

Computer Vision

Linear Filtering and Edge Detection

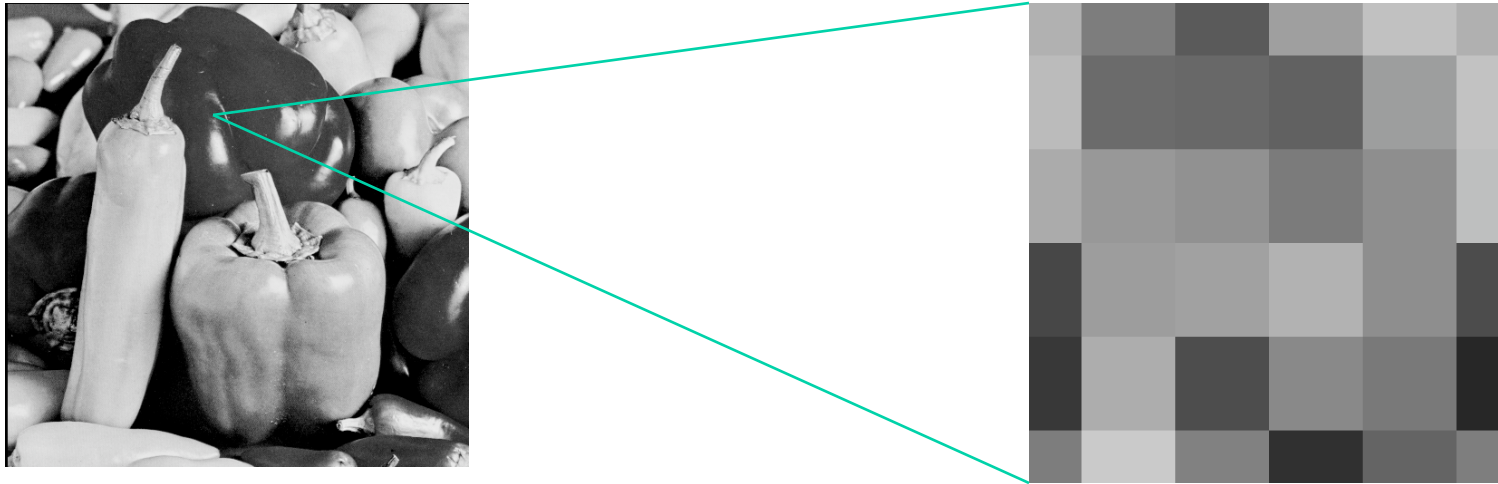
Professor Hager
<http://www.cs.jhu.edu/~hager>

Outline

- Image noise
- Filtering by Convolution
- Properties of Convolution
- Derivative Operators
- The Canny Edge Detector

Goals: To understand the properties of common linear filtering operations on gray-scale images and to understand how they can be used to solve problems

IMAGE NOISE



Cameras are not perfect sensors
Scenes never quite match our expectations

Noise Models

- Noise is commonly modeled using the notion of “additive white noise.”
 - Scalar example: $I(x) = I^*(x) + n(x)$
 - Images: $I(i,j,t) = I^*(i,j,t) + n(i,j,t)$
 - Note that $n(i,j,t)$ is independent of $n(i',j',t')$ unless $i'=i,j'=j,t'=t$.
 - Typically we assume that n (noise) is independent of image location --- that is, it is i.i.d

Gaussian
Noise:
 $\sigma=1$



9/17/08

Gaussian
Noise:
 $\sigma=16$



Properties of Noise Processes

- Properties of temporal image noise:

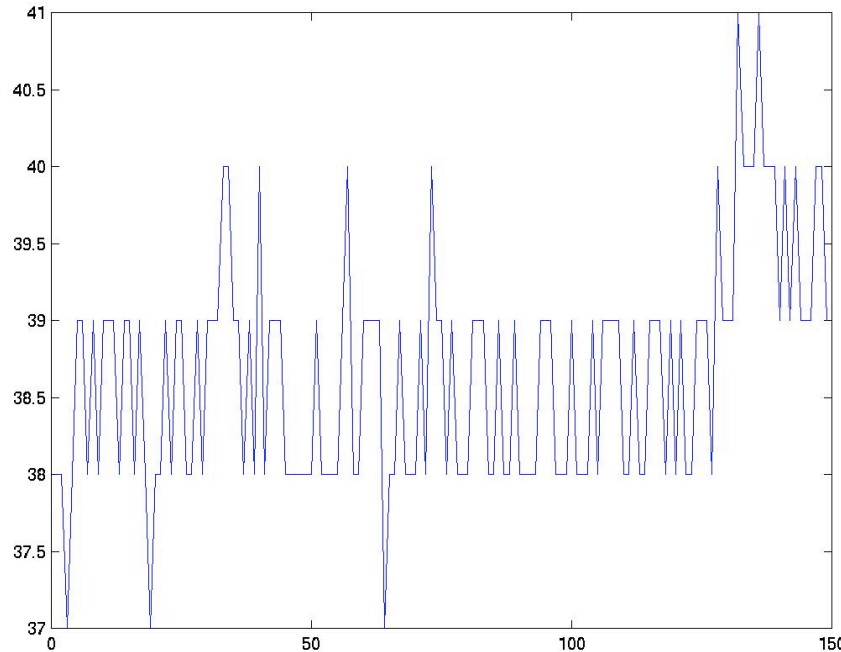
$$\text{Mean } m(i,j) = \sum I(i,j,t)/n$$

$$\text{Standard Deviation } s_{i,j} = \text{Sqrt}(\sum (m(i,j) - I(i,j,t))^2/n)$$

$$\text{Signal-to-noise Ratio } \frac{m(i,j)}{s_{i,j}}$$

Image Noise

- An experiment: take several images of a static scene and look at the pixel values



mean = 38.6
std = 2.99

$\text{Snr} = 38.6/2.99 \approx 13$
 $\text{max snr} = 255/3 \approx 85$

PROPERTIES OF TEMPORAL IMAGE NOISE (i.e., successive images)

If standard deviation of grey values at a pixel is s for a pixel for a single image, then the laws of statistics states that for independent sampling of grey values, for a temporal average of n images, the standard deviation is:

$$\frac{s}{\text{Sqrt}(n)}$$

For example, if we want to double the signal to noise ratio, we could average 4 images.

Temporal vs. Spatial Noise

- It is common to assume that:
 - spatial noise in an image is consistent with the temporal image noise
 - the spatial noise is independent and identically distributed
- Thus, we can think of the image itself as an additive noise process
- Averaging is a common way to reduce noise
 - instead of temporal averaging, how about spatial?
- For example, for a pixel in image I at i, j

$$I'(i, j) = 1/9 \sum_{i'=i-1}^{i+1} \sum_{j'=j-1}^{j+1} I(i', j')$$

Convolution

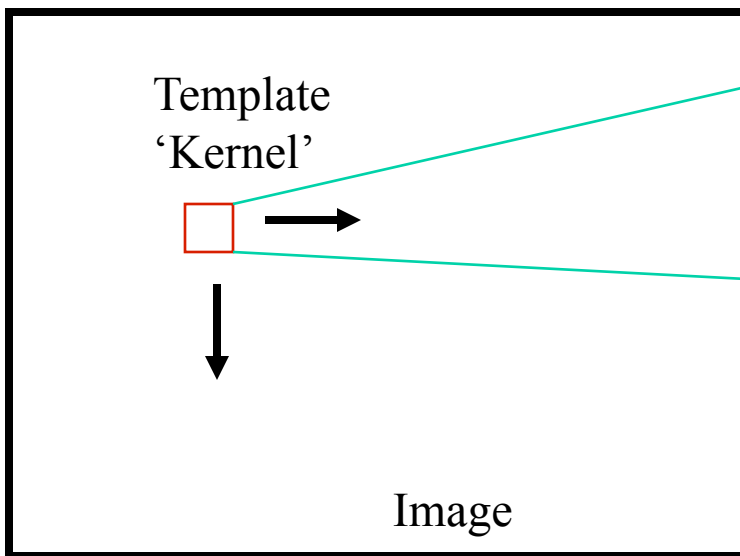
Convolution is the generalization of this “averaging” process. As we’ll see, it can do more than average.

$$(I * K)(x,y) = \sum_i \sum_j I(x-i,y-j) K(i,j)$$

Note these indices run backwards --- this can sometimes fool you down the road!

We often write $I' = I * K$ to represent the convolution of I by K . K is referred to as the *kernel* of the convolution (or sometimes the “stencil” in the discrete case).

DISCRETE CONVOLUTION



T1	T2	T3
T4	T5	T6
T7	T8	T9

3x3
Template

I1	I2	I3
I4	I5	I6
I7	I8	I9

Local
Image
Neighborhood

$$\begin{aligned} I &= T9 \times I1 + T8 \times I2 + T7 \times I3 \\ &+ T6 \times I4 + T5 \times I5 + T4 \times I6 \\ &+ T3 \times I7 + T2 \times I8 + T1 \times I9 \end{aligned}$$

Convolution

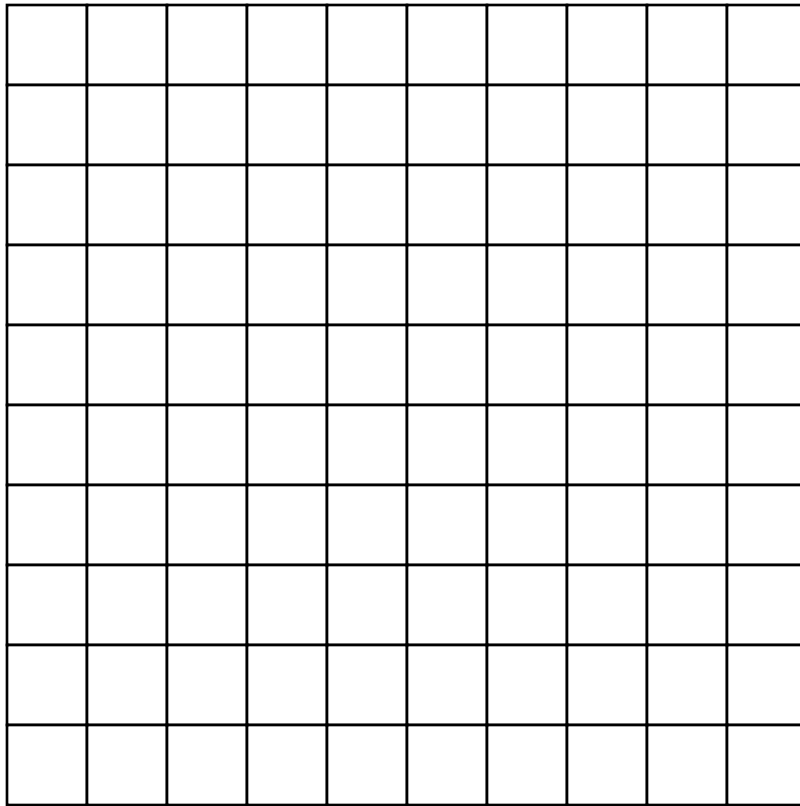


Image (I)

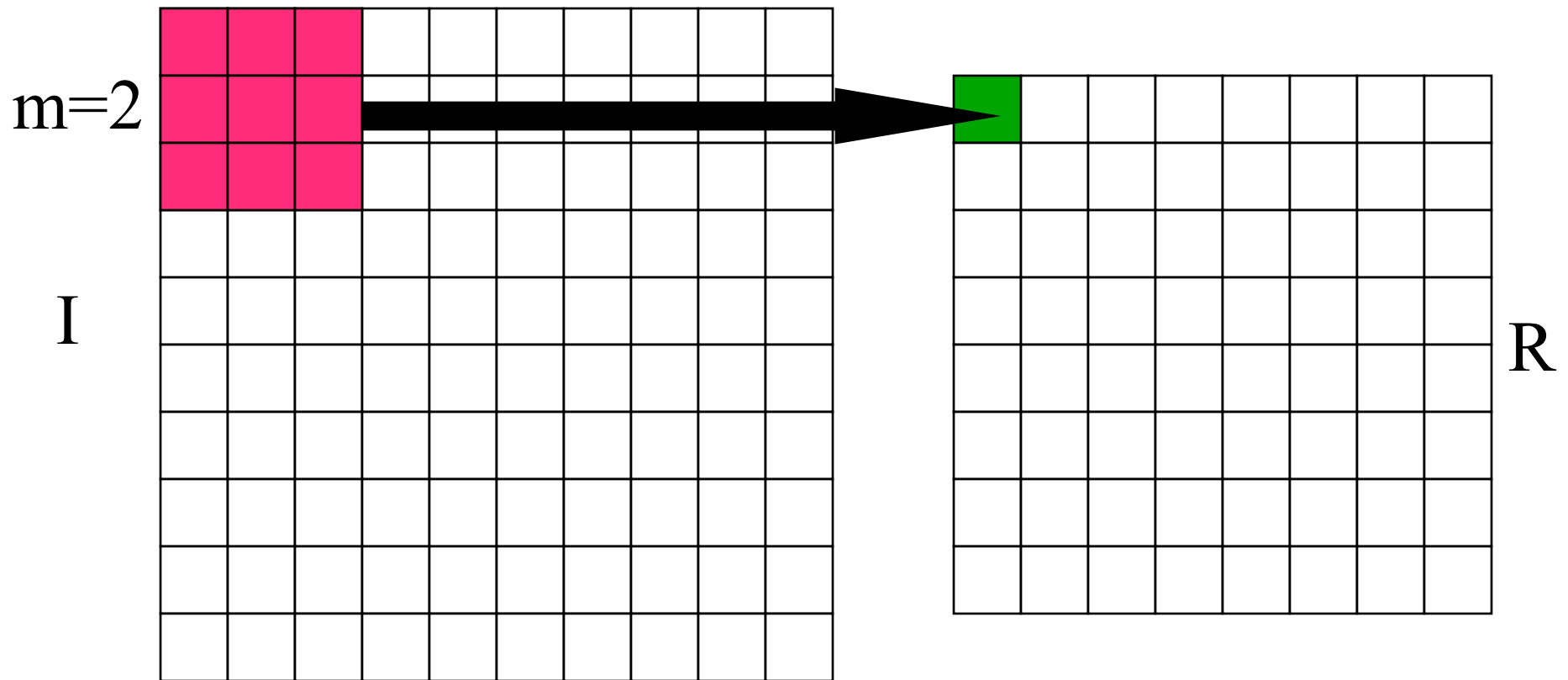
*

1	2	1
2	4	2
1	2	1

Kernel (K)

Note: Typically Kernel is relatively small in vision applications.

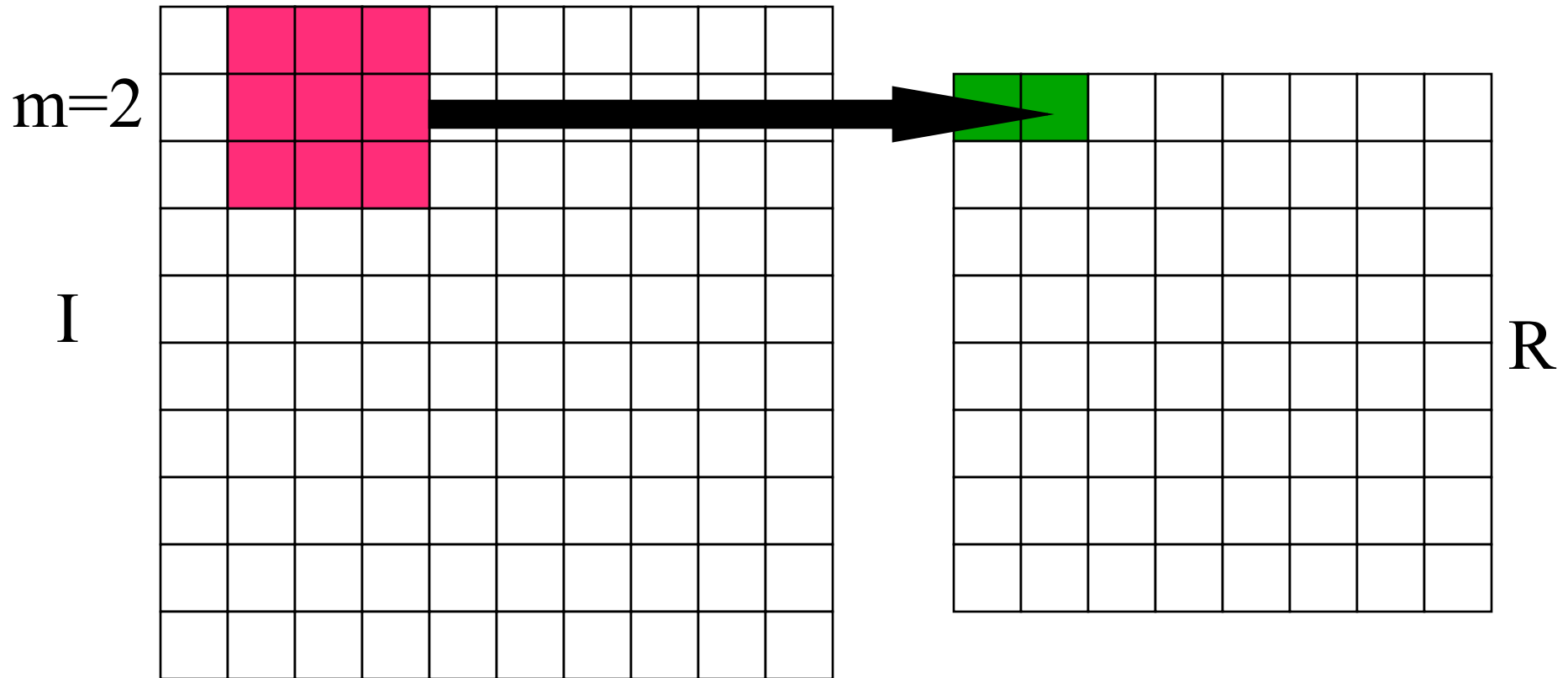
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

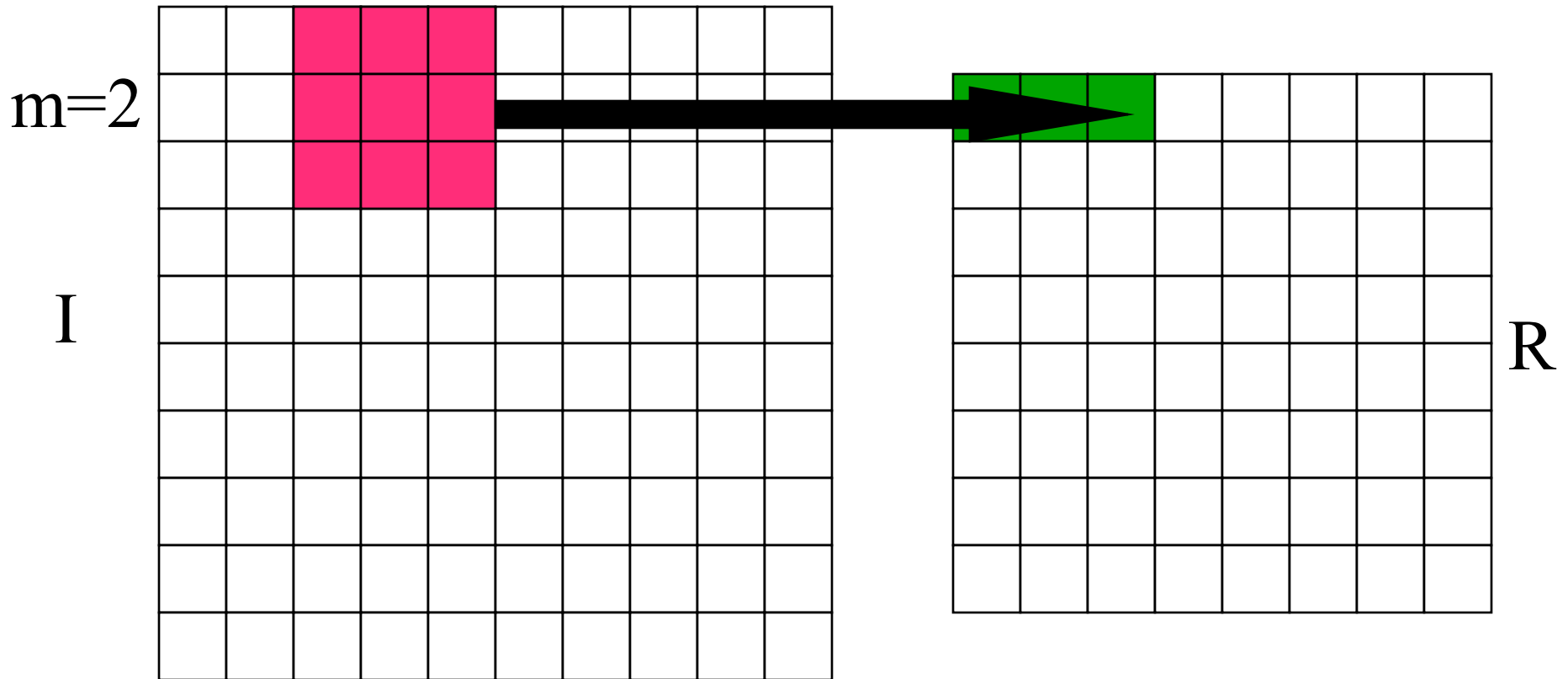
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

