Digital Image Processing

Daniel T. McGuiness, PhD

Version: γ.2024.2

MCI

Table of Contents

Lecture Structure

Table of Contents

Lecture Structure

- The goal of this lecture is to give you the fundamentals of digital image processing and understanding of mathematical principles.
- This lecture is a total of 4 SWS with a total of sixty (60) hours.
- There are two (2) assignments for this course

  1st will be a pre-defined work which is individual based.
  2nd will be group based.

> You are to come up with a project that uses DIP using Python.

- You will work with a group of up to three (3) or two (2).
- You are to come up with a group and decide on your topic.

Lecture Structure

- The individual assignment focuses on understanding DIP principles.
- The assignment is uploaded to SAKAI for you to work on along with what is required of you for submission.
  - The assignment contains questions where applications of DIP will be needed.
- The deadline is the end day of **last lecture before presentations**.

A Help in Colour

Due to the nature of the topic, some aspects are to be presented in a colour spectrum some student may not be able to perceive. In situations like this, please let me know if there are some diagrams or some colour choices making the lecture illegible via mail and I will send you a colour correct version based on the condition.

Lecture Structure

- For your project use Python.
- Some possible project ideas:
  - License plate detection,
  - Handwriting detection,
  - Signature verification,
  - Face detection,
  - Image to text conversion,
  - Barcode detection,
  - Convert sudoku drawings to computer code.
  - Book detection.

> The use of AI/ML is allowed as long as clear explanation is given and its process is understood.

Lecture Structure

- The last three (3) appointments are reserved for group presentations.
- You will do a presentation in front of the class for 20 mins.
- The next 20 mins following your presentation will be the Q&A.
- The Q&A will involve two (2) questions from your relevant work.
- You are also to submit a report with your project detailing the work.

> Each student needs to declare the part the student worked on.

- i.e., Student A has done the writing, edge detection
- i.e., Student B has done the data analysis, figure generation.
- You are to submit your reports and all relevant resources to SAKAI no later than 2 weeks before your assigned presentation.

Lecture Structure

Assessment Type	Overall Points	Breakdown	%
Homework	40		
		Report	20
		Solution(s)	60
		Code Analysis	20
Group Project	60		
		Report	40
		Presentation	40
		Q & A	20

Table 1: Assessment Grade breakdown for the lecture.

Lecture Structure

Covered Topic	Appointment
Mathematical Fundamentals	1
Perception	2
Camera	2-3
Display	4
Noise	4-5
Histogram Operations	6
Morphological Operations	7
Blurring Filters	8
Feature Analysis	9
Edge Detection	10
Neural Networks for Image Processing	11-12
Group Assignment Presentations	13-15

Table 2: Distribution of materials across the semester.

Lecture Structure

Mathematical Fundamentals

- 2D Convolution,
- Discrete Fourier Transform,
- Sampling Theorem

Perception

- Colour Blindness,
- Colour Standards,
- Colour Models

Lecture Structure

Cameras

- Used sensors,
- Lenses,
- Sensitivity

Displays

- Dithering,
- Interlacing,
- Display Technologies

Noise

- Types of noises,
- Modelling Noises,
- Random Noise generation

Histogram Operations

- Colour Channels,
- Masking,
- Dynamic Range

Morphological Operations

- Opening,
- Closing,
- Erosion,
- Dilation.

Blurring Filters

- Gaussian Blurring,
- Multivariate Distribution,
- Bilinear Filtering

Feature Analysis

- ORB Feature Extractor,
- Adaptive Threshold,
- Scale Invariant Feature Transform.

Lecture Structure

Egde Detection

- Defining an Edge to the computer,
- Types of Kernels,
- Canny Edge Detection.

Neural Networks for Image Processing

- Defining ANNs,
- OCR,
- ResNet.

Books

- Forsyth, Ponce "*Computer Vision: A Modern Approach*" Prentice-Hall, 2003.
- Young I. "*Fundamentals of Image Processing*" Delft 1998.
- Szeliski R. "*Computer Vision: Algorithms and Applications*" Springer 2022.
- Nixon M. et. al "*Feature Extraction and Image Processing for Computer Vision*" Academic press 2019.
- Gonzalez R. "*Digital Image Processing*" Pearson 2009

White Papers

- Luminera "*Getting it Right: Selecting a Lens for a Vision System*",
- Luminera "*The Complete Guide to Industrial Camera Lenses*",
- Fowler B, et. al, *Read Noise Distribution Modeling for CMOS Image Sensors*.
- Oxford Instruments *Understanding Read Noise in sCMOS Cameras*.

