Digital Image Processing

Prof. Tiago Vieira, PhD

Universidade Federal de Alagoas

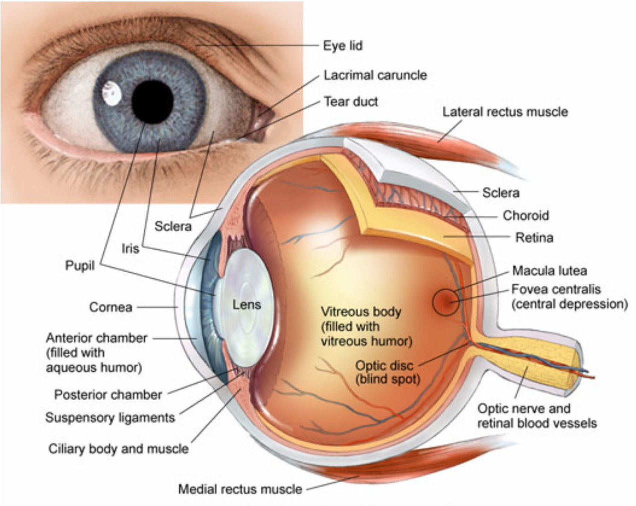
tvieira@ic.ufal.br

October 11, 2019

Contents

The human eye.

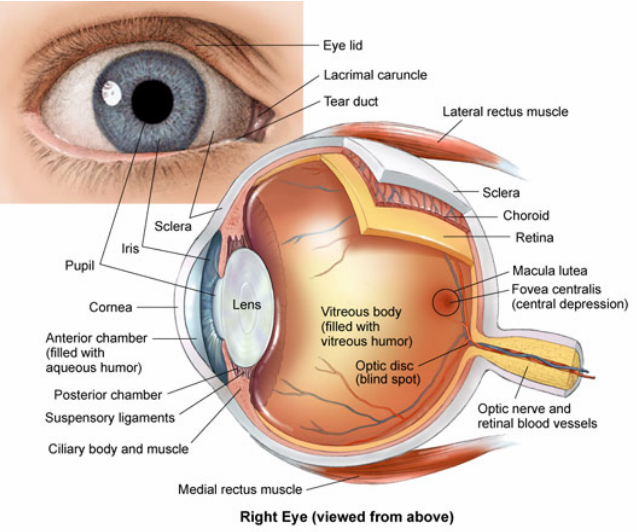
- Diameter $\approx 20mm$.
- Cornea/sclera; Choroid; Retina.



Right Eye (viewed from above)

The human eye.

- Iris.
- Pupil
  ($\phi \in [2, 8]$mm).
- Lens.



Retina.

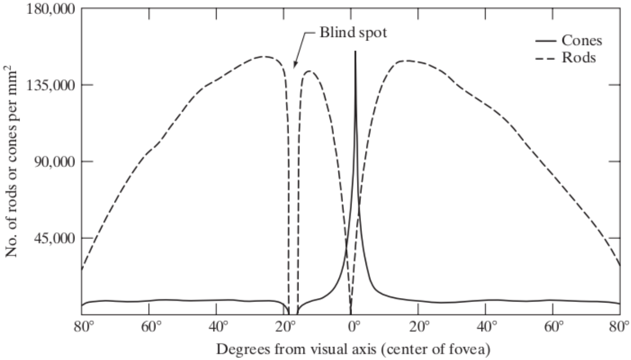
- Innermost membrane of the eye.
- Contains elements sensitive to light:

Light sensitiveness:

- Cones.
  - $6 - 7$ million.
  - More densely distributed around the fovea ($\phi \approx 1.5$mm).
  - Highly sensitive to color.
  - Photopic vision (bright-light vision).

Light sensitiveness:

- Rods.
  - $75 - 150$ million.
  - Scotopic vision (low level of illumination).
  - Give an overview of the Field-of-View (FOV).
  - Do not sense color.

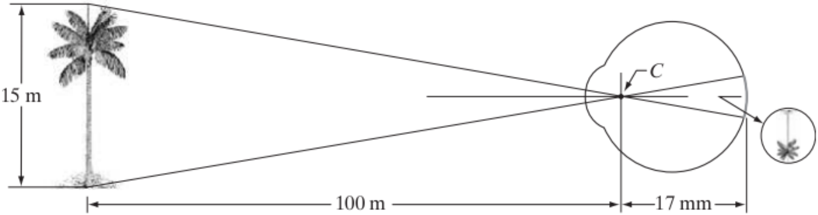


- Suppose that the fovea is a $1.5 \times 1.5mm^2$ square $\rightarrow$ and the density of cones $150k$ elem./mm^2. Then we have $\approx 337k$ elements.
- A medium resolution CCD chip no larger than 5 mm $\times$ 5 mm has this amount of elements.

Current electronic imaging sensors are comparable to the eye's resolution.

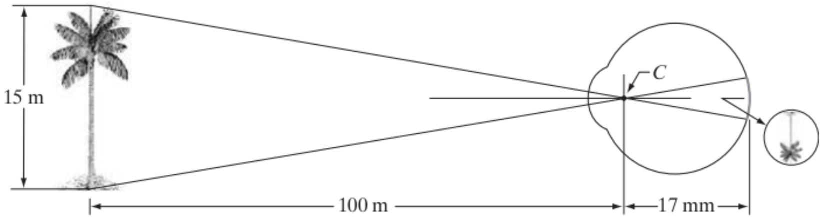
The lens:

- Fixed w.r.t. the retina (≈ 17 mm).
- Range of focal length $\approx [14, 17]$ mm.
- $h = 17 \cdot 15/100 = 2.55$ mm.



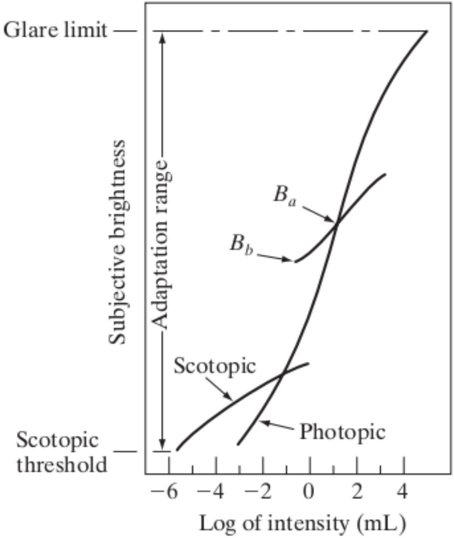
The lens:

- The retinal image is focused primarily on the fovea.
- Perception senses the relative excitation of light receptors.
- Light is converted into electric signals, interpreted by the brain.



