

Chapter 1

Introduction

Computer Vision

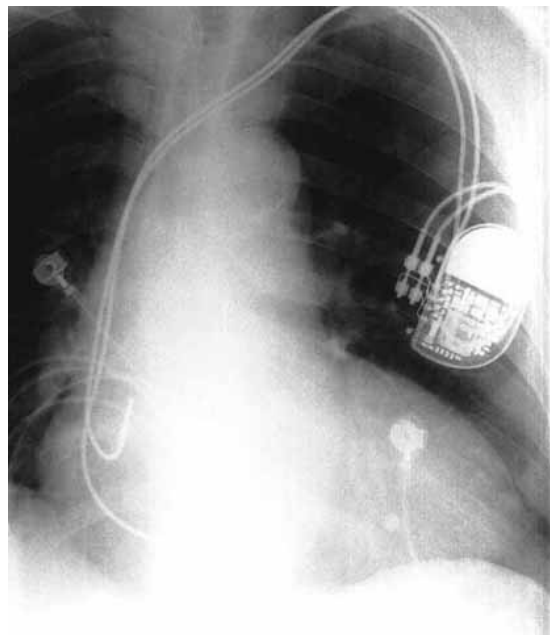
Acquisition and processing of visual information by the computer.

Applications involve tasks that require

- High rate of processing of information
- Access to and use of large data bases
- Work in hostile environments

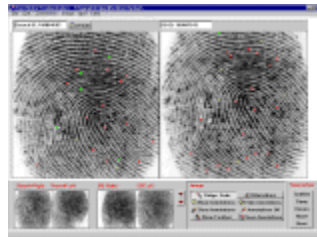
Applications

- **Medical systems**, e.g. diagnosis of tumor, aid to neurosurgeon during brain surgery.

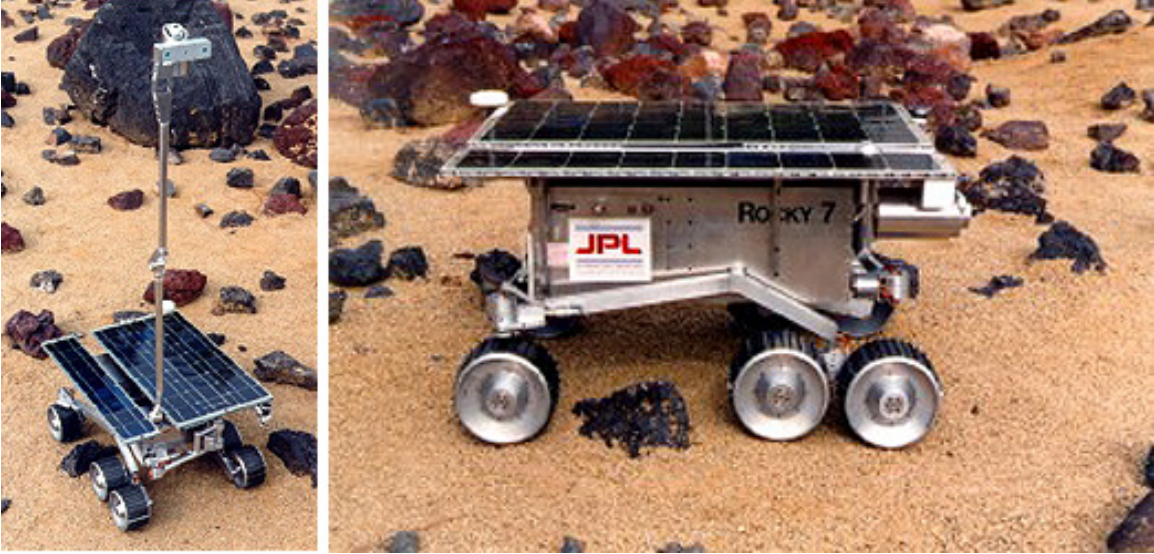




- **Security systems**, e.g. to identify people by facial scan, vein in hands.
- Manufacturing systems for quality control.
- Law enforcement, e.g. automatic identification of fingerprints, identifies traffic violators.



- Defense, e.g. automatic target tracking.
- Satellite information gathering, e.g. process images obtained from satellite for making maps, predicting weather, military reconnaissance
- Space applications, e.g. planetary explorations.



