Digital Image Processing

CLASS NOTES

Mahanth Yalla
M. Tech-AI, IISc

Preface

These notes are based on the lectures delivered by **Prof. Rajiv Soundararajan (ECE) and Prof. Soma Biswas (EE)** in the course **E9 241 - Digital Image Processing** at Indian Institute of Science (IISc) Bengaluru - Aug Semester 2025. The notes are intended to be a concise summary of the lectures and are not meant to be a replacement for the lectures.

Disclaimer

These notes are not official and may contain errors. Please refer to the official course material for accurate information.

"I cannot guarantee the correctness of these notes. Please use them at your own risk".

- Mahanth Yalla

Contribution

If you find any errors or have any suggestions, please feel free to open an issue or a pull request on this GitHub repository, I will be happy to incorporate them.

Feedback

If you have any feedback or suggestions on the notes, please feel free to reach out to me via social media or through mail mahanthyalla [at] {iisc [dot] ac [dot] in , gmail [dot] com}.

Acknowledgements: I would like to thank Prof. Rajiv Soundararajan (ECE) and Prof. Soma Biswas (EE) for delivering the lectures and providing the course material. I would also like to thank the TAs for their help and support.

Contents

Chapter 1

Binary Image Processing

1.1 Introduction

Definition 1.1.1: Image Definition

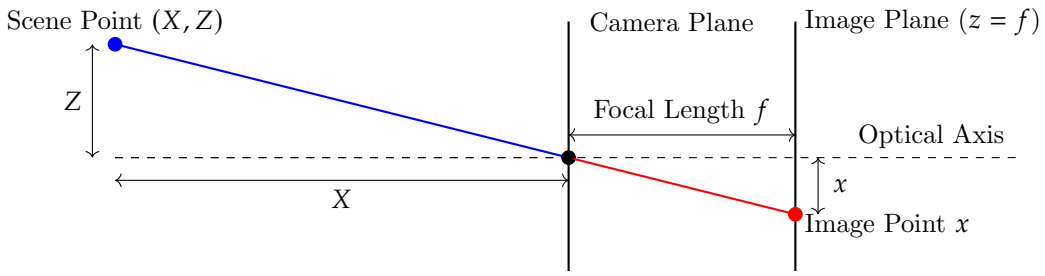
An **image** is a projection of a 3D object onto a 2D surface. This dimensional reduction causes loss of certain 3D information, which is generally very hard to recover.

1.2 Image Formation

The process of **image formation** models how a 3D scene is mapped to a 2D image plane through a camera model. A pinhole camera serves as the simplest abstraction to study this process.

1.2.1 Pinhole Camera Model (2D)

In the 2D case, the relation between a point (X, Z) in the scene and its projection (x) on the image plane is derived by similar triangles:



from the diagram, we have:

$$\frac{x}{f} = \frac{X}{Z} \quad \Rightarrow \quad x = f \cdot \frac{X}{Z}$$

Here, f is the focal length of the pinhole camera.

Note:-

we use Capital letters for 3D coordinates and lowercase for 2D image coordinates and observe that the Z component is lost.

1.2.2 Pinhole Camera Model (3D)

Extending to 3D coordinates (X, Y, Z), the projection onto the image plane gives:

$$x = f \cdot \frac{X}{Z}, \quad y = f \cdot \frac{Y}{Z}$$

This shows how 3D geometry is mapped to 2D via perspective projection.

1.2.3 Homogeneous Coordinates

Homogeneous coordinates are used to express projection as a matrix multiplication:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Note:-

Homogeneous coordinates are crucial because they unify perspective projection, translation, and rotation into matrix multiplications.

1.2.4 Intrinsic Matrix

The **intrinsic parameters** capture camera-specific properties such as focal length and principal point offset:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

where (f_x, f_y) are focal lengths in pixel units and (c_x, c_y) is the principal point.

1.2.5 Extrinsic Matrix

The **extrinsic parameters** describe the camera's position and orientation in the world:

$$M = [R|t]$$

where R is a 3×3 rotation matrix and t is a 3×1 translation vector.

Note:-

The rotation matrix R can be defined using an angle α in 2D as:

$$R(\alpha) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

This matrix rotates a point by α radians in the plane. In 3D, rotation can be performed about each axis using three matrices:

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}, \quad R_y(\phi) = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix}, \quad R_z(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A general 3D rotation can be represented by multiplying these matrices:

$$R = R_z(\psi) R_y(\phi) R_x(\theta)$$

where θ, ϕ, and ψ are rotation angles about the x, y, and z axes, respectively.