Lecture Structure

Lecture Notes

- Applied Multi-variable Statistical Analysis "*Lesson 4: Multivariate Normal Distribution*",
- Statistical Theory and Methods I "*Chapter 3: Multivariate Distributions*", Stephen M. Stigler
- The Discrete Fourier Transform "*Signal Processing & Filter Design*", Stephen Roberts.
- Procedural Generation: 2D Perlin Noise *Game Programming*, Mount .E, Eastman R.
- Foundations of computer vision: Lecture notes, Carreira-Perpinan M.
- Computer Vision, CMU School of Computer Science
- Computer Vision, University of Cambridge
- Computer Vision, NYU Computer Science

Web Resources

- Scikit-image documentation
- OpenCV documentation
- Pillow (fork of PIL) documentation

Mathematical Fundamentals

Table of Contents

Learning Outcomes

(LO1) An Overview of Mathematical Methods,
(LO2) Description of Analogue and Digital,
(LO3) Fourier Analysis Overview,
(LO4) Convolution Introduction.

Mathematical Fundamentals

- Computer Vision encompasses multiple disciplines, including digital image processing, cameras and displays.
- To better prepare, it is important to refresh/learn some fundamental mathematical principles & concepts.

Concepts and Principles

- Convolution
- Fourier Analysis
  - Properties
  - Discrete Fourier Transform
- Shannon-Nyquist Sampling Theorem
- A brief introduction to Information Theory
  - Entropy in Information Theory

Mathematical Fundamentals

- Convolution, mathematically is defined as:

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau)\, g(t - \tau)\, \mathrm{d}\tau.$$

> In layman's terms convolution is just fancy multiplication.

Example

Imagine you manage a hospital treating patients with a single disease.

You have:

Treatment Plan `3` Every patient gets 3 units of the cure on their first day.

Patient List `[1, 2, 3, 4, 5]` Your patient count for the week (1 person Monday, 2 people on Tuesday, etc.).

How much medicine do you use each day?

Solution

The answer is a quick multiplication:

$$\text{Plan} \times \text{Patients} = \text{Daily Usage}$$

`3` × `[1, 2, 3, 4, 5]` = `[3, 6, 9, 12, 15]`

Multiplying the plan by the patient list gives usage for upcoming days:

`[3, 6, 9, 12, 15]`

Everyday multiplication of (3×4) means using the plan with a single day of patients:

`[3]` × `[4]` = `[12]`

Example

Now the disease mutates and needs multi-day treatment. A new plan:

Plan: `[3, 2, 1]`

Meaning:

- 3 units of the cure on day one,
- 2 units on day two,
- 1 unit on day three.

Given the same patient schedule of:

Patient: `[1, 2, 3, 4, 5]`

what's our medicine usage each day?

Solution

Let's see

- On day 1, 1 patient `A` comes in. It's their first day, so 3 units.
- On day 2, `A` gets 2 units (second day), but two new patients (`B1` & `B2`) arrive, who get 3 each ($2 \times 3 = 6$).
  - The total is $2 + (2 \times 3) = 8$ units.
- On Wednesday, it's trickier: The patient `A` finishes (1 unit, her last day), the `B1` and `B2` get 2 units (2 * 2), and there are 3 new Wednesday people....

The patients are overlapping and it's hard to track. How can we organise this calculation?

Solution

An idea worth considering is to reverse the order of the patient list:

New Patient List: `[5, 4, 3, 2, 1]`

Next, imagine we have 3 separate rooms where we apply the proper dose:

Rooms: `[3, 2, 1]`

On your first day, you walk into the first room and get 3 units of medicine. The next day, you walk into room #2 and get 2 units. On the last day, you walk into room #3 and get 1 unit. There's no rooms afterwards, and your treatment is done.

Solution

To calculate the total medicine usage, line up the patients and walk them through the rooms:

```
Monday
----------------------------
Rooms                  3 2 1
Patients       5 4 3 2 1

Usage                  3
```

On Monday (our first day), we have a single patient in the first room. `A` gets 3 units, for a total usage of 3.

Makes sense, right?

Solution

On Tuesday, everyone takes a step forward:

```
Tuesday
----------------------------
Rooms                  3 2 1
Patients ->      5 4 3 2 1

Usage                  6 2      = 8
```

The first patient is now in the second room, and there's 2 new patients in the first room. We multiply each room's dose by the patient count, then combine.