1.1 Energy Sources for Image Acquisition

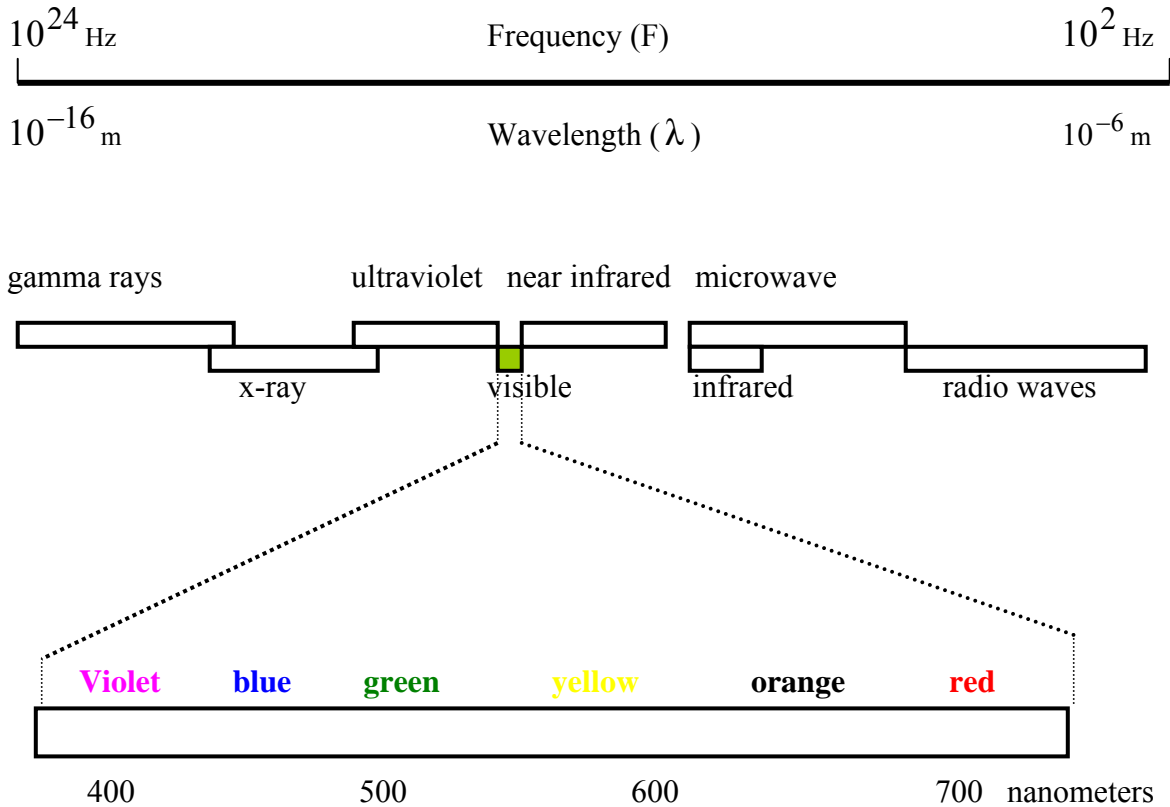
Imaging: sensing the environment and representing sensed data as an image

- **Passive imaging:** uses energy sources that are already present in the scene – e.g. light from the sun energy.
- **Active imaging:** uses artificial energy source to probe the environment, e.g. radiation in medical field

Light energy

- (a) inherently safe
- (b) can be generated reliably and cheaply,
- (c) easy to control and
- (d) can be detected easily using various sensors.

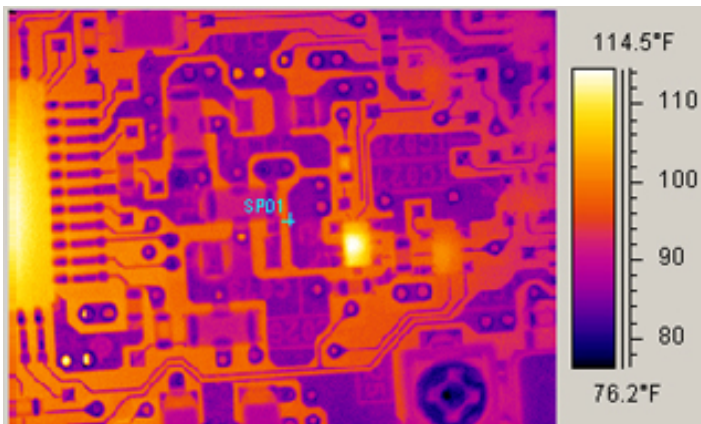
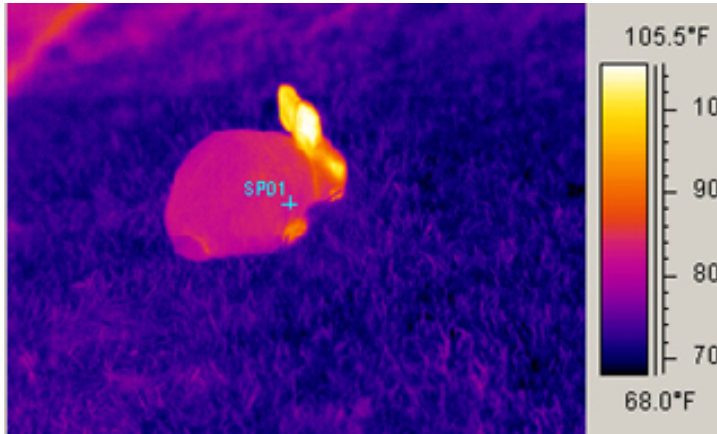
However, all regions of electromagnetic (EM) spectrum are suited to imaging.



