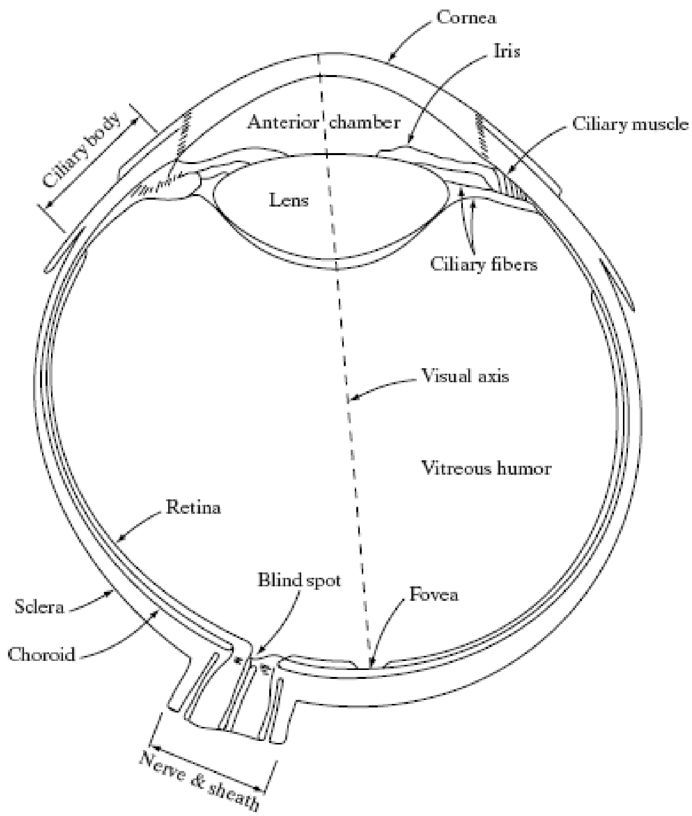
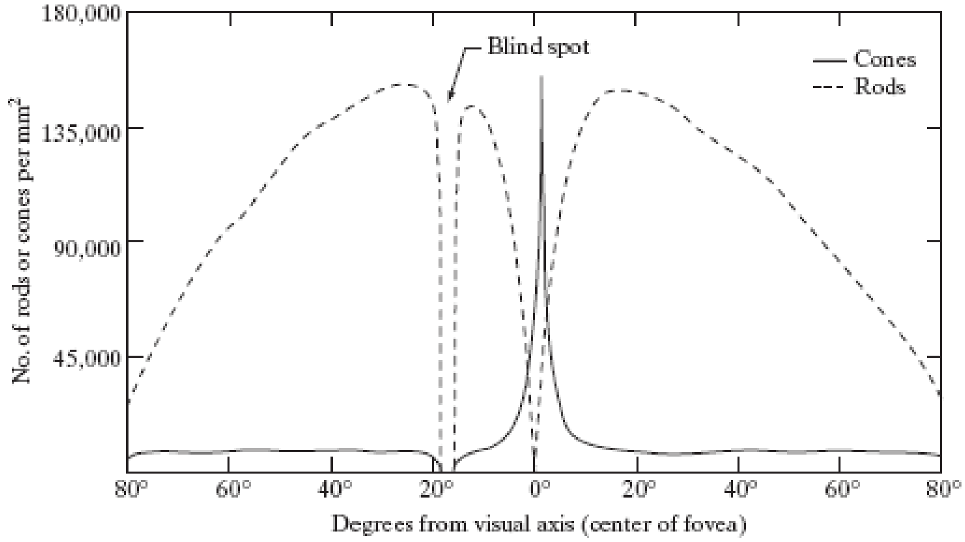
Chapter 2 Digital Image Fundemantals

Human Eye: It is nearly a sphere with an average diameter of 20 mm enclosed in three membranes: cornea, sclera and the choroid and retina as shown below.



