Pre-processing

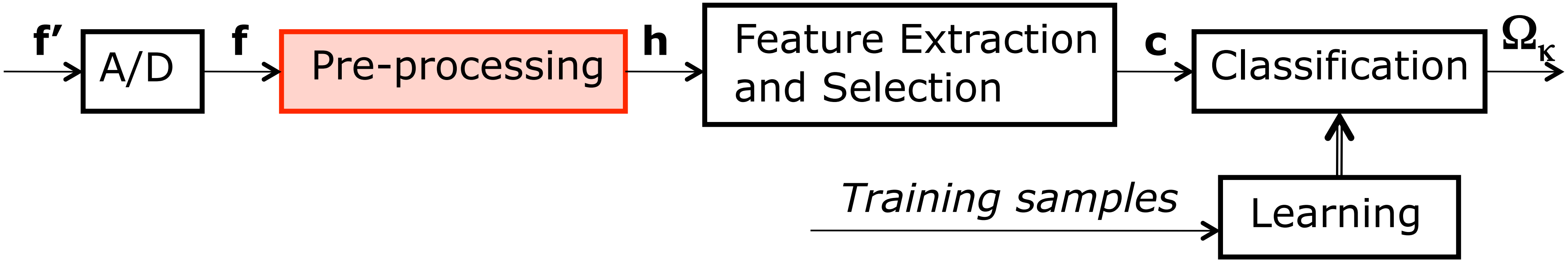
Filtering

Dr. Elli Angelopoulou

Lehrstuhl für Mustererkennung (Informatik 5)

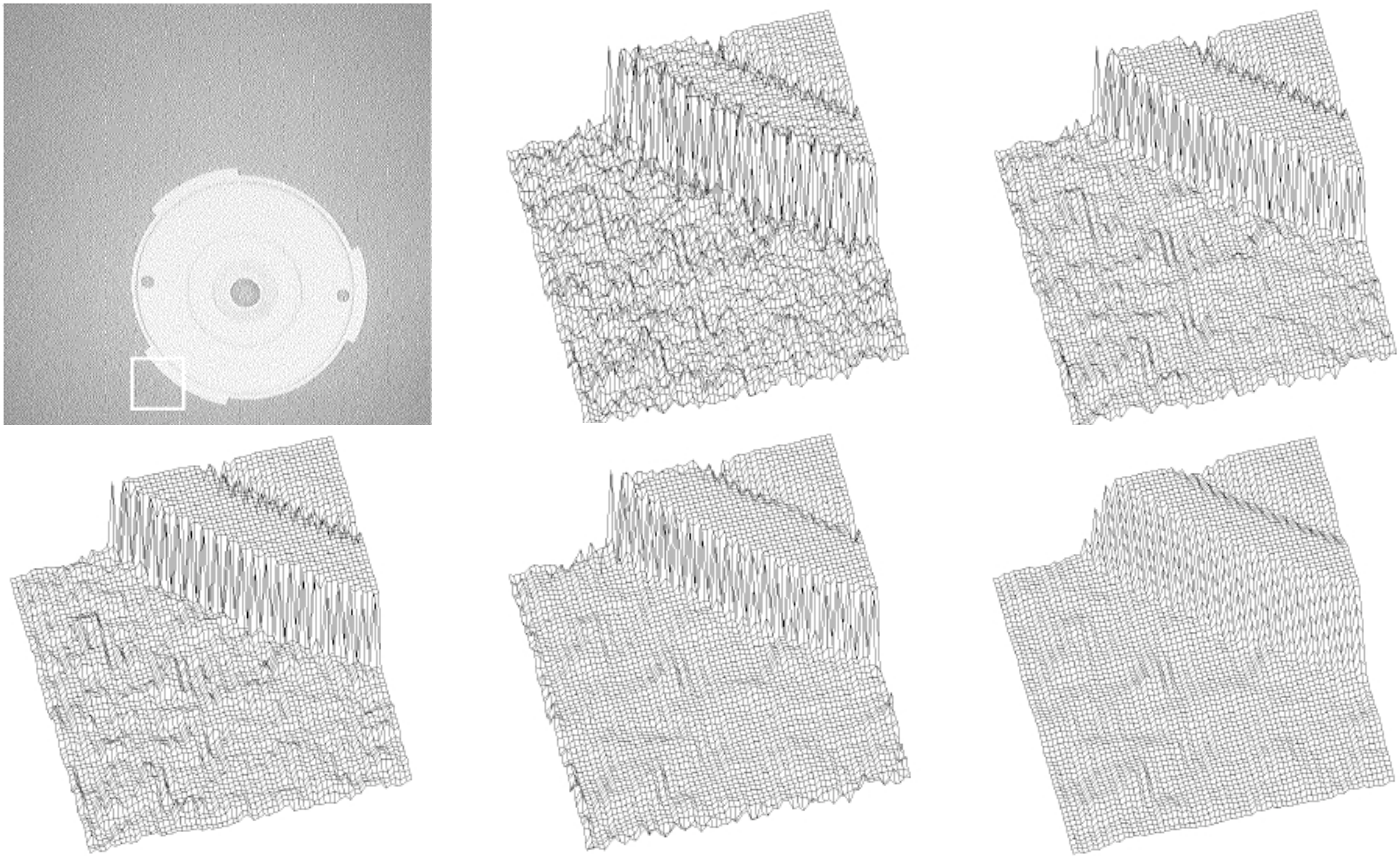
Friedrich-Alexander-Universität Erlangen-Nürnberg

Pattern Recognition Pipeline



- The goal of pre-processing is to transform a signal f to another signal h so that the resulting signal h
  - makes subsequent processing easier
  - makes subsequent processing better (more accurate)
  - makes subsequent processing faster
- Already studied *histogram equalization* and *thresholding*.

Pre-processing Example



Noise Sources

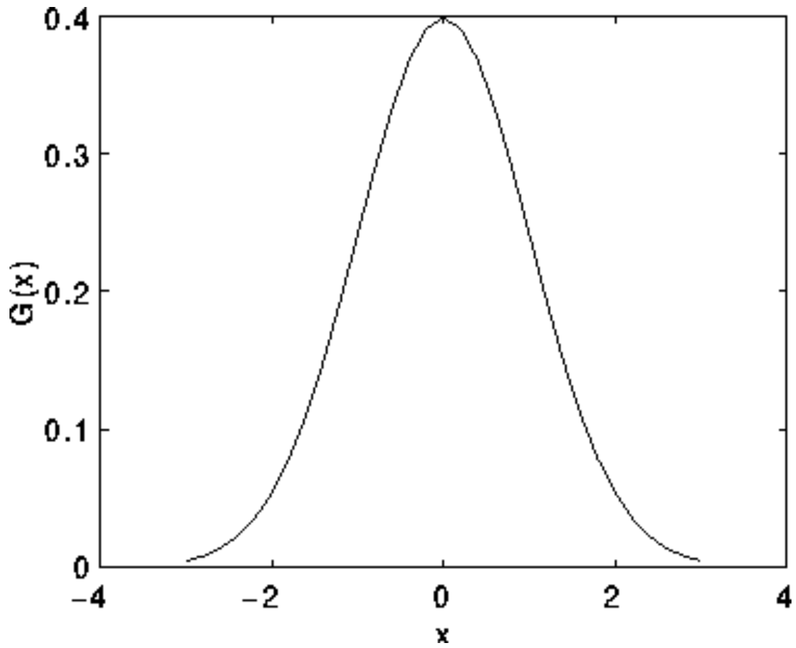
- Photon noise: variation in the #photons falling on a pixel per time interval T.
- Saturation: each pixel can only generate a limited amount of charge.
- Blooming: saturated pixel can overflow to neighboring pixels.
- Thermal noise: heat can free electrons and generate a response when there is none.
- Electronic noise.
- Burned pixels.
- Black is not black.



- Keep in mind: Camera response may not be linear over the number of photons falling on a surface (camera gamma)

Detector Noise

- Source of noise: the discrete nature of radiation, i.e. the fact that each imaging system is recording an image by counting photons.
- Can be modeled as an independent additive noise which can be described by a zero-mean Gaussian.



Salt and Pepper Noise

- A common form of noise is caused by *data drop-out noise.*
- It is also known as commonly referred to as intensity spikes, speckle or salt and pepper noise.
- Sources of error:
  - Errors in the data transmission.
  - Burned pixels: the corrupted pixels are either set to the maximum value (which looks like snow in the image) or are set to zero ("peppered" appearance), or a combination of the two.
  - Single bits are flipped over.
- Isolated/localized noise. It only affects individual pixels

Filtering

- Most of the images we capture are noisy
- Goal:

Noisy Image$_{in}$ → Filter → Clean Image$_{out}$

- This notion of filtering is more general and can be used in a wide range of transformations that we may want to apply to images.

Image$_{in}$ → Filter → Image$_{out}$

- Mathematically, a filter *H* can be treated as a function on an input image *I*:

$$H(I) = R$$

- Note: We use the terms *filter* and *transformation* interchangeably

Linear Transformation

- A transformation H is **linear** if, for any inputs $I_1(x,y)$ and $I_2(x,y)$ (in our case input images), and for any constant scalar α we have:

$$H(\alpha I_1(x,y)) = \alpha H(I_1(x,y))$$

and

$$H(I_1(x,y) + I_2(x,y)) = H(I_1(x,y)) + H(I_2(x,y))$$

- This means:
  - Multiplication in the input corresponds to multiplication in the output
  - Filtering an additive image is equivalent to filtering each image separately and then adding the results.