Note: $\lambda f = c = 10^8 \text{ m/s}$

- Images acquired at different wavelengths have different properties.
- The visible portion of the spectrum is between 400 and 700 nanometers (10^{-9} m).
- In the x-ray region ($\lambda \cong 10^{-10}$ m), the wave has sufficient energy to penetrate into material and therefore x-ray images reveal internal structure of objects and human body.
- Gamma rays are highly penetrating. Typically a patient drinks a substance that is “tagged” with radioactive tracer. This tracer is taken up in varying amounts by different body tissues according to level of tissue activity. A gamma camera collects ray photons emitted by body tissues, and diseased tissue such as a tumor will appear as a bright region in the image.

- Infrared (IR) radiations ($\lambda > 700 \text{ nm}$) are emitted from warm objects, and IR imaging can be used to locate people, moving vehicles, etc. in total darkness



- Synthetic aperture radar (SAR) imaging with $\lambda = 1 - 100 \text{ cm}$, use an artificially generated source of microwaves that are unaffected by clouds and provide information about planes, and surfaces of other planets (e.g. venus).

1.2 The Electronic Camera

A camera uses a lens to focus part of the visual environment onto a sensor.

$$\frac{x}{f} = -\frac{X}{Z-f} \quad \text{and} \quad \frac{y}{f} = -\frac{Y}{Z-f} \quad (1.1)$$

negative sign indicates that the image points are inverted.

The magnification factor μ specifies the lens magnifying power as

$$\mu = \frac{y}{Y} = (\text{image size})/(\text{object size}) \quad (1.2)$$

Usually, the magnifying power of a lens is expressed in terms of its focal length f . Ignoring the inversion (i.e. the negative sign in front of X and Y), the focal length in term of magnification factor is obtained from (1.1) and (1.2) as

$$f = \frac{\mu Z}{\mu + 1} \quad (1.3)$$

where Z is the distance of the object from the image plane.

