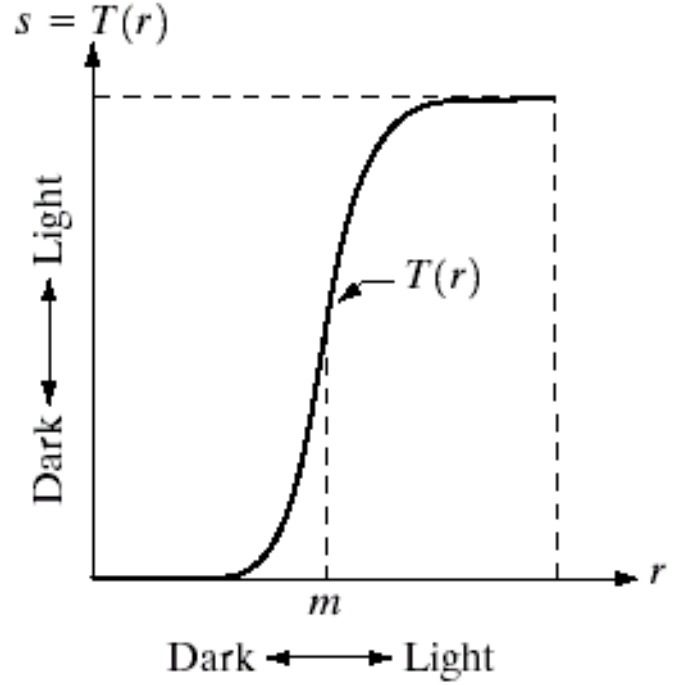
- **Intensity transformation**
- **Edge Detection**
- **Texture Features**
- **Corner Detection**
- **SIFT, Shape Context, LBP, …**

Reading: Szeliski Sec. 3.1, 3.2, 7.1, 7.2

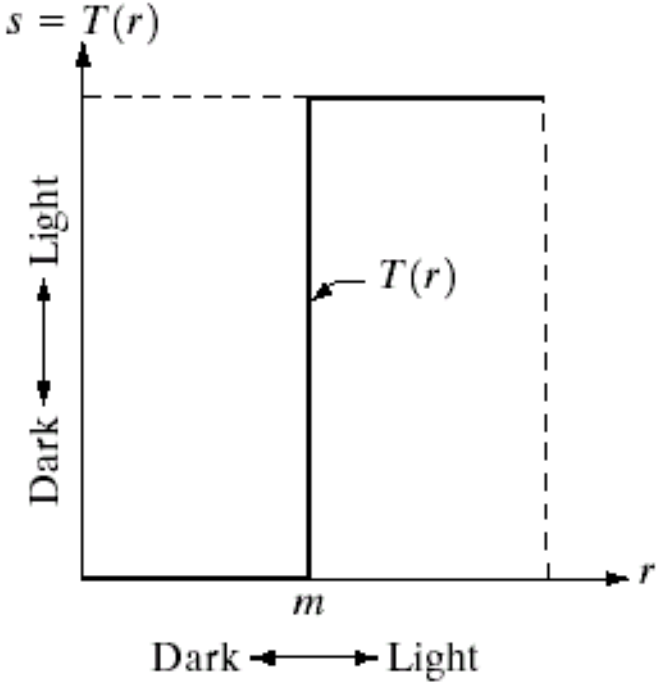
Grayscale Transformation

Point processing: s = T(r)

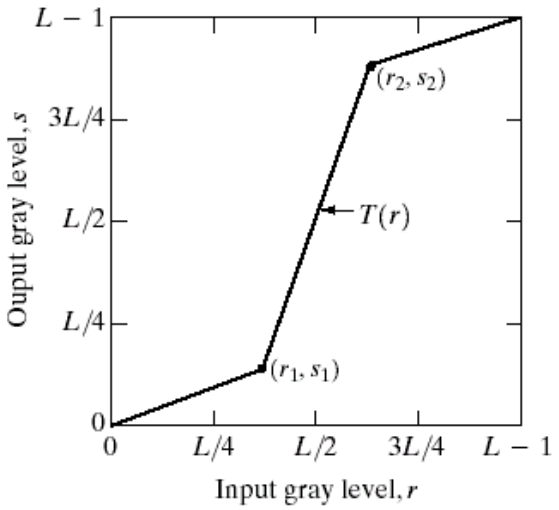
Contrast stretching

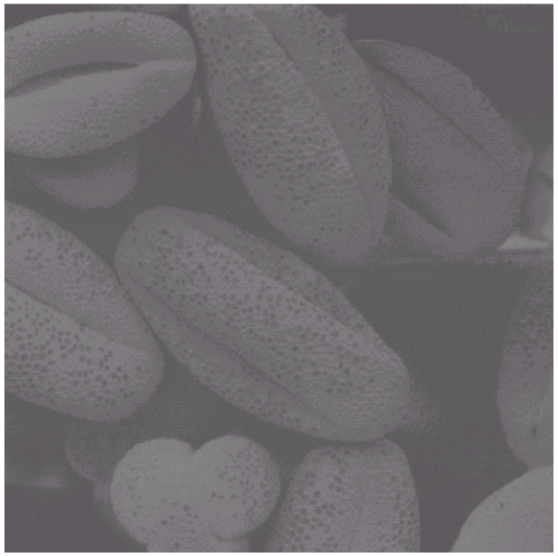


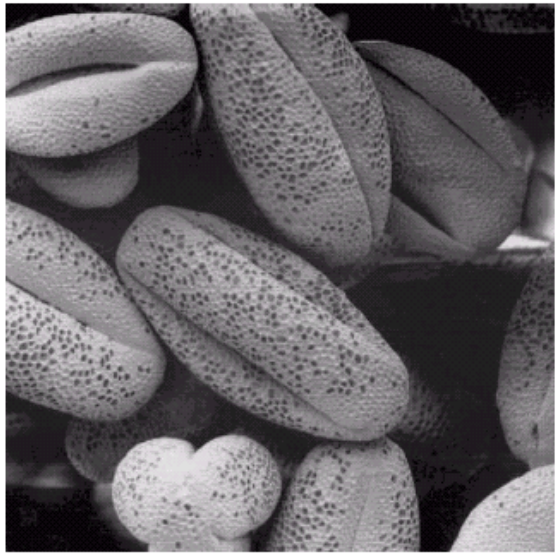
Thresholding



Contrast Stretching







a b
c d

FIGURE 3.10 Contrast stretching. (a) Form of transformation function. (b) A low-contrast image. (c) Result of contrast stretching. (d) Result of thresholding. (Original image courtesy of Dr. Roger Heady, Research School of Biological Sciences, Australian National University, Canberra, Australia.)

Algorithm 5.1: Histogram equalization

1. For an $N \times M$ image of G gray-levels (often 256), create an array H of length G initialized with 0 values.
2. Form the image histogram: Scan every pixel and increment the relevant member of H—if pixel p has intensity g_p, perform

$$H[g_p] = H[g_p] + 1 \, .$$

3. Form the cumulative image histogram H_c:

$$\begin{aligned} H_c[0] &= H[0] \, , \\ H_c[p] &= H_c[p-1] + H[p] \, , \quad p = 1, 2, \ldots, G-1 \, . \end{aligned}$$

4. Set

$$T[p] = \text{round} \left(\frac{G-1}{NM} H_c[p] \right) \, .$$

(This step obviously lends itself to more efficient implementation by constructing a look-up table of the multiples of $(G-1)/NM$, and making comparisons with the values in H_c, which are monotonically increasing.)

5. Rescan the image and write an output image with gray-levels g_q, setting

$$g_q = T[g_p] \, .$$

Histogram Equalization Example

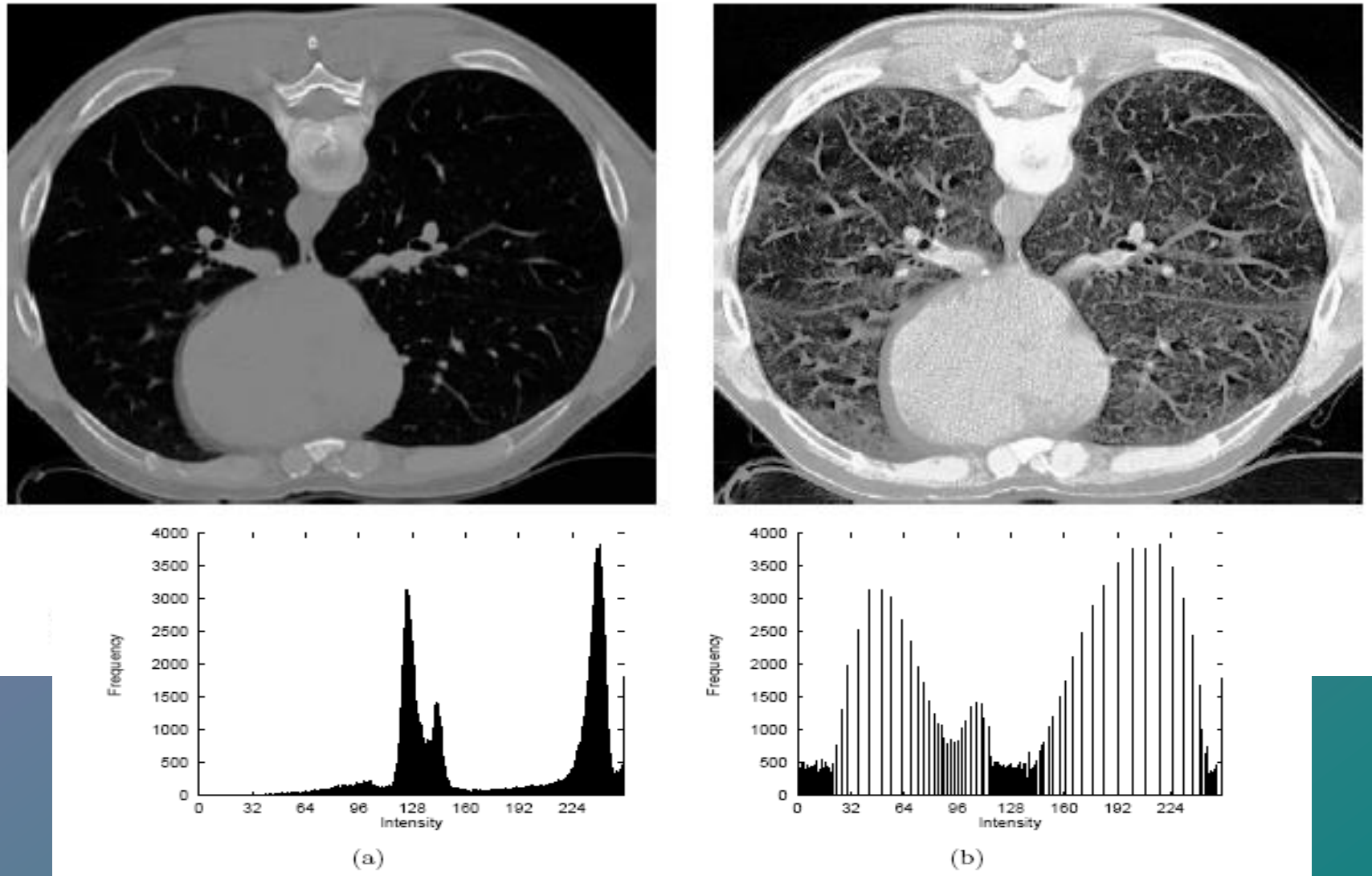


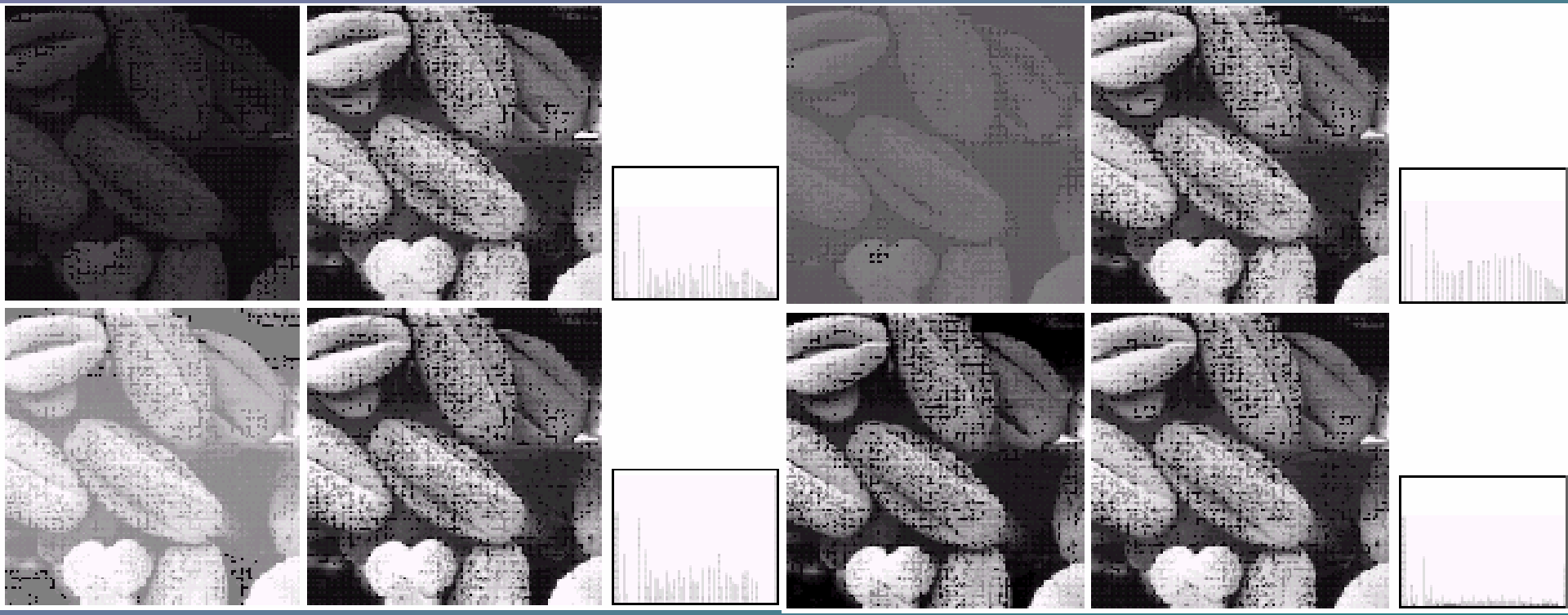
Figure 5.4: Histogram equalization: Original and equalized histograms corresponding to Figure 5.3a,b.

Histogram Equalization Example

Original image *Pout*

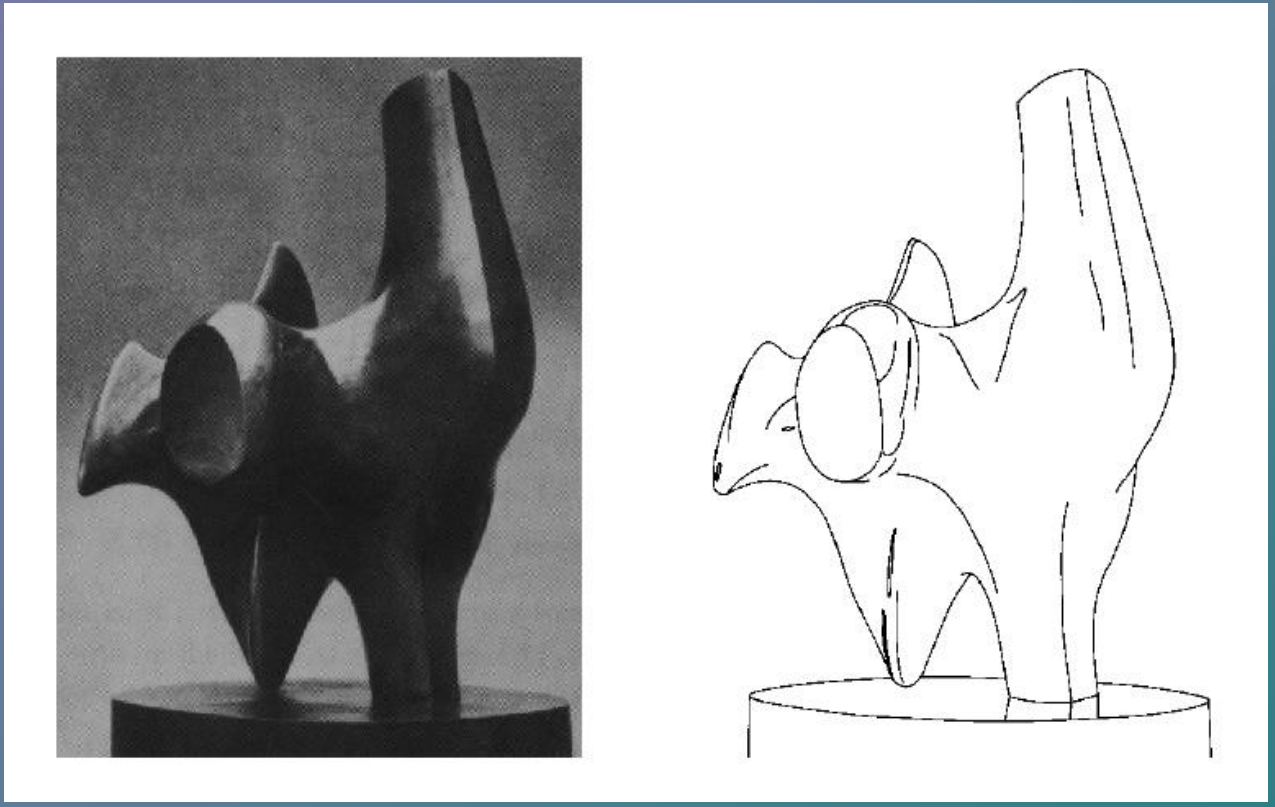
Pout
after histogram equalization

Histogram Equalization Example



It is commonly used for light normalization for image comparison or pattern recognition under different lighting conditions.

Edge detection



- Convert a 2D image into a set of curves
  - Extracts salient features of the scene
  - More compact than pixels

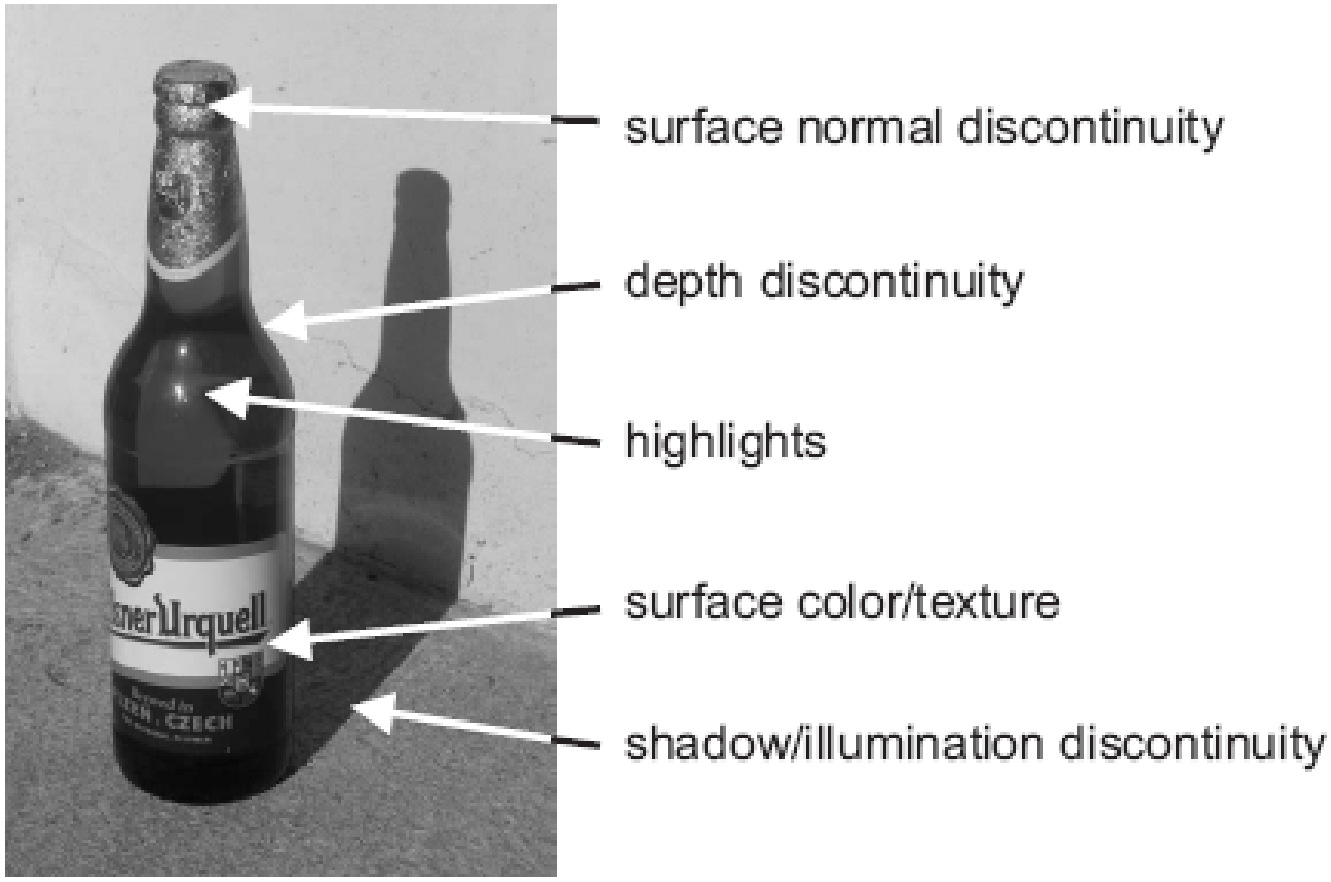


Figure 5.16: Origin of edges, i.e., physical phenomena in the image formation process which lead to edges in images.

Edge Types

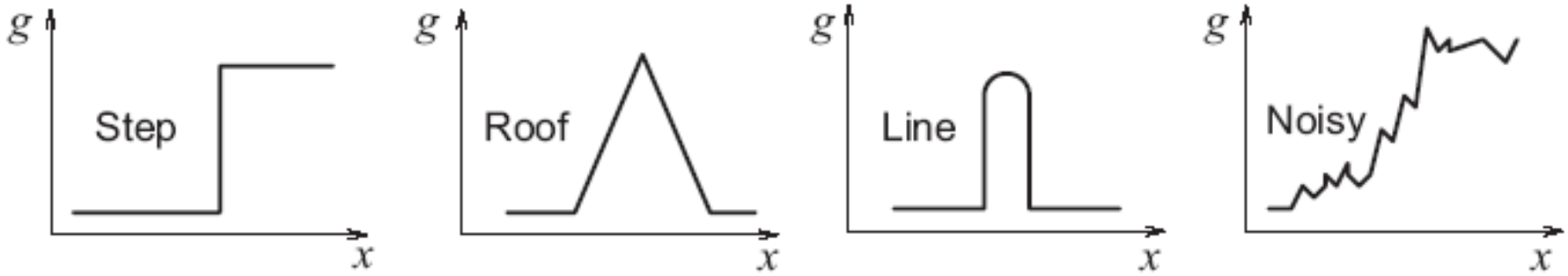


Figure 5.19: Typical edge profiles.

Image gradient

- The gradient of an image:

$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]$$

- The gradient points in the direction of most rapid change in intensity

- The gradient direction is given by:

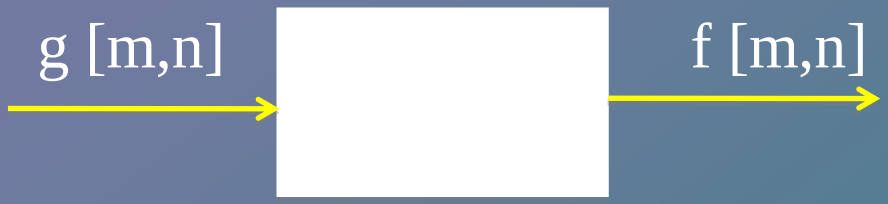
$$\theta = \tan^{-1}\left(\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x}\right)$$

  - how does this relate to the direction of the edge?

