Digital Image Fundamentals

Elements of visual perception

- Any artificial system is benchmarked against the human visual system
- Structure of the human eye (Figure 2.1)
  - Nearly a sphere with an average diameter of approximately 20mm
  - Enclosed by various membranes
  - Cornea and sclera outer cover
    - Cornea: Transparent convex anterior portion of the outer fibrous coat of the eyeball that covers the iris and the pupil and is continuous with the sclera
    - Sclera: Tough white fibrous outer envelope of tissue covering all of the eyeball except the cornea
  - Choroid
    - Dark-brown vascular coat of the eye next to sclera towards the interior of the eye
    - Network of blood vessels to provide nutrition to eye
    - Heavily pigmented to reduce the amount of extraneous light entering the eye, and backscatter within optical globe
    - Divided into ciliary body and iris diaphragm
    - Iris expands or contracts to control the amount of light entering the eye
    - Central opening of iris (pupil) varies from 2 to 8 mm in diameter
    - Front of iris contains the visible pigment of the eye whereas the back contains the black pigment
  - Lens
    - Made up of concentric layers of fibrous cells
    - Suspended by fibers attached to the ciliary body
    - Colored by slightly yellow pigmentation that increases with age
    - Excessive clouding of lens, called cataract, leads to poor color discrimination and loss of clear vision
    - Absorbs about 8% of visible light, and most of IR and UV waves, possibly causing damage in excessive amounts
  - Retina
    - Delicate, multilayered, light-sensitive membrane lining the inner eyeball and connected by the optic nerve to the brain
    - Distribution of discrete light receptors allows for pattern discrimination
    - Cones
      - 6 to 7 million photoreceptors, packed at about 150,000 cones per sq mm in the center of fovea
      - Located in central portion of retina (fovea)
      - Highly sensitive to color
      - Each one is connected to its own nerve end, allowing for fine resolution detail
      - Responsible for photopic or bright light vision
    - Rods
      - 75 to 150 million over the retinal surface with several rods connected to a single nerve end, providing coarse resolution
      - General overall view of the scene
      - Not involved in color vision and sensitive to low levels of illumination
      - Responsible for scotopic or low light vision
- Image formation in eye
  - Flexible lens

- Lens changes refractive power depending on the distance of object, by flexing the fibers in ciliary muscles
- Retinal image reflected on the fovea
- Perception involves transformation of radiant energy into electrical impulses decoded by brain

- Brightness adaptation and discrimination
  - Digital images displayed as a discrete set of intensities
  - Range of human eye is about 10^{10} different light intensity levels, from scotopic threshold to the glare limit
  - Subjective brightness (as perceived by humans) is a logarithmic function of light intensity incident on the eye (Fig 2.4)
  - Visual system cannot operate over the enormous range simultaneously
  - Brightness adaptation
    - Change in overall sensitivity of perceived brightness
    - Number of distinct intensity level that can be perceived simultaneously is small compared to number of levels that can be perceived
    - Brightness adaptation level – current sensitivity level of the visual system
  - Weber ratio
    - Measure of contrast discrimination ability
    - Background intensity given by I
    - Increment of illumination for short duration at intensity ΔI (Fig 2.5)
    - ΔI_c is the increment of illumination when the illumination is visible half the time against background intensity I
    - Weber ratio is given by $\Delta I_c/I$
    - A small value of $\Delta I_c/I$ implies that a small percentage change in intensity is visible, representing good brightness discrimination
    - A large value of $\Delta I_c/I$ implies that a large percentage change is required for discrimination, representing poor brightness discrimination
    - Typically, brightness discrimination is poor at low levels of illumination and improves at higher levels of background illumination (Fig 2.6)
  - Mach bands
    - Scalloped brightness pattern near the boundaries shown in stripes of constant intensity (Figure 2.7)
    - The bars themselves are useful for calibration of display equipment
  - Simultaneous contrast
    - A region's perceived brightness does not depend simply on intensity
    - Lighter background makes an object appear darker while darker background makes the same object appear brighter (Fig 2.8)
  - Optical illusions
    - Show the compensation achieved by eye
    - Figure 2.9