Cornea: is a tough transparent tissue covering the anterior surface and the opaque membrane sclera does the remainder of the optic globe.
Choroid: It lies directly under the sclera and it contains a network of **blood vessels** as the major source of nutrition for the eye. At its anterior, the choroid is divided into the **ciliary** body and the **iris** diagram.
Iris: Through contraction and expansion, it controls the light entering the eye. Central opening of the iris is called **pupil** with a diameter varying 2-8 mm. Front of the iris contains colored pigments, which are unique for each person and the inside contains black pigments.
Lens are made up of concentric layers of fibrous cells and its suspended by fibers attached to the ciliary body. They are colored with slightly yellow pigments, which increases with age and extreme cases result in cataracts. They absorb ~8% of the visible light spectrum with relatively higher percentage higher frequencies (shorter wavelengths). Both ultraviolet and infrared light are absorbed significantly by the lens system to protect the eye, however, excessive amounts damage the eye.
Retina: it lines the inside wall's entire posterior portion. When the eye is properly focused, light from an object outside is imaged on the retina. Pattern vision is handled by the distribution of light receptors distributed over the surface of the retina. There are two classes of light receptors: **Cones** and **rods** and their distribution is shown below.

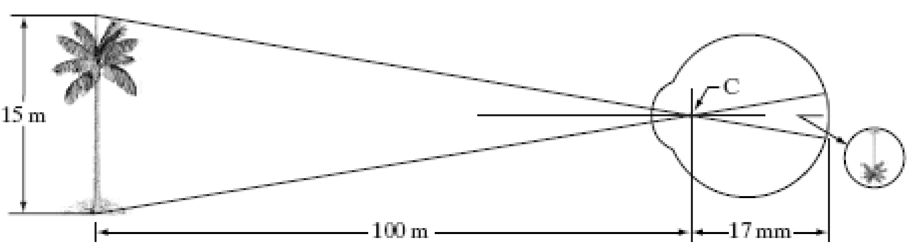


Cones: There are 6-7 million of them in each eye and heavily concentrated on the center of fovea and they are highly sensitive to color. Humans can resolve fine details with cones since they are individually connected to a nerve of their own. Muscles rotate the eyeball until the image falls on the fovea. Cone vision is called **photopic** or **bright light vision**.

Rods: There are 75-150 million rods per eye and they are distributed across the retinal surface. Several rods are connected to a single nerve and thus they cannot carry detailed information. They tend to give overall picture of the field of view. They are not involved in color vision and are sensitive to low levels of illumination. Example: Brightly colored objects in daylight appear as colorless forms under moonlight since only the rods are stimulated to result in **scotopic** or **dim-light vision**. There are no receptors where the optic nerve is connected and it is called **blind-spot**.

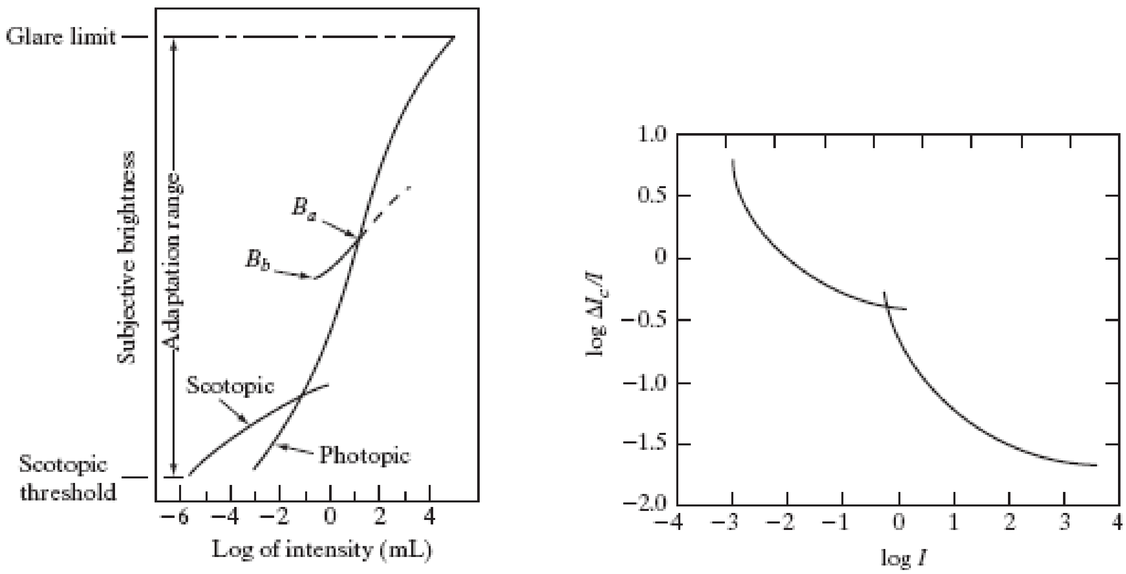
Fovea: It is a circular indentation of about 1.5 mm in diameter. To make it easy to discuss it in relation to camera setups, it is approximated as a 1.5x1.5 mm square array. The density of cones in fovea is 150,000 elements per square mm resulting in 337,000 sensor array. A charge-coupled CCD imaging device needs 5x5 mm array size to have such a resolution.