1.2.6 Projection Matrix

The complete mapping from 3D world coordinates to 2D image plane is:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = KM \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

with scaling factor s.

Claim 1.2.1 Projection Matrix

The projection matrix $P = K@M$ completely defines the mapping from world coordinates to image coordinates, which depends on the intrinsic and extrinsic parameters of the camera. (i.e., focal length, principal point, rotation, and translation vectors.)

Definition 1.2.1: Camera Calibration

The process of estimating the intrinsic and extrinsic parameters of a camera from observed images of a known calibration pattern.

Calibration ensures accurate geometric measurements from images.

1.3 Image Representation

1.3.1 Pixel Values

A digital image is represented as a 2D array of **pixels (picture elements)**, each encoding intensity (grayscale) or color.

Definition 1.3.1: Image & Pixel

An **image** is a 2D array

$$I : \{0, \ldots, M-1\} \times \{0, \ldots, N-1\} \to \{0, \ldots, K-1\}$$

Each element $I[i, j]$ is a **pixel** with **intensity** (gray level) in $\{0, \ldots, K-1\}$, where, 0 represents black and $K-1$ represents white. For a 8 bit storage, $B = 8$, then maximum intensity can be calculated as $K = 2^B = 2^8 = 256$, 0 represents black and 255 represents white.

For normalized intensity, use

$$I_n[i, j] = \frac{I[i, j]}{(K-1)} \in [0, 1]$$

.

1.3.2 Color Spaces

Different **color spaces** represent pixel values differently:

- RGB: additive primary colors.
- HSV: hue, saturation, value (closer to human perception).
- YCbCr: luminance and chrominance separation. (where Y is intensity)

1.3.3 Feature Extraction and Descriptors

Features capture essential information in images such as edges, corners, or textures. **Descriptors** encode features into numerical representations (e.g., SIFT, HOG) for matching and recognition.

1.3.4 Sampling and Quantization

The ideal irradiance signal is sampled on a discrete grid and quantized to a finite set of gray levels.

- **Sampling:** Selecting discrete spatial points to represent an image.
  choose integer pixel sites (i, j); the image becomes $I[i, j]$ on an $M \times N$ grid.
- **Quantization:** Mapping continuous intensity values into discrete levels
  map real intensities to $\{0, 1, \ldots, K-1\}$, e.g., $K = 256$ for 8-bit grayscale.

Binary images: special case $K = 2$ with intensities $\{0, 1\}$ (or $\{0, 255\}$ in 8-bit storage).

Note:-

Undersampling causes aliasing; inadequate quantization leads to loss of detail.

1.3.5 Image Formats

Typical resolutions include 256×256, 512×512, **1920x1080**, etc.As we can calculate the number of Bytes required to store these images, we find:

- For 256×256 images: $256 \times 256 \times 1 = 65,536$ Bytes (assuming 8-bit grayscale).
- For 512×512 images: $512 \times 512 \times 1 = 262,144$ Bytes (assuming 8-bit grayscale).
- For 1920×1080 images: $1920 \times 1080 \times 3 = 6,220,800$ Bytes (assuming 24-bit RGB).

which is nearly 6.3 MB for a full HD image (1920x1080 3-channel RGB image). Hence, image storage can be quite substantial, necessitating efficient compression techniques.

Few common **Formats** include:

- JPEG: lossy compressed format.
  The most common for photographs to save space, as the human eye is less sensitive to high-frequency details. The compression is achieved by discarding some image data, but the benefit is a significantly reduced file size. (e.g. the full HD image (1920x1080) can be compressed from **6.3 MB** to around less than a **1 MB**)
- PNG: lossless compression, supports transparency.
- BMP: uncompressed raster format.
- TIFF: flexible format supporting various compressions.
- GIF: supports animation and transparency (limited color palette).
- WEBP: modern format providing lossy and lossless compression.
- HEIF: high efficiency image format, supports advanced features.
- AVIF: image format based on AV1 compression, offering high quality at smaller file sizes.
- EXR: high dynamic range (HDR) image format, supports wide color gamuts and high bit depths.
- DNG: raw image format for digital photography, preserves original sensor data.
- PPM: portable pixmap format, simple uncompressed color image format.

1.3.6 Binary Images and Thresholding

Definition 1.3.2: Binary Image

A **binary image** is an image that consists of only two colors, typically black and white. Each pixel in a binary image is represented by a single bit, where 0 represents black and 1 represents white.