Solution

```
Wednesday
----------------------------
Rooms                  3 2 1
Patients ->        5 4 3 2 1
Usage                  9 4 1    = 14

Thursday
-----------------------------
Rooms                  3 2 1
Patients ->          5 4 3 2 1
Usage                 12 6 2    = 20

Friday
-----------------------------
Rooms                  3 2 1
Patients ->            5 4 3 2 1
Usage                 15 8 3    = 26
```

Solution

It's intricate, but we figured it out, right? We can find the usage for any day by reversing the list, sliding it to the desired day, and combining the doses.

The total day-by-day usage looks like this (don't forget Sat and Sun, since some patients began on Friday):

```
Plan      *  Patient List   = Total Daily Usage

[3 2 1]   *  [1 2 3 4 5]    = [3 8 14 20 26 14 5]
              M T W T F        M T W  T  F  S  S
```

This calculation is the convolution of the plan and patient list. It's a fancy multiplication between a list of input numbers and a "program".

Example

Write a script which does convolution of the following two (2) arrays:

$$A = [1,\ 1,\ 2,\ 2,\ 1] \qquad\qquad B = [1,\ 1,\ 1,\ 3]$$

Solution

```python
import numpy as np
def convolve_1d(signal, kernel):
    kernel = kernel[::-1]
    k = len(kernel)
    s = len(signal)
    signal = [0]*(k-1)+signal+[0]*(k-1)
    n = s+(k-1)
    res = []
    for i in range(s+k-1):
        res.append(np.dot(signal[i:(i+k)], kernel))
    return res
```

```python
A = [1,1,2,2,1]
B = [1,1,1,3]

print(convolve_1d(A, B))
```

- An operation on two functions (f and g) that produces $f * g$.

It expresses how the shape of one is modified by the other.

- There are several notations to indicate convolution with the most common is:

$$c = f(t) * g(t) = (f * g)(t),$$

- In 2D continuous space (i.e., analogue):

$$
\begin{aligned}
c(x, y) &= f(x, y) * g(x, y), \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\chi, \xi)\, g(x-\chi, y-\xi)\, \mathrm{d}\chi\, \mathrm{d}\xi.
\end{aligned}
$$

- In 2D discrete space (i.e., digital):

$$
\begin{aligned}
c[m, n] &= f[m, n] * g[m, n], \\
&= \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} f[j, k]\, g[m-j, n-k].
\end{aligned}
$$

Mathematical Fundamentals

- It is the single most important technique in Digital Signal Processing.
- Using the strategy of impulse decomposition, systems are described by a signal called the impulse response.
- Convolution is important because it relates the three (3) signals of interest:
  1. Input signal,
  2. Output signal,
  3. Impulse response.

Commutative

- The order in which we convolve two signals does not change the result:

$$f\left(t\right)*g\left(t\right)=g\left(t\right)*f\left(t\right)$$

Distributive

- if there are three signals $f\left(t\right), g\left(t\right), h\left(t\right)$, then the convolution of $f\left(t\right)$ is distributive:

$$f\left(t\right)*\left[g\left(t\right)+h\left(t\right)\right]=\left[f\left(t\right)*g\left(t\right)\right]+\left[f\left(t\right)*h\left(t\right)\right]$$

Associative

- The way in which the signals are grouped in a convolution does not change the result:

$$f(t) * [g(t) * h(t)] = [f(t) * g(t)] * h(t)$$

Shift Property

- The convolution of a signal with a time shifted signal results a shifted version of that signal.
- i.e.,

$$f(t) * g(t) = y(t)$$

- Then according to the shift property of convolution:

$$f(t) * f(t - T_0) = y(t - T_0)$$

- Similarly:

$$f(t - T_0) * f(t) = y(t - T_0)$$

- Therefore:

$$f(t - T_1) * f(t - T_2) = y(t - T_1 - T_2)$$

Mathematical Fundamentals

$$
\underbrace{\begin{pmatrix}
0_{\times 1} & 1_{\times 0} & 1_{\times 1} & 1 & 0 & 0 & 0 \\
0_{\times 0} & 0_{\times 1} & 1_{\times 0} & 1 & 1 & 0 & 0 \\
0_{\times 1} & 0_{\times 0} & 0_{\times 1} & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}}_{I}
*
\underbrace{\begin{pmatrix}
1 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 1
\end{pmatrix}}_{K}
=
\underbrace{\begin{pmatrix}
\boxed{1} & 4 & 3 & 4 & 1 \\
1 & 2 & 4 & 3 & 3 \\
1 & 2 & 3 & 4 & 1 \\
1 & 3 & 3 & 1 & 1 \\
3 & 3 & 1 & 1 & 0
\end{pmatrix}}_{I * K}
$$

Figure 1: A visual representation of how convolution works in 2D.

