CHAPTER 1
INTRODUCTION

The proverb 'One picture is worth of thousand words' expressing correctly the amount of information contained in a single picture. Pictures (images) play an important role in the organization of our society as a mass communication medium. Most media (e.g. news papers, TV, cinema) use pictures (still or moving) as information carriers. The tremendous volume of optical information and the needed for its processing and transmission paved the way to image processing by digital computers.

Interest in digital image processing methods stems form two principal application areas.

- Improvement of pictorial information for human interpretation
- Processing of scene data for autonomous machine perception

One of the first application of image processing techniques in the first category was in improving digitized newspaper pictures sent by submarine cable between London and New York. Introduction of the Bart lane cable picture transmission system in the early 1920s reduced the time required to transport a picture across the Atlantic from more than a week to less than three hours.

The relevant efforts started around 1964 at the Jet population laboratory (Pasadena, California) with the digital processing of satellite images coming from moon. Soon, a new branch of science called Digital Image Processing enlarged. Since then, it has exhibited a tremendous growth and created an important technological impact in several areas, e.g. in telecommunication, TV broadcasting, the printing and graphic art industry, medicine and scientific research.

Digital image processing concerns the transmission of an image to a digital format and its processing by digital computers. Both input and output
of a digital image processing system are digital images. Digital image analysis is related to the description and recognition of the digital image component. Its input is a digital image and its output is a symbolic image description. In many cases digital image analysis techniques simulate human vision functions. Therefore, the term computer vision can be used as equivalent to (or the superset of) digital image analysis. Human vision is a very complex neuro-physiological process. Its characteristics are only partially known, despite the tremendous progress that has been made in this area in the past decades. Therefore, its simulation by digital image analysis and computer vision is a very difficult task. In general, the techniques used in the digital image analysis and computer vision differ greatly from the human visual perception mechanisms, although both have similar goals.

From 1964 until present the field of image processing has grown vigorously. In addition to applications in the space program, digital image processing techniques now are used to solve a variety of problems. Although often, unrelated these problems commonly require methods capable of enhancing pictorial information for human interpretation and analysis. In medicine, for instance computer procedures enhance the contrast or code the intensity levels into color which eases interpretation of x-rays and other biomedical images. Geographers use same or similar techniques to study pollution pattern form aerial and satellite images. Image enhancement and restoration procedures are used to process degraded images of unrecoverable objects or experimental results too expensive to duplicate. In archeology, image processing methods have successfully restored blurred pictures that were the only available records of rare artifacts lost or damaged after being photographed. In physics and related fields computer techniques routinely enhances image of experiments in areas such as high energy plasmas and electron microscopy. Similarly successful applications of image processing concepts can be found in astronomy, biology, nuclear medicine, law enforcement, defense, and industrial applications.
The second major area of digital image processing techniques mentioned at the beginning of this section is in solving problems dealing machine perception. In this case, interest focuses on procedures for extracting from image information in a form suitable for computer processing. Often, this information bears little resemblance to visual features that human-being use in interpreting the content of an image. Examples of the type of information used in machine perception are statistical moments, Fourier transform coefficients, and multidimensional distance measures.

Typical problems in machine perception that routinely utilize image processing techniques are automatic character recognition, industrial machine vision for product assembly and inspection, military recognizance, automatic processing of finger prints, screening of x-rays, blood samples and machine processing of aerial and satellite imagery for weather prediction and crop assessment.

1.1. DIGITAL IMAGE DEFINITIONS

A digital image $a[m,n]$ described in a 2D discrete space is derived from an analog image $a(x,y)$ in a 2D continuous space through a sampling process that is frequently referred to as digitization. The mathematics of that sampling process will be described in Section 1.4. For now we will look at some basic definitions associated with the digital image. The effect of digitization is shown in Figure 1.1. The 2D continuous image $a(x,y)$ is divided into $N$ rows and $M$ columns. The intersection of a row and a column is termed a pixel. The value assigned to the integer coordinates $[m,n]$ with $\{m=0,1,2,\ldots,M-1\}$ and $\{n=0,1,2,\ldots,N-1\}$ is $a[m,n]$. In fact, in most cases $a(x,y)$—which we might consider to be the physical signal that impinges on the face of a 2D sensor—is actually a function of many variables including depth ($z$), color($\lambda$), and time($t$). Unless otherwise stated, we will consider the case of 2D, monochromatic, static images in this chapter.
Figure 1.1: Digitization of a continuous image. The pixel at coordinates \([m=10, n=3]\) has the integer brightness value 110.

The image shown in Figure 1.1 has been divided into \(N = 16\) rows and \(M = 16\) columns. The value assigned to every pixel is the average brightness in the pixel rounded to the nearest integer value. The process of representing the amplitude of the 2D signal at a given coordinate as an integer value with \(L\) different gray levels is usually referred to as amplitude quantization or simply quantization.

1.1.1. Common Values

There are standard values for the various parameters encountered in digital image processing. These values can be caused by video standards, by algorithmic requirements, or by the desire to keep digital circuitry simple. Table 1.1 gives some commonly encountered values.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Typical values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>(N)</td>
<td>256, 512, 1024, 2048, 4096</td>
</tr>
<tr>
<td>Columns</td>
<td>(M)</td>
<td>256, 512, 1024, 2048</td>
</tr>
<tr>
<td>Gray Levels</td>
<td>(L)</td>
<td>2, 6.4, 256, 1024, 4096</td>
</tr>
</tbody>
</table>

Table 1.1: Common values of digital image parameters

Quite frequently we see cases of \(M=N=2K\) where \(\{K = 8, 9, 10\}\). This can be motivated by digital circuitry or by the use of certain algorithms such as the (fast) Fourier transform.
The number of distinct gray levels is usually a power of 2, that is, \( L = 2^B \), where \( B \) is the number of bits in the binary representation of the brightness levels. When \( B > 1 \) we speak of a \textit{gray-level image}; when \( B = 1 \) we speak of a \textit{binary image}. In a binary image there are just two gray levels which can be referred to, for example, as "black" and "white" or "0" and "1".