Shift-Invariant Transformation

- A transformation *H* is **shift-invariant** if for every pair (x_0, y_0) and for every input image $I(x,y)$, such that

$$H(I(x,y)) = R(x,y)$$

we get

$$H(I(x - x_0, y - y_0)) = R(x - x_0, y - y_0)$$

- This means that the filter *H* does not change as we shift it in the image (as we move it from one position to the next).

Convolution

- If a transformation (or filter) is linear shift-invariant (LSI) then one can apply it in a systematic manner over every pixel in the image.
- **Convolution** is the process through which we apply **linear shift-invariant filters** on an image.

- Convolution is defined as:

$$R(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} H(x-i, y-j) I(i,j)$$

and is denoted as:

$$R = H * I$$

Another Look at Convolution

- Filtering often involves replacing the value of a pixel in the input image *F* with the weighted sum of its neighbors.
- Represent these weights as an image, ***H***
- ***H*** is usually called the **kernel**
- The operation for computing this weighted sum is called **convolution.**

$$R = H * I$$

- Convolution is:
  - commutative, $H * I = I * H$
  - associative, $H_1 * (H_2 * I) = (H_1 * H_2) * I$
  - distributive, $(H_1 + H_2) * I = (H_1 * I) + (H_2 * I)$

Smoothing via Simple Averaging

- One of the simplest filters is the mean filter: $H = \begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$
- In this case, $R(x,y) = \sum_{i=-1}^{1} \sum_{j=-1}^{1} I(x-i, y-j) H(i,j)$
- It is used for removing image noise, i.e. for smoothing.

Original image

$*$  $=$

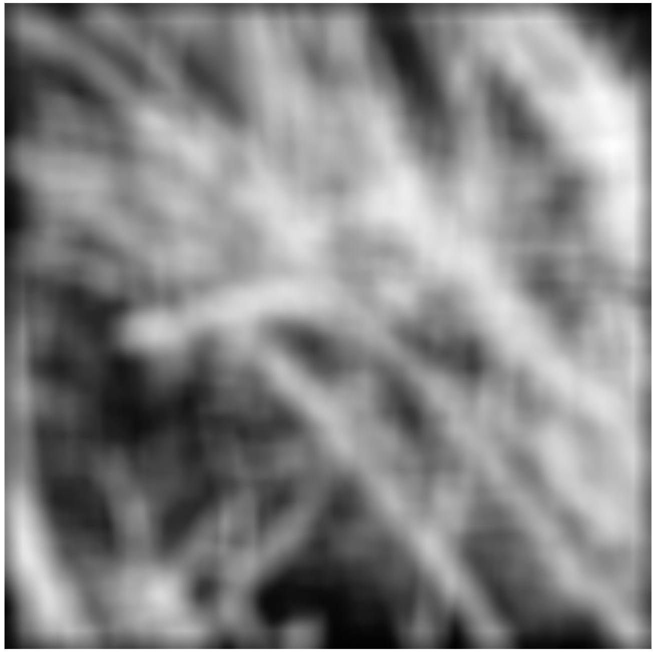


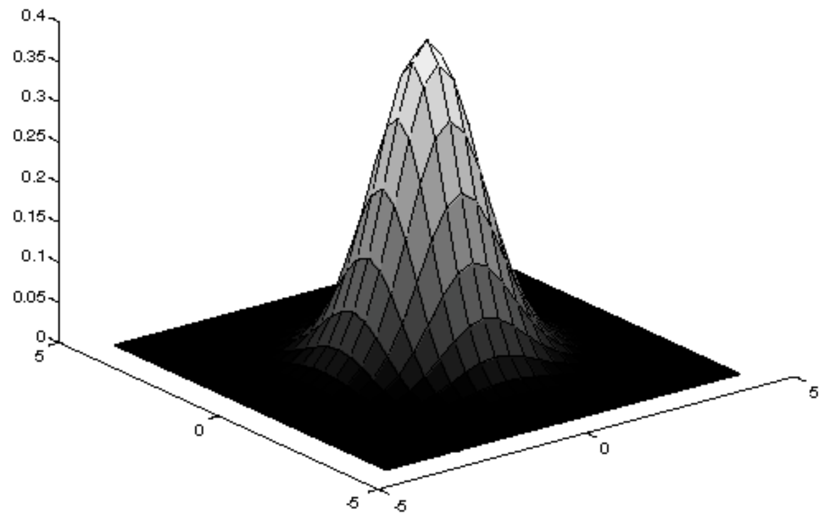
Image after mean filtering (25x25 kernel)

Gaussian Smoothing

- Idea: Use a weighted average. Pixels closest to the central pixel are more heavily weighted.

- The Gaussian function has exactly that profile.

- Gaussian also better approximates the behavior of a defocused lens.



Isotropic Gaussian Filter

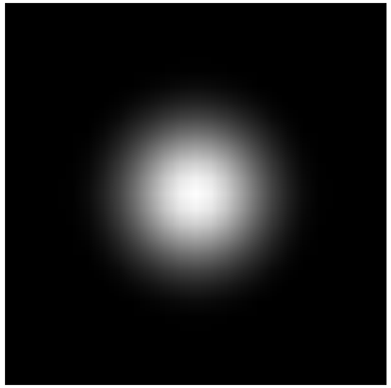
- To build a filter *H*, whose weights resemble the Gaussian distribution, assign the weight values on the matrix *H* according to the Gaussian function:

$$H(i,j) = e^{-(i^2 + j^2)/2\sigma^2}$$

$$H_{Gauss} = \begin{bmatrix} 1/16 & 1/8 & 1/16 \\ 1/8 & 1/4 & 1/8 \\ 1/16 & 1/8 & 1/16 \end{bmatrix}$$

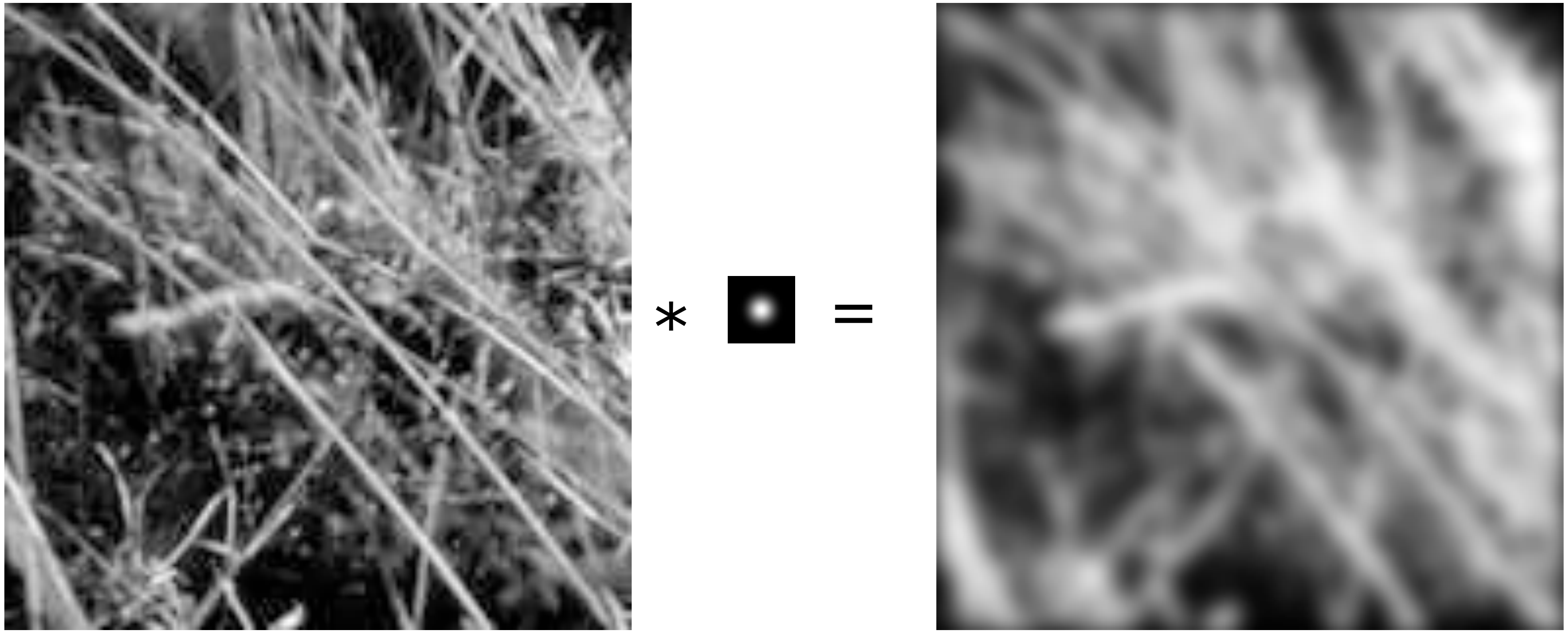
- Small σ, almost no effect, weights at neighboring points are negligible.
- Large σ, blurring, neighbors have almost the same weight as the central pixel.
- Commonly used σ values: Let w be the size of the kernel *H*. Then σ=w/5.

  For example for a 3x3 kernel, σ=3/5=0.6



Gaussian Smoothing Example

- Compared to mean filtering, Gaussian filtering exhibits no “ringing” effect.