- The *edge strength* is given by the gradient magnitude

$$\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

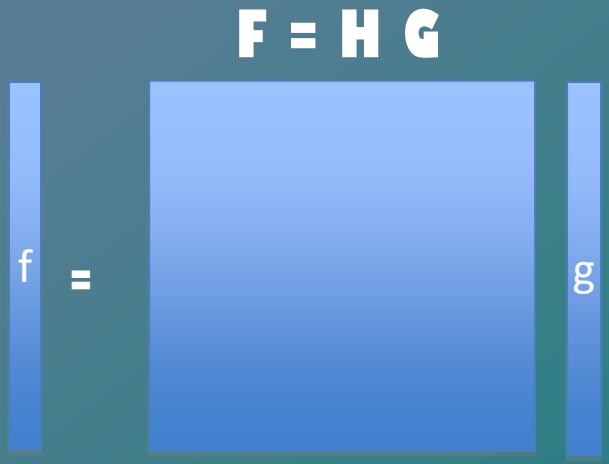
Linear filtering



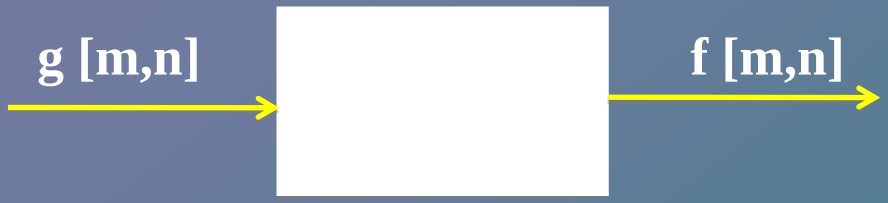
For a linear system, each output is a linear combination of all the input values:

$$f[m,n] = \sum_{k,l} h[m,n,k,l]\,g[k,l]$$

In matrix form:

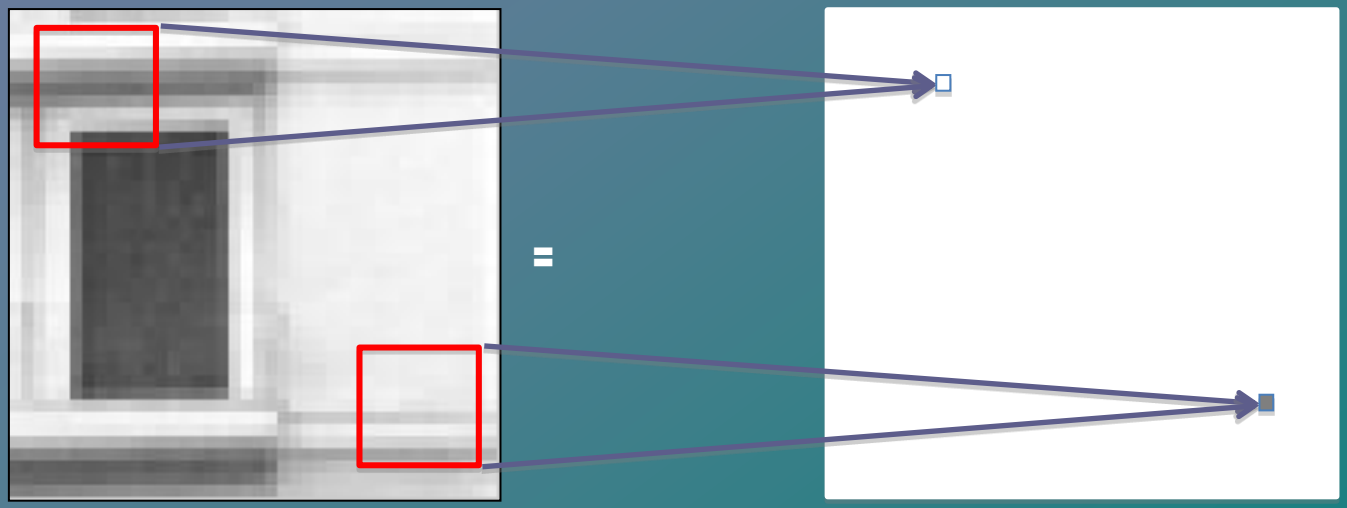


Linear filtering



In vision, many times, we are interested in operations that are spatially invariant. For a linear spatially invariant system:

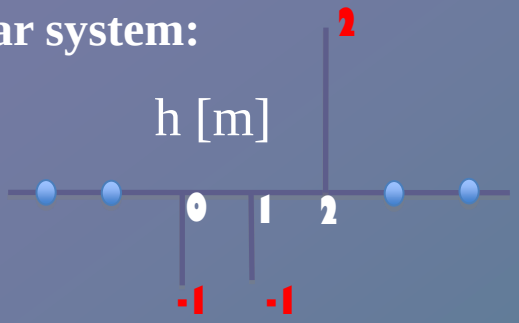
$$f[m,n] = h \otimes g = \sum_{k,l} h[m-k, n-l] g[k,l]$$



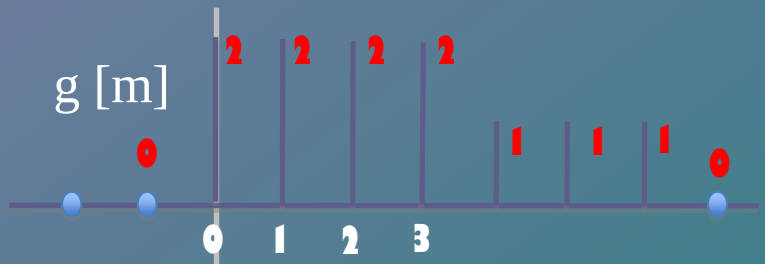
Linear filtering

$$f[m,n] = h \otimes g = \sum_{k,l} h[m-k, n-l] g[k,l]$$

Linear system:

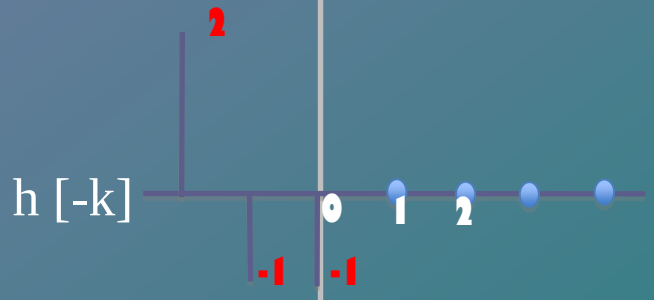


Input:



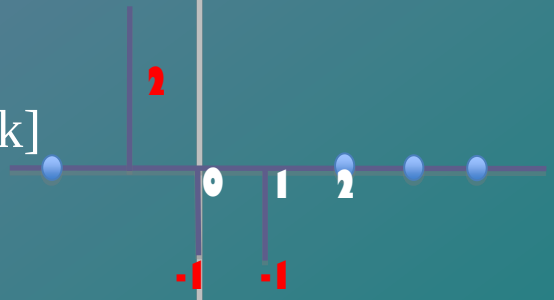
Output?

$$f[m=0] = \sum_k h[-k] g[k]$$



f [m=0]=-2

$$f[m=1] = \sum_k h[1-k] g[k]$$

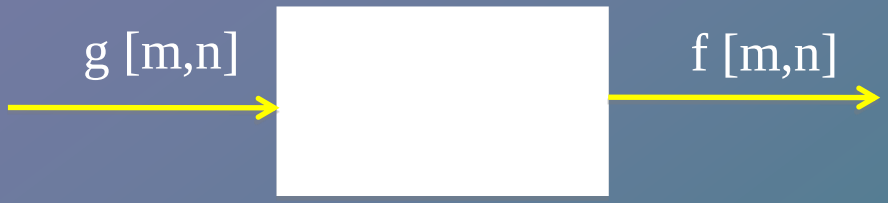


f [m=1]=-4

$$f[m=2] = \sum_k h[2-k] g[k]$$

f [m=2]=0

Linear filtering



For a linear spatially invariant system

$$f[m,n] = h \otimes g = \sum_{k,l} h[m-k, n-l]\, g[k,l]$$

m=0 1 2 …

111	115	113	111	112	111	112	111
135	138	137	139	145	146	149	147
163	168	188	196	206	202	206	207
180	184	206	219	202	200	195	193
189	193	214	216	104	79	83	77
191	201	217	220	103	59	60	68
195	205	216	222	113	68	69	83
199	203	223	228	108	68	71	77

g[m,n]

⊗

-1	2	-1
-1	2	-1
-1	2	-1

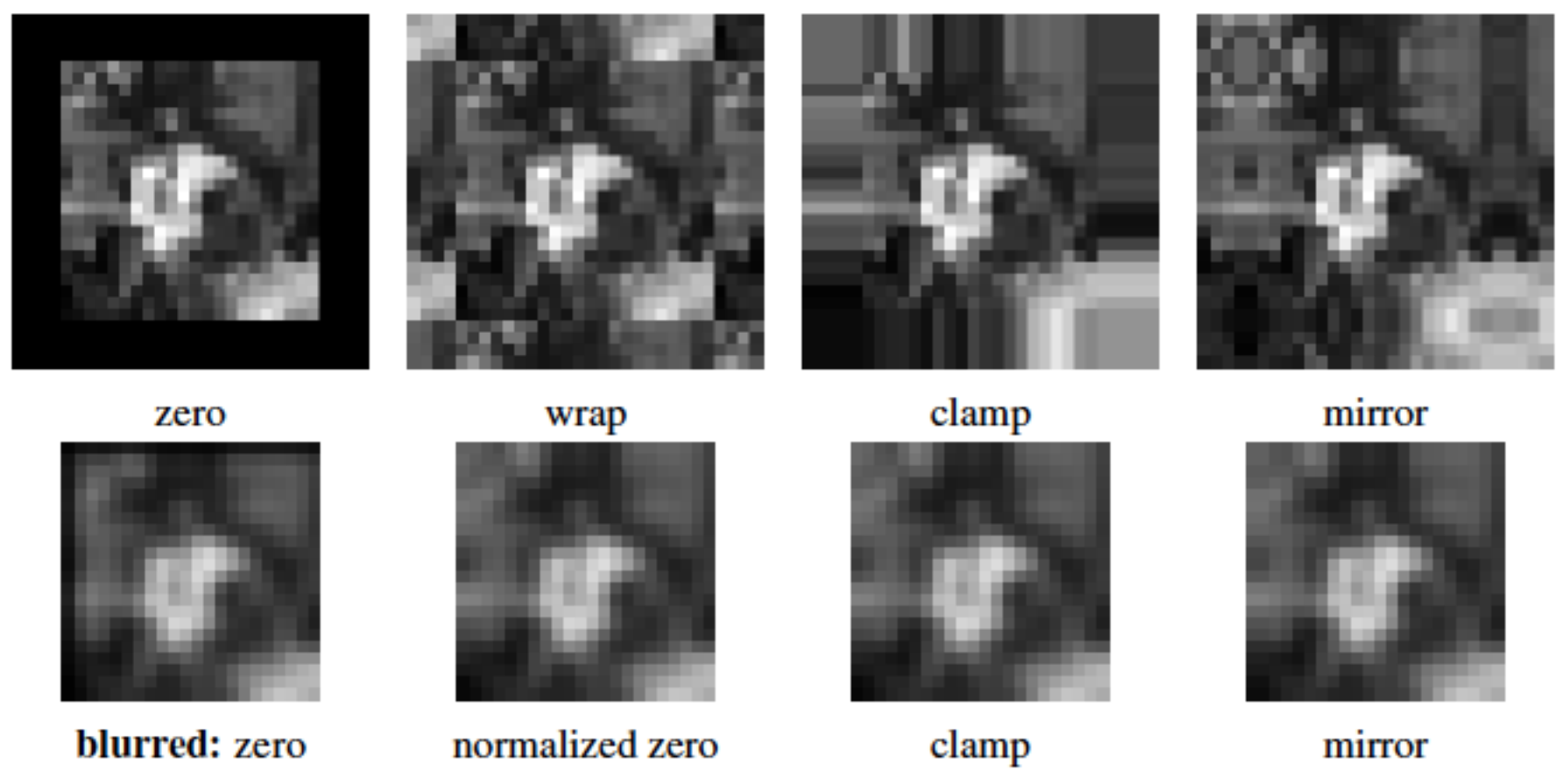
h[m,n]

=

?	?	?	?	?	?	?	?
?	-5	9	-9	21	-12	10	?
?	-29	18	24	4	-7	5	?
?	-50	40	142	-88	-34	10	?
?	-41	41		-175	-71	0	?
?	-24	37		-224	-120	-10	?
?	-23	33		-217	-134	-23	?
?	?	?	?	?	?	?	?

f[m,n]

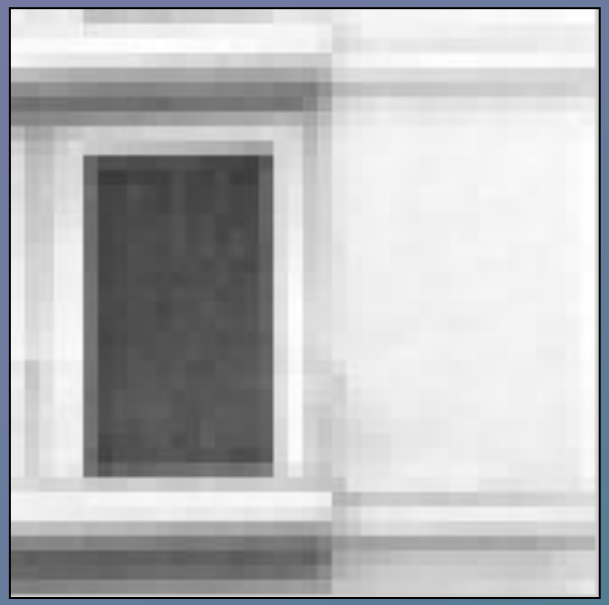
Borders



From Rick's book

Impulse

$$f[m,n] = h \otimes g = \sum_{k,l} h[m-k, n-l] g[k,l]$$



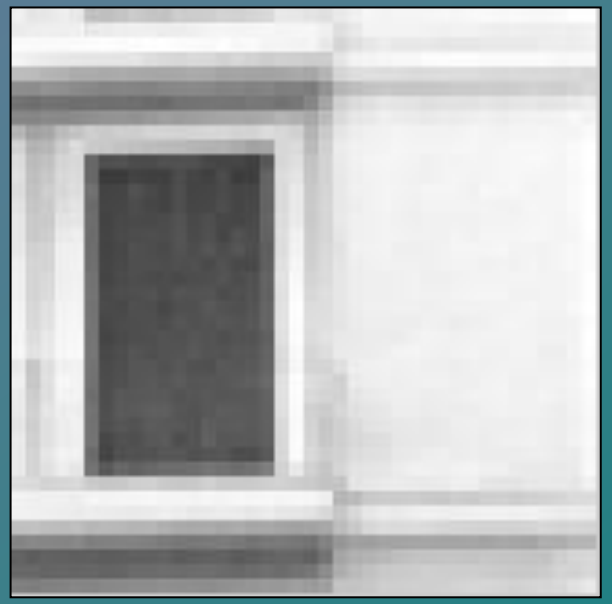
g[m,n]

$\otimes$

0	0	0	0	0
0	0	0	0	0
0	0	1	0	0
0	0	0	0	0
0	0	0	0	0

h[m,n]

=

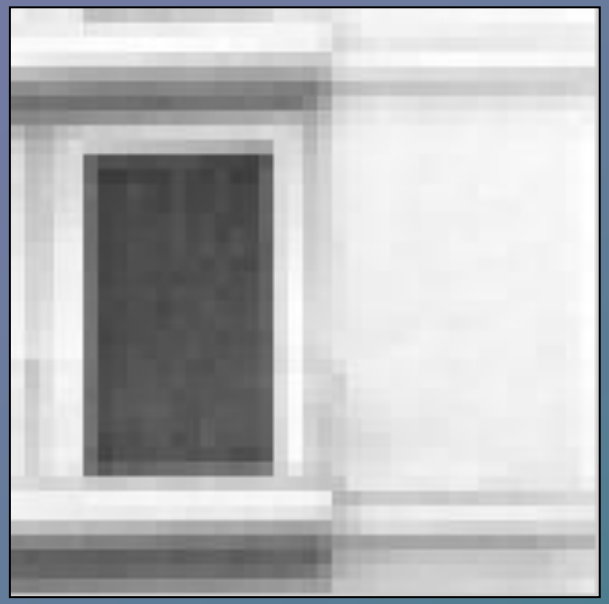


f[m,n]

Shift

$$f[m,n] = h \otimes g = \sum_{k,l} h[m-k, n-l] g[k,l]$$

2 pixels →



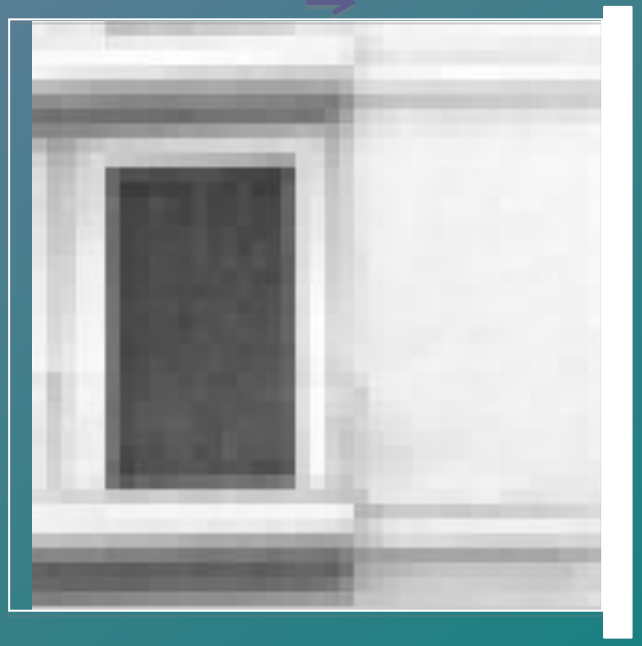
g[m,n]

$\otimes$

0	0	0	0	0
0	0	0	0	0
0	0	0	0	1
0	0	0	0	0
0	0	0	0	0

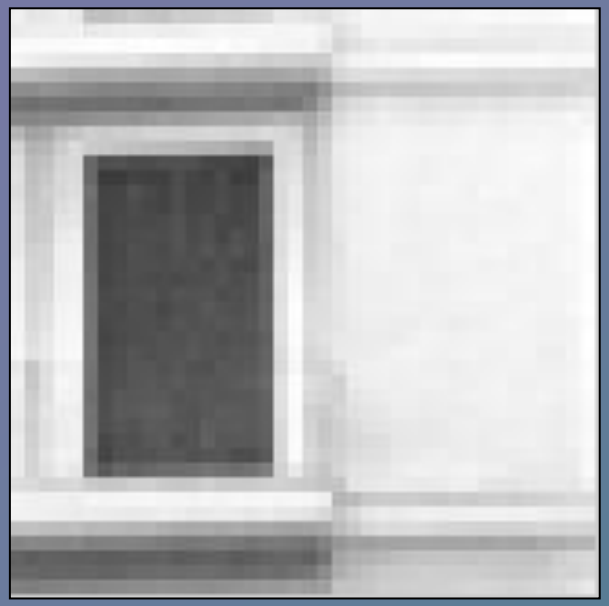
h[m,n]

=



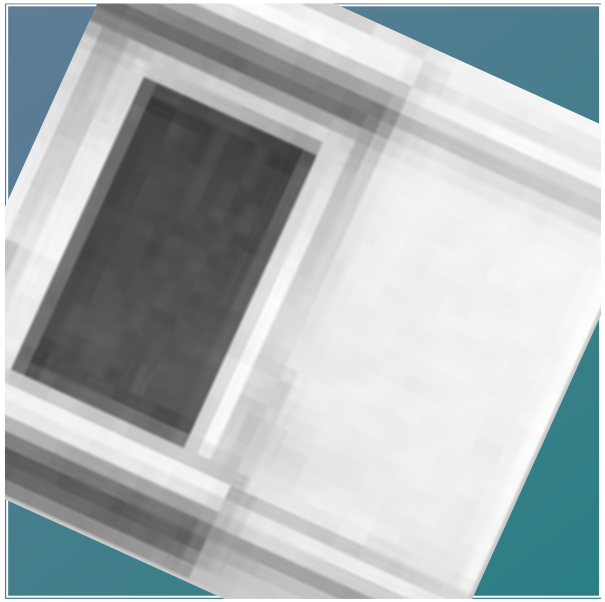
f[m,n]

Image rotation



$\otimes$? =

h[m,n]



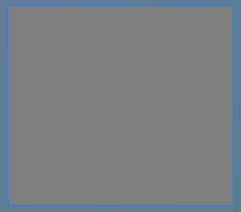
g[m,n]

It is linear, but not a spatially invariant operation. It cannot be done via convolution.

Rectangular filter

g[m,n]

⊗



h[m,n]

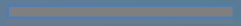
=

f[m,n]

Rectangular filter

g[m,n]

⊗



h[m,n]

=

f[m,n]

Rectangular filter

g[m,n]

⊗

h[m,n]

=

f[m,n]

Differentiation and Convolution

- Recall

$$\frac{\partial f}{\partial x} = \lim_{\varepsilon \to 0} \left(\frac{f(x+\varepsilon, y)}{\varepsilon} - \frac{f(x,y)}{\varepsilon} \right)$$

- We could approximate this as

$$\frac{\partial f}{\partial x} \approx \frac{f(x_{n+1}, y) - f(x_n, y)}{\Delta x}$$

- Now this is linear and shift invariant, so must be the result of a convolution.

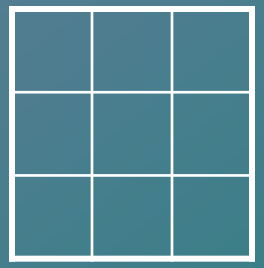
- (which is obviously a convolution; it's not a very good way to do things, as we shall see)

The discrete gradient

- How can we differentiate a *digital* image F(x,y)?
  - Option 1: reconstruct a continuous image, then take gradient
  - Option 2: take discrete derivative (finite difference)

$$\frac{\partial F}{\partial x}[x, y] \approx F[x+1, y] - F[x, y]$$

- How would you implement this as a cross-correlation?



H

The Sobel Operator

- Better approximations of the derivatives exist
  - The *Sobel* operators below are commonly used

$\frac{1}{8}$

-1	0	1
-2	0	2
-1	0	1

H_x

$\frac{1}{8}$

1	2	1
0	0	0
-1	-2	-1

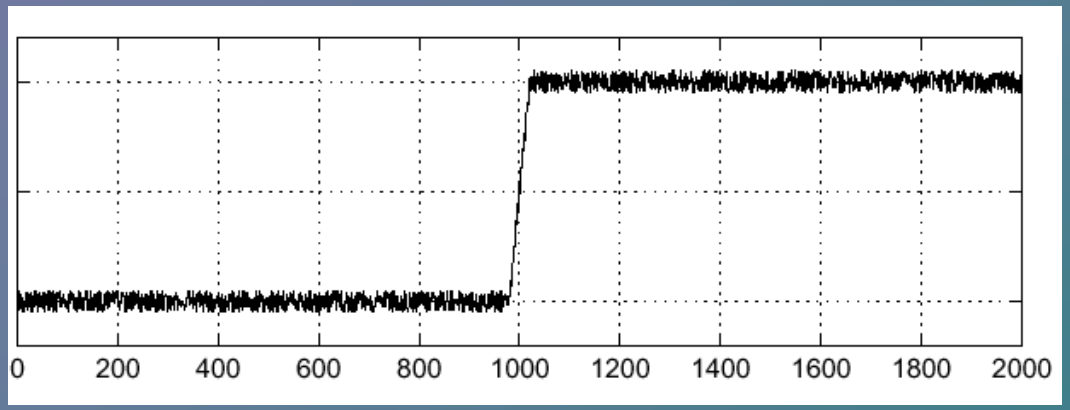
H_y

- The standard defn. of the Sobel operator omits the 1/8 term
  - doesn't make a difference for edge detection
  - the 1/8 term is needed to get the right gradient value, however

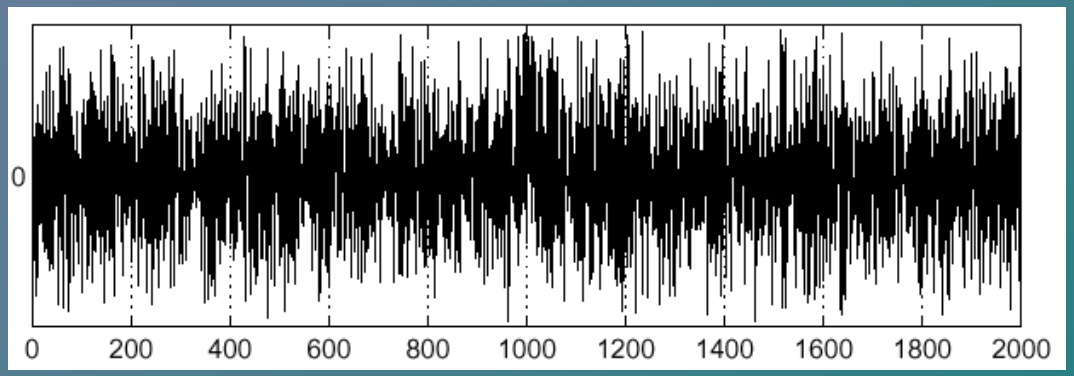
Effects of noise

- Consider a single row or column of the image
  - Plotting intensity as a function of position gives a *signal*

$f(x)$



$\frac{d}{dx}f(x)$



- Where is the edge?

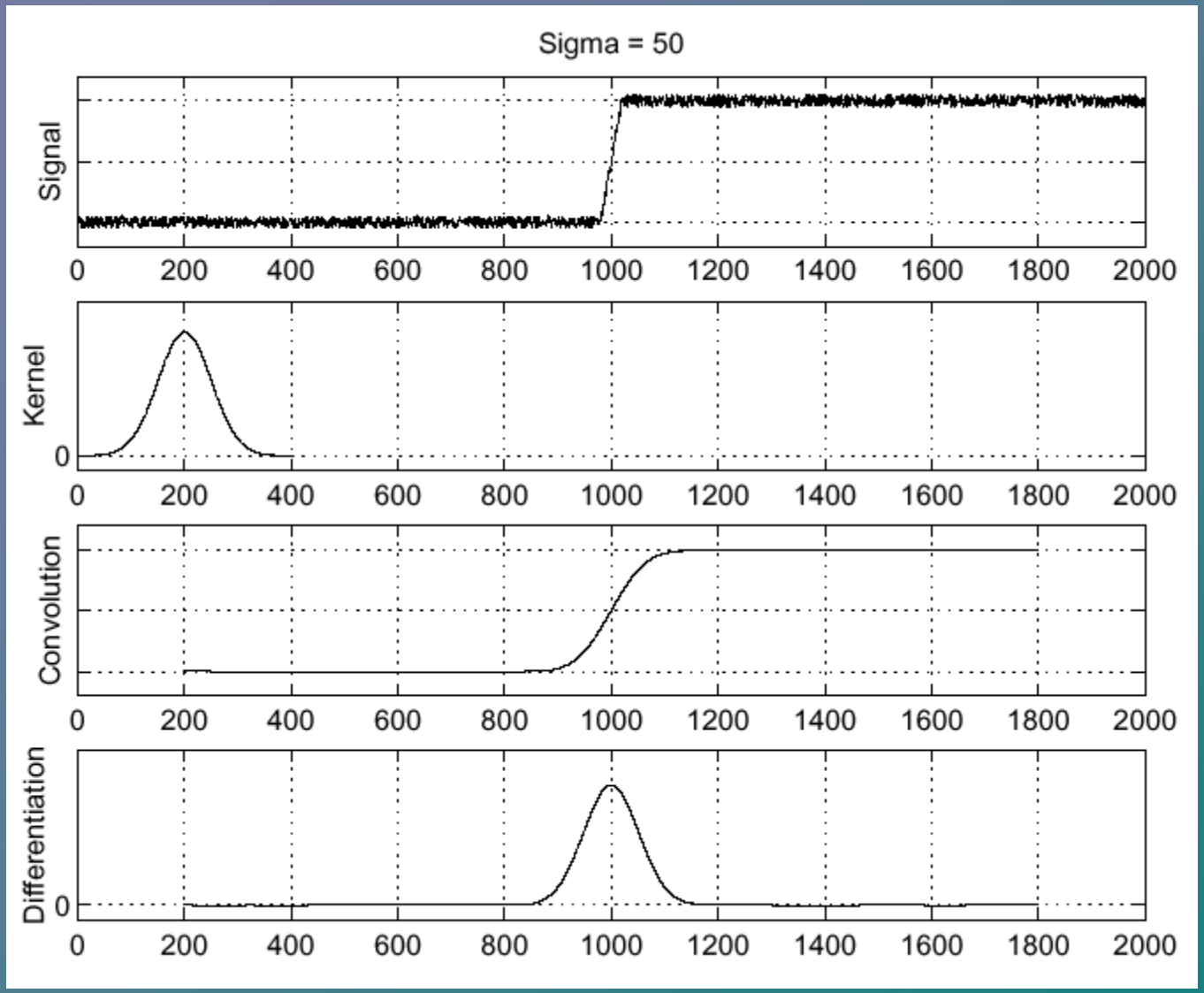
Solution: smooth first

f

h

$h \star f$

$\frac{\partial}{\partial x}(h \star f)$



- Where is the edge? Look for peaks in $\frac{\partial}{\partial x}(h \star f)$

Derivative theorem of convolution

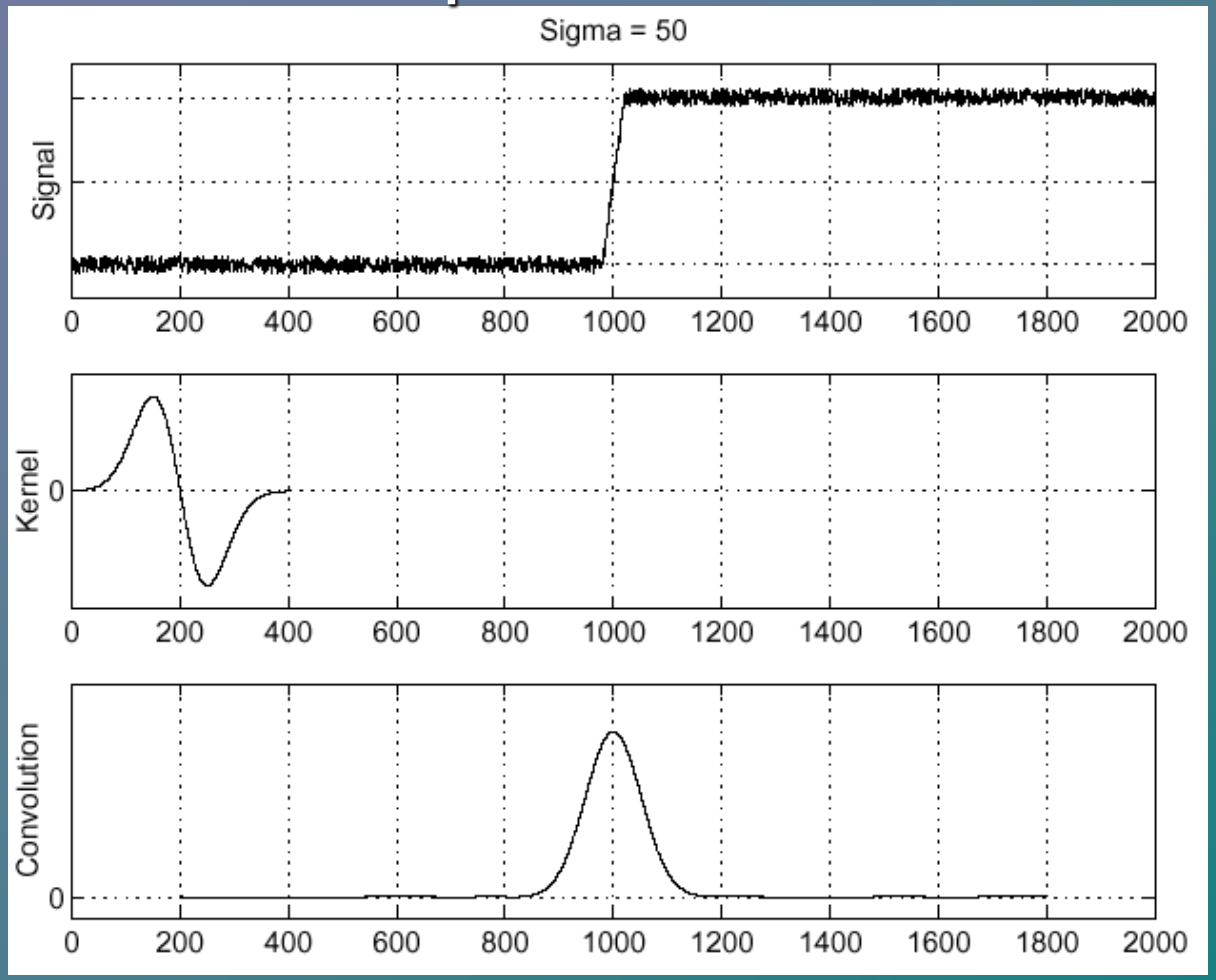
$$\frac{\partial}{\partial x}(h \star f) = (\frac{\partial}{\partial x}h) \star f$$