Light and electromagnetic spectrum

- Visible light vs the complete spectrum
- Wavelength (λ) and frequency (ν) are related using the constant c for speed of light

$$\lambda = \frac{c}{\nu}$$

- Energy of various components in EM spectrum is given by

$$E = h\nu$$

  where h is Planck's constant

- Units of measurements
  - Frequency is measured in Hertz (Hz)
  - Wavelength is measured in meters; also microns (μm or 10^{-6}m) and nanometers (10^{-9}m)
  - Energy is measured in electron-volt
- Photon
  - Massless particles whose stream in a sinusoidal wave pattern forms energy
  - Energy is directly proportional to frequency
    * Higher frequency energy particles carry more energy
    * Radio waves have less energy while gamma rays have more energy, making gamma rays more dangerous to living organisms
- Visible spectrum
  - 0.43μm (violet) to 0.79μm (red)
  - VIBGYOR regions
  - Colors are perceived because of light reflected from an object
  - Absorption vs reflectance of colors
    * An object appears white because it reflects all colors equally
- Achromatic or monochromatic light
  - No color in light
  - Amount of energy describes *intensity*
    * Quantified by *gray level* from black through various shades of gray to white
  - Monochrome images also called gray scale images
- Chromatic light
  - Spans the energy spectrum from 0.43 to 0.79 μm
  - Described by three basic quantities: radiance, luminance, brightness
  - Radiance
    * Total amount of energy flowing from a light source
    * Measured in Watts
  - Luminance
    * Amount of energy perceived by an observer from a light source
    * Measured in lumens (lm)
  - Brightness
    * Subjective descriptor of light perception
    * Achromatic notion of intensity
    * Key factor in describing color sensation
  - Light emitted from a regular light bulb contains a lot of energy in the IR band; newer bulbs give the same amount of luminance with less energy consumption

- A new 26 watt bulb has the same luminance as an old 100 watt bulb

Image sensing and acquisition

- Illumination source and absorption/reflectance of objects in the scene
  - Illumination energy is reflected from or transmitted through objects
  - X-ray images
- Transform incoming energy into voltage by a combination of input electrical power and sensor material responsive to the type of energy being detected
- Output voltage waveform from sensor is digitized to get a discrete response
- Image acquisition using a single sensor
  - Exemplified by a photodiode
    - Outputs voltage waveform proportional to incident light
  - Selectivity can be improved by using filters
    - A green (pass) filter will allow only the green light to be sensed
  - 2D image acquired by relative displacement of the sensor in both x and y directions
  - Single sensor mounted on a axle to provide motion perpendicular to the motion of object being scanned; also called microdensitometer
  - Slow and relatively antiquated
- Image acquisition using sensor strips
  - Strips in the form of in-line arrangement of sensors
  - Imaging elements in one direction
  - Used in flat-bed scanners and photocopy machines
  - Basis for computerized axial tomography
- Image acquisition using sensor arrays
  - Individual sensors in the form of a 2D array
  - CCD array in video cameras
  - Response of each sensor is proportional to the integral of light energy projected onto the surface of sensor
  - Noise reduction is achieved by integrating the input light signal over a certain amount of time
  - Complete image can be obtained by focusing the energy pattern over the surface of array
- A simple image formation model
  - Denote images by 2D function $f(x, y)$
  - x and y are spatial coordinates on a plane and $f(x, y)$ is a positive scalar/vector quantity to represent the energy at that point
  - Image function values at each point are positive and finite

$$0 < f(x, y) < \infty$$

  - $f(x, y)$ is characterized by two components

    Illumination: Amount of source illumination incident on the scene being viewed; denoted $i(x, y)$

    Reflectance: Amount of illumination reflected by objects in the scene; denoted $r(x, y)$

  - The product of illumination and reflectance yields $f(x,y) = i(x,y) \cdot r(x,y)$ such that $0 < i(x,y) < \infty$ and $0 < r(x,y) < 1$
    * Reflectance is bounded by 0 (total absorption) and 1 (total reflectance)
    * For images formed by transmission rather than reflection (X-rays), reflectivity function is replaced by transmissivity function with the same limits

- Intensity of monochrome image at any coordinate is called the gray level l of the image at that point
  - The range of l is given by
$$L_{\min} \leq l \leq L_{\max}$$
  - The interval $[L_{\min}, L_{\max}]$ is called the gray scale
  - It is common to shift this interval to $[0, L-1]$ where $l = 0$ is considered black and $l = L - 1$ is considered white, with intermediate values providing different shades of gray