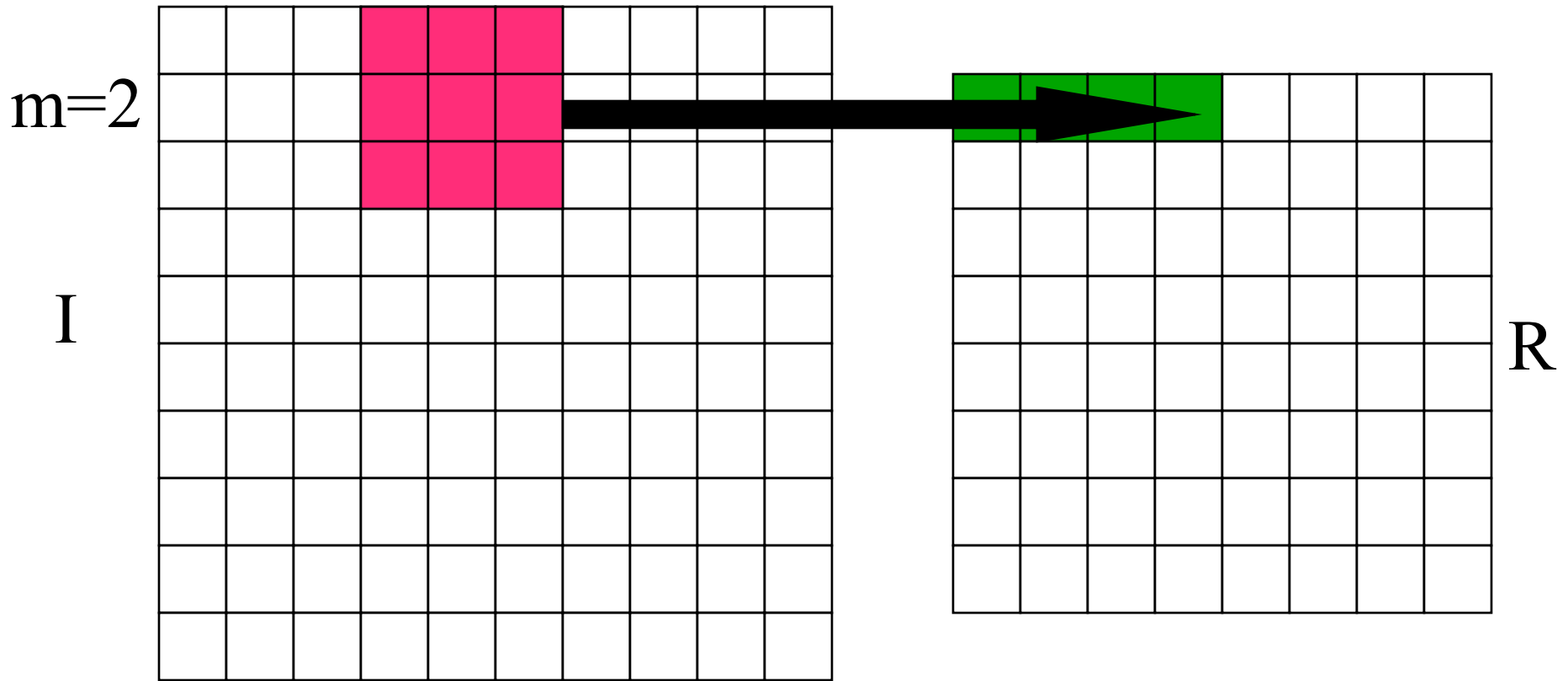
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

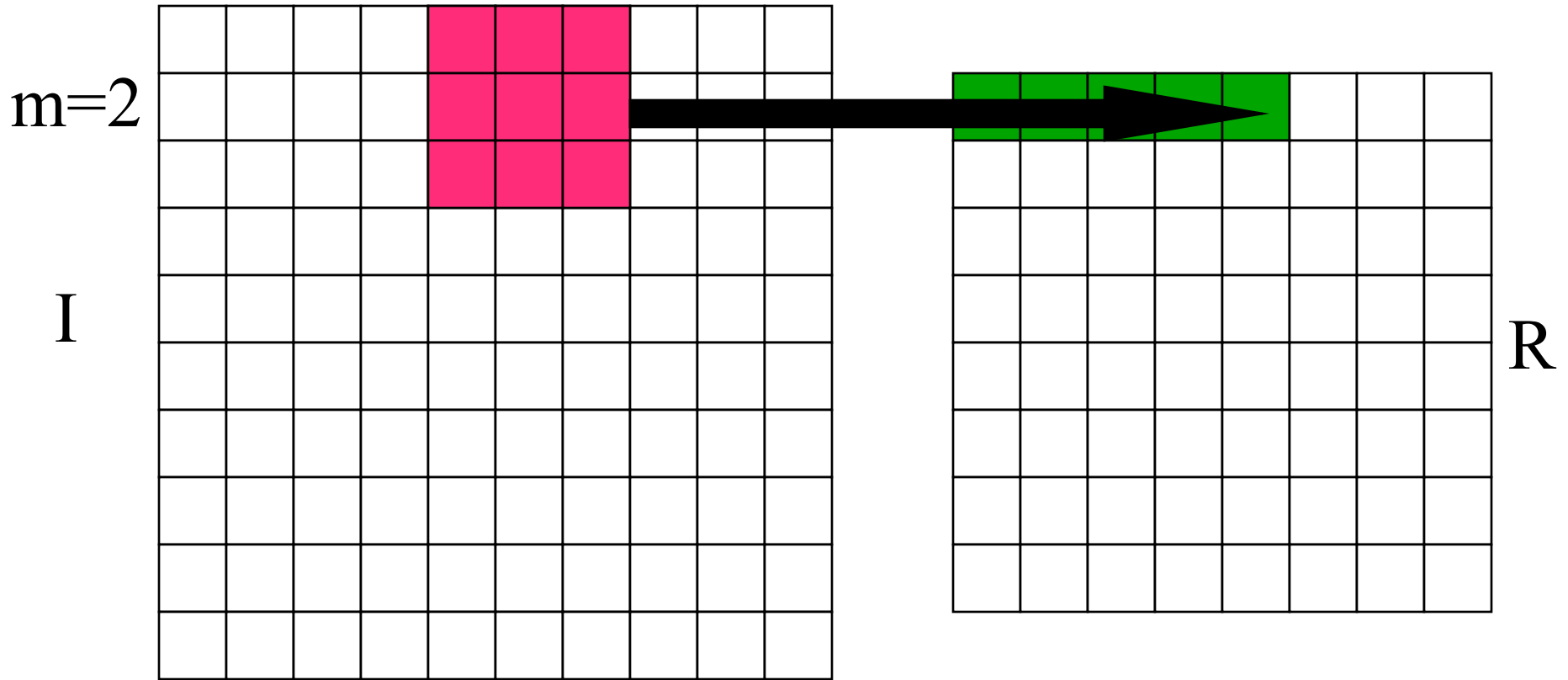
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

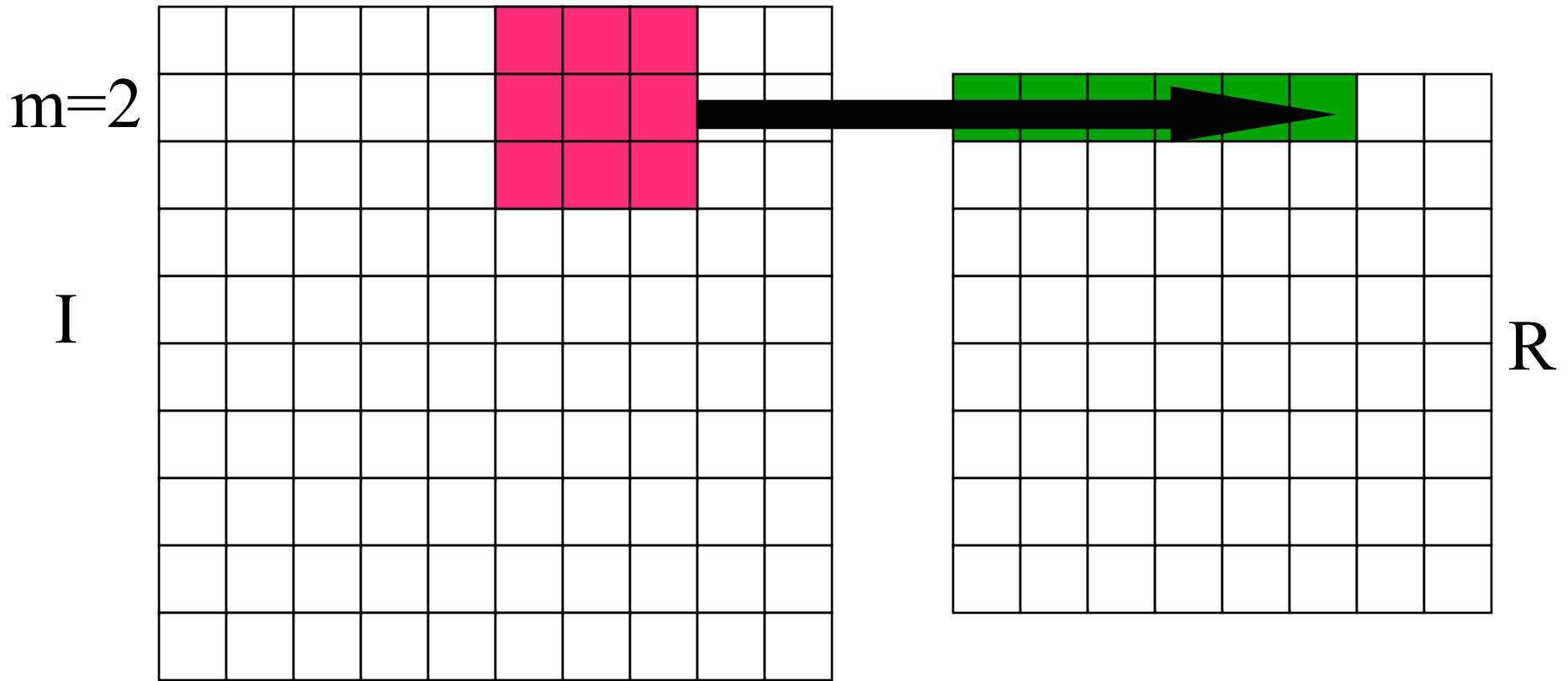
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

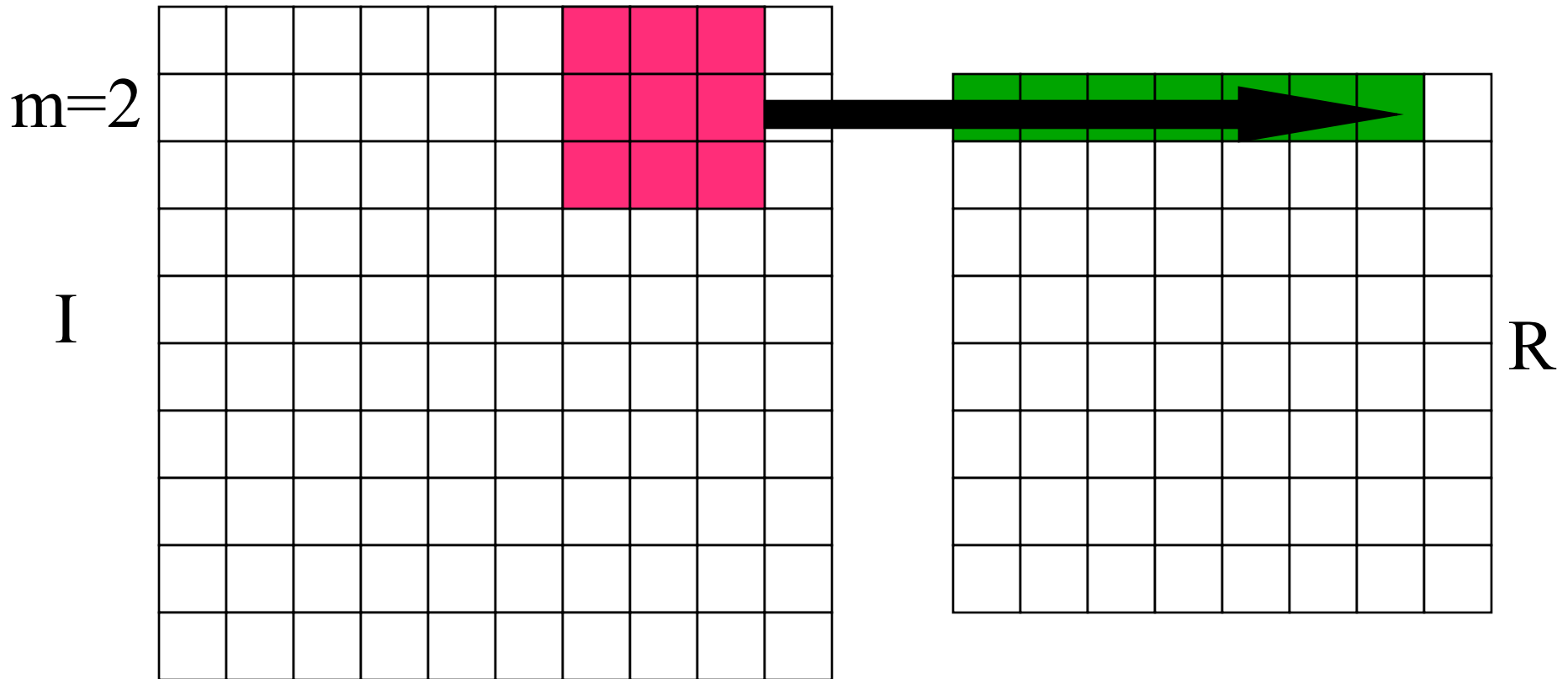
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

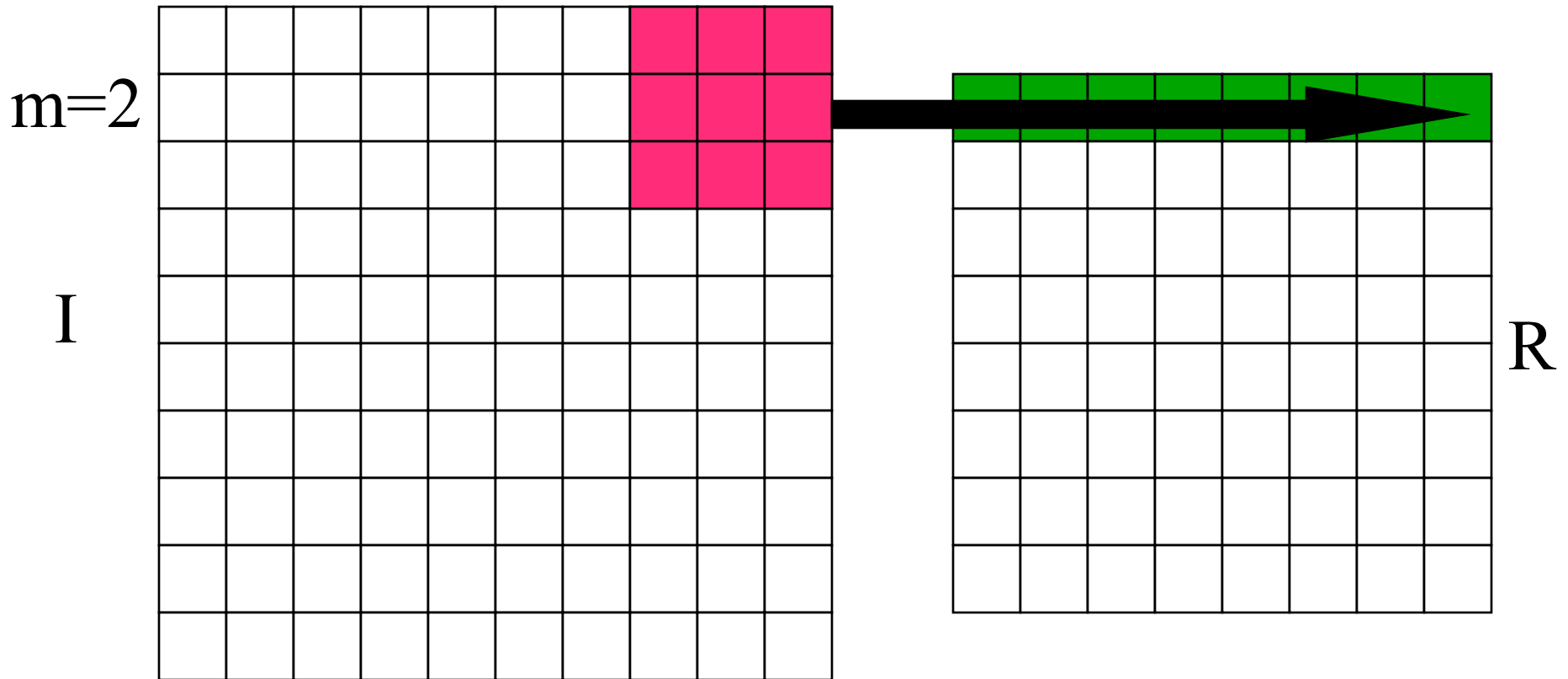
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

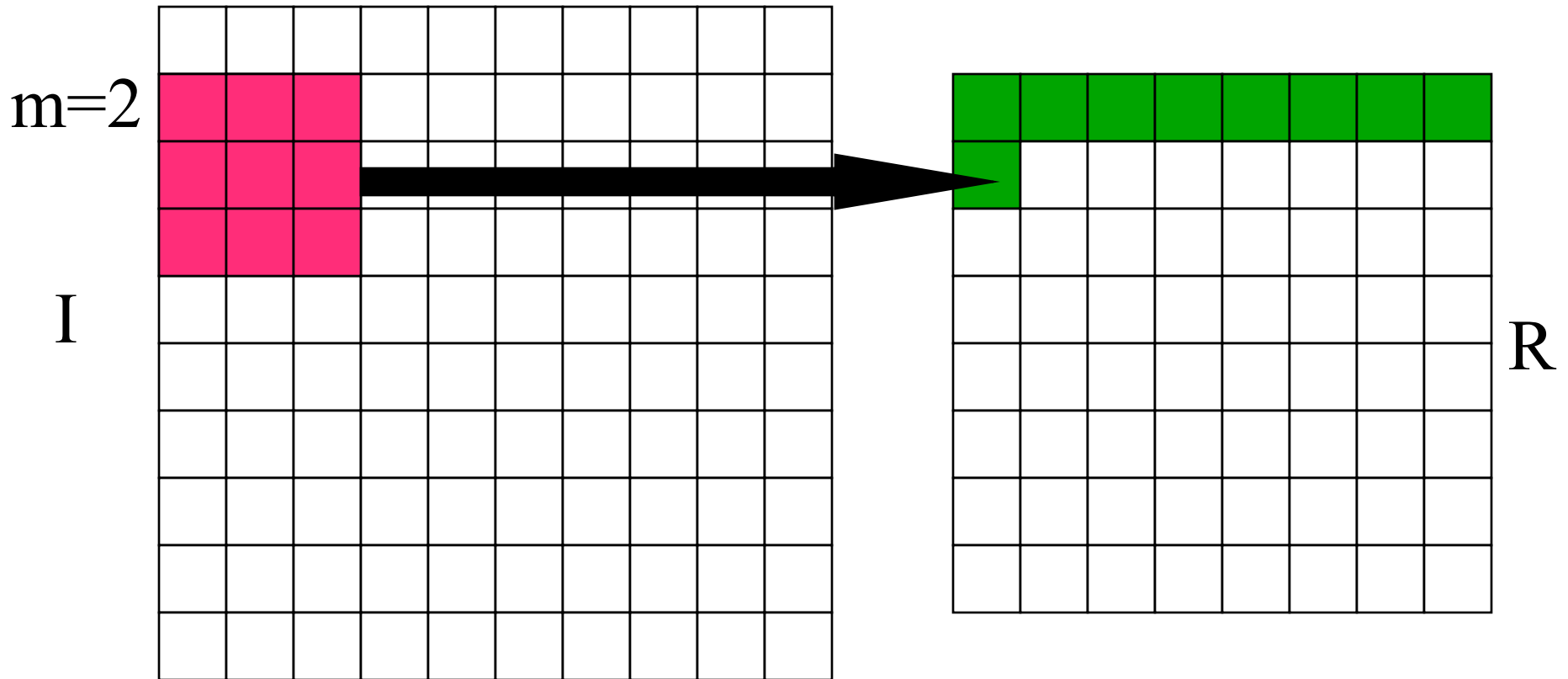
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