1.1.2. Characteristics of Image Operations

There are varieties of ways to classify and characterize image operations. The reason for doing so is to understand what type of results we might expect to achieve with a given type of operation or what might be the computational burden associated with a given operation.

1.1.2.1. Types of Operations

The types of operations that can be applied to digital images to transform an input image \( a[m,n] \) into an output image \( b[m,n] \) (or another representation) can be classified into three categories as shown in Table 1.2.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Characterization</th>
<th>Generic Complexity: Pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Point}</td>
<td>the output value at a specific coordinate is dependent only on the input value at that same coordinate.</td>
<td>constant</td>
</tr>
<tr>
<td>\textit{Local}</td>
<td>the output value at a specific coordinate is dependent on the input values in the \textit{neighborhood} of that same coordinate.</td>
<td>( P^2 )</td>
</tr>
<tr>
<td>\textit{Global}</td>
<td>the output value at a specific coordinate is dependent on all the values in the input image.</td>
<td>( N^2 )</td>
</tr>
</tbody>
</table>

\textbf{Table 1.2:} Types of image operations. Image size = \( N \times N \); neighborhood size = \( P \times P \) Note that the complexity is specified in operations \textit{per pixel}. 

\textit{Introduction to Image Processing}
1.1.2.2. 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 as illustrated in Figure 1.1. This results in the type of sampling shown in Figure 1.3(a) & (b).
- **Hexagonal sampling** – An alternative sampling scheme is shown in Figure 1.3c and is termed hexagonal sampling.

Both sampling schemes have been studied extensively [1] 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 \( b[m=m_0, n=n_0] \) based upon the pixel values in the neighborhood of \( a[m=m_0, n=n_0] \). Some of the most

---

*Department of Computer Science & Technology, S.K. University, Anantapur*
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 1.3(a) Figure 1.3(b) Figure 1.3(c)
Rectangular sampling Rectangular sampling Hexagonal sampling
4-connected 8-connected 6-connected

1.2. TOOLS

Certain tools are central to the processing of digital images. These include mathematical tools such as convolution, Fourier analysis, and statistical descriptions, and manipulative tools such as chain codes and run codes. We will present these tools without any specific motivation.

1.2.1. Convolution

There are several possible notations to indicate the convolution of two (multidimensional) signals to produce an output signal. The most common are

\[ c = a \oplus b = a \ast b \]  \hspace{1cm} (1)

We shall use the first form, \( c = a \oplus b \), with the following formal definitions. In 2D continuous space:
Introduction to Image Processing

\[
c(x, y) = a(x, y) \ast b(x, y) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} a(j, k) b(x + j - x_{0}, y + k - y_{0})
\]

In 2D discrete space:

\[
c[m, n] = a[m, n] \ast b[m, n] = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a[j, k] b[m - j, n - k]
\]

### 1.2.2. Properties of Convolution

There are a number of important mathematical properties associated with convolution.

- Convolution is **commutative**.
  \[
c = a \ast b = b \ast a
\]

- Convolution is **associative**.
  \[
c = a \ast (b \ast d) = (a \ast b) \ast d = a \ast b \ast d
\]

- Convolution is **distributive**.
  \[
c = a \ast (b + d) = (a \ast b) - (a \ast d)
\]

Where \( a, b, c, \) and \( d \) are all images, either continuous or discrete.

### 1.2.3. Statistics

In image processing it is quite common to use simple statistical descriptions of images and sub-images. The notion of a statistic is intimately connected to the concept of a probability distribution, generally the distribution of signal amplitudes. For a given region—which could conceivably be an entire image—the define of probability distribution function of the brightness in that region and the probability density function of the
brightness in that region are explained. A digitized image $a[m,n]$ is used to represent the above function.

1.2.3.1. Probability Distribution Function of the Brightness

The probability distribution function, $P(a)$, is the probability that a brightness chosen from the region is less than or equal to a given brightness value $a$. As $a$ increases from $-\alpha$ to $+\alpha$, $P(a)$ increases from 0 to 1. $P(a)$ is monotonic, non-decreasing in $a$ and thus $dP/da \geq 0$.

1.2.3.2. Probability Density Function of the Brightness

The probability that a brightness in a region falls between $a$ and $a+\Delta a$, given the probability distribution function $P(a)$, can be expressed as $p(a)\Delta a$ where $p(a)$ is the probability density function:

$$p(a)\Delta a = \frac{dP(a)}{da} \Delta a$$

(7)

Because of the monotonic, non-decreasing character of $P(a)$ we have that:

$$p(a) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} p(a) da = 1$$

(8)

For an image with quantized (integer) brightness amplitudes, the interpretation of $a$ is the width of a brightness interval. Constant width intervals is considered. The brightness probability density function is frequently estimated by counting the number of times that each brightness occurs in the region to generate a histogram, $h[a]$. The histogram can then be normalized so that the total area under the histogram is 1 (eq. (8)). Said another way, the $p[a]$ for a region is the normalized count of the number of pixels, $\Lambda$, in a region that have quantized brightness $a$:

$$p[a] = \frac{1}{\Lambda} h[a] \quad \text{with} \quad \Lambda = \sum_{a} h[a]$$

(9)
Both the distribution function and the histogram as measured from a region are a statistical description of that region. It must be emphasized that both $P[a]$ and $p[a]$ should be viewed as estimates of true distributions when they are computed from a specific region. That is, we view an image and a specific region as one realization of the various random processes involved in the formation of that image and that region. In the same context, the statistics defined below must be viewed as estimates of the underlying parameters.