Image sampling and quantization

- Sensors output a continuous voltage waveform whose amplitude and spatial behavior are related to the physical phenomenon being sensed
  - Need to convert continuous sampled data to discrete/digital form using sampling and quantization
- Basic concepts in sampling and quantization
  - Figure 2.16
  - Continuous image to be converted into digital form
  - Image continuous with respect to x and y coordinates as well as amplitude
  - Sampling: Digitizing the coordinate values
  - Quantization: Digitizing the amplitude values
  - Issues in sampling and quantization, related to sensors
    * Electrical noise
    * Limits on sampling accuracy
    * Number of quantization levels
- Representing digital images
  - Continuous image function of two variables s and t denoted by $f(s,t)$
    * Convert $f(s,t)$ into a digital image by sampling and quantization
    * Sample the continuous image into a 2D array $f(r,c)$ with M rows and N columns, using integer values for discrete coordinates r and c: $r = 0, 1, 2, \ldots, M-1$, and $c = 0, 1, 2, \ldots, N-1$
    * Matrix of real numbers, with M rows and N columns
  - Concepts/definitions

    Spatial domain: Section of the real plane spanned by the coordinates of an image

    Spatial coordinates: Discrete numbers to indicate the locations in the plane, given by a row number r and column number c
  - Image representation (Fig. 2.18)
    * As a 3D plot of (x,y,z) where x and y are planar coordinates and z is the value of f at (x,y)
    * As intensity of each point, as a real number in the interval $[0,1]$
    * As a set of numerical values in the form of a matrix (we'll use this in our work)
      · We may even represent the matrix as a vector of size $MN \times 1$ by reading each row one at a time into the vector

- Conventions
  - Origin at the top left corner
  - c increases from left to right
  - r increases from top to bottom
  - Each element of the matrix array is called a pixel, for picture element
- Definition of sampling and quantization in formal mathematical terms
  - Let $\mathcal{Z}$ and $\Re$ be the set of integers and real numbers
  - Sampling process is viewed as partitioning the xy plane into a grid
    - Coordinates of the center of each grid are a pair of elements from the Cartesian product $\mathcal{Z}^2$
    - $\mathcal{Z}^2$ denotes the set of all ordered pairs of elements (z_i, z_j) such that $z_i, z_j \in \mathcal{Z}$
  - $f(r, c)$ is a digital image if
    - $(r, c) \in \mathcal{Z}^2$, and
    - f is a function that assigns a gray scale value (real number) to each distinct pair of coordinates (r, c)
    - If gray scale levels are integers, $\mathcal{Z}$ replaces $\Re$; image is a 2D function with integer coordinates and amplitudes
- Decision about the size and number of gray scales
  - No requirements for M and N, except that they be positive integers
  - Gray scale values are typically powers of 2 because of processing, storage, and sampling hardware considerations
$$L = 2^k$$
  - Assume that discrete levels are equally spaced and in the interval $[0, L-1]$ – dynamic range of the image
    - Dynamic range is the ratio of maximum measurable intensity to the minimum detectable intensity
    - Upper limit is determined by saturation and the lower limit is determined by noise
  - Contrast – Difference in intensity between the highest and lowest intensity levels in the image
  - High dynamic range – gray levels span a significant portion of range
  - High contrast – Appreciable number of pixels are distributed across the range
- Number of bits required to store an image – $M \times N \times k$
  - For $M = N$, this yields $M^2 k$
- 8-bit image

- Spatial and gray-level resolution
  - Spatial resolution determined by sampling
    - Smallest discernible detail in an image
    - Stated as line pairs per unit distance, or dots per unit distance
      - Construct a chart with alternate black and white vertical lines, each of width W units
      - Width of each line pair is $2W$, or $1/2W$ lines pairs per unit distance
      - Dots per inch is common in the US
    - Important to measure spatial resolution in terms of spatial units, not just as the number of pixels
    - Lower resolution images are smaller
  - Gray-level resolution determined by number of gray scales
    - Smallest discernible change in gray level
    - Most common number is 8 bits (256 levels)
  - Subsampling
    - Possible by deleting every other row and column (Fig 2-19)
    - Possible by averaging a pixel block

- Resampling by pixel replication (Fig 2-20)
- Changing the number of gray levels (Fig 2-21)
  * False contouring – Effect caused by insufficient number of gray scale levels
  * More visible in images with 16 or less intensity levels
- Amount of detail in an image (Fig 2-22)
  * Frequency of an image