Original image

Image after Gaussian filtering (25x25 kernel)

“Ringing” effect

Original image

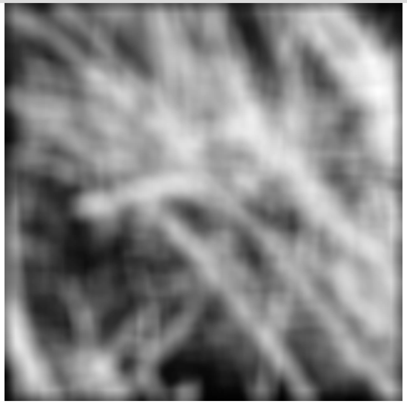


Image after Mean filtering (25x25 kernel)

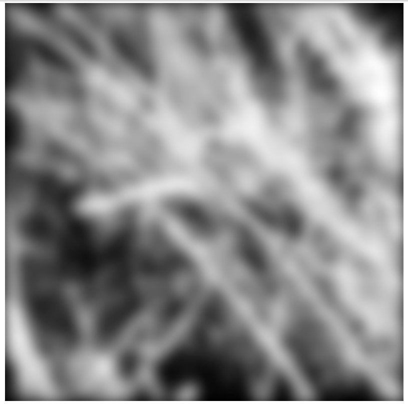
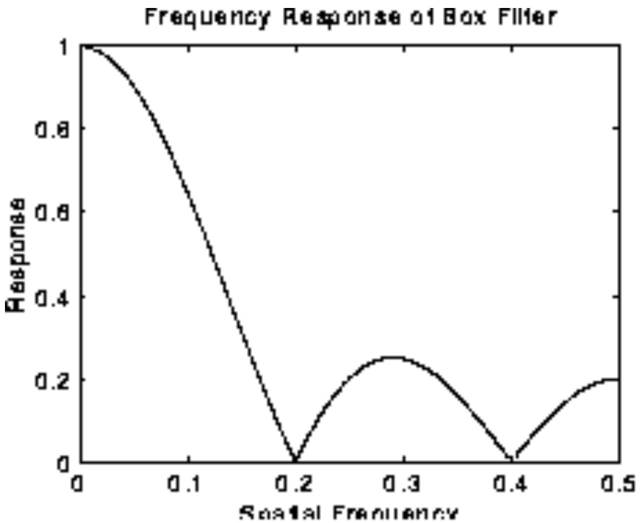
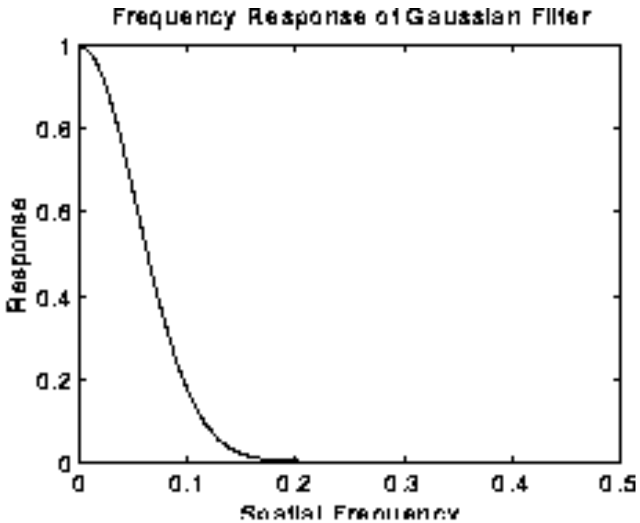


Image after Gaussian filtering (25x25 kernel)

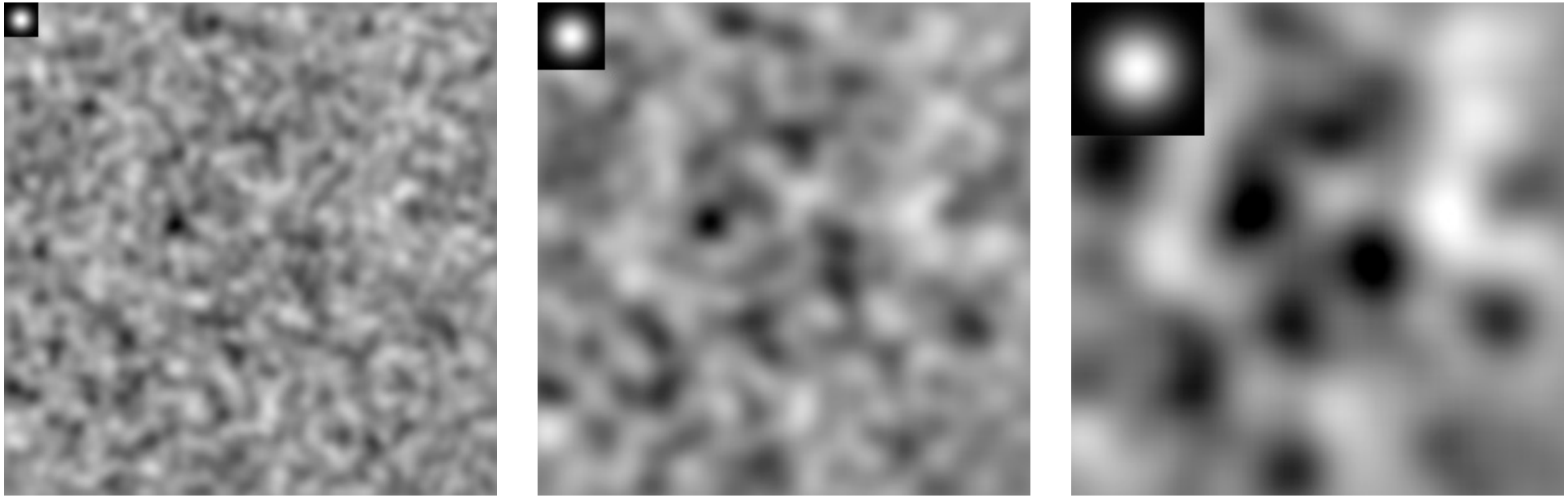




A close look at the frequency response of the two filters show that: compared to Gaussian filtering, mean filtering exhibits oscillations

The Effect of σ

- Different σ values affect the amount of blurring, but also emphasize different characteristics of the image.



Non-Linear Smoothing

- The **median** filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it is representative of its surroundings.
- It replaces a pixel value with the median of all pixel values in the neighborhood.
- It is a relatively slow filter, because it involves sorting.
- Can **not** be implemented via convolution.

Smoothing Examples

Original image

Image after 9x9 Mean filtering

Image after 9x9 Gaussian filtering

Mean Filter

Original image

Image after 3x3 Mean filtering

Image after 7x7 Mean filtering

Image after applying 3 times 3x3 Mean filtering

- Mean filtering is sensitive to outliers.
- It typically blurs edges.
- It often causes a ringing effect.

Gaussian Filtering and Salt & Pepper Noise

Original image

Image with salt-pepper noise (1% prob. that a bit is flipped)

Image after 5x5 Gaussian filtering, σ=1.0

Image after 9x9 Gaussian filtering, σ=2.0

- Gaussian filtering works very well for images affected by Gaussian noise
- It is not very effective in removing Salt and Pepper noise.

Median Filtering and Salt & Pepper Noise

Original image

Image with salt-pepper noise (5% prob. that a bit is flipped)

Image after 3x3 Median filtering

Image after 7x7 Median filtering

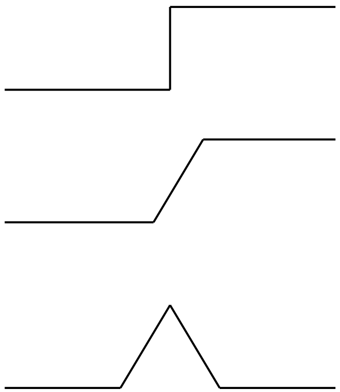
Image after applying 3 times 3x3 Median filtering

- Median filtering preserves high spatial frequency details.
- It works well when less than half of the pixels in the smoothing window have been affected by noise.
- It is not very effective in removing Gaussian noise.

Edges

- An edge is:
  - A significant change in intensity values.
  - Related to object boundaries, patterns (brick wall), shadows, etc.
  - A property attached to each pixel.
  - Calculated using the image intensities of neighboring pixels.

- Examples of 1D Edges



50	50	50	50	100	100	100	100	step edge
50	50	60	70	80	90	100	100	ramp edge
50	50	50	50	100	100	50	50	roof edge

Edge Detection Example

Original images

Images after edge detection

Edge Detection Steps

1. Noise Smoothing
   - Suppress as much noise as possible without destroying edge information.

2. Edge Enhancement
   - Design a filter that gives high responses at edges and low response at non-edge pixels.

3. Edge Localization
   - Decide which high responses of the edge filter are responses to true edges and which ones are caused by noise or other artifacts.

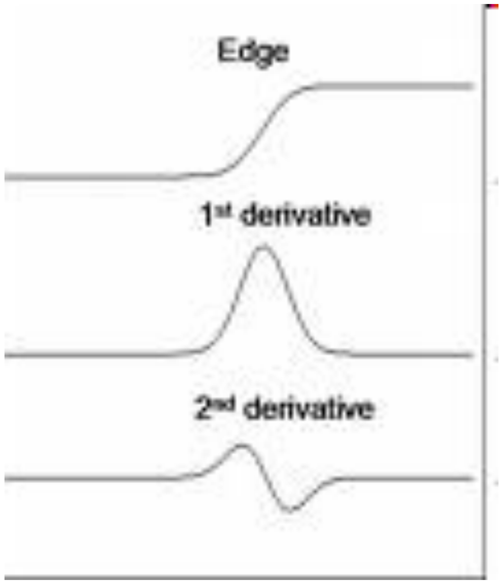
Types of Edge Detection

- Detecting edges is equivalent to detecting changes in intensity values.
- How do we detect change?

