ECE 661
Purdue Univ.
Avi Kak
Lecture 8

Visual Perception and Edge Detection

References: ① "Digital Picture Processing", Vol. 1, by Rosenfeld and Kak
② "Three-Dimensional Computer Vision" by Faugeras

The slides shown in class to illustrate the properties of the human visual system are from Chapter 3 of the Rosenfeld & Kak reference

- Practically all information in an image is encoded in the form of gray level variations — contrast differences — in the image.
- The material that follows consists of some useful-to-know and, at the same time, peculiar — and occasionally illusory — aspects of the human visual system with regard to detecting contrast differences in images.
- With regard to the detection of contrast differences, the first question to raise is: What's the smallest gray level difference discernible to the human eye? In other words, suppose the gray levels on the two sides of an edge are I and $I + \delta I$, what's the smallest value of δI that would be discernible to the human eye? Unfortunately, there is no simple answer to this question. The smallest discernible δI depends on several factors, as you'll see in what follows.
- The larger the value of I, the larger δI needs to be in order for it to be discernible. This is known as Weber's Law. The law says that the response of the human visual system is logarithmic. If R denotes the response, by Weber's Law, $\boxed{R \propto \log I}$, implying that $\boxed{\delta R \propto \frac{\delta I}{I}}$. For any given value for δR, this formula tells us that the larger the I, the greater δI must also be in order to be discernible.
- The smallest possible value for δI depends significantly on adaptation, meaning the ambient lighting that the human visual system has become accustomed to. At one end of the spectrum you get what's known as dark adaptation by placing a human in a completely dark room for about an hour. With this adaptation the value of δI will be the smallest (and the human will experience a diminished ability to see colors). Vision under dark adaptation is referred to as scotopic vision. At the other end of the spectrum, you have photopic vision.
- Another factor that complicates our perception of contrast differences is that our visual system does NOT possess an absolute sense about gray levels in images. Given two image regions of exactly the same gray level, one may appear brighter or darker than the other depending upon the gray levels in the adjoining regions. This is known as the Phenomenon of Simultaneous Contrast. I'll illustrate it in class with slides of images.
- Another way to illustrate "Simultaneous Contrast" is through a Benussi Ring. The ring consists of pixels of exactly the same gray level. I'll show how by drawing a line that appears to cut the ring in two halves, the two parts suddenly look to be of different gray levels.

- Whether or not you can see a contrast difference also depends on the sizes of the image regions with different gray levels. This spatial-resolution related aspect of our visual system is referred to as our visual acuity. Our visual acuity is the greatest in the central part of ~~our~~ the eye's visual field — what we see in this part of the visual field is referred to as foveal vision — and falls off rapidly toward the periphery. [Interestingly, for the same spatial resolution, the human eye is much more sensitive to a given contrast difference δI in the peripheral vision than in the foveal vision.]

- One of the best ways to illustrate acuity is to show to a human observer a 2D pattern in which the gray levels in the first row vary sinusoidally with a logrithmically increasing frequency. The different rows show the same pattern but with decreasing amplitude for the sinusoidal variations. If x is the horizontal axis and y the vertical, such a pattern is described by $\boxed{A(y) \sin(2\pi f(x) x)}$ with the frequency $\boxed{f(x) \propto \log x}$ and $\boxed{A(y) \propto \frac{1}{y}}$. Such a pattern will be shown in class. It is placed at a distance from the observer so that the frequences subtended at the eye correspond to the numbers you will see below the pattern. From such experiments, one concludes that, in the foveal vision, the human visual system is most sensitive to sinusoidal variations around 5 to 10 cycles per degree. The envelope that one perceives of such a pattern is the Modulation Transfer Function (MTF) of the eye. The MTF tells us that the human visual system ~~is~~ acts as a bandpass filter.

- The bandpass nature of the human visual system implies that it should have a natural tendency to enhance edges in images. This in indeed the case, as you'll see in class. These illusory bands are known as Mach bands.

- Another illusory phenomenon exhibited by the human visual system is known as Subjective Contours. One sees edges in portions of an image even when there are NO real edges there. This is caused by our compelling need to make sense of what we see.

So far we talked about perception of contrast differences. Next we talk about perception of patterns.