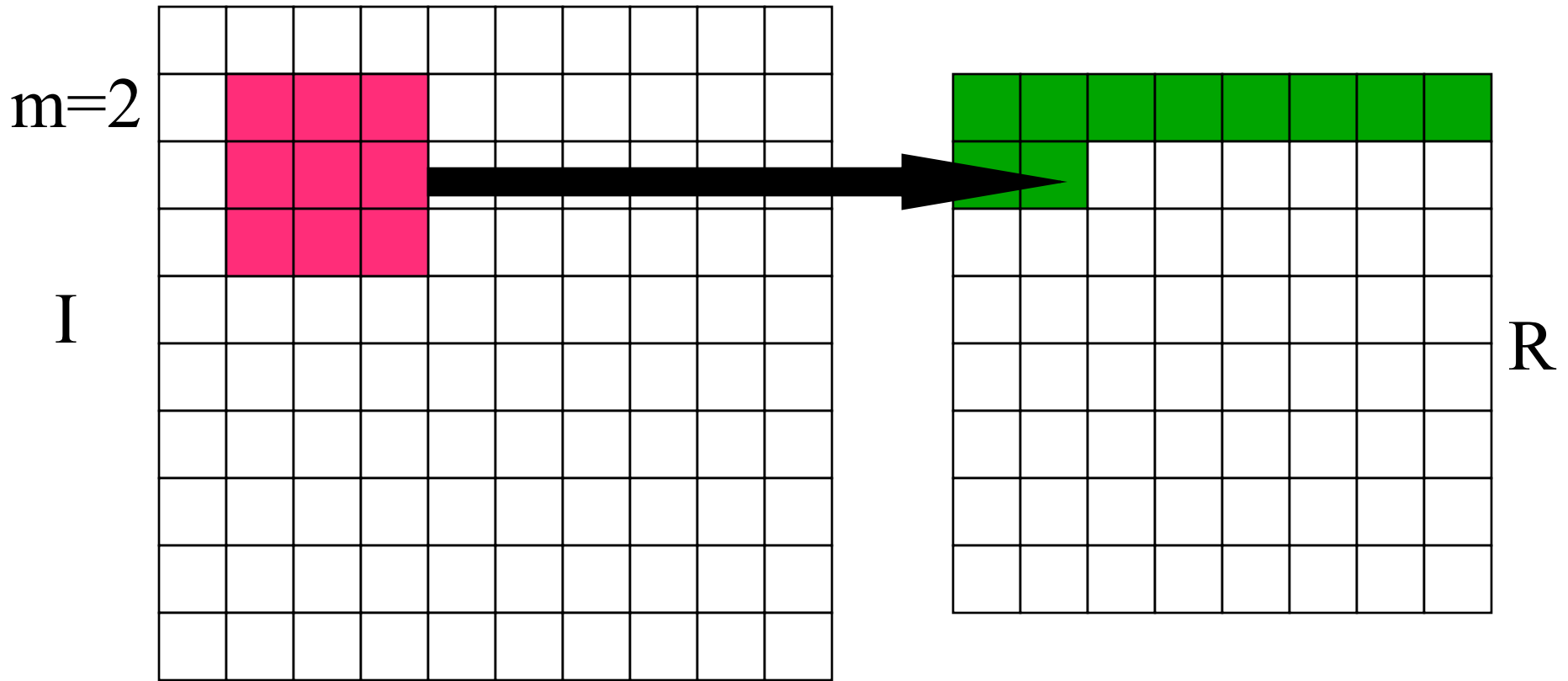
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

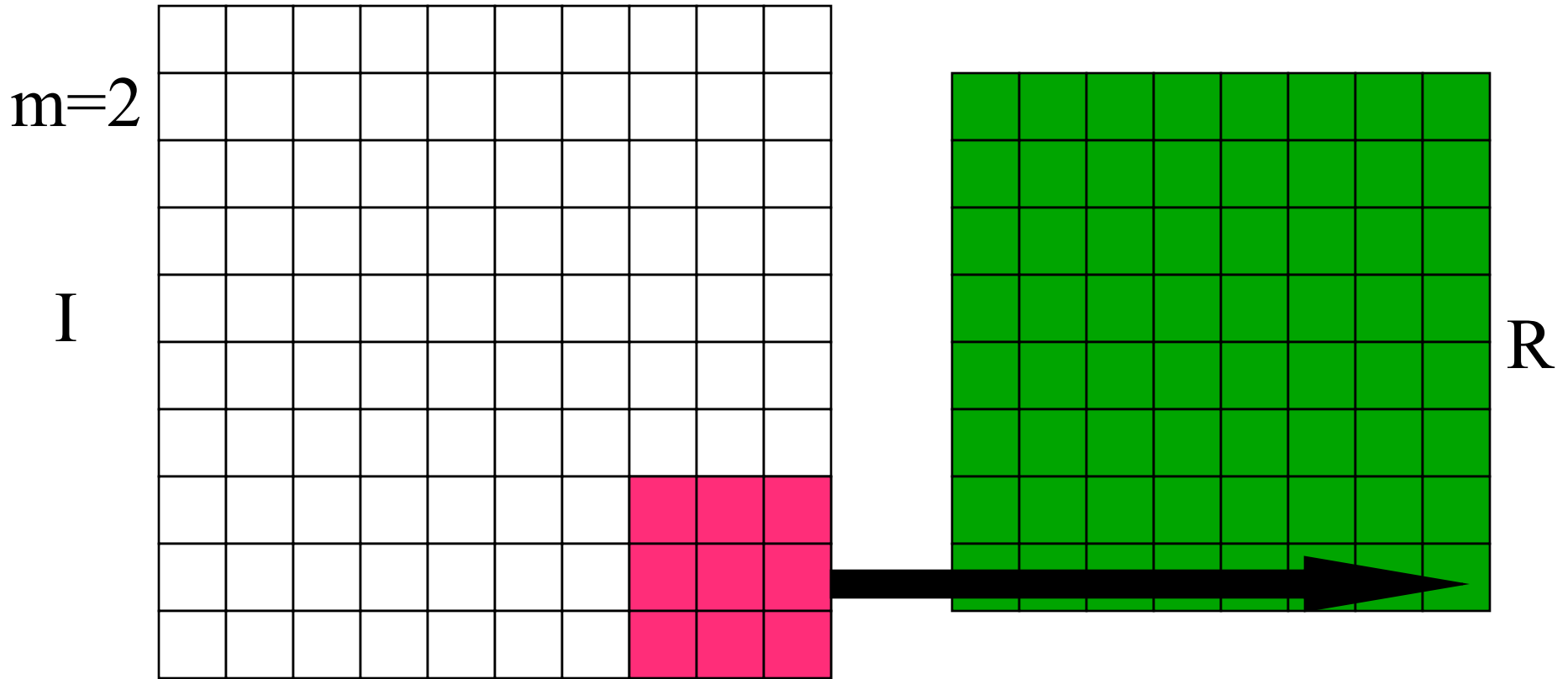
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

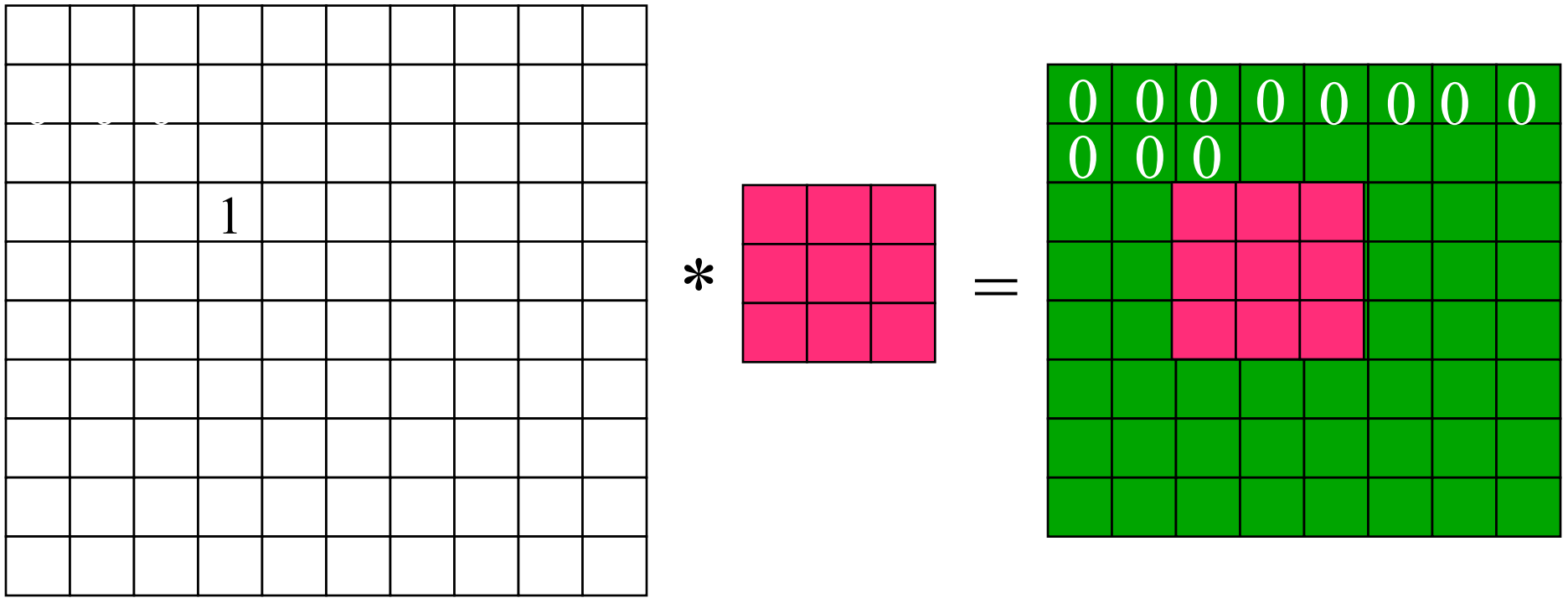
Convolution: $R = K * I$



Kernel size
is $m+1$ by $m+1$

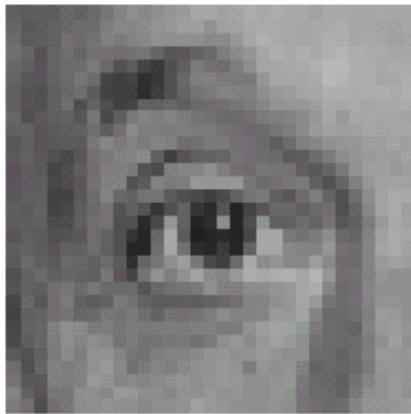
$$R(i, j) = \sum_{h=-m/2}^{m/2} \sum_{k=-m/2}^{m/2} K(h, k) I(i-h, j-k)$$

Impulse Response

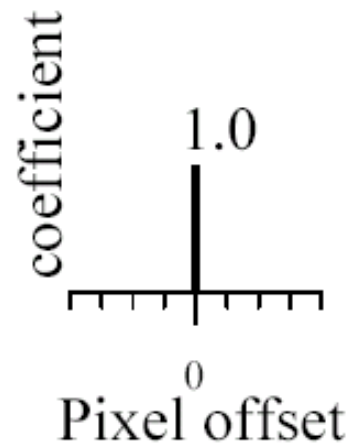


Note that the “self-replicating” property of the impulse response is what require the “flip” in the template when it is applied

Linear filtering (warm-up slide)

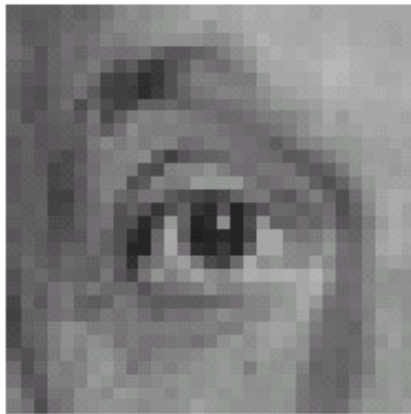


original

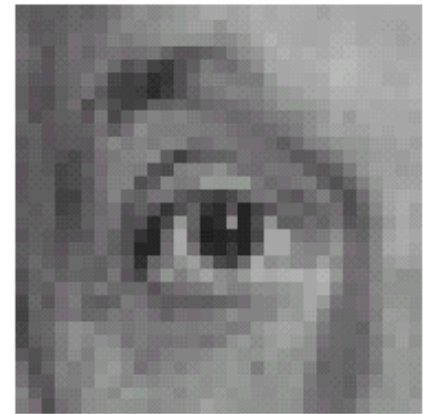
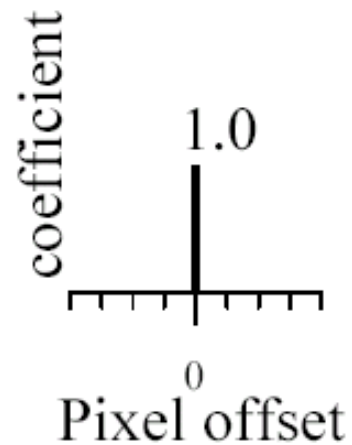


?

Linear filtering (warm-up slide)

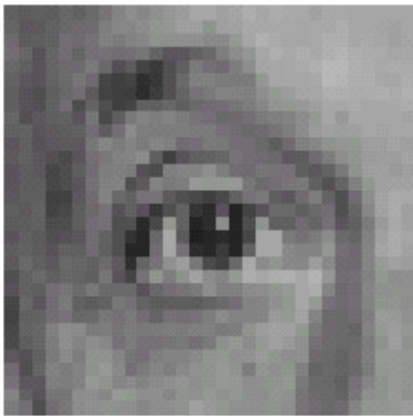


original

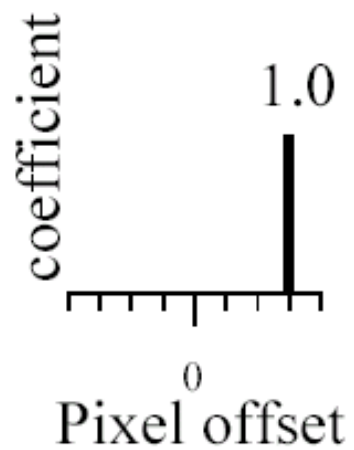


Filtered
(no change)

Linear filtering

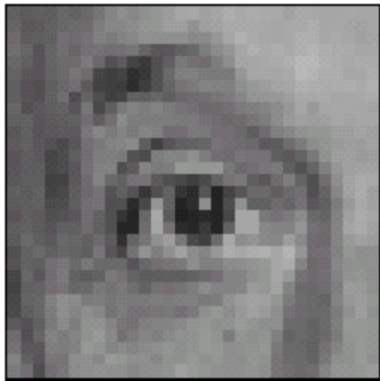


original

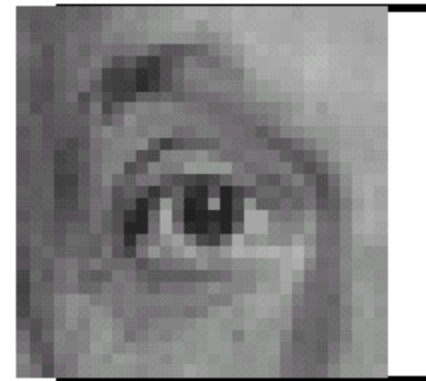
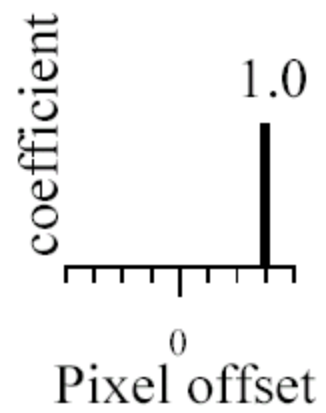


?

shift

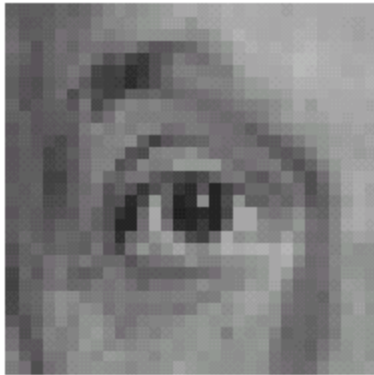


original

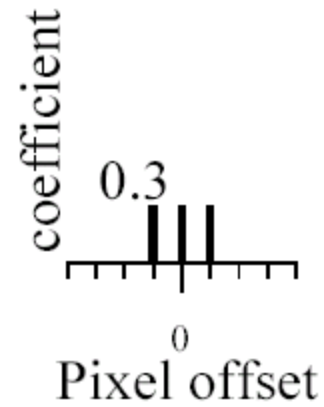


shifted

Linear filtering

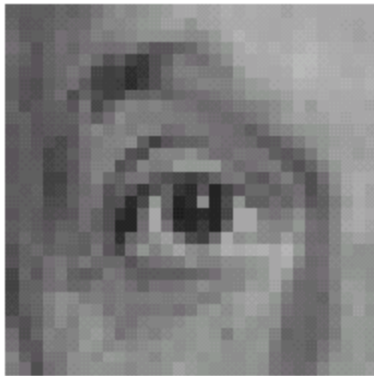


original

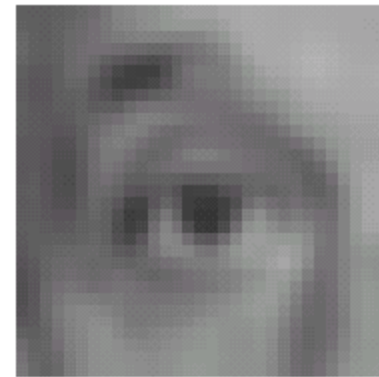
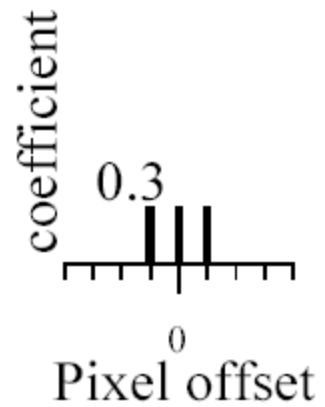


?

Blurring

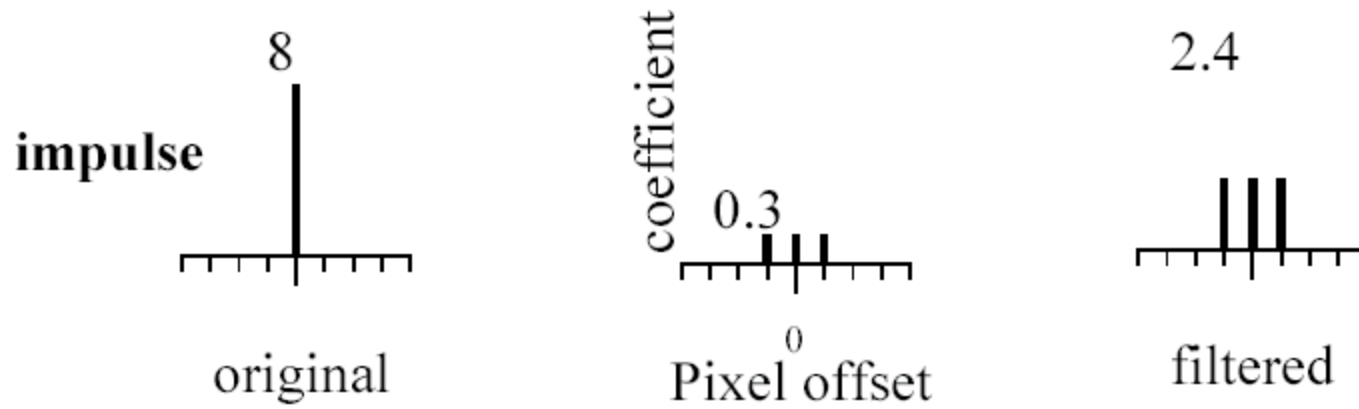


original

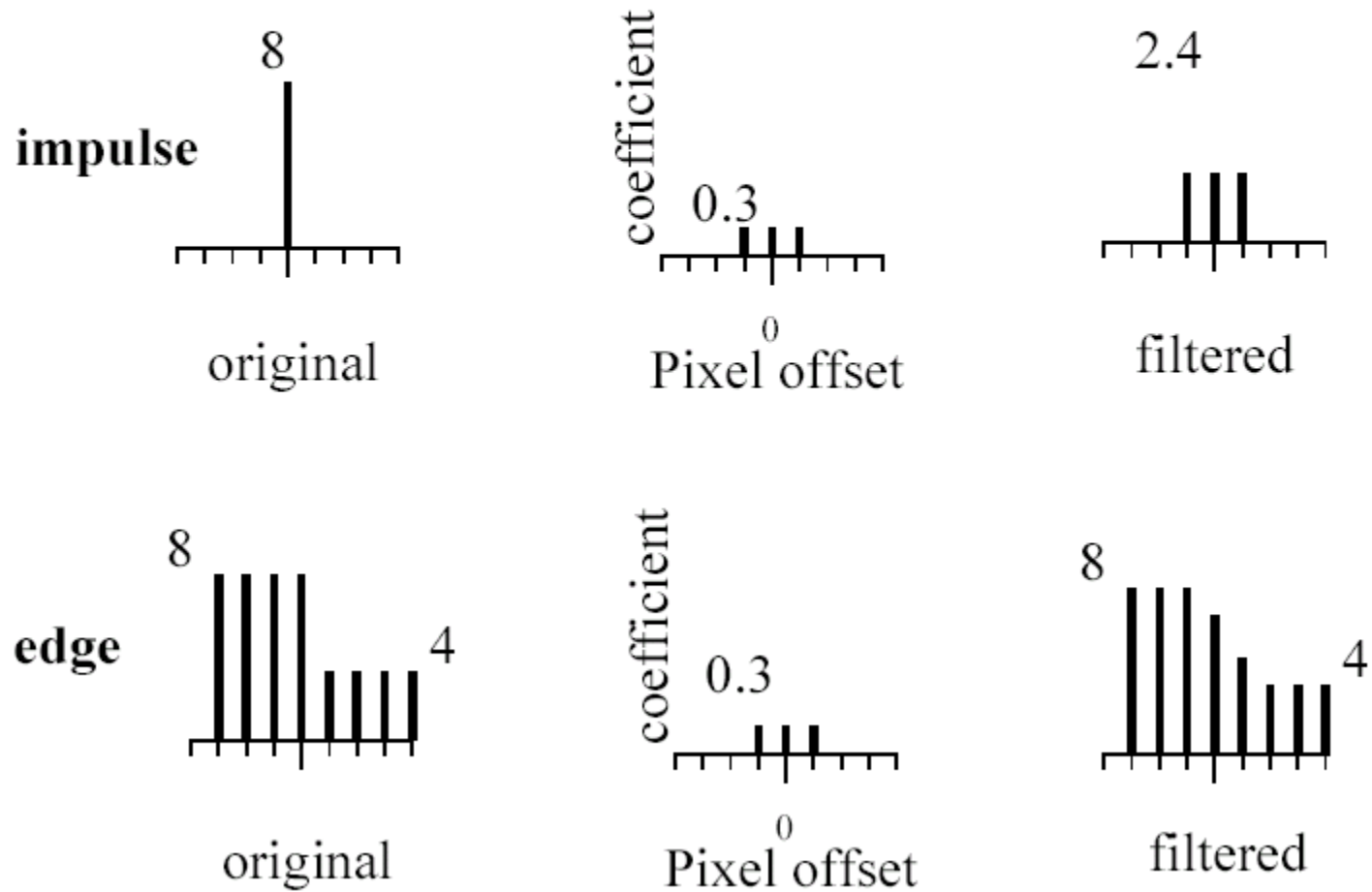


Blurred (filter applied in both dimensions).

