Seminar 1 - Lectures 1, 2 and 3
Computer Vision 1, Master AI, 2025

Martin Oswald, Dimitris Tzionas, Qi Bi, Roan van Blanken

1 Exercise 1: Pinhole Camera

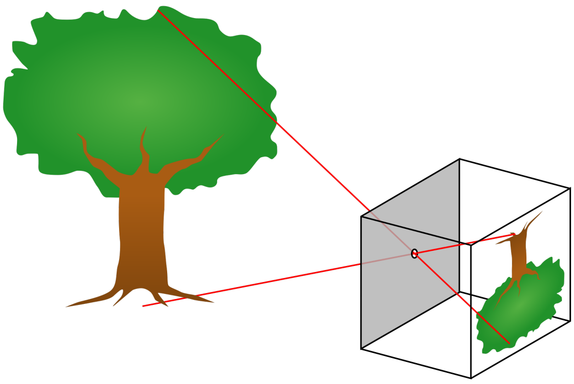
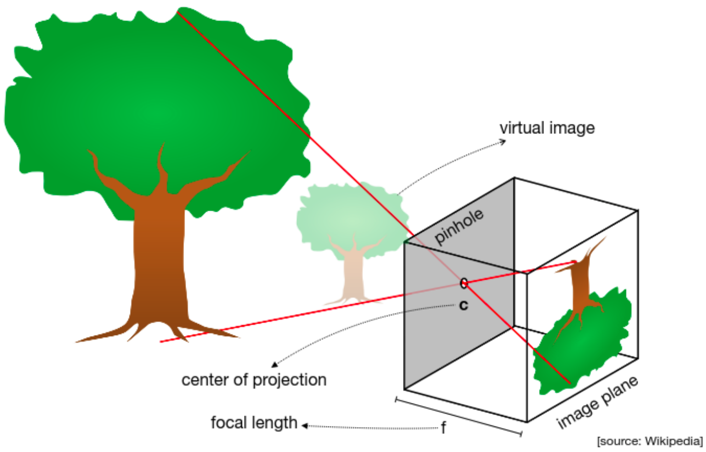


Figure 1: Pinhole camera model (source: Wikipedia)

The camera obscura or pinhole image leverages a natural optical phenomenon. The camera has the shape of a box, light from an object enters through a small hole (the pinhole) in the front and produces an image on the back camera wall.

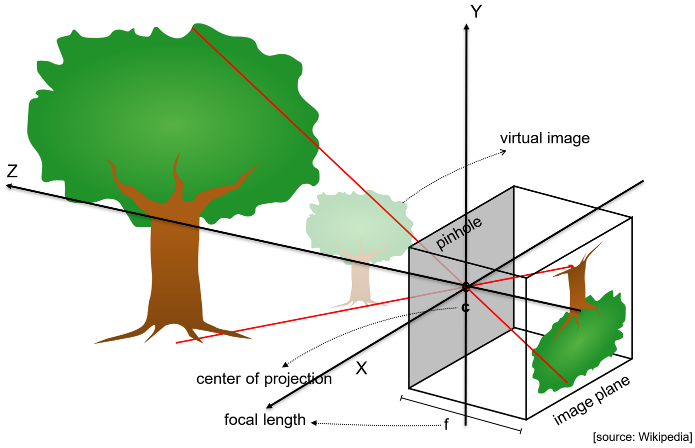
Q.1.a Annotate the Figure 1 with the focal length, center of projection and image plane. What is the virtual image in this context, and where would you place it?

Answer: See this graph:



Q.1.b Following part (a), consider the 3D Euclidean coordinates where the pinhole is located at the origin $\mathbf{O} = [0, 0, 0]^\top$. The Z-axis aligns with the camera axis and points outward from the camera. The Y-axis points vertically upward, and the X-axis points horizontally to the left. Note that this forms a right-handed coordinate system. Draw and label these axes on the provided figure.

Answer: See this graph:



Q.1.c Given a focal length $f = 15$, calculate the coordinates $\mathbf{p}$ of the projection on the image plane for the point $\mathbf{P} = [30, 15, 45]^\top$. Additionally, determine the coordinates $\mathbf{p}_v$ on the virtual image plane.

