

# UNIT 1

## Fundamentals of Image Processing

(8 Hours) — Image Processing (CACS404)

### ⚠ Note: This may contain errors — please verify

These notes were generated with the help of AI and **may contain errors** (wrong formulas, numerical slips, missing details or syllabus mismatches). Please cross-check every important definition, equation and worked example with your prescribed textbook, class notes and lecturer before relying on them for exams or practicals.

## 1.1 Image Representation

An **image** is a two-dimensional function  $f(x, y)$ , where  $x$  and  $y$  are spatial (plane) coordinates and the amplitude of  $f$  at any point  $(x, y)$  is called the **intensity** or **gray level** of the image at that point. When  $x$ ,  $y$  and the amplitude values are all **finite and discrete**, the image is a **digital image**. Digital Image Processing (DIP) is the processing of digital images by means of a digital computer.

A digital image is stored as an  $M \times N$  matrix of real numbers. Each element is a **pixel** (picture element, pel).

| Image type           | Pixel values                      | Description                                                              |
|----------------------|-----------------------------------|--------------------------------------------------------------------------|
| Binary (1-bit)       | 0 or 1                            | Black and white only; used for masks, text, segmentation output          |
| Gray-scale (8-bit)   | 0 – 255                           | 0 = black, 255 = white, in-between = shades of gray                      |
| Colour (RGB, 24-bit) | 3 values (R, G, B), each 0 – 255  | Three matrices (planes) stacked together; $\approx 16.7$ million colours |
| Indexed              | Index into a colour map (palette) | Compact; used in GIF and 8-bit PNG                                       |

**Number of gray levels:**  $L = 2^k$ , where  $k$  is the number of bits per pixel. **Storage required** (in bits) is  $b = M \times N \times k$ .

### Example 1.1 — Storage requirement

How many bytes are needed to store a  $1024 \times 768$  image with 8 bits per pixel? What about a 24-bit colour image of the same size?

### Solution

Gray-scale:  $b = 1024 \times 768 \times 8 = 6,291,456$  bits = 786,432 bytes = **768 KB**.

Colour (24-bit):  $1024 \times 768 \times 24$  bits =  $786,432 \times 3$  bytes = 2,359,296 bytes = **2.25 MB**.

## Spatial and Intensity Resolution

- **Spatial resolution** — the smallest discernible detail; measured in pixels per unit distance (dpi, pixels/cm). More pixels  $\rightarrow$  finer detail.
- **Intensity (gray-level) resolution** — the smallest discernible change in intensity; depends on the number of bits  $k$ . Too few levels produce **false contouring**.

## 1.2 Basic Relationships Between Pixels

### Neighbours of a Pixel

For a pixel  $p$  at coordinates  $(x, y)$ :

- **4-neighbours  $N_4(p)$** :  $(x+1, y)$ ,  $(x-1, y)$ ,  $(x, y+1)$ ,  $(x, y-1)$ .
- **Diagonal neighbours  $N_D(p)$** :  $(x+1, y+1)$ ,  $(x+1, y-1)$ ,  $(x-1, y+1)$ ,  $(x-1, y-1)$ .
- **8-neighbours  $N_8(p)$** :  $N_4(p) \cup N_D(p)$ .

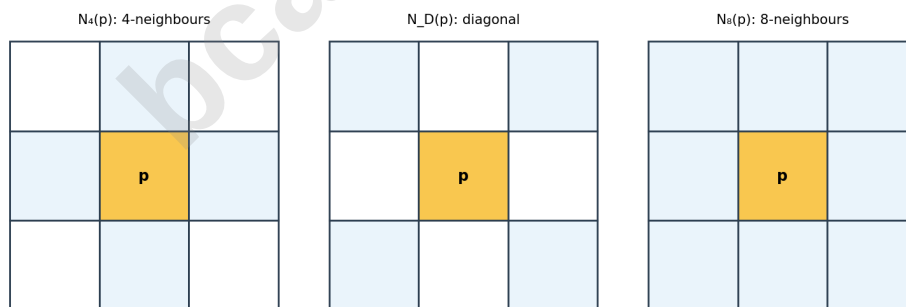


Figure 1.1 — 4-neighbours, diagonal neighbours and 8-neighbours of pixel  $p$

## Adjacency, Connectivity, Regions and Boundaries

Let  $V$  be the set of gray-level values used to define adjacency (for a binary image,  $V = \{1\}$ ).