1.2.3.3. Average

The average brightness of a region is defined as the sample mean of the pixel brightness within that region. The average, $m_o$, of the brightness over the $\Lambda$ pixels within a region $(x_i)$ is given by:

$$m_o = \frac{1}{\Lambda} \sum_{(m,n) \in x_i} o[m,n]$$

(10)

Alternatively, formulation based upon the (unnormalized) brightness histogram is used $h(a) = \Lambda * p(a)$, with discrete brightness values $a$. This gives:
The average brightness, \( m_a \), is an estimate of the mean brightness, \( u_a \), of the underlying brightness probability distribution.

1.2.3.4. Standard Deviation

The unbiased estimate of the standard deviation, \( s_a \), of the brightness within a region \( \Omega \) with \( \Lambda \) pixels is called the sample standard deviation and is given by:

\[
\begin{align*}
  s_a &= \sqrt{\frac{1}{\Lambda - 1} \sum_{m,n \notin \Omega} (a[m,n] - m_a)^2} \\
       &= \sqrt{\frac{\sum_{m,n \notin \Omega} (a[m,n]^2 - (m_a)^2)}{\Lambda - 1}}
\end{align*}
\]

Using the histogram formulation gives:

\[
  s_a = \sqrt{\frac{\sum_{\alpha} \alpha^2 \cdot h[\alpha] - \Lambda \cdot m_a^2}{\Lambda - 1}}
\]

The standard deviation, \( s_a \), is an estimate of \( \sigma_a \) of the underlying brightness probability distribution.

1.2.3.5. Coefficient-of-Variation

The dimensionless coefficient-of-variation, \( CV \), is defined as:
1.2.3.6. Percentiles

The percentile, \( p\% \), of an unquantized brightness distribution is defined as that value of the brightness \( a \) such that or equivalently

\[
P_i(a) = p^a \quad (15)
\]

\[
\int_{-\infty}^{a} p_i(x) \, dx = p^a
\]

Three special cases are frequently used in digital image processing.

- 0% the minimum value in the region
- 50% the median value in the region
- 100% the maximum value in the region

All three of these values can be determined from Figure 1.4.

1.2.3.7. Mode

The mode of the distribution is the most frequent brightness value. There is no guarantee that a mode exists or that it is unique.

1.2.3.8. Signal–to–Noise Ratio

The signal–to–noise ratio, SNR, can have several definitions. The noise is characterized by its standard deviation, \( s_n \). The characterization of the signal can differ. If the signal is known to lie between two boundaries, \( a_{\text{min}} < a < a_{\text{max}} \), then the SNR is defined as
\[ SNR = 20 \log_{10} \left( \frac{O_{\text{max}}}{O_{\text{mean}}} \right) \text{ dB} \]  \hspace{1cm} (16)

If the signal is not bounded but has a statistical distribution then two other definitions are known:

Stochastic signal –

\[ S & \text{N} \text{ inter-dependent} \quad SNR = 20 \log_{10} \left( \frac{m_S}{s_S} \right) \text{ dB} \]  \hspace{1cm} (17)

\[ S & \text{N} \text{ independent} \quad SNR = 20 \log_{10} \left( \frac{m_S}{s_S} \right) \text{ dB} \]  \hspace{1cm} (18)

where \( m_a \) and \( s_a \) are defined above.

The various statistics are given in Table 1.3 for the image and the region shown in Figure 1.5.

![Regions the interior of the Circle](image)

**Figure 1.5.** Regions the interior of the Circle

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Image</th>
<th>ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>137.5</td>
<td>179.3</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>49.5</td>
<td>4.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>56</td>
<td>202</td>
</tr>
<tr>
<td>Median</td>
<td>144</td>
<td>220</td>
</tr>
<tr>
<td>Minimum</td>
<td>24</td>
<td>226</td>
</tr>
<tr>
<td>Mode</td>
<td>62</td>
<td>220</td>
</tr>
<tr>
<td>SNR (dB)</td>
<td>33.3</td>
<td>33.3</td>
</tr>
</tbody>
</table>

**Table 1.3.** Statistics from Figure (1.5)
A SNR calculation for the entire image based on eq. (16) is not directly available. The variations in the image brightness that lead to the large value of $s_{\text{SNR}} (=49.5)$ are not, in general, due to noise but to the variation in local information. With the help of the region there is a way to estimate the SNR. $s_{\text{SNR}} (=4.0)$ and the dynamic range, $a_{\text{max}}-a_{\text{min}}$, for the image ($=241-56$) are used to calculate a global SNR ($=33.3\text{dB}$). The underlying assumptions are the signal is approximately constant in that region and the variation in the region is therefore due to noise, the noise is the same over the entire image with a standard deviation given by $s_n = s_{\text{SNR}}$.

1.2.4. Contour Representations

When dealing with a region or object, several compact representations are available that can facilitate manipulation of, and measurements on the object. An image representation of the object is shown in Figure 1.6 Several techniques exist to represent the region or object by describing its contour.

1.2.4.1. Chain Code

This representation is based upon the work of Freeman [2,5]. For the standard implementation of the chain code consider a contour pixel to be an object pixel that has a background (non-object) pixel as one or more of its 4-connected neighbors. See Figures 1.3. The codes associated with eight possible directions are the chain codes and, with $x$ as the current contour pixel position, the codes are generally defined as

$$
\begin{align*}
\text{Chain codes} &= 3 \ 2 \ 1 \\
&\quad 4 \ x \ 0 \\
&\quad 5 \ 6 \ 7 \\
\end{align*}
$$

(19)
Figure 1.6. Region (Shaded) as it is transformed from continuous to discrete form and the considered as a contour or run lengths illustrated in altering colors.