  Differentiation
- Image is a 2D function

  => partial derivative in x

  & partial derivative in y



- If we take the 1st derivative we have **Gradient-based** edge detectors.
- If we take the 2nd derivative we have **Laplacian** edge detectors (look for zero-crossings).

Gradient-Based Edge Detection

- The gradient vector **G**(x,y), at an image pixel *I(x,y)* is:

$$\mathbf{G}(x,y)=\left(\frac{\partial I(x,y)}{\partial x},\frac{\partial I(x,y)}{\partial y}\right)=(I_x(x,y),I_y(x,y))$$

- The gradient vector points in the direction of maximum change.
- Its orientation (its angle with the x-axis) is given by:

$$\theta=\tan^{-1}\left(I_y(x,y)\big/I_x(x,y)\right)$$

- Its magnitude is given by: $\|\mathbf{G}(x,y)\|=\sqrt{I_x^2(x,y)+I_y^2(x,y)}$

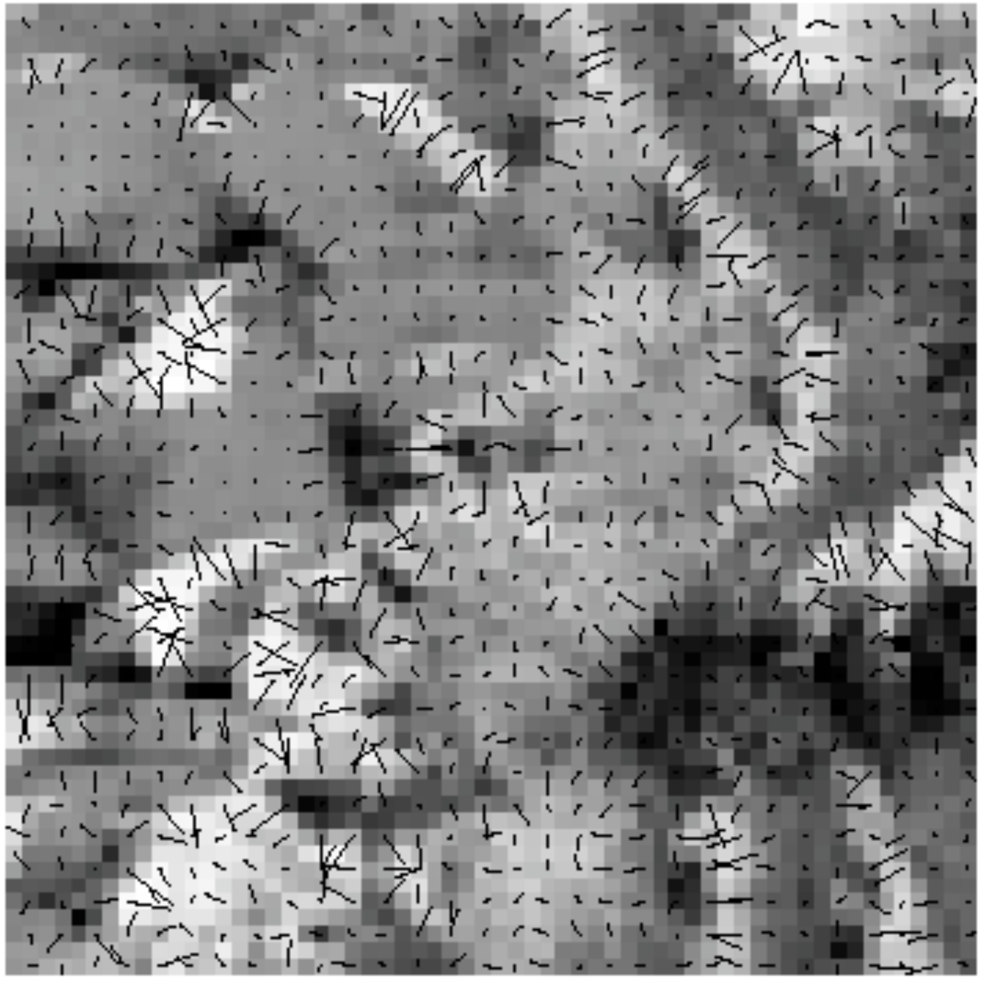
  or its approximations:

$$\|\mathbf{G}(x,y)\|\approx|I_x(x,y)|+|I_y(x,y)|$$

$$\|\mathbf{G}(x,y)\|\approx\max(I_x(x,y),I_y(x,y))$$

Gradient Vector Image

- An image showing the gradient vectors themselves.
- The length of the gradient vector corresponds to its magnitude.



Implementation

- By definition:

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

- In the discrete world differentiation is approximated by finite differencing:

$$I_x(x,y) = \partial I(x,y)/\partial x \approx \frac{I[x,y] - I[x-\Delta x,y]}{\Delta x}$$

- But since our smallest step is $\Delta x = 1$:

$$I_x(x,y) = \partial I(x,y)/\partial x = I[x,y] - I[x-1,y]$$
$$I_y(x,y) = \partial I(x,y)/\partial y = I[x,y] - I[x,y-1]$$

Implementation (continued)

- We can express this operation in a kernel form:

$$H_x = I_x = \begin{bmatrix} -1 & +1 \end{bmatrix} \qquad H_y = I_y = \begin{bmatrix} -1 \\ +1 \end{bmatrix}$$

- To make it less susceptible to noise we use the values of two consecutive rows or columns.

$$H_x = I_x = \begin{bmatrix} -1 & +1 \\ -1 & +1 \end{bmatrix} \qquad H_y = I_y = \begin{bmatrix} -1 & -1 \\ +1 & +1 \end{bmatrix}$$

- These kernels, however, evaluate an approximation of the derivative at half-pixel locations, $I_x[x-1/2, y]$ and $I_y[x, y-1/2]$

Common Edge Masks

- Prewitt edge detection masks

$$P_x = \begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix} \qquad P_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{bmatrix}$$

- Sobel edge detection masks

$$S_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \qquad S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}$$

Gradient Edge Detection Process

- Given an input image I, the gradient-based edges are computed as follows:

1. Compute $I_x = H_x * I$
2. Compute $I_y = H_y * I$
3. Compute $\|\mathbf{G}(x, y)\|$ using your favorite method
4. If $\|\mathbf{G}(x, y)\| \geq t$

   then pixel (x,y) is an edge-pixel (*edgel*)

   compute the angle θ for that pixel.

Gradient Edge Detector Example

Original image

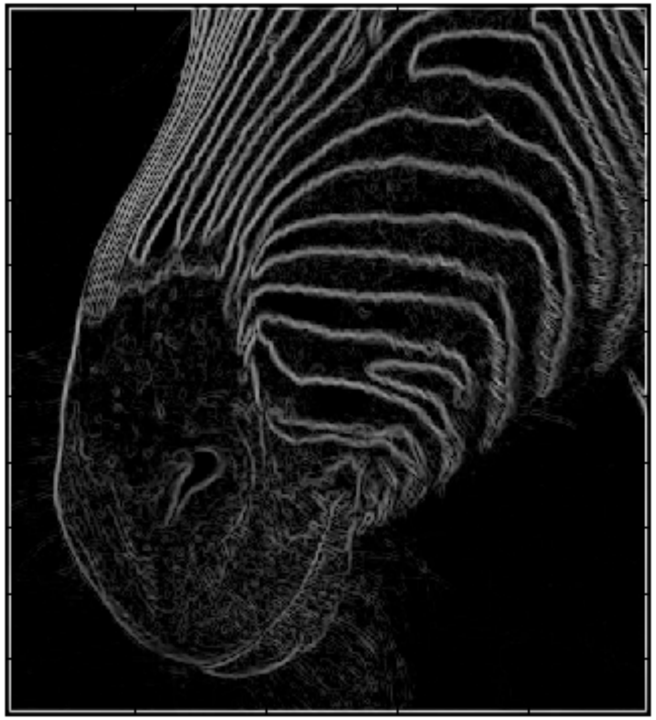


Image after edge detection

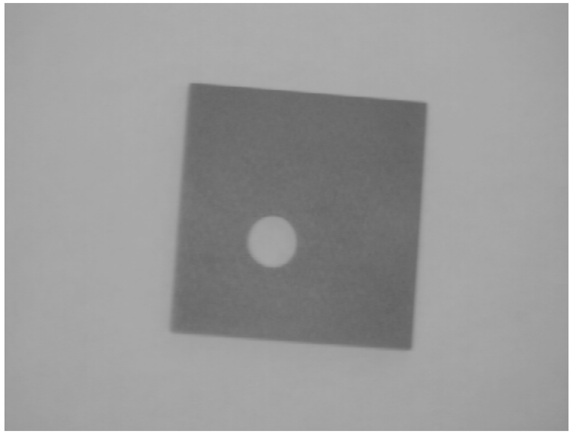
Canny Edge Detector

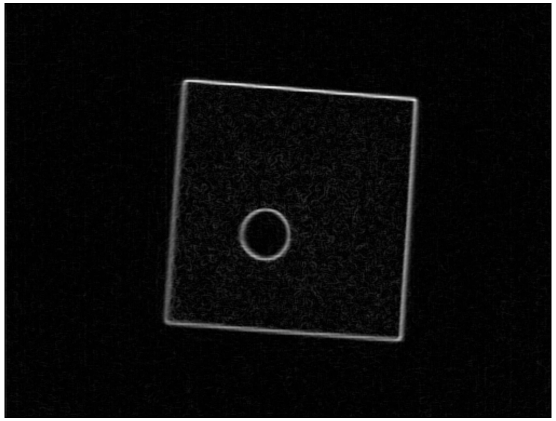
- After a a gradient-based edge image is created, the Canny method uses optimization to systematically clean noise effects. It uses two separate optimization processes:
  1. Non-maximum suppression