- This saves us one operation:

f

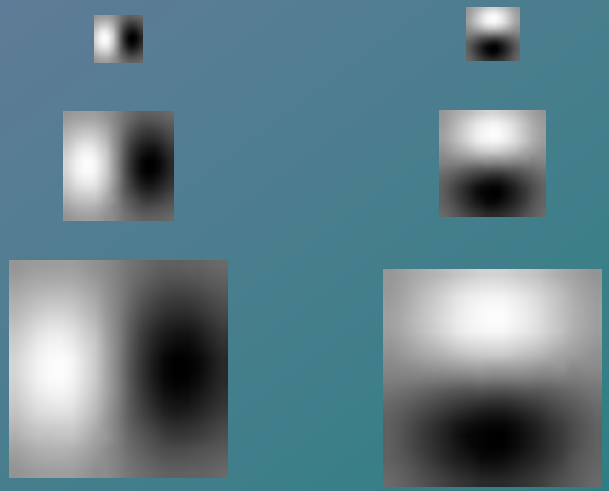
$\frac{\partial}{\partial x}h$

$(\frac{\partial}{\partial x}h) \star f$

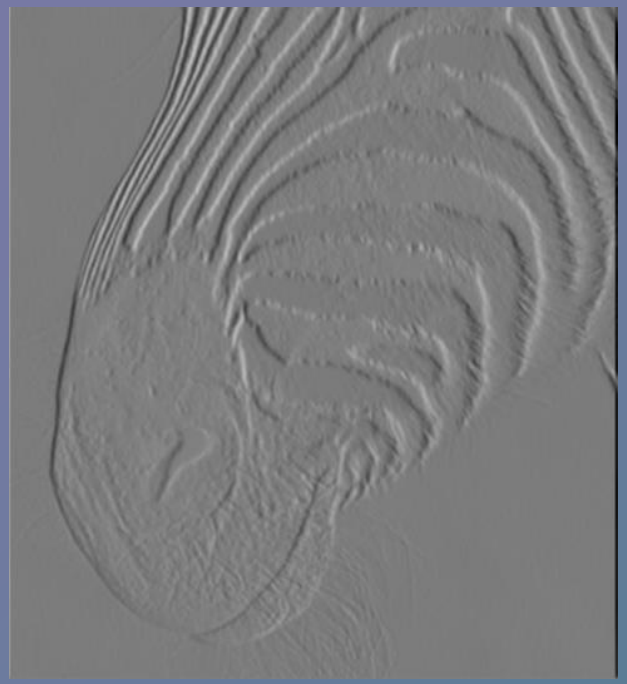


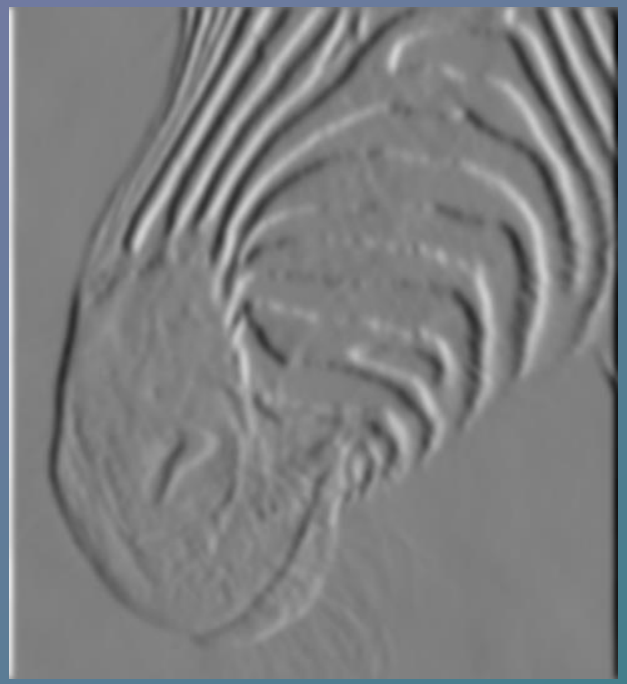
Smoothing and Differentiation

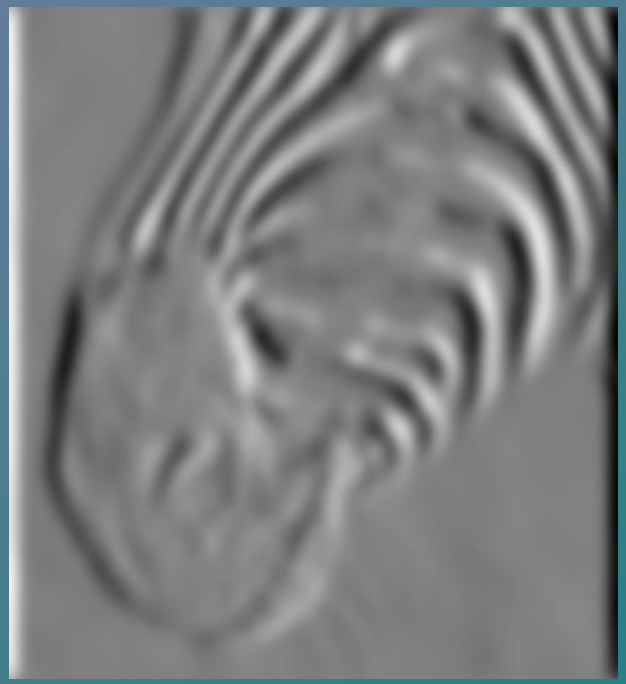
- Issue: noise
  - smooth before differentiation
  - two convolutions to smooth, then differentiate?
  - actually, no - we can use a derivative of Gaussian filter
    - because differentiation is convolution, and convolution is associative



Edge Detection at Different Scales







σ =1 pixel

σ = 3 pixels

σ = 7 pixels

The scale of the smoothing filter affects derivative estimates, and also the semantics of the edges recovered.

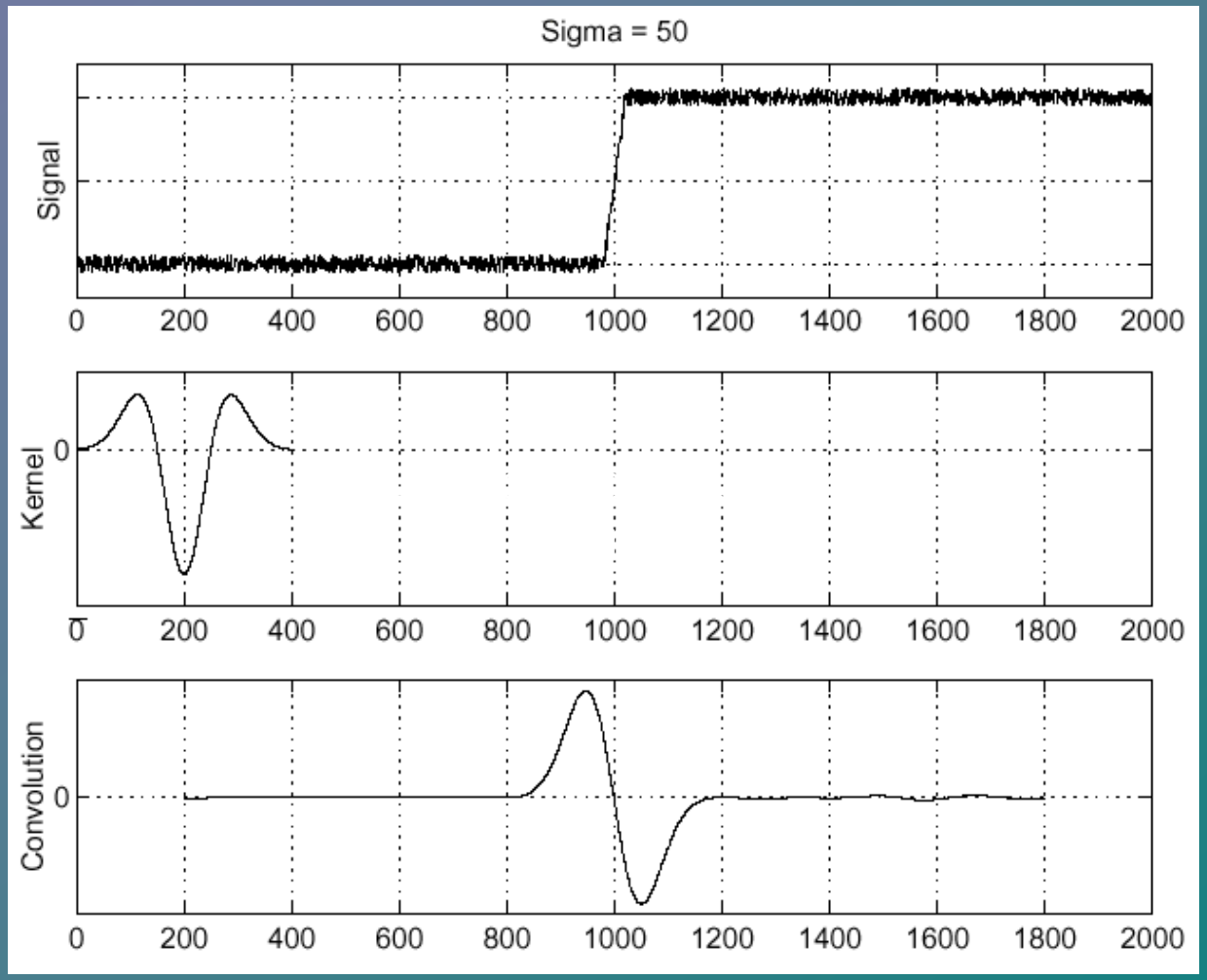
Laplacian of Gaussian

- Look for zero-crossings of $\frac{\partial^2}{\partial x^2}(h \star f)$

f

$\frac{\partial^2}{\partial x^2} h$

$(\frac{\partial^2}{\partial x^2} h) \star f$



1D Example in LoG Result

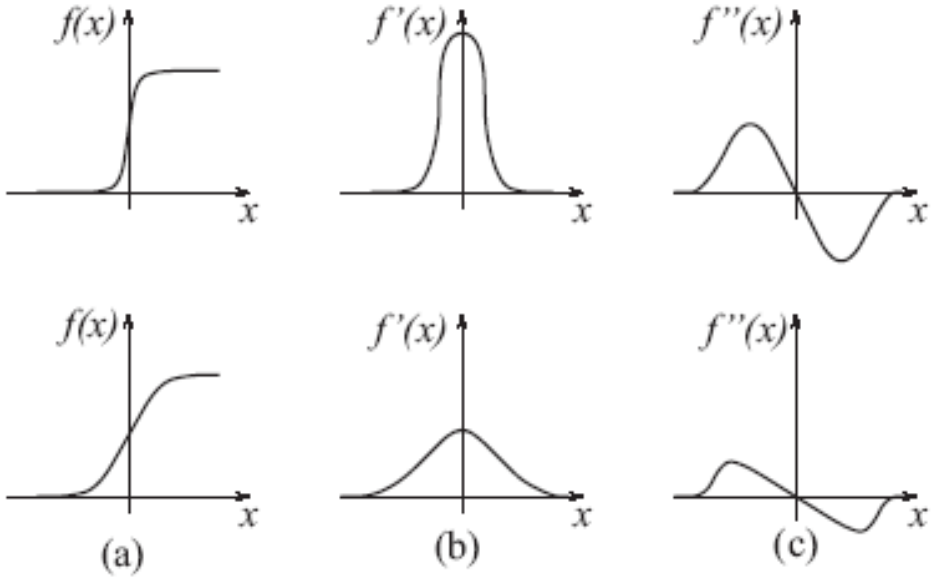
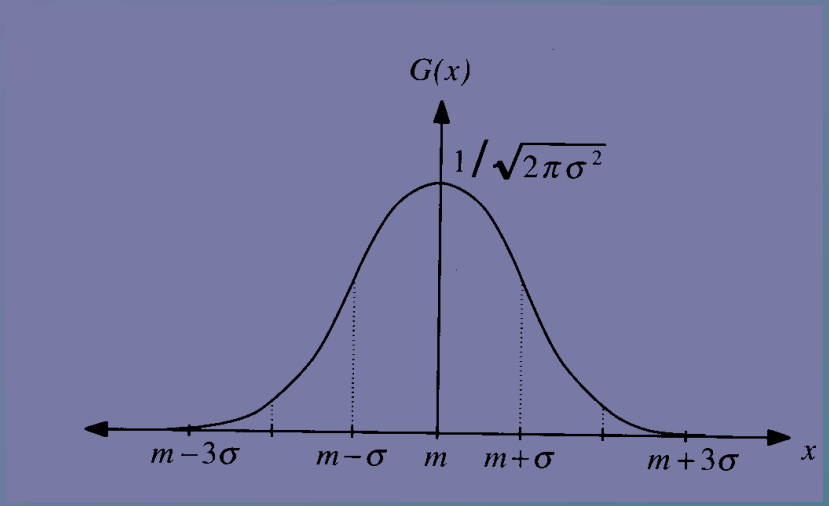


Figure 5.22: 1D edge profile of the zero-crossing.

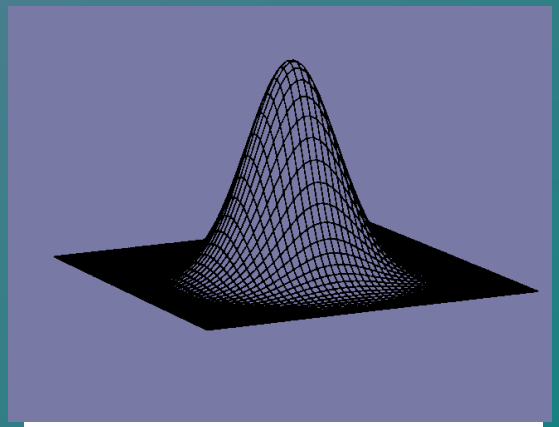
Gaussian Functions in 1-D and 2-D



$$G_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

Gaussians are Separable

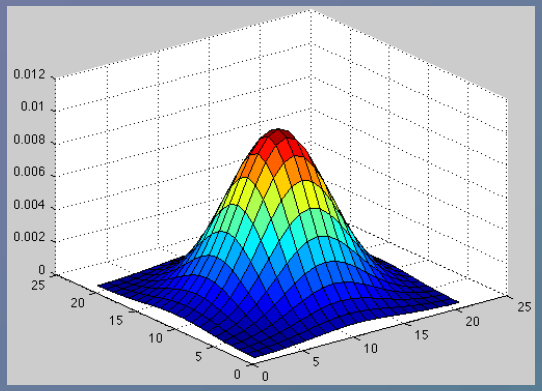
$$\begin{aligned} G_\sigma(x,y) &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x^2)}{2\sigma^2}\right)\right) \times \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^2)}{2\sigma^2}\right)\right) \end{aligned}$$



$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$$

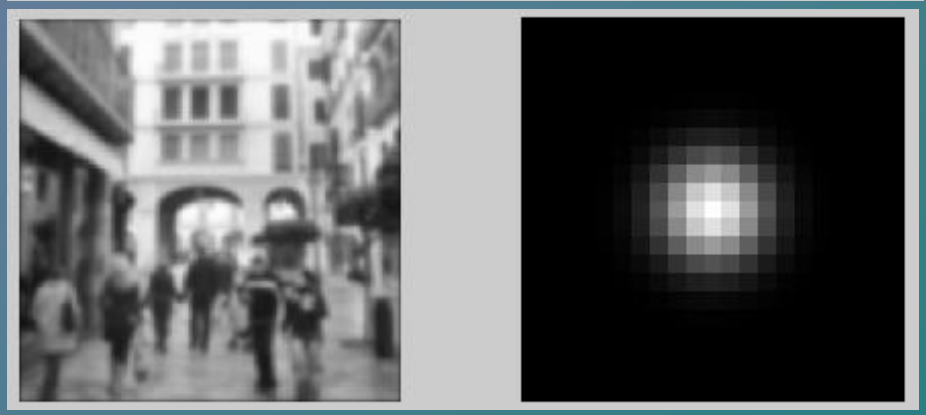
Gaussian filtering

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

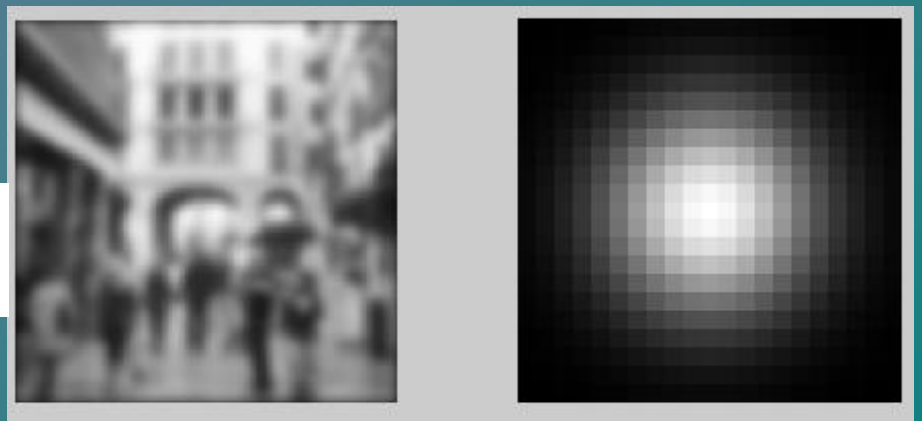


$$f[m,n] = I \otimes G = \sum_{k,l} I[m-k, n-l] G[k,l]$$

σ=1



σ=2



σ=4

SECOND ORDER EDGE OPERATOR

MARR-HILDRETH OPERATOR

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

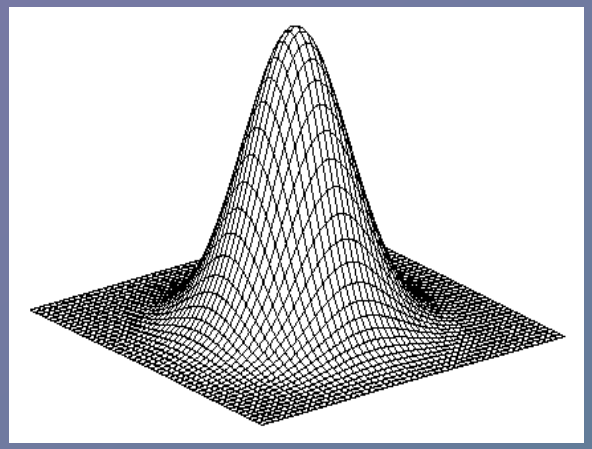
0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1

Combine this with Gaussian smoothing.

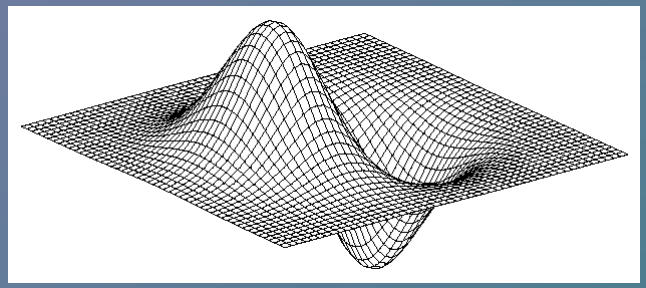
Detect edge at 'Zero-Crossing' looking for positive and negative peaks on either side.

2D edge detection filters



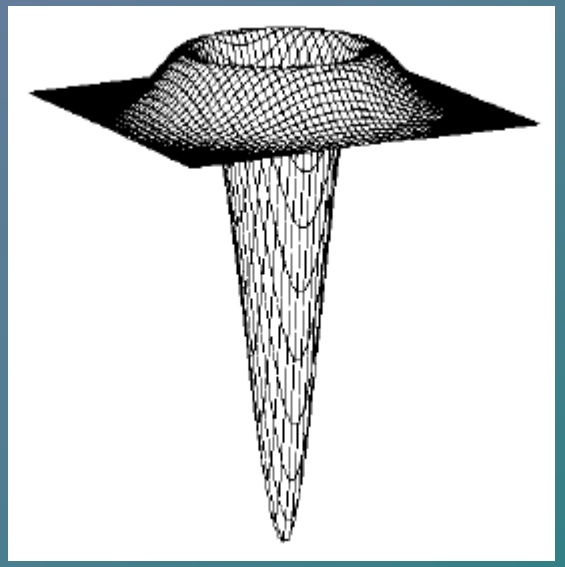
Gaussian

$$h_\sigma(u,v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}}$$



derivative of Gaussian

$$\frac{\partial}{\partial x} h_\sigma(u,v)$$

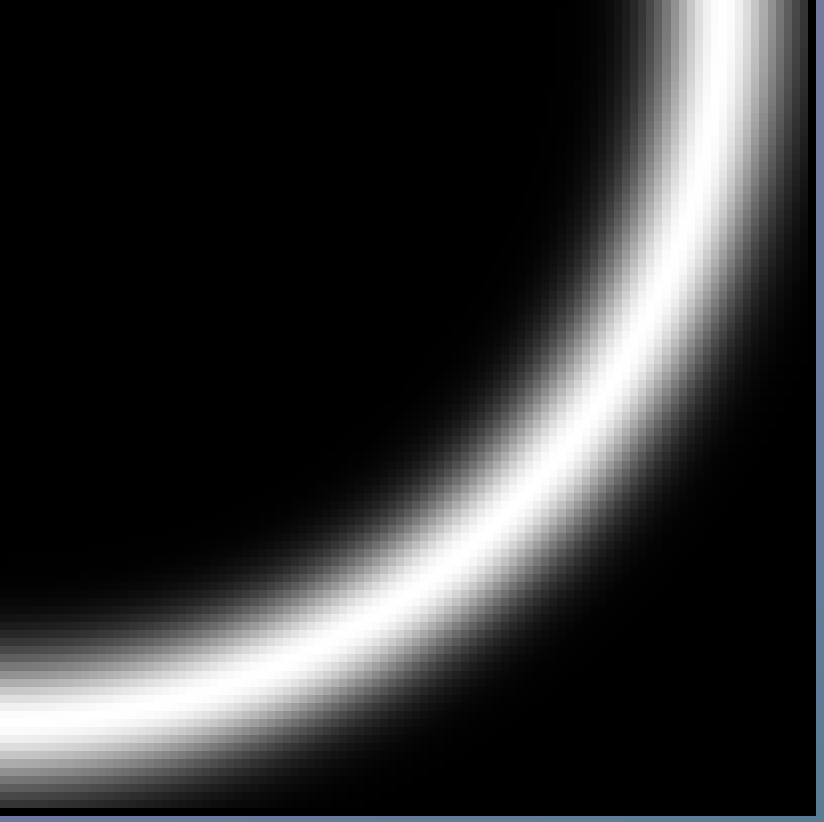


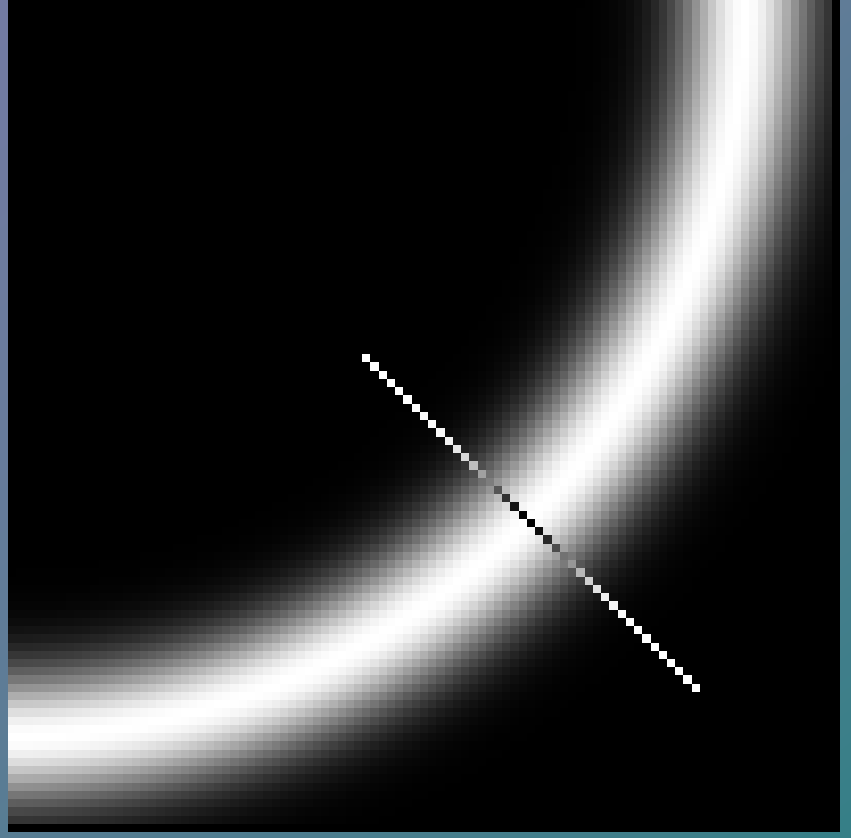
Laplacian of Gaussian

$$\nabla^2 h_\sigma(u,v)$$

- ∇^2 is the **Laplacian** operator:

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

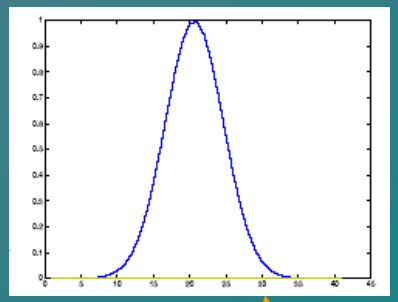


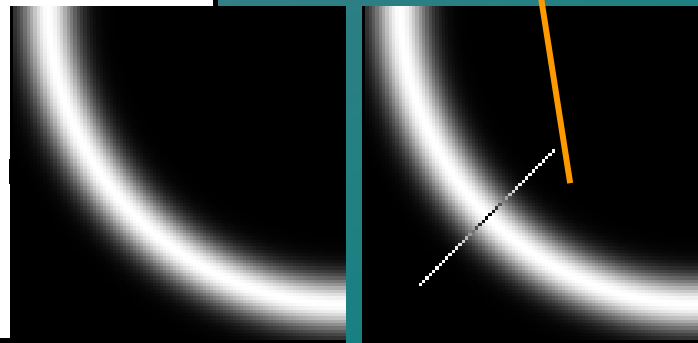


We wish to mark points along the curve where the magnitude is biggest. We can do this by looking for a maximum along a slice normal to the curve (non-maximum suppression). These points should form a curve. There are then two algorithmic issues: at which point is the maximum, and where is the next one?