• Image interpolation
  - Basic tool used extensively in tasks such as zooming, shrinking, rotating and geometric corrections
  - Process of using known data to estimate values at unknown locations
  - Zooming and shrinking considered as image resampling methods
    * Zooming $\Rightarrow$ oversampling
    * Shrinking $\Rightarrow$ undersampling
  - Zooming
    * Create new pixel locations
    * Assign gray levels to these pixel locations
    * Pixel replication
      · Special case of nearest neighbor interpolation
      · Applicable when size of image is increased by an integer number of times
      · New pixels are exact duplicates of the old ones
    * Nearest neighbor interpolation
      · Assign the gray scale level of nearest pixel to new pixel
      · Fast but may produce severe distortion of straight edges, objectionable at high magnification levels
      · Better to do *bilinear interpolation* using a pixel neighborhood
    * Bilinear interpolation
      · Use four nearest neighbors to estimate intensity at a given location
      · Let (r, c) denote the coordinates of the location to which we want to assign an intensity value $v(r, c)$
      · Bilinear interpolation yields the intensity value as

$$v(r, c) = C_1 c + C_2 r + C_3 rc + C_4$$

      where the four coefficients are determined from the four equations in four unknowns that can be written using the four nearest neighbors of point (r, c)
      · Better results with a modest increase in computing
    * Bicubic interpolation
      · Use 16 nearest neighbors of a point
      · Intensity value for location (r, c) is given by

$$v(r, c) = \sum_{i=0}^{3} \sum_{j=0}^{3} C_{ij} r^i c^i$$

      where the 16 coefficients are determined from the 16 equations in 16 unknowns that can be written using the 16 nearest neighbors of point (r, c)
      · Bicubic interpolation reduces to bilinear form by limiting the two summations from 0 to 1
      · Bicubic interpolation does a better job of preserving fine detail compared to bilinear
      · Standard used in commercial image editing programs
    * Other techniques for interpolation are based on splines and wavelets

- Shrinking
  * Done similar to zooming
  * Equivalent to pixel replication is row-column deletion
  * Aliasing effects can be removed by slightly blurring the image before reducing it

Basic relationships between pixels

- Neighbors of a pixel p
  - 4-neighbors $(N_4(p))$
    * Four vertical and horizontal neighbors for pixel p at coordinates (r, c) are given by the pixels at coordinates

      $$N_4(p) = \{(r+1, c), (r, c+1), (r-1, c), (r, c-1)\}$$

    * Each 4-neighbor is at a unit distance from p
    * Some neighbors may be outside of the image if p is a boundary pixel
  - 8-neighbors $(N_8(p))$
    * Non-uniform distance from p
    * Include $N_4(p)$ as well as the pixels along the diagonal given by

      $$N_D(p) = \{(r+1, c+1), (r-1, c+1), (r-1, c-1), (r+1, c-1)\}$$

    * Effectively, we have $N_8(p) = N_4(p) + N_D(p)$
- Adjacency, connectivity, regions, boundaries
  - Pixels are connected if they are neighbors and their gray scales satisfy a specified criteria of similarity
  - Adjacency
    * Defined using a set of gray-scale values V
    * $V = \{1\}$ if we refer to adjacency of pixels with value 1 in a binary image
    * In a gray scale image, the set V may contain more values
    * 4-adjacency
      · Two pixels p and q with values from V are 4-adjacent if $q \in N_4(p)$
    * 8-adjacency
      · Two pixels p and q with values from V are 8-adjacent if $q \in N_8(p)$
    * m-adjacency (mixed adjacency)
      · Modification of 8-adjacency
      · Two pixels p and q with values from V are m-adjacent if
        1. $q \in N_4(p)$, or
        2. $q \in N_D(p)$ and the set $N_4(p) \cap N_4(q)$ has no pixels whose values are from V
      · Eliminates the ambiguities arising from 8-adjacency (Fig 2-25)
  - Path
    * A digital path or curve from pixel $p(r, c)$ to $q(r', c')$ is a set of adjacent pixels from p to q, given by

      $$(r_0, c_0), (r_1, c_1), \ldots, (r_n, c_n)$$

      where $(r, c) = (r_0, c_0)$ and $(r', c') = (r_n, c_n)$ and pixels at (r_i, c_i) and (r_{i-1}, c_{i-1}) are adjacent
    * Length of the path is given by the number of pixels in such a path
    * Closed path, if $(r_0, c_0) = (r_n, c_n)$
    * 4-, 8-, or m- paths depending on the type of adjacency defined