1.2.4.2. Chain Code Properties

- Even codes \{0,2,4,6\} correspond to horizontal and vertical directions; odd codes \{1,3,5,7\} correspond to the diagonal directions.

- Each code can be considered as the angular direction, in multiples of 45 deg., that we must move to go from one contour pixel to the next.

- The absolute coordinates \([m,n]\) of the first contour pixel (e.g. top, leftmost) together with the chain code of the contour represent a complete description of the discrete region contour.

- When there is a change between two consecutive chain codes, then the contour has changed direction. This point is defined as a corner.
1.2.4.3. "Crack" Code

An alternative to the chain code for contour encoding is to use neither the contour pixels associated with the object nor the contour pixels associated with background but rather the line, the "crack", in between. This is illustrated with an enlargement of a portion of Figure 1.6 in Figure 1.7.

The "crack" code can be viewed as a chain code with four possible directions instead of eight.

\[
\begin{align*}
\text{Crack code} & = 2 \quad x \quad 0 \\
& = 3
\end{align*}
\]

Figure 1.7. (a). Object including part to be studied. (b) Contour pixels re used in the chain code or diagonally shaded. The “Crack” is shown with the thick black line

The chain code for the enlarged section of Figure 1.7b, from top to bottom, is \{5,6,7,7,0\}. The crack code is \{3,2,3,0,3,0,0\}.

1.2.4.4. Run codes

A third representation is based on coding the consecutive pixels along a row—a run—that belong to an object by giving the starting position of the run and the ending position of the run. Such runs are illustrated in Figure 1.6. There are a number of alternatives for the precise definition of the positions. Which alternative should be used depends upon the application.
1.3 HUMAN VISUAL SYSTEM AND PERCEPTUALITY

Many image processing applications are intended to produce images that are to be viewed by human observers (as opposed to, say, automated industrial inspection.) 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 i) the human visual system is not well understood, ii) no objective measure exists for judging the quality of an image that corresponds to human assessment of image quality, and, iii) the "typical" human observer does not exist. Nevertheless, research in perceptual psychology has provided some important insights into the visual system.

Before the output of an imaging system is presented to a human observer, it is essential to consider how it is transformed into information by the viewer. Understanding of the visual perception process is important for developing measures of image fidelity. Which aid in the design and evaluation of image processing algorithms and imaging systems. Visual image data itself represents spatial distribution of physical quantities such as luminance and spatial frequencies of an object. The perceived information may be represented by attributes such as brightness, color and edges.

Light is the electromagnetic radiation that stimulates our visual response. It is expressed as a spectral energy distribution $L(\lambda)$, where $\lambda$ is the wavelength that lies in the visible region, 350nm to 780nm, of the electromagnetic spectrum. Light received from an object can be written as

$$I(\lambda) = \rho(\lambda) \alpha(\lambda)$$  \hspace{1cm} (21)

Where $\rho(\lambda)$ represents the reflectivity or transmissivity of the object and $\alpha(\lambda)$ is the incident energy distribution. The illumination range over which the visual system can operate is brought 1 to $10^{10}$, or 10 orders of magnitude. The retina of the human eye contains two types of photo receptors called rods and
cones. The rods which are more in number (at out 100 million) are long and thin. They provide scotopic vision, which is the visual response at the lower orders of magnitude of illumination. The cones, which are fewer in number (about 6.5 million) are shorter and thicker and are less sensitive than rods. They provide photopic vision, which is the visual response for higher orders of magnitude of illumination. In the intermediate region of illumination, both rods and cones are active and provide mesopic vision. The cones are responsible for colour vision. The pupil of the eye acts as an aperture. In bright light it is about 2mm in diameter and acts as a low pass filter. (Fig 1.8)

![Cross Section of Eye](image1.png)  ![Relative Luminous Efficiency Function](image2.png)

Figure 1.8 (a) Cross Section of Eye  Figure 1.8. (b) Relative Luminous efficiency

The luminance or intensity of a spatially distributed object with light distribution $I(X,Y,\lambda)$ is given as

$$f(s,\gamma)=\int I(X,Y,\lambda)(\lambda)\,d\lambda$$  (22)

Where $V(\lambda)$ is called the relative luminous efficiency function of the visual system. For the human eye, $V(\lambda)$ is a belt shaped curve who characteristics depend on whether it is scotopic or photopic vision. The luminance of an object is independent of the luminances of the surrounding objects. The brightness (or apparent brightness) of an object is the perceived luminance and depends on the luminance of the surround. Two objects with different surroundings could have identical luminances but different brightness.
1.3.1 PRECEPTUALITY CONCEPT

Our perception is sensitive to luminance contrast rather than the absolute luminance values themselves. According to Weber's law if the luminance \( f_s \) of an object is just noticeably different from the luminance \( f_o \) of its surround, then their ratio is \( \frac{|f_s - f_o|}{f_o} = \text{constant} \). This can also be written as

\[
\frac{\Delta f}{f} = d(\log f) \Delta C \quad \text{where} \quad \Delta f \text{ is small for just noticeably different luminances.}
\]

The spatial interaction of luminances from an object and its surround creates a phenomenon called the mach band effect. The Mach band effect measures the response of the visual system in spatial coordinates. A direct measurement of the MTF is possible by considering a sinusoidal grating of varying contract (ratio of the maximum to minimum intensity) and spatial frequency. Human visual system is most sensitive to mid-frequencies and least sensitive to high frequencies. The impulse response of the visual system, manifests a phenomenon known as lateral inhibition. The impulse response values represent the relative spatial weighting by the receptors, rods and tones.