Mathematical Fundamentals

- Converting from a continuous 2D data $a(x, y)$ to its digital representation $a[x, y]$ requires the process of sampling.
- An ideal sampling system is defined as the image $a(x, y)$ multiplied by an ideal 2D impulse train $\delta(x, y)$:

$$
\begin{aligned}
b[m, n] &= a(x, y) \sum_{m=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \delta(x - mX_0, y - nY_0) \\
&= \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} a(mX_0, nY_0)\, \delta(x - mX_0, y - nY_0)\,.
\end{aligned}
$$

  where X_0 and Y_0 are the sampling distance or intervals and δ is the Dirac delta function.
- If you were to sample in square shapes $X_0 = Y_0$ where you could think of each individual block a pixel

Mathematical Fundamentals

- To reconstruct a continuous analog signal from its sampled version accurately, the sampling rate must be at least twice the highest frequency present in the signal.
- This ensures that there are enough samples taken per unit of time to capture all the details of the original waveform without introducing aliasing, which can cause distortion or artifacts in the reconstructed signal.

$$f_{\mathrm{s}} \geq 2f_{\mathrm{m}}$$

where f_{s} is the signal frequency, f_{m} is the maximum sample frequency.

> This is only a theoretical limit, not a practical one.

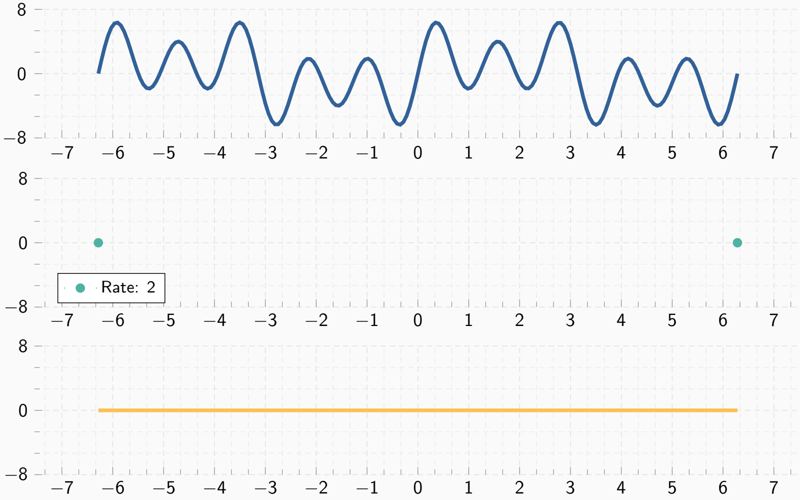

Figure 2: The effects of signal reconstruction on the sampling rate.

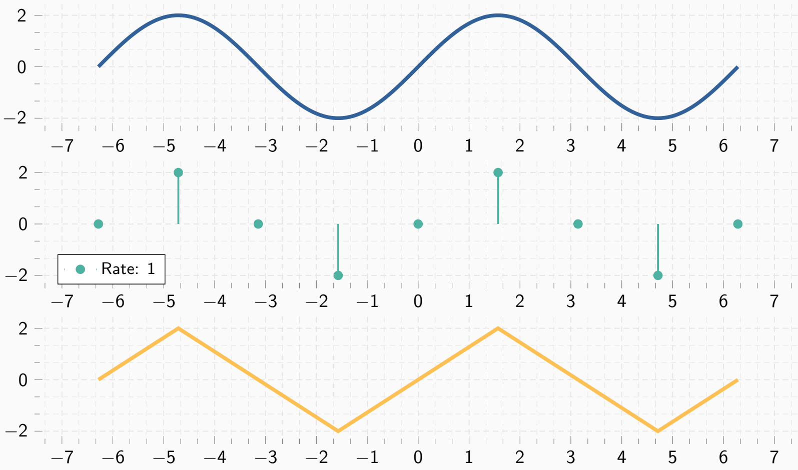

Figure 3: Reconstruction of the signal with 1 times the signal frequency.