The simplest method to obtain binary images is **thresholding**:

$$B(x,y) = \begin{cases} 1 & I(x,y) \geqslant T \\ 0 & I(x,y) < T \end{cases}$$

where T is the threshold.

1.3.7 Gray Level Histograms

The histogram of grayscale values is a fundamental tool to analyze and design thresholding algorithms.

Claim 1.3.1 Histogram-based Thresholding

If the histogram shows two well-separated peaks, the optimal threshold lies near the valley between them.

1.4 Histogram of an Image

Definition 1.4.1: Gray-Level Histogram

A **gray-level histogram** is a representation of the distribution of pixel intensities in a grayscale image. It counts the number of pixels for each intensity level, providing insights into the image's contrast and brightness.

Mathematical Formulation Let a grayscale image be

$$I : \{0,\dots,M-1\} \times \{0,\dots,N-1\} \to \{0,\dots,K-1\}$$

. The *histogram* counts occurrences at each gray level, given as a function as

$$H : \{0,\dots,K-1\} \to \{0,\dots,MN\}$$

such that

$$H(k) = \text{no of occurrences of gray level } k$$

where $k \in \{0,\dots,K-1\}$

$$H(k) \;=\; \#\{(i,j) : I[i,j] = k\} \qquad k = 0,\dots,K-1$$

also,

$$\sum_{k=0}^{K-1} H(k) = MN.$$

The *normalized histogram* (PMF) is

$$p(k) = \tfrac{H(k)}{MN}$$

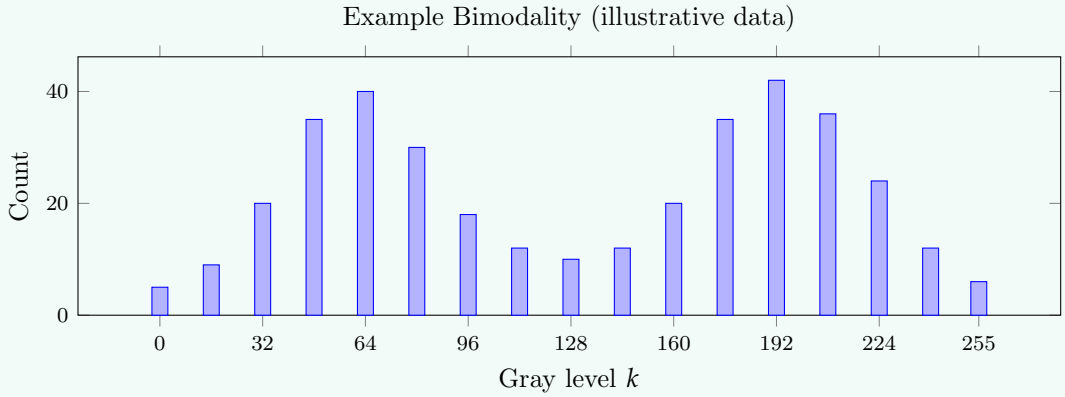
and

$$\sum_{k} p(k) = 1$$

Example 1.4.1 (Interpreting Histogram Shapes)

Given an image whose histogram exhibits a tall peak near intensity 40 (dark shades), and another smaller, broad peak near 200 (bright shades), we deduce the image features a dark background with a bright foreground—ideal for segmentation via thresholding.

- **Dark image:** $p(k)$ concentrated near $k \approx 0$.
- **Bright image:** $p(k)$ concentrated near $k \approx K-1$.
- **Bimodal image:** two peaks (e.g., dark foreground on bright background) $\Rightarrow$ suitable for a *single* global threshold.



Note:-

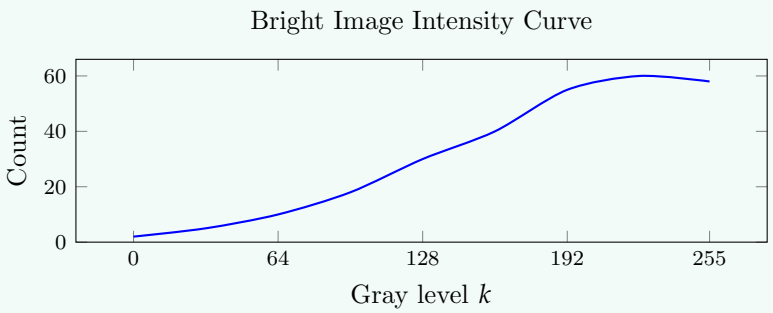
Understanding the histogram profile assists in identifying whether an image is underexposed, overexposed, well-contrasted, or subject to other illumination artifacts.