Image Formation: It is the same principle as in ordinary optics as shown below and the image is upside down.



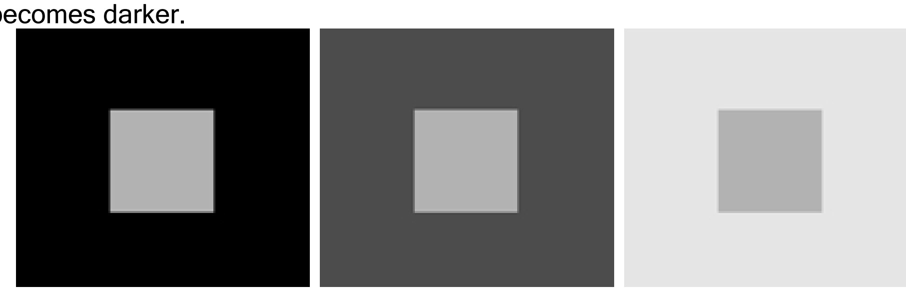
The distance between the center of lens and the retina (focal length) varies from 14-17 mm as the refractive power of the lens increases from its minimum to maximum. When the eye focuses on objects more than 3 m, the lens has the lowest refractive power. On the other hand, it has the highest refraction, for nearby objects. In the picture above, the image will be formed 15/100=h/17=2.55 mm.

The range of light intensity levels to which the human visual system can adapt is on the order of 10^{10} from the scotopic threshold to the glare limit. Experimental evidence indicates that ***subjective brightness*** (intensity as *perceived* by the human visual system) is a logarithmic function of the light intensity incident on the eye. The quantity $\Delta I_c / I$, where ΔI_c is the increment of illumination discriminable 50% of the time with background illumination I, is called the Weber ratio. A small value $\Delta I_c / I$ of means that a small percentage change in intensity is discriminable. This represents “good” brightness discrimination. Conversely, a large value $\Delta I_c / I$ of means that a large percentage change in intensity is required. This represents “poor” brightness discrimination.



Subjective brightness as a function of log intensity (miliLambert) and Typical Weber ratio as a function of intensity.

The above curve shows that brightness discrimination is poor (the Weber ratio is large) at low levels of illumination, and it improves significantly (the Weber ratio decreases) as background illumination increases. Two branches in the curve reflect the fact that at low levels of illumination vision is carried out by activity of the rods, whereas at high levels (showing better discrimination) vision is the function of cones.
Simultaneous Contrast: Even if all objects have the same intensity, but they appear darker as the background becomes darker.



The electromagnetic spectrum is expressed in terms of wavelength, frequency, or energy. Wavelength λ and frequency v are related by $\lambda = c / v$ where $c = 2.998 * 10^{10}$ m/s, (speed of light). The energy of the various components of the electromagnetic spectrum is given by the expression *E=h.v*, where *h* is Planck’s constant.The units of wavelength are meters, with the terms *microns* (denoted μm) and *nanometers* being used.