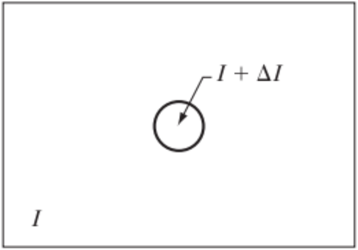
Brightness adaptation and discrimination:

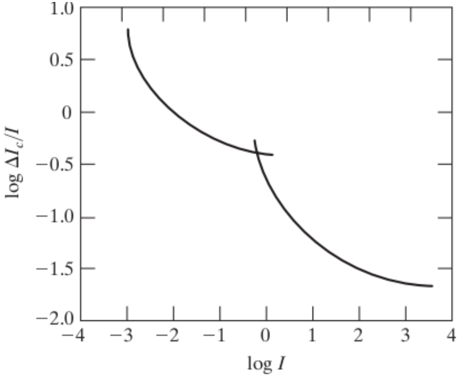
- The eye can adapt to a huge amount of intensity levels ($\backsim 10^{10}$).
- Subjective brightness is a logarithmic function of the light intensity incident in the human eye.



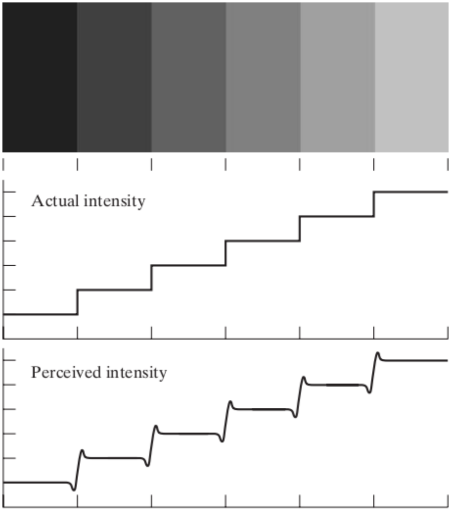
Eye ability to discriminate between *changes* in illumination intensities.

- Weber ratio $= \Delta I_c / I$.

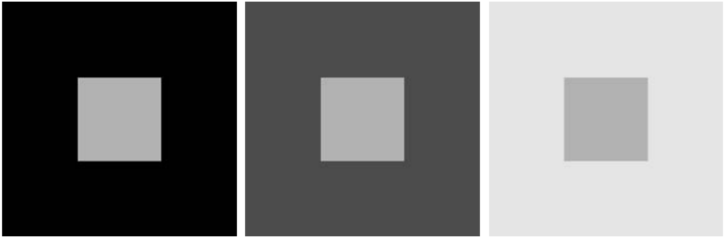




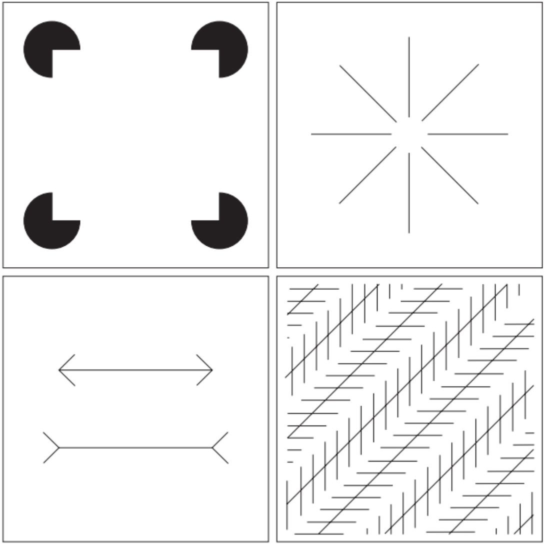
Perceived brightness is not a simple function of intensity (Mach bands).



Perceived brightness is not a simple function of intensity (Simultaneous contrast).



Optical illusions:



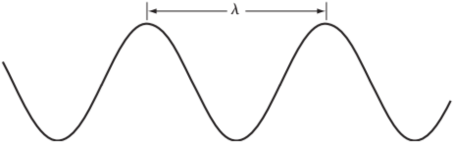
Light

Electromagnetic radiation can be expressed in terms of wavelength (λ), frequency (ν) or energy (E).

$$c = \lambda\nu \tag{1}$$

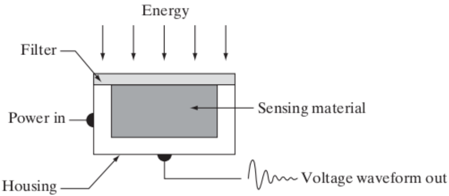
$$E = h\nu \tag{2}$$

Where $c \approx 3 \cdot 10^8$ m/s is the speed of light and $h = 6.62 \cdot 10^{-34} m^2 kg/s$ is the Plank constant.



- If a sensor can be developed that is capable of detecting energy radiated by a band of the electromagnetic spectrum, we can image events of interest in that band.
- Example: water molecule ($\approx 10^{-10} m$). We need a source emitting in the far ultraviolet or soft X-ray region.
- Remember that other sources can be used for imaging, such as sound, electron microscopy and synthetic images.

1. Sensor arrangements:



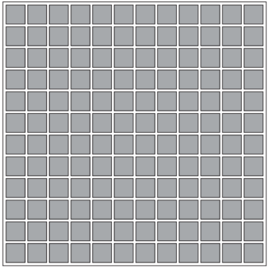


Image acquisition with single sensor.

- Photo-diode.
- Displacement is necessary to generate 2D images.
- *Microdensiometers*.
- Laser with moving mirrors.

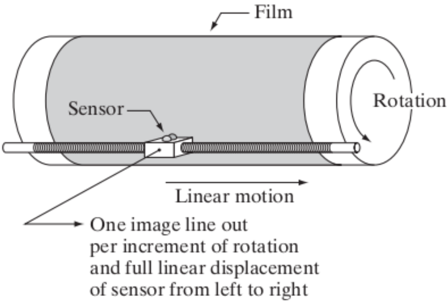
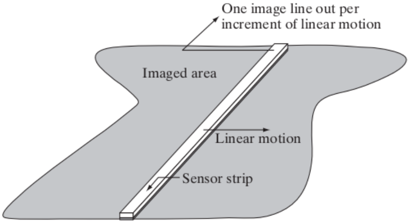
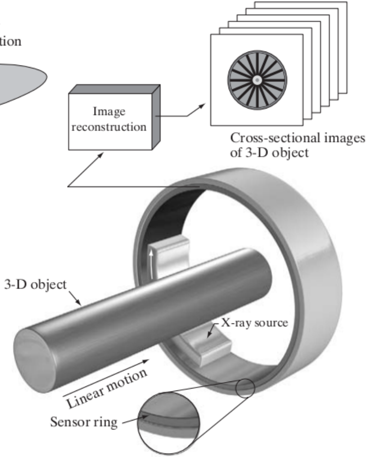


Image acquisition using sensor strips.