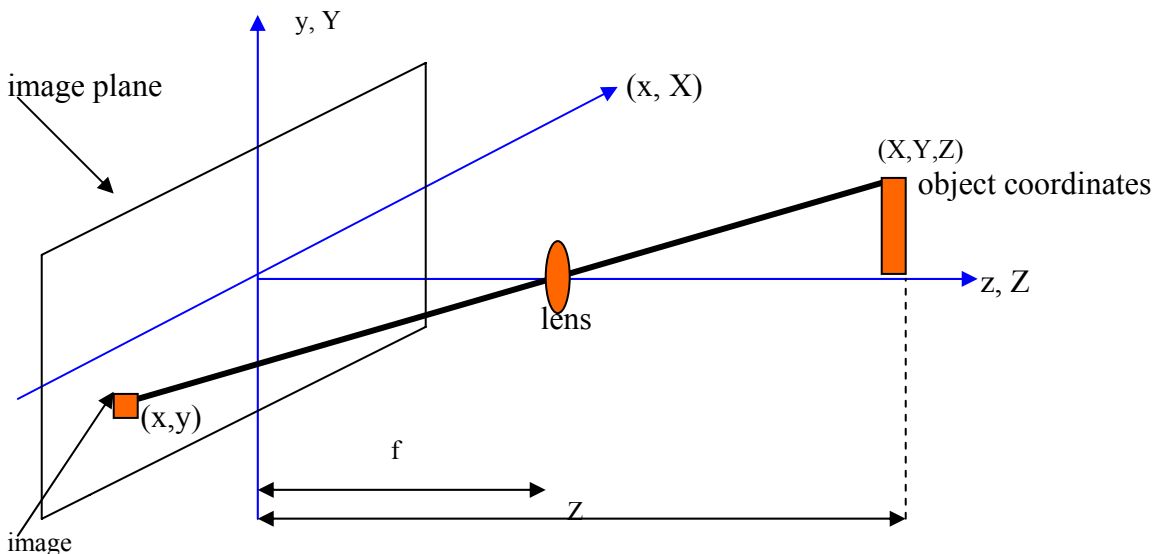


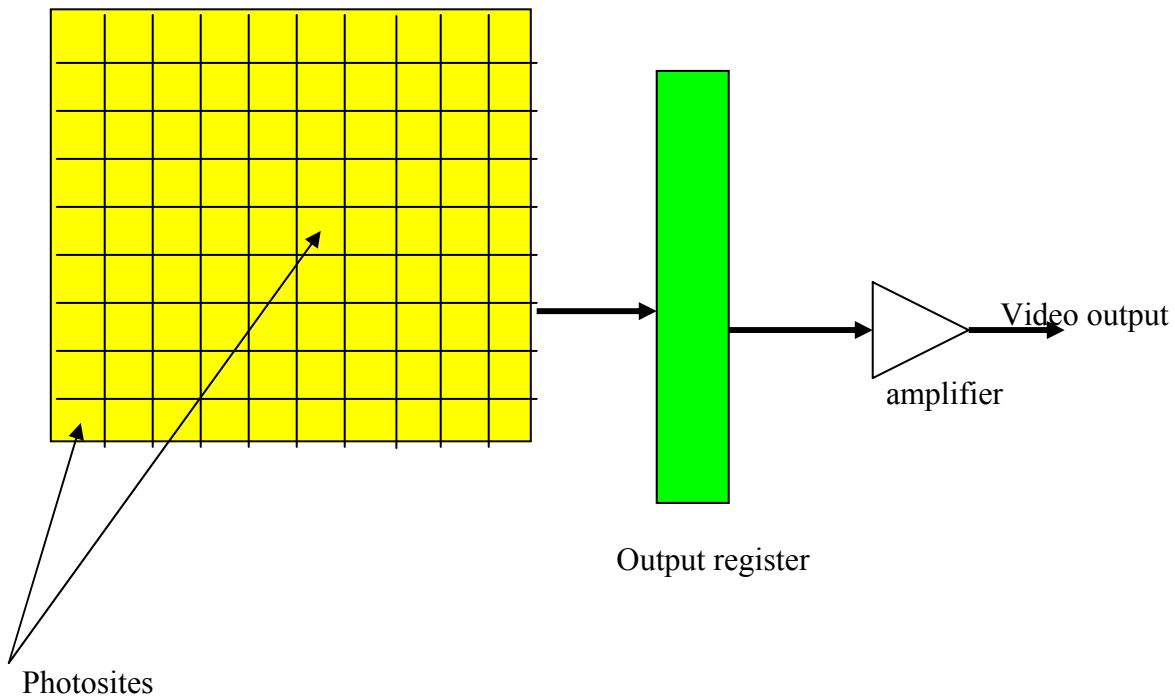
Fig. 1.2 Image geometry

The light gathering capacity of a camera lens is determined by its **aperture**. This is less than the lens diameter, and is usually made smaller by means of an adjustable circular hole or diaphragm.

The aperture is expressed in terms of the ratio of focal length to aperture diameter D_a , i.e. f / D_a , and is referred to as the “*f* number”. Most cameras have a sequence of fixed ***f* numbers**, i.e. f2.8, f4, f5.6, f8, f11 that progressively halve the total amount of light.

Most lenses suffer from defects and distortions. A charged coupled device, or **CCD**, is a semiconductor sensor that is used in imaging applications since it does not suffer from **geometric distortion** and has linear response to the light.

A CCD camera consists of a two dimensional array of discrete silicon elements or photosites. The area of the array is typically 1 cm^2 . When light falls on a CCD, each photosite accumulates an amount of electronic charge proportional to the illumination time and intensity (up to a limit of about one million charge carriers). In order to retrieve image data, the accumulated charge must be shifted into an output register, and then to an amplifier. The photosites are connected in sequence to the register at a transfer clock. The transfer process is efficient and there is no loss of charge.



1.3 Human Vision and Computer Vision

Human vision system components

eye (camera) and

brain (processor), connected by optic nerves and works as follows:

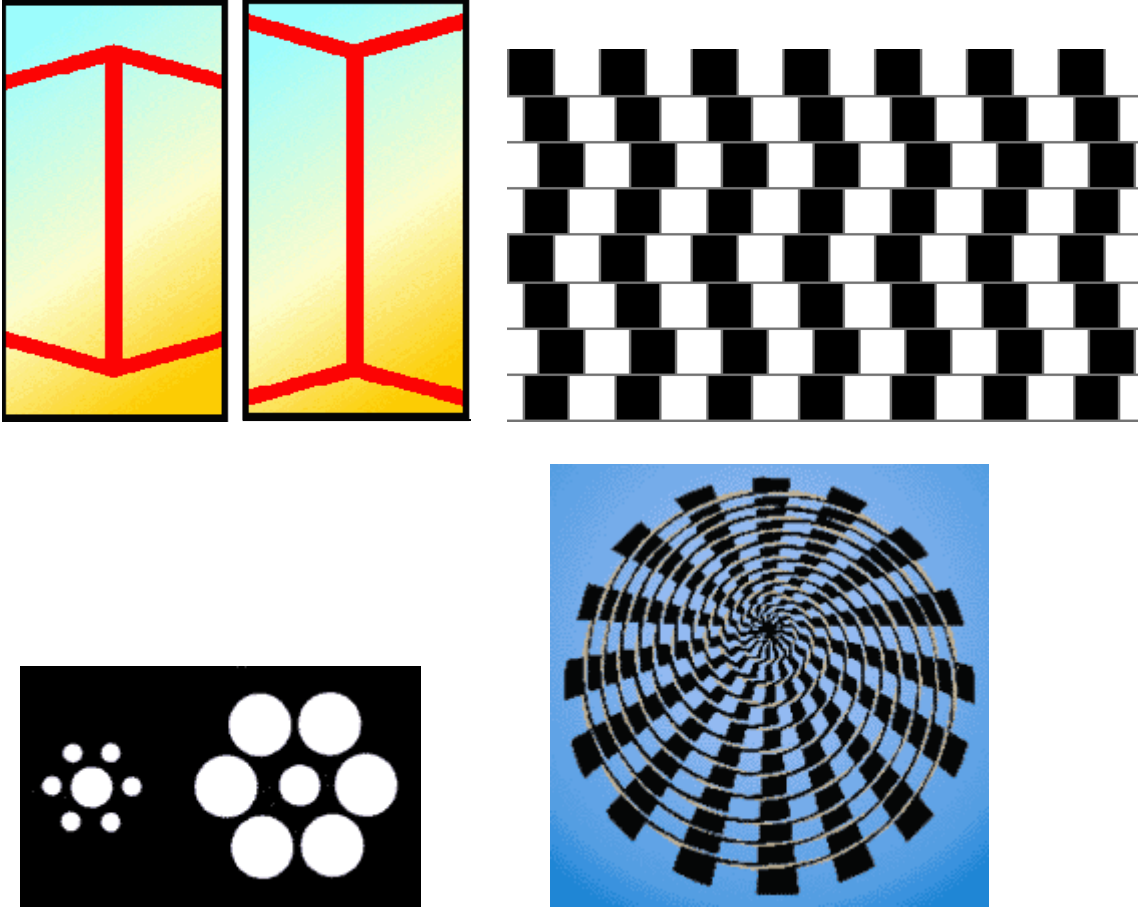
1. Light is focused by the eye's lens onto sensors located on retina
2. These sensors respond and send electrical signal over the optic nerves to the brain
3. The brain uses the signals to create neurological patterns (images).

Human vision is very complex, and is integrated and interdependent on various sensors. The degree of coupling is adaptive – it depends on the brain and signals from other body organs in addition to the eye.

Factors affecting human visual performance are fatigue, previous training, knowledge, etc. most of which do not exist in computer vision.

Important parameters in comparing human and computer vision are:

- **Adaptability:** This is the ability to modify operations according to environment parameters. Human vision is highly adaptable due to the brain in the loop. It can modify the system characteristics in an on-line fashion, e.g. can interpret an image even if it is partially obstructed. Computer vision is non-adaptable, but is repeatable. It can perform repetitive and identical tasks with a high degree of certainty.
- **Judgment and decision making:** These can be (a) value judgment based on heuristics and pseudo quantified factors, or (b) factual judgment based on quantified measurable parameters. Human vision can be effective in (a) for inspection tasks involving feature, i.e. shape. However, psychological and neurological inputs can lead to misinterpretation, e.g. a line segment may appear long or curved. Computer vision is effective in factual judgment, and is more consistent than human vision.



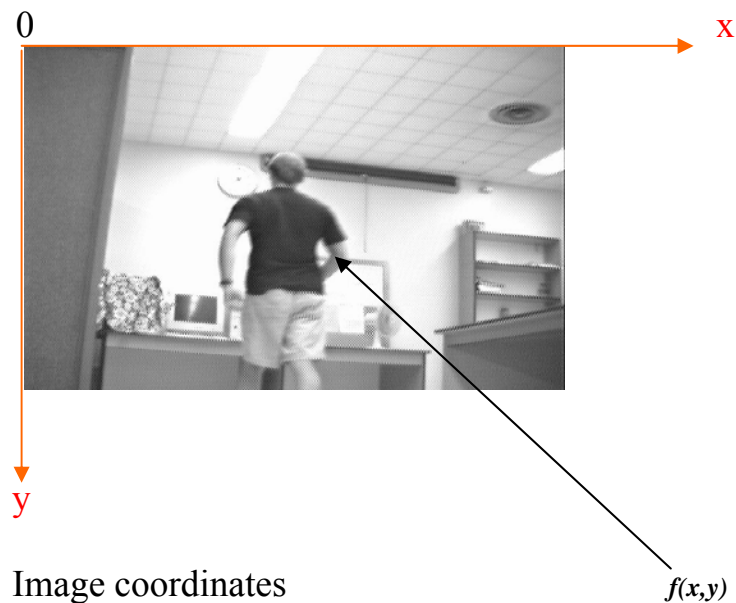
- **Quality of measurements:** Consistency of results and level of precision of computer vision is superior. Level of precision is only limited by the number of bits available to encode the integer representing the light intensity. It has no errors due to fatigue or distraction.
- **Image acquisition speed:** In computer vision image acquisition time depends on (a) the size of the image array, (b) processing time of the frame grabber, (c) type of camera. Electronic (solid state) cameras can acquire an image in few microseconds. Human time acquisition is about 0.06 second – TV frame changes of 30 image per second cannot be detected by human.
- **Spectrum response:** Human vision can only see a limited range of wavelength (see Fig. 1.1). Computer vision can create a new image based on combination of information from different parts of the