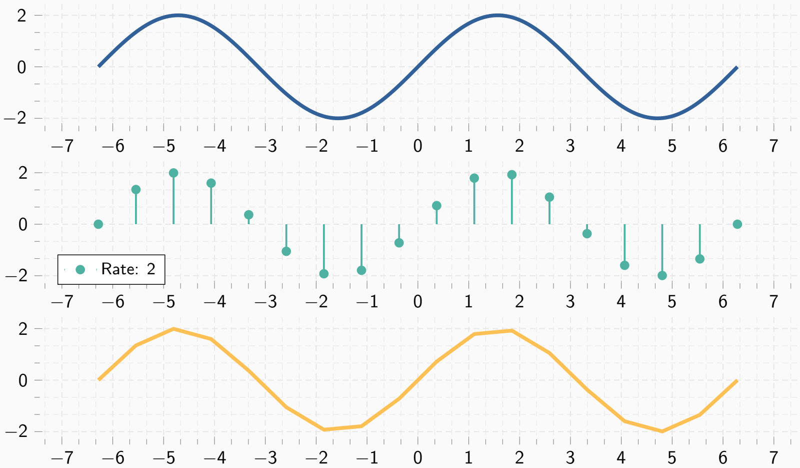

Figure 4: Reconstruction of the signal with 2 times the signal frequency.

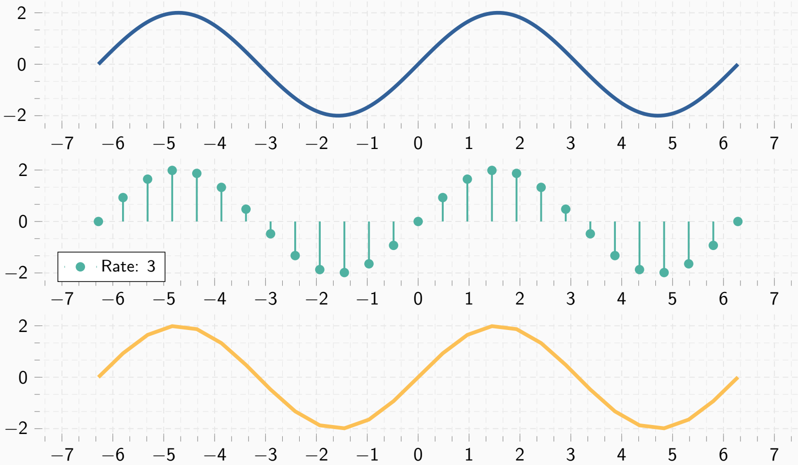

Figure 5: Reconstruction of the signal with 3 times the signal frequency.

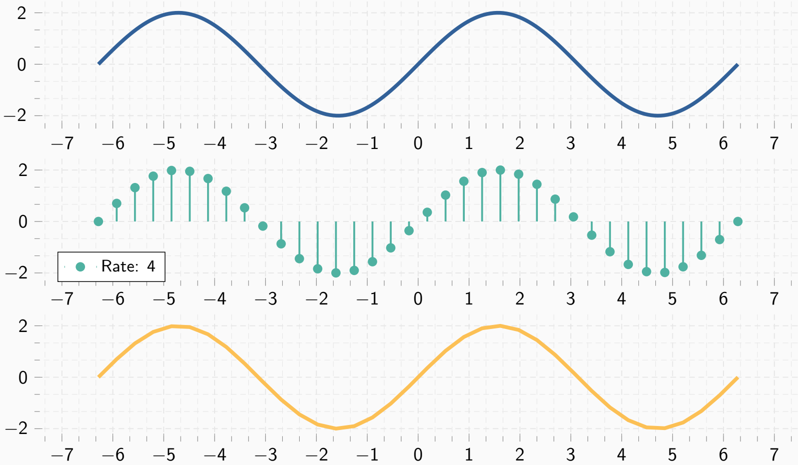

Figure 6: Reconstruction of the signal with 4 times the signal frequency.