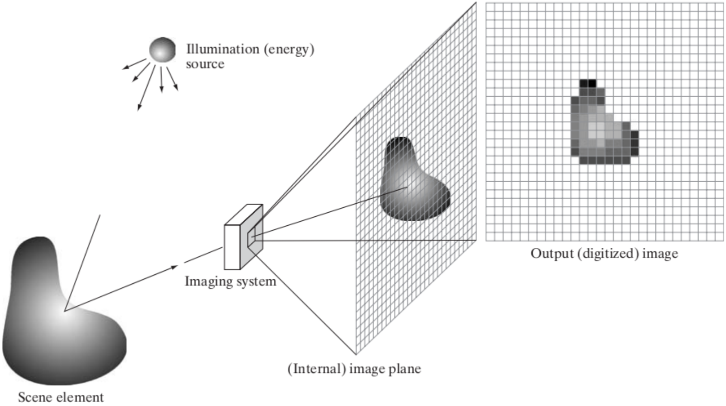
- Airborne imaging applications.
- Comp. Axial Tomography (CAT) scan.



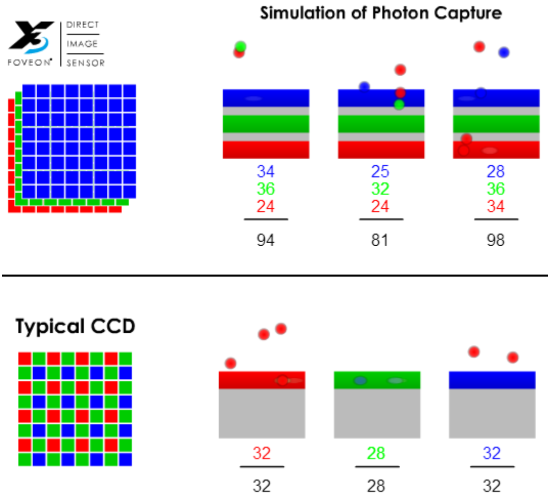


Sensor arrays.

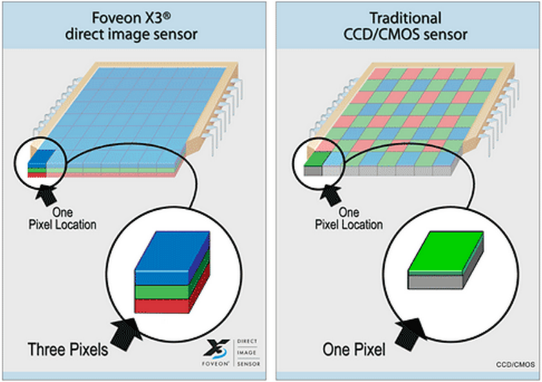
- No motion necessary.
- Typical in cameras.
- $4k \times 4k = 16$ Mpx.



Sensor arrays.



Sensor arrays.



Sensor arrays.

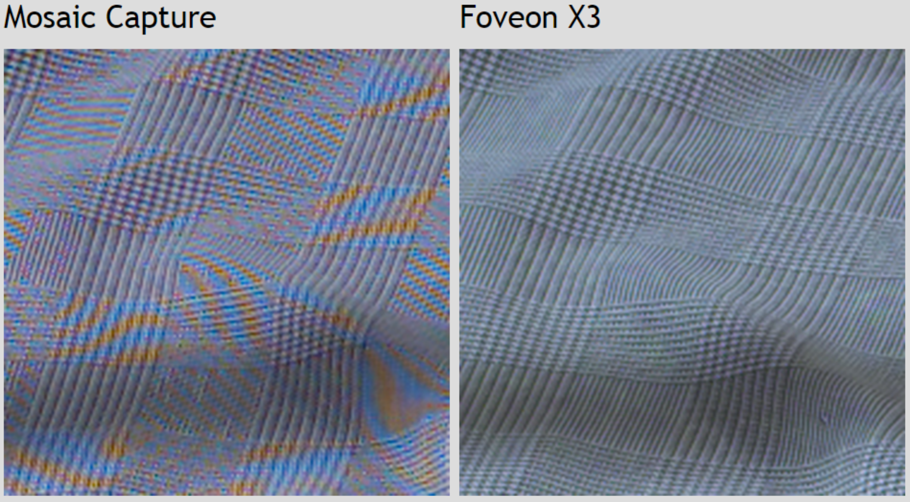


Image formation model:

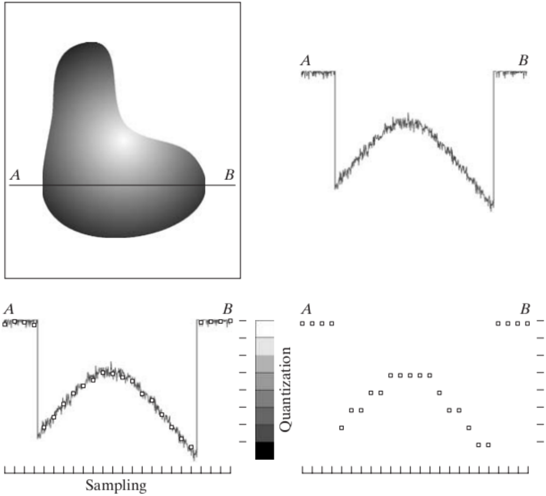
- $0 \leq f(x, y) < \infty$, image. Determined by the product between:
  - $i(x, y)$, illumination, determined by the light source $(0 \leq i(x, y) < \infty)$.
  - $r(x, y)$, reflectance, determined by object's characteristics $(0 \leq r(x, y) \leq 1)$.