Answer: The coordinates of the projection $\mathbf{p}$ on the image plane can be calculated using the general formula:

$$\mathbf{p} = \left[-f\frac{X}{Z}, -f\frac{Y}{Z}, -f\right]^\top$$

For the given point $\mathbf{P} = [30, 15, 45]^\top$ and focal length $f = 15$:

$$\mathbf{p} = \left[-15\frac{30}{45}, -15\frac{15}{45}, -15\right]^\top = [-10, -5, -15]^\top$$

For the virtual image plane, which typically mirrors the real image plane across the origin, the coordinates $\mathbf{p}_v$ are:

$$\mathbf{p}_v = [10, 5, 15]^\top$$

2 Exercise 2: Image Representation

Figure 2: First picture to be scanned, stored and recreated in digital pixels (source: Wikipedia)

A digital image is an image composed of picture elements (pixels). Each pixel has a finite and discrete numeric quantity representing the local intensity. In practice, the set of quantities is determined by the number of bits available per pixel. Depending on whether the image resolution is fixed, it may be a vector or raster type. Here we focus on the latter.

Q.2.a Consider a function $f : X \mapsto Y$. How are digital images represented as a function f? What are common sets for Y and their associated image types?

Answer: Digital images can be represented as a function $f : X \mapsto Y$, where X is the set of pixel coordinates in the image, and Y represents the set of possible values for each pixel, typically corresponding to intensity or color.
Common sets for Y include:

- **Binary Images:** $Y = \{0, 1\}$, where each pixel is either black (0) or white (1).
- **Grayscale Images:** $Y = \{0, \ldots, 255\}$, where each pixel has a value between 0 (black) and 255 (white), representing different shades of gray.
- **RGB Images:** $Y = \{0, \ldots, 255\}^3$, where each pixel is represented by three values (R, G, B), each ranging from 0 to 255, corresponding to the intensity of red, green, and blue color channels.

Q.2.b What are the differences between changes to the quantization level and changes to the spatial resolution? How do these changes affect the visual perception of the image?

Answer: The quantization level tells us how many bits are used to represent a single pixel. The spatial resolution describes the number of pixels composing the image. Reducing either the spatial resolution and quantization may result in a smaller size for storage and lower bit-rate for transfer or processing. The trade off is a low spatial resolution image may look blurry or pixelated. A low quantization level may lose nuances on shading and may introduce artifacts.

3 Exercise 3: Camera Projection

Homogeneous Coordinates

Q.3.a Use the directions defined in Exercise 1, convert the following Cartesian 2D vector into homogeneous coordinates: $\mathbf{v} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

Answer: The homogeneous vector $\mathbf{v}_h = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$. This is a 3D vector.

Q.3.b Convert the following homogeneous coordinate into Euclidean coordinates.

$$\mathbf{v}_h = \begin{pmatrix} 6 \\ 9 \\ 12 \\ 3 \end{pmatrix}$$

Answer: To convert a 4D homogeneous coordinate $\mathbf{v}_h = \begin{pmatrix} x & y & z & w \end{pmatrix}^\top$ into a 3D Euclidean coordinate, divide each of the first three elements by the fourth element w:

$$\mathbf{v} = \begin{pmatrix} x/w \\ y/w \\ z/w \end{pmatrix}$$

For the given $\mathbf{v}$, this results in:

$$\mathbf{v} = \begin{pmatrix} 6/3 \\ 9/3 \\ 12/3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

So note that a 3D Euclidean point is represented by a 4D homogeneous vector.

Q.3.c In the lecture, the extrinsic matrix $[R \,|\, t]$ is discussed. When projecting a 3D point onto a 2D image, what are the dimensions of the matrix $[R \,|\, t]$? Additionally, what are the dimensions of the matrices R and t?

Answer: The dimensions of $[R \,|\, t]$ are 3×4 (horizontal $\times$ vertical). R has dimensions 3×3. t has dimensions 3×1.

Q.3.d Explain why homogeneous coordinates are necessary.

Answer: Homogeneous coordinates are ubiquitous in computer graphics because they allow common vector operations such as translation, rotation, scaling and perspective projection to be represented as a matrix by which the vector is multiplied. To keep the mathematics simple, we desire to capture all these types of projections into one linear operator, that is in one matrix: $[R \,|\, t]$. Informally, by using linear operators, we can easily combine several matrices and just multiply them if we would like to do several operations.