Non-maximum suppression

At q, we have a maximum if the value is larger than those at both p and at r. Interpolate to get these values.



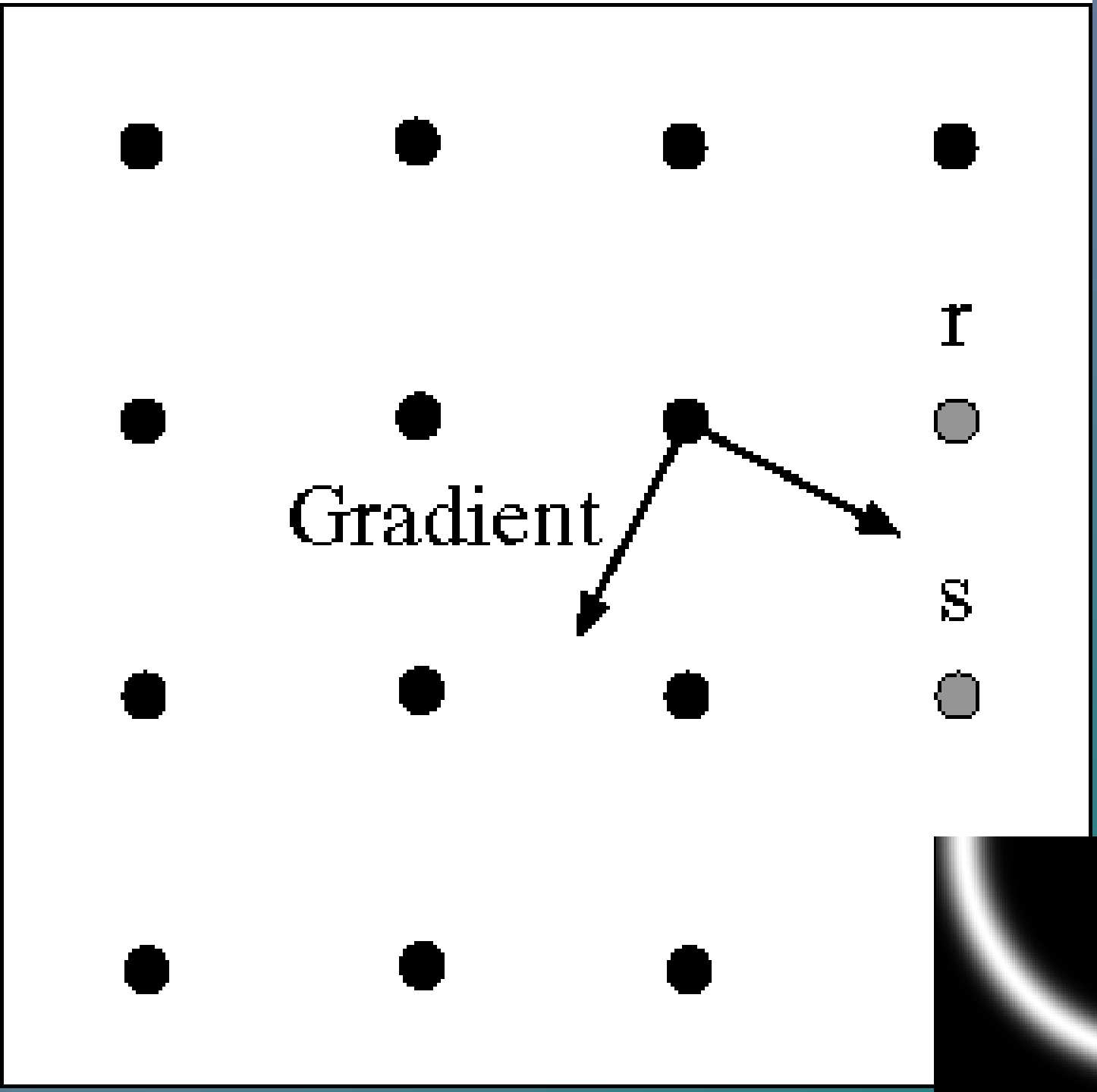


Non-Maximum Suppression

```
While there are points with high gradient
that have not been visited

  Find a start point that is a local maximum in the
    direction perpendicular to the gradient
    erasing points that have been checked

  while possible, expand a chain through
    the current point by:
      1) predicting a set of next points, using
        the direction perpendicular to the gradient

      2) finding which (if any) is a local maximum
        in the gradient direction

      3) testing if the gradient magnitude at the
        maximum is sufficiently large

      4) leaving a record that the point and
        neighbours have been visited

      record the next point, which becomes the current point

  end

end
```

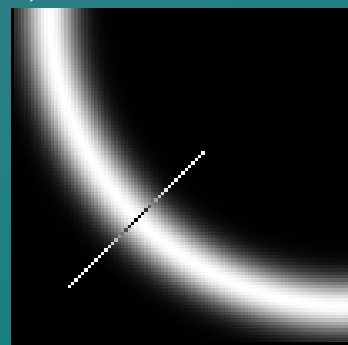
Simple Gradient-Based Edge Detection

```
form an estimate of the image gradient

obtain the gradient magnitude from this estimate

identify image points where the value
  of the gradient magnitude is maximal
  in the direction perpendicular to the edge
  and also large; these points are edge points
```



Predicting the next edge point

Assume the marked point is an edge point. Then we construct the tangent to the edge curve (which is normal to the gradient at that point) and use this to predict the next points (here either r or s).



Algorithm 5.4: Canny edge detector

1. Convolve an image f with a Gaussian of scale σ.
2. Estimate local edge normal directions $\mathbf{n}$ using equation (5.58) for each pixel in the image.
3. Find the location of the edges using equation (5.60) (non-maximal suppression).
4. Compute the magnitude of the edge using equation (5.61).
5. Threshold edges in the image with hysteresis (Algorithm 6.5) to eliminate spurious responses.
6. Repeat steps (1) through (5) for ascending values of the standard deviation σ.
7. Aggregate the final information about edges at multiple scale using the 'feature synthesis' approach.

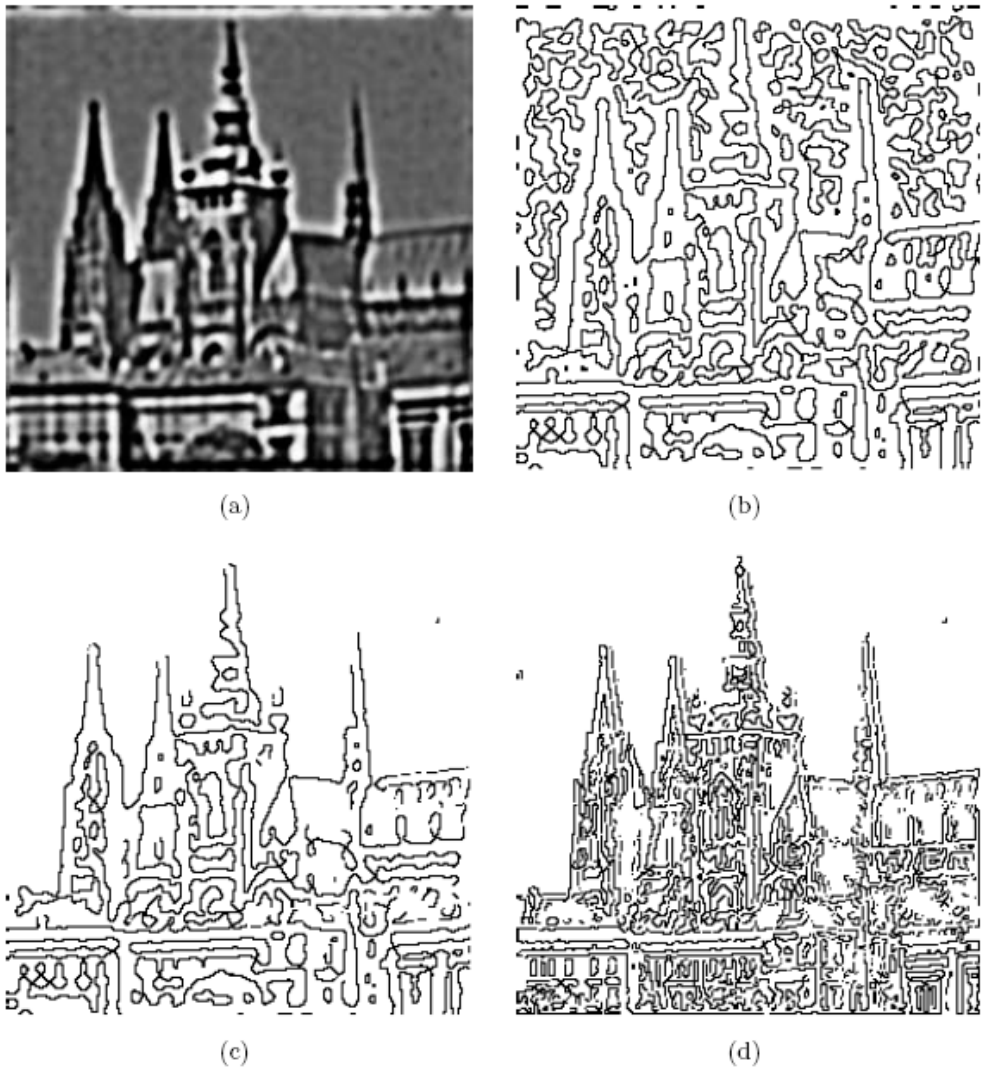


Figure 5.23: Zero-crossings of the second derivative, see Figure 5.10a for the original image. (a) DoG image ($\sigma_1 = 0.10, \sigma_2 = 0.09$), dark pixels correspond to negative DoG values, bright pixels represent positive DoG values. (b) Zero-crossings of the DoG image. (c) DoG zero-crossing edges after removing edges lacking first-derivative support. (d) LoG zero-crossing edges ($\sigma = 0.20$) after removing edges lacking first-derivative support—note different scale of edges due to different Gaussian smoothing parameters.

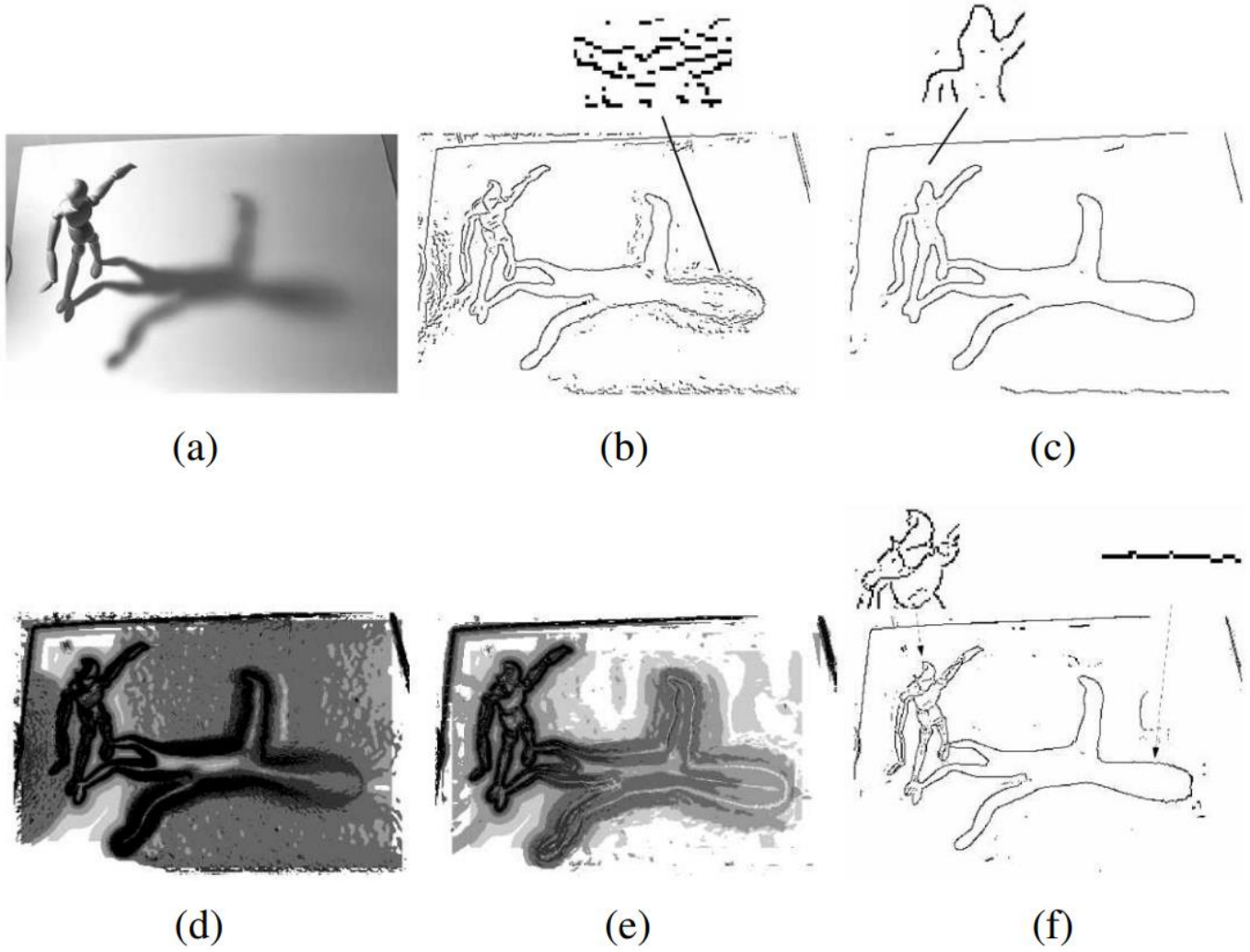


Figure 7.33 *Scale selection for edge detection (Elder and Zucker 1998) © 1998 IEEE: (a) original image; (b–c) Canny/Deriche edge detector tuned to the finer (mannequin) and coarser (shadow) scales; (d) minimum reliable scale for gradient estimation; (e) minimum reliable scale for second derivative estimation; (f) final detected edges.*

Human Boundary Detection

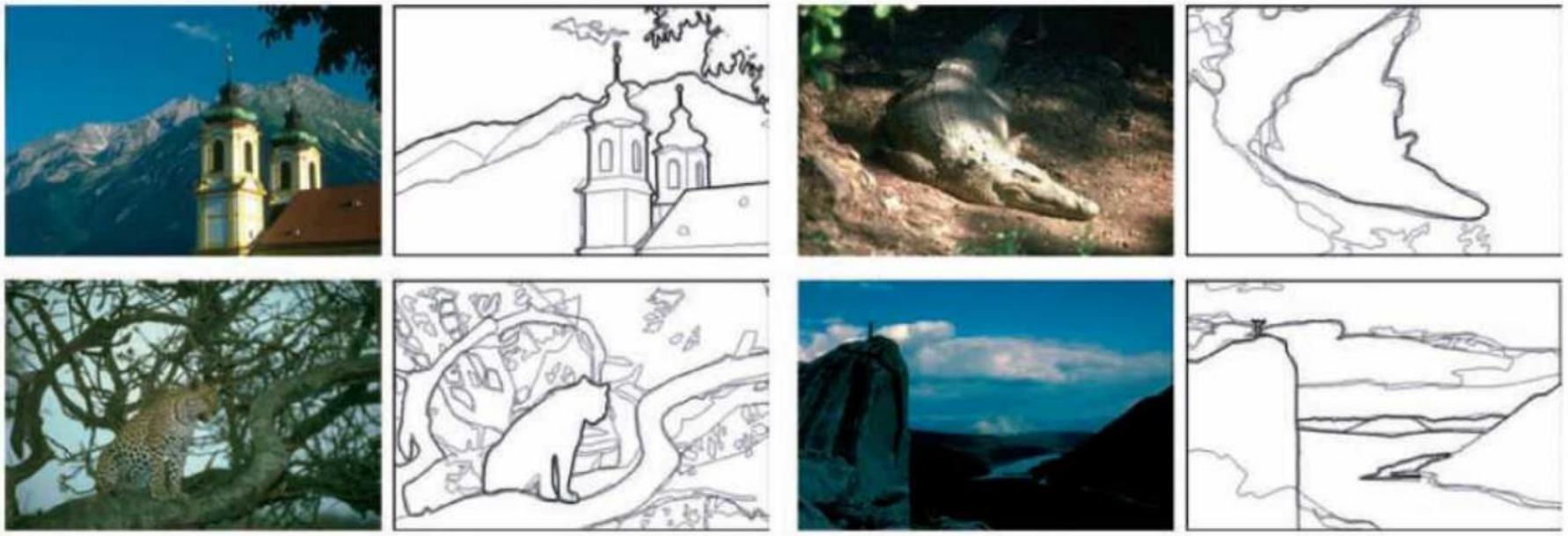


Figure 7.32 *Human boundary detection (Martin, Fowlkes, and Malik 2004) © 2004 IEEE. The darkness of the edges corresponds to how many human subjects marked an object boundary at that location.*

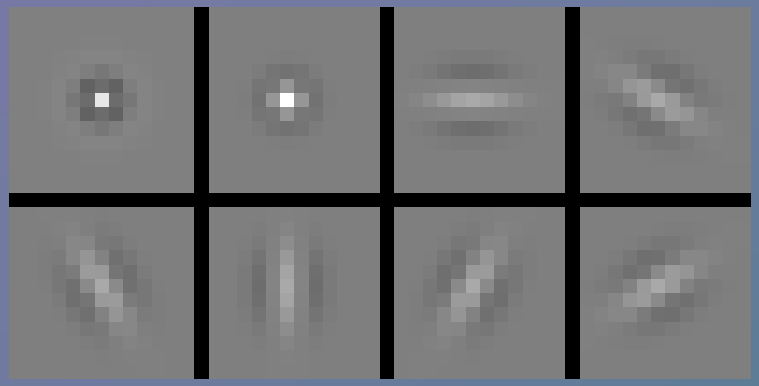
Combining Edge Feature Cues

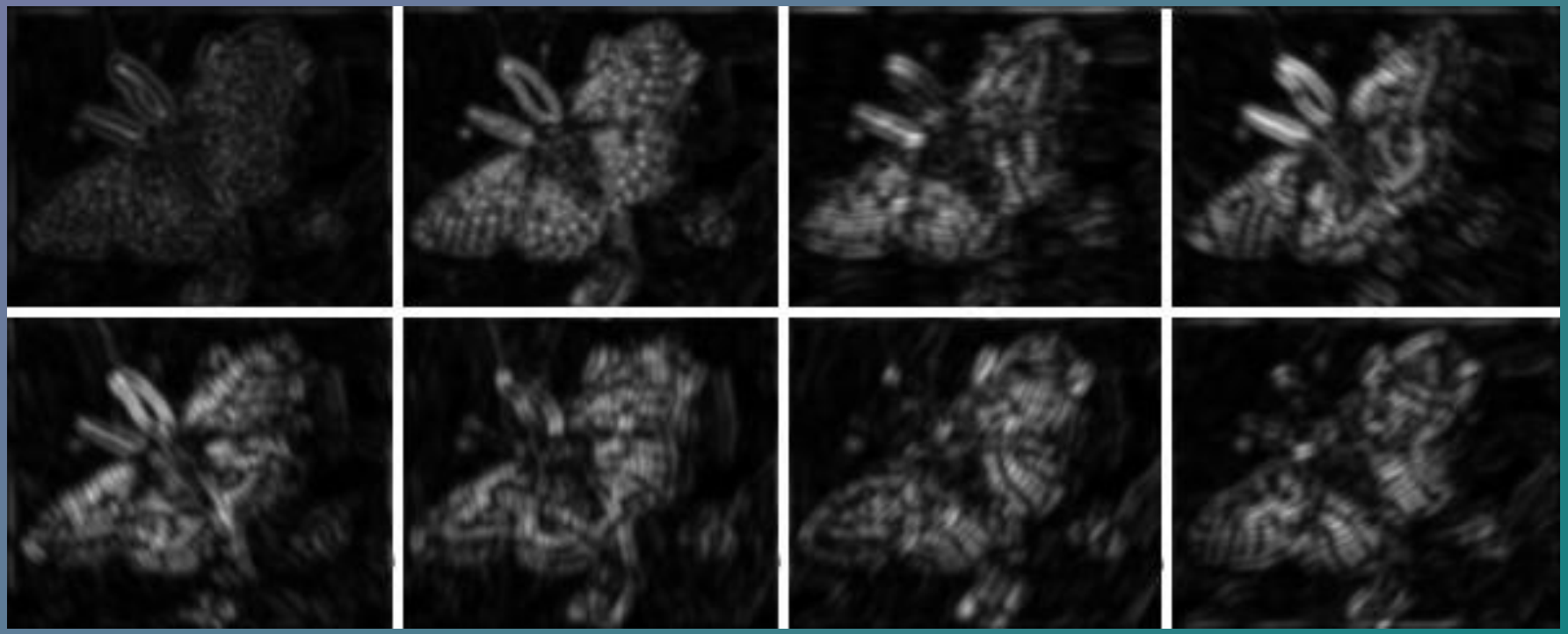
brightness gradient (BG), color gradient (CG), texture gradient (TG), and combined (BG+CG+TG) detectors

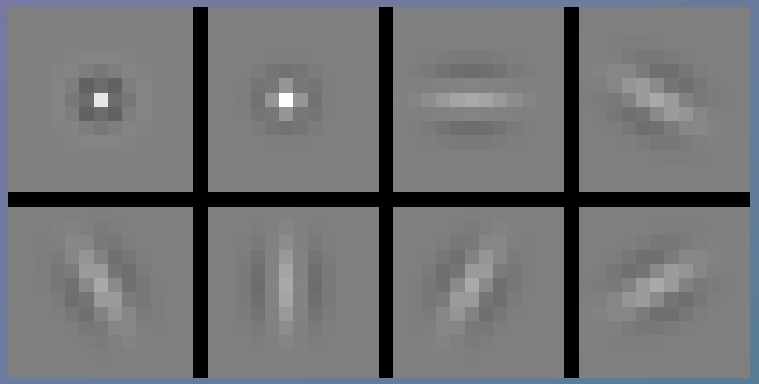
Martin, D., Fowlkes, C., and Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(5):530–549.

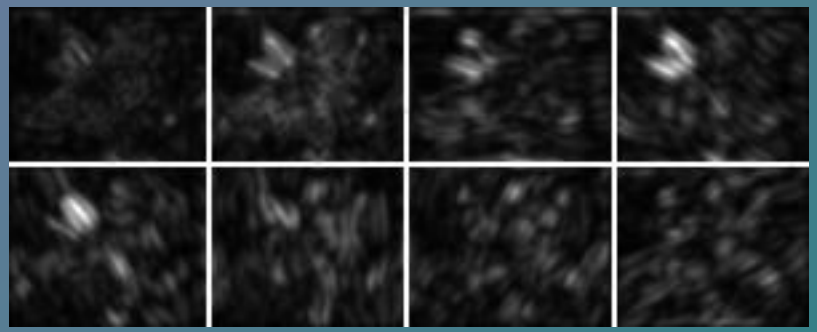
Representing textures

- Textures are made up of quite stylised subelements, repeated in meaningful ways
- Representation:
  - find the subelements, and represent their statistics
- But what are the subelements, and how do we find them?
  - recall normalized correlation
  - find subelements by applying filters, looking at the magnitude of the response
- What filters?
  - experience suggests spots and oriented bars at a variety of different scales
  - details probably don't matter
- What statistics?
  - within reason, the more the merrier.
  - At least, mean and standard deviation
  - better, various conditional histograms.









The Laplacian Pyramid

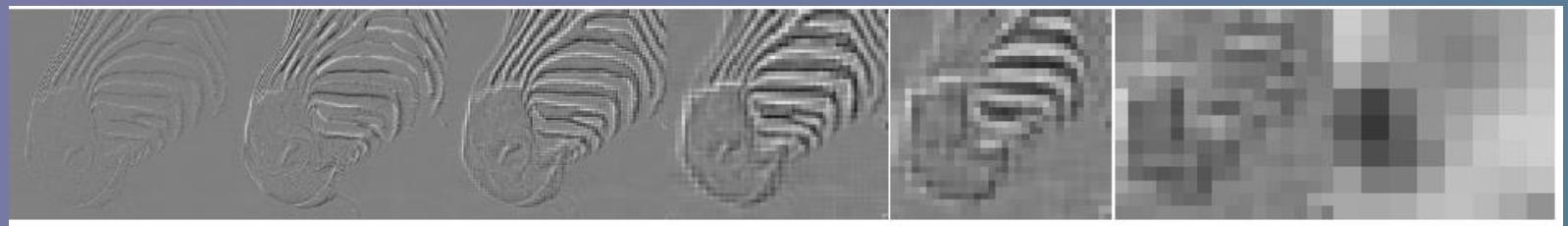
- Synthesis
  - preserve difference between upsampled Gaussian pyramid level and Gaussian pyramid level
  - band pass filter - each level represents spatial frequencies (largely) unrepresented at other levels
- Analysis
  - reconstruct Gaussian pyramid, take top layer

$$P_{\text{Laplacian}}(\mathcal{I})_m = P_{\text{Gaussian}}(\mathcal{I})_m$$

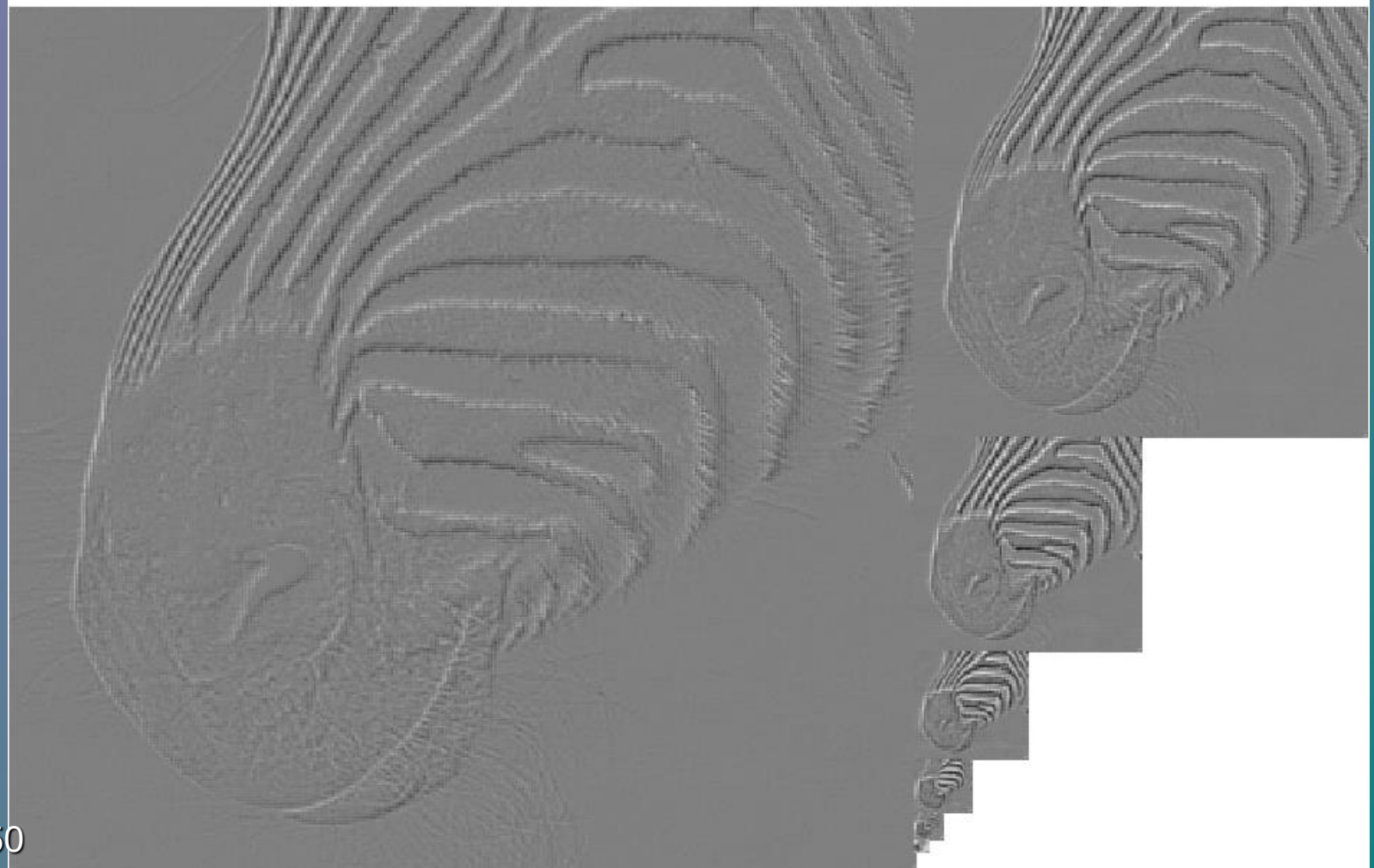
(where m is the coarsest level) and

$$\begin{aligned} P_{\text{Laplacian}}(\mathcal{I})_k &= P_{\text{Gaussian}}(\mathcal{I})_k - S^{\uparrow}(P_{\text{Gaussian}}(\mathcal{I})_{k+1}) \\ &= (Id - S^{\uparrow}S^{\downarrow}G_{\sigma})P_{\text{Gaussian}}(\mathcal{I})_k \end{aligned}$$