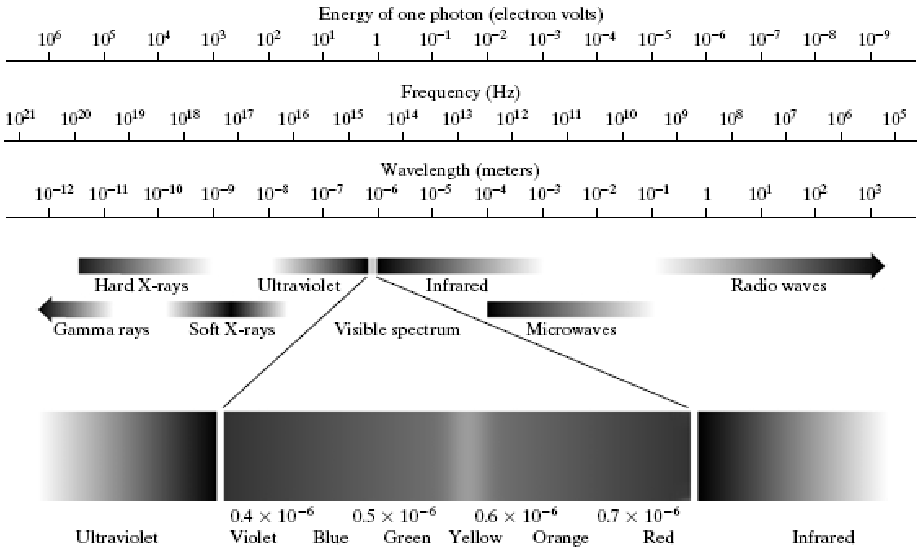


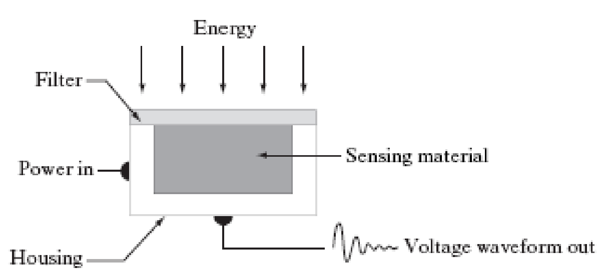
FIGURE 2.10 The electromagnetic spectrum. The visible spectrum is shown zoomed to facilitate explanation, but note that the visible spectrum is a rather narrow portion of the EM spectrum.

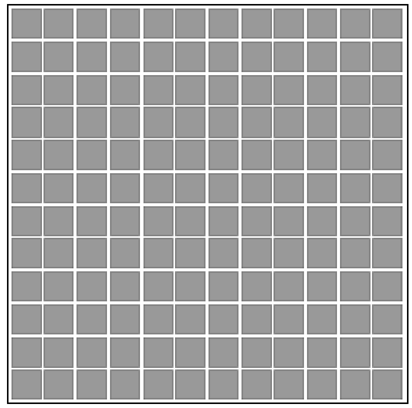
The colors that humans perceive in an object are determined by the nature of the light *reflected* from the object. For example, green objects reflect light with wavelengths primarily in the 500 to 570 nm range while absorbing most of the energy at other wavelengths. Light that is void of color is called *achromatic* or *monochromatic* light. The only attribute of such light is its *intensity*, or amount. The term *gray level* generally is used to describe monochromatic intensity because it ranges from black, to grays, and finally to white.

Three basic quantities are used to describe the quality of a chromatic light source: radiance; luminance; and brightness.
***Radiance** (in watts)* is the total amount of energy that flows from the light source. .
Luminance, measured in lumens (lm), gives a measure of the amount of energy an observer *perceives* from a light source. For example, light emitted from a source operating in the far infrared region of the spectrum could have significant energy (radiance), but an observer would hardly perceive it; its luminance would be almost zero.
Brightness is a subjective descriptor of light perception that is practically impossible to measure.

IMAGE ACQUISITION and QUANTIZATION

.
Depending on the nature of the source, illumination energy is reflected from, or transmitted through, objects and the constructed image is acquired by sensors. There are three different sensor configurations used in imaging: single, line, and array sensors.





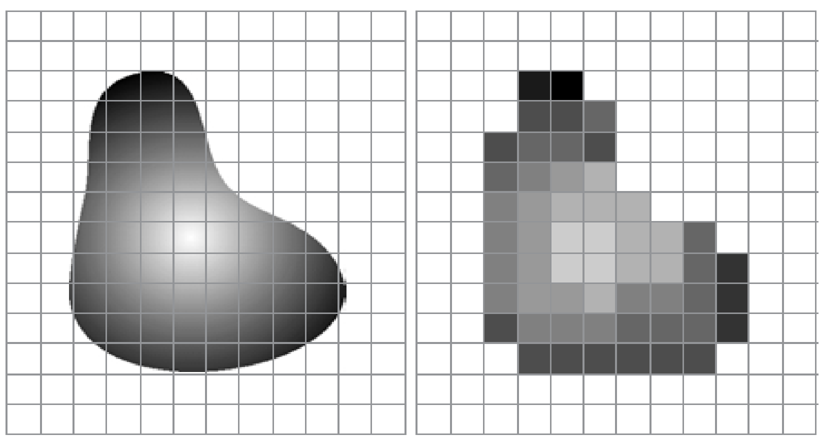
Images are formed and represented by: $f(x,y) = i(x,y).r(x,y)$ with the property:

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

where $i(x,y)$ is the illumination and

$r(x,y)$ the reflectance components with the range $0 < r(x,y) < 1$.

