2. IMAGE FORMATION AND IMAGE MODELS

Light - a carrier of information

In this unit we shall consider the most common type of digital images, where pixel values correspond to luminance.

Light is the carrier of information about environment. Light emanating from a given source (e.g. Sun) interacts with objects and other components of the environment. The following types of interactions occur:

- Absorption
- Diffraction
- Refraction
- Reflection

These interactions alter the spectral and spatial characteristics of light. As regards objects, their surface material (colour, fabric) determines which portion of the spectrum is reflected; surface texture determines how coherently light is reflected.

Thus light reflected from surfaces in the direction of a sensor (natural or human-made) carries with it information about environment which it acquired during the interactions. Gibson (see [Bruce et al, Ch.11]) called a portion of all the light rays coming from the environment and hitting the receptor array the *ambient optic array*. Any movement, of self or within the environment, changes the optic array. All this static an dynamic information carried by light needs to be interpreted in order to make sense of the environment.

Light sensing

Human retina

Light sensing takes place on the retina (see Unit 1). When the sufficient number of light photons hit a receptor (rod or cone), a chemical reaction is effected (*bleaching*), which in turn generates electric impulses. These are then carried on along neurons to various parts of the brain. The sensor response (measured by the *firing rate* of a neuron) is proportional to the amount of light received. Sometimes the responses from several light sensors are polled together - these receptors are then said to define a *receptive field* on the retina. The size of receptive fields increases towards the periphery.

Electronic receptors

Electronic receptors (normally *light sensitive diodes*, LSD) work in a similar way to rods and cones - they generate electric current in proportion to the amount of light. Analogue-to-Digital Converters (ADCs) convert the voltage magnitude to a number (hence Digital), which is then stored in a computer memory, for further processing. LSDs are normally arranged in a regular rectangular grid (array). The resulting digital image is also rectangular.

Significant parameters

The most important factors in image formation are:

- the relationship between the amount of light and the sensor response (a *response function*)
- the relationship between the sensor response and the digital value (*brightness sampling*)
- receptive field sizes (*spatial resolution*)
- geometry of sensors (*spatial sampling*)

These parameters will be further considered here for digital images (see [Bruce et al Ch.1] for a discussion of these concepts in biological systems). Prior to this, a notion of an *image as a function* will be introduced.

Image as a function

This is the fundamental abstraction of an image: its mathematical representation. A digital image is represented by a *discrete* function, i.e. it comes in “chunks” rather than as a continuum. Different functions can represent the same image, depending on what characteristics are important.

A two dimensional monochrome digital images can be expressed by a function (let’s call it I for Image) of two spatial variables x,y (coordinates); the value of the function is a number (b)which corresponds to image brightness at a spatial coordinate (x,y):

$$I(x,y) = b$$

In colour images, the “value” at location (x,y) is a triplet (a *vector*) of values ($\bar{c}$), normally the red, the green and the blue component of the colour, i.e.

$$I(x,y) = \bar{c} = [\ r\ \ g\ \ b\ ].$$

Sometimes there will be more than three components (or *spectral bands*), for example in remote sensing near- and far-infrared bands can be used.

If a sequence of 2-dimensional *temporal images* (i.e. images taken over a period of time) needs to be considered, an image sequence will be a function of three variables:

$$I(x,y, t) = b$$

For three dimensional monochrome (i) or colour (ii) images, the functions will depend also on the depth (z) variable:

$$\text{(i) } I(x,y,z) = b \qquad\qquad \text{(ii) } I(x,y,z) = \bar{c}$$

This representation, of an image as a function, makes it easy to define various operations on images in a functional form. For example the expression:

$$I(x,y) + 20, \text{ for all } (x,y)$$

is a convenient way of specifying that value 20 should be added to all image pixels. The expression:

$$I'(x,y) = I(x,y) + 20, \text{ for all } (x,y)$$

means that after the addition, image values will be stored in a new image I’ at the location (x,y).

With this notation introduced, we return to digital image formation, which can be described by a *digitising model*.