A simple monochrome vision model is as shown in fig. 1.9

![Figure 1.9. Monochrome vision model](image)

Light enters the eye, whose optical characteristics can be represented by a low pass filter \( H_1(\xi_1, \xi_2) \). The spatial response of the eye, represented by the relative luminous efficiency function \( V(\lambda) \), yields the luminance distribution \( f(x,y) \). The non-linear response of the rods and cones, represented by the point non-linearity \( g(*) \) yields contrast \( c(x,y) \). The lateral inhibition phenomenon is represented by a spatially invariant, linear system whose frequency response is \( H(\xi_1, \xi_2) \). Its output is the apparent brightness \( b(x,y) \).
For an optically well-corrected eye, the low pass filter has a much shower drop-off with increasing frequency that of lateral inhibition mechanism.

The methods which are based on models of the HVS usually utilize some of the conventional techniques such as block DCTs with perceptually based bit allocation scheme. The ultimate user of a lossy compressed image communication system is a human being. But the visual system is not that simple. The frequency sensitivity of the human visual system (represented by modulation transfer function (MTF)) can be utilized. This function describes the human eye's sensitivity to sine wave gratings at various frequencies. As DCT-based coders use cosine inputs, it is not directly usable for DCT-based coders. Hence, threshold based system is to be used.

Given that the minimum viewing distance is fixed, it is possible to determine a static just noticeable difference (JND) threshold for each frequency band. The concept of determining a JND threshold for each location in an image is central to many of the perceptually based approaches for image compression. An other property known as contrast sensitivity can also be included in the perceptual based coders. Contrast sensitivity is a way to measure the effect of the detectability threshold of noise on a constant background. This phenomenon depends on the gray level of the back ground as well as the contrast level of the noise. For the Human Visual System, this is a non-linear function. Contrast sensitivity can be included implicitly by developing functions of measuring threshold versus local brightness.

In addition to these effects, another useful property for compression is Visual masking (or contrast masking). Masking is the Phenomena where the detectability of a signal component depends on the presence or absence of other signal components in its immediate vicinity. All of these properties depend on the local scene content and therefore have the desired property of local control. The local JND level depends on the MTF of the human visual system, the local brightness sensitivity, and a measure of the local texture. In order to achieve the highest subjective quality at a given bit rate, all of these properties must be utilized.
It is well known that the statistics of image signals are quite non-stationary and the fidelity of the reconstructed images demanded by the human eye differs from pixel to pixel. Consequently, the essential task of perceptual coding is to effectively adopt the coding algorithm to the sensitivity of the human eye. A variety of methods have been proposed to incorporate certain Psychovisual properties of the human Visual system (HVS) into image coding algorithms. Some methods exploit the sensitivity of the HVS to spatial frequency for adopting the quantizer step size in frequency domain. The other methods tried to make efficient use of spatial masking effects to hide distortion in spatial domain. The just-noticeable distortion (JND) is a key concept in perceptual coding. JND threshold should be such that the distortion presented in the reconstructed image is minimally perceptible and appears to be uniformly distributed.

The individual components of the eye work in a manner similar to a camera. Each part plays a vital role in providing clear vision. So think of the eye as a camera with the cornea, behaving much like a lens cover. As the eye's main focusing element, the cornea takes widely diverging rays of light and bends them through the pupil, the dark, round opening in the center of the colored iris. The iris and pupil act like the aperture of a camera. Next in line is the lens which acts like the lens in a camera, helping to focus light to the back of the eye. Note that the lens is the part which becomes cloudy and is removed during cataract surgery to be replaced by an artificial implant nowadays.

Figure 1.10 Working process of the Human Eye
The very back of the eye is lined with a layer called the retina which acts very much like the film of the camera. The retina is a membrane containing photoreceptor nerve cells that lines the inside back wall of the eye. The photoreceptor nerve cells of the retina change the light rays into electrical impulses and send them through the optic nerve to the brain where an image is perceived. The center 10% of the retina is called the macula. This is responsible for your sharp vision, your reading vision. The peripheral retina is responsible for the peripheral vision. As with the camera, if the "film" is bad in the eye (i.e. the retina), no matter how good the rest of the eye is, you will not get a good picture. The human eye is remarkable. It accommodates to changing lighting conditions and focuses light rays originating from various distances from the eye. When all of the components of the eye function properly, light is converted to impulses and conveyed to the brain where an image is perceived.
Anterior Chamber

The cavity in the front part of the eye between the lens and cornea is called the Anterior Chamber. It is filled with Aqueous, a water-like fluid. This fluid is produced by the ciliary body and drains back into the blood circulation through channels in the chamber angle. It is turned over every 100 minutes.

Chamber Angle

Located at the junction of the cornea, iris, and sclera, the anterior chamber angle extends 360 degrees at the perimeter of the iris. Channels here allow aqueous fluid to drain back into the blood circulation from the eye. May be obstructed in glaucoma.

Ciliary Body

A structure located behind the iris (rarely visible) which produces aqueous fluid that fills the front part of the eye and thus maintains the eye pressure. It also allows focusing of the lens.

Conjunctiva

A thin lining over the sclera, or white part of the eye. This also lines the inside of the eyelids. Cell in the conjunctiva produce mucous, which helps to lubricate the eye.

Cornea

The transparent, outer "window" and primary focusing element of the eye. The outer layer of the cornea is known as epithelium. Its main job is to protect the eye. The epithelium is made up of transparent cells that have the ability to regenerate quickly. The inner layer of the cornea is also made up of transparent tissue, which allows light to pass.
Hyaloid Canal

A narrow channel that runs from the optic disc to the back surface of the lens. It serves an embryologic function prior to birth but none afterwards.

Iris

Inside the anterior chamber is the iris. This is the part of the eye which is responsible for one's eye color. It acts like the diaphragm of a camera, dilating and constricting the pupil to allow more or less light into the eye.

Pupil

The dark opening in the center of the colored iris that controls how much light enters the eye. The colored iris functions like the iris of a camera, opening and closing, to control the amount of light entering through the pupil.