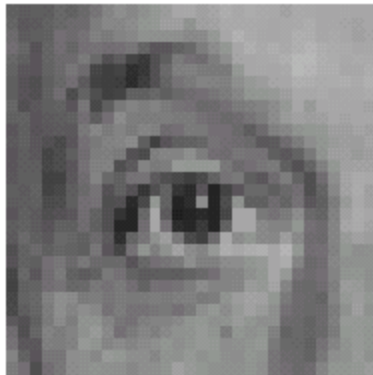
Blur examples



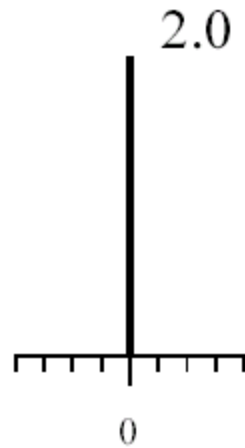
Blur examples



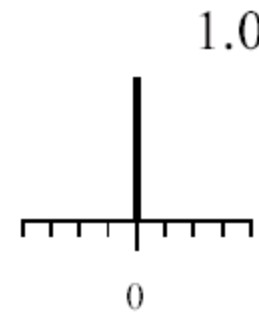
Linear filtering (warm-up slide)



original

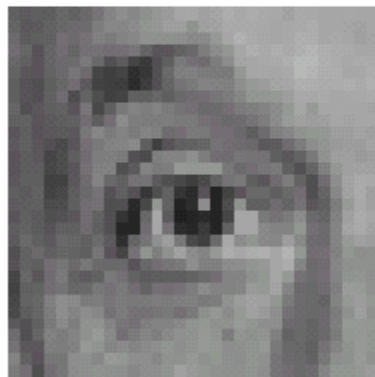


—

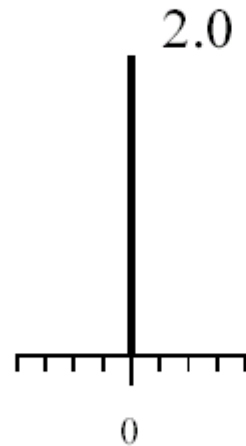


?

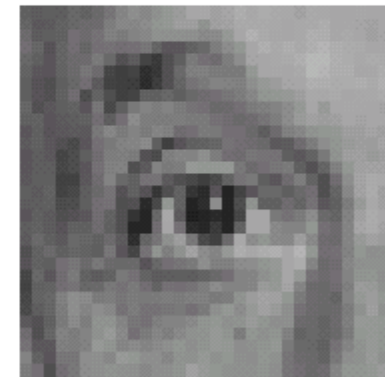
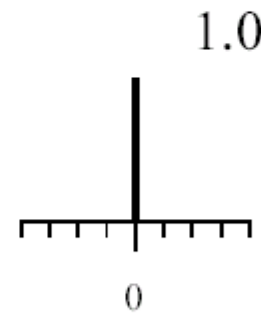
Linear filtering (no change)



original

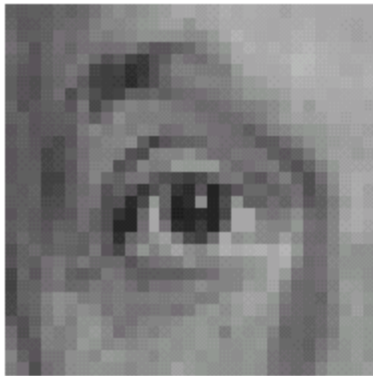


—

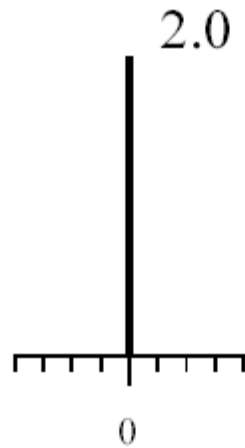


Filtered
(no change)

Linear filtering



original

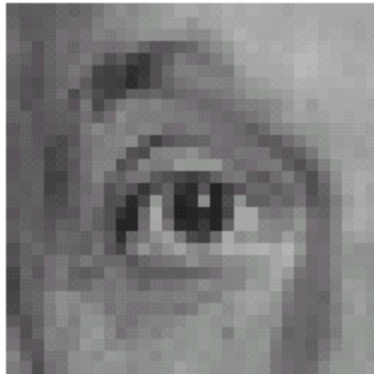


—

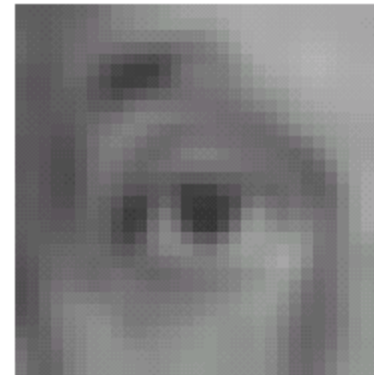
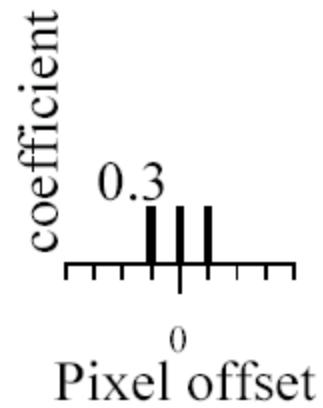


?

(remember blurring)

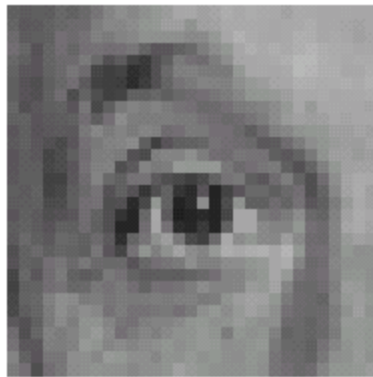


original

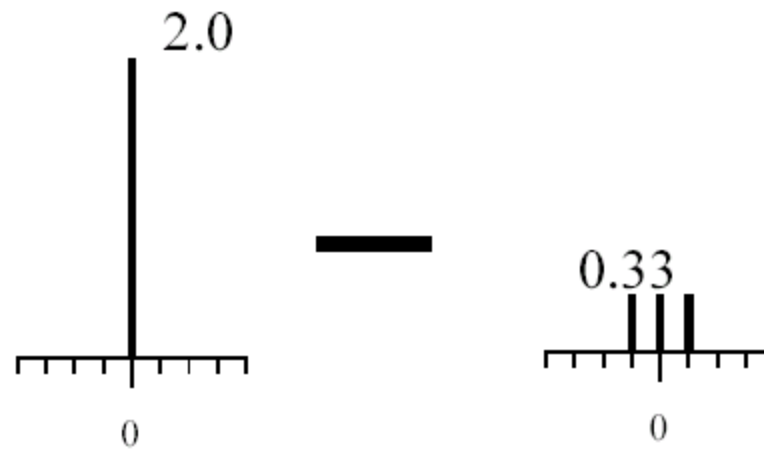


Blurred (filter applied in both dimensions).

Sharpening

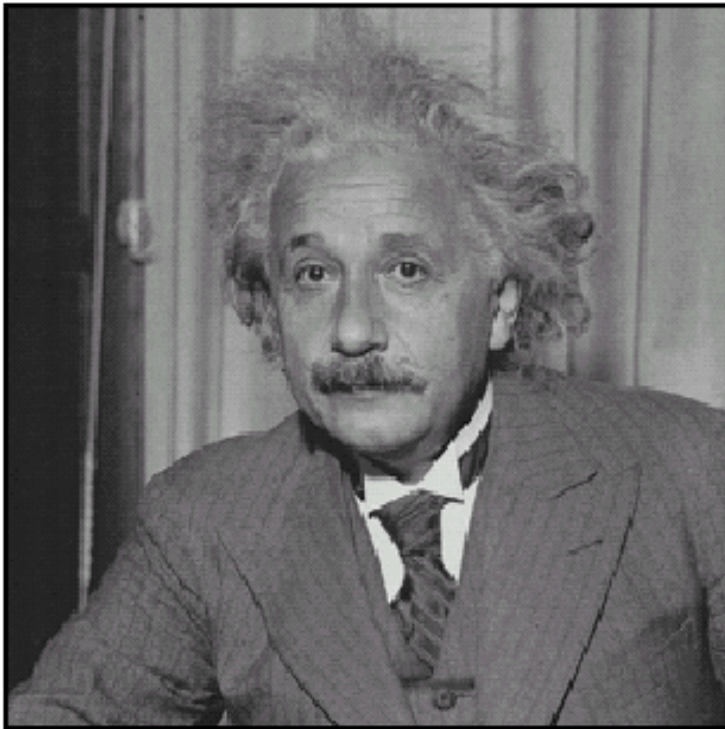


original

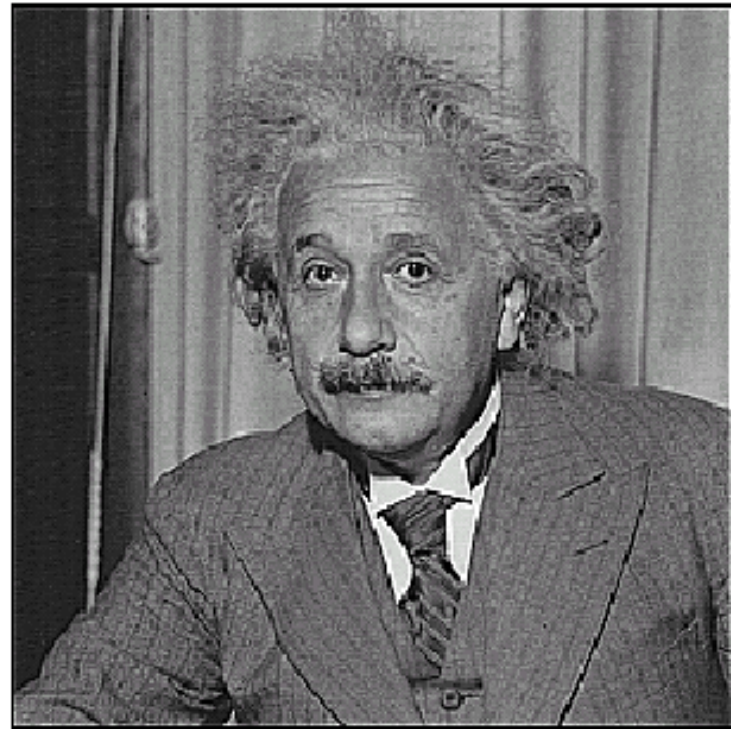


Sharpened
original

Sharpening



before



after

9/17/08

How to Reduce Noise

- For a pixel in image I at I,j

$$I'(i, j) = 1/9 \sum_{i'=i-1}^{i+1} \sum_{j'=j-1}^{j+1} I(i', j')$$

- Computing this for every pixel location is the *convolution* of the image I with the *template* (or *kernel*) consisting of a 3x3 array of 1's.
- Note that is this $O(n^2m^2)$ for an nxn image and mxm template.
- Note we have to normalize the template to 1 to make sure we don't introduce any scaling into the image.

Smoothing by Averaging

Kernel: 



9/17/08

CS 461, Copyright © D. K. Hager
Downloaded from <http://www.cse.cmu.edu/~hager/>

Some Convolution Facts

- Convolution is
 - Associative
 - Commutative
 - A linear operator
- With a fixed kernel, convolution is also *shift invariant*, meaning if we translate the signal, we translate (but do not otherwise change) the response.
 - Linearity plus shift invariance can be used to derive convolution!
- We are using a *discrete convolution*; we will see this is not always consistent with an underlying *continuous convolution* that we may wish to implement
- Convolution is formally defined on unbounded images and kernels.
 - padding schemes:
 - pad with zeros (same size vs. full size)
 - compute only legal values



time for a Matlab
demo!

Another View of Convolution

- Suppose we consider the convolution template as a “vector”:
 - $T = [T_1, T_2, T_3 \dots T_n]$
- Likewise, consider a region of the image to which the convolution is applied as a vector
 - $I = [I_1, I_2, \dots, I_n]$
- Then the value of the convolution at a point is just the “dot product”
 - $v = \text{flip}(T) \cdot I$
- Thus, we can also think of convolution as a kind of “pattern match” where regions of the image that are “similar” to T respond more strongly than those that are dissimilar (up to a scale factor)
- We’ll see this makes sense when thinking of Fourier transforms...

Understanding Convolution

- Another way to think about convolution is in terms of how it changes the *frequency distribution* in the image.
- Recall the *fourier* representation of a function
 - $F(u) = \int f(x) e^{-2\pi i u x} dx$
 - recall that $e^{-2\pi i u x} = \cos(2\pi u x) - i \sin(2\pi u x)$
 - Also we have $f(x) = \int F(u) e^{2\pi i u x} du$
 - $F(u) = |F(u)| e^{i \angle F(u)}$
 - a decomposition into magnitude ($|F(u)|$) and phase $\angle F(u)$
 - If $F(u) = a + i b$ then
 - $|F(u)| = (a^2 + b^2)^{1/2}$ and $\angle F(u) = \text{atan2}(a,b)$
 - $|F(u)|^2$ is the *power spectrum*
- Questions: what function takes many many many terms in the Fourier expansion?

Understanding Convolution

Discrete Fourier Transform (DFT)

$$F[u, v] \equiv \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I[x, y] e^{-\frac{2\pi}{N} j (xu+yv)}$$

Inverse DFT

$$I[x, y] \equiv \frac{1}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F[u, v] e^{\frac{+2\pi}{N} j (ux+vy)}$$

Implemented via the “Fast Fourier Transform” algorithm (FFT)

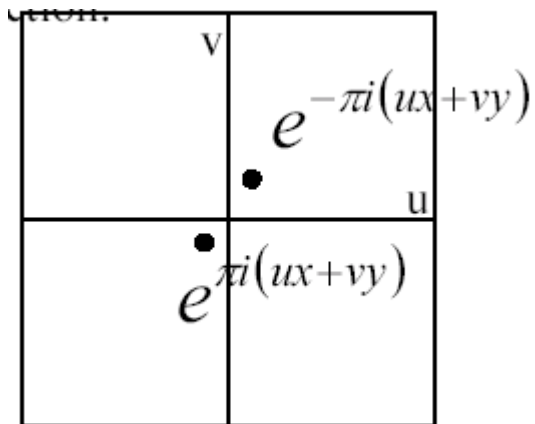
Fourier basis element

$$e^{-i2\pi(ux+vy)}$$

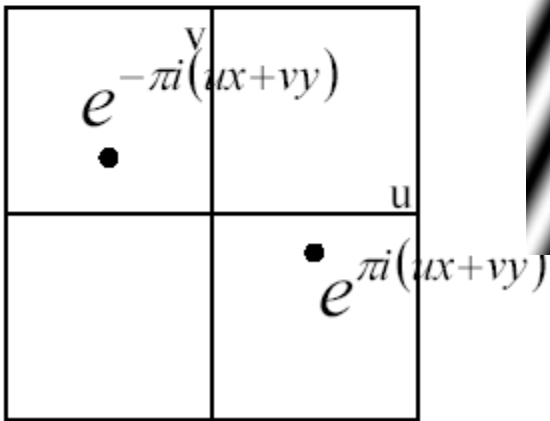
Transform is sum of orthogonal basis functions

Vector (u,v)

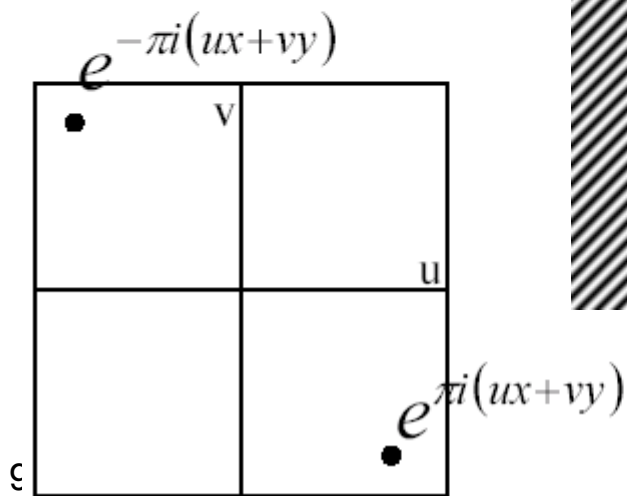
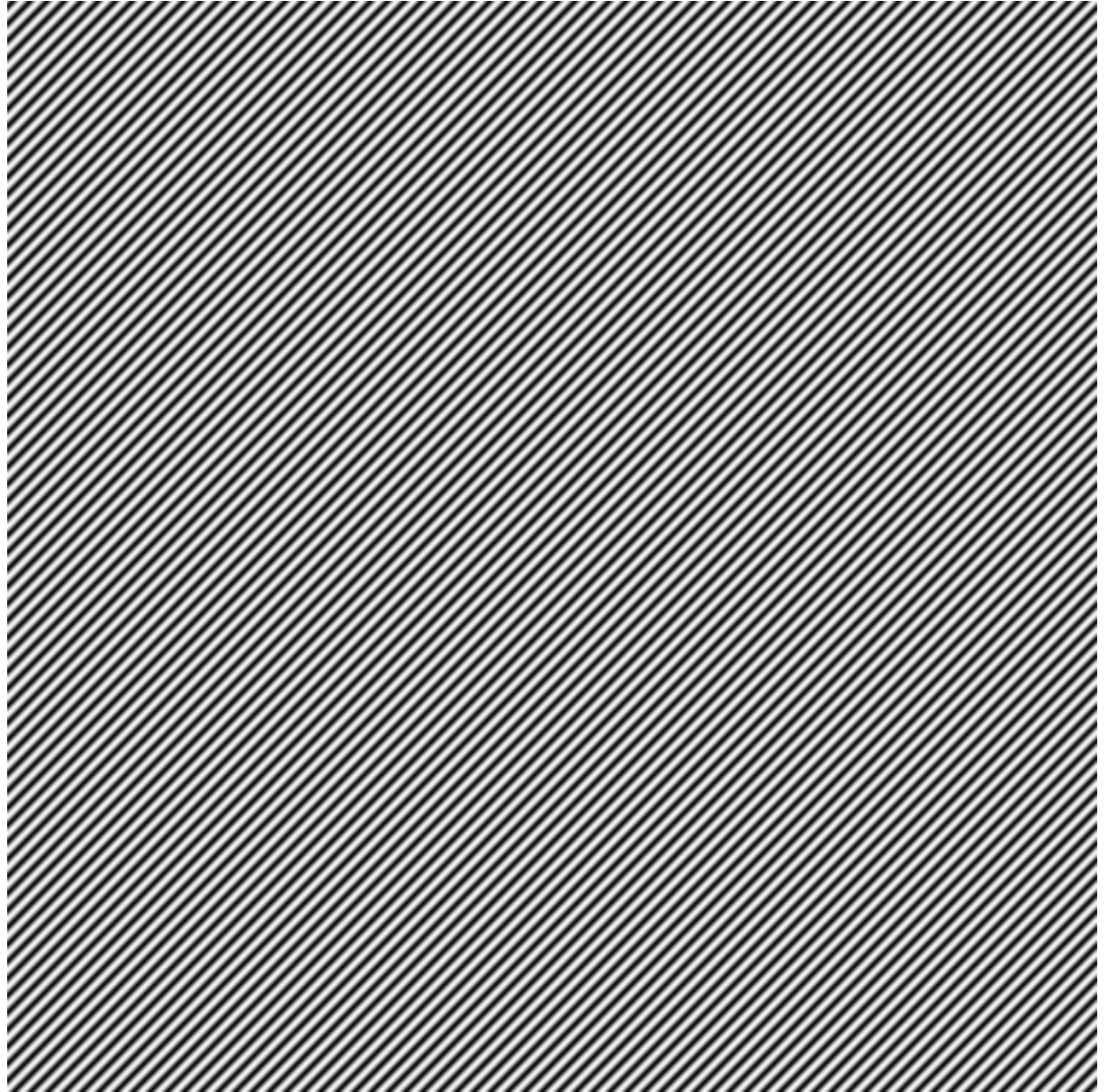
- Magnitude gives frequency
- Direction gives orientation.