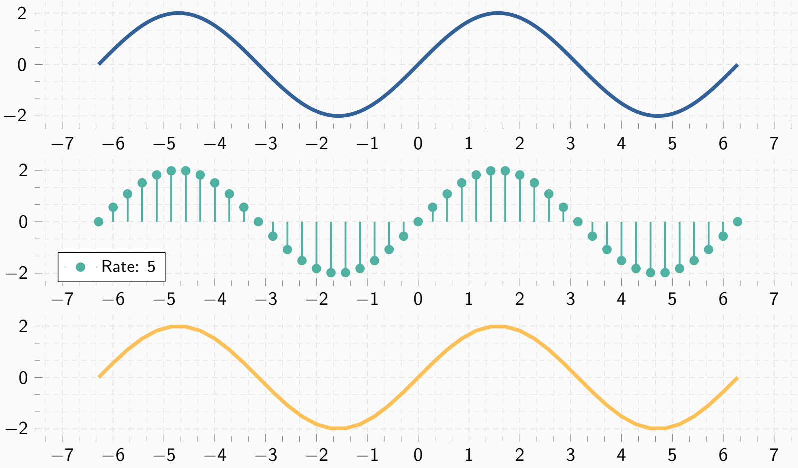

Figure 7: Reconstruction of the signal with 5 times the signal frequency.

Reconstruction of an Audio Signal

- In practice, doubling frequency is not enough recreate the signal.
- Approaching Nyquist frequency will create a siren like sound, and reaching exact frequency will record a pulse-wave approximation of a sine wave at an amplitude that will vary based on phase.
- Even at 4 times the sampling will only reconstruct a triangle wave and shifting the phase will create tonal distortion.
- For practical cases at least 6 times sampling rate is needed to accurately reconstruct the sine wave

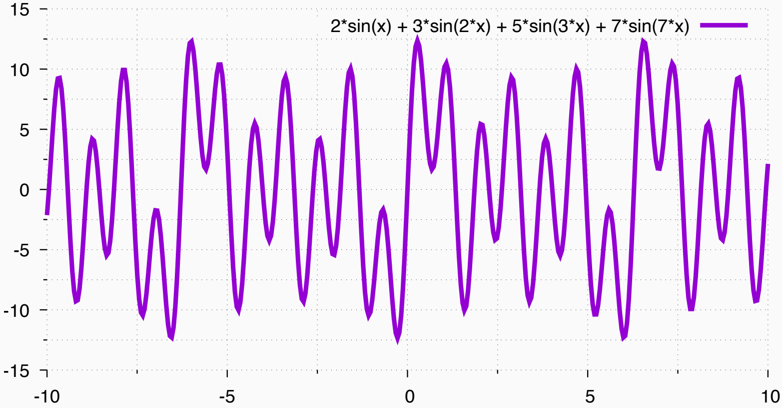

Figure 8: A sample signal with containing sample sine waves.

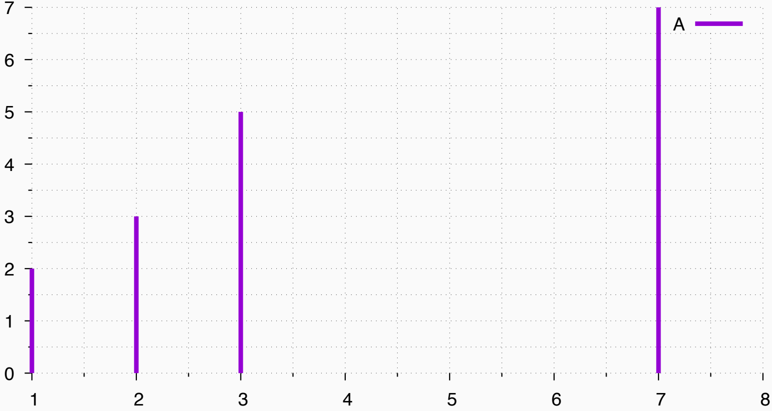

Figure 9: The FFT of the previous complex signal.

Mathematical Fundamentals

- Two (2) key problems arise when conducting spectral analysis of finite, discrete time series (not an infinite time series):

Aliasing we only resolve frequencies lower than the Nyquist frequency and frequencies higher than this get aliased to lower frequencies.

Spectral Leakage we assume that all waveforms stop and start at $= 0$ and $=$, but in the real world, many of these wave numbers may not complete a full integer number of cycles throughout the domain, causing spectral leakage to other wave numbers.

Aliasing

- If the initial samples are not sufficiently closely spaced to represent high-frequency components present in the underlying function, then the DFT values will be corrupted by aliasing.
- The solution is either to increase the sampling rate (if possible) or to pre-filter the signal in order to minimise its high-frequency spectral content.

Leakage

- The continuous Fourier transform of a periodic waveform requires the integration to be performed over the interval $-\infty$ to $+\infty$ or over an integer number of cycles of the waveform.
- If we attempt to complete the DFT over a non-integer number of cycles of the input signal, then we might expect the transform to be corrupted in some way.