- Connected pixels
  - Let S represent a subset of pixels in an image
  - Two pixels p and q are connected in S if there is a path between them consisting entirely of pixels in S
  - For any pixel p in S, the set of pixels connected to it in S form a *connected component* of S
  - If there is only one connected component of S, the set S is known as a *connected set*
- Region
  - Let R be a subset of pixels in the image
  - R is a region of the image if R is a connected set
  - Two regions R_i and R_j are *adjacent* if their union forms a connected set
  - Regions that are not adjacent are *disjoint*
  - Foreground and background
    - Let an image contain K disjoint regions $R_k, k = 1, 2, \ldots, K$, none of which touch the image border
    - Let R_u be the union of all the K regions and let $(R_u)^c$ be its complement
    - All the pixels in R_u form the *foreground* in image
    - All the pixels in $(R_u)^c$ form the image *background*
  - The boundary of a region R is the set of pixels in the region that have one or more neighbors that are not in R
    - The set of pixels within the region on the boundary are also called *inner border*
    - The corresponding pixels in the background are called *outer border*
    - This distinction will be important in border-following algorithms
    - If R is an entire rectangular image, its boundary is the set of pixels in the first and last rows and columns
    - An image has no neighbors beyond its borders
- Edge
  - Gray level discontinuity at a point
  - Formed by pixels with derivative values that exceed a preset threshold

- Distance measures
  - Properties of distance measure D, with pixels $p(r, c)$, $q(r', c')$, and $z(r'', c'')$
    - $D(p, q) \geq 0$; $D(p, q) = 0 \Leftrightarrow p = q$
    - $D(p, q) = D(q, p)$
    - $D(p, z) \leq D(p, q) + D(q, z)$
  - Euclidean distance
    $$D_e(p, q) = \sqrt{(r - r')^2 + (c - c')^2}$$
    Also represented as $||p - q||$
  - City-block distance (D_4 distance)
    $$D_4(p, q) = |r - r'| + |c - c'|$$
    Pixels with $D_4 = 1$ are 4-neighbors
  - Chessboard distance (D_8 distance)
    $$D_8(p, q) = \max(|r - r'|, |c - c'|)$$
    Pixels with $D_8 = 1$ are 8-neighbors
  - D_4 and D_8 distances are independet of any path between points as they are based on just the position of points
  - D_m distances are based on m-adjacency and depend on the shortest m path between the points

Mathematical tools for digital image processing

- Array vs matrix operations

- An array operation on images is carried on a per pixel basis
- Need to make a distinction between array and matrix operations
- Consider the following 2×2 images

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

- Array product of these two images is

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$$

- Matrix product is given by

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

- Assume array operations in this course, unless stated otherwise

• Linear vs nonlinear operations

- Consider a general operator H that produces an output image $g(r, c)$ for a given input image $f(r, c)$

$$g(r, c) = H[f(r, c)]$$

- H is a linear operator if

$$\begin{aligned} H[a_i f_i(r, c) + a_j f_j(r, c)] &= a_i H[f_i(r, c)] + a_j H[f_j(r, c)] \\ &= a_i g_i(r, c) + a_j g_j(r, c) \end{aligned}$$

where a_i and a_j arbitrary constants, and f_i and f_j are two images of the same size

- Linear operators have the following properties

Additivity Output of linear operator on the sum of two images is same as the sum of output of linear operator applied to those images individually

Homogeneity Output of linear operation to constant times an image is the same as constant times the output of linear operation to the images

- Let H be the summation operator, then, we have

$$\begin{aligned} \sum [a_i f_i(r, c) + a_j f_j(r, c)] &= \sum a_i f_i(r, c) + \sum a_j f_j(r, c) \\ &= a_i \sum f_i(r, c) + a_j \sum f_j(r, c) \\ &= a_i g_i(r, c) + a_j g_j(r, c) \end{aligned}$$

showing that summation operator is linear

- Now consider the max operation to get the maximum value of any pixel in the images, and the following two images

$$\begin{bmatrix} 0 & 2 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 6 & 5 \\ 4 & 7 \end{bmatrix}$$

Let $a_1 = 1$ and $a_2 = -1$

* The left hand side of the equation evaluates to

$$\begin{aligned} \max \left\{ (1) \begin{bmatrix} 0 & 2 \\ 2 & 3 \end{bmatrix} + (-1) \begin{bmatrix} 6 & 5 \\ 4 & 7 \end{bmatrix} \right\} &= \max \left\{ \begin{bmatrix} -6 & -3 \\ -2 & -3 \end{bmatrix} \right\} \\ &= -2 \end{aligned}$$

    * The right hand side evaluates to

$$\begin{aligned}(1)\max\left\{\begin{bmatrix}0 & 2\\ 2 & 3\end{bmatrix}\right\}+(-1)\max\left\{\begin{bmatrix}6 & 5\\ 4 & 7\end{bmatrix}\right\} &= 3+(-1)7\\ &= -4\end{aligned}$$