Lens

The part of the eye immediately behind the iris that performs delicate focusing of light rays upon the retina. In persons under 40, the lens is soft and pliable, allowing for fine focusing from a wide variety of distances. For individuals over 40, the lens begins to become less pliable, making focusing upon objects near to the eye more difficult. This is known as presbyopia.

Macula

The part of the retina which is most sensitive, and is responsible for the central (or reading) vision. It is located near the optic nerve directly at the back of the eye (on the inside). This area is also responsible for color vision.

Optic Disc

The position in the back of the eye where the nerve (along with an artery and vein) enters the eye corresponds to the "blind spot" since there are no rods or cones in these location. Normally, a person does not notice this
blind spot since rapid movements of the eye and processing in the brain compensate for this absent information. This is the area that the ophthalmologist studies when evaluating a patient for glaucoma, a condition where the optic nerve becomes damaged often due to high pressure within the eye. As it looks like a cup when viewed with an ophthalmoscope, it is sometimes referred to as the Optic Cup.

**Optic Nerve**

The optic nerve is the structure which takes the information from the retina as electrical signals and delivers it to the brain where this information is interpreted as a visual image. The optic nerve consists of a bundle of about one million nerve fibers.

**Retina**

The membrane lining the back of the eye that contains photoreceptor cells. These photoreceptor nerve cells react to the presence and intensity of light by sending an impulse to the brain via the optic nerve. In the brain, the multitude of nerve impulses received from the photoreceptor cells in the retina are assimilated into an image.

**Sclera**

The white, tough wall of the eye. Few diseases affect this layer. It is covered by the episclera (a fibrous layer between the conjunctiva and sclera) and conjunctiva, and eye muscles are connected to this.

**Vitreous**

Next in our voyage through the eye is the vitreous. This is a jelly-like substance that fills the body of the eye. It is normally clear. In early life, it is firmly attached to the retina behind it. With age, the vitreous becomes more water-like and may detach from the retina. Often, little clumps or strands of the jelly form and cast shadows which are perceived as "floaters". While
frequently benign, sometimes floaters can be a sign of a more serious condition such as a retinal tear or detachment and should be investigated with a thorough ophthalmologic examination.

1.3.2 Brightness 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.

1.3.2.1. Wavelength Sensitivity

The perceived intensity as a function of \( \lambda \), the spectral sensitivity, for the "typical observer" is shown in Figure 1.13 [2].

![Figure 1.13 Spectral Sensitivity of the "typical" human observer](image)

\[ R = \log(I_o) \] (23)
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 1.14 ab.

A horizontal line through the top portion of Figure 1.14a shows a linear increase in objective brightness (Figure 1.14b) 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 1.14b) but a linear increase in subjective brightness.

![Figure 1.14. (a). (Top) Brightness step $\Delta I = \frac{k}{x}$ (Bottom) Brightness step $\Delta I = k \times x$ (b). Actual brightnesses plus interpolated values](image)

The Mach band effect is visible in Figure 1.14. 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, there is a slight decrease in brightness compared to the true physical value. After the step there is 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.
1.3.3. Spatial Frequency Sensitivity

If the constant intensity (brightness) \( I_0 \) is replaced by a sinusoidal grating with increasing spatial frequency (Figure 1.15a), it is possible to determine the spatial frequency sensitivity. The result is shown in Figure 1.15b [3, 4].

![Figure 1.15](image.png)

**Figure 1.15.** (a). Sinusoidal test grating (b). 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 \( \tan(1_{\text{deg}}) = 0.87 \) cm on the computer screen. Thus the spatial frequency of maximum response \( f_{\text{max}} = 10 \) cycles/0.87 cm = 11.46 cycles/cm at this viewing distance. Translating this into a general formula gives

\[
f_{\text{max}} = \frac{10}{d \cdot \tan(1^{\circ})} = \frac{57.29}{d} \text{ cycles/cm}
\]

where \( d = \text{viewing distance measured in cm.} \)

1.3.4. Color Sensitivity

Human color perception is an exceedingly complex topic. As such is present a brief introduction here. The physical perception of color is based upon three color pigments in the retina.
1.3.4.1. Standard Observer

Based upon psychophysical measurements, standard curves have been adopted by the CIE (Commission Internationale de l'Eclairage) as the sensitivity curves for the "typical" observer for the three "pigments" $V(\lambda)$, $U(\lambda)$, $X(\lambda)$. These are shown in Figure 1.16. These are not the actual pigment absorption characteristics found in the "standard" human retina but rather sensitivity curves derived from actual data [5].

![Figure 1.16. Standard observer spectral sensitivity curves](image)

For an arbitrary homogeneous region in an image that has an intensity as a function of wavelength (color) given by $I(\lambda)$, the three responses are called the tristimulus values

$$X = \int_0^\infty I(\lambda) V(\lambda) \, d\lambda, \quad Y = \int_0^\infty I(\lambda) U(\lambda) \, d\lambda, \quad Z = \int_0^\infty I(\lambda) X(\lambda) \, d\lambda. \quad (25)$$

1.3.4.2. CIE Chromaticity Coordinates

The chromaticity coordinates which describe the perceived color information are defined as

$$X = \frac{X}{X - Y - Z}, \quad Y = \frac{Y}{X - Y - Z}, \quad Z = 1 - (X + Y) \quad (26)$$
The red chromaticity coordinate is given by $x$ and the green chromaticity coordinate by $y$. The tristimulus values are linear in $I(\lambda)$ and thus the absolute intensity information has been lost in the calculation of the chromaticity coordinates $\{x,y\}$. All color distributions, $I(\lambda)$, that appear to an observer as having the same color will have the same chromaticity coordinates.