Here u and v are larger than in the previous slide.

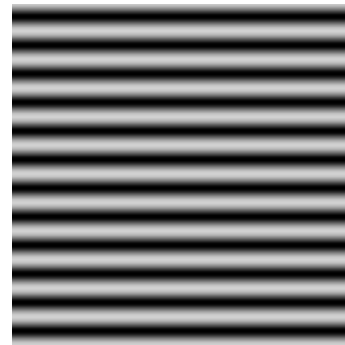
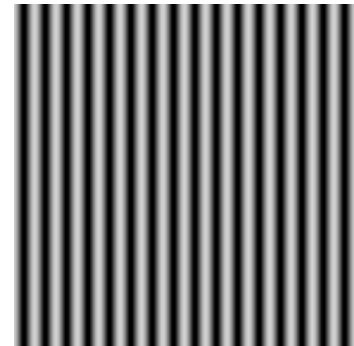


And larger still...

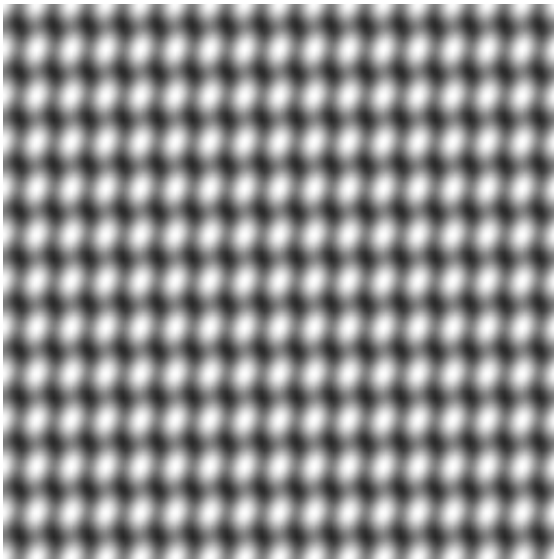


The Fourier “Hammer”

“Power Spectrum”

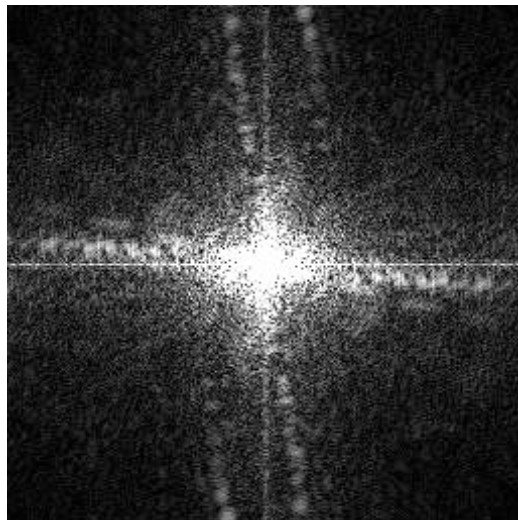


Linear Combination:

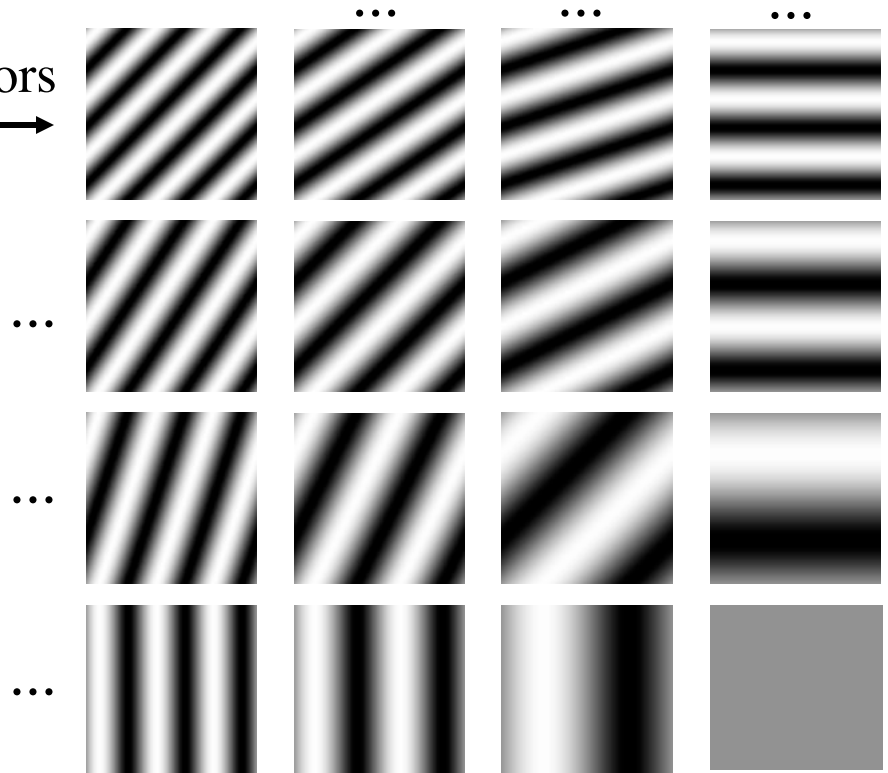


Basis vectors

Frequency Decomposition



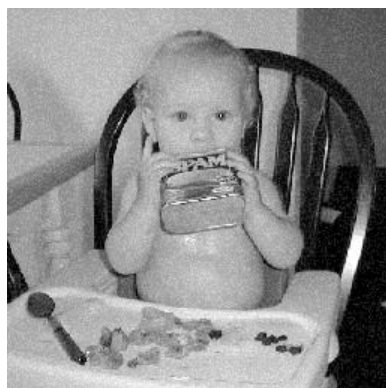
All Basis Vectors
→
←
Example



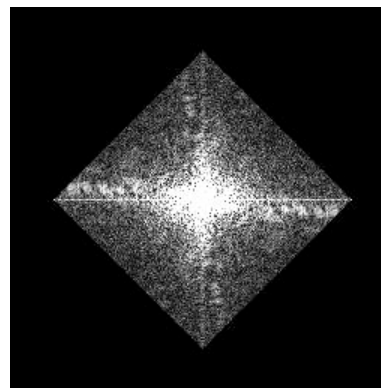
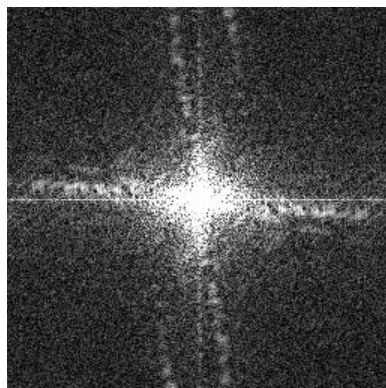
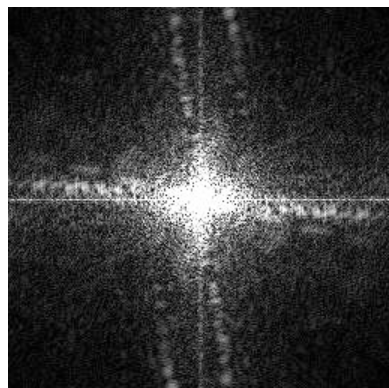
intensity \sim that frequency's coefficient



Using Fourier Representations



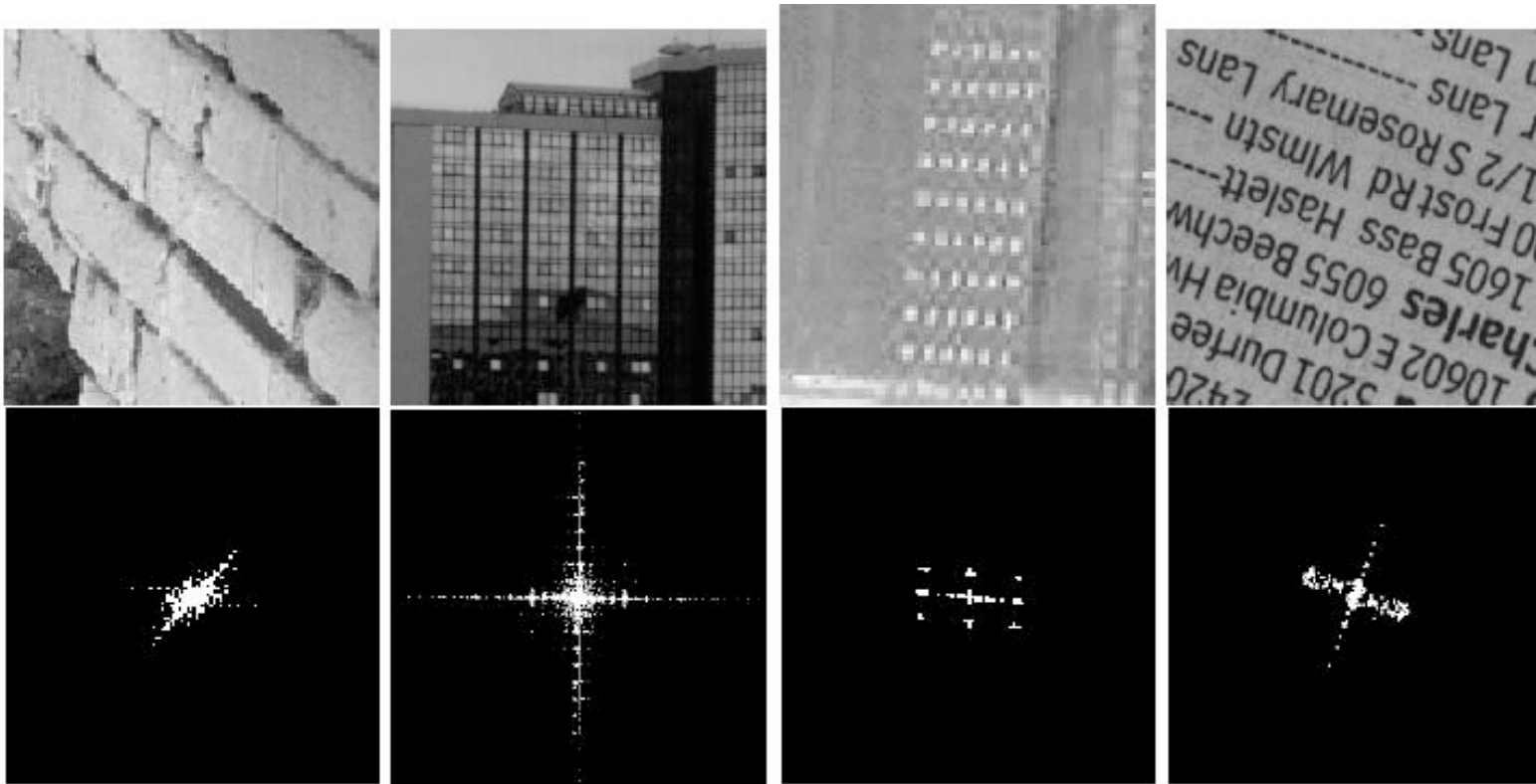
Smoothing



Data Reduction: only use *some* of the existing frequencies

Using Fourier Representations

Dominant Orientation



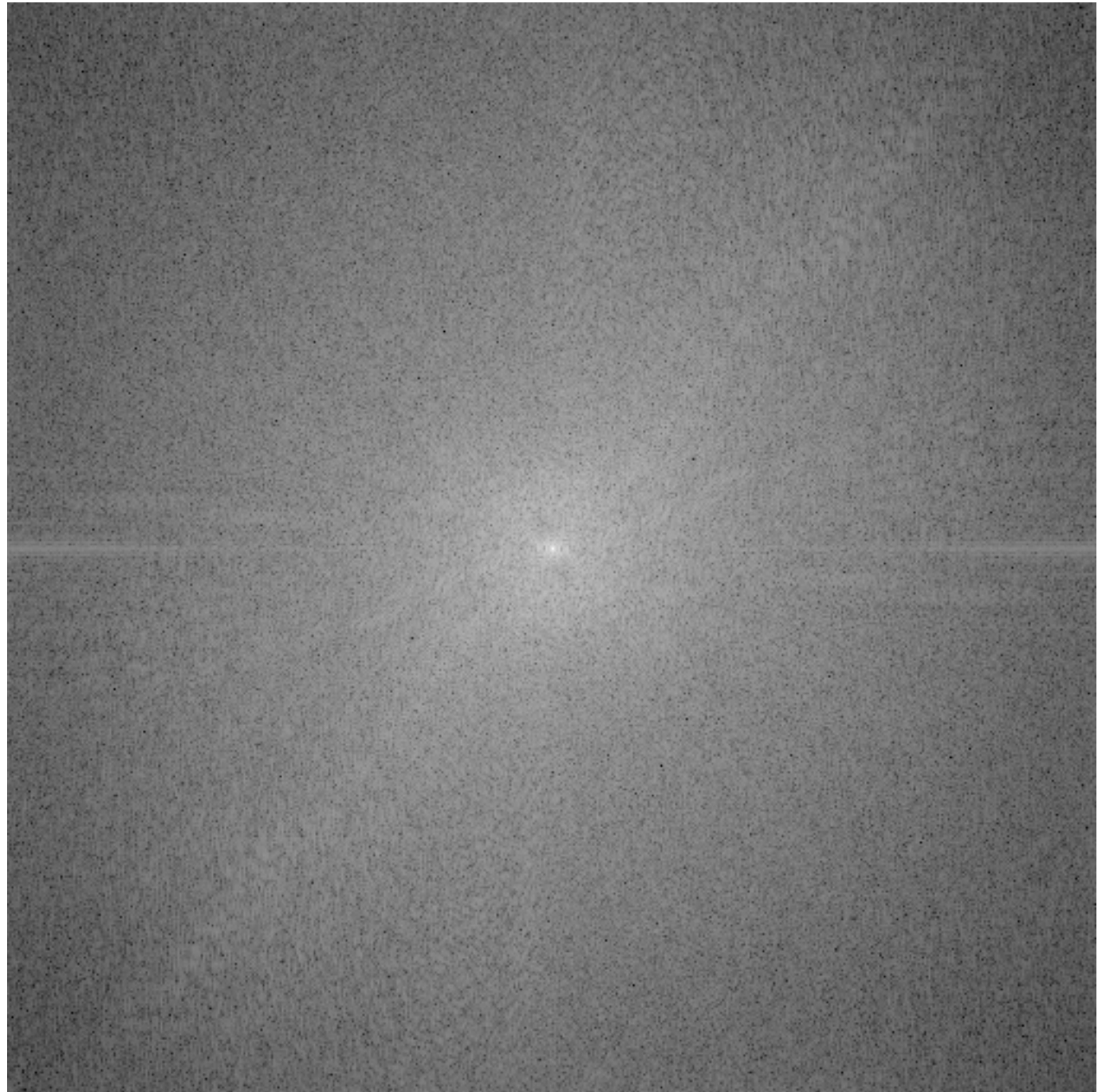
Limitations: not useful for local segmentation



9/17/08

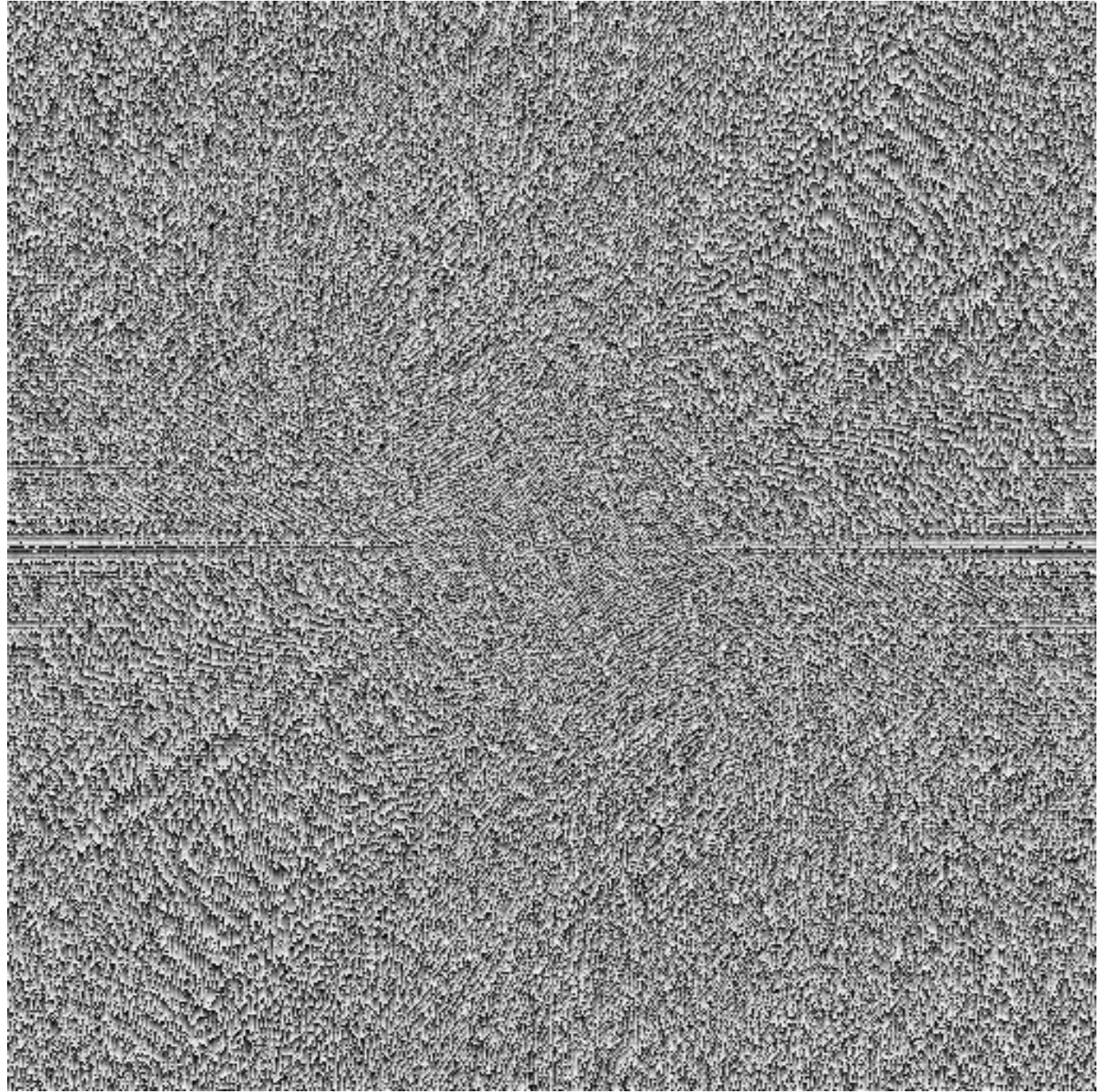
CS 461 Copyright © D. K. Hager
Sponsored by Intel