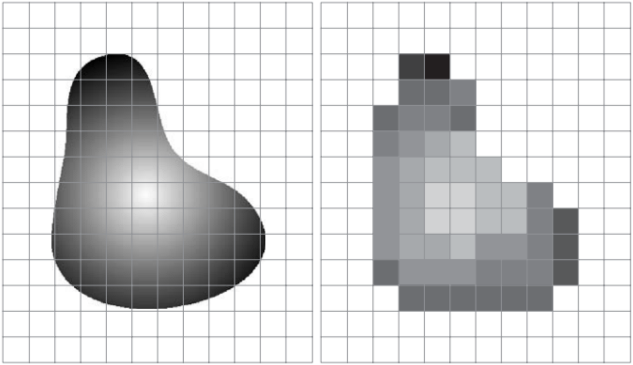
$$f(x, y) = i(x, y) \times r(x, y) \tag{3}$$

- I, intensity level, value of the image at (x, y).
- $L_{min} \leq I \leq L_{max}$, interval, gray scale.
- Typically, the gray scale is shifted to $[0, L-1]$, where $I = 0$ is black and $I = L - 1$ means white.
- Typically, $L = 2^k$. For `uint8` image format, $k = 8$ bits and there are 256 levels.

Generation of digital images from sampled data.

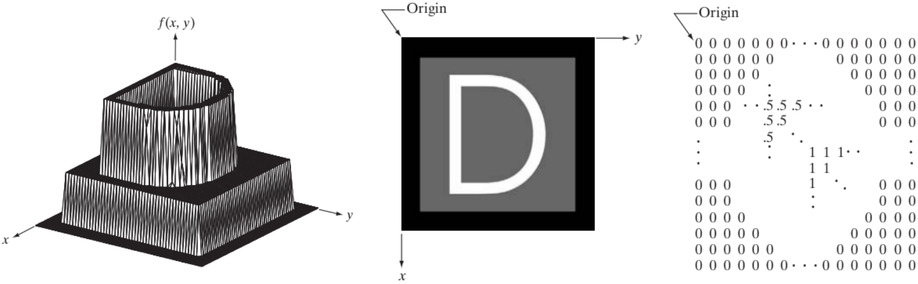
- Sampling.
- Quantization.





Spatial domain:

- Spatial coordinates (x, y) are different from sensor coordinates (u, v).
- $x = 0, \ldots, M - 1$.
- $y = 0, \ldots, N - 1$.



An image is an $M \times N$ matrix, where each cell contains an intensity.

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0, N-1) \\ f(1,0) & f(1,1) & \cdots & f(1, N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1, N-1) \end{bmatrix}$$

Consider a chart:

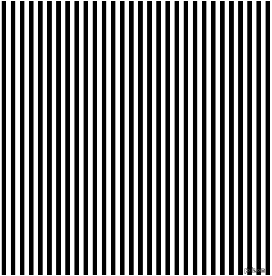
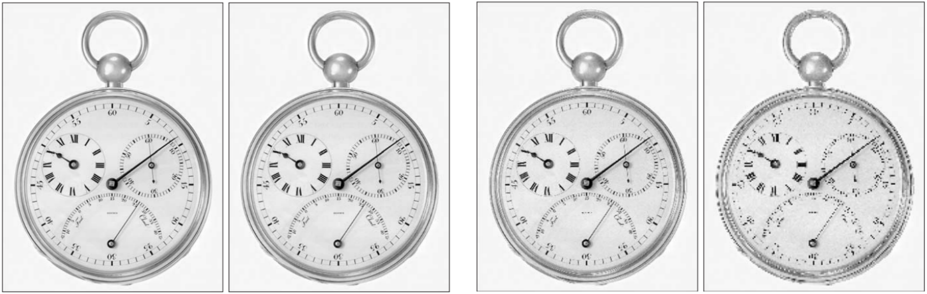


Image resolution is the largest number of *discernible* line pairs per unit distance (dots per inch – *dpi*).

- Newspapers (75 dpi).
- Magazines (133 dpi).
- Glossy brochures (175 dpi).

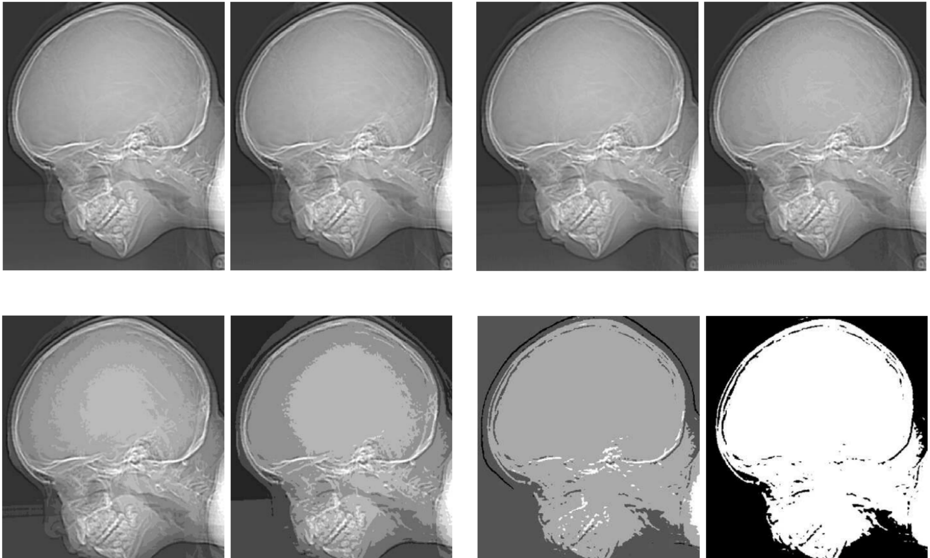
- An image with 1024×1024 resolution has how many dpi?
- One must state the spatial dimension encompassed by the image.



Intensity resolution

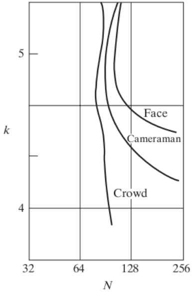
Smallest discernible change in intensity level.

Number of intensity levels: 256, 128, 64, 32, 16, 8, 4, 2. Observe the *false contouring*.



Concept of image quality is subjective. Where N is dimension and 2^k gray levels.

- Isopreference curves in the Nk-plane.
- Tend to be more vertical as the detail in the image increases.
- I. e., the more detail, the less the intensity levels needed.



Rough rule of thumb

Images of size 256×256 pixels with 64 intensity levels and printed on a size format on the order of 5×5 cm are about the lowest spatial and intensity resolution images that can be expected to be reasonably free of objectionable sampling checkerboards and false contouring.

Interpolation

Often used in:

- Zooming.
- Shrinking.
- Rotating.
- Geometric corrections.