512 256 128 64 32 16 8



512 256 128 64 32 16 8



Building a Laplacian Pyramid

```
Form a Gaussian pyramid

Set the coarsest layer of the Laplacian pyramid to be
  the coarsest layer of the Gaussian pyramid

For each layer, going from next to coarsest to finest

  Obtain this layer of the Laplacian pyramid by
    upsampling the next coarser layer, and subtracting
    it from this layer of the Gaussian pyramid

end
```

Synthesis: Obtaining an Image from Its Laplacian Pyramid

```
Set the working image to be the coarsest layer

For each layer, going from next to coarsest to finest

  Upsample the working image and add the current layer
    to the result

  Set the working image to be the result of this operation

end
The working image now contains the original image
```

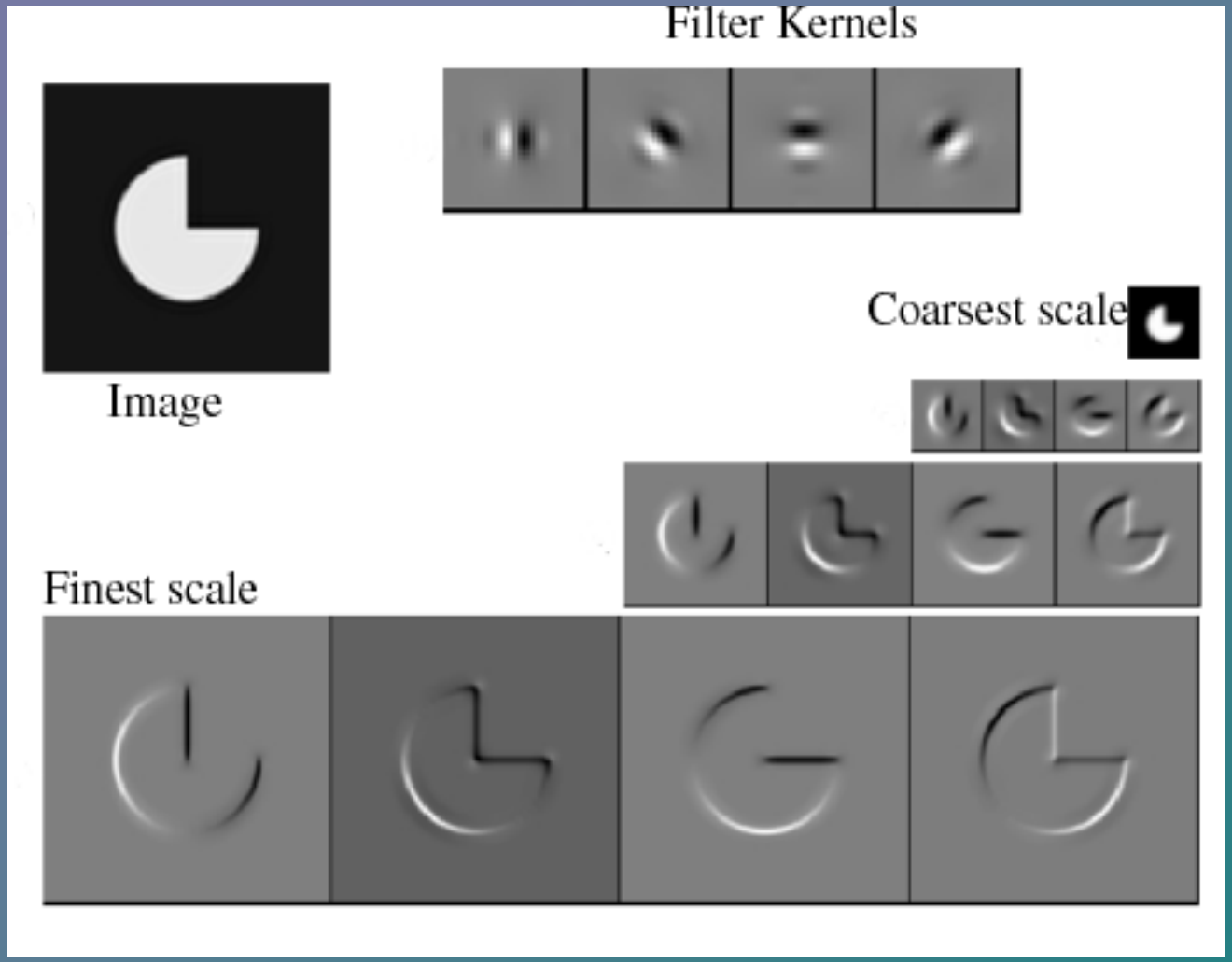
Local Spatial Frequency Analysis and Gabor Filters

- Gabor filter responds strongly at points in an image where there are components that locally have a particular spatial frequency and orientation.
- Gabor filters come in pairs, often referred to as quadrature pairs; one of the pair recovers symmetric components in a particular direction, and the other recovers antisymmetric components.
- The mathematical form of the symmetric kernel is

$$G_{symmetric}(x, y) = \cos\left(k_x x + k_y y\right) \exp - \left\{ \frac{x^2 + y^2}{2\sigma^2} \right\}$$

- The antisymmetric kernel has the form

$$G_{antisymmetric}(x, y) = \sin\left(k_0 x + k_1 y\right) \exp - \left\{ \frac{x^2 + y^2}{2\sigma^2} \right\}$$



Reprinted from “Shiftable MultiScale Transforms,” by Simoncelli et al., IEEE Transactions on Information Theory, 1992, copyright 1992, IEEE

Corner (Interest Point) Detection

- Correspondence problem
- Aperture problem

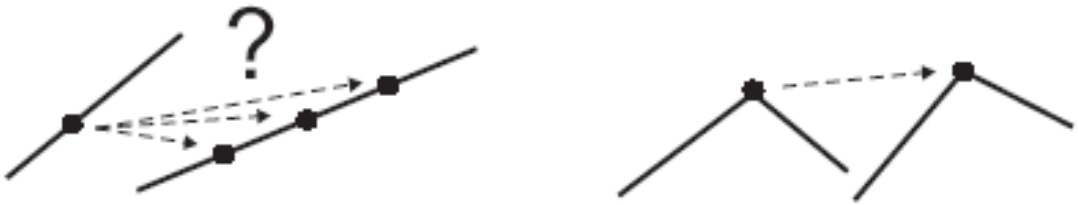


Figure 5.34: Ambiguity of lines for matching and unambiguity of corners.

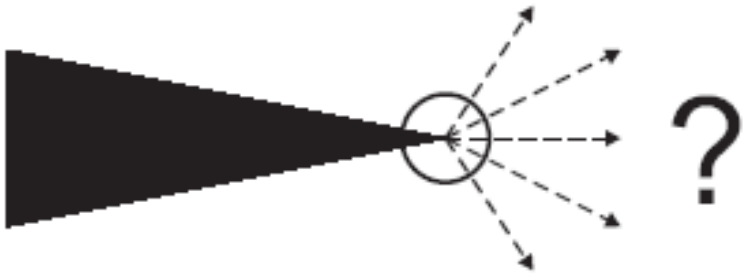
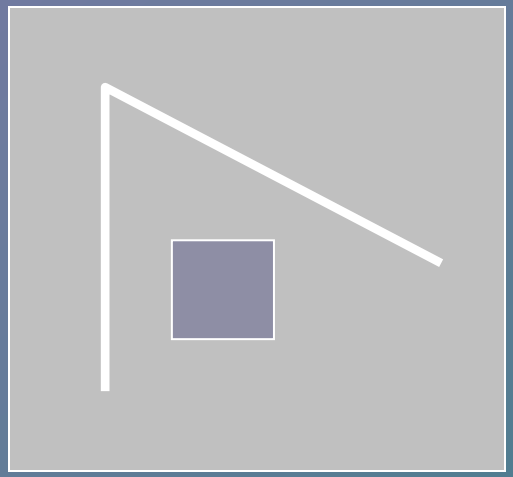


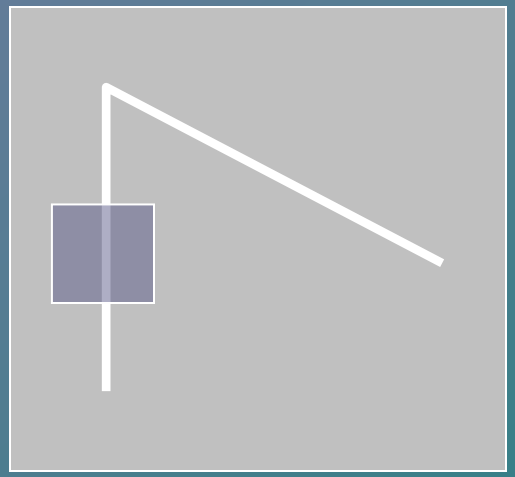
Figure 5.35: Ambiguity of edge detector at the corner tip.

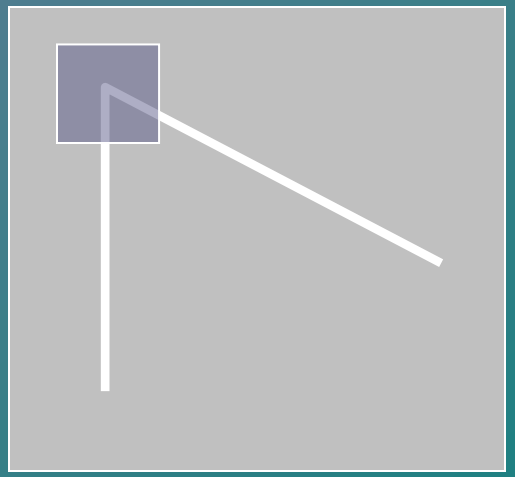
Local measures of uniqueness

Suppose we only consider a small window of pixels

- What defines whether a feature is a good or bad candidate?





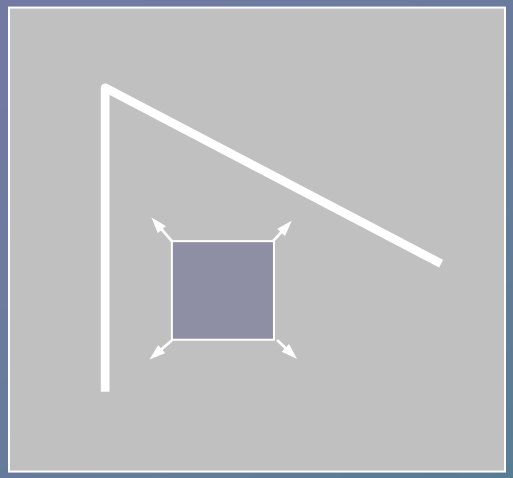


Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

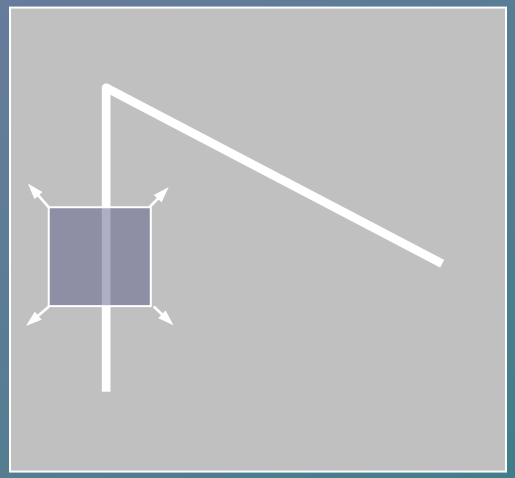
Corner detection

Local measure of feature uniqueness

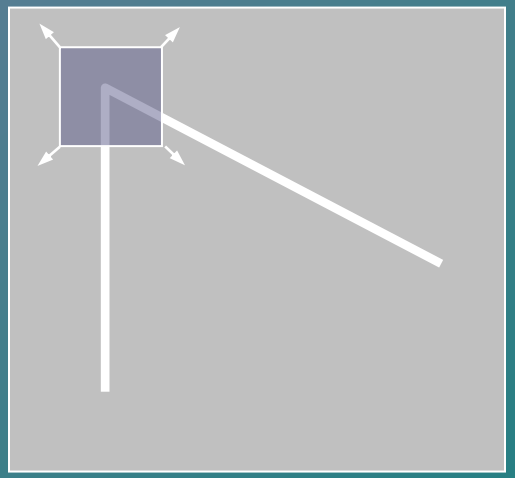
- How does the window change when you shift it?
- Shifting the window in *any direction* causes a *big change*



"flat" region:
no change in all directions



"edge":
no change along the edge direction



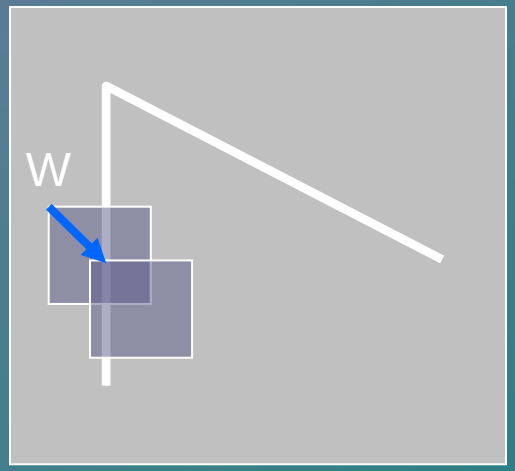
"corner":
significant change in all directions

Slide adapted from Darya Frolova, Denis Simakov, Weizmann Institute.

Corner detection: the math

Consider shifting the window W by (u,v)

- how do the pixels in W change?
- compare each pixel before and after by summing up the squared differences (SSD)
- this defines an SSD "error" of *E(u,v)*:



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

Small motion assumption

Taylor Series expansion of I:

$$I(x+u, y+v) = I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \text{higher order terms}$$

If the motion (u,v) is small, then first order approx is good

$$I(x+u, y+v) \approx I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

$$\approx I(x,y) + [I_x \; I_y] \begin{bmatrix} u \\ v \end{bmatrix}$$

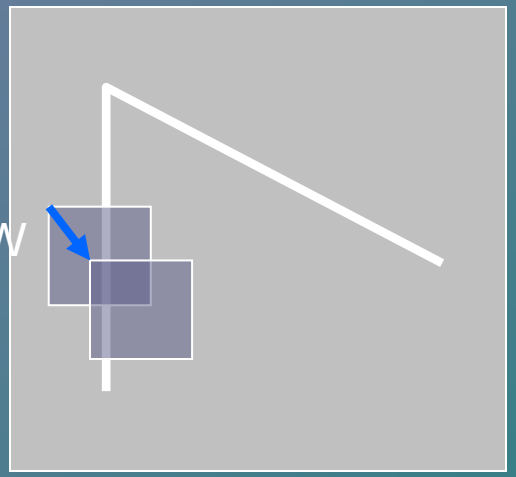
$$\text{shorthand: } I_x = \frac{\partial I}{\partial x}$$

Plugging this into the formula on the previous slide…

Feature detection: the math

Consider shifting the window W by (u,v)

- how do the pixels in W change?
- compare each pixel before and after by summing up the squared differences
- this defines an "error" of E(u,v):



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

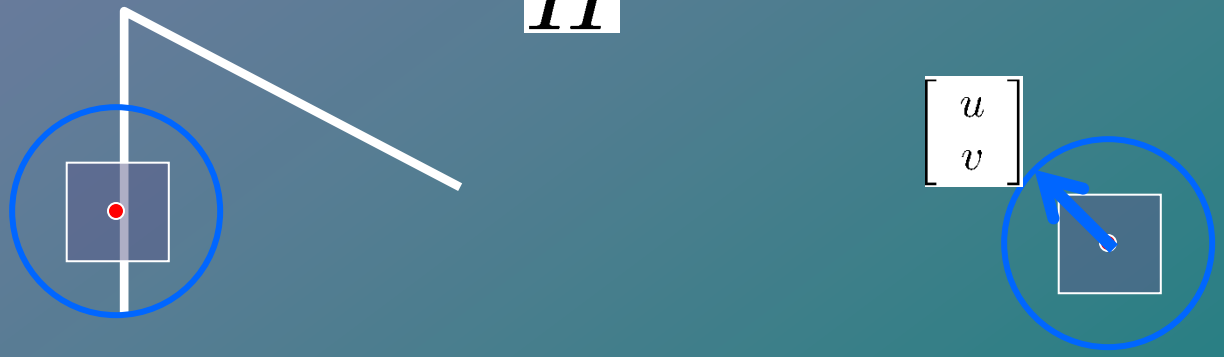
$$\approx \sum_{(x,y)\in W} [I(x,y) + [I_x \; I_y] \begin{bmatrix} u \\ v \end{bmatrix} - I(x,y)]^2$$

$$\approx \sum_{(x,y)\in W} \left[[I_x \; I_y] \begin{bmatrix} u \\ v \end{bmatrix} \right]^2$$

Corner detection: the math

This can be rewritten:

$$E(u,v) = \sum_{(x,y)\in W} [u\ v] \underbrace{\begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix}}_{H} \begin{bmatrix} u \\ v \end{bmatrix}$$



For the example above

- You can move the center of the green window to anywhere on the blue unit circle
- Which directions will result in the largest and smallest E values?
- We can find these directions by looking at the eigenvectors of *H*

Feature detection: the math

This can be rewritten:

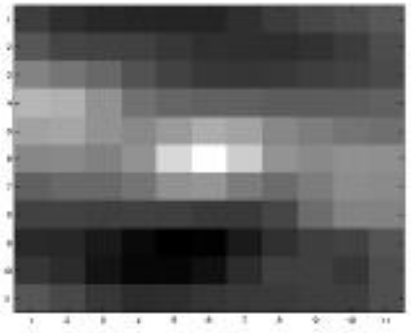
$$E(u,v) = \sum_{(x,y)\in W} [u\ v] \underbrace{\begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix}}_{H} \begin{bmatrix} u \\ v \end{bmatrix}$$

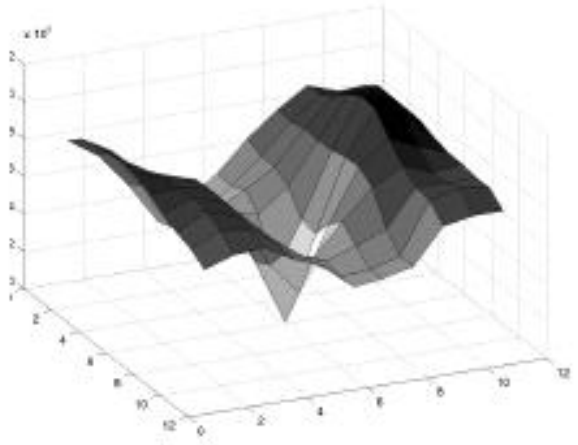
Eigenvalues and eigenvectors of H

- Define shifts with the smallest and largest change (E value)
- x_+ = direction of largest increase in E.
- λ_+ = amount of increase in direction x_+
- x_- = direction of smallest increase in E.
- λ_- = amount of increase in direction x_+

$$Hx_+ = \lambda_+ x_+$$
$$Hx_- = \lambda_- x_-$$

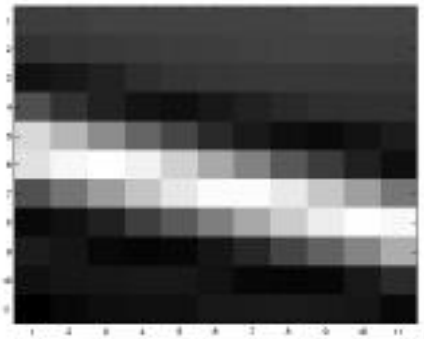
Interest Point Detection

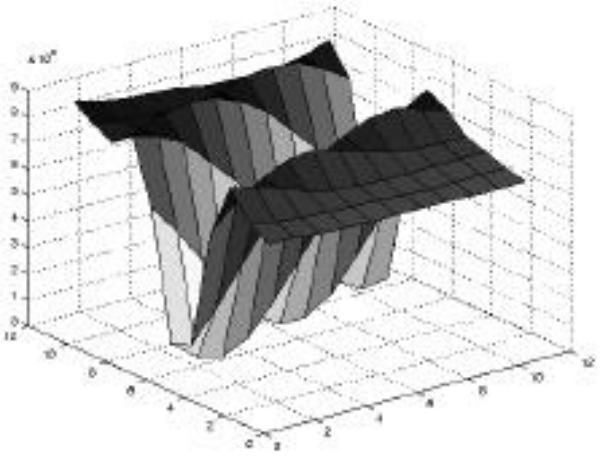




Both λ_+ and λ_- are large

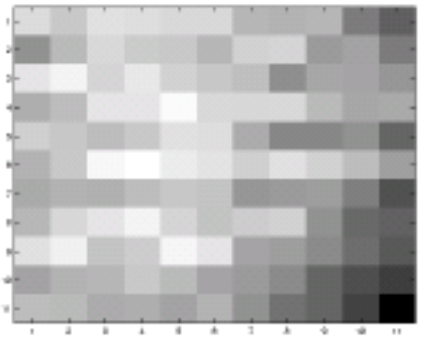
Interest Point Detection

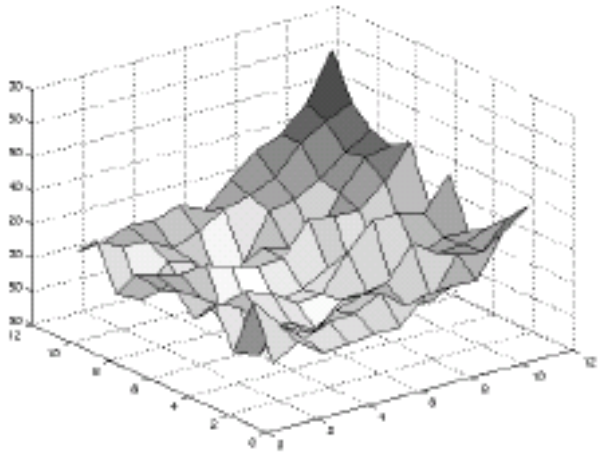




Large λ_+ and small λ_-

Interest Point Detection





Small λ_+ and small λ_-

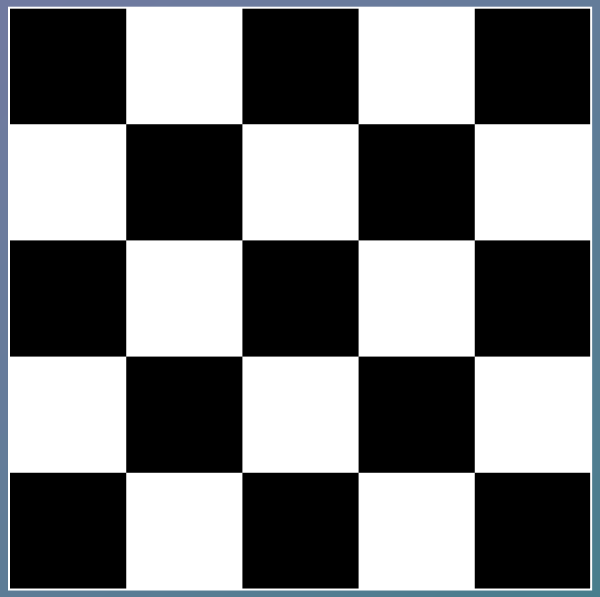
Corner detection: the math

How are λ_+, x_+, λ_-, and x_+ relevant for feature detection?

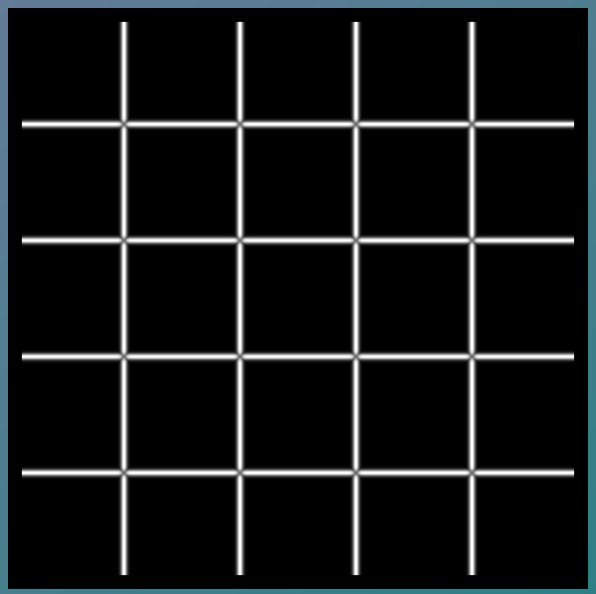
- What's our feature scoring function?