FOR THOSE WHO ARE INTERESTED: A quick reminder on a more formal definition of linearity: An operator P is linear if: $P(x + y) = P(x) + P(y)$ and $P(\alpha x) = \alpha P(x)$. Now let us do a 2D example, were we define the operator T to translate (displace) a point in 2D space (cartesian coordinates) by 2 units in the x direction: $T(x) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix}$. We can easily check that according to the definition above, T is NOT a linear operator: $T(\alpha x) = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix} \neq \alpha T(x)$.
We can solve this problem by adding a dimension to the vector and to the operator, and change to homogeneous coordinates. The homogeneous vector becomes: $v_h = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$ and and the homo-

geneous operator for the same translation becomes (refer to lecture 2): $T_h = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, and use normal matrix algebra you can check that the translation operator in this form is linear, e.g.:
$T_h(\alpha v_h) = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha \end{pmatrix} = \begin{pmatrix} \alpha x_1 + 2\alpha \\ \alpha x_2 \\ \alpha \end{pmatrix} = \alpha \begin{pmatrix} x_1 + 2 \\ x_2 \\ 1 \end{pmatrix} = \alpha T_h(v_h)$.

Camera Translations and Rotations

Q.3.e Write the homogeneous 3×4 matrix for the following transformation: The camera is translated by $+7$ units in the X direction.

Answer: A general representation for a 3×4 homogeneous matrix involving the rotation and translation of a camera is:

$$\begin{pmatrix} R_{3\times3} & T_{3\times1} \end{pmatrix}$$

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

For a translation of the camera by $+7$ units along the X-axis, we note that this translation is applied to the entire scene relative to the camera, meaning everything is translated by -7 units in the X direction. The translation vector is $T = (-7, 0, 0)^\top$, and R is the identity matrix. Thus, the homogeneous transformation matrix is:

$$\begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Q.3.f Write the homogeneous 3×4 matrix for the following transformation: Rotate by 30 degrees along the X-axis, with the rotation being counter-clockwise when looking along the positive direction of X.

Answer: For a rotation about the X-axis by an angle θ, the general 3×4 homogeneous matrix is given by:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \end{pmatrix}$$

Given $\theta = 30°$, we substitute $\cos 30° = \sqrt{3}/2$ and $\sin 30° = 1/2$, resulting in:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{3}/2 & -1/2 & 0 \\ 0 & 1/2 & \sqrt{3}/2 & 0 \end{pmatrix}$$

Since this is purely a rotation, the translation vector is $\mathbf{T} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

Q.3.g Write the homogeneous 3×4 matrix for the combination of the rotation followed by the translation as described above.

Answer: When combining these transformations, we apply the rotation first, followed by the translation. The combined homogeneous 3×4 matrix is:

$$\begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & \sqrt{3}/2 & -1/2 & 0 \\ 0 & 1/2 & \sqrt{3}/2 & 0 \end{pmatrix}$$

This matrix represents the sequence of rotating the camera by 30 degrees around the X-axis and then translating it by -7 units along the X-axis.

It is also possible to obtain this result by matrix multiplication of $T \times R$ but we have to be careful with the matrix dimensions now, therefore we need to change the rotation matrix a bit. The first step is a Rotation R, rotate a 4D homogeneous vector in a rotated 4D homogeneous vector. A general representation for 4x4 matrices involving rotation and translation is

$$\begin{pmatrix} R_{3\times 3} & T_{3\times 1} \\ 0_{1\times 3} & 1_{1\times 1} \end{pmatrix} \tag{1}$$

where R is 3×3 rotation matrix, and T is a 3×1 translation matrix. So now we obtain with matrix multiplication:

$$\begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{3}/2 & -1/2 & 0 \\ 0 & 1/2 & \sqrt{3}/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & \sqrt{3}/2 & -1/2 & 0 \\ 0 & 1/2 & \sqrt{3}/2 & 0 \end{pmatrix} \tag{2}$$