- **4-adjacency**:  $p$  and  $q$  (both with values in  $V$ ) are 4-adjacent if  $q \in N_4(p)$ .
- **8-adjacency**:  $p$  and  $q$  are 8-adjacent if  $q \in N_8(p)$ .
- **m-adjacency (mixed)**:  $p$  and  $q$  are m-adjacent if (i)  $q \in N_4(p)$ , or (ii)  $q \in N_D(p)$  and the set  $N_4(p) \cap N_4(q)$  has no pixels with values in  $V$ . m-adjacency removes the ambiguity

(multiple paths) that arises with 8-adjacency.

- **Path:** a sequence of distinct pixels  $p_1, p_2, \dots, p_n$  where consecutive pixels are adjacent.
- **Connected component:** set of pixels in a subset  $S$  connected to a given pixel by a path lying completely in  $S$ .
- **Region:** a connected subset of an image. **Boundary (border):** pixels of a region having at least one neighbour outside the region.

## Distance Measures

For pixels  $p(x, y)$  and  $q(s, t)$ , a valid distance function  $D$  satisfies:  $D(p, q) \geq 0$ ;  $D(p, q) = D(q, p)$ ; and  $D(p, z) \leq D(p, q) + D(q, z)$ .

| Distance         | Formula                          | Shape of equidistant points |
|------------------|----------------------------------|-----------------------------|
| Euclidean $D_e$  | $\sqrt{[(x - s)^2 + (y - t)^2]}$ | Circle                      |
| City-block $D_4$ | $ x - s  +  y - t $              | Diamond                     |
| Chessboard $D_8$ | $\max( x - s ,  y - t )$         | Square                      |

### Example 1.2 — Distances

Find  $D_e$ ,  $D_4$  and  $D_8$  between  $p(2, 3)$  and  $q(5, 7)$ .

#### Solution

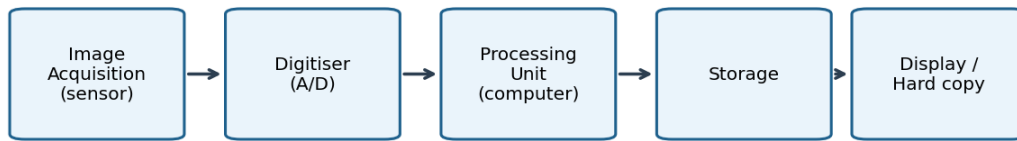
$$D_e = \sqrt{(3^2 + 4^2)} = \sqrt{25} = 5$$

$$D_4 = |2-5| + |3-7| = 3 + 4 = 7$$

$$D_8 = \max(3, 4) = 4$$

## 1.3 Elements of a DIP System

### Elements of a Digital Image Processing System



*Scene → sensor → digital image → algorithms → memory / network → output device*

*Figure 1.2 — Block diagram of a digital image processing system*

- **Image acquisition:** a physical sensor (CCD/CMOS camera, scanner, X-ray, satellite sensor) senses energy from the scene, and a digitiser converts the analogue signal to digital form.
- **Processing hardware/software:** a computer (PC to supercomputer, GPU) running image processing algorithms.
- **Mass storage:** short-term (RAM, frame buffers), on-line (disks) and archival (tapes, cloud). Image files are large, so storage is critical.
- **Display:** monitors, printers, projectors, film recorders.
- **Networking:** for transmission of images; bandwidth is the main limitation, which motivates compression (Unit 4).