spectrum, e.g. IR plus visible range can be combined to identify an object emitting heat.

- **2D and 3D Capabilities:** Human vision has 3D capabilities superior to computer vision. Two cameras can be used in computer vision (called stereo vision) to provide distance information. However, this process is not straightforward.

1.4 Image Representation

An image can be represented mathematically by $f(x, y)$ where x and y denote the coordinates, $f(x, y)$ denotes the intensity (brightness) at the point (x, y) . In the case of color images, $f(x, y)$ is a multi-valued function represented by a vector.



The function $f(x, y)$ must be translated into a discrete array of numerical data for computer processing. The resulting digital representation is only an approximation of the original image. The **translation of $f(x, y)$ into digital form** is done by the processes of **sampling and quantization** both of which are carried by an electronic circuit called analog to digital converter (ADC).

Sampling: The process of measuring (or obtaining) the values of the image function $f(x, y)$ at discrete intervals in space is called sampling. Each sample is for a small square area of the image known as pixel (pixel element), where

pixels are indexed by integer values x and y . A CCD camera sensor has a discrete array of sensors, and so it is already sampled. However, in video cameras these samples are converted into an analog signal for compatibility with the majority of video equipment.

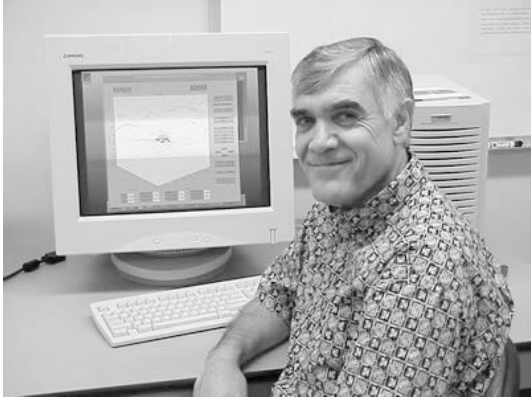
A single frame of a video signal is already discrete in y dimension and consists of an integer number of lines of data. Video standards have specified sample rates, e.g. RS-170 has 485 lines of data and each frame must have an aspect ratio of 4/3 and therefore there are $485 * (4/3) = 646$ samples per line. In practice few lines are trimmed to give an array of 640 by 480 pixels. Digital still picture camera is already in discrete form and there is no need to sample in x direction. The CCD cameras typically produce images of **1024 by 768**, **1280 by 1024**, etc. These are compatible with computer display standards, e.g. SVGA.

The **spatial resolution** of an image is the size (say in mm) of a pixel in the image, i.e. the area in the scene that is represented by a single pixel in the image. In medicine, the sampling is generally coarse to reduce the amount of gamma ray emitted into the patient's body. For example, chest X-ray of a patient may be represented by a 64 x 64 pixel array.

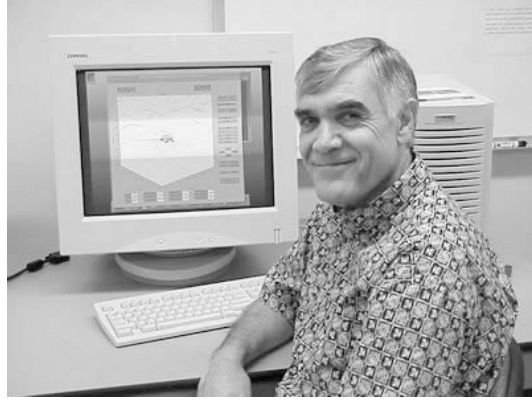
Quantization: As discussed above sampling involves digitizing the arguments x and y in the image function $f(x, y)$. Quantization is the process of digitizing the value of the function. The accuracy of representing the brightness is determined by the number of quantization levels. In black and white images, the levels correspond to shades of gray which are referred to as gray levels. For efficient processing gray levels g are a power of two, i.e.