4 Exercise 4: Color

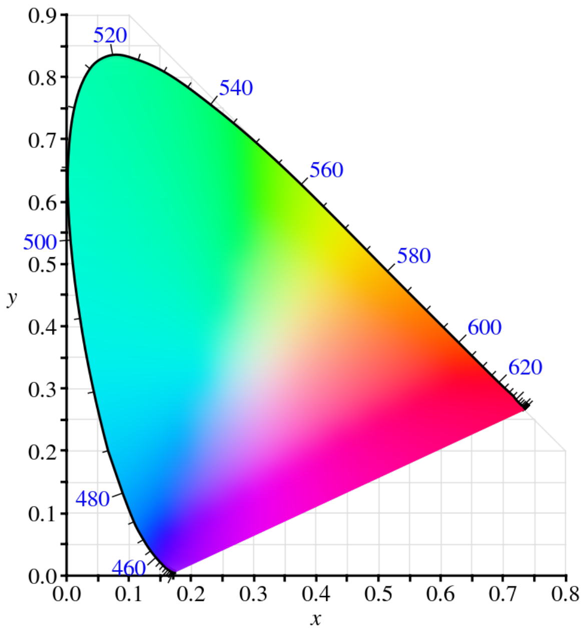


Figure 3: Chromaticity diagram

To calculate the color of light sources, the following intuitive color models are used: intensity V, hue H, saturation S and chromaticity xy, Let's assume, for simplicity reasons, that sunlight S is given by the CIE values $X = Y = Z = 100$. Further, let $X = 100$, $Y = 100$ and $Z = 150$ be the values for a given artificial lamp A.

Q.4.a Calculate the intensity V (in the same units as the X, Y, Z values) for the two light sources S and A.

Answer: The intensity I is calculated as:

$$I = \frac{X + Y + Z}{3}$$

For sunlight S:

$$I_S = \frac{100 + 100 + 100}{3} = 100$$

For the artificial lamp A:

$$I_A = \frac{100 + 100 + 150}{3} = 117$$

Q.4.b Calculate the chromaticity values $x = \dfrac{X}{X + Y + Z}$, $y = \dfrac{Y}{X + Y + Z}$ for the two light sources S and A, and plot these on the chromaticity diagram given in Figure 3.

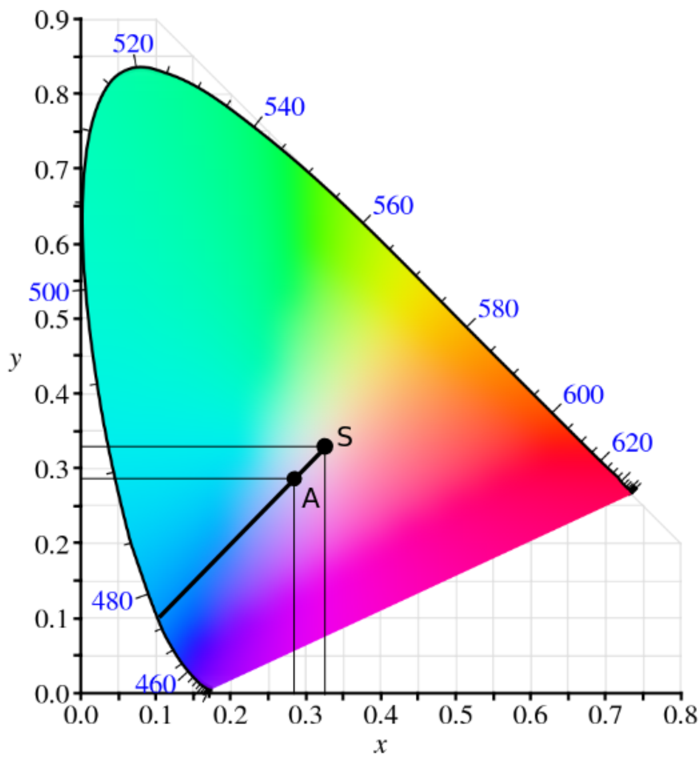
Answer: For sunlight S:

$$x_S = \frac{100}{300} = \frac{1}{3}, \quad y_S = \frac{100}{300} = \frac{1}{3}, \quad z_S = \frac{100}{300} = \frac{1}{3}$$

For the artificial lamp A:

$$x_A = \frac{100}{350} \approx 0.286, \quad y_A = \frac{100}{350} \approx 0.286, \quad z_A = \frac{150}{350} \approx 0.428$$

See this figure:



Q.4.c What is the estimated hue of light source A with S as reference white light?

Answer: The line from S through A intersects the boundary at around 476 nm, which represents the hue (dominant wavelength) of A with S as the reference white light.

Q.4.d Rank the light sources with respect to their saturation.