     A single real edge may appear as having wide ridges around it.
     Non-maximum suppression thins such ridges downto 1-pixel wide edges.
  2. Hysteresis thresholding

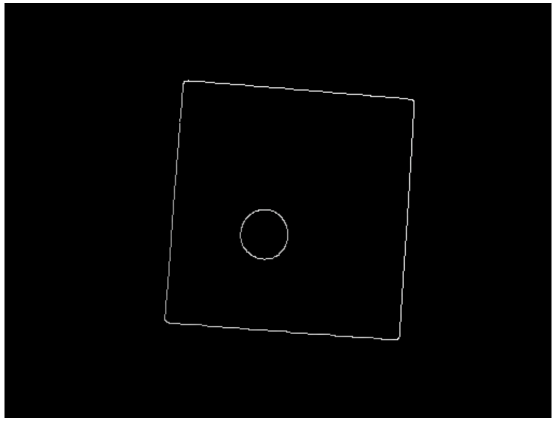
     Use a pair of threshold values. The high threshold is used as a first rough screening. For the edge pixels that survive this first screening, follow chains (contours) of edges. Use those edgels on the chain which are above the second, lower, threshold.
- Canny proved that this is the optimal edge detection method.
- Due to the optimization post-processing, it is slower than the basic gradient-based edge detectors.

Sobel versus Canny



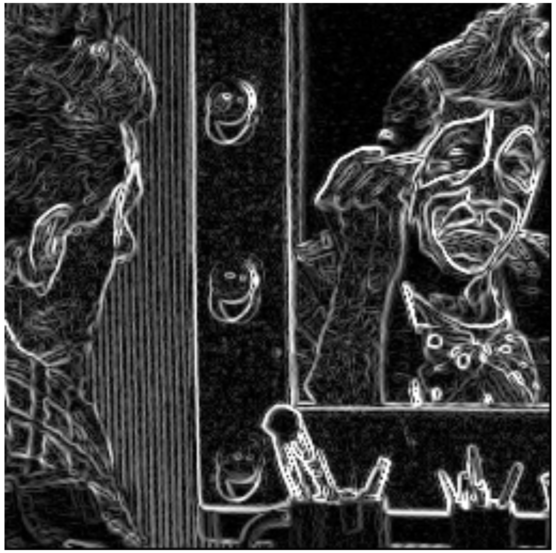


Sobel

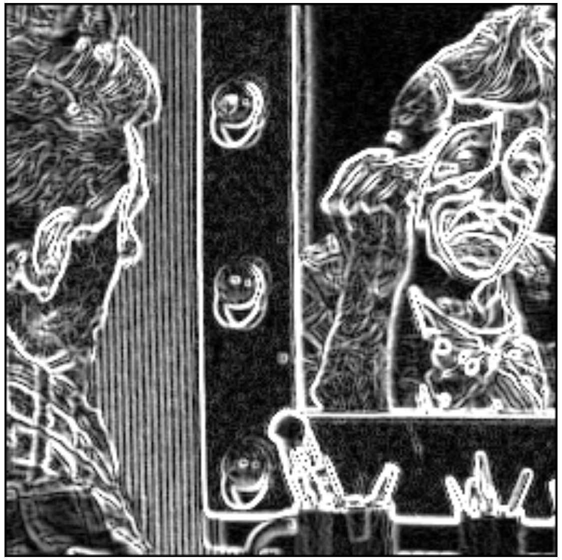


Canny

Roberts vs. Sobel

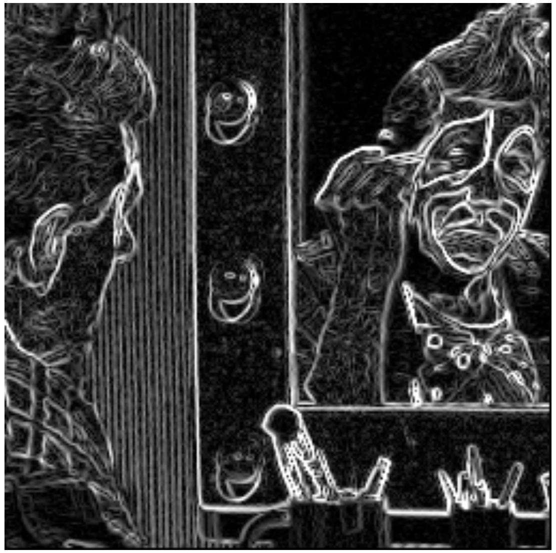


Roberts



Sobel

Roberts vs. Canny



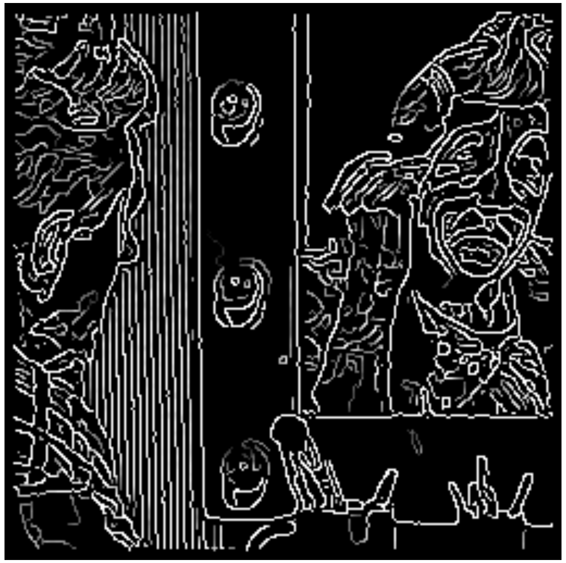
Roberts

Canny

$\sigma = 1, t_l = 1, t_h = 255$

Canny Edge Detector

Canny
$\sigma = 1, t_l=220\ , t_h= 255$



Canny
$\sigma = 1, t_l=1\ , t_h= 128$

Canny
$\sigma = 2, t_l=1\ , t_h= 128$

LSI Filtering - Review

- We are focusing on linear shift-invariant (LSI) filters.

- Such filters are typically applied to an image through convolution:

$$R(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} H(x-i, y-j) I(i,j)$$

- In practice, convolution can be seen as computing the weighted sum of a (2k+1)x(2k+1) neighborhood centered around pixel (x,y), where the filter H contains the applied weights.

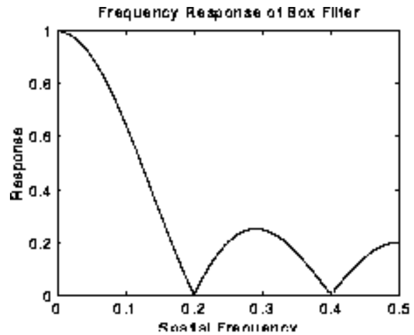
$$R(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x-i, y-j) H(i,j)$$

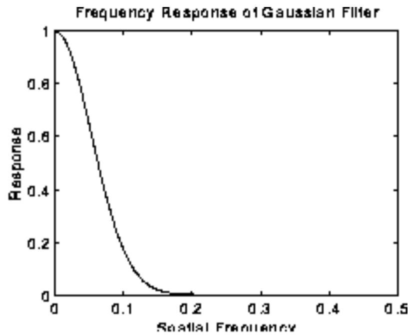
Convolution and LSI filtering - Review

- Important Properties of Convolution:
  - commutativity, $H * I = I * H$
  - associativity, $H_1 * (H_2 * I) = (H_1 * H_2) * I$
  - distributivity, $(H_1 + H_2) * I = (H_1 * I) + (H_2 * I)$

- A very common application of filtering is for noise removal.

- Two LSI smoothing filters are:
  - Mean filter
  - Gaussian filter





- They are also known as low-pass filters, because in the frequency domain, they allow only transfer the low frequency information in the output image.

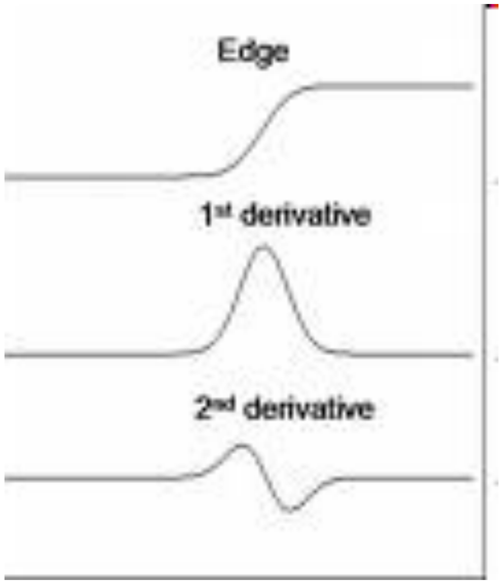
Edge Detection - Review

Original images

Images after edge detection

Types of Edge Detection - Review

- Detecting edges is equivalent to detecting changes in intensity values.
- How do we detect change?
  Differentiation
- Image is a 2D function
  => partial derivative in x
  & partial derivative in y



- If we take the 1st derivative we have **Gradient-based** edge detectors.
- If we take the 2nd derivative we have **Laplacian** edge detectors (look for zero-crossings).