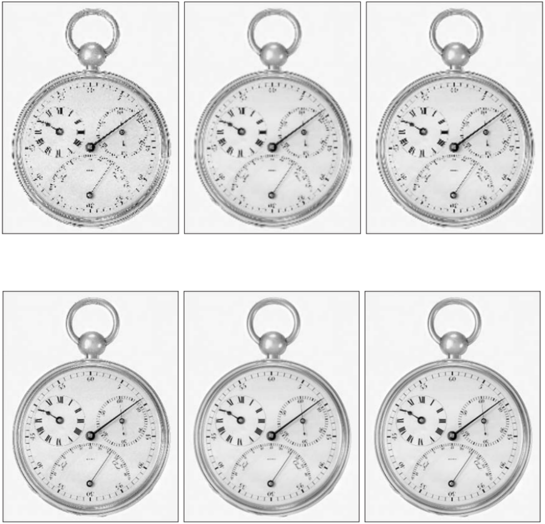
Interpolation methods:

- Nearest-neighbor.
    - Simple but introduces artifacts.
- Bilinear interpolation - New point between 4 points (coefficients).

$$v(x, y) = ax + by + cxy + d. \tag{4}$$

- Bicubic interpolation (standard, used in *Photoshop*) - 16 coefficients.

$$v(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j. \tag{5}$$



a b c
d e f

FIGURE 2.24 (a) Image reduced to 72 dpi and zoomed back to its original size (3692 × 2812 pixels) using nearest neighbor interpolation. This figure is the same as Fig. 2.20(d). (b) Image shrunk and zoomed using bilinear interpolation. (c) Same as (b) but using bicubic interpolation. (d)–(f) Same sequence, but shrinking down to 150 dpi instead of 72 dpi [Fig. 2.24(d) is the same as Fig. 2.20(c)]. Compare Figs. 2.24(e) and (f), especially the latter, with the original image in Fig. 2.20(a).

Basic relationships between pixels

- $N_4(p)$ - the set of 4-*neighbors* of pixel p have coordinates

$$(x+1, y), (x-1, y), (x, y+1), (x, y-1).$$

- $N_D(p)$ - the set of 4-*diagonal neighbors* of pixel p whose coordinates are

$$(x+1, y+1), (x-1, y-1), (x+1, y-1), (x-1, y+1).$$

- $N_8(p) = N_4(p) \cup N_D(p)$.
- Neighbor coordinates may fall outside the image range.

Let V be the set of values defining adjacency (Binary image, $V = 1$).

- 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 $q \in N_8(p)$.

Let V be the set of values defining adjacency (Binary image, $V = 1$).

- m-*adjacency* (mixed adjacency). Two pixels p and q with values from V are m-adjacent if
  - $q \in N_4(p)$, or
  - $q \in N_D(p)$ *and* the set $N_4(p) \cap N_4(q)$ has no pixels whose values are from V.

$$\begin{array}{ccc} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \qquad \begin{array}{ccc} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \qquad \begin{array}{ccc} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}$$

Mixed adjacency is a modification of 8-adjacency. It is introduced to eliminate the ambiguities that often arise when 8-adjacency is used.

A (digital) path (or curve) from pixel p with coordinates (x, y) to pixel q with coordinates (s, t) is a sequence of distinct pixels with coordinates

$$(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$$

where:

- $(x_0, y_0) = (x, y)$.
- $(x_n, y_n) = (s, t)$.
- And pixels (x_i, y_i) and (x_{i-1}, y_{i-1}) are adjacent for $1 \leq i \leq n$.
- $n =$ length of the path.

If $(x_0, y_0) = (x_n, y_n)$ the path is closed. Paths are defined as 4-, 8-, or m-paths.

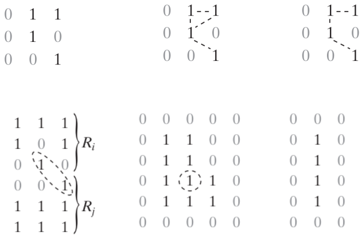
Let S represent a subset of pixels in an image.

- Two pixels p and q are said to be connected in S if there exists a path between them consisting entirely of pixels in S.
- For any pixel p in S, the set of pixels that are connected to it in S is called a connected component of S.
- If it only has one connected component, then set S is called a connected set.

Connected set

A subset R of pixels in an image is called a *region* of the image if R is a connected set.

Two regions R_i and R_j are said to be adjacent if their union forms a connected set.



a b c
d e f

FIGURE 2.25 (a) An arrangement of pixels. (b) Pixels that are 8-adjacent (adjacency is shown by dashed lines; note the ambiguity). (c) m-adjacency. (d) Two regions (of 1s) that are adjacent if 8-adjecency is used. (e) The circled point is part of the boundary of the 1-valued pixels only if 8-adjacency between the region and background is used. (f) The inner boundary of the 1-valued region does not form a closed path, but its outer boundary does.

Boundary

The boundary (border or contour) of a region R is the set of pixels that have at least one background neighbor.

For pixels p, q and z, with respective coordinates (x, y), (s, t), and (v, w), D is a distance function or metric if

- $D(p, q) \geq 0$ ($D(p, q) = 0$ if $p = q$).
- $D(p, q) = D(q, p)$, and
- $D(p, z) < D(p, q) + D(q, z)$.

The Euclidean distance between p and q is defined as

$$D_e(p, q) = \left[(x - s)^2 + (y - t)^2\right]^{1/2}.$$

Other distances:

- City-block distance

$$D_4(p,q) = |x-s| + |y-t|.$$

		2		
	2	1	2	
2	1	0	1	2
	2	1	2	
		2		

- The D_8 distance

$$D_8(p,q) = \max(|x-s|, |y-t|).$$

2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

- D_4 and D_8 distances involve only coordinates of the points.
- D_m (m-adjacency)) distance between 2 points = shortest m-path between the points.
- Now the pixels in between matter.

	p_3	p_4
p_1	p_2	
p		

	0	1
0	1	
1		

pp_2p_4

	0	1
1	1	
1		

$pp_1p_2p_4$

	1	1
1	1	
1		

$pp_1p_2p_3p_4$

Mathematical Tools

Attention:

- Array operation is made pixel-by-pixel.
- Matrix operations are carried out using matrix theory.
- Henceforth assume array operation unless stated otherwise.

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

Array product:

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

Matrix product:

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

Linear versus Nonlinear operations:

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

$$H[f(x, y)] = g(x, y). \tag{6}$$

- H is linear if

$$H[a_i f_i(x, y) + a_j f_j(x, y)] = a_i H[f_i(x, y)] + a_j H[f_j(x, y)]. \tag{7}$$

- Where $a_i, a_j, f_i(x, y)$ and $f_j(x, y)$ are arbitrary constants and images (of the same size).