Images are quantized (digitized) through a 2-D sampling process and depending upon the number of quantizer levels they have varying resolution.



Types of neighborhoods: Neighborhood operations play a key role in modern digital image processing. It is therefore important to understand how images can be sampled and how that relates to the various neighborhoods that can be used to process an image.

> * Rectangular sampling - In most cases, images are sampled by laying a rectangular grid over an image.

> * Hexagonal sampling - An alternative sampling scheme is shown in Figure 3c and is termed hexagonal sampling.

Both sampling schemes have been studied extensively and both represent a possible periodic tiling of the continuous image space. We will restrict our attention, however, to only rectangular sampling as it remains, due to hardware and software considerations, the method of choice.

Local operations produce an output pixel value $b[m=m_o,n=n_o]$ based upon the pixel values in the *neighborhood* of $a[m=m_o,n=n_o]$. Some of the most common neighborhoods are the 4-connected neighborhood and the 8-connected neighborhood in the case of rectangular sampling and the 6-connected neighborhood in the case of hexagonal sampling illustrated in Figure 3.

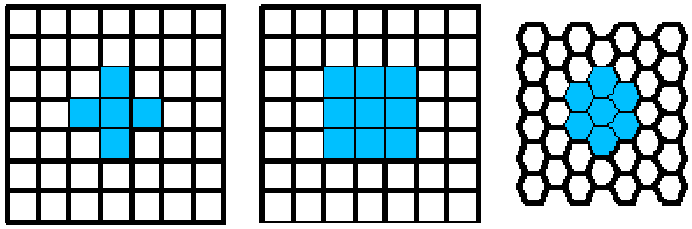


Figure 3. Rectangular sampling Rectangular sampling exagonal sampling 4-connected 8-connected 6-connected

Video Parameters:

We do not propose to describe the processing of dynamically changing images in this introduction. It is appropriate--given that many static images are derived from video cameras and frame grabbers-- to mention the standards that are associated with the three standard video schemes that are currently in worldwide use - NTSC, PAL, and SECAM. This information is summarized in Table 3.

	Aspect Ratio	*Interlace*	*Frames/s*	*Total/Active Lines*	*BW (MHz)*
NTSC (USA,Japan, Canada,Mexico)	4:3	2:1	29.97	525/480	4.2
PAL (Great Britain)	4:3	2:1	25	625/580	5.5
PAL (Germany, Austria, Italy)	4:3	2:1	25	625/580	5.0
PAL (China)	4:3	2:1	25	625/580	6.0
SECAM (France,Russia)	4:3	2:1	25	625/580	6.0

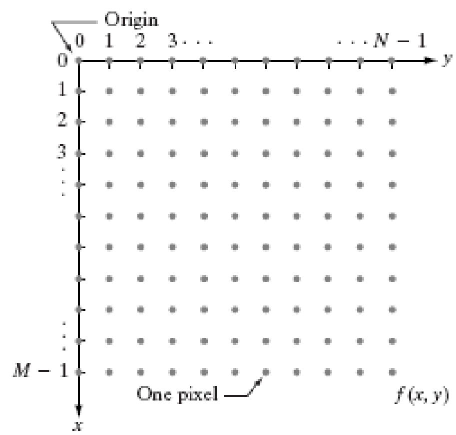
Image aspect ratio = image width /image height

Table 3: Standard video parameters

In an interlaced image the odd numbered lines (1,3,5,...) are scanned in half of the allotted time (e.g. 20 ms in PAL) and the even numbered lines (2,4,6,...) are scanned in the remaining half. The image display must be coordinated with this scanning format. (See Section 8.2.) The reason for interlacing the scan lines of a video image is to reduce the perception of flicker in a displayed image. If one is planning to use images that have been scanned from an interlaced video source, it is important to know if the two half-images have been appropriately "shuffled" by the digitization hardware or if that should be implemented in software. Further, the analysis of moving objects requires special care with interlaced video to avoid "zigzag" edges.

The number of rows (*N*) from a video source generally corresponds one-to-one with lines in the video image. The number of columns, however, depends on the nature of the electronics that is used to digitize the image. Different frame grabbers for the same video camera might produce *M* = 384, 512, or 768 columns (pixels) per line.