Types of Histograms and Their Interpretation Different histogram profiles relate directly to the visual impression and underlying properties of an image:

- **Bright Images:** Most pixel values clustered towards the higher end of the gray level range.
- **Dark Images:** Values crowded in the lower end, leading to overall darker visual output.
- **Dual Peak Model:** Exhibits two pronounced peaks, often corresponding to distinct foreground and background regions; common in images fit for binarization.
- **Flat Histogram:** Pixels uniformly distributed across gray levels—rare in natural images, may occur in images heavily corrupted with noise.
- **Equal Histograms:** Result from histogram equalization operations to improve contrast.

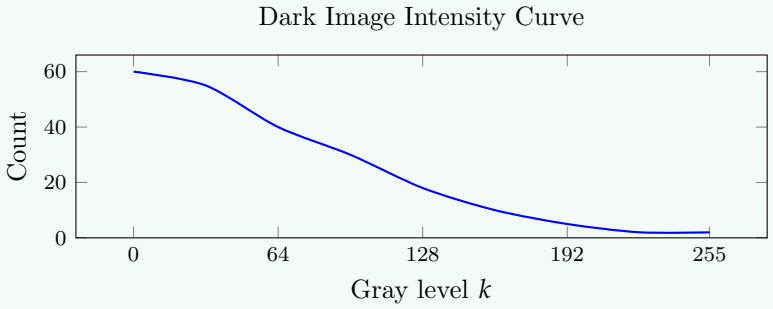
Example 1.4.2 (Histogram Interpretation)

- A **bright image** shows histogram concentrated on high intensity values.



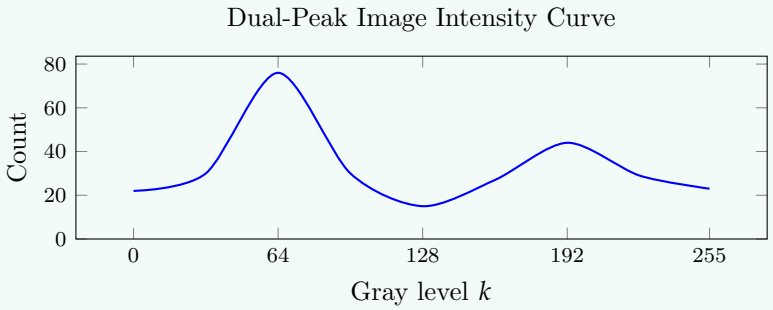
Intensity profile of a bright image: curve shifted towards high gray levels

- A **dark image** shows histogram concentrated on low values.



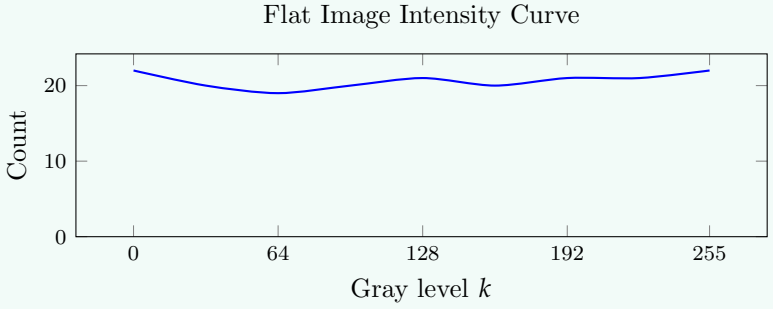
Intensity profile of a dark image: curve shifted towards low gray levels

- A **dual-peak image** indicates presence of both dark (background) and bright (foreground) regions.



Intensity profile of a dual-peak image: two distinct peaks at low and high gray levels

- A **flat histogram** corresponds to uniformly distributed intensities.



Intensity profile of a flat image: uniform distribution across all gray levels

At a Glance:

Histogram Type	Typical Scene	Segmentation Implication
Dark-skewed	Low-light or dark objects	Threshold near lower gray levels
Bright-skewed	Bright background/lighting	Threshold near higher gray levels
Bimodal	Foreground vs. background	Global threshold often effective
Flat/noisy	Low contrast/high noise	Consider contrast stretch or adaptive threshold

1.5 Otsu's Binarization

Otsu's thresholding is a non-parametric, unsupervised method for automatic image binarization. It determines the optimal threshold to separate foreground and background by maximizing inter-class variance.