Gradient-Based Edge Detection - Review

- The gradient vector **G**(x,y), at an image pixel *I(x,y)* is:

$$\mathbf{G}(x,y) = \left(\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y} \right) = (I_x(x,y), I_y(x,y))$$

- The gradient vector points in the direction of maximum change.
- Its orientation (its angle with the x-axis) is given by:

$$\theta = \tan^{-1}\left(I_y(x,y) \big/ I_x(x,y) \right)$$

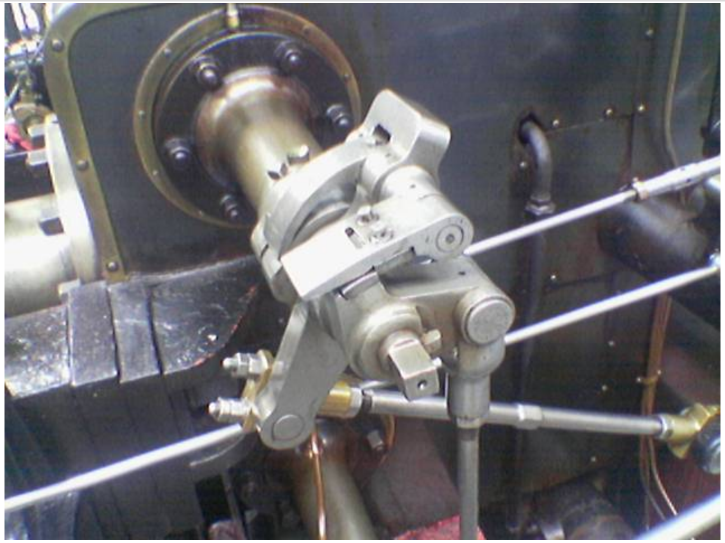
- Its magnitude is given by: $\|\mathbf{G}(x,y)\| = \sqrt{I_x^2(x,y) + I_y^2(x,y)}$

  or its approximations:

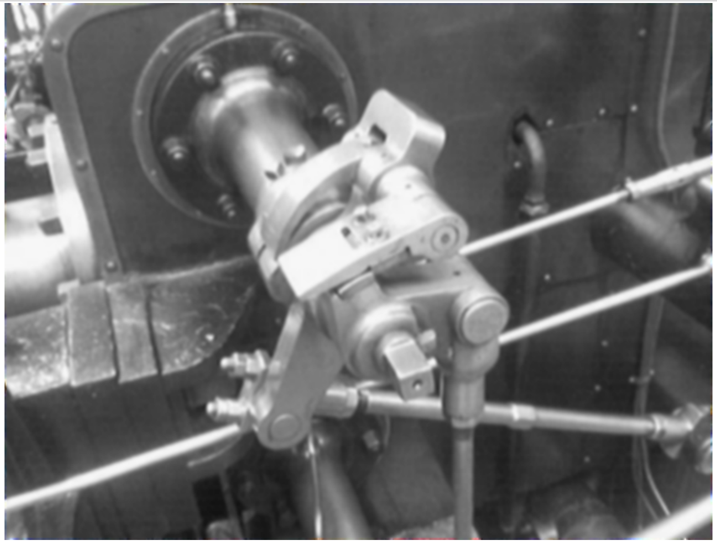
$$\|\mathbf{G}(x,y)\| \approx |I_x(x,y)| + |I_y(x,y)|$$

$$\|\mathbf{G}(x,y)\| \approx \max(I_x(x,y), I_y(x,y))$$

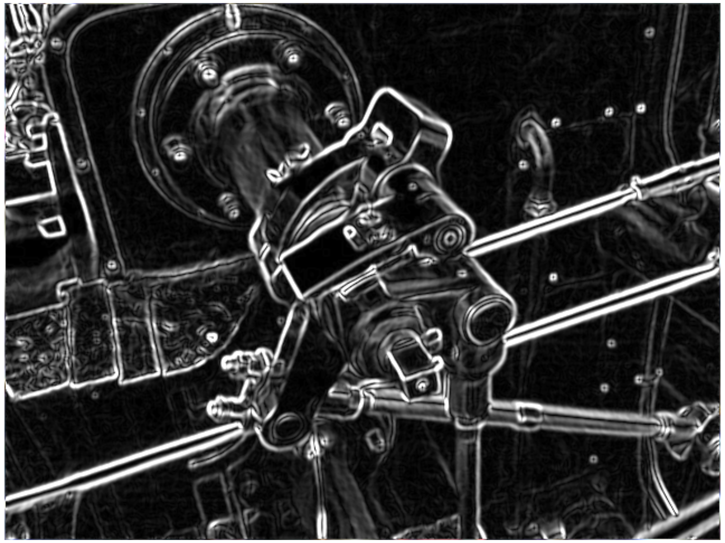
Gradient-Based Edge Detector Example



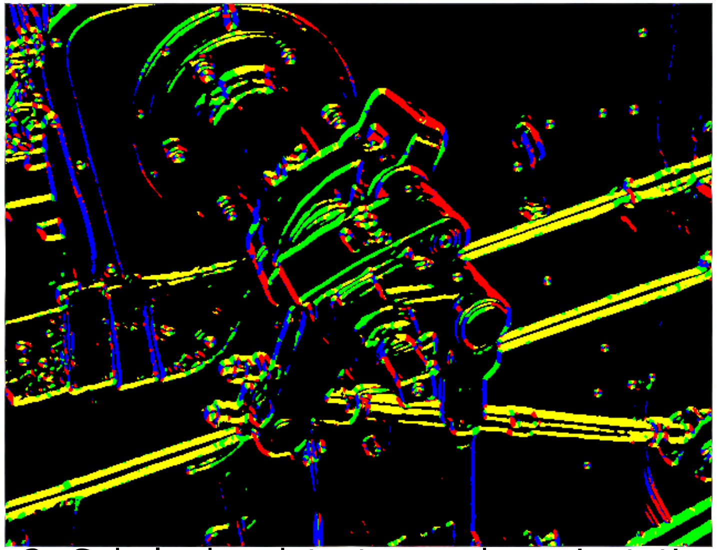
Original image



Step 1: Conversion to grayscale and smoothing with 5x5 Gaussian



Step 2: Sobel edge detector – edge magnitude image



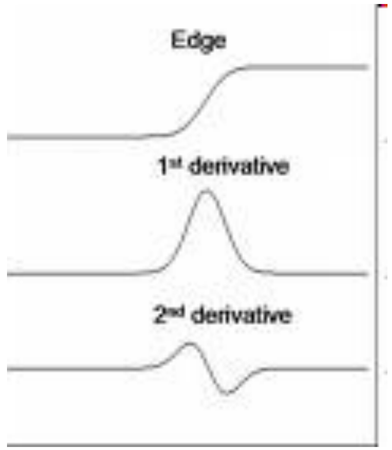
Step 2: Sobel edge detector – edge orientation image

Second Order Derivative

- Another way to detect an extremal first derivative is to look for a zero-valued 2nd derivative.

- A popular calculus tool that gives the magnitude of change in a bivariate function without direction information is the Laplacian.

$$\nabla^2(I(x,y)) = \left(\frac{\partial^2 I(x,y)}{\partial x^2} + \frac{\partial^2 I(x,y)}{\partial y^2} \right)$$



- Note that the result of the Laplacian is a scalar.

Laplacian Implementation

- Again differentiation is approximated by finite differencing.

$$
\begin{aligned}
\partial I^2(x,y)/\partial x^2 &= \partial(I_x(x,y))/\partial x \\
&= \partial(I[x,y] - I[x-1,y])/\partial x \\
&= \partial(I[x,y])/\partial x - \partial(I[x-1,y])/\partial x \\
&= (I[x+1,y] - I[x,y]) - (I[x,y] - I[x-1,y]) \\
&= I[x+1,y] - 2I[x,y] + I[x-1,y]
\end{aligned}
$$

- Written as a mask, we get:

$$
H_x = {}^2 I_x = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}
$$

Laplacian Implementation

- Similarly, for the 2nd partial derivative with respect to y, we get:

$$H_y = {}^{2}I_y = \begin{bmatrix} 0 & +1 & 0 \\ 0 & -2 & 0 \\ 0 & +1 & 0 \end{bmatrix}$$

- By adding the two together, we get the Laplacian mask:

$$H_{Lap} = {}^{2}I_x + {}^{2}I_y = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- If we want to use all 8 neighbors, we can use:

$$H_{Lap} = \begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix}$$

Simple Laplacian Example

- When we convolve an image that contains a significant change in values (i.e. edge) with a Laplacian kernel, we get a new image with negative values on one side of the edge and positive values on the other side of the edge.