- Arithmetic operations
  - Array operations between corresponding pixel pairs

$$\begin{aligned}s(r,c) &= f(r,c)+g(r,c)\\ d(r,c) &= f(r,c)-g(r,c)\\ p(r,c) &= f(r,c)\times g(r,c)\\ v(r,c) &= f(r,c)\div g(r,c)\end{aligned}$$

    * The two input images as well as the output image are of the same size in each of these operations
  - Example: Addition of noisy images for noise reduction
    * Let $g(r,c)$ denote a corrupted image formed by the addition of uncorrelated noise $\eta(r,c)$ to a noiseless image $f(r,c)$

$$g(r,c)=f(r,c)+\eta(r,c)$$

    * Form an average image by averaging K different noisy images

$$\bar{g}(r,c)=\frac{1}{K}\sum_{i=1}^{K}g_i(r,c)$$

    * Since the noise is uncorrelated, the expected value is

$$E\{\bar{g}(r,c)\}=f(r,c)$$

    * The variances are related by

$$\sigma^2_{\bar{g}(r,c)}=\frac{1}{K}\sigma^2_{\eta(r,c)}$$

    * The standard deviation at any point in the average image is

$$\sigma_{\bar{g}(r,c)}=\frac{1}{\sqrt{K}}\sigma_{\eta(r,c)}$$

    * As K increases, the variance at each location (r,c) decreases
    * In practice, the images $g_i(r,c)$ must be *registered* for expected addition to approach $f(r,c)$
    * Image averaging as above is important in astronomy where images under low light level cause sensor noise to render single images virtually useless for analysis
      · Figure 2.26
    * Addition provides a discrete version of continuous integration
  - Image subtraction to enhance differences
    * Figure 2.27
    * Change detection via image subtraction
    * Mask mode radiography

$$g(r,c)=f(r,c)-h(r,c)$$

      · $h(x,y)$ is the mask or X-ray image of a patient's body captured by intensified TV camera, located opposite an X-ray source
      · Inject an X-ray contrast medium into a patient's bloodstream, taking a series of *live images*, and subtracting the mask from the live stream

        - Areas that are different between $f(r,c)$ and $h(r,c)$ appear in the output stream as enhanced detail
        - Over time, the process shows the propagation of contrast medium through various arteries
        - Figure 2.28
  - Image multiplication for shading correction
    - Let the sensor produce an image $g(r,c)$ that is product of a *perfect* image $f(r,c)$ with a shading function $h(r,c)$
    - If h is known, we can obtain $f(r,c)$ by dividing g by h
    - We can obtain an approximation to h by imaging a target of constant intensity
    - Figure 2.29
  - Image multiplication for masking or ROI operations
    - Figure 2.30
  - Pixel saturation
    - Most image representations used by us are in the range $[0, 255]$
    - Addition and subtraction may yield values in the range $[-255, 510]$
    - Change the minimum value of each pixel to 0

      $$f_m = f - min(f)$$

    - Scale the image in the range $[0, K]$ by

      $$f_s = K[f_m / \max(f_m)]$$

    - The discussion is applicable to images in ImageMagick that are in the range `[0,MAXRGB]`

- Set and logical operations
  - Basic set operations
    - Let A be a set composed of ordered pairs of real numbers
    - If $a = (a_1, a_2)$ is an element of A, we say $a \in A$
    - If a is not an element of A, we have $a \notin A$
    - The set with no elements in called null or empty set and is denoted by $\emptyset$
    - Set is specified by the contents of two braces: $\{\cdot\}$
    - $C = \{w | w = -d, d \in D\}$
    - Elements of sets could be coordinates of pixels (ordered pairs) representing regions (objects) in an image
    - If every element of A is also an element of B, then, $A \subseteq B$
    - Union of two sets A and B is denoted by $C = A \cup B$
    - Intersection of two sets A and B is denoted by $D = A \cap B$
    - Two sets A and B are disjoint or mutually exclusive if they have no common elements, or $A \cap B = \emptyset$
    - Set universe U is the set of all elements in a given application
      - If you are working with the set of real numbers, the set universe is the real line containing all real numbers
      - In image processing, the universe is typically the rectangle containing all pixels in an image
    - Complement of a set is the set of elements not in A