Image labeling convention is traditionally upper left to lower right as shown below. The notation used in representing a frame of image is in a matrix form:



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

Subsampling of images create a lower number of pixels (bits when quantized) as shown below.

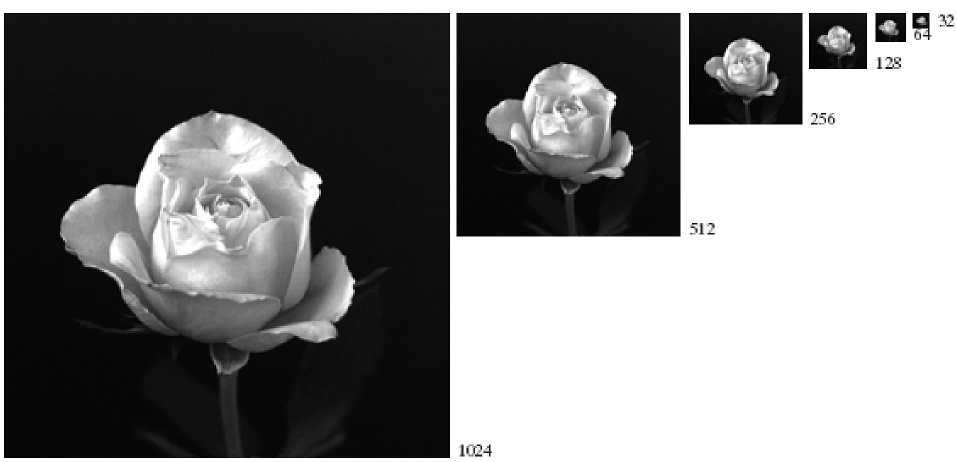


FIGURE 2.19 A 1024 × 1024, 8-bit image subsampled down to size 32 × 32 pixels. The number of allowable gray levels was kept at 256.

Distance Measures: Performance quality in image processing usually involves a distance measure. For pixels *p, q, z* in locations (*x,y*), (*s,t*), and (*v,w*), the distance measure is a distance function or metric if:

(a) $D(p,q) \geq 0 \quad with\ D(p,q) = 0 \quad iff \quad p = q$

(b) $D(p,q) = D(q,p)$

(c) $D(p,z) \leq D(p,q) + D(q,z)$

Euclidean Distance: $D(p,q) = \{(x-s)^2 + (y-t)^2\}^{1/2}$

City-Block Distance D_4: $D_4 = |x-s| + |y-t|$

Chessboard Distance D_8 $D_8 = \max((|x-s|,|y-t|\}$

Examples: $D_4 \leq 2$ and $D_8 \leq 2$

$$\begin{array}{ccccc} & & 2 & & \\ & 2 & 1 & 2 & \\ 2 & 1 & 0 & 1 & 2 \\ & 2 & 1 & 2 & \\ & & 2 & & \end{array} \qquad \begin{array}{ccccc} 2 & 2 & 2 & 2 & 2 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{array}$$

PERCEPTION

Many image processing applications are intended to produce images that are to be viewed by human observers. It is therefore important to understand the characteristics and limitations of the human visual system--to understand the "receiver" of the 2D signals. At the outset it is important to realize that :

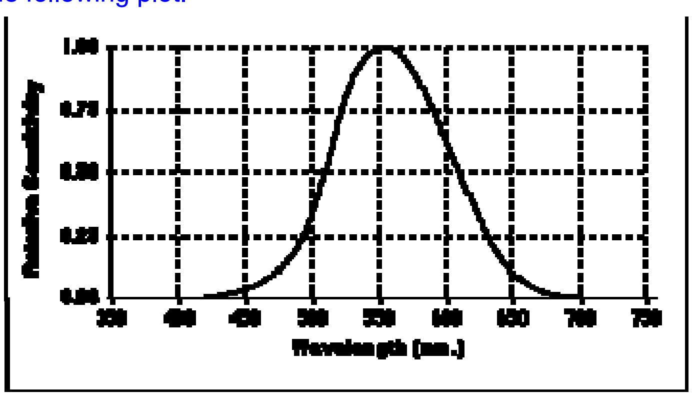
1) Human visual system is not well understood,
2) No objective measure exists for judging the quality of an image that corresponds to human assessment of image quality, and,
3) The "typical" human observer does not exist.