$$g = 2^b \tag{1.4}$$

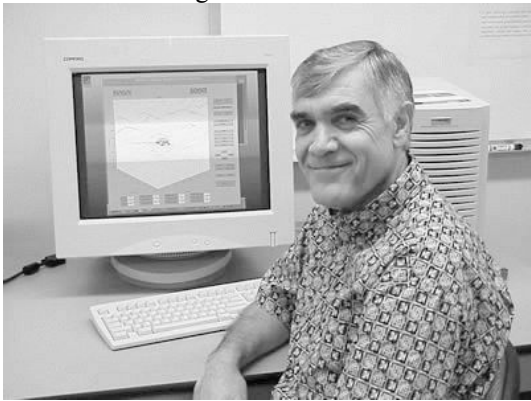
where b is the number of bits used for quantization, and is typically 8, giving images with 256 gray levels ranging from 0 (black) to 255 (white). The number of 8-bit bytes to store an image of height h and width w pixels is $2^{b-3} wh$. For example a 512 by 512 image with 256 gray levels will require 262,144 bytes of storage.



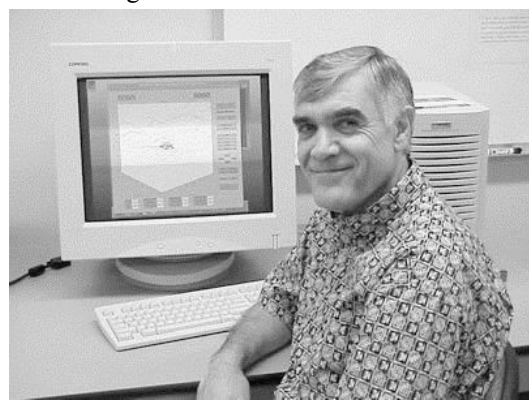
$g = 256$



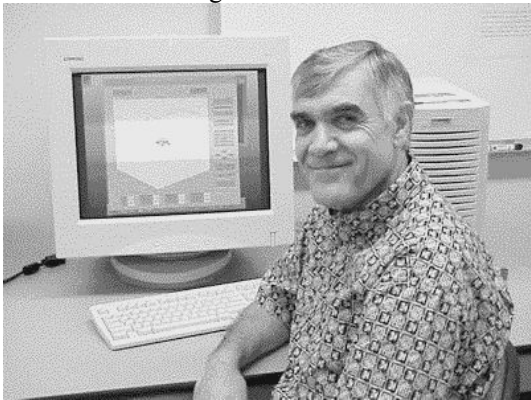
$g = 128$



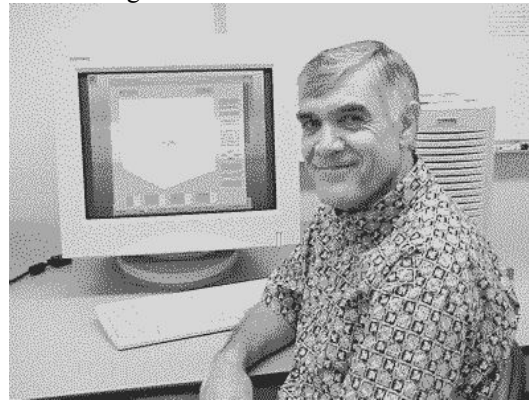
$g = 64$



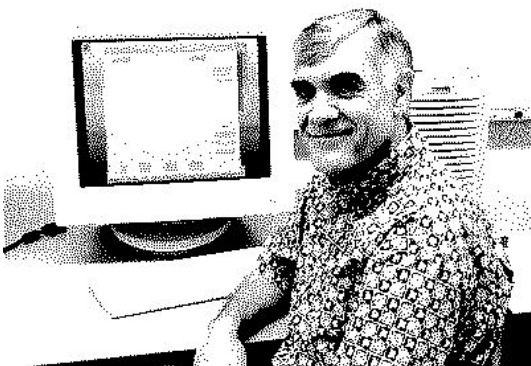
$g = 16$



$g = 8$



$g = 4$



$g = 2$

1.5 Color Images

Color images are modeled as three band, usually red-blue-green or RGB, of monochrome image data. When the image is displayed, the brightness information is shown on the screen by picture elements that emit light energy corresponding to the particular color. If each band is 8-bit (or 256 shades), then there will be $256^3 = 16,777,216$ colors.