- For example:

Input image

2	2	2	2	2	2	8	8	8	8
2	2	2	2	2	2	8	8	8	8
2	2	2	2	2	2	8	8	8	8
2	2	2	2	2	2	8	8	8	8
2	2	2	2	2	2	8	8	8	8
2	2	2	2	2	2	8	8	8	8

Image after the Laplacian

0	0	0	0	0	6	−6	0	0	0
0	0	0	0	0	6	−6	0	0	0
0	0	0	0	0	6	−6	0	0	0
0	0	0	0	0	6	−6	0	0	0
0	0	0	0	0	6	−6	0	0	0
0	0	0	0	0	6	−6	0	0	0

zero crossing

Laplacian of Gaussian

- The computation of 2nd order derivatives is very sensitive to noise.
- Solution: Smooth first the image I with a Gaussian H_{Gauss} and then apply the Laplacian H_{Lap} on the image.

$$R_{LapEdge} = H_{Lap} * (H_{Gauss} * I)$$

- Convolution is associative.

$$R_{LapEdge} = (H_{Lap} * H_{Gauss}) * I$$

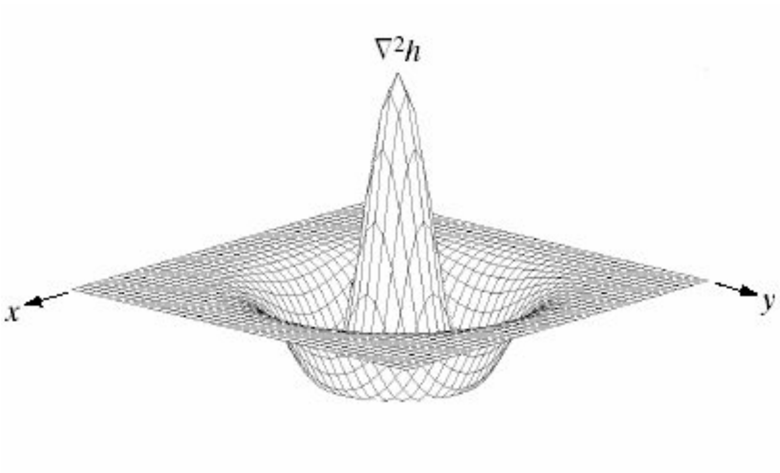
- The combined filter (H_{Lap} * H_{Gauss}) is nothing more than computing the Laplacian of the Gaussian (LoG):

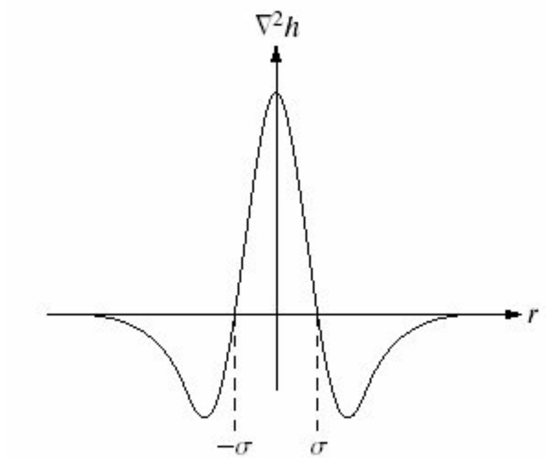
$$\nabla^2(G_{auss}(x,y)) = \nabla^2(\mathrm{e}^{(-(x^2+y^2)/2\sigma^2})$$

$$= \frac{(x^2+y^2-\sigma^2)}{\sigma^4}(\mathrm{e}^{(-(x^2+y^2)/2\sigma^2})$$

LoG Kernel

- The LoG function, $\nabla^2(G_{auss}(x,y))$ looks like a "mexican hat".

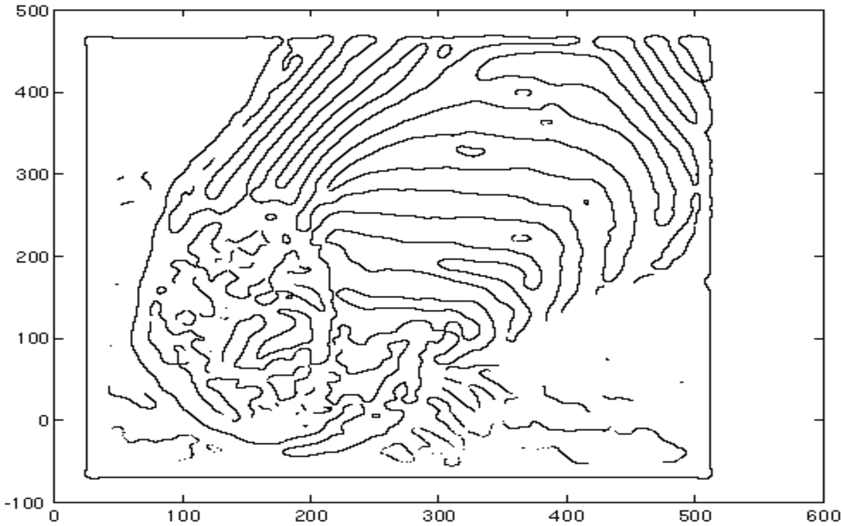




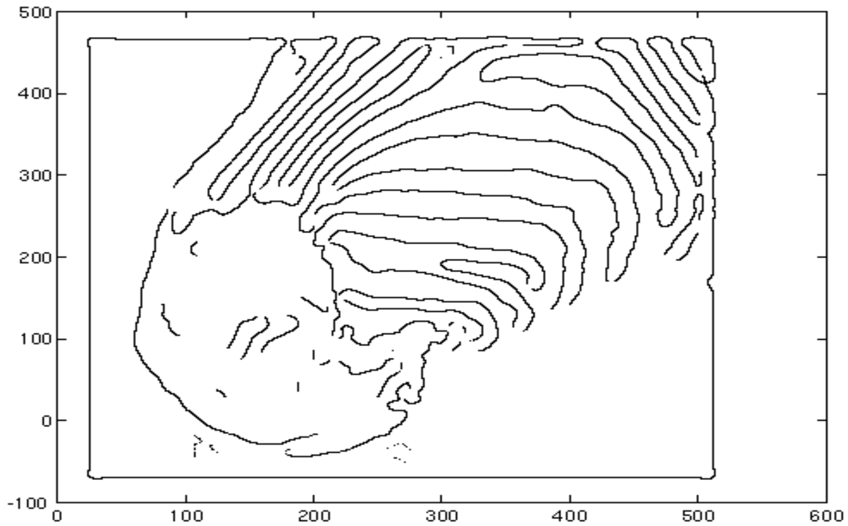
- $\nabla^2(G_{auss}(x,y))$ can also be approximated by a convolution kernel:

$$H_{LoG} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}$$

Examples of LoG Zero Crossings

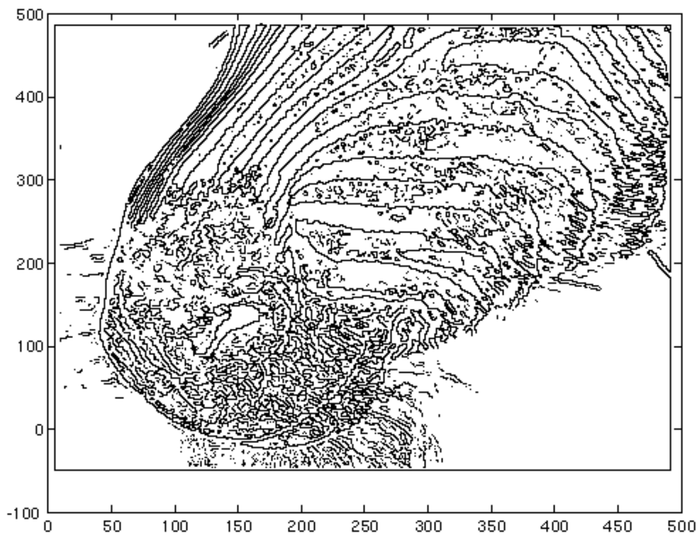


$\sigma = 4$

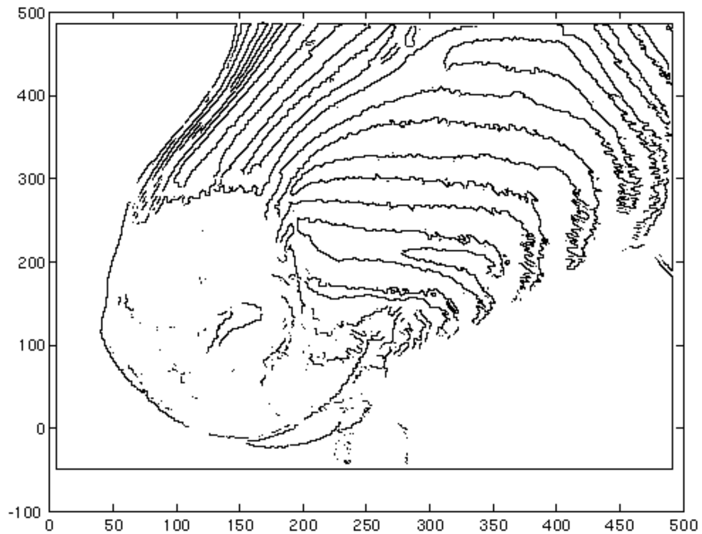


contrast=1

contrast=4



$\sigma = 2$



Smoothing and Differentiation

- The concepts of first smoothing and then differentiating generalizes to all edge detection methods (both 1st and 2nd order derivative methods).

- Convolution is associative, so we can always create a combined filter and convolve (filter) the image only once.