Brightness Sensitivity:

- Wavelength sensitivity
- Stimulus sensitivity

There are several ways to describe the sensitivity of the human visual system. To begin, let us assume that a homogeneous region in an image has an intensity as a function of wavelength (color) given by $I(\lambda)$ Further let us assume that $I(\lambda) = I_0$, a constant.

Wavelength sensitivity: Perceived intensity as a function of λ, the spectral sensitivity, for the "typical observer" is shown in the following plot.

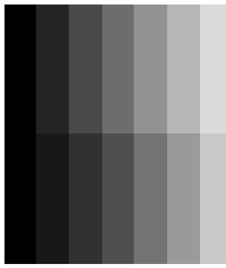


Spectral Sensitivity of the "typical" human observer

Stimulus sensitivity: If the constant intensity (brightness) I_o is allowed to vary then, to a good approximation, the visual response, R, is proportional to the logarithm of the intensity. This is known as the Weber-Fechner law: $R = \log(I_0)$

The implications of this are easy to illustrate. Equal perceived steps in brightness, $\Delta R = k$ require that the physical brightness (the stimulus) increases exponentially. This is illustrated in Figure 11ab.

A horizontal line through the top portion of Figure 11a shows a linear increase in objective brightness (Figure 11b) but a logarithmic increase in subjective brightness. A horizontal line through the bottom portion of Figure 11a shows an exponential increase in objective brightness (Figure 11b) but a linear increase in subjective brightness.



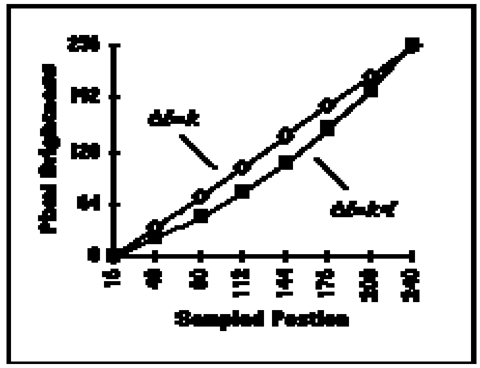
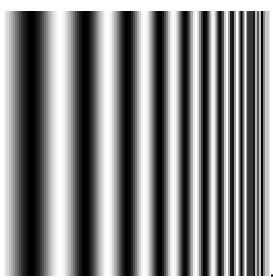


Figure 11a Figure 11b (top) Brightness step $\Delta R = k$ Actual brightnesses plus interpolated values (bottom) Brightness step $\Delta I = k*I$

Although the physical brightness is constant across each vertical stripe, the human observer perceives an "undershoot" and "overshoot" in brightness at what is physically a step edge. Thus, just before the step, we see a slight decrease in brightness compared to the true physical value. After the step we see a slight overshoot in brightness compared to the true physical value. The total effect is one of increased, local, *perceived* contrast at a step edge in brightness.

Spatial Frequency Sensitivity: If the constant intensity (brightness) I_o is replaced by a sinusoidal grating with increasing spatial frequency (Figure 12a), it is possible to determine the spatial frequency sensitivity. The result is shown in Figure 12b .



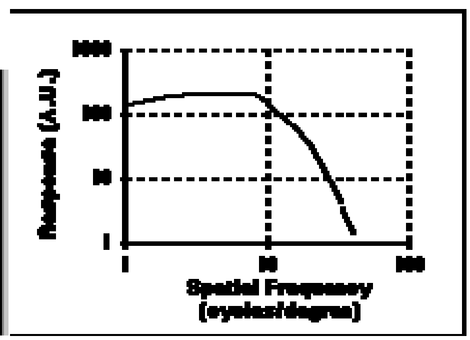


Figure 12a Figure 12b Sinusoidal test grating Spatial frequency sensitivity

To translate these data into common terms, consider an "ideal" computer monitor at a viewing distance of 50 cm. The spatial frequency that will give maximum response is at 10 cycles per degree. (See Figure 12b.) The one degree at 50 cm translates to 50 tan(1deg.) = 0.87 cm on the computer screen. Thus the spatial frequency of maximum response f_{max} = 10 cycles/0.87 cm = 11.46 cycles/cm at this viewing distance. Translating this into a general formula gives:

$$f_{\max} = \frac{10}{d.\tan(1^o)} = \frac{572.9}{d}\ Hz.$$

where d = viewing distance measured in cm.