Digitising model

A “real” image (i.e. a pattern of light falling onto receptors) is a *continuous* function. Image capture (by natural or man-made system) corresponds to taking by measurements (*samples*), at certain intervals (regularly or irregularly spaced). As the image is *spatially sampled* in this way, the brightness at each sample is *quantified* into a number of different levels (e.g. *grey levels*).

Thus digitisation involves two important factors, which will be discussed below:
- spatial sampling

- quantisation of grey levels

Sampling interval (spatial resolution)

The first question is: for a given height and width of a real 2D image (which, normally, is a projection of a 3D scene), how many pixels should be used to adequately represent the scene?

The more pixels there are, the higher resolution, hence the closer to the original. Degradation increases with decreasing spatial resolution. However, the amount of memory required for storage increases with increase in spatial resolution.

Probably the best answer to this question is given by the Sampling Theorem.

Sampling Theorem (Nyquist Criterion)

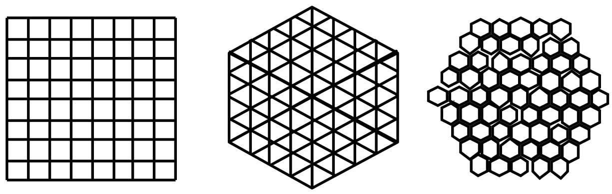
To fully represent detail (the rate of brightness change) in an original image, it has to be sampled at a rate at least twice as high as the highest spatial frequency of the detail.

It is wasteful to sample at a rate faster than twice of the finest detail to be resolved in the image, i.e. if the image size is 10 cm by 10 cm and the smallest detail is 1 cm, each pixel should have size of 0.5 cm, resulting in digital image of 20x20 pixels.

The effects of different spatial resolution are shown in the figures following this unit.

Tessellation

Image acquisition process can be thought of as approximating image samples over a finite cells partitioning the image. The cells are called pixels. The pattern is called *tessellation*. The most common tessellations are rectangular, triangular and hexagonal:



Grey level quantisation (brightness resolution)

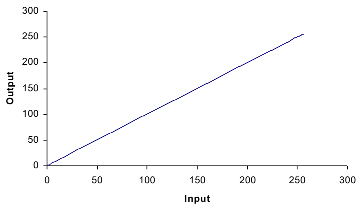
The next question is: how many grey levels to chose to represent accurately the image brightness?

This question is much harder to answer, because it depends on many factors: are images intended for further computer processing or for viewing by a human observer? Are they intended for accurate representation of the scene (in which case good *blend* is essential); or for detection (in which case good local contrast is important and fewer grey level values could be beneficial). Typically viewing will require at least 64 grey levels, whereas for simple image interpretation tasks, such as separating objects from background, 2 grey levels will be just fine.

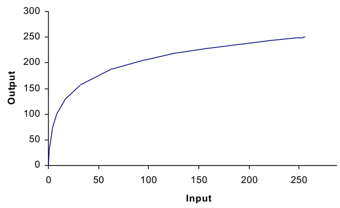
In terms of computer memory, the grey level resolution is determined by the *number of bits per pixel* (bpp), for example with 8 bpp, the grey level resolution is 2^8 = 256 levels (typical for digital images). Increment by one bit doubles brightness resolution.

The effects of different grey level resolution are shown in the figures following this unit.

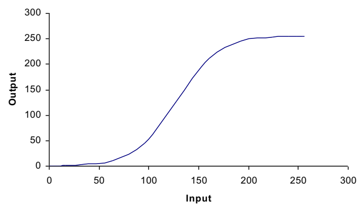
A very important factor affecting the choice of brightness resolution is the nature of the brightness *response function (or brightness transfer function)*, which defines how input brightnesses are transformed to image values. Graphs of several typical transfer functions are shown below. The choice depends on the dynamic range (i.e. the minimum and maximum brightness) of the scene luminances, the dynamic range of the sensors and the number of bits per pixel available for the image.



Linear



Logarithmic



Gamma

Other mathematical models of the image

Image analysis uses many approaches, and depending on the approach, different mathematical models are used to describe and to analyse the image. The following models are most common, most of them will be considered in further lectures.