## Fundamental Steps in DIP

Image acquisition → preprocessing/enhancement → restoration → colour processing → wavelets/multiresolution → compression → morphological processing → segmentation → representation & description → object recognition. (Outputs of the first stages are images; later stages output attributes extracted from images.)

## 1.4 Elements of Visual Perception

### Structure of the Human Eye

- **Cornea and sclera** — transparent front cover and tough outer white coat.
- **Iris and pupil** — control the amount of light entering the eye ( $\approx 2$  mm to 8 mm).
- **Lens** — focuses the image on the retina by changing its shape.
- **Retina** — inner light-sensitive layer containing the photoreceptors.

- **Cones** (6–7 million, concentrated in the **fovea**) — photopic (bright-light) and colour vision. Each cone has its own nerve ending, giving fine detail.
- **Rods** (75–150 million, spread over the retina) — scotopic (dim-light) vision, no colour, share nerve endings, give overall picture.
- **Blind spot** — where the optic nerve leaves the retina; no receptors.

## Brightness Adaptation, Discrimination and Weber Ratio

The eye can adapt to a huge range of intensities (about  $10^{10}$ ), but at any given adaptation level it discriminates only a much smaller range. **Weber ratio** =  $\Delta I_c / I$ , where  $\Delta I_c$  is the increment of illumination discriminable 50% of the time against a background  $I$ . A *small* Weber ratio means good brightness discrimination.

Two perception phenomena show that perceived brightness is **not** a simple function of intensity: the **Mach band effect** and **simultaneous contrast** (a gray patch looks darker on a bright background than on a dark one).

## 1.5 Simple Image Formation Model

Images are formed by illumination reflected (or transmitted) from objects. The intensity at  $(x, y)$  is modelled as the product:

$$f(x, y) = i(x, y) \cdot r(x, y), \text{ with } 0 < i(x, y) < \infty \text{ and } 0 < r(x, y) < 1$$

- **$i(x, y)$**  — illumination: amount of source light incident on the scene (depends on the source).
- **$r(x, y)$**  — reflectance: fraction of light reflected by the object (0 = total absorption, 1 = total reflectance). Depends on the object.
- Typical reflectance values: black velvet  $\approx 0.01$ , stainless steel  $\approx 0.65$ , white wall paint  $\approx 0.80$ , silver plate  $\approx 0.90$ , snow  $\approx 0.93$ .

For a monochrome image, gray level  $\ell = f(x, y)$  lies in  $[L_{\min}, L_{\max}]$ ; the interval is shifted numerically to  $[0, L - 1]$ , with 0 = black and  $L - 1$  = white.

## 1.6 Sampling and Quantization

The output of most sensors is a continuous voltage waveform. To create a digital image we must convert this continuous data into digital form in two steps:

- **Sampling** — digitising the *coordinate* values ( $x$  and  $y$ ). Determines spatial resolution ( $M \times N$ ).

- **Quantization** — digitising the *amplitude* values into  $L = 2^k$  discrete levels. Determines intensity resolution.

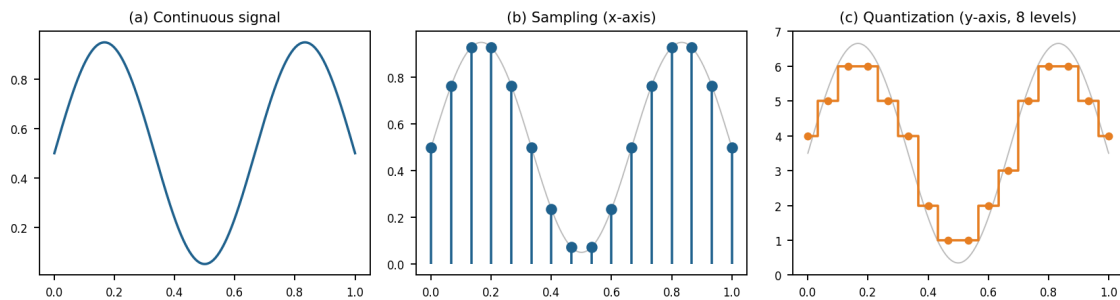


Figure 1.3 — Sampling and quantization of a 1-D scan line

**Sampling theorem (Nyquist):** a band-limited signal can be reconstructed exactly if the sampling frequency is at least twice the highest frequency in the signal ( $f_s \geq 2f_{\max}$ ). Under-sampling causes **aliasing** — high frequencies appear as false low frequencies (e.g., Moiré patterns).