Want *E(u,v)* to be *large* for small shifts in *all* directions

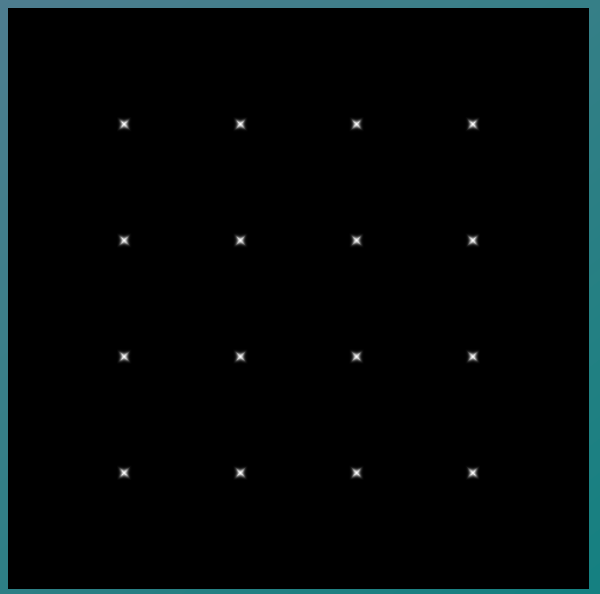
- the *minimum* of *E(u,v)* should be large, over all unit vectors [u v]
- this minimum is given by the smaller eigenvalue (λ_-) of *H*



I



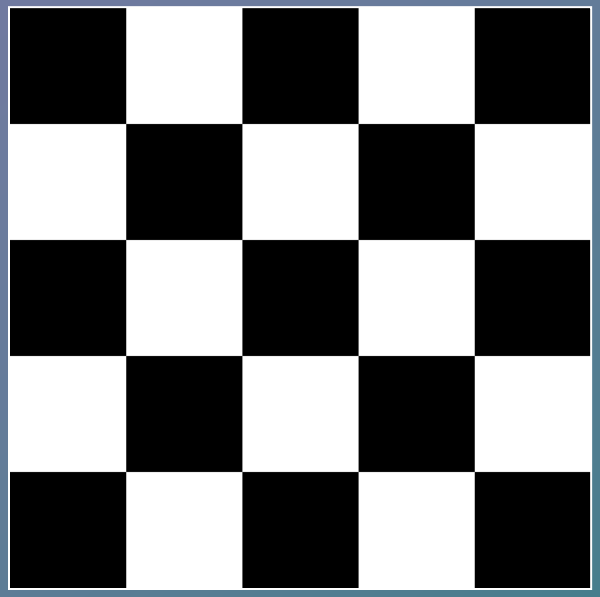
λ_+



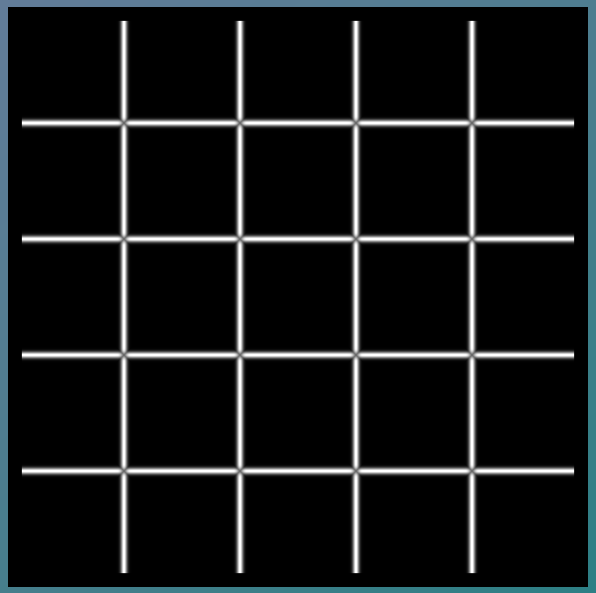
λ_-

Corner detection summary

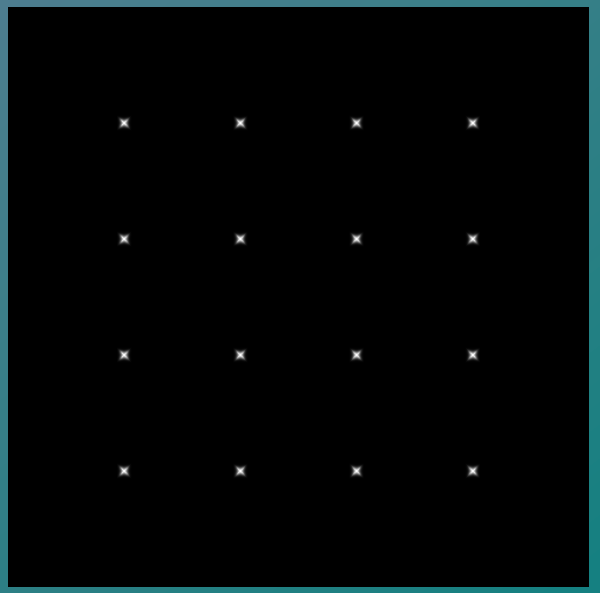
- Compute the gradient at each point in the image
- Create the *H* matrix from the entries in the gradient
- Compute the eigenvalues.
- Find points with large response (λ_- > threshold)
- Choose those points where λ_- is a local maximum as features



I



λ_+



λ_-

The Harris operator

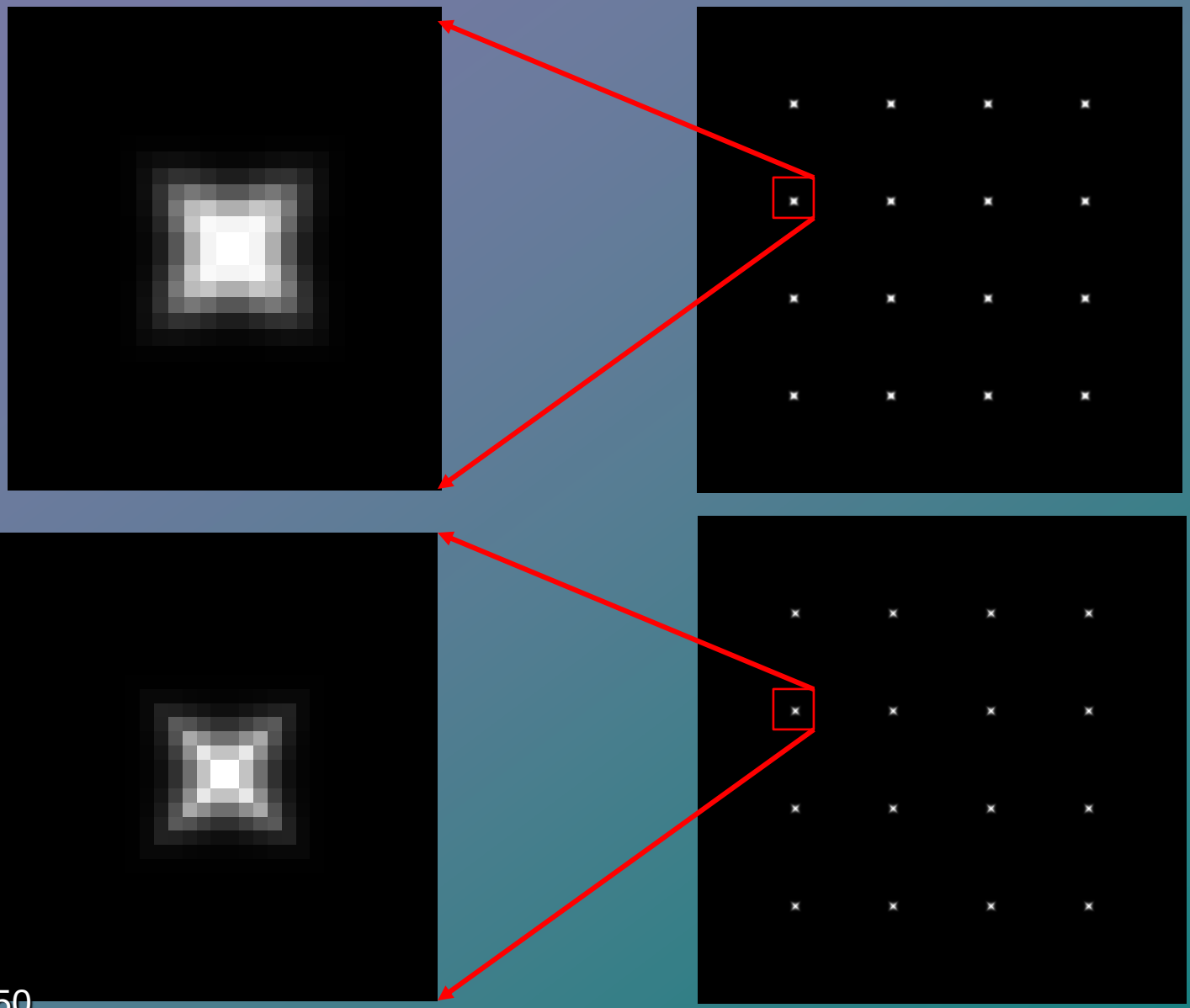
λ_- is a variant of the "Harris operator" for feature detection

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

$$= \frac{determinant(H)}{trace(H)}$$

- The *trace* is the sum of the diagonals, i.e., $trace(H) = h_{11} + h_{22}$
- Very similar to λ_- but less expensive (no square root)
- Called the "Harris Corner Detector" or "Harris Operator"
- Lots of other detectors, this is one of the most popular

The Harris operator



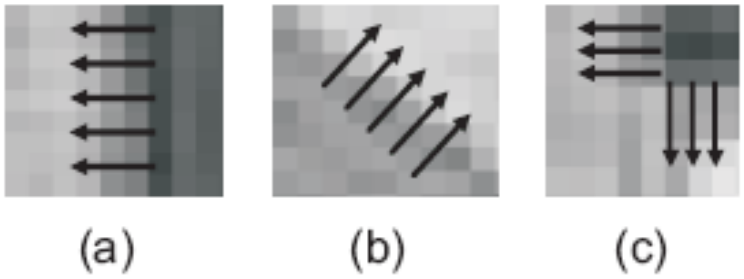
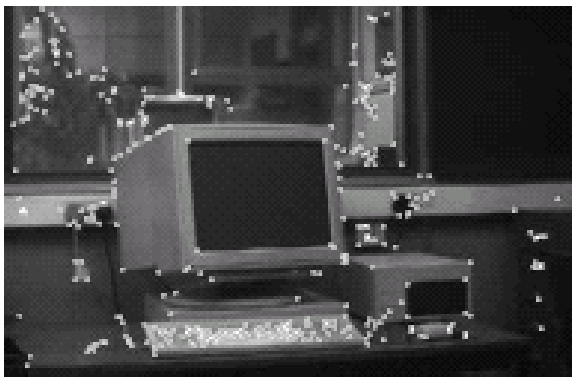
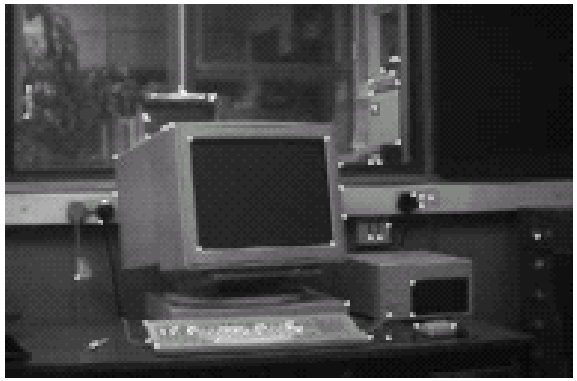


Figure 5.36: Illustration of the decision within Harris corner detector according to eigenvalues of the local structure matrix. (a), (b) Ridge detected, no corner at this position. (c) Corner detected.

The corner detection algorithm we have been describing is due to Harris (1987). It is necessary to calculate A at every pixel and mark corners where the quantity $\lambda_1\lambda_2 - \kappa(\lambda_1+\lambda_2)^2$ exceeds some threshold ($\kappa \approx 0.04$ makes the detector a little "edge-phobic"). Note that $\det A = \lambda_1\lambda_2$ and trace $A = \lambda_1 + \lambda_2$.



Low threshold



High threshold

Harris Detector: Mathematics

Measure of corner response:

$$R = \det \mathbf{A} - k(\mathit{trace}\,\mathbf{A})^2$$

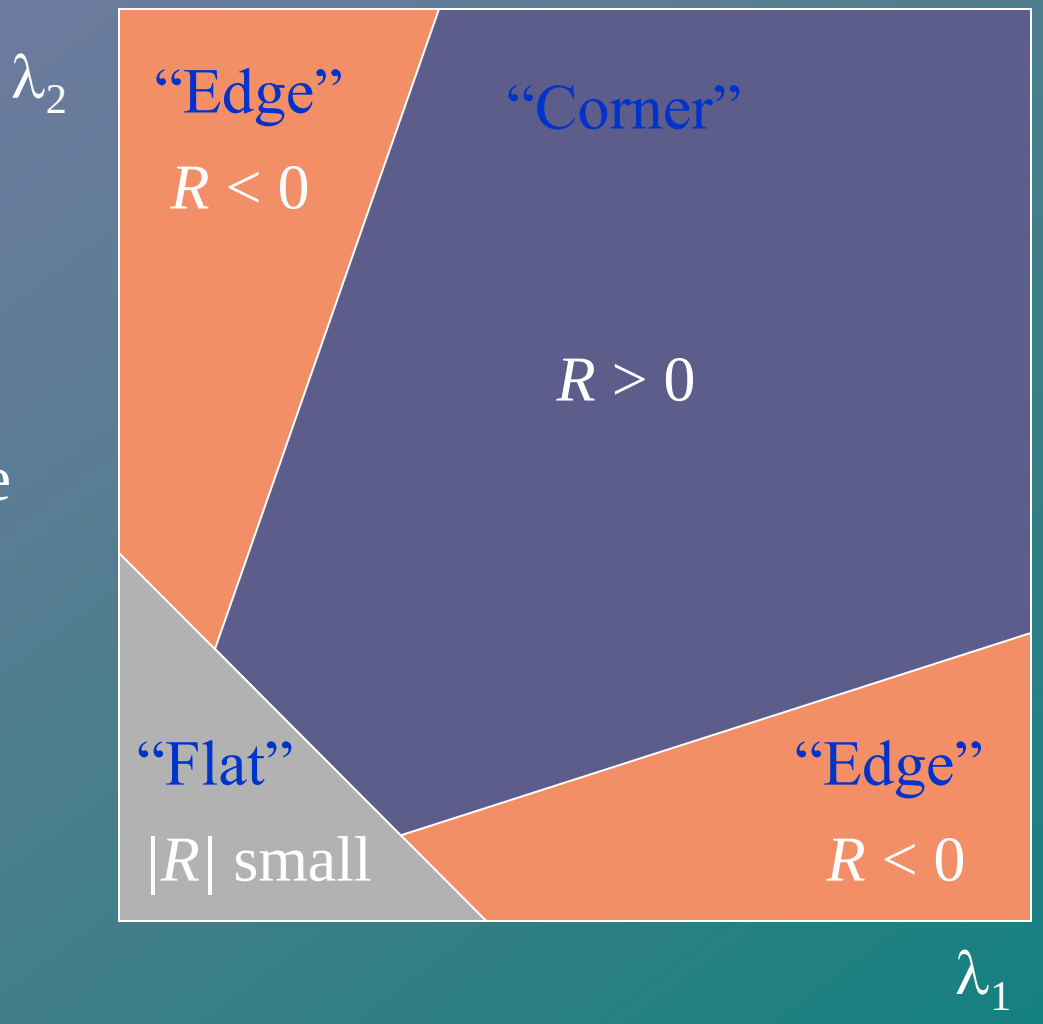
$$\det \mathbf{A} = \lambda_1 \lambda_2$$

$$\mathit{trace}\,\mathbf{A} = \lambda_1 + \lambda_2$$

(k – empirical constant, k = 0.04-0.06)

Harris Detector

- R depends only on eigenvalues of M
- R is large for a corner
- R is negative with large magnitude for an edge
- $|R|$ is small for a flat region



Harris Corner Detector

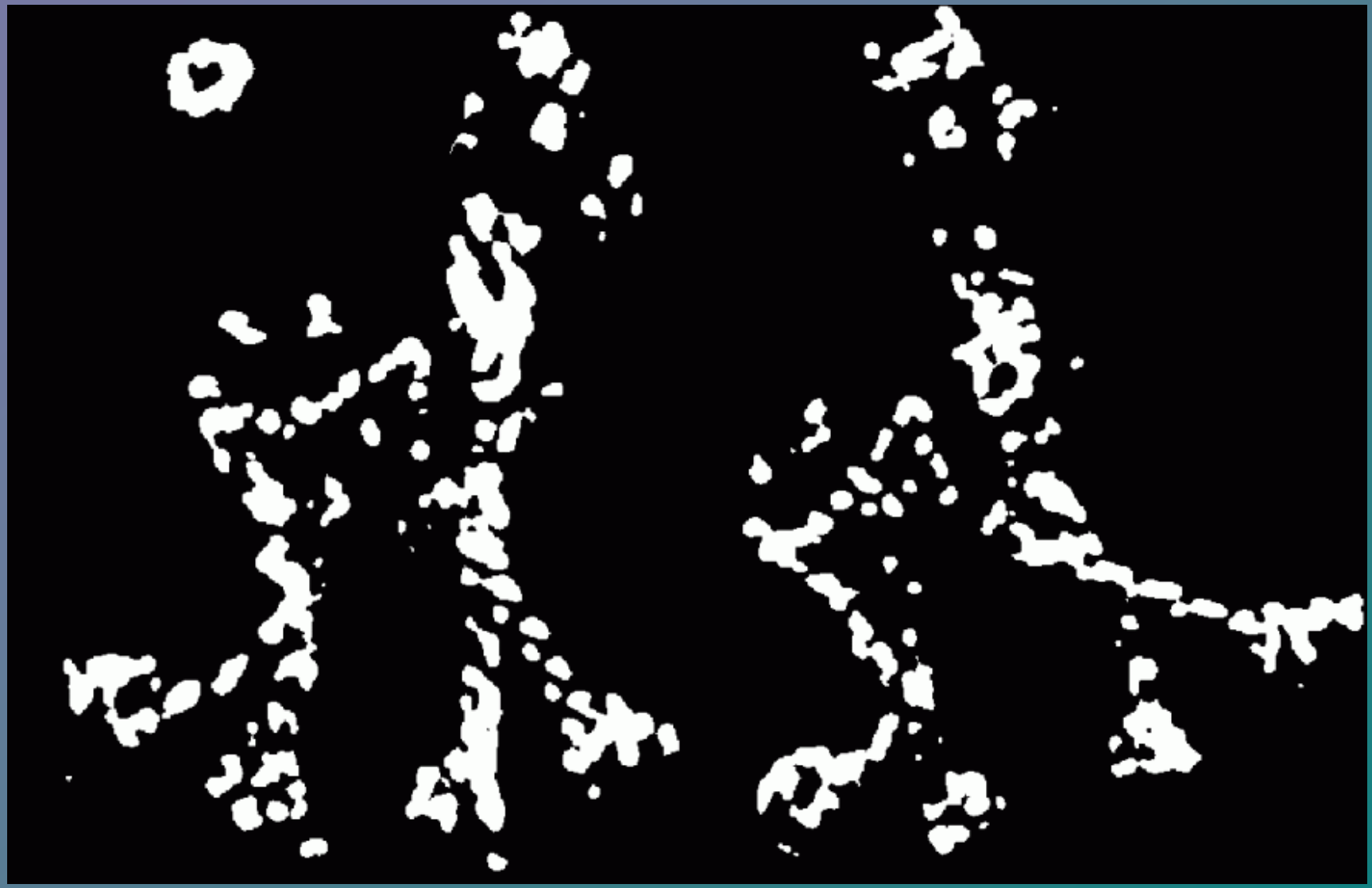
Algorithm 5.5: Harris corner detector

1. Filter the image with a Gaussian.
2. Estimate intensity gradient in two perpendicular directions for each pixel, $\frac{\partial f(x,y)}{\partial x}$, $\frac{\partial f(x,y)}{\partial y}$. This is performed by twice using a 1D convolution with the kernel approximating the derivative.
3. For each pixel and a given neighborhood window:
   - Calculate the local structure matrix A.
   - Evaluate the response function $R(A)$.
4. Choose the best candidates for corners by selecting a threshold on the response function $R(A)$ and perform non-maximal suppression.

Harris Detector: Workflow

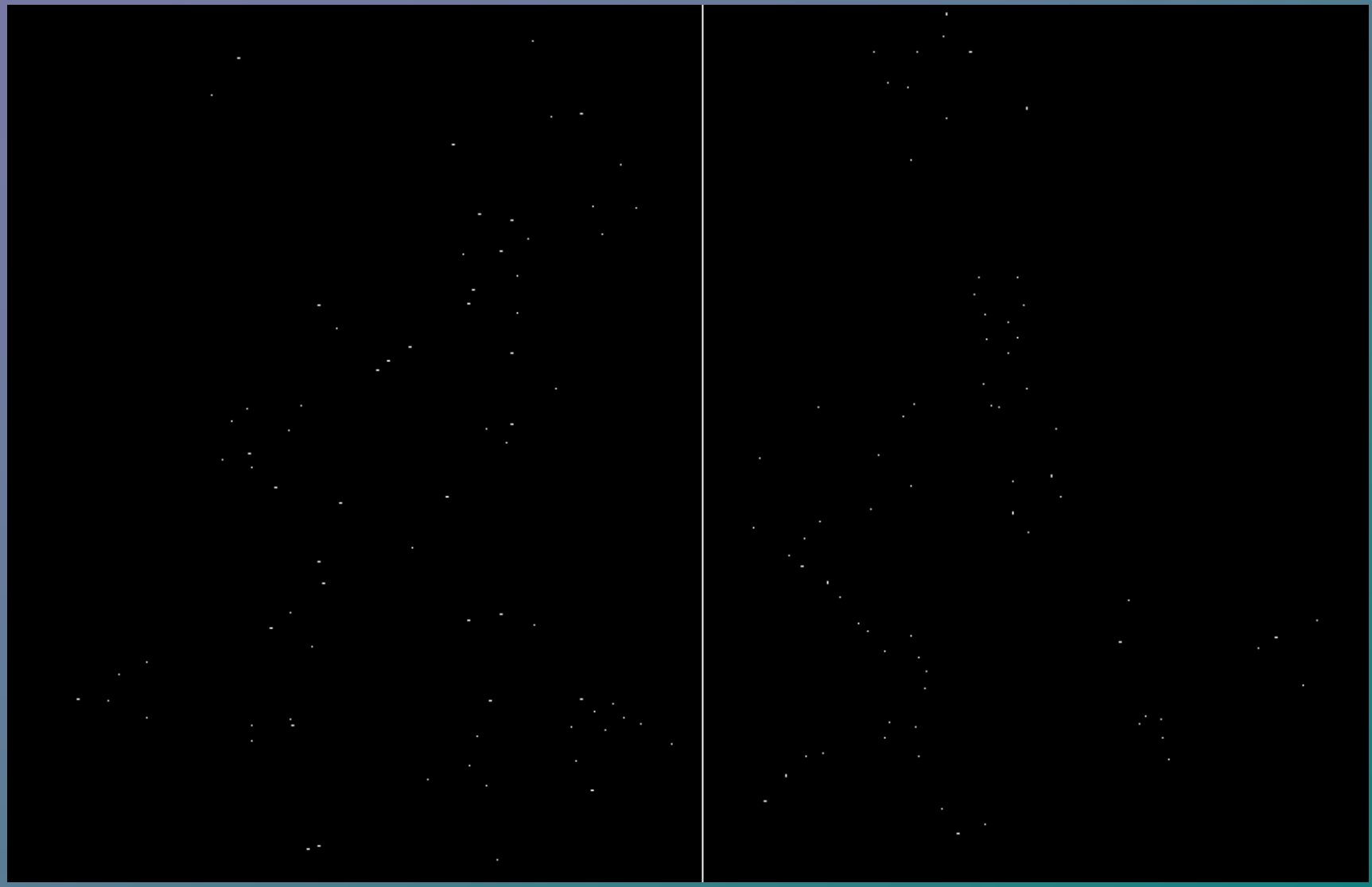
Harris Detector: Workflow

Find points with large corner response: R>threshold



Harris Detector: Workflow

Take only the points of local maxima of R



Harris Detector: Workflow

Example of Harris Corner Detection

Figure 5.37: Example of Harris corners in the image. *Courtesy of Martin Urban, Czech Technical University in Prague, who used such images for 3D reconstruction. A color version of this figure may be seen in the color inset—Plate 7.*

Feature Invariance

Suppose you **rotate** the image by some angle

- Will you still pick up the same features?

What if you change the brightness?

Scale?

Scale invariant detection

Suppose you're looking for corners

Key idea: find scale that gives local maximum of f

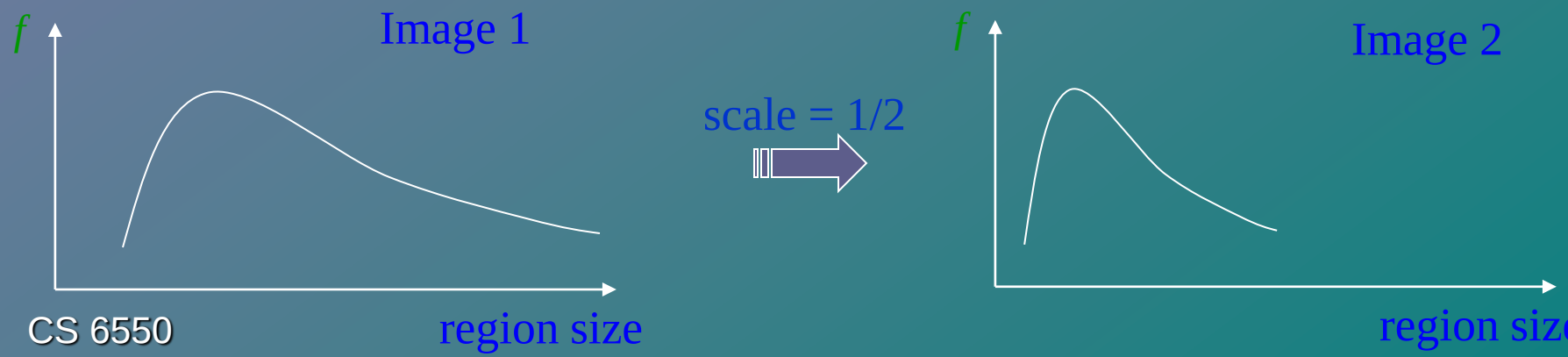
- f is a local maximum in both position and scale
- Common definition of f: Laplacian
  (or difference between two Gaussian filtered images with different sigmas)

Scale Invariant Detection

- Solution:
  - Design a function on the region (circle), which is "scale invariant" (the same for corresponding regions, even if they are at different scales)

Example: average intensity. For corresponding regions (even of different sizes) it will be the same.

  - For a point in one image, we can consider it as a function of region size (circle radius)



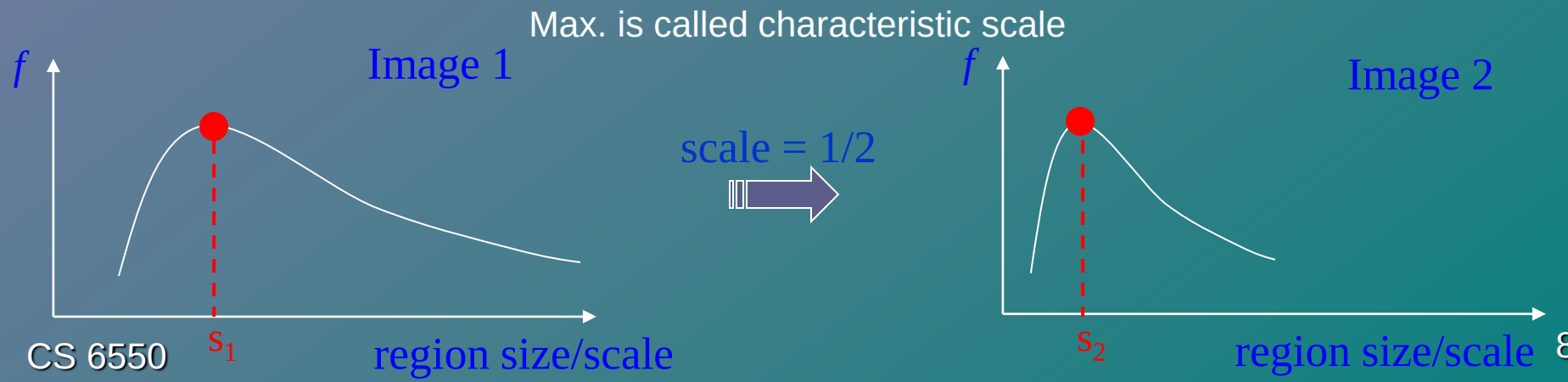
Scale Invariant Detection

- Common approach:

  Take a local maximum of this function

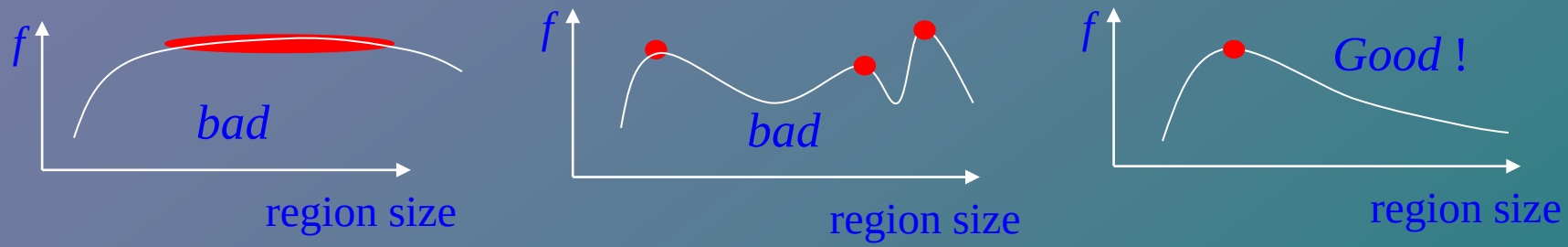
  Observation: region size (scale), for which the maximum is achieved, should be *invariant* to image scale.