- Linear example: The sum operator ($\sum$):

$$\begin{aligned}\sum\left[a_i f_i(x, y) + a_j f_j(x, y)\right] &= \sum a_i f_i(x, y) + \sum a_j f_j(x, y)\\ &= a_i \sum f_i(x, y) + a_j \sum f_j(x, y)\\ &= a_i g_i(x, y) + a_j g_j(x, y)\end{aligned}$$



- Non-linear example; the Max operator.
- Another example: Absolute image differencing:

- linear operations are important due to their theoretical and practical results applicable to image processing.
- Nonlinear systems are not nearly as well understood, so their scope of application is more limited.

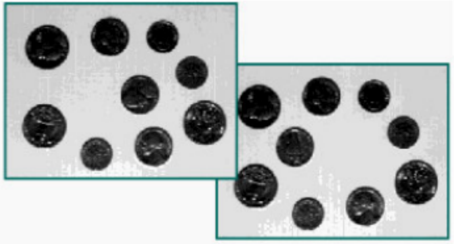
- The arithmetic operations $(+, -, \times, \div)$ are performed pixelwise in array operations, between images of the same size.

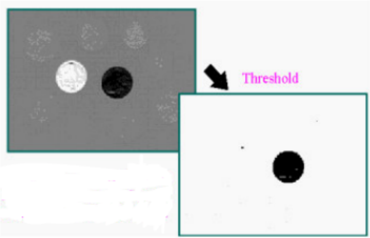
Example of noise reduction using average. Assume:

1. An image $f(x, y)$ corrupted by noise:
$$g(x, y) = f(x, y) + \eta(x, y). \tag{8}$$
2. The noise is uncorrelated between image acquisitions.
3. The noise has zero mean value.

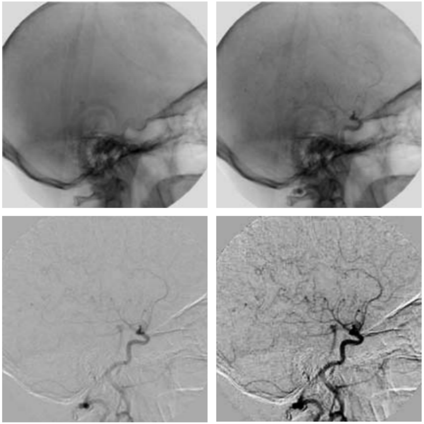
Example of image subtraction application.

- Subtract images containing differences or movement.
- The subtraction operation highlights differences.





Another example:



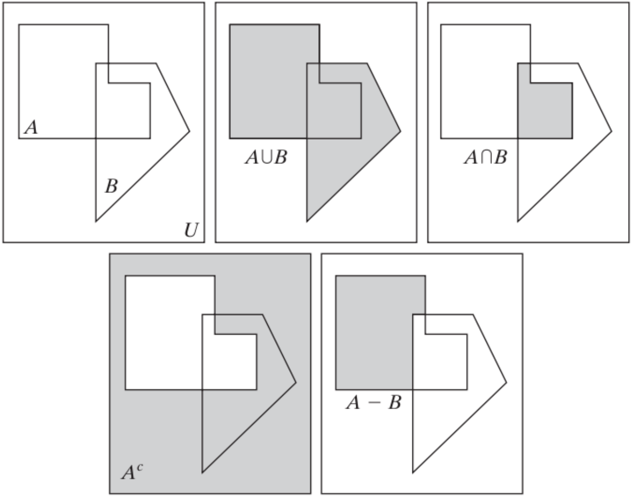
a	b
c	d

FIGURE 2.28 Digital subtraction angiography. (a) Mask image. (b) A live image. (c) Difference between (a) and (b). (d) Enhanced difference image. (Figures (a) and (b) courtesy of The Image Sciences Institute, University Medical Center, Utrecht, The Netherlands.)

Attention!

The image resulting form arithmetic operations should (often) be scaled to $[0, 255]$!

Basic set operations:



Consider set A formed of triplets (x, y, z), where

- x, y are spatial coordinates.
- z is intensity.

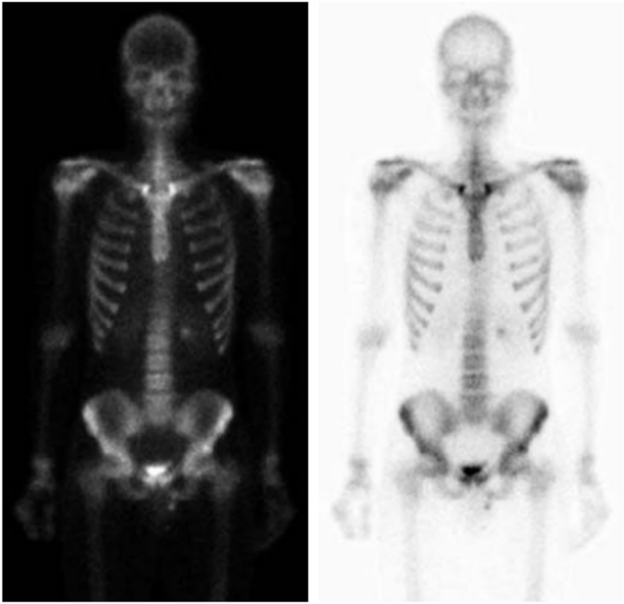
The *complement* of A is defined as

$$A^c = (x, y, K - z)|(x, y, z) \in A,$$

where

$$K = 2^k - 1.$$

e. g., $K = 255$ for 8 bits.



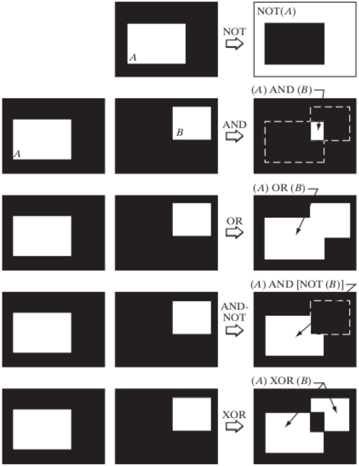
The union of two gray-scale sets A and B may be defined as the set

$$A \cup B = \left\{ \max_{z}(a, b) | a \in A, b \in B \right\} \tag{9}$$

[1]

[1] $B = 3 \times \mu_A$

Logical operators.



Spatial operations are performed directly on the pixels of a given image.