Figure 1.4 — A smooth gradient quantized to 8, 4, 2 and 1 bits: fewer levels cause visible false contours

*Note: Reducing the number of samples gives a checkerboard effect; reducing the gray levels gives false contouring. Images with large detail need more samples; images with smooth regions can tolerate fewer gray levels (Isopreference curves, Huang 1965).*

## 1.7 Colour Fundamentals and Models

Colour is the perceptual result of light in the visible spectrum ( $\approx 400$  nm violet to  $\approx 700$  nm red) striking the retina. Three basic quantities describe a chromatic light source: **radiance** (total energy from the source, in watts), **luminance** (energy perceived by the observer, in lumens) and **brightness** (subjective, impossible to measure).

The cones are of three types, roughly 65% sensitive to red, 33% to green and 2% to blue (but blue cones are the most sensitive). Hence **red, green and blue are the primary colours of light**. Mixing two primaries gives the **secondary colours**: Yellow = R + G, Cyan = G + B, Magenta = R + B.

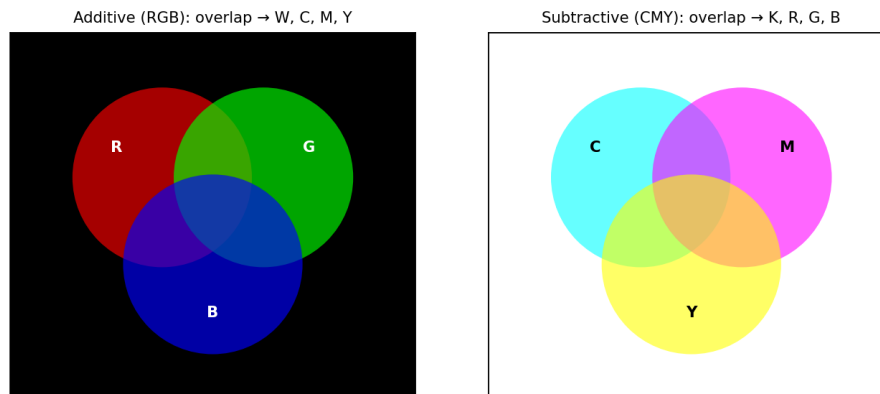


Figure 1.5 — Additive (RGB) and subtractive (CMY) colour mixing

### (a) RGB Model

Based on a Cartesian coordinate system; the colour sub-space is a cube with R, G, B on three axes. Black is at the origin (0,0,0), white at (1,1,1) opposite it, and gray values lie along the diagonal. Used in monitors, cameras and scanners. A pixel with 8 bits per channel has a depth of 24 bits (the **full-colour** image).

### (b) CMY and CMYK Models

Cyan, magenta and yellow are the secondary colours of light and the primaries of pigments; used in printers. Conversion (normalised values in [0,1]):  $C = 1 - R$ ,  $M = 1 - G$ ,  $Y = 1 - B$ . Equal C, M, Y should give black, but real inks give a muddy brown, so a **K** (black) ink is added: CMYK, which also saves coloured ink.

### (c) HSI Model

Describes colour in the way humans interpret it: **Hue** (dominant wavelength, the 'pure colour' as an angle), **Saturation** (purity — degree of dilution with white light) and **Intensity** (brightness). It decouples intensity from colour information, so gray-scale algorithms can be applied to the I channel only.