      $$A^c = \{w | w \notin A\}$$

    - Difference of two sets A and B, denoted $A - B$, is defined as

      $$A - B = \{w | w \in A, w \notin B\} = A \cap B^c$$

    - A^c can be defined in terms of universe as

      $$A^c = U - A$$

    - Figure 2.31 for operations with binary images

- Operations with gray scale images
  - Gray scale image pixels can be represented as a set of 3-tuples (r, c, m) where m is the magnitude and r, c are row and column number of pixels
  - Define complement of A as $A^c = \{(r, c, K - m) | (r, c, m) \in A\}$, and K is the maximum gray scale value
  - Now, the negative of an image is given by its complement
  - Union of two gray-scale sets is given by

$$A \cup B = \left\{ \begin{array}{c} \max \\ m \end{array} (a, b) | \, a \in A, b \in B \right\}$$

  - Figure 2.32
- Logical operations
  - Foreground (1-valued) and background (0-valued) sets of pixels
  - Regions or objects can be defined as composed of foreground pixels
  - Consider two regions A and B composed of foreground pixels
  - $\vee$, $\wedge$, and $\neg$ logical operations
  - $\neg A$ is the set of pixels in the image that are not in region A (background pixels and foreground pixels from regions other than A)
  - Figure 2.33

- Spatial operations
  - Performed directly on the pixels of a given image
  - Single pixel operations
    - Simplest operation to alter the value of individual pixels based on intensity
    - Expressed as a transform function of the form

$$s = T(z)$$

      where z is the intensity of a pixel in the original image and s is the intensity of the corresponding pixel in the processed image
    - Negative of an image
  - Neighborhood operations
    - Let S_{rc} denote the set of coordinates of a neighborhood centered at a point (r, c) in an image f
    - Generate a corresponding pixel at (r, c) by processing all the pixels in S_{rc}
    - Average value of pixels, centered at (r, c) where S_{rc} is delimited by a rectangle of size $m \times n$

$$g(r, c) = \frac{1}{mn} \sum_{(r', c') \in S_{rc}} f(r', c')$$

    - Local blurring (Figure 2.35) to eliminate small details
  - Geometric spatial transformations and image registration
    - Modify the spatial relationship between pixels in an image
    - Also called rubber-sheet transformations
    - Consists of two basic operations
      1. A spatial transformation of coordinates
      2. Intensity interpolation that assigns intensity values to spatially transformed pixels
    - The transform can be expressed as

$$(r, c) = T\{(r', c')\}$$

      where (r', c') are pixel coordinates in original image and (r, c) are corresponding pixel coordinates in the transformed image

    * The transformation $(r, c) = T\{(r', c')\} = (r'/2, c'/2)$ shrinks the original image to half its size in both directions
    * Affine transform to scale, rotate, translate, or shear a set of coordinate points depending on the value chosen for the elements of matrix $\mathbf{T}$

$$[r\ c\ 1] = [r'\ c'\ 1]\mathbf{T} = [r'\ c'\ 1] \begin{bmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & 1 \end{bmatrix}$$

    * Table 2.2 and Figure 2.36
    * Matrix representation allows us to concatenate together a sequence of operations
    * Above transformations allow us to relocate pixels in an image
    * We may also have to change intensity values at the new locations, possibly by intensity interpolation (zooming)
  - Image registration
    * Estimating the transformation function and use it to register the input and output images
    * The image against which we perform registration is called the reference image
    * Used when two images need to be aligned when two images of same object are taken at different time or with different sensors
    * Tie points or control points
      · Corresponding points whose locations are known precisely in the input images
      · Can be applied manually or detected automatically by sophisticated algorithms
      · Some sensors may produce a set of known points, called *reseau marks*, directly on images to be used as guides for tie points
    * Transformation function can be estimated based on modeling
      · Given a set of four tie points in an input image and a reference image
      · A simple model based on bilinear approximation gives

$$\begin{aligned} x &= c_1 v + c_2 w + c_3 vw + c_4 \\ y &= c_5 v + c_6 w + c_7 vw + c_8 \end{aligned}$$

      where (v, w) and (x, y) are coordinates of the points in the input and reference images, respectively
      · With four pairs of points, we can write eight equations and use them to solve for the eight unknown coefficients $c_1, c_2, \ldots, c_8$
      · The coefficients are the model to transform pixels of one image into locations of the other to achieve registration

- Vector and matrix operations
  - Used routinely in multispectral image processing
  - Each pixel in an RGB Image can be organized in the form of a column vector