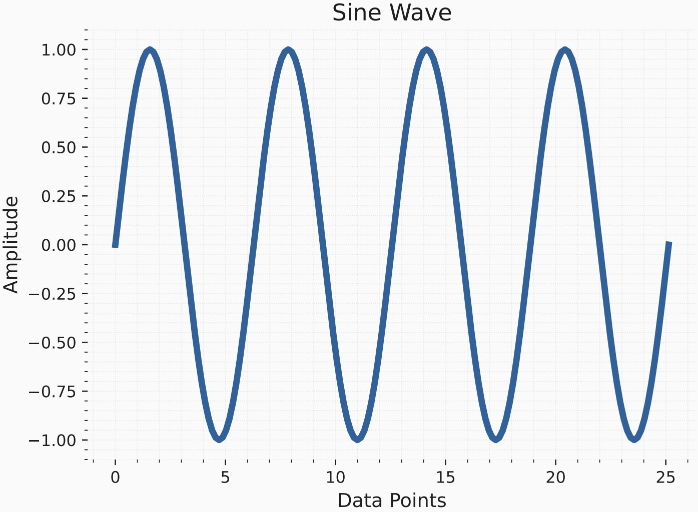

Figure 10: An example of a sine wave with four (4) complete cycles.

- Computing the discrete power spectrum gives:

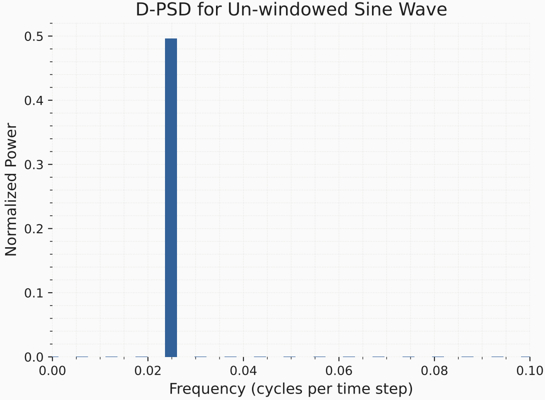

Figure 11: The PSD of an un-windowed sine wave.

- As expected, a single spectral peak corresponding to the frequency of our sine wave.
- Let's see what happens if we apply a window to our sine wave that cuts off the sine wave such that the sine function does not complete an integer number of cycles within the time domain.

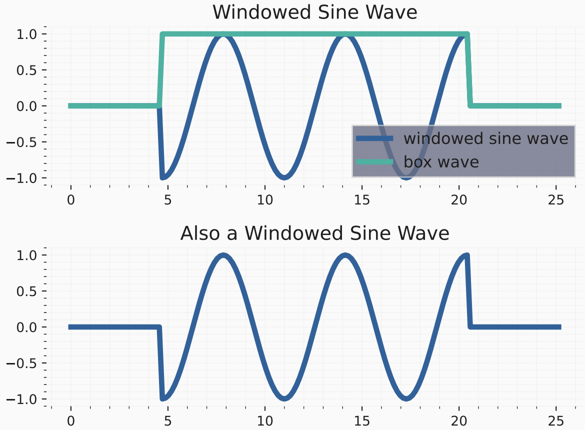

Figure 12: Windowed sine wave.

- To demonstrate what spectral leakage is, we will now compute the discrete power spectrum of the windowed sine wave to see what happens.

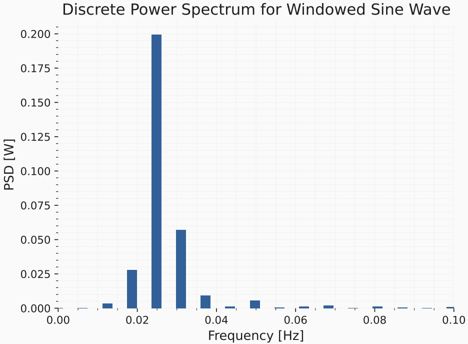

Figure 13: PSD of a windowed sine wave.

Example

Below is a signal with 1 Hz, Amplitude of 1 and 8 Sampling points.