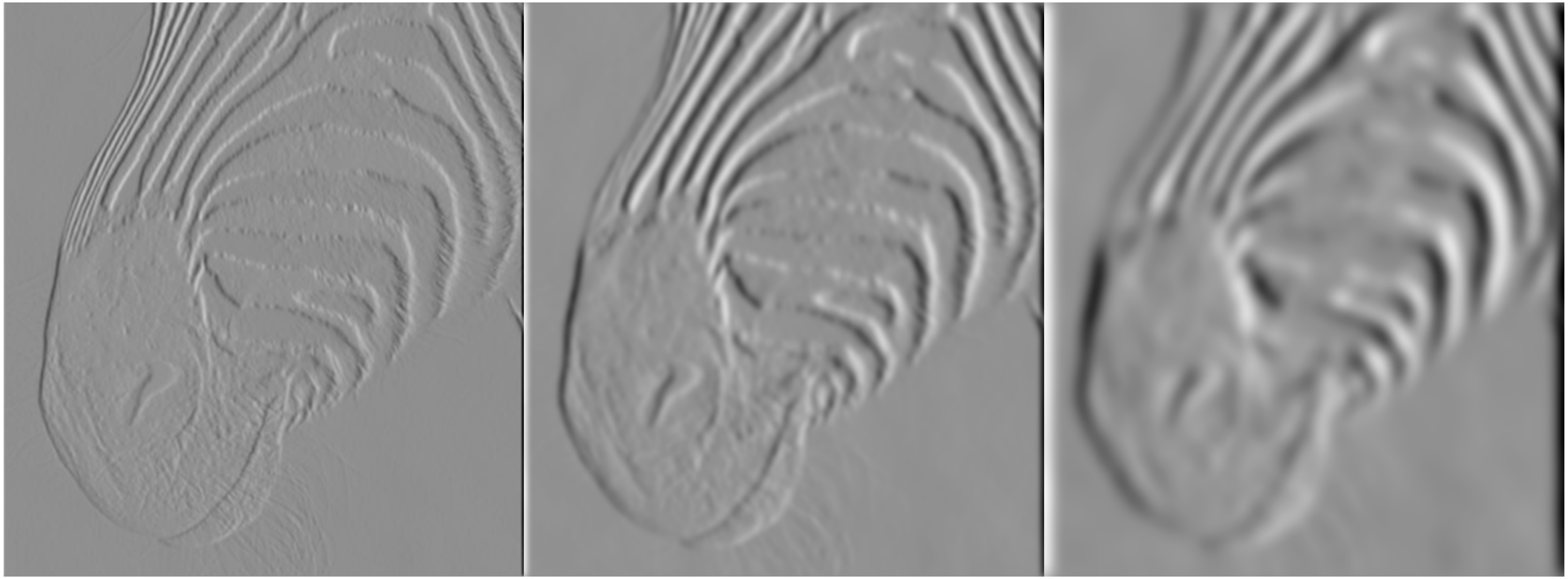
$$R = H_{edge} * (H_{smooth} * I) = (H_{edge} * H_{smooth}) * I = H * I$$

where $H = H_{edge} * H_{smooth}$

- By using different degrees of smoothing (Gaussian with different σ values or mean filters of different sizes, i.e. 3x3, 5x5, 7x7, etc.) we can obtain a hierarchy, a pyramid, of images with different levels of detail.

Different Scales

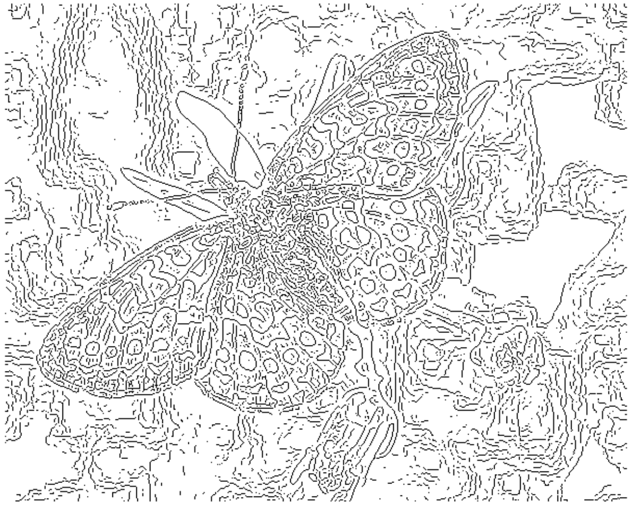
- The scale of the smoothing filter affects the derivative estimates as well as the semantics of the recovered edges



No smoothing 3x3 filter 7x7 filter

Different Scales

Original image



Fine scale, high threshold

Coarse scale, low threshold

Coarse scale, high threshold

Comments on Filtering

- Design Decisions:
  - Size of filter. There is no single good size. It depends on he size of the objects in the image.
  - Speed versus accuracy: (Gaussian vs. Median, Gradient-based vs. Laplacian-based, Canny vs. Sobel)
- Systematic approach: try different resolutions
  - Either create a formal model for each resolution and study the change of the model at different resolutions.
  - Or maintain a tree (pyramid) of images at different resolutions.
- Multi-resolution example:

  Apply an edge detector at different resolutions of Gaussians.

  Perform numerical optimization to find the best response for the particular image.

  Optimal for edges corrupted by white noise.

Gaussian Pyramid Example

Sharpening

- A very common filtering operation for contrast enhancement in images is *image sharpening*.
- The goal of image sharpening is to produce a more visually pleasing image:
  - Texture and finer details are made more prominent
  - The image looks sharper, crisper.

Sharpening - continued

- Image sharpening almost always involves improving the parts of the image where a sudden change in intensity or color occur, since this is where inaccuracies are introduced by the digital data capturing process.
- What filtering operation do we know that gives a high response at sudden changes in intensity or color?
- Edge Detector, H_{edge}
- A simple way to achieve sharpening is to superimpose the original image with the magnitude of the edge image.

$$R = I + c(I * H_{edge})$$

UnSharp Mask

- Most image processing software packets perform sharpening using the UnSharp Mask (USM).
- It is based on an old photographic film technique.
- It is called unsharp masking, because it first blurs the image (unsharpens it)

$$R_1 = I * H_{smooth}$$

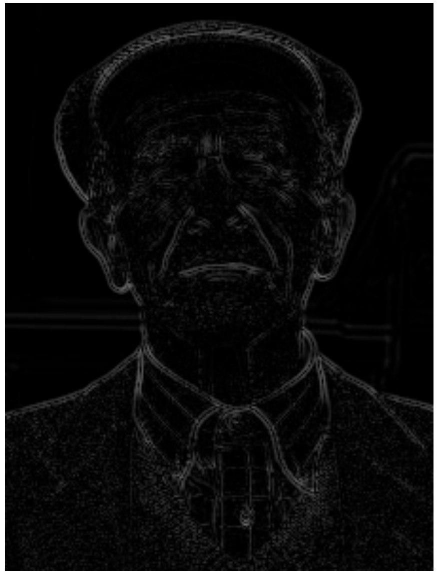
- An unsharp mask, UM, for the entire image is created by thresholding the absolute difference of the original and the blurred image.

$$UM(x,y) = \begin{cases} 1 & \text{if } \left| I(x,y) - R_1(x,y) \right| > \theta \\ 0 & \text{otherwise} \end{cases}$$

UnSharp Mask - continued

- The unsharp mask is then scaled (to achieve the desired visual effect) and added to the original image. The scaling factor *c* is often called *amount*.

$$R_2 = I + cUM$$



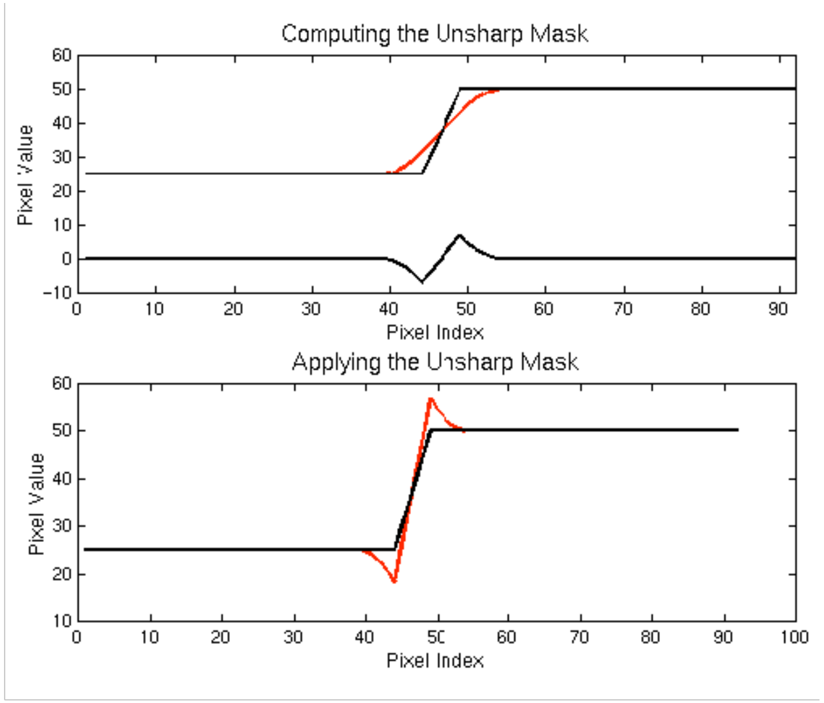


Image Sources

1. “Image with salt & pepper noise”, Marko Meza.
2. “Set of images of Roberts vs. Canny vs. Sobel”, Hypermedia Image Processing Reference at the University of Edinburgh.
3. “LoG plots”, Simon Yu Ming, http://hi.baidu.com/simonyuee/blog/item/446a911bf43cc91c8618bf8f.html
4. Many of the smoothing and edge detection images are from the slides by D.A. Forsyth, University of California at Urbana-Champaign.
5. The bird sharpening example was done using Adobe Photoshop Lightroom, http://mansurovs.com/how-to-properly-sharpen-images-in-lightroom.
6. The unsharp mask example is copyrighted by Sean T. McHugh, http://www.cambridgeincolour.com/tutorials/unsharp-mask.htm