Answer: $S < A$

Q.4.e Plot the colors produced through the mixture of S and A.

Answer: The colors produced by the mixture of S and A lie on the line connecting S and A in the chromaticity diagram.

5 Exercise 5: Reflection

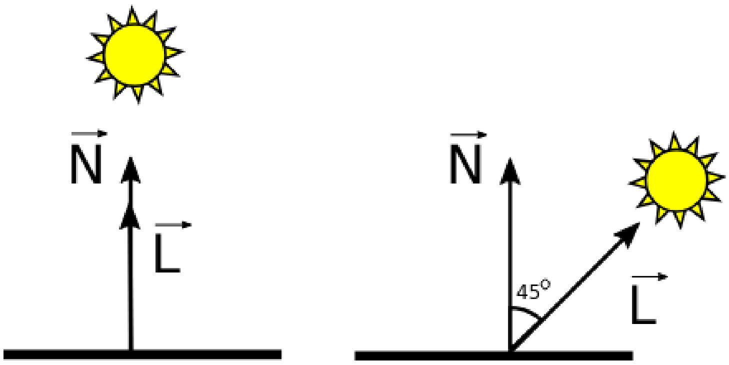


Figure 4: Light source position scenarios: aligned with the surface normal (left) and at an angle of 45° (right).

Consider the color of a matte, non-glossy surface under white light illumination. The color at a specific point on the surface is described by the following reflection model:

$$R = Ik_R \cos\theta, \quad G = Ik_G \cos\theta, \quad B = Ik_B \cos\theta$$

where:

- I is the intensity of the white light source,
- k_R, k_G, and k_B represent the fractions of red, green, and blue light reflected by the surface (i.e., the inherent color of the surface),
- $\cos\theta = \vec{n} \cdot \vec{l}$ is the dot product of two unit vectors: $\vec{n}$ (the surface normal) and $\vec{l}$ (the direction of the light source).

Refer to Figure 4 for visual scenarios where the light source is either aligned with the surface normal or at an angle of 45° to it.

Q.5.a Assume that the surface is flat and homogeneously colored. Explain why the intensity is higher when the surface normal coincides with the direction of the light source compared to when it is observed at an angle relative to the light source direction.

Answer: The perceived color intensity is given by:

$$R = Ik_R \cos\theta, \quad G = Ik_G \cos\theta, \quad B = Ik_B \cos\theta$$

where I is the intensity of the light source, and k_R, k_G, k_B are the surface's albedo constants for red, green, and blue, respectively.

The intensity depends on $\cos\theta$, where θ is the angle between the surface normal $\vec{n}$ and the light source direction $\vec{l}$. Consider two scenarios: (1) when the light source direction coincides with the surface normal, and (2) when they form an angle of 45° (refer to Figure 4).

In the first case, $\cos\theta = \cos 0 = 1$, resulting in maximum intensity. In the second case, $\cos\theta = \cos 45° = \frac{\sqrt{2}}{2}$, leading to lower intensity.

Q.5.b Assume that the color of the surface is yellow, i.e., $R = 100$, $G = 100$, and $B = 10$. Explain what will happen to the values of R, G, and B if the intensity of the light source diminishes. Assume that the camera response to intensity is linear and that intensity is linearly coded.

Answer: The intensity of each color component is given by:

$$R = Ik_R \cos\theta, \quad G = Ik_G \cos\theta, \quad B = Ik_B \cos\theta$$

where I is the light source intensity. As the light source intensity diminishes, the values of R, G, and B will decrease proportionally.

This means that the resulting set of colors will lie on a line connecting the origin with the point $RGB(100, 100, 10)$ in the RGB-color space (see Figure 5).