If a tunable source of pure color is used (such as a dye laser), then the intensity can be modeled as $I(\lambda) = d(\lambda - \lambda_0)$ with $d(\cdot)$ as the impulse function. The collection of chromaticity coordinates $\{x,y\}$ that will be generated by varying $\lambda_0$ gives the CIE chromaticity triangle as shown in Figure 1.17.

![Figure 1.17. Chromaticity diagram containing the CIE chromaticity triangle associated with pure spectral colors and triangle associated with CRT phosphors.](image)

Pure spectral colors are along the boundary of the chromaticity triangle. All other colors are inside the triangle. The chromaticity coordinates for some standard sources are given in Table 1.4.

<table>
<thead>
<tr>
<th>Source</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluorescent lamp ($\lambda_0 = 4800 , \text{K}$)</td>
<td>0.35</td>
<td>0.37</td>
</tr>
<tr>
<td>Sun ($\lambda_0 = 6000 , \text{K}$)</td>
<td>0.32</td>
<td>0.33</td>
</tr>
<tr>
<td>Red Phosphor (cathodoluminescent)</td>
<td>0.68</td>
<td>0.32</td>
</tr>
<tr>
<td>Green Phosphor (zinc cadmium sulfide)</td>
<td>0.28</td>
<td>0.60</td>
</tr>
<tr>
<td>Blue Phosphor (zinc sulfide)</td>
<td>0.15</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 1.4: Chromaticity coordinates for standard sources.
The description of color on the basis of chromaticity coordinates not only permits an analysis of color but provides a synthesis technique as well. Using a mixture of two color sources, it is possible to generate any of the colors along the line connecting their respective chromaticity coordinates. Since there is no negative number of photons, this means the mixing coefficients must be positive. Using three color sources such as the red, green, and blue phosphors on CRT monitors leads to the set of colors defined by the interior of the "phosphor triangle" shown in Figure 1.17.

The formulas for converting from the tristimulus values \((X,Y,Z)\) to the well-known CRT colors \((R,G,B)\) and back are given by:

\[
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix} = 
\begin{bmatrix}
1.9107 & -0.5526 & -0.2883 \\
-0.9843 & 1.9984 & -0.0283 \\
0.0583 & -0.1185 & 0.8986
\end{bmatrix} \cdot
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}
\] (27)

and

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} = 
\begin{bmatrix}
0.6067 & 0.1736 & 0.2001 \\
0.2988 & 0.5868 & 0.1143 \\
0.0000 & 0.0661 & 1.1149
\end{bmatrix} \cdot
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\] (28)

As long as the position of a desired color \((X,Y,Z)\) is inside the phosphor triangle in Figure 1.17, the values of \(R, G,\) and \(B\) as computed by eq. (27) will be positive and can therefore be used to drive a CRT monitor.

It is incorrect to assume that a small displacement anywhere in the chromaticity diagram (Figure 1.17) will produce a proportionally small change in the perceived color. An empirically-derived chromaticity space where this property is approximated is the \((u',v')\) space.
The description of the human visual system presented above is couched in standard engineering terms. This could lead one to conclude that there is sufficient knowledge of the human visual system to permit modeling the visual system with standard system analysis techniques. Two simple examples of optical illusions, shown in Figure 1.18, illustrate that this system approach would be a gross oversimplification. Such models should only be used with extreme care.

![Image of optical illusions](image)

**Figure 1.18. Optical illusions**

The left illusion induces the illusion of gray values in the eye that the brain "knows" does not exist. Further, there is a sense of dynamic change in the image due, in part, to the saccadic movements of the eye. The right illusion, Kanizsa's triangle, shows enhanced contrast and false contours [3] neither of which can be explained by the system-oriented aspects of visual perception described above.
1.4. IMAGE SAMPLING

Converting from a continuous image \( a(x,y) \) to its digital representation \( b[m,n] \) requires the process of sampling. In the ideal sampling system \( a(x,y) \) is multiplied by an ideal 2D impulse train

\[
b_{\text{ideal}}[m,n] = a(x,y) \bullet \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x - mX_0, y - nY_0) \tag{30}
\]

\[
= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a(mX_0, nY_0) \delta(x - mX_0, y - nY_0)
\]

where \( X_0 \) and \( Y_0 \) are the sampling distances or intervals and \( \delta(*) \) is the ideal impulse function. (At some point, of course, the impulse function \( d(x,y) \) is converted to the discrete impulse function \( d[m,n] \).) Square sampling implies that \( X_0 = Y_0 \). Sampling with an impulse function corresponds to sampling with an infinitesimally small point. This, however, does not correspond to the usual situation as illustrated in Figure 1.1. To take the effects of a finite sampling aperture \( p(x,y) \) into account, we can modify the sampling model as follows

\[
b[m,n] = (a(x,y) \circ p(x,y)) \bullet \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(x - mX_0, y - nY_0) \tag{31}
\]

The combined effect of the aperture and sampling are best understood by examining the Fourier domain representation.

\[
B(\Omega, \Psi) = \frac{1}{4\pi^2} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} A(\Omega - m\Omega_x, \Psi - m\Psi_x) \ast P(\Omega - m\Omega_x, \Psi - m\Psi_y) \tag{32}
\]

where \( \Omega_x = 2\pi/X_0 \) is the sampling frequency in the \( x \) direction and \( \psi_y = 2\pi/Y_0 \) is the sampling frequency in the \( y \) direction. The aperture \( p(x,y) \) is frequently square, circular, or Gaussian with the associated \( P(\Omega, \Psi) \). (See Table 1.5.)
### T.1 Rectangle

\[ R_{ab}(x, y) = \frac{1}{4ab} e^{-\gamma x^2 - \beta y^2} \]

\[ F \leftarrow \frac{\sin(ab \omega, \theta)}{ab \omega, \theta} \]