1.5.1 Probability and Statistical Prerequisites

Basic Probability Concepts Consider an image histogram with L gray levels $(0, 1, ..., L-1)$. Let p_k denote the normalized probability for gray level k:

$$p_k = \frac{n_k}{N}$$

where n_k is the number of pixels with gray level k, and N is the total pixel count.

Class Probabilities, Means, and Variances

For a given threshold T:
Class Probability

- Probability that the pixel belongs to class 0 (background)

$$\omega_0(T) = \sum_{k=0}^{T} p_k \quad \text{(Probability of class 0 - background - black pixels)}$$

- Probability that the pixel belongs to class 1 (foreground)

$$\omega_1(T) = \sum_{k=T+1}^{L-1} p_k \quad \text{(Probability of class 1 - foreground - white pixels)}$$

Class Means

- Probability that the pixel takes values k given that it belongs to class 0

$$\mu_0(T) = \sum_{k=0}^{T} k \cdot p(k|C_0) \quad \text{(Mean of class 0)}$$

$$\text{where} \quad p(k|C_0) = \frac{p(k, C_0)}{p(C_0)} = \frac{p_k p(C_0|k)}{p(C_0)}$$

$$\text{for} k \in \{0 \dots T\} \quad p(k|C_0) = \frac{p_k}{\omega_0(T)}$$

$$\mu_0(T) = \frac{1}{\omega_0(T)} \sum_{k=0}^{T} k p_k$$

$$\text{also, for} k \in \{T+1 \dots L-1\} \quad p(k|C_0) = 0$$

$$\mu_0(T) = \frac{1}{\omega_0(T)} \sum_{k=0}^{L-1} k p_k$$

- Probability that the pixel takes values k given that it belongs to class 1

$$\mu_1(T) = \sum_{k=T+1}^{L-1} k \cdot p(k|C_1) \quad \text{(Mean of class 1)}$$

$$\text{where} \quad p(k|C_1) = \frac{p(k, C_1)}{p(C_1)} = \frac{p_k p(C_1|k)}{p(C_1)}$$

$$\text{for} k \in \{T+1 \dots L-1\} \quad p(k|C_1) = \frac{p_k}{\omega_1(T)}$$

$$\mu_1(T) = \frac{1}{\omega_1(T)} \sum_{k=T+1}^{L-1} k p_k$$

$$\text{also, for} k \in \{0 \dots T\} \quad p(k|C_1) = 0$$

$$\mu_1(T) = \frac{1}{\omega_1(T)} \sum_{k=0}^{L-1} k p_k$$

- Overall Image Mean

$$\begin{aligned} \mu_T &= \sum_{k=0}^{L-1} k p_k \\ &= \sum_{k=0}^{T} k p_k + \sum_{k=T+1}^{L-1} k p_k \\ &= \mu_0(T)\omega_0(T) + \mu_1(T)\omega_1(T) \end{aligned}$$

Class Variances

- Variance of class 0

$$\begin{aligned} \sigma_0^2(T) &= \sum_{k=0}^{T} (k - \mu_0(T))^2 p(k|C_0) \\ &= \sum_{k=0}^{T} (k - \mu_0(T))^2 \frac{p_k}{\omega_0(T)} \end{aligned}$$

- Variance of class 1

$$\begin{aligned}\sigma_1^2(T) &= \sum_{k=T+1}^{L-1} (k-\mu_1(T))^2 p(k|C_1) \\ &= \sum_{k=T+1}^{L-1} (k-\mu_1(T))^2 \frac{p_k}{\omega_1(T)}\end{aligned}$$

- Total Image Variance

$$\begin{aligned}\sigma^2(T) &= \sum_{k=0}^{L-1}(k-\mu_T)^2 p_k \\ &= \sum_{k=0}^{T}(k-\mu_T)^2 p_k + \sum_{k=T+1}^{L-1}(k-\mu_T)^2 p_k \\ &= \sigma_0^2(T) + \sigma_1^2(T) + \omega_0(T)[\mu_0(T)-\mu_T]^2 + \omega_1(T)[\mu_1(T)-\mu_T]^2\end{aligned}$$

1.5.2 Otsu's Binarization: Statement and Algorithm

Definition 1.5.1: Otsu's Binarization

The process of determining the threshold T^* that maximizes the inter-class variance between foreground and background, thus optimally segmenting a bimodal histogram image.

Bibliography