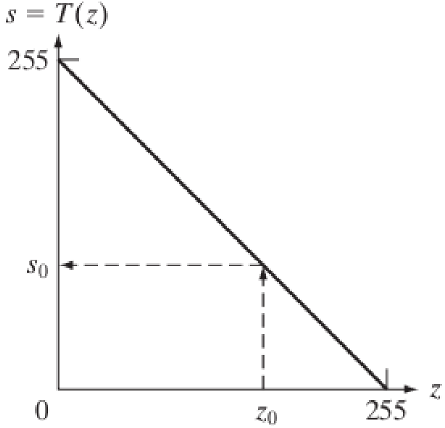
The operations can be:

- Single-pixel.
- Neighborhood.
- Geometric spatial.

Single pixel operations

- Alter the value of the pixels based on their intensities.

$$s = T(z). \tag{10}$$

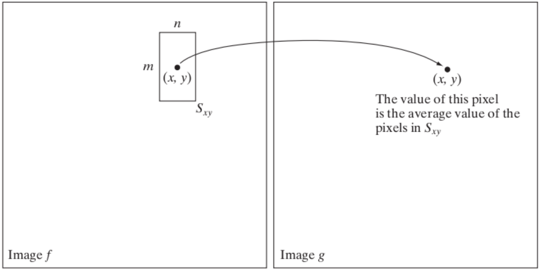


Neighborhood operations

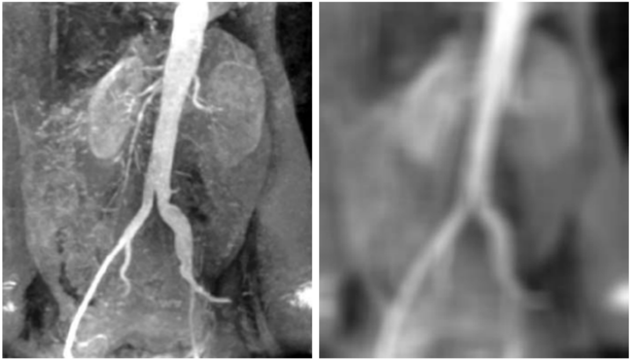
- Each pixel on the output image depends on the neighborhood of the corresponding (same position) pixel in the input image.

Example: Local averaging.

$$g(x,y) = \frac{1}{mn} \sum_{(x,y)\in S_{xy}} f(r,c). \tag{11}$$



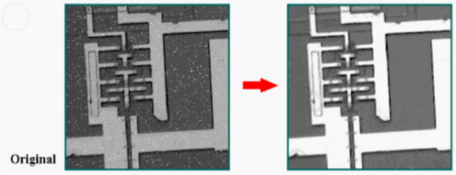
The net effect is to perform local blurring in the original image.



Median filter:

- Order the neighboring pixels in ascending order and sets the output to the median value.

- Good for filtering high valued noises:



Geometric spatial operations

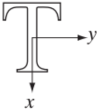
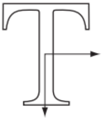
- Change the relation between pixel positions.
- Two steps:
  - Spatial transformation.

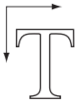
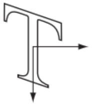
$$(x, y) = T\{(u, v)\} \tag{12}$$

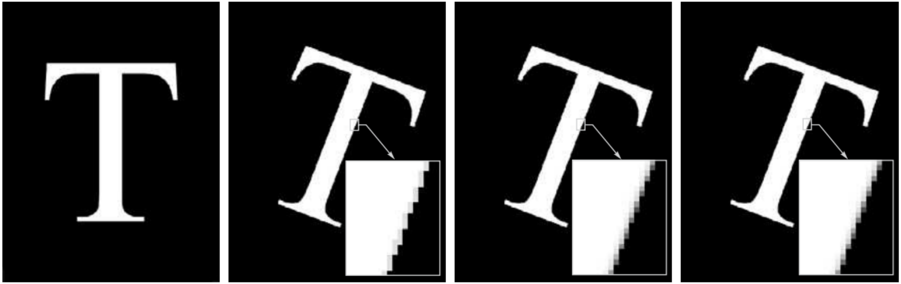
    - This process often creates pixels nonexistent in the original image, leading to the necessity of
  - Interpolation:
    - Replication (nearest-neighbor), bi-linear, $\sin(x)/x$, etc. . .

A general form that can *scale*, *translate*, *rotate* or *sheer* a set of coordinate points is the *Affine* transform:

$$[x \ \ y \ \ 1] = [v \ \ w \ \ 1] \, \mathbf{T} = [v \ \ w \ \ 1] \begin{bmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & 1 \end{bmatrix}$$

Transformation Name	Affine Matrix, T	Coordinate Equations	Example
Identity	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v$ $y = w$	
Scaling	$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = c_x v$ $y = c_y w$	
Rotation	$\begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v\cos\theta - w\sin\theta$ $y = v\cos\theta + w\sin\theta$	

Translation	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix}$	$x = v + t_x$ $y = w + t_y$	
Shear (vertical)	$\begin{bmatrix} 1 & 0 & 0 \\ s_v & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v + s_v w$ $y = w$	
Shear (horizontal)	$\begin{bmatrix} 1 & s_h & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x = v$ $y = s_h v + w$	