This is the
magnitude
transform
of the
cheetah
pic



9/17/08

This is the
phase
transform
of the
cheetah
pic



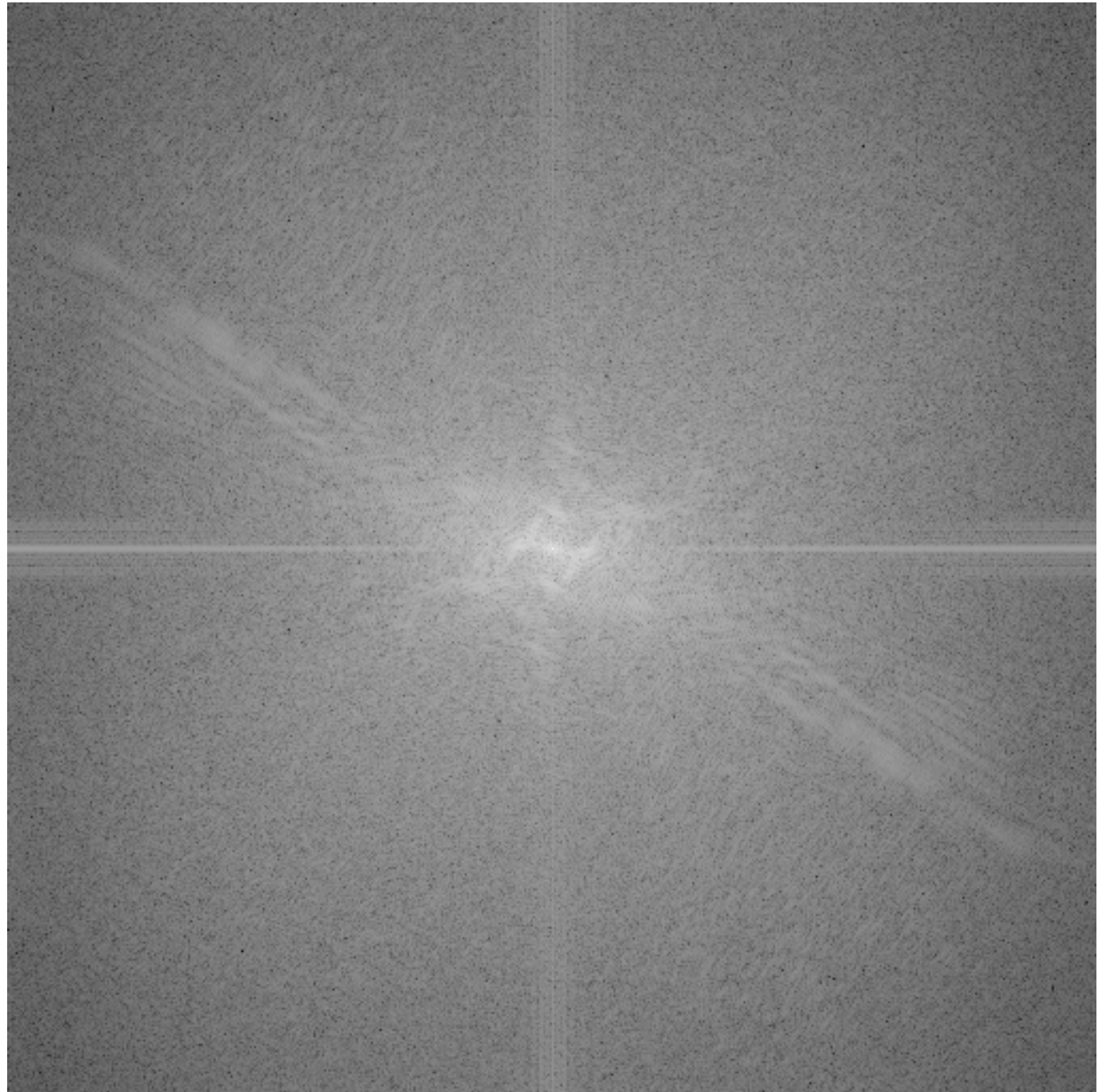
9/17/08



9/17/08

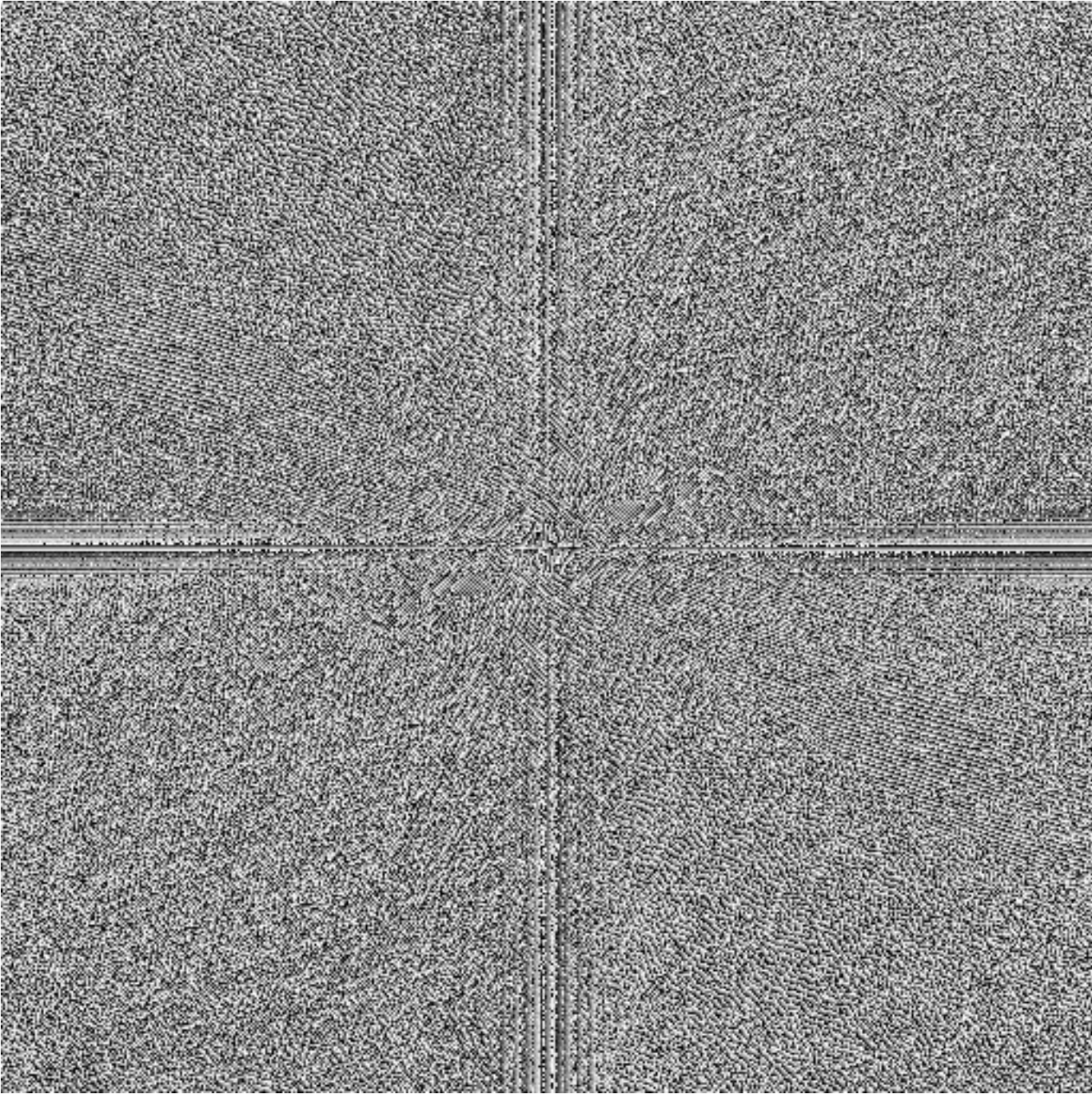
CS 461 Copyright © D. K. Hager
Downloaded by D. K. Hager

This is the
magnitude
transform
of the
zebra pic



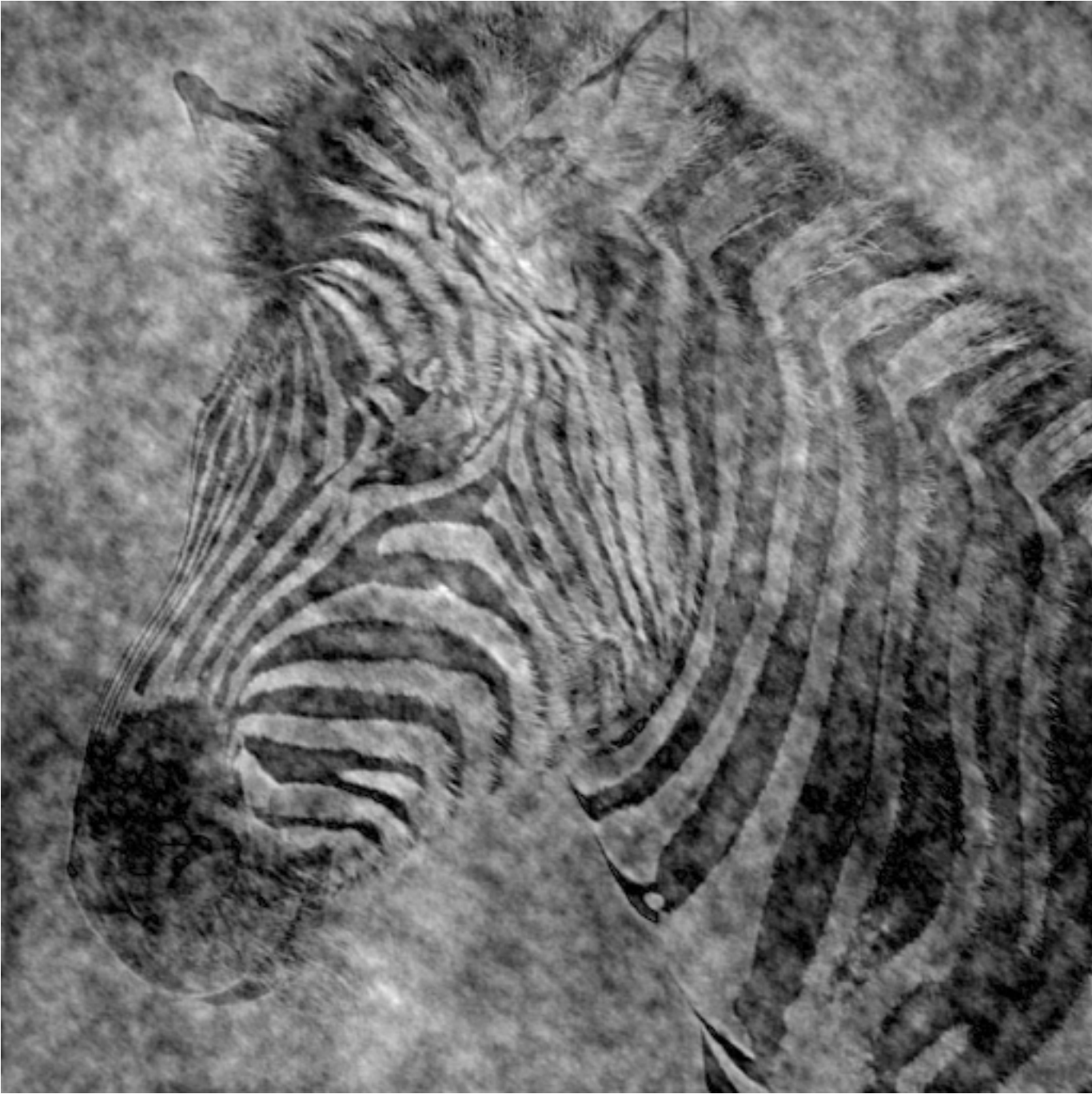
9/17/08

This is the
phase
transform
of the
zebra pic



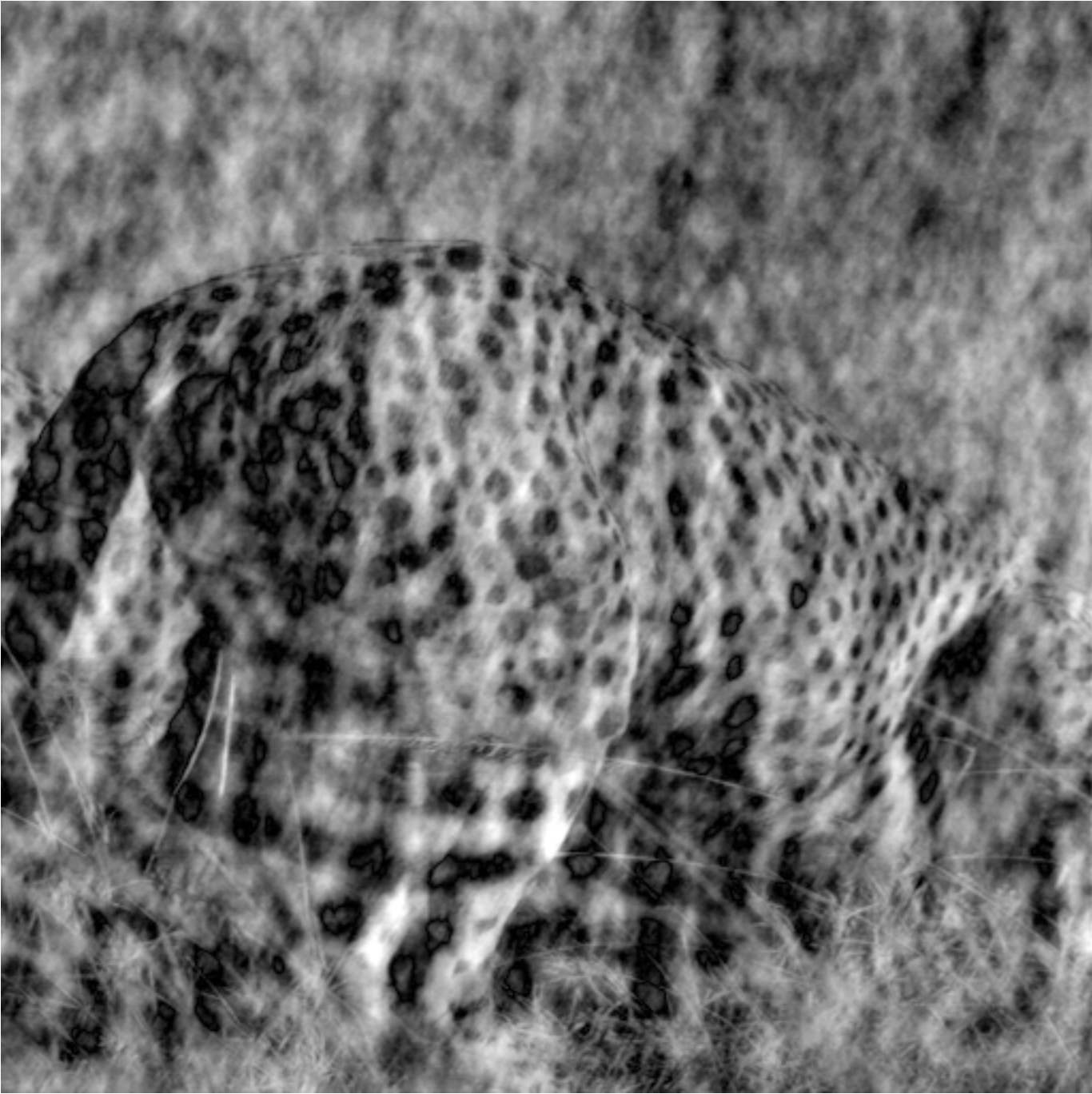
9/17/08

Reconstruction
with zebra
phase,
cheetah
magnitude



9/17/08

Reconstruction
with cheetah
phase, zebra
magnitude



9/17/08

CS 461 Copyright © D. K. Hager

The Fourier Transform and Convolution

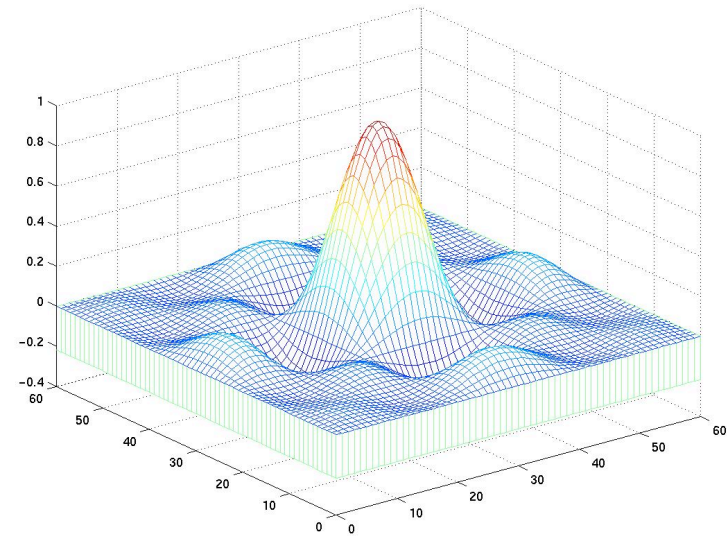
- If H and G are images, and $F(\cdot)$ represents Fourier transform, then

$$F(H * G) = F(H)F(G)$$

- Thus, one way of thinking about the properties of a convolution is by thinking of how it modifies the frequencies of the image to which it is applied.
- In particular, if we look at the power spectrum, then we see that convolving image H by G attenuates frequencies where G has low power, and amplifies those which have high power.
- This is referred to as the **Convolution Theorem**

The Properties of the Box Filter

F(mean filter) =



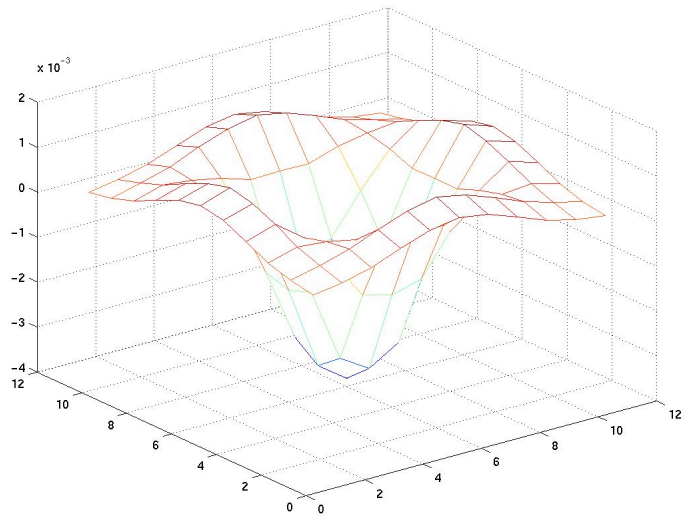
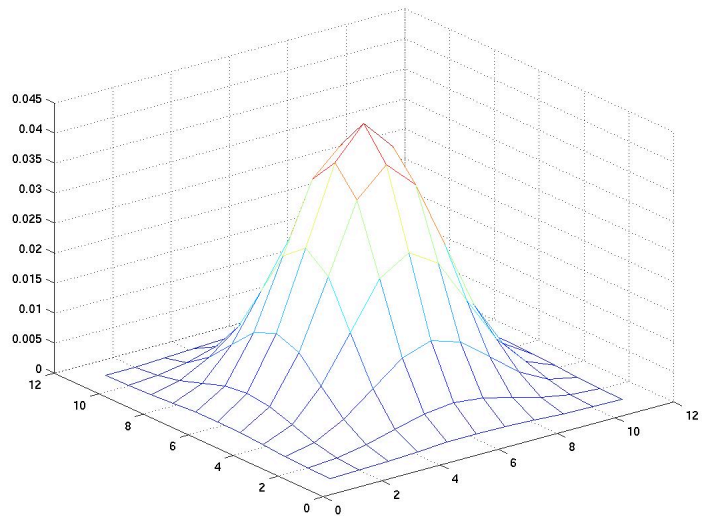
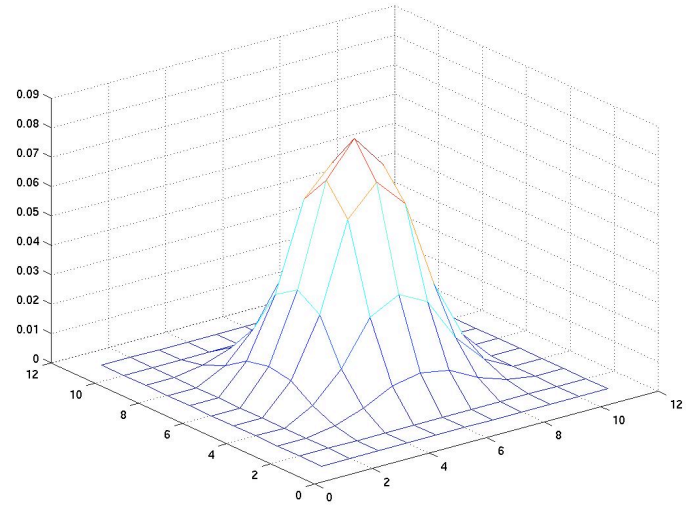
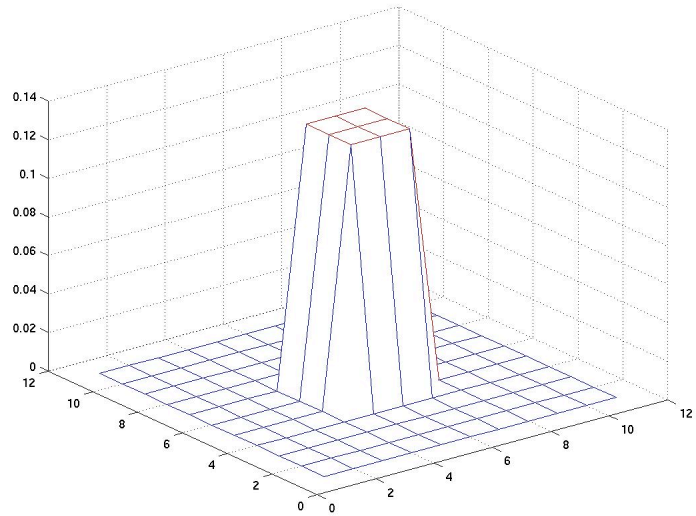
Thus, the mean filter enhances low frequencies but also has “side lobes” that admit higher frequencies

The Gaussian Filter: A Better Noise Reducer

- Ideally, we would like an averaging filter that removes (or at least attenuates) high frequencies beyond a given range

$$g(i, j; \sigma) = e^{-\frac{(i^2 + j^2)}{2\sigma^2}}$$

- It is not hard to show that the FT of a Gaussian is again a Gaussian.
 - What does this imply?
- Note that in general, we truncate --- a good general rule is that the width (w) of the filter is at least such that $w > 5 s$. Alternatively we can just stipulate that the width of the filter determines s (or vice-versa).
- Note that in the discrete domain, we truncate the Gaussian, thus we are still subject to ringing like the box filter.