Important: this scale invariant region size is found in each image independently!



Scale Invariant Detection

- A “good” function for scale detection: has one stable sharp peak



- For usual images: a good function would be the one with contrast (sharp local intensity change)

Scale Invariant Detection

- Functions for determining scale

$$f = \text{Kernel} * \text{Image}$$

Kernels:

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

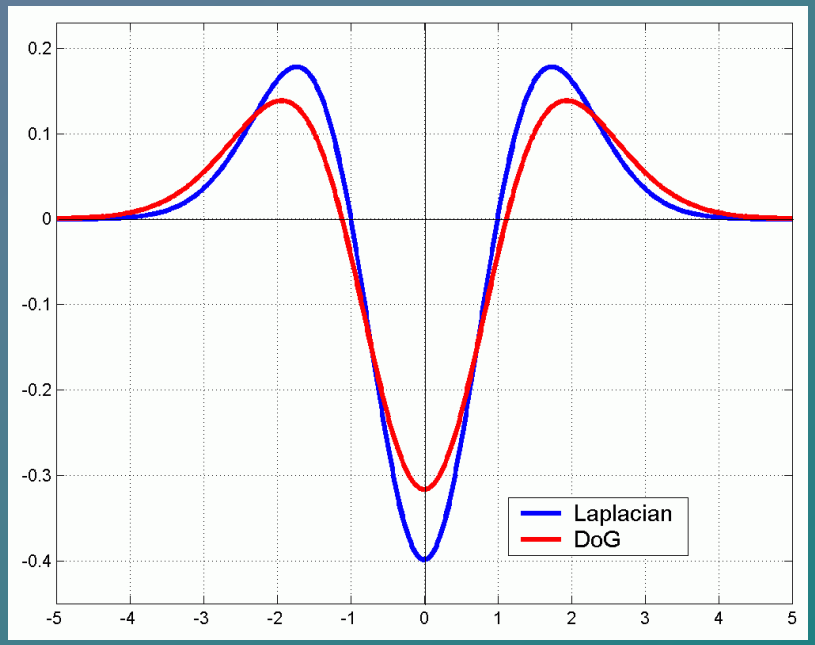
(Laplacian)

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

(Difference of Gaussians)

where Gaussian

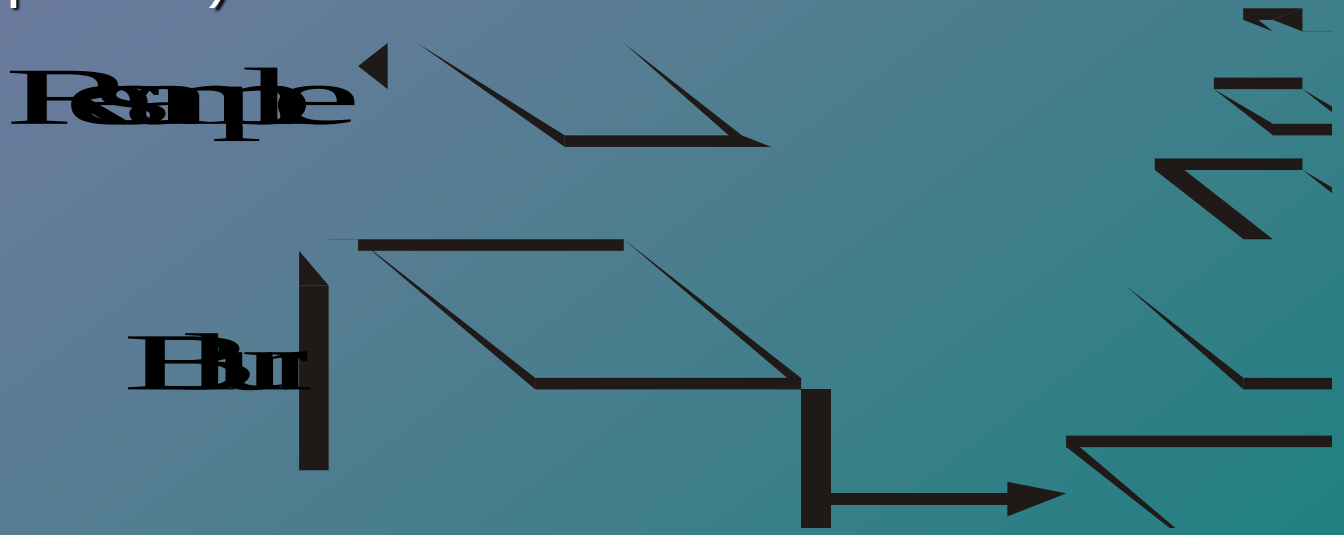
$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$



Note: both kernels are invariant to *scale* and *rotation*

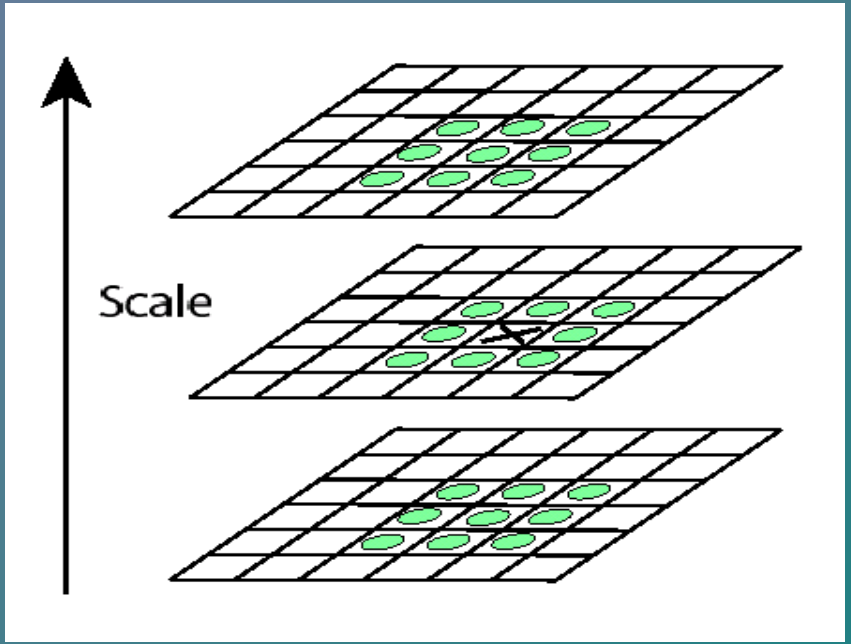
Build Scale-Space Pyramid

- All scales must be examined to identify scale-invariant features
- An efficient function is to compute the Difference of Gaussian (DOG) pyramid (Burt & Adelson, 1983) (or Laplacian)



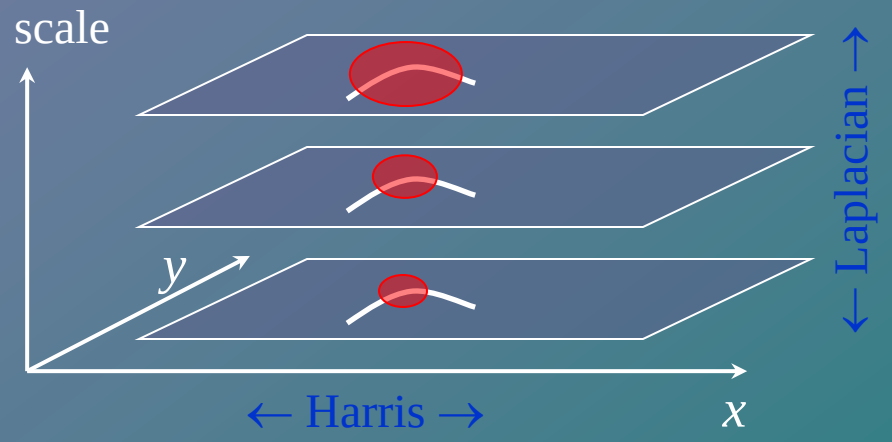
Key point localization

- Detect maxima and minima of difference-of-Gaussian in scale space

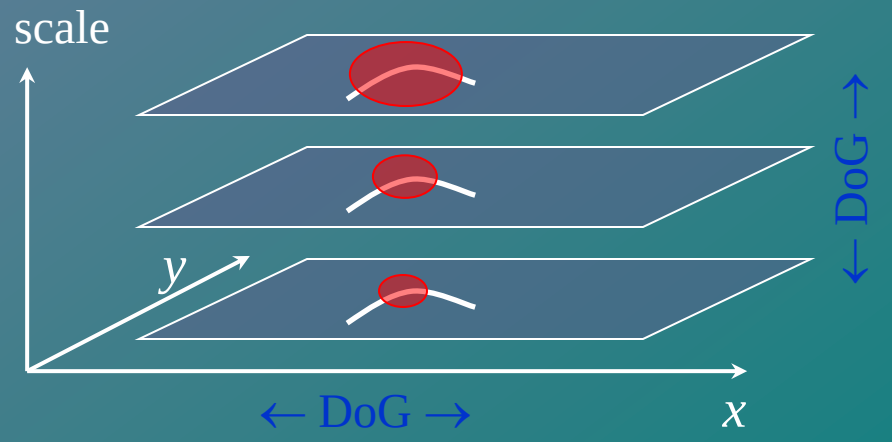


Scale Invariant Detectors

- Harris-Laplacian[1]
  Find local maximum of:
  - Harris corner detector in image space
  - Laplacian in scale

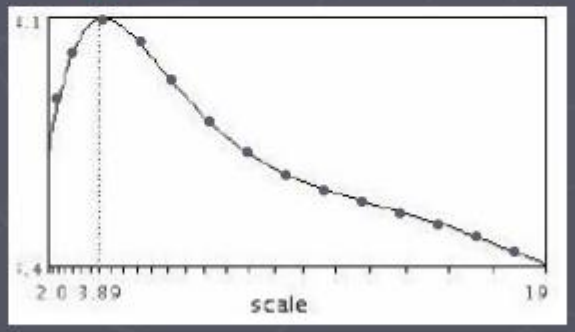


- SIFT (Lowe)[2]
  Find local maximum of:
  - Difference of Gaussians in space and scale

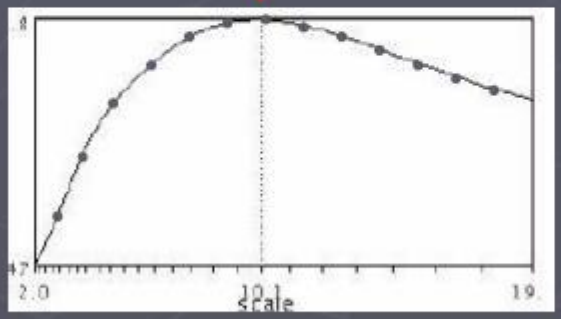


[1] K.Mikolajczyk, C.Schmid. "Indexing Based on Scale Invariant Interest Points". ICCV 2001
[2] D.Lowe. "Distinctive Image Features from Scale-Invariant Keypoints". Accepted to IJCV 2004

Automatic scale selection



$f(I_{i_1 \ldots i_m}(x,\sigma))$



$f(I_{i_1 \ldots i_m}(x',\sigma'))$

Obtaining Location, Radius, Orientation for a Corner Detector

Assume a fixed scale parameter k
Apply a corner detector to the image $\mathcal{I}$
Initialize a list of patches
For each corner detected
  Write (x_c, y_c) for the location of the corner
  Compute the radius r for the patch at (x_c, y_c) as

$$r(x_c, y_c) = \underset{\sigma}{\text{argmax}} \ \nabla^2_\sigma \mathcal{I}(x_c, y_c)$$

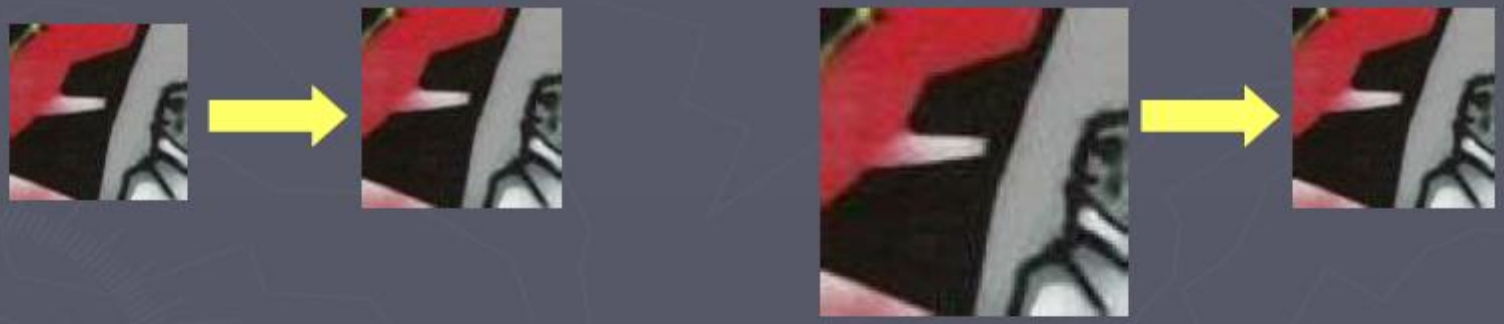
  by computing $\nabla^2_\sigma \mathcal{I}(x_c, y_c)$ for a variety of values of σ, interpolating these values, and maximizing
  Compute an orientation histogram $H(\theta)$ for gradient orientations within a radius kr of (x_c, y_c).
  Compute the orientation of the patch θ_p as

$$\theta_p = \underset{\theta}{\text{argmax}} \ H(\theta).$$

  If there is more than one theta that maximizes this histogram, make one copy of the patch for each.
  Attach (x_c, y_c, r, θ_p) to the list of patches for each copy

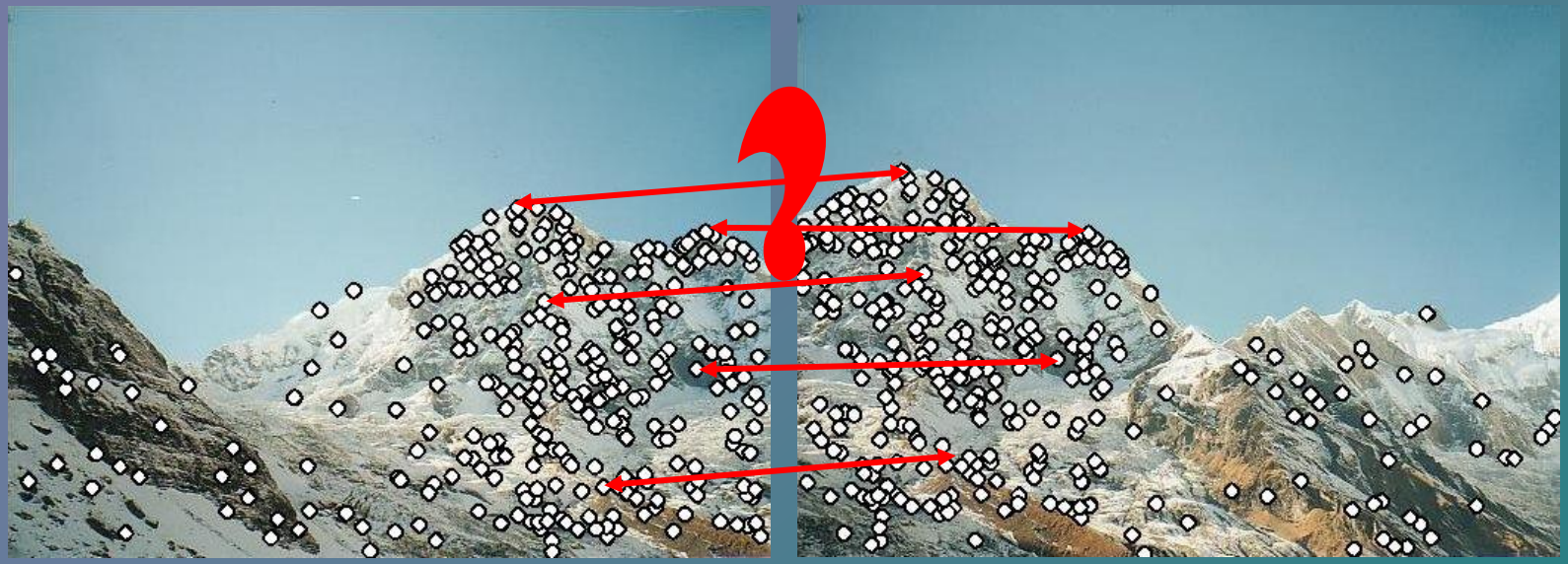
Automatic scale selection

Normalize: rescale to fixed size



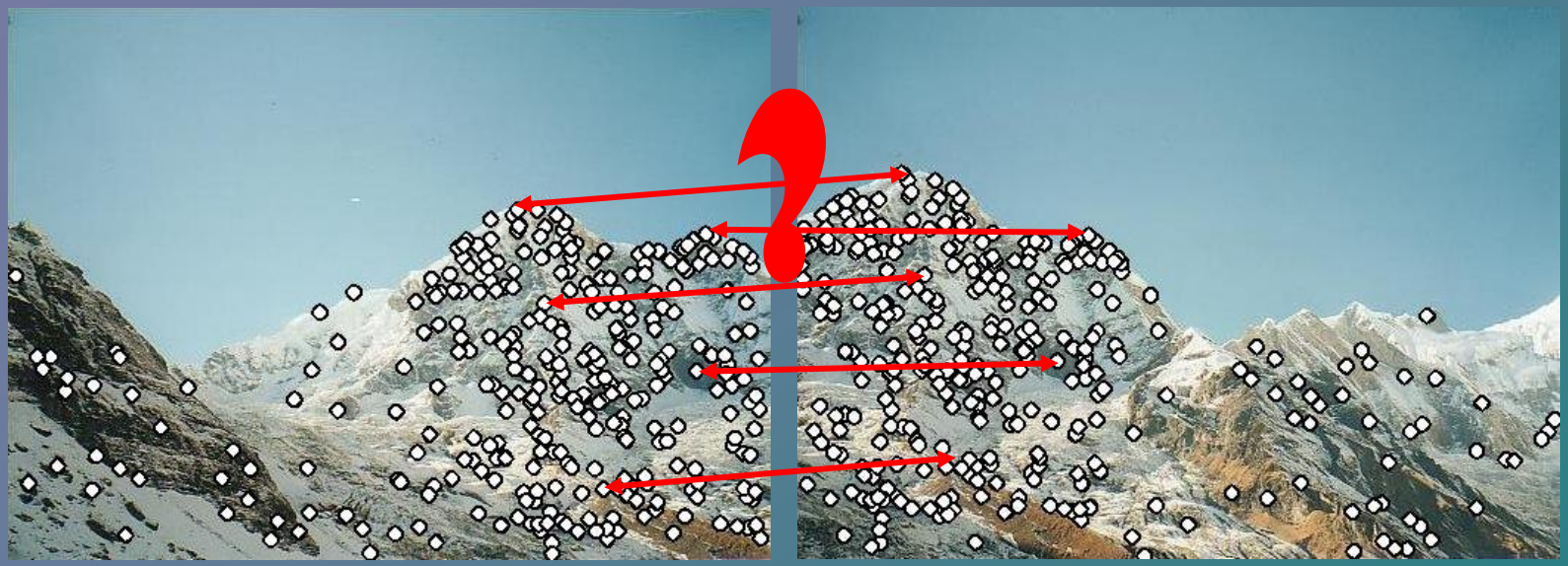
Feature descriptors

We know how to detect good points
Next question: **How to match them?**



Feature descriptors

We know how to detect good points
Next question: **How to match them?**



Lots of possibilities (this is a popular research area)

- Simple option: match square windows around the point
- State of the art approach: SIFT
  - David Lowe, UBC http://www.cs.ubc.ca/~lowe/keypoints/

Invariance

Suppose we are comparing two images I_1 and I_2

- I_2 may be a transformed version of I_1
- What kinds of transformations are we likely to encounter in practice?

We'd like to find the same features regardless of the transformation

- This is called transformational ***invariance***
- Most feature methods are designed to be invariant to
  - Translation, 2D rotation, scale
- They can usually also handle
  - Limited 3D rotations (SIFT works up to about 60 degrees)
  - Limited affine transformations (some are fully affine invariant)
  - Limited illumination/contrast changes

How to achieve invariance

Need both of the following:

1. Make sure your detector is invariant
   - Harris is invariant to translation and rotation
   - Scale is trickier
     - common approach is to detect features at many scales using a Gaussian pyramid (e.g., MOPS)
     - More sophisticated methods find "the best scale" to represent each feature (e.g., SIFT)
2. Design an invariant feature *descriptor*
   - A descriptor captures the information in a region around the detected feature point
   - The simplest descriptor: a square window of pixels
     - What's this invariant to?
   - Let's look at some better approaches…

Rotation invariance for feature descriptors

Find dominant orientation of the image patch

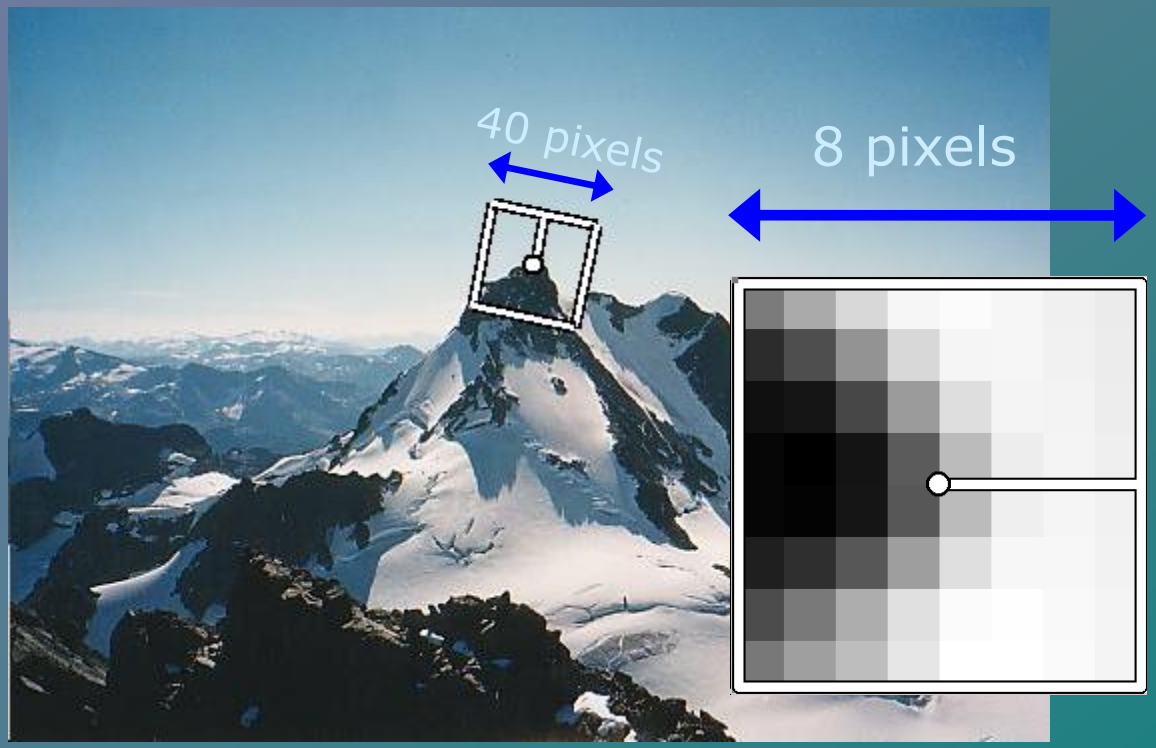
- This is given by $\mathbf{x}_+$, the eigenvector of $\mathbf{H}$ corresponding to λ_+
  - λ_+ is the *larger* eigenvalue
- Rotate the patch according to this angle

Figure by Matthew Brown

Multiscale Oriented PatcheS descriptor

Take 40x40 square window around detected feature

- Scale to 1/5 size (using prefiltering)
- Rotate to horizontal
- Sample 8x8 square window centered at feature
- Intensity normalize the window by subtracting the mean, dividing by the standard deviation in the window



Adapted from slide by Matthew Brown

Detections at multiple scales

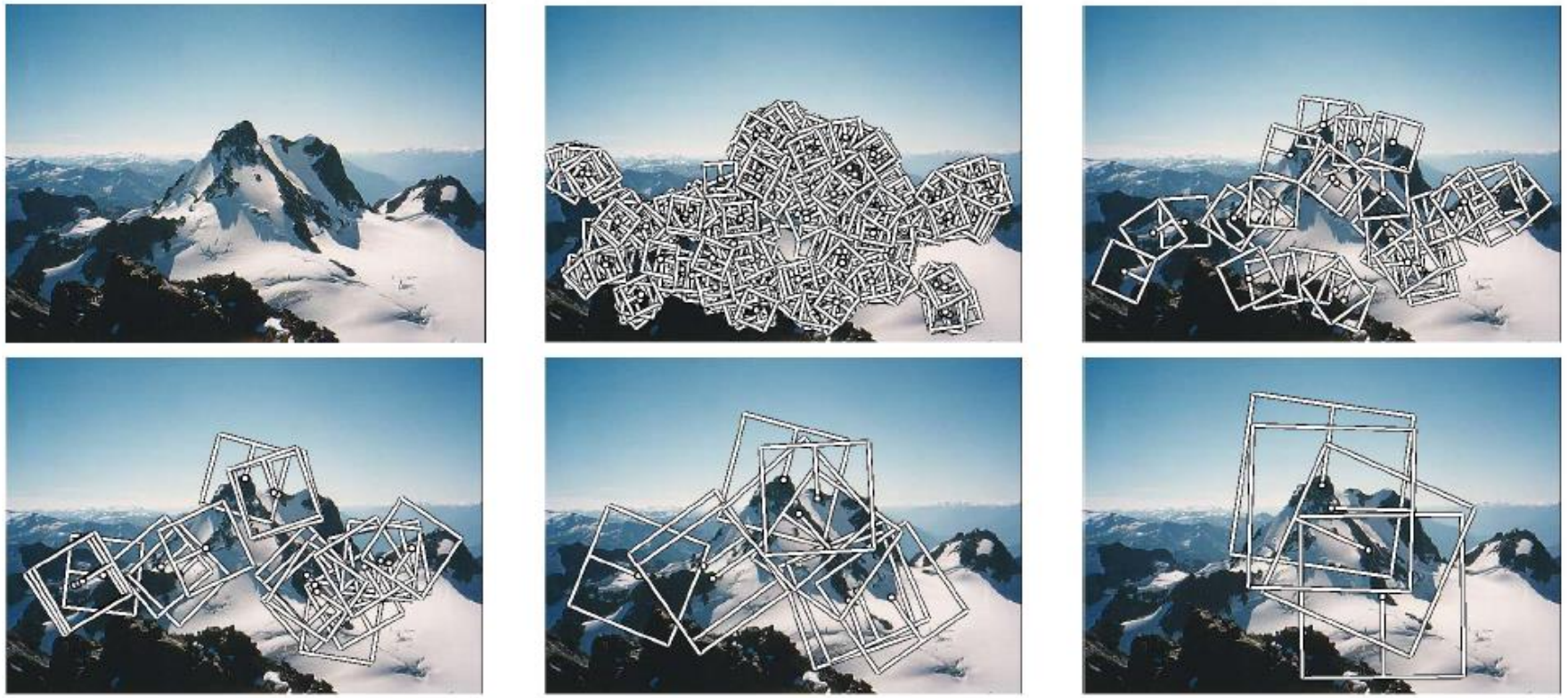
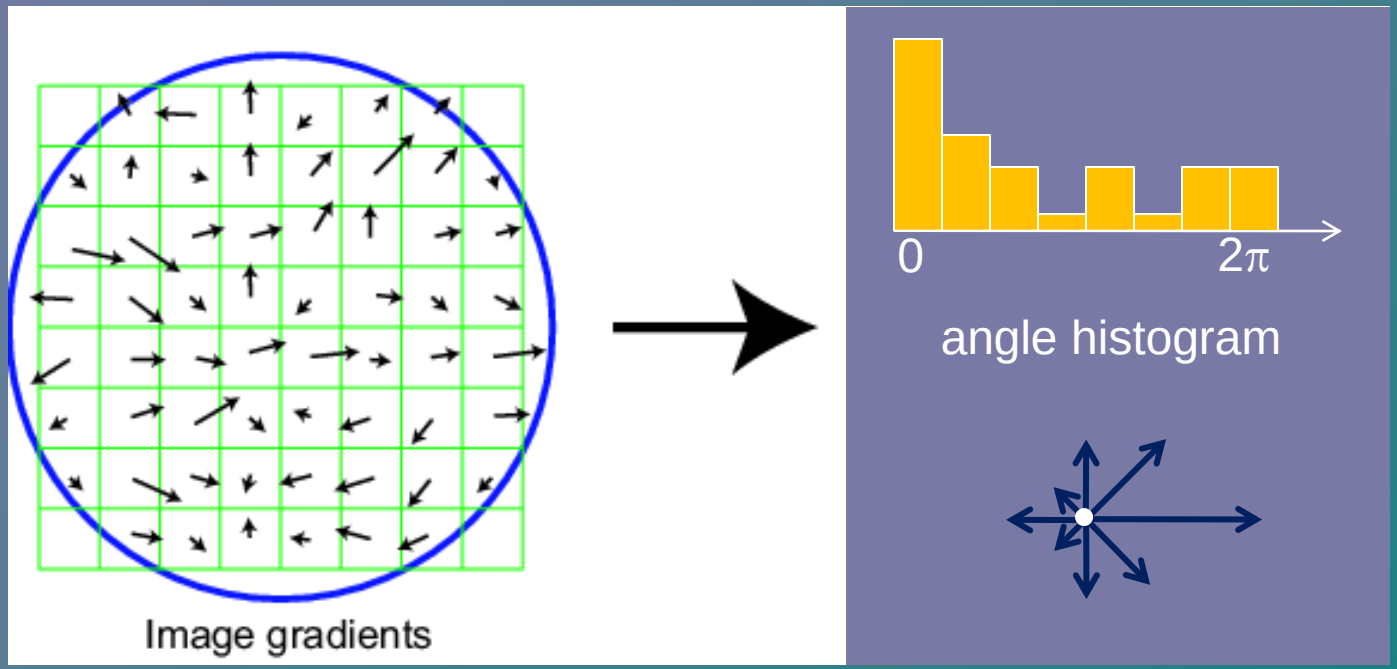


Figure 1. Multi-scale Oriented Patches (MOPS) extracted at five pyramid levels from one of the Matier images. The boxes show the feature orientation and the region from which the descriptor vector is sampled.