Statistical model :	describes the image in terms of statistics of its grey levels
Geometric model :	describes how three dimensions are projected into two.
Radiometric model :	shows how the imaging geometry, light sources and reflectance properties of objects affect the light measurement at the sensor.
Spatial frequency model:	describes how spatial variations of the image may be characterised in a transform domain.
Colour model :	describes how different image values are related to image colours.

Practical aspects

The students should look further at:

- Methods of image acquisition (imaging devices such as image digitisers, monochrome and colour digital cameras, range sensors, colour film), see for example [Ballard & Brown, 2.3]

- Image formats (e.g. GIF, TIFF, JPEG, PPM), see for example [Umbaugh, 1.8]

Further reading and exploration

Sonka, M. et al (1993), Chapter 2.
Gonzalez, R.C. & Woods, R.E. (1992), Chapter 2.

HIPR

Worksheets -> Image Arithmetic.

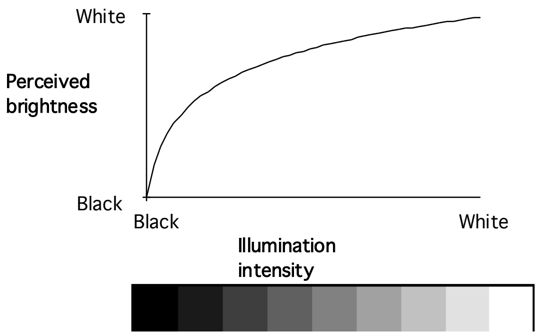
CVIP

Compression -> Preprocessing -> Spatial quantization
Compression -> Preprocessing -> Gray level quantization
Utilities -> Subtract

Things to think about

- How does the change in size of the receptive fields affects foveal and peripheral perception, and especially sensitivity and acquity?
- Compare and contrast spatial resolution in digital and biological images.
- Compare and contrast grey level resolution in digital and biological images.
- Think through this biological solution to the following dynamic range problem. The brighter the light falling onto retinal receptor cells, the stronger is their response. The response is measured by the "firing rate" - the number of electric impulses per second fired by the cell. The maximum firing rate is around 1000 / sec. It is known from experiments that humans can perceive light luminances in the range from 10^{-4} to 10^{4} lux (9 orders of magnitude). If we assume that the lowest light level of 10^{-4} lux generates one impuls, then, if the cell response were linear, it would need to fire 10^{8} times per seconds, which is inconsistent with the experimental facts.

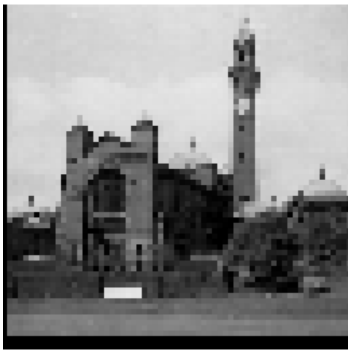
  In fact, the response is non-linear; it is logarithmic and its magnitude depends on the ratios of luminances rather on their differences. This relationship can be illustrated by the graph below. Explain the nature of this relationship mathematically.



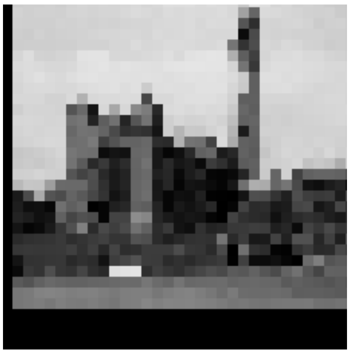
You will find that a number of terms in this Unit (and throughout) are not explicitly defined. This has been done deliberately, to encourage you to study the recommended topics in books and other media. Familiarity with these terms will be expected for assessed work, including an examination.

The effects of different spatial resolution. All images have 256 grey levels.

256 x 256 pixels



128 x 128 pixels



32 x 32 pixels

The effects of different grey level resolution. All images are 256 x 256 pixels.

64 grey levels

16 grey levels

4 grey levels