**RGB → HSI** (R, G, B normalised to [0, 1]):

- $I = (R + G + B) / 3$
- $S = 1 - 3 \cdot \min(R, G, B) / (R + G + B)$
- $H = \theta$  if  $B \leq G$ , else  $H = 360^\circ - \theta$ , where  $\theta = \cos^{-1} \{ \frac{1}{2}[(R - G) + (R - B)] / [(R - G)^2 + (R - B)(G - B)]^{1/2} \}$

### Example 1.3 — Colour conversion

Convert the normalised RGB pixel (0.8, 0.4, 0.2) into (a) CMY and (b) HSI.

#### Solution

(a)  $C = 1 - 0.8 = 0.2$ ,  $M = 1 - 0.4 = 0.6$ ,  $Y = 1 - 0.2 = 0.8 \rightarrow \mathbf{CMY = (0.2, 0.6, 0.8)}$

(b)  $I = (0.8 + 0.4 + 0.2)/3 = \mathbf{0.467}$ .  $S = 1 - 3(0.2)/1.4 = 1 - 0.429 = \mathbf{0.571}$ .

Numerator =  $\frac{1}{2}[(0.4) + (0.6)] = 0.5$ ; denominator =  $\sqrt{[0.16 + (0.6)(0.2)]} = \sqrt{0.28} = 0.529$ ; ratio = 0.945;  $\theta = \cos^{-1}(0.945) \approx 19.1^\circ$ . Since  $B \leq G$ ,  $\mathbf{H \approx 19.1^\circ}$  (orange).

### (d) Luminance–Chrominance Models (YIQ, YCbCr)

Separate luminance (Y) from colour difference signals.  $\mathbf{Y = 0.299R + 0.587G + 0.114B}$ . Used in TV broadcasting (YIQ — NTSC; YUV — PAL) and in JPEG/MPEG (YCbCr), where the chrominance can be down-sampled because the eye is less sensitive to colour detail.

| Model       | Components                       | Used in                             |
|-------------|----------------------------------|-------------------------------------|
| RGB         | Red, Green, Blue                 | Monitors, cameras, scanners         |
| CMY / CMYK  | Cyan, Magenta, Yellow (, Black)  | Colour printing                     |
| HSI / HSV   | Hue, Saturation, Intensity/Value | Colour image analysis, segmentation |
| YIQ / YCbCr | Luminance + two chrominance      | TV, JPEG, MPEG                      |

## 1.8 Image File Formats

| Format          | Compression              | Features / Use                                                    |
|-----------------|--------------------------|-------------------------------------------------------------------|
| BMP             | None (or RLE)            | Windows bitmap; simple, large files                               |
| GIF             | Lossless (LZW)           | Max 256 colours (palette), animation, transparency; web icons     |
| JPEG / JPG      | Lossy (DCT)              | Photographs; adjustable quality; no transparency                  |
| PNG             | Lossless (Deflate)       | Alpha transparency, 24/48-bit colour; replaces GIF                |
| TIFF            | Lossless / lossy options | Flexible, multi-page; publishing, scanning, medical               |
| PGM / PPM / PBM | None                     | Netpbm portable gray / pixel / bit maps; simple, used in teaching |
| RAW             | None / lossless          | Unprocessed sensor data from cameras                              |
| DICOM           | Varies                   | Medical imaging with patient metadata                             |

Every file has a **header** (width, height, bit depth, compression type, palette) followed by the pixel data. Lossless formats retain the exact pixel values; lossy formats discard information to reduce size.

## 1.9 Image Operations

### Classification by Pixels Involved

- **Point operations:** output at  $(x, y)$  depends only on the input at  $(x, y)$ :  $g(x, y) = T[f(x, y)]$  (negative, thresholding, brightness change).
- **Local (neighbourhood) operations:** output depends on a neighbourhood of  $(x, y)$  (smoothing, sharpening, median filter).
- **Global operations:** output depends on the whole image (Fourier transform, histogram equalization).