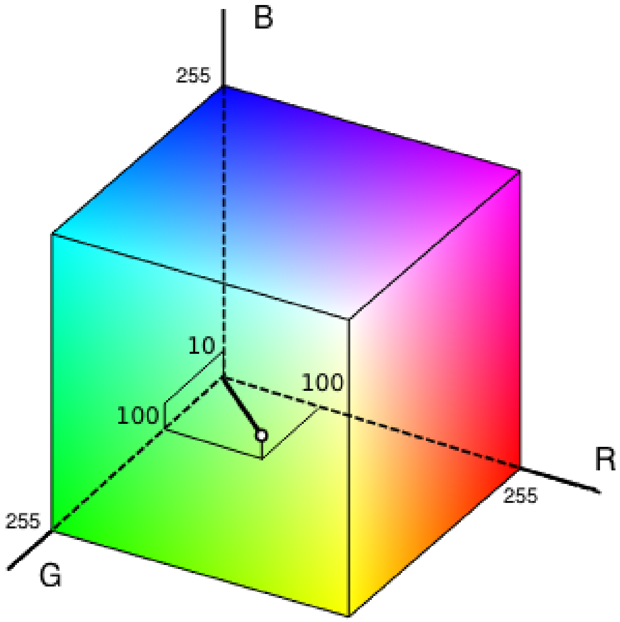


Figure 5: The position of color $RGB(100, 100, 10)$ in the RGB-color space

Q.5.c In the case of a curved (non-flat) surface, assuming no occlusion or interreflection, indicate where the colors will be positioned in the RGB-color space. Explain your answer.

Answer: The positioning of colors in the RGB-color space for a curved surface is similar to that for a flat surface. The colors will lie along a line connecting the origin to the point representing the surface's color in the RGB-color space. This is because, despite the curvature, the reflection model remains the same, and the color intensities still vary proportionally with the light source intensity I. Thus, the curvature does not change the general relationship between R, G, and B, but it may cause different parts of the surface to reflect light at different intensities, leading to a distribution of colors along this line.

Q.5.d A simple color invariant is given by R/G. Prove that R/G is independent of the intensity of the light source I, object geometry, and the direction of the light source.

Answer: The ratio R/G can be expressed as:

$$\frac{R}{G} = \frac{Ik_R \cos\theta}{Ik_G \cos\theta} = \frac{k_R}{k_G}$$

Here, I (the light source intensity), $\cos\theta$ (related to object geometry and light source direction), and $Ik_s \cos^n \alpha$ cancel out, proving that R/G depends only on the surface's inherent color properties, k_R and k_G, and is therefore independent of the light source intensity, geometry, and direction.

Q.5.e Consider the same surface, but assume it is glossy (instead of matte). The reflection model is now given by:

$$R = Ik_R \cos\theta + Ik_s \cos^n \alpha, \quad G = Ik_G \cos\theta + Ik_s \cos^n \alpha, \quad B = Ik_B \cos\theta + Ik_s \cos^n \alpha$$

where k_s is the specular reflection coefficient, and $\cos^n \alpha$ depends on the glossiness and viewing angle.

Prove that $\frac{R-G}{R-B}$ is a color invariant for shiny surfaces.

Answer: The reflection model introduces a specular component $Ik_s \cos^n \alpha$, but the ratio $\frac{R-G}{R-B}$ can still be shown to be invariant:

$$\frac{R-G}{R-B} = \frac{(Ik_R \cos\theta + Ik_s \cos^n \alpha) - (Ik_G \cos\theta + Ik_s \cos^n \alpha)}{(Ik_R \cos\theta + Ik_s \cos^n \alpha) - (Ik_B \cos\theta + Ik_s \cos^n \alpha)} = \frac{I(k_R \cos\theta - k_G \cos\theta)}{I(k_R \cos\theta - k_B \cos\theta)} = \frac{k_R - k_G}{k_R - k_B}$$

This proves that $\frac{R-G}{R-B}$ remains a color invariant, as the specular reflection component $Ik_s \cos^n \alpha$ cancels out in the ratio.

6 Exercise 6: Histograms & Images

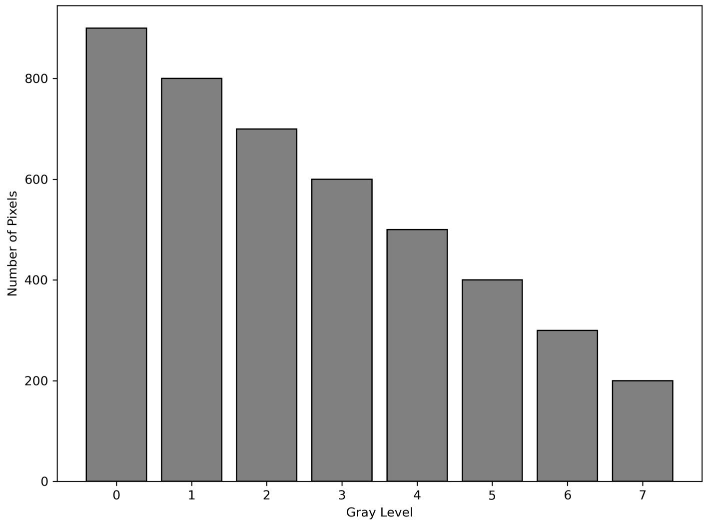
Histograms
To calculate the histogram equalization of an image, we use the histogram of the image, binned by gray levels (L). Consider the distribution of pixel values in a 3-bit grayscale image given in Table 1. The total number of gray levels is 8 ($2^3 = 8$), ranging from 0 to 7.