- That brings us to our final topic in Visual Perception: The Gestalt Laws of Perceptual Organization. These laws address our ability to group together the more elementary objects in images for the formation of higher-level constructs. The most important of these laws are: ① Law of Similarity; ② Law of Proximity; ③ Law of Good Continuation; and ④ Law of Closure. The law of similarity says that similar things tend to group together in our vision. Here "similar" is to be understood in terms of properties such as brightness, color, slope, size, etc. The Law of Proximity says that closely clustered entities tend to group together. The law of good continuation says that when curves cross or branch, parts that represent smooth continuations group together. The law of closure tells us that closed figures tend to be seen as units, unless the perception is overridden by one of the other laws. I'll show examples of these laws in class.

- When the Gestalt laws allow for more than one interpretation for figure-ground separation, only one of those interpretations leaps at you at any given time. We can force our perception to see a different interpretation through a cognitive mechanism called the Gestalt Switch. I'll illustrate this in class with the Rubin's Vase and the Kopfermann Cube examples.

- Note that Gestalt Laws also apply to patterns that change with time

Edge Detection With the Sobel Operator

- Edge detection is fundamentally about detecting sudden changes in the gray levels. The simplest way to do so is to calculate at each pixel the gradient vector of the gray levels: $\boxed{\nabla f(x,y) = \hat{x}\frac{\partial f}{\partial x} + \hat{y}\frac{\partial f}{\partial y}}$
The magnitude (the norm) of the gradient, $\boxed{\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}}$, gives us the strength of the edge at that point and $\boxed{\theta = \tan^{-1}\left(\frac{\partial f}{\partial y} \big/ \frac{\partial f}{\partial x}\right)}$ the direction of the normal to the edge. [The direction of the gradient is always along the max rate of change.]
- Different digital approximations to the calculation of the gradient give us different edge detection operators, the most popular being the Sobel Operator. For Sobel, you implement the $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ calculations by $\boxed{\text{convolving}}$ an image pixel array with the 3x3 masks $\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$, respectively

- Apply the Sobel Operator to the image: →
Your output should consist of two 10 x 10 arrays, one for the magnitude of the gradient and the other for the direction of the edge (make sure you add 90° to θ to get the direction of the edge). Apply a threshold to the magnitude values for the final edge image.

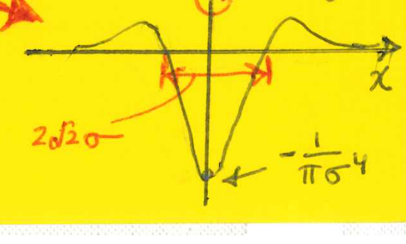
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	1	1	0	0	1	1	0	0
0	0	1	1	0	0	1	1	1	0
0	0	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0

Edge Detection With the LoG Operator

- LoG stands for Laplacian-of-Gaussian.

- Every edge operator must include some smoothing to counteract the noise enhancing property of derivative calculations. [Can you see what part of Sobel amounts to smoothing?]

- For LoG, the smoothing consists of convolving the image $f(x,y)$ with a Gaussian function $g(x,y)$. That is, we compute $\boxed{\iint f(x',y')\, g(x-x', y-y')\, dx'dy'}$ where $\boxed{g(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}}$
And, the derivative operation consists of applying the isotropic Laplacian operator $\boxed{\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}}$ to the smoothed data. The edges correspond to the zero-crossings of the Laplacian operator output.
- In practice, we combine the smoothing and the derivative operations into a single operator $\boxed{h(x,y) = \left\{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right\} g(x,y) = \frac{-1}{2\pi\sigma^4}\left(2 - \frac{x^2+y^2}{\sigma^2}\right) e^{-\frac{x^2+y^2}{2\sigma^2}}}$ →



LoG in 1-D:  edge $\xrightarrow[\frac{1}{\sqrt{2\pi}} e^{-x^2/2\sigma^2}]{\text{smoothed w/}}$  $\xrightarrow{\partial^2/\partial x^2}$ 

- The digital approximation to the LoG operator is expressed as $\boxed{h(m,n) = A\left(1 - k\frac{m^2+n^2}{\sigma^2}\right)e^{-\frac{m^2+n^2}{2\sigma^2}}}$ with $\boxed{-N \leq m,n \leq N}$, giving us a $(2N+1)\times(2N+1)$ operator. As you know, σ controls the extent of smoothing. To set σ, we assume a unit sampling interval between the pixels. As shown in the figure at the bottom of the previous page, the radius of the central lobe in the xy-plane is $\sqrt{2}\sigma$. So if we set $\sigma = \sqrt{2}$, the radius will be roughly 2 pixels. Even though the analytical $h(x,y)$ extends over the entire xy-plane, we truncate $h(m,n)$ at 3 times the radius, that is at $3\sqrt{2}\sigma$. So when $\sigma=\sqrt{2}$, we get $N=6$, which results in a 13×13 operator. Next, we set k so that the normalization condition $\boxed{\sum\sum_{m,n=-N}^{N} h(m,n) = 0}$ is satisfied. Finally, we set A so that the integer quantization of the values of $h(m,n)$ retain the essential properties of the operator. Note when $\sigma = \sqrt{2}$ $\boxed{|h(0,0)| = \frac{1}{\pi\sigma^4} = 0.08}$. So a straightforward integer quantization would cause all of the operator to disappear. We make A large enough so that the farthest values of h are close to zero.