9/17/08

Smoothing by Averaging

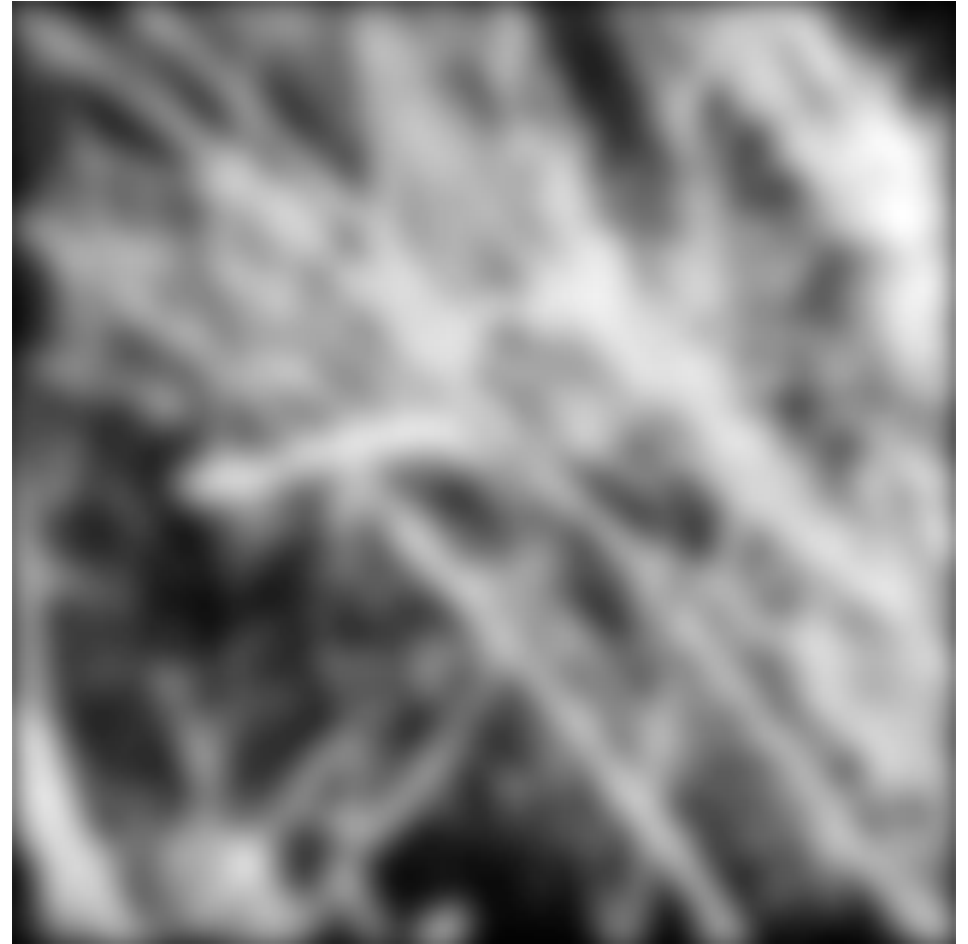
Kernel: 



9/17/08

Smoothing with a Gaussian

Kernel: 

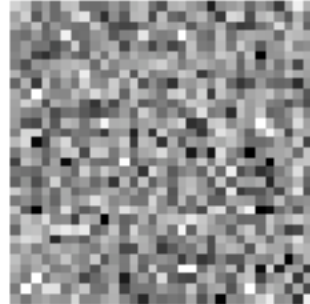
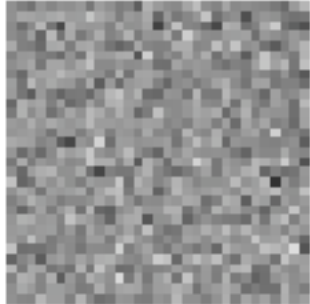
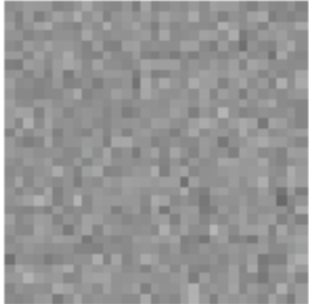


9/17/08

$\sigma=0.05$

$\sigma=0.1$

$\sigma=0.2$



no
smoothing



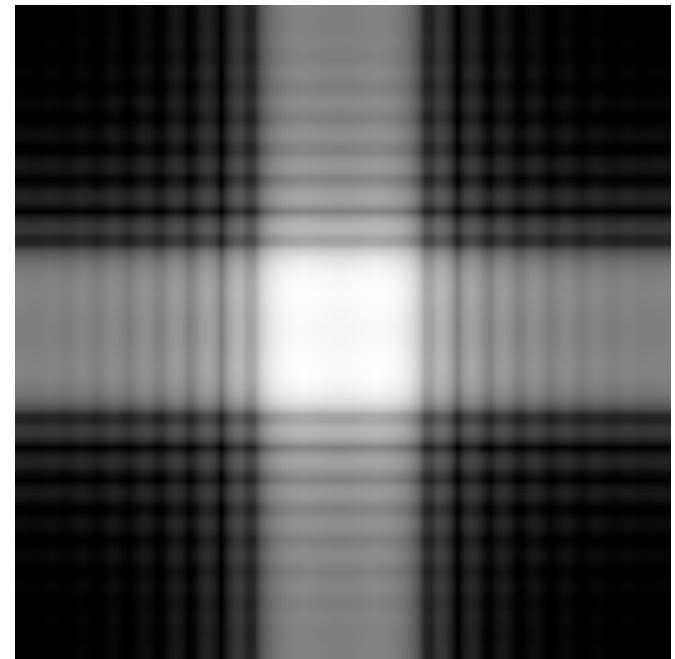
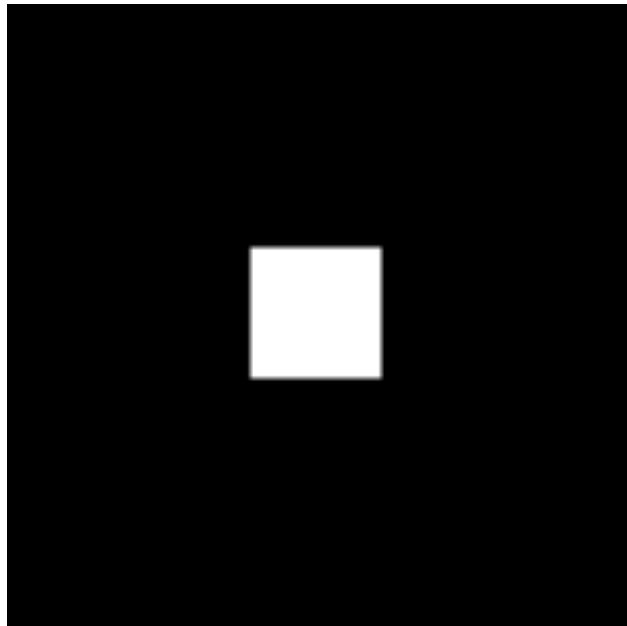
$\sigma=1$ pixel



$\sigma=2$ pixels

The effects of smoothing
Each row shows smoothing with gaussians of different width; each column shows different realizations of an image of gaussian noise.

Why Not a Frequency Domain Filter?



Gabor Filters

- Fourier decompositions are a way of measuring “texture” properties of an image, but they are global
- Gabor filters are a “local” way of getting image frequency content

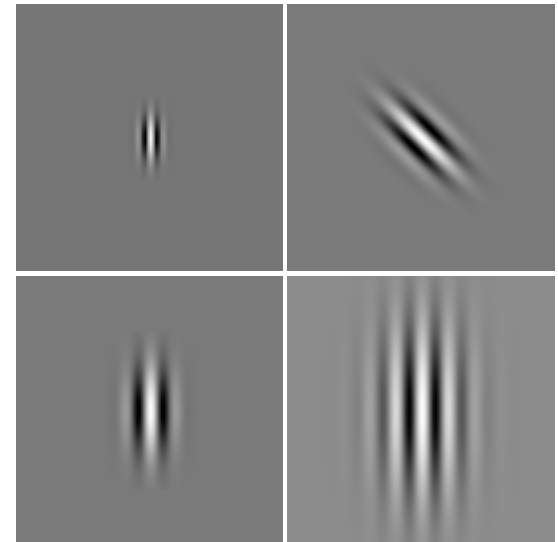
$g(x,y) = s(x,y) w(x,y)$ === a “sin” and a “weight”

$$s(x,y) = \exp(-i (2 \pi (x u + y v)))$$

$$w(x,y) = \exp(-1/2 (x^2 + y^2)/ \sigma^2)$$

Now, we have several choices to make:

1. u and v defines frequency and orientation
2. σ defines scale (or locality)



Thus, Gabor filters for texture can be computationally expensive as we often must compute many scales, orientations, and frequencies

Computational Issues: Separability

- Recall that convolution is associative. Suppose I use the templates $g_x = \exp(-i^2/2 s^2)$ and $g_y = \exp(-j^2/2 s^2)$ Then
 - $g_x * (g_y * I) = (g_x * g_y) * I$
 - but, it is not hard to show that the first convolution is simply the 2-D Gaussian that we defined previously!
- In general, this means that we can “separate” the 2-D Gaussian convolution into 2 1-D convolutions with great computational cost savings.
- A good exercise is to show that the box filter is also separable.

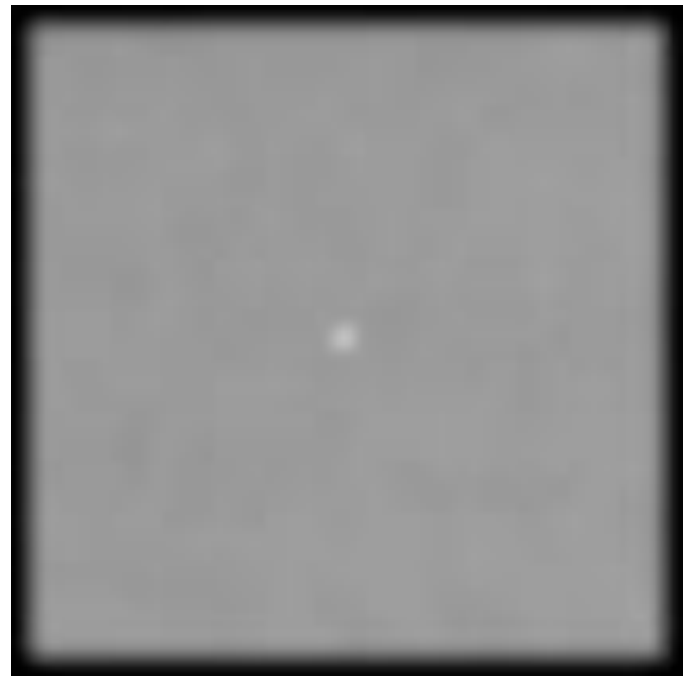
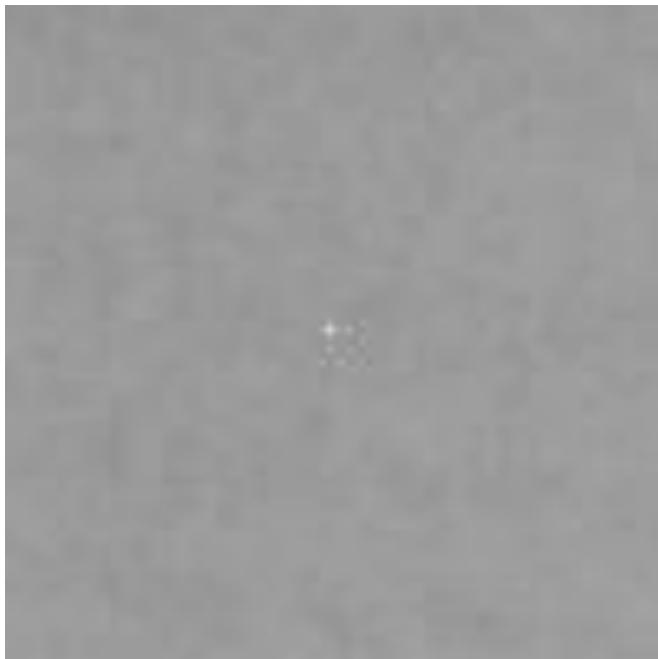
Computational Issues: Minimizing Operations

- Note that for a 256 gray level image, we can *precompute* all values of the convolution and avoiding multiplication.
- For the box filter, we can implement any size using $4n$ additions per pixel.
- Also note that, by the central limit theorem, repeated box filter averaging yields approximations to a Gaussian filter.
- Finally, note that a sequence of filtering operations can be collapsed into one by associativity.
 - in general, this is not a win, but we'll see examples where it is ...

Other Types of Noise

- Impulsive noise
 - randomly pick a pixel and randomly set to a value
 - saturated version is called salt and pepper noise
- Quantization effects
 - Often called noise although it is not statistical
- Unanticipated image structures
 - Also often called noise although it is a real repeatable signal.

Limitations of Linear Operators



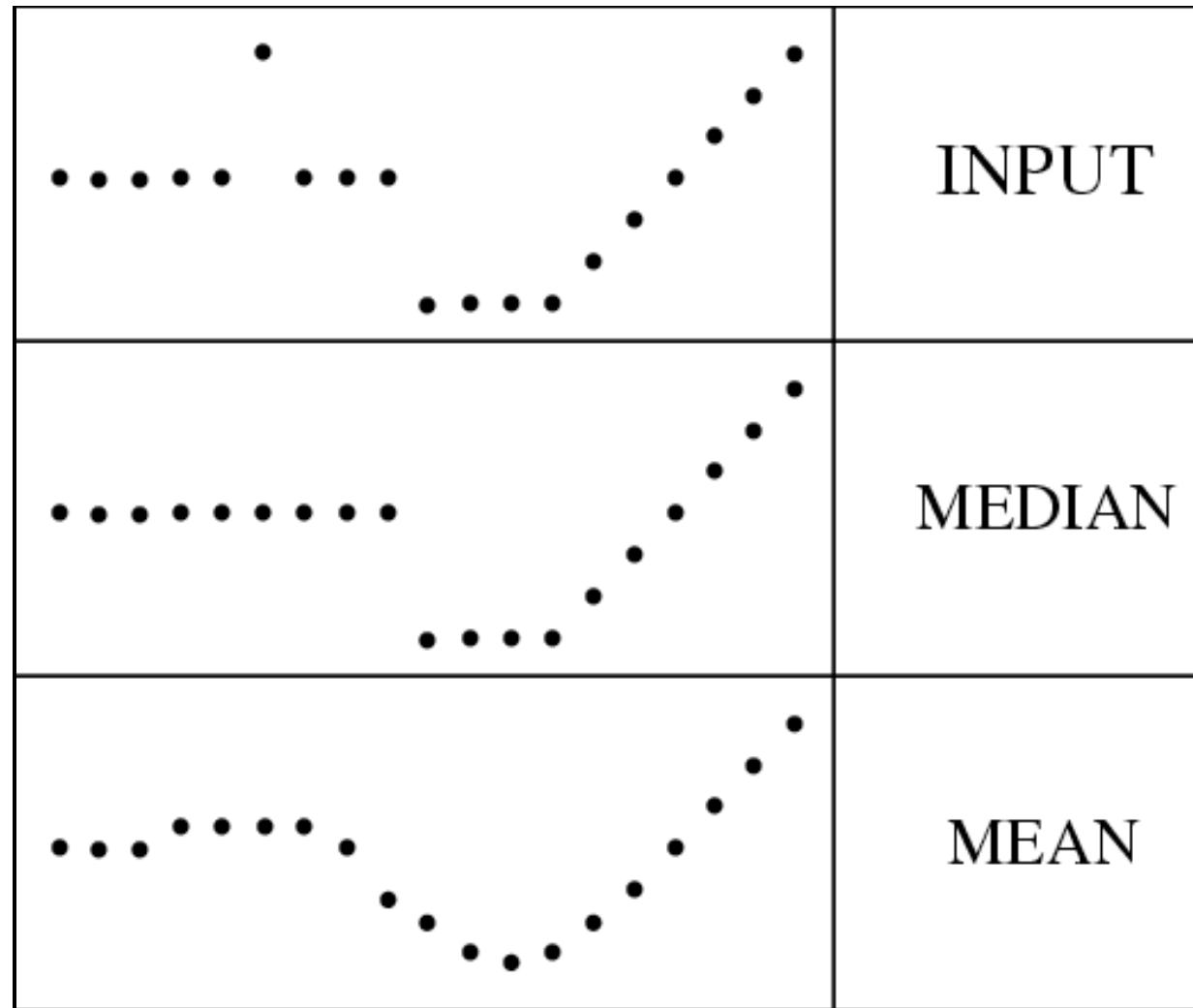
Nonlinear Filtering: The Median Filter

Suppose I look at the local statistics and replace each pixel with the *median* of it's neighbors:



Median filters : example

filters have width 5 :



Digitization Effects

- The “diameter” d of a pixel determines the highest frequency representable in an image

$$l = 1 / 2d$$

- Real scenes may contain higher frequencies resulting in aliasing of the signal.
- In practice, this effect is often dominated by other digitization artifacts.
- One problem in particular is differing sampling rates between digitizer and camera readout of a row.
- This is issue when *resampling* an image.

Filter Pyramids

- An Exercise:
 - Suppose I have $G(s)$ and I perform $(I * G(s)) * G(s)$
 - Hint: think about the convolution theorem and the FFT
- Suppose I want to subsample images
 - subsampling reduces the highest frequencies
 - averaging reduces noise
 - Can I average and resample and reduce noise while not losing desirable frequencies?

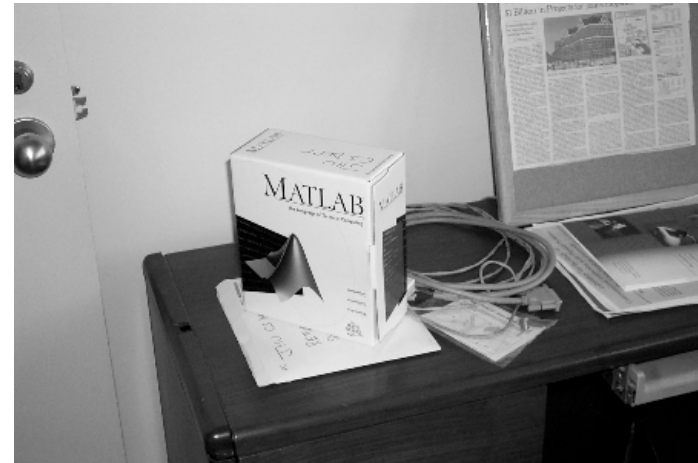
Gaussian Pyramid

- Algorithm:
 - 1. Filter with $G(s)$
 - 2. Resample at every other pixel
 - 3. Repeat



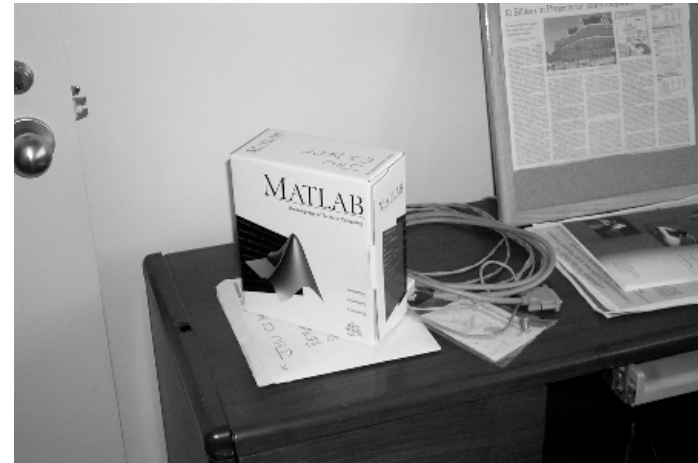
What Else Can You Do With Convolution?

- Thus far, we've only considered convolution kernels that are smoothing filters.
- Consider the following kernel:
 - $[1, -1]$



What Else Can You Do With Convolution?