Gray Level	0	1	2	3	4	5	6	7
No. of Pixels	900	800	700	600	500	400	300	200

Table 1: Pixel distribution of a random 3-bit grayscale image

Q.6.a Draw the histogram of the given 3-bit gray levels. What can you infer from the given gray level information? What type of image do you think this represents? Can you reconstruct the image using just the information given in Table 1? Justify your answer.

Answer: See this figure:



The histogram suggests a gradient with more pixels concentrated at lower gray levels, indicating that the image is likely rather dark with a gradual shift towards lighter regions. However, the histogram alone does not provide sufficient information to reconstruct the image, as it only shows the frequency of each gray level and lacks spatial information, such as the location of each pixel in the image.

Q.6.b Calculate the probabilities of pixels for each gray level.

Answer: The probability of a pixel belonging to a specific gray level is calculated by dividing the number of pixels at that gray level by the total number of pixels. Let N represent the total number of pixels in the image. Given the values from Table 1, the probabilities for each gray level are calculated as follows:

$$\text{Probability}(i) = \frac{\text{No. of pixels at gray level } i}{N}$$

The total number of pixels $N = 900 + 800 + 700 + 600 + 500 + 400 + 300 + 200 = 4400$.

Thus, the probabilities are:

Gray Level	0	1	2	3	4	5	6	7
Probability	0.2045	0.1818	0.1591	0.1364	0.1136	0.0909	0.0682	0.0455

Table 2: Probability distribution of gray levels in a random 3-bit grayscale image

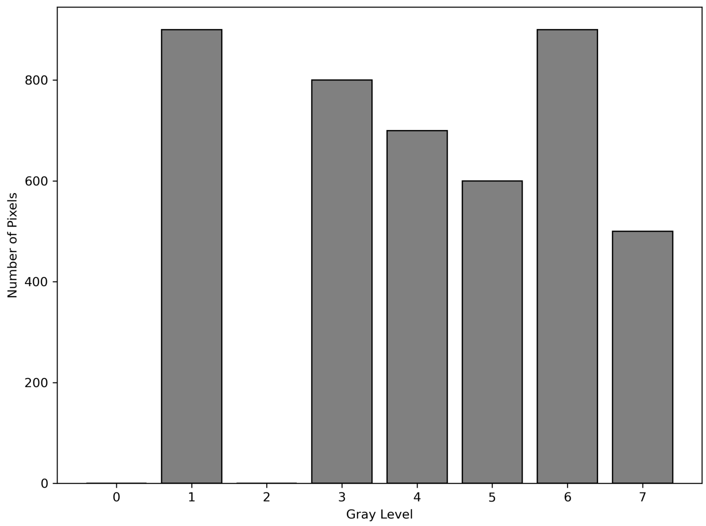
Q.6.c Equalize the histogram and create a table showing the new gray levels and their corresponding pixel counts. Then, draw the equalized histogram.

Answer: First, calculate the cumulative distribution function (CDF) for the given gray levels and use it to determine the new gray levels after equalization. The value L in the table refers to the number of gray levels in the image. In this case, for a 3-bit grayscale image, $L = 8$ (since $2^3 = 8$). The resulting table should look like this:

Gray Level	k	n	$P_k = \frac{n}{N}$	CDF	CDF * (L - 1)	$\hat{k}$	New Pixel Count
0	0	900	0.2045	0.2045	1.43	1	900
1	1	800	0.1818	0.3864	2.70	3	800
2	2	700	0.1591	0.5455	3.82	4	700
3	3	600	0.1364	0.6818	4.77	5	600
4	4	500	0.1136	0.7955	5.57	6	900
5	5	400	0.0909	0.8864	6.20	6	-
6	6	300	0.0682	0.9545	6.68	7	500
7	7	200	0.0455	1.0000	7.00	7	-
$N = \sum n$		4400			$N = \sum n$		4400