### Types of Operations

- **Arithmetic:** addition (noise averaging, brightening), subtraction (change detection, background removal), multiplication (masking, shading correction), division (ratio images).
- **Logical:** AND, OR, NOT, XOR on binary images — masking, region-of-interest extraction.
- **Geometric:** translation, scaling, rotation, shearing and flipping. Needs **interpolation** (nearest-neighbour, bilinear, bicubic) for the new pixel grid.
- **Set operations:** union, intersection, complement, difference on regions (basis for morphology).

## 1.10 Brightness, Contrast, Hue, Saturation and the Mach Band Effect

- **Brightness** — overall lightness or darkness of an image; adding a constant to every pixel increases it:  $g = f + b$ .
- **Contrast** — difference between the brightest and darkest regions; scaling by a constant  $a > 1$  raises it:  $g = a \cdot f + b$ . Often defined as  $(I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$ .
- **Hue** — the attribute of a colour associated with the dominant wavelength (red, green, yellow ...).
- **Saturation** — purity of the colour; a pure spectral colour is 100 % saturated, a gray has 0 %.

## Mach Band Effect

Near the boundary between two regions of different but constant intensities, the visual system perceives an **undershoot** on the dark side and an **overshoot** on the bright side — bands that do not exist physically. They arise from **lateral inhibition** in the retina: each receptor's response is reduced by the activity of its neighbours, which exaggerates edges. Discovered by Ernst Mach (1865).

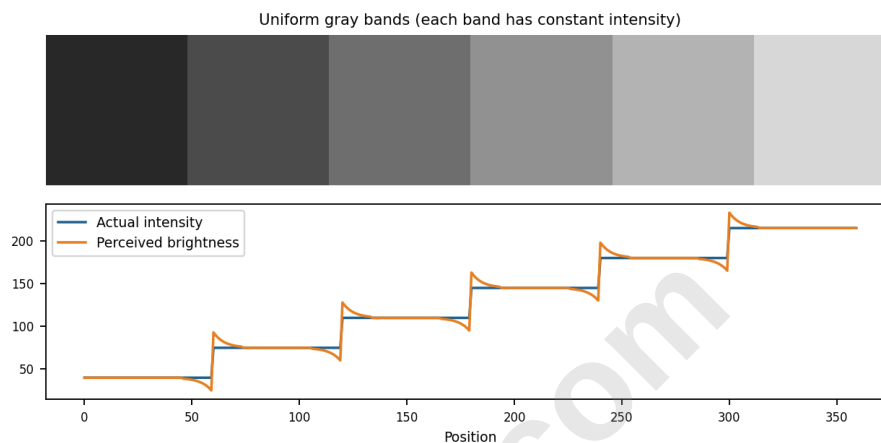


Figure 1.6 — Mach bands: perceived brightness (orange) overshoots and undershoots at each step edge

*Note: Mach bands show that the human eye acts like an edge-enhancing (high-pass) system, which is why edge sharpening makes images look 'crisper'.*

### Key Takeaways

- ✓ A digital image is an  $M \times N$  matrix;  $L = 2^k$  gray levels; storage =  $M \cdot N \cdot k$  bits.
- ✓ Neighbourhoods  $N_4$ ,  $N_D$ ,  $N_8$ ; adjacency 4/8/m; distances  $D_e$ ,  $D_4$  (city-block),  $D_8$  (chessboard).
- ✓ Image model:  $f(x, y) = i(x, y) \cdot r(x, y)$ . Sampling digitises coordinates, quantization digitises amplitude; Nyquist:  $f_s \geq 2f_{\max}$ .
- ✓ RGB (display), CMY/CMYK (print), HSI (human-oriented), YCbCr (compression).  $C = 1 - R$  etc.
- ✓ Mach band effect = edge exaggeration by lateral inhibition; Weber ratio measures brightness discrimination.
- ✓ Know the common file formats (BMP, GIF, JPEG, PNG, TIFF) and the point / local / global classification of operations.

*Reminder: these notes may contain errors — verify with your textbook and lecturer.*

— End of Unit 1 —