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

  - An RGB image of size $M \times N$ can be represented by three component images of this size each, or by a total of MN 3D vectors
  - A general multispectral image with n component images will give us n-dimensional vectors
  - The Euclidean distance D between a pixel vector $\mathbf{z}$ and an arbitrary point $\mathbf{a}$ in n-dimensional space is defined by the vector product

$$\begin{aligned} D(\mathbf{z}, \mathbf{a}) &= \sqrt{(\mathbf{z} - \mathbf{a})^T (\mathbf{z} - \mathbf{a})} \\ &= \sqrt{(z_1 - a_1)^2 + (z_2 - a_2)^2 + \cdots + (z_n - a_n)^2} \end{aligned}$$

- D is sometime referred to as vector norm and may be denoted by $||\mathbf{z} - \mathbf{a}||$
- Pixel vectors are useful in linear transformations, represented as

$$\mathbf{w} = \mathbf{A}(\mathbf{z} - \mathbf{a})$$

where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{z}$ and $\mathbf{a}$ are column vectors of size $n \times 1$

- Image transforms
  - All the operations so far work directly on the pixels in *spatial domain*
  - Some operations may be done by transforming the image into a *transformation domain* and applying the inverse transform to bring it back to spatial domain
  - A 2D linear transform my be expressed in the general form as

$$T(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) r(x,y,u,v)$$

where $f(x,y)$ is the input image and $r(x,y,u,v)$ is a forward transformation kernel; the equation is evaluated for $u = 0, 1, 2, \ldots, M-1$ and $v = 0, 1, 2, \ldots, N-1$

  - The image can be transformed back to spatial domain by applying the inverse transform as

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} T(u,v) s(x,y,u,v)$$

  - Figure 2.39
- Probabilistic methods
  - We may treat intensity values as random quantities
  - Let $z_i = 0, 1, 2, \ldots, L-1$ be the values of all possible intensities in an $M \times N$ image
  - The probability $p(z_k)$ of intensity level z_k in the image is given by

$$p(z_k) = \frac{n_k}{MN}$$

where n_k is the number of pixels at intensity level z_k

  - Clearly

$$\sum_{k=0}^{L-1} p(z_k) = 1$$

  - The mean intensity of the image is given by

$$m = \sum_{k=0}^{L-1} z_k p(z_k)$$

  - The variance of intensities is

$$\sigma^2 = \sum_{k=0}^{L-1} (z_k - m)^2 p(z_k)$$

    * Variance is a measure of spread of values of z around the mean, so it is a useful measure of image contrast
  - nth moment of random variable z about the mean is defined as

$$\mu_n(z) = \sum_{k=0}^{L-1} (z_k - m)^n p(z_k)$$

    * $\mu_0(z) = 1$
    * $\mu_1(z) = 0$
    * $\mu_2(z) = \sigma^2$
    * Figure 2.41

- Aliasing and Moiré Patterns
  - Functions with finite area under the curve can be represented in terms of sines and cosines of various frequencies
  - The sine/cosine component with the highest frequency determines the highest "frequency content" of the function
  - Band-limited functions
    * Highest frequency is finite
    * Function is of unlimited duration
  - Shannon's sampling theorem
    * *If the function is sampled at a rate equal to or greater than twice its highest frequency, it is possible to recover completely the original function from its samples.*
  - Aliasing
    * Result of *undersampling* a function
    * Additional frequency components (*aliased frequencies*) are introduced into the sampled function
  - Impossible to satisfy the sampling theorem in practice
    * Can only work with sampled data that are finite in duration
    * Model the process of converting a function of unlimited duration into a function of finite duration by multiplying using a *gating function* that is 1 for some interval and 0 elsewhere
    * This function itself has frequency components that extend to infinity
    * Act of limiting the duration of band-limited function causes it to cease being band-limited, violating the key condition of sampling theorem
  - Principle approach to reduce the aliasing effects on an image is to reduce its high-frequency components by blurring the image, *prior to sampling*
    * Aliasing is always present in a sampled image
    * Effect of aliased frequencies can be seen in the form of so-called Moiré patterns
  - Moiré effect
    * A periodic function may be sampled at a rate equal to or exceeding twice its highest frequency
    * Possible to recover the function from its samples provided that the sampling captures exactly an integer number of periods of the function
    * Figure 2-24
      · Two identical periodic patterns of equally spaced vertical bars, rotated in opposite directions and superimposed upon each other by multiplying the two images
      · Moiré pattern is caused by breakup of periodicity, as a 2D sinusoidal (aliased) waveform running in a vertical direction