### T.2 Pyramid

\[ R_{ab}(x, y) \triangleq R_{ab}(x, y) \]

\[ F \leftarrow \frac{\sin(ab \omega, \theta)}{ab \omega, \theta} \]

### T.3 Cylinder

\[ P_u(x) = \frac{u(x^2 - r^2)}{\pi a^2} \]

\[ F \leftarrow 2 \left( J_0(\omega) \right)^2 \]

### T.4 Cone

\[ P_u(x) \triangleq P_u(x) \]

\[ F \leftarrow \frac{4 J_1(\omega)^2}{\omega} \]
### Table 1.5: 2D Images and their Fourier Transforms
1.4.1. Sampling Density for Image Processing

To prevent the possible aliasing (overlapping) of spectral terms that is inherent in eq. (32) two conditions must hold:

\[ | A(u, v) | = 0 \quad \text{for} \quad |u| > u_c \quad \text{and} \quad |v| > v_c \quad (33) \]

Nyquist Sampling Frequency

\[ \Omega_s > 2 \cdot u_c \quad \text{and} \quad \Psi_s > 2 \cdot v_c \quad (34) \]

where \( u_c \) and \( v_c \) are the cutoff frequencies in the \( x \) and \( y \) direction, respectively. Images that are acquired through lenses that are circularly-symmetric, aberration free, and diffraction-limited will, in general, be band limited. The lens acts as a low pass filter with a cutoff frequency in the frequency domain given by

\[ u_c = v_c = \frac{\lambda}{\sqrt{4F^2 - 1}} \quad (35) \]

where \( NA \) is the numerical aperture of the lens and \( \lambda \) is the shortest wavelength of light used with the lens [7]. If the lens does not meet one or more of these assumptions then it will still be bandlimited but at lower cutoff frequencies than those given in eq.(35). When working with the F-number (\( F \)) of the optics instead of the \( NA \) and in air (with index of refraction = 1.0), eq. (35) becomes:

\[ u_c = v_c = \frac{\lambda}{\sqrt{4F^2 - 1}} \quad (36) \]
1.4.1.1. Sampling Aperture

The aperture \( p(x,y) \) described above will have only a marginal effect on the final signal if the two conditions eqs. (34) and (35) are satisfied. Given, for example, the distance between samples \( X_0 \) equals \( Y_0 \) and a sampling aperture that is not wider than \( X_0 \), the effect on the overall spectrum—due to the \( A(u,v)P(u,v) \) behavior implied by eq.(31)—is illustrated in Figure 1.19 for square and Gaussian apertures.

The spectra are evaluated along one axis of the 2D Fourier transform. The Gaussian aperture in Figure 1.19 has a width such that the sampling interval \( X_0 \) contains \( \pm 3\sigma \) (99.7%) of the Gaussian. The rectangular apertures have a width such that one occupies 95% of the sampling interval and the other occupies 50% of the sampling interval. The 95% width translates to a fill factor of 90% and the 50% width to a fill factor of 25%.

![Figure 1.19. Aperture spectrum \( p(u,v) = 0 \) for frequencies up to half the Nyquist frequency](image)

1.4.2. Sampling Density for Image Analysis

The "rules" for choosing the sampling density when the goal is image analysis—as opposed to image processing—are different. The fundamental difference is that the digitization of objects in an image into a collection of pixels introduces a form of spatial quantization noise that is not band limited.
This leads to the following results for the choice of sampling density when one is interested in the measurement of area and (perimeter) length.

### 1.4.2.1. Sampling for Area Measurements

Assuming square sampling, $X_0 = Y_0$ and the unbiased algorithm for estimating area which involves simple pixel counting, the $CV$ (see eq. (38)) of the area measurement is related to the sampling density by [6, 7]:

$$2^D: \lim_{S \to \infty} CV(S) = k_1 S^{-\frac{1}{2}} \quad 3^D: \lim_{S \to \infty} CV(S) = k_3 S^{-2}$$

and in $D$ Dimension

$$\lim_{S \to \infty} CV(S) = k_3 S^{-(D-1)}$$

where $S$ is the number of samples per object diameter. In 2D the measurement area, in 3D volume, and in $D$-dimensions hyper volume.

### 1.4.2.2. Sampling for Length Measurements

Again assuming square sampling and algorithms for estimating length based upon the Freeman chain-code representation, the $CV$ of the length measurement is related to the sampling density per unit length as shown in Figure 1.20 (see [7].)

![Figure 1.20. CV of length for measurement for various algorithms](image-url)
The curves in Figure 1.20 were developed in the context of straight lines but similar results have been found for curves and closed contours. The specific formulas for length estimation use a chain code representation of a line and are based upon a linear combination of three numbers

\[ L = \alpha \cdot N_e - \beta \cdot N_o + \gamma \cdot N_c \]  

where \( N_e \) is the number of even chain codes, \( N_o \) the number of odd chain codes, and \( N_c \) the number of corners. The specific formulas are given in Table 1.6.

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Formula</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \gamma )</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel count</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>Freeman</td>
<td>1</td>
<td>( \sqrt{2} )</td>
<td>0</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Kulp</td>
<td>0.9-81</td>
<td>0.9-81 ( \sqrt{2} )</td>
<td>0</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Corner count</td>
<td>0.980</td>
<td>1.426</td>
<td>-0.591</td>
<td>2.1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.6. Length estimation formulas based on chain code counts (\( N_e, N_o, N_c \))

1.4.2.3. Conclusions on Sampling

If one is interested in image processing, one should choose a sampling density based upon classical signal theory, that is, the Nyquist sampling theory. If one is interested in image analysis, one should choose a sampling density based upon the desired measurement accuracy (bias) and precision (CV). In a case of uncertainty, one should choose the higher of the two sampling densities (frequencies).