Table 3: Histogram equalization with new pixel counts after equalization

See this figure:



Images

1	1	1	3	3	4	4	1
4	0	0	5	1	2	0	1
0	1	3	1	2	0	2	0
3	1	2	2	2	1	1	2
1	4	2	2	2	3	5	3
0	2	1	0	2	1	3	0
0	2	1	1	0	0	4	1
0	4	0	0	0	2	1	1

Table 4: 3-bit grayscale image patch

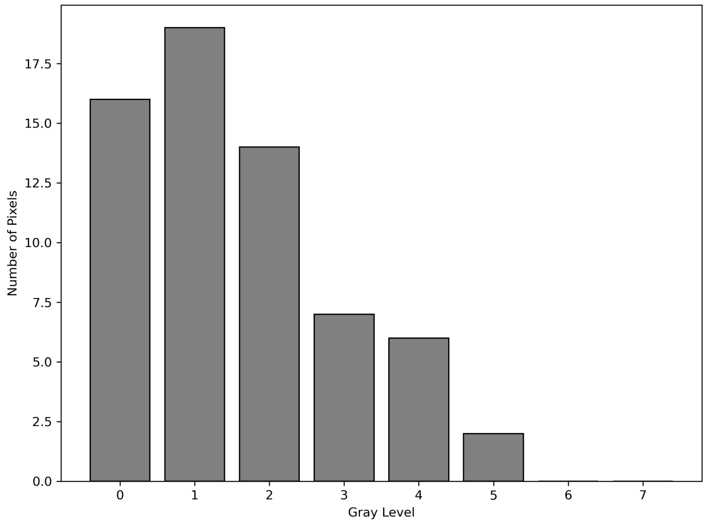
Q.6.d Consider the 3-bit grayscale image patch given in Table 4. Calculate the pixel distribution like in Table 1 and draw the histogram.

Answer: Pixel distribution:

Gray Level	0	1	2	3	4	5	6	7
No. of Pixels	16	19	14	7	6	2	0	0

Table 5: Pixel distribution of the 3-bit grayscale image patch

See this figure for the histogram based on this distribution.



Q.6.e Equalize the histogram and then back-project the equalized histogram to obtain the new image. Draw 3-bit grayscale image patch after histogram equalization

Answer: First, calculate the cumulative distribution function (CDF) for the given gray levels and use it to determine the new gray levels after equalization. The resulting table should look like this:

Gray Level	k	n	$P_k = \frac{n}{N}$	CDF	CDF * (L - 1)	$\hat{k}$	New Pixel Count
0	0	16	0.2500	0.2500	1.7500	2	16
1	1	19	0.2969	0.5469	3.8281	4	19
2	2	14	0.2188	0.7656	5.3594	5	14
3	3	7	0.1094	0.8750	6.1250	6	7
4	4	6	0.0938	0.9688	6.7812	7	8
5	5	2	0.0312	1.0000	7.0000	7	-
6	6	0	0.0000	1.0000	7.0000	7	-
7	7	0	0.0000	1.0000	7.0000	7	-
$N = \sum n$		64			$N = \sum n$		64

Table 6: Histogram equalization with new pixel counts after equalization

Next, back-project the equalized histogram to obtain the new image, as shown in Table 7.

4	4	4	6	6	7	7	4
7	2	2	7	4	5	2	4
2	4	6	4	5	2	5	2
6	4	5	5	5	4	4	5
4	7	5	5	5	6	7	6
2	5	4	2	5	4	6	2
2	5	4	4	2	2	7	4
2	7	2	2	2	5	4	4

Table 7: 3-bit grayscale image patch after histogram equalization

Q.6.f So far, you have equalized only grayscale images. How would you apply the algorithm to an RGB image? Discuss your answer.

Answer: To apply histogram equalization to an RGB image, first convert the image from RGB to HSV color space, which separates the color information (Hue and Saturation) from the intensity information (Value). Then, perform histogram equalization on the V (Value) channel to enhance the image's contrast while preserving its color characteristics. After equalizing the V channel, recombine it with the original H and S channels, and convert the image back to RGB color space. This approach improves the brightness and contrast of the image without altering the original colors.