Digital Laplacian with no smoothing

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

Edge Detection With the Canny Operator

- The Canny edge detector is derived by optimizing a convolutional function $h(x,y)$ with respect to three criteria: ① the SNR (Signal-to-Noise Ratio) at the true location of the edge; ② the localization of the detected edge vis-a-vis that of the true edge; and ③ the distance between where the true edge is detected and the nearest spurious edge (we want to maximize this distance, obviously).
- The three criteria are best cast as an exercise in 1-D optimization and then the result extended to 2D. For 1-D, we seek an operator $h(x)$ that when convolved with a 1-D signal $f(x)$ would return a function whose peaks would correspond to the locations of the jump discontinuities in $f(x)$ subject to the optimization of the 3 criteria above.
- We model $f(x)$ as $\boxed{f(x) = A\,U(x) + n(x)}$ where $U(x)$ is a step function and $n(x)$ the input noise. Assuming $n(x)$ to be zero-mean white, we have $\boxed{E\{n(x)\}=0}$ and $\boxed{E\{n(x)\,n(x+\tau)\} = \sigma_0^2\,\delta(\tau)}$ where $\delta(\tau)$ is the Dirac-delta function and σ_0^2 the noise variance.
- Since $h(x)$ is a convolutional operator, we have $\boxed{f_{out}(x) = \int_{-\infty}^{\infty} f(x-\alpha)\,h(\alpha)\,d\alpha = A\int_{-\infty}^{x} h(\alpha)\,d\alpha + \int_{-\infty}^{\infty} n(x-\alpha)\,h(\alpha)\,d\alpha}$ on account of $U(x)$. We can write this for $x=0$ as $f_{out}(0) = S + N$ where $\boxed{S = A\int_{-\infty}^{0} h(\alpha)\,d\alpha}$ is the signal at the true location of the edge and $\boxed{N = \int_{-\infty}^{\infty} n(-\alpha)\,h(\alpha)\,d\alpha}$ the noise at the same point. We have $E\{N\}=0$ and $E\{N^2\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\{n(-\alpha)\,n(-\beta)\}\,h(\alpha)\,h(\beta)\,d\alpha\,d\beta = \sigma_0^2\int_{-\infty}^{\infty} h^2(\alpha)\,d\alpha = 2\sigma_0^2\int_{-\infty}^{0} h^2(\alpha)\,d\alpha$ where the last equality follows from recognizing that $h(x)$ must be a derivative-like operator that has the property $\boxed{h(x) = -h(-x)}$.
- We now define $\boxed{SNR = \frac{S}{(E\{N^2\})^{1/2}} = \frac{A}{\sigma_0}\frac{\int_{-\infty}^{0} h(\alpha)\,d\alpha}{\left(\int_{-\infty}^{0} h^2(\alpha)\,d\alpha\right)^{1/2}}}$. Expressing SNR as $\frac{A}{\sigma_0}\Sigma(h)$, we get for the first of the three criteria to be optimized: $\boxed{\Sigma(h) = \frac{\int_{-\infty}^{0} h(\alpha)\,d\alpha}{\left[\int_{-\infty}^{0} h^2(\alpha)\,d\alpha\right]^{1/2}}}$
- For the second criterion we want find x where $f_{out}(x)$ has its first peak. That is, we want to find x_0 where $\boxed{\frac{\partial f_{out}(x)}{\partial x} = 0}$. We then want to maximize $\boxed{\frac{1}{E\{|x_0|^2\}}}$. This leads to the criterion: $\boxed{\Lambda(h) = \frac{|h'(0)|}{\left[\int_{-\infty}^{0} h'^2(\alpha)\,d\alpha\right]^{1/2}}}$
- Similarly, the third criterion result in $\boxed{X(h) = 2\pi\left[\frac{\int_{-\infty}^{0} h'^2(\alpha)\,d\alpha}{\int_{-\infty}^{0} h''(\alpha)\,d\alpha}\right]^{1/2}}$