Scale Invariant Feature Transform

Basic idea:

- Take 16x16 square window around detected feature
- Compute edge orientation (angle of the gradient - 90°) for each pixel
- Throw out weak edges (threshold gradient magnitude)
- Create histogram of surviving edge orientations

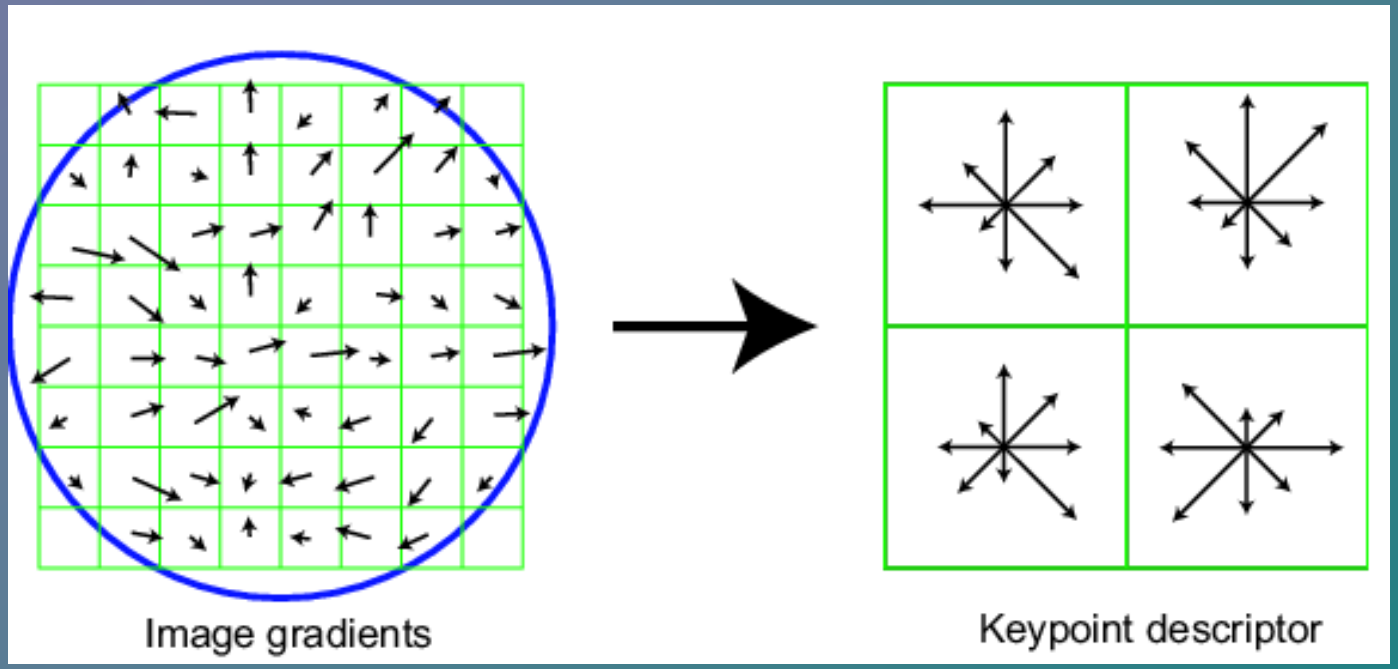


Adapted from slide by David Lowe

SIFT descriptor

Full version

- Divide the 16x16 window into a 4x4 grid of cells (2x2 case shown below)
- Compute an orientation histogram for each cell
- 16 cells * 8 orientations = 128 dimensional descriptor

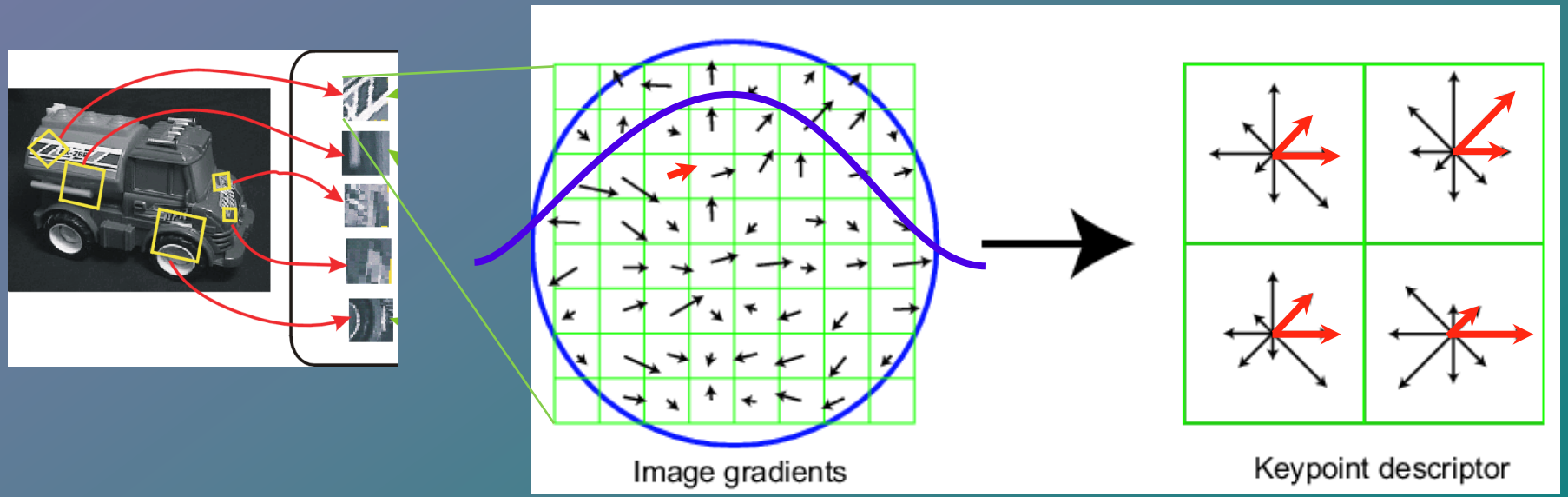


showing only 2x2 here but is 4x4

Adapted from slide by David Lowe

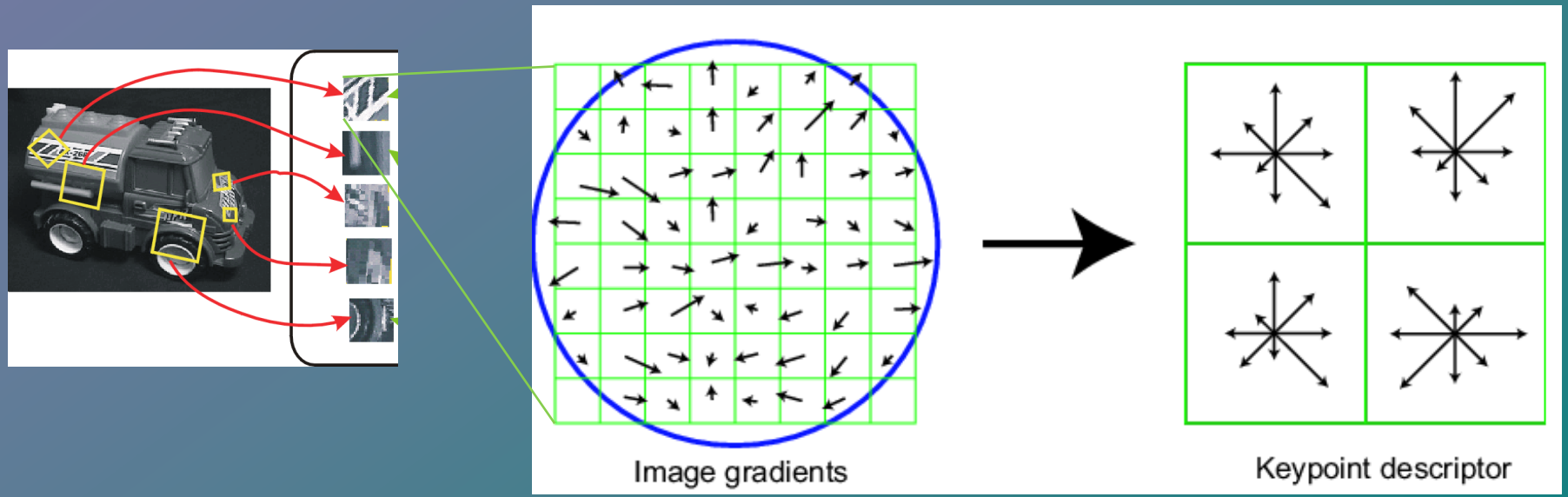
Ensure Smoothness

- Gaussian weight
- Trilinear interpolation
  - a given gradient contributes to 8 bins: 4 in space times 2 in orientation



Reduce effect of illumination

- 128-dim vector normalized to 1
- Threshold gradient magnitudes to avoid excessive influence of high gradients
  - after normalization, clamp gradients >0.2
  - renormalize



SIFT Descriptor Computing Flow

Given an image $\mathcal{I}$, and a patch with center (x_c, y_c),
radius r, orientation θ, and parameters n, m, q, k and t.
For each element of the $n \times n$ grid centered at (x_c, y_c) with spacing kr
Compute a weighted q element histogram of the averaged
gradient samples at each point of the $m \times m$ subgrid,
as in Algorithm 5.5.
Form an $n \times n \times q$ vector v by concatenating the histograms.
Compute $u = v/\sqrt{v \cdot v}$.
Form w whose i'th element w_i is $\min(u_i, t)$.
The descriptor is $d = w/\sqrt{w \cdot w}$.

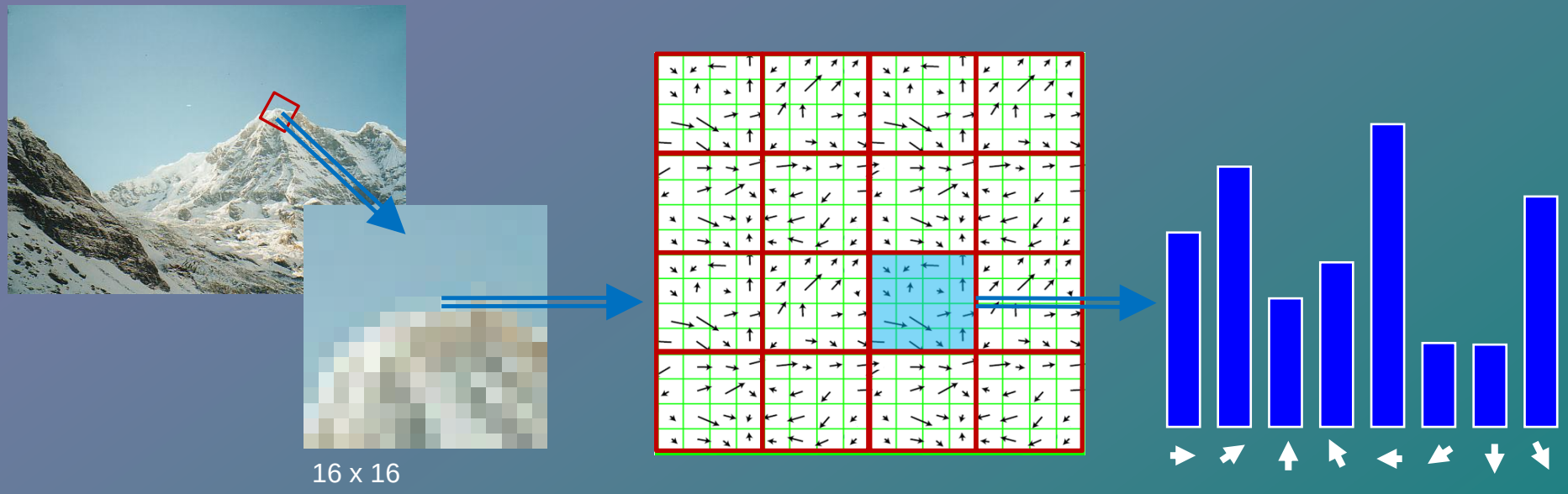
Weighted Oriented Histogram

Given a grid cell $\mathcal{G}$ for patch with center $\boldsymbol{c} = (x_c, y_c)$ and radius r

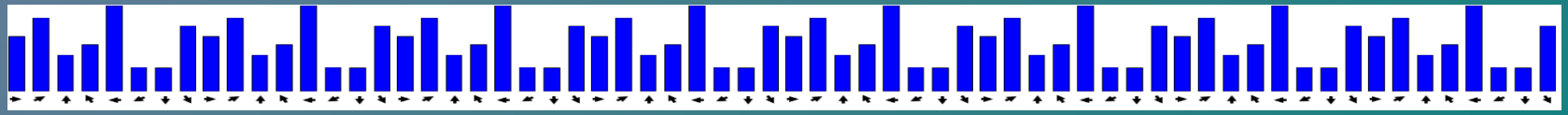
Create an orientation histogram
For each point $\boldsymbol{p}$ in an $m \times m$ subgrid spanning $\mathcal{G}$
  Compute a gradient estimate $\nabla\mathcal{I}\,|_{\boldsymbol{p}}$ estimate at $\boldsymbol{p}$
    as a weighted average of $\nabla\mathcal{I}$, using bilinear weights centered at $\boldsymbol{p}$.
  Add a vote with weight $\|\nabla\mathcal{I}\|\frac{1}{r\sqrt{2\pi}}\exp\left(-\frac{\|\boldsymbol{p}-\boldsymbol{c}\|^2}{r^2}\right)$
    to the orientation histogram cell for the orientation of $\nabla\mathcal{I}$.

SIFT descriptor

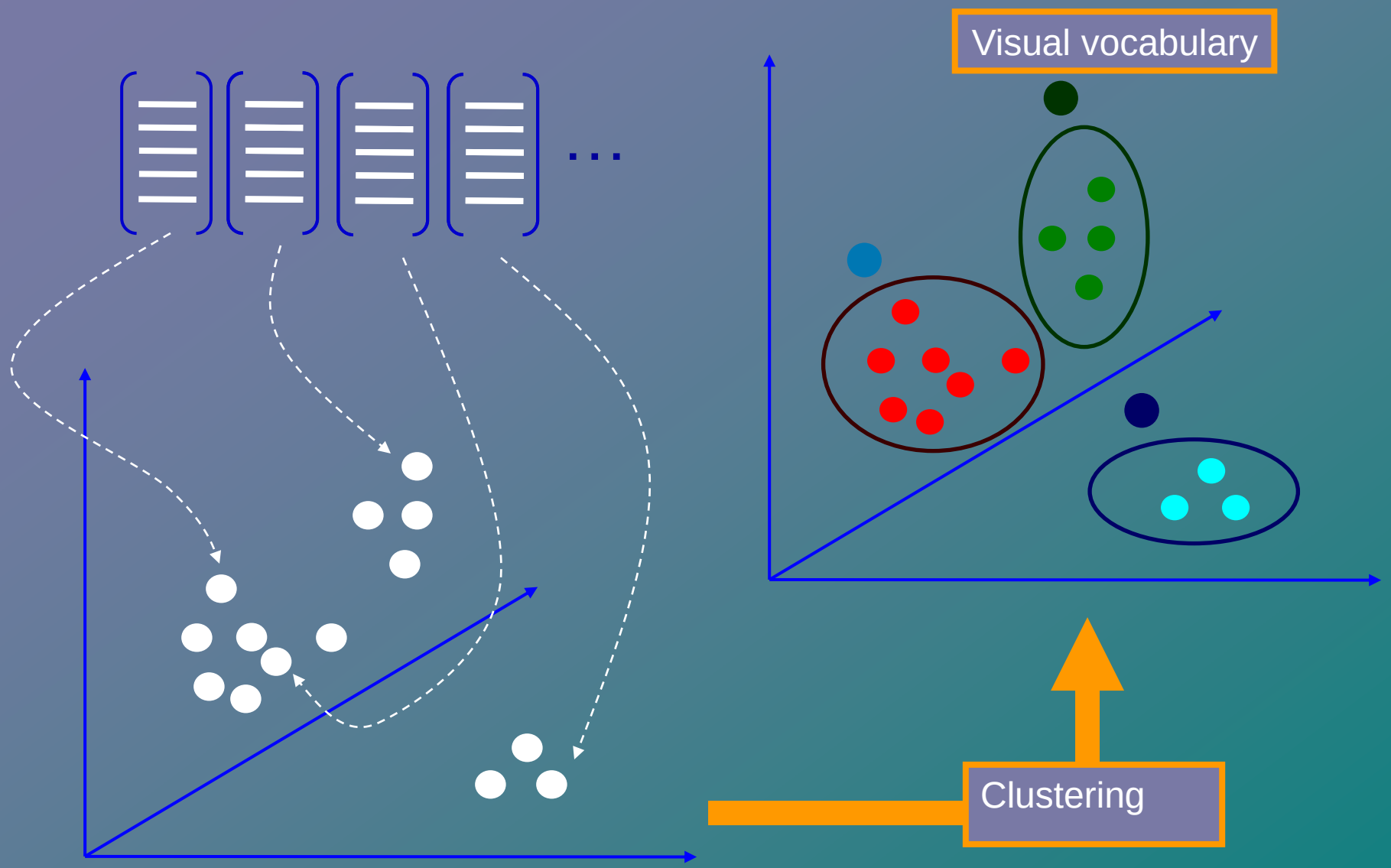
- SIFT (Scale Invariant Feature Transform)



The result: 128 **dimensions feature vector.**

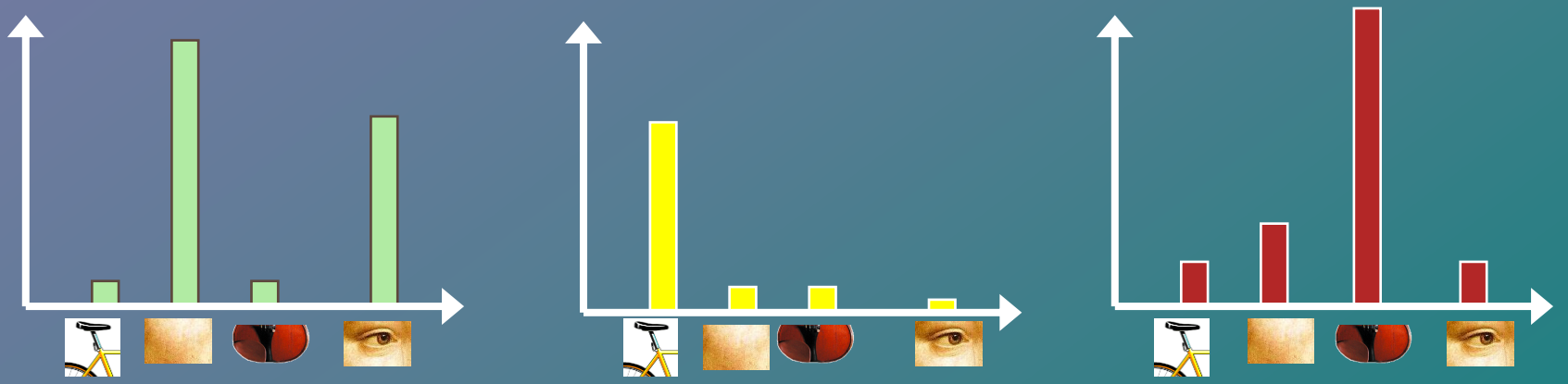


Learning the visual vocabulary

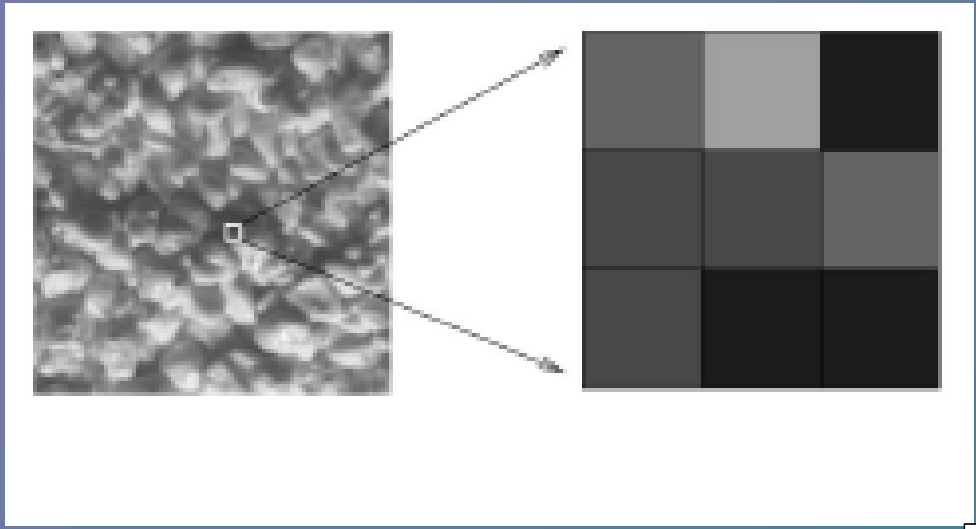


Bag-of-Words

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”



Local Binary Pattern (LBP)



85	99	21
54	54	86
57	12	13

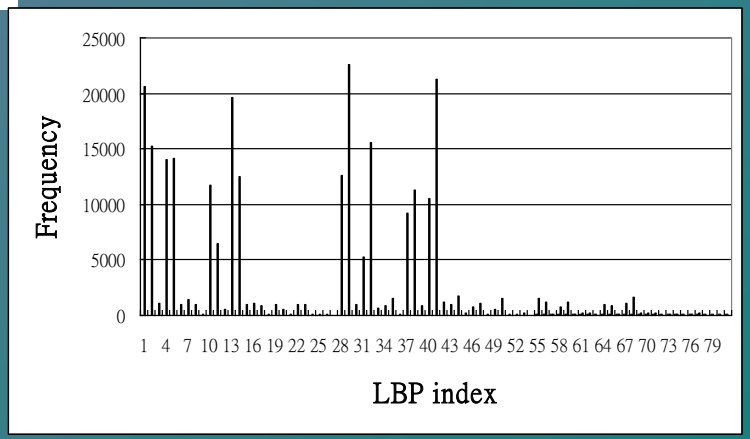
Sub-Image

1	1	0
1		1
1	0	0

11010011
11101001
11110100
⋮
10100111

$(00111101)_2 \rightarrow (61)_{10}$

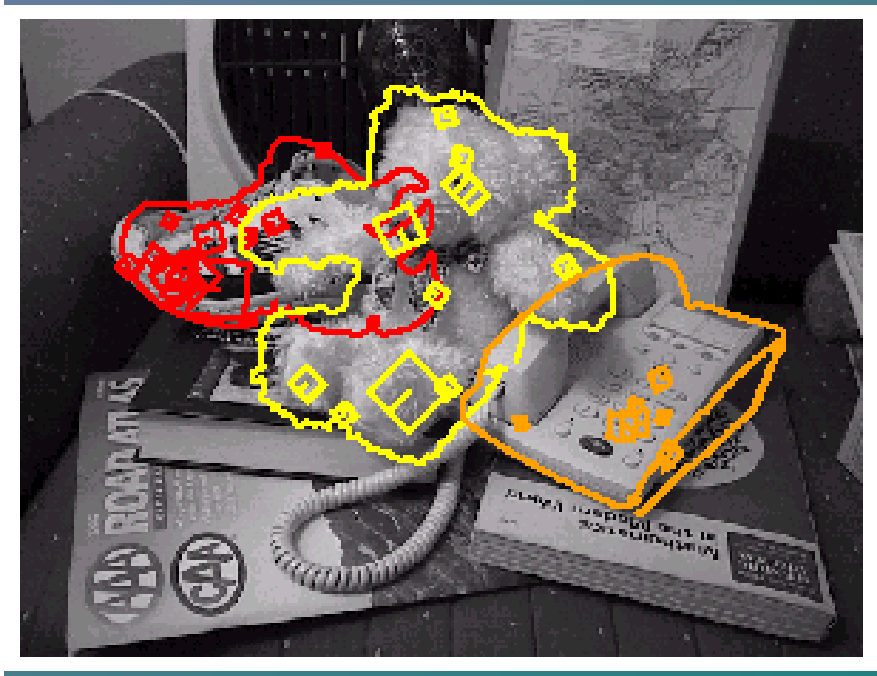
LBP index



Uniform LBP

- An important special case of LBP is the *uniform LBP. A* LBP descriptor is called *uniform if and only if at most two* bitwise transition between 0 and 1 over the circulated binary feature.
- **For example:**
- 00000000 (0 transition), 11100011 (2 transitions) are uniform
- 01010000 (4 transitions), 01110101 (6 transitions) are non-uniform

Object recognition



Images from David Lowe

Summary

- Things to take away from this unit
  - Brightness transformation
  - Edge detection by differentiation
  - Image gradients
  - Laplacian operator
    - Laplacian of Gaussian (LoG)
  - Canny edge detector (basic idea)
    - Effects of varying sigma parameter
  - Texture Features
  - Corner Detection
  - Interest Point Descriptors (SIFT, LBP, …)