- Thus far, we've only considered convolution kernels that are smoothing filters.
- Consider the following kernel:
 - $[1; -1]$



Physical causes of edges

1. Object boundaries
2. Surface normal discontinuities
3. Reflectance (albedo) discontinuities
4. Lighting discontinuities

Object Boundaries

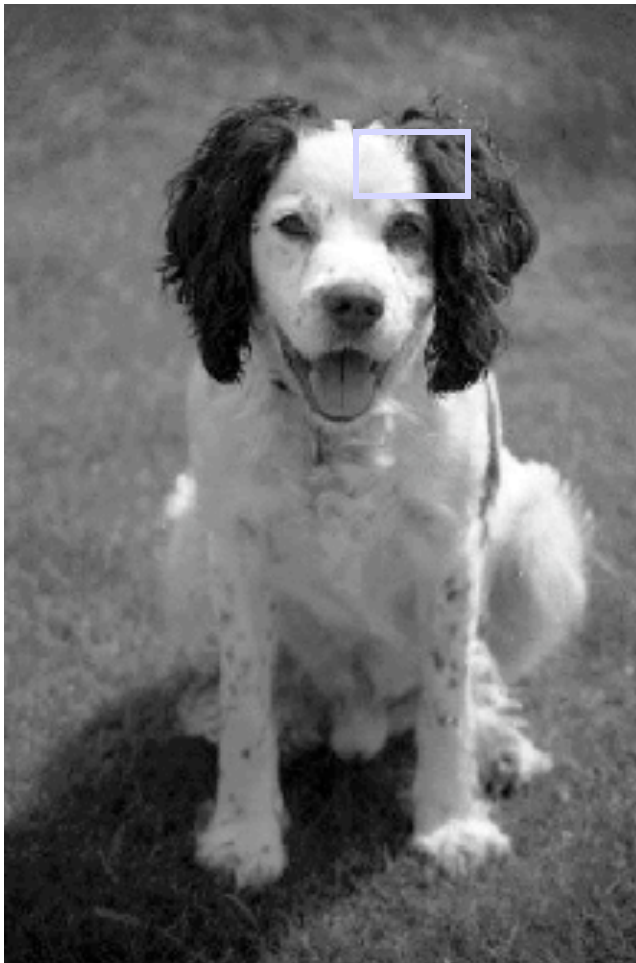


Surface normal discontinuities



9/17/06

Boundaries of material properties



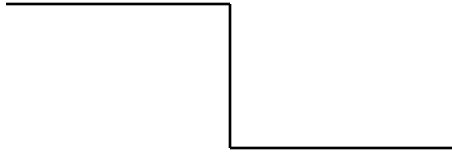
9/17/06

Boundaries of lighting



9/17/06

Edge Types



Step



Ridge

Which of these do you suppose a derivative filter detects best?



Roof

The Image Gradient

- Recall from calculus for a function of two variables $f(x,y)$:
 - The gradient: points in the direction of maximum increase.
 - Its magnitude is proportional to the rate of increase.
 - The total derivative in the direction $n = n \cdot \nabla f$
- The kernel $[-1,1]$ is a way of computing the x derivative
- The kernel $[-1;1]$ is a way of computing the y derivative

Some Other Interesting Kernels

The Roberts Operator

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

The Prewitt Operator

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}$$
$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

Some Other Interesting Kernals

The Sobel Operator $\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$

The Laplacian Operator $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

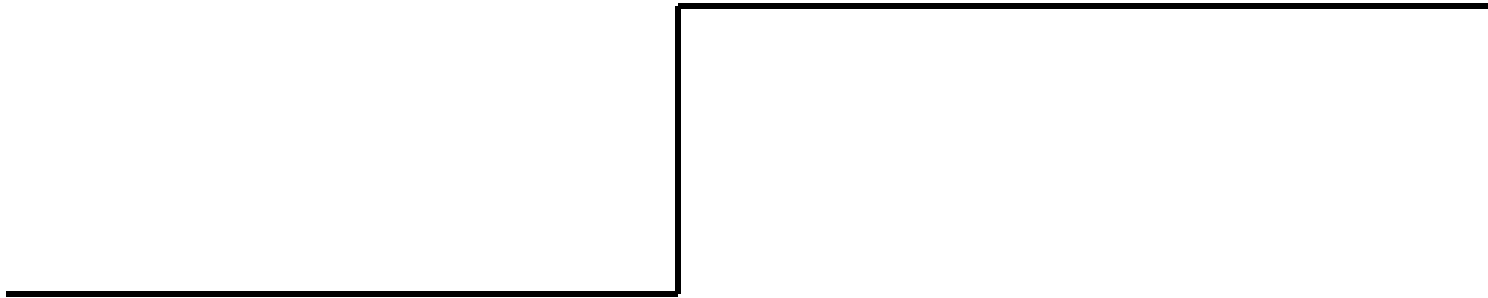
A good exercise: derive the Laplacian from 1-D derivative filters.

Note the Laplacian is rotationally symmetric!

Edge is Where Change Occurs

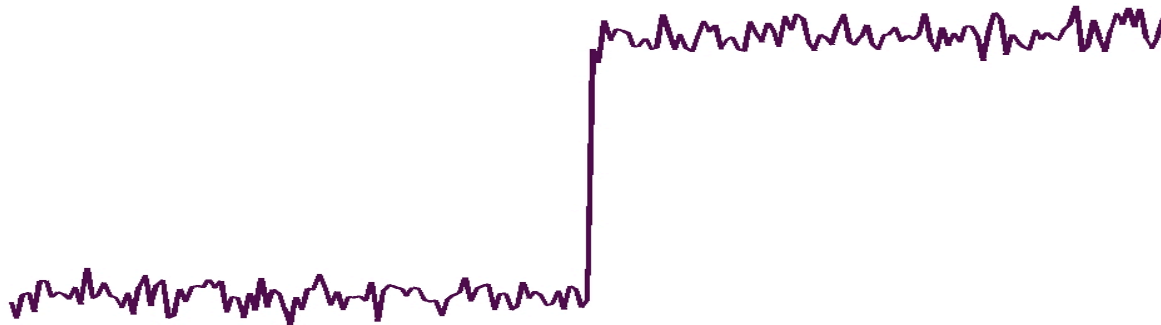
1-D

- Change is measured by derivative in 1D
 - Biggest change, derivative has maximum magnitude
 - Or 2nd derivative is zero.



Noisy Step Edge

- Derivative is high everywhere.
- Must smooth before taking gradient.

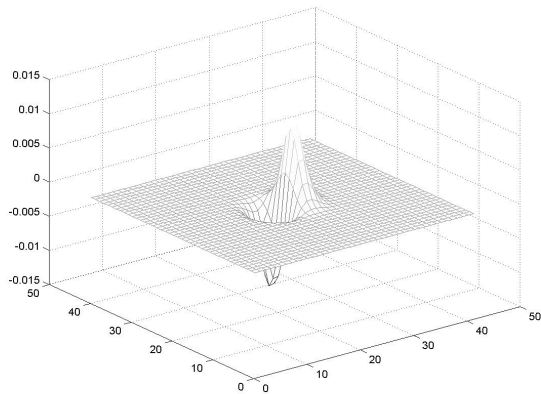


Smoothing Plus Derivatives

- One problem with differences is that they by definition reduce the signal to noise ratio (can you show this?)
- Recall smoothing operators (the Gaussian!) reduce noise.
- Hence, an obvious way of getting clean images with derivatives is to combine derivative filtering and smoothing: e.g.

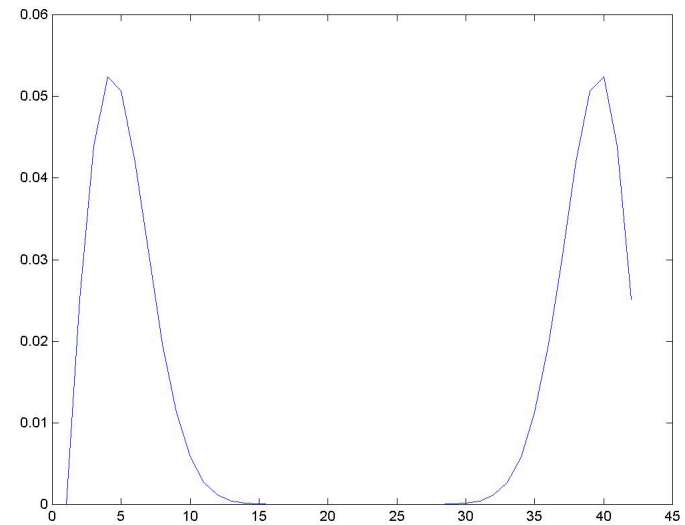
$$- G * D_x * I = D_x * G * I$$

The Fourier Spectrum of DOG



Derivative of a Gaussian

PS of central slice



Properties of the DoG operator

- Now, going back to the directional derivative:
 - $D_u(f(x,y)) = f_x(x,y)u_1 + f_y(x,y)u_2$
- Now, including a Gaussian convolution, we see
 - $D_u[G*I] = D_u[G]*I = [u_1G_x + u_2G_y]*I = u_1G_y*I + u_2G_x*I$
- The two components $I*G_x$ and $I*G_y$ are the *image gradient*
- Note the directional derivative is maximized in the direction of the gradient
- (note some authors use DoG as “Difference of Gaussian” which we’ll run into soon)

Algorithm: Simple Edge Detection

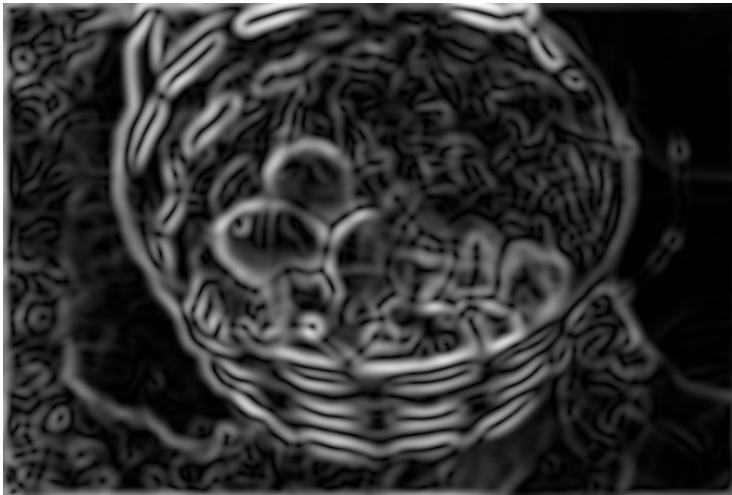
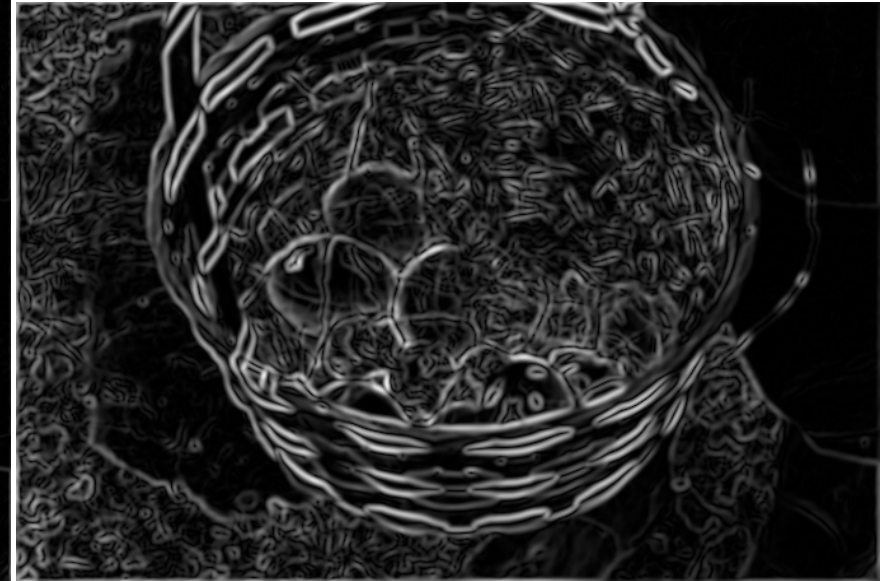
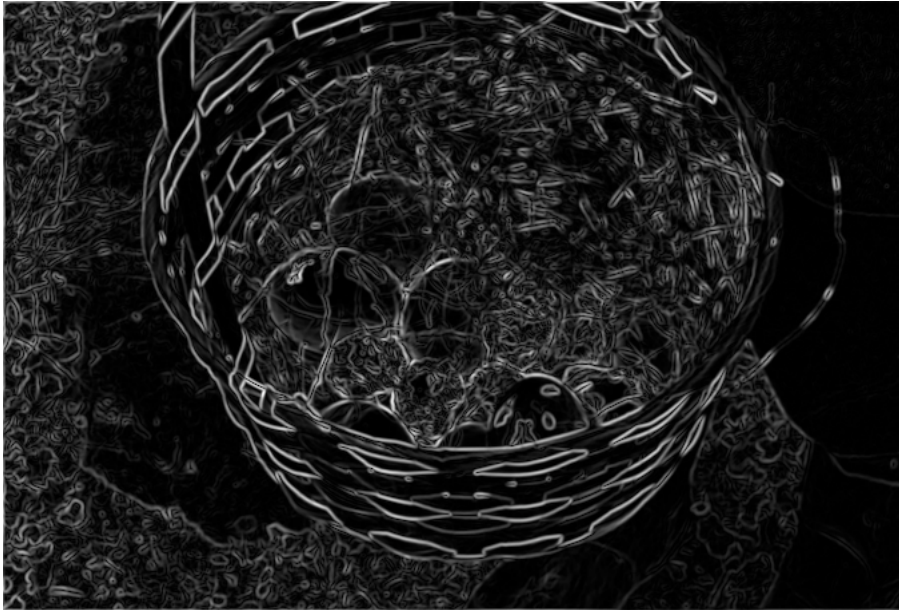
1. Compute $I_x = I_g * (G(s) * G(s)' * [1,-1;1,-1])$
2. Compute $I_y = I_g * (G(s) * G(s)' * [1,-1;1,-1]')$
3. Compute $I_{mag} = \text{sqrt}(I_x .* I_x + I_y .* I_y)$
4. Threshold: $I_{res} = I_{mag} > t$

It is interesting to note that if we wanted an edge detector for a specific direction of edges, we can simply choose the appropriate projection (weighting) of the component derivatives.

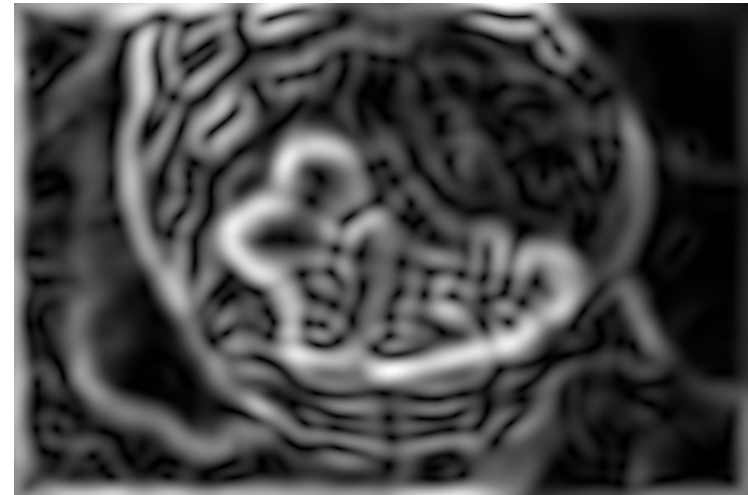
sigma = 1

Example

sigma = 2



sigma = 5



sigma = 10

9/17/08

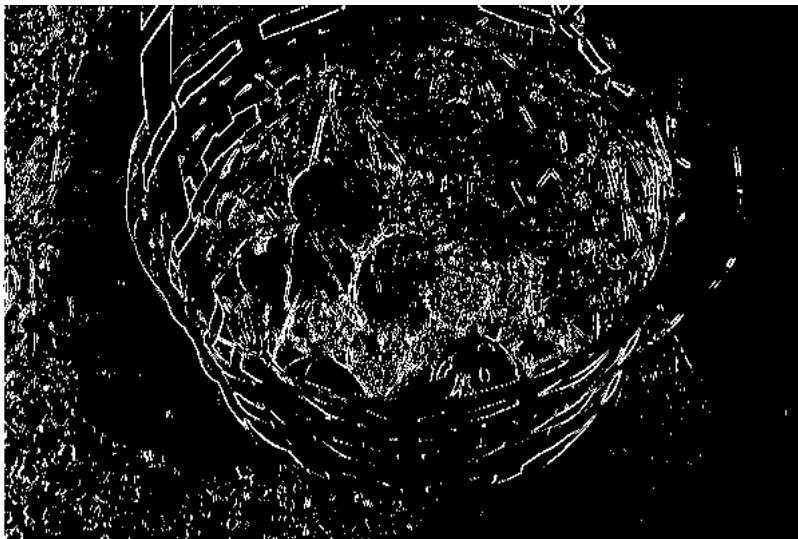
Laplacian Pyramid Algorithm

- Create a Gaussian pyramid by successive smoothing with a Gaussian and down sampling
- Set the coarsest layer of the Laplacian pyramid to be the coarsest layer of the Gaussian pyramid
- For each subsequent layer $n+1$, compute
 - $L(n+1) = G(n+1) - \text{upsample}(G(n))$
- In general, the idea of using a series of Gaussians with different values of s leads to the idea of *scale space* in computer vision

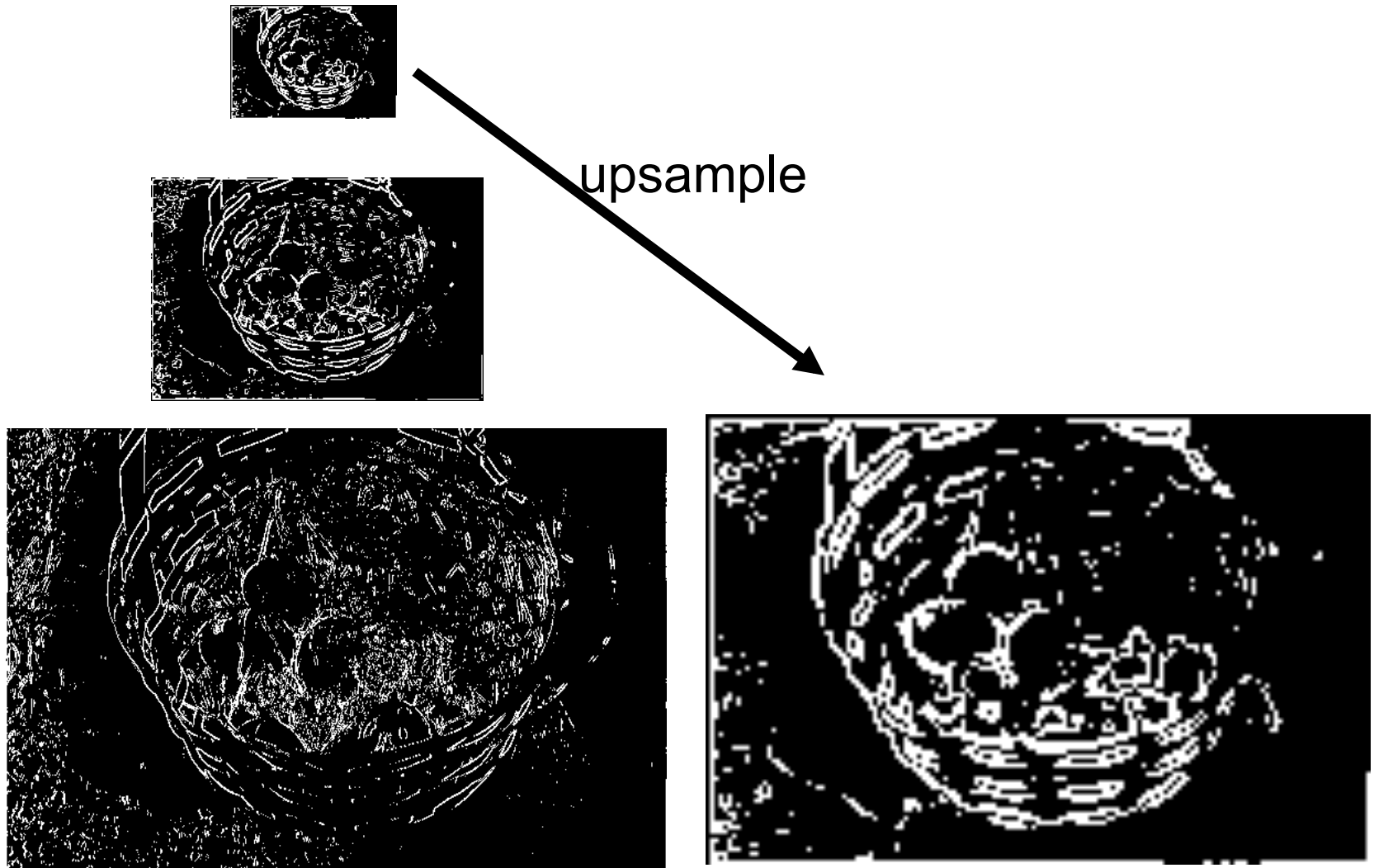
$$L(s) = I * G(s)$$

- i.e. a continuous family of images

Laplacian of Gaussian Pyramid



Laplacian of Gaussian Pyramid



Summary

- Linear filters as
 - Smoothing
 - Gradient operators
 - Laplacian
- Fourier transforms
 - What they represent
 - Convolution Theorem
- Computational Issues
 - Association
 - Separability