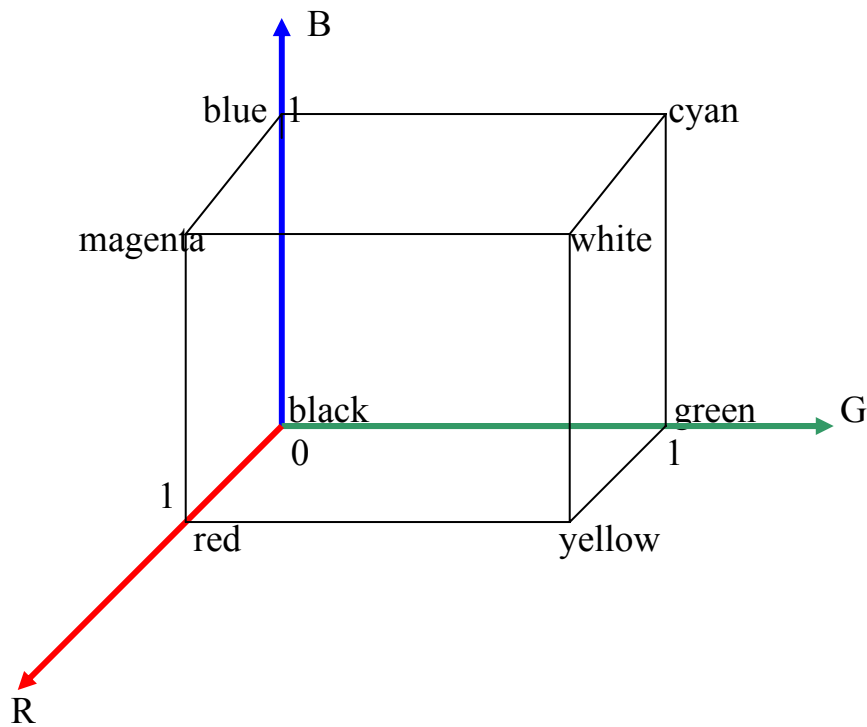


Fig. 1.5 Color coordinates



RGB



RGB →BGR

The cyan, magenta and yellow, or **CMY** color model is sometimes used as the primary colors, and is derived from the normalized RGB colors as $C = 1 - R$, $M = 1 - G$ and $Y = 1 - B$. By mixing C, M and Y, it is theoretically possible to produce any color. However in practice, black is not satisfactorily produced, and a fourth color denoted by K and representing black is added which results in the **CMYK** model. This is used for printing of hard copies on the color printers.

The RGB color information can be transformed into another 3 dimensional color model called hue-saturation-intensity, or **HSI**. In this model one dimension is used for brightness and the other two dimensions represent relative amount of different colors. A benefit of this transformation is that it creates a more human way of describing colors.

Hue (H) is color (e.g. orange, brown, etc),
saturation (S) is a measure of how much white the color is, and
intensity (I) is the brightness of the color.

For example, a “deep bright orange” would have a large intensity (or brightness), a hue of orange and a high value of saturation (that is “deep”). A rather complicated transformation exists for converting RGB into HSI.



Correct colors



Low Hue



High Hue



Correct colors



Low lightness



High lightness



Low Saturation



High Saturation

There are also several internationally recognized standards for colors. For computer imaging, the International Telecommunication Union-Radio (ITU-R, previously CCIR), has specified standards for digital video. This is based on one **luminance signal (Y)**, and **two color difference signals U,V** (sometimes called Cr, Cb). The transformation for a 24 bits/pixel RGB is

$$\begin{aligned} Y &= 0.3R + 0.587 G + 0.114B \\ U &= 0.5R - 0.419G - 0.0813B + 128 \\ V &= -0.169R - 0.331G + 0.5B + 128 \end{aligned} \tag{6}$$

Note that RGB in the above equations are not normalized, and range from 0 to 255. The human eye is less sensitive to color variations than to intensity variations, so YUV allows the encoding of luminance (Y) information at full bandwidth and chrominance (UV) information at half bandwidth.

There are also several internationally recognized standards for colors. For computer imaging, the International Telecommunication Union-Radio (ITU-R, previously CCIR), has specified standards for digital video. This is based on one **luminance signal (Y)**, and **two color difference signals U,V** (sometimes called Cr, Cb). The transformation for a 24 bits/pixel RGB is

$$\begin{aligned} Y &= 0.3R + 0.587 G + 0.114B \\ U &= 0.5R - 0.419G - 0.0813B + 128 \\ V &= -0.169R - 0.331G + 0.5B + 128 \end{aligned} \tag{6}$$

Note that RGB in the above equations are not normalized, and range from 0 to 255. The human eye is less sensitive to color variations than to intensity variations, so YUV allows the encoding of luminance (Y) information at full bandwidth and chrominance (UV) information at half bandwidth.

Appendix to Chapter 1
The story of YUV (YCbCr)