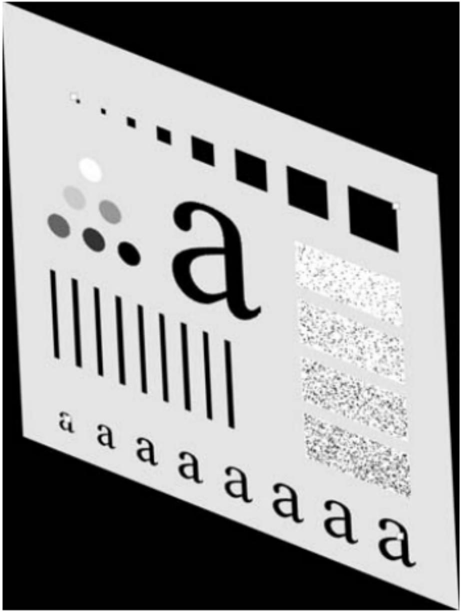
a b c d

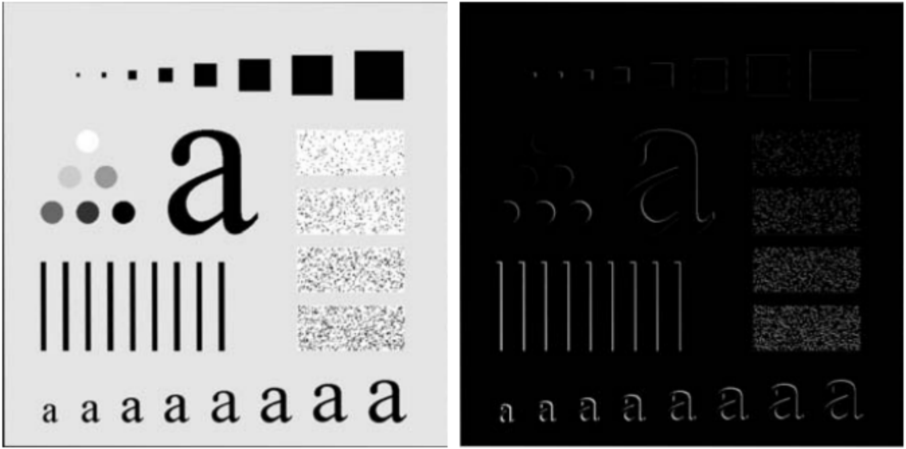
FIGURE 2.36 (a) A 300 dpi image of the letter T. (b) Image rotated 21° using nearest neighbor interpolation to assign intensity values to the spatially transformed pixels. (c) Image rotated 21° using bilinear interpolation. (d) Image rotated 21° using bicubic interpolation. The enlarged sections show edge detail for the three interpolation approaches.

Image registration

Input and output images are given and the transformation must be estimated.

- Examples: Change detection in satellite images.
- Before comparison, images must be corrected to fix acquisition misalignment.
- *Ground Control Points* (GCP) are often used in aerial photography.
- Small metallic objects embedded in the sensor produce *known* points that can also be used as reference.
- Image descriptors (SIFT, SURF, etc).





Input images :

Stitched Output:

- So far, processing was presented only in the spatial domain (directly upon pixels).
- Another tools arise when transforming images into other domains, such as frequency, as illustrated below.

$$T(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) r(x,y,u,v). \tag{13}$$

where:

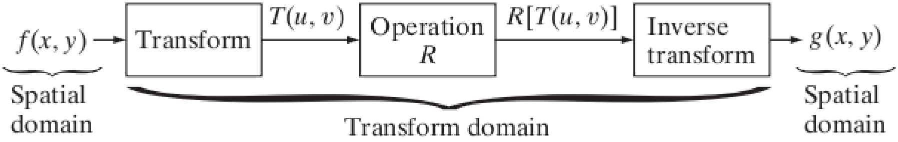
- $f(x,y)$ - input image.
- $r(x,y,u,v)$ - forward transformation kernel.
- x, y - spatial variables.
- M, N row and column dimensions of $f(x,y)$.
- $u = 0, \ldots, M-1$ and $v = 0, \ldots, N-1$ form the transformed coordinate system.

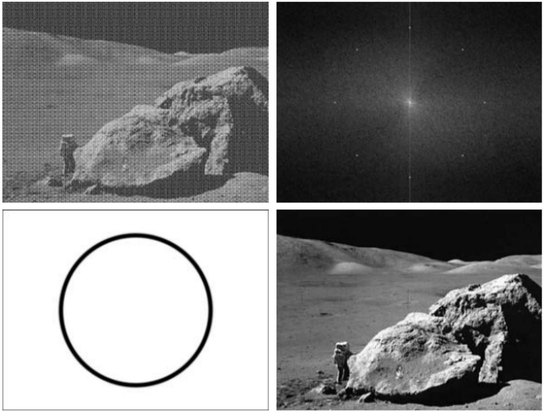
Given $T(u,v)$, $f(x,y)$ can be recovered from the inverse transform:

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} T(u,v)s(x,y,u,v). \tag{14}$$

where:

- $f(x,y)$ - recovered image.
- $s(x,y,u,v)$ - backward transformation kernel.
- x, y - spatial variables.
- M, N row and column dimensions of $f(x,y)$.
- $x = 0, \ldots, M-1$ and $y = 0, \ldots, N-1$ form the spatial coordinate system.


$f(x, y)$
Transform
$T(u, v)$
Operation R
$R[T(u, v)]$
Inverse transform
$g(x, y)$
Spatial domain
Transform domain
Spatial domain



Intensity values in an image can be considered random quantities:

- Let $z_i, i = 0, \ldots, L-1$, denote the values of all possible intensities in an $M \times N$ digital image.
- The probability $p(z_k)$, of intensity level z_k occurring is

$$p(z_k) = \frac{n_k}{MN}, \tag{15}$$

  where n_k is the number of occurrences of z_k and MN is the total number of pixels.
- Note that $\sum_{k=0}^{L-1} p(z_k) = 1$.

Given $p(z_k)$, other image characteristics can be obtained, such as

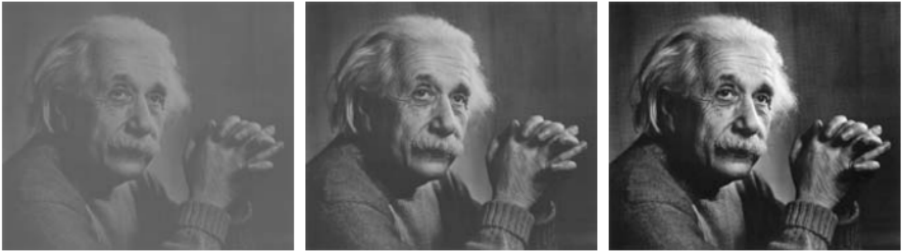
- Image mean
$$m = \sum_{k=0}^{L-1} z_k p(z_k).$$
- Variance
$$\sigma^2 = \sum_{k=0}^{L-1} (z_k - m)^2 \, p(z_k).$$
- nth moment of random variable z about the mean
$$\mu_n(z) = \sum_{k=0}^{L-1} (z_k - m)^n \, p(z_k).$$

Notice that

- $\mu_0(z) = 1$.
- $\mu_1(z) = 0$.
- $\mu_2(z) = \sigma^2$.

- The mean and variance are more easily associated with the intensity distribution on an image.
- Higher moments, however, are more subtle.
- Example, μ_3:
  - $\mu_3 > 0$, distribution skewed to the right.
  - $\mu_3 = 0$, distribution symmetrically skewed.
  - $\mu_3 < 0$, distribution skewed to the left.

Example of an image with low, medium and high contrast with, respectively, 14, 32 and 50 standard deviations.



Probabilistic methods are broadly used in image processing:

- Intensity transformations.
- Restoration.
- Segmentation.
- Texture description.
- Object recognition.