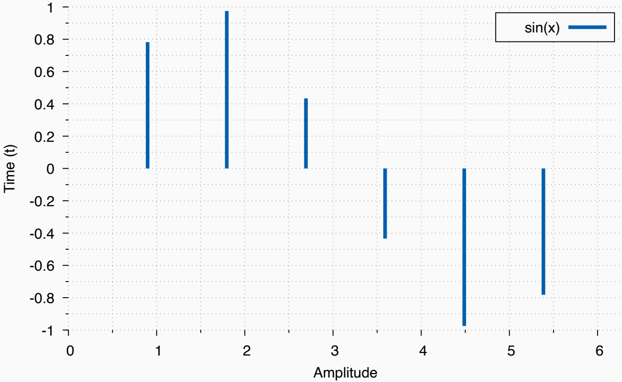

Figure 14: A Sampled Sine wave.

Solution

As it is a single sine function with 1 Hz, we expect a single value of 1 in the frequency domain (at 1 Hz).

The sampling points will sample the signal and retrieve the following data points as shown in the array below:

$$x_k = \left[0, 0.707, 1, 0.707, 0, -0.707, -1, -0.707\right]$$

Once we have these sampling points (x_n), we can turn our attention to the DFT formula:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-(\mathbf{j}\, 2\pi k n)/N}$$

where X_k is the k^{th} frequency bin.

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_1:

$$X_1 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\, 2\pi\, (1)\, n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\, 2\pi)/\ N} \\ e^{-(\mathbf{j}\, 4\pi)/\ N} \\ e^{-(\mathbf{j}\, 6\pi)/\ N} \\ e^{-(\mathbf{j}\, 8\pi)/\ N} \\ e^{-(\mathbf{j}\, 10\pi)/\ N} \\ e^{-(\mathbf{j}\, 12\pi)/\ N} \\ e^{-(\mathbf{j}\, 14\pi)/\ N} \end{bmatrix} = 0 - \mathbf{j}\, 4 \quad \blacksquare$$

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_2:

$$X_2 = \sum_{n=0}^{7} x_n \cdot e^{-(j\,2\pi\,(2)\,n)/\;N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(j\,4\pi)/\;N} \\ e^{-(j\,8\pi)/\;N} \\ e^{-(j\,12\pi)/\;N} \\ e^{-(j\,16\pi)/\;N} \\ e^{-(j\,20\pi)/\;N} \\ e^{-(j\,24\pi)/\;N} \\ e^{-(j\,28\pi)/\;N} \end{bmatrix} = 0 \quad \blacksquare$$

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_3:

$$X_3 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\,2\pi\,(3)\,n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,6\pi)/\ N} \\ e^{-(\mathbf{j}\,12\pi)/\ N} \\ e^{-(\mathbf{j}\,18\pi)/\ N} \\ e^{-(\mathbf{j}\,24\pi)/\ N} \\ e^{-(\mathbf{j}\,30\pi)/\ N} \\ e^{-(\mathbf{j}\,36\pi)/\ N} \\ e^{-(\mathbf{j}\,42\pi)/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

Mathematical Fundamentals

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_4:

$$X_4 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\,2\pi\,(4)\,n)/\,N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,8\pi)/\,N} \\ e^{-(\mathbf{j}\,16\pi)/\,N} \\ e^{-(\mathbf{j}\,24\pi)/\,N} \\ e^{-(\mathbf{j}\,32\pi)/\,N} \\ e^{-(\mathbf{j}\,40\pi)/\,N} \\ e^{-(\mathbf{j}\,48\pi)/\,N} \\ e^{-(\mathbf{j}\,56\pi)/\,N} \end{bmatrix} = 0 \quad \blacksquare$$

Mathematical Fundamentals

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_5:

$$X_5 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\,2\pi\,(5)\,n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,10\pi)/\ N} \\ e^{-(\mathbf{j}\,20\pi)/\ N} \\ e^{-(\mathbf{j}\,30\pi)/\ N} \\ e^{-(\mathbf{j}\,40\pi)/\ N} \\ e^{-(\mathbf{j}\,50\pi)/\ N} \\ e^{-(\mathbf{j}\,60\pi)/\ N} \\ e^{-(\mathbf{j}\,70\pi)/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_6:

$$X_6 = \sum_{n=0}^{7} x_n \cdot e^{-(\mathbf{j}\, 2\pi\,(6)\, n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\, 12\pi)/\ N} \\ e^{-(\mathbf{j}\, 24\pi)/\ N} \\ e^{-(\mathbf{j}\, 36\pi)/\ N} \\ e^{-(\mathbf{j}\, 48\pi)/\ N} \\ e^{-(\mathbf{j}\, 60\pi)/\ N} \\ e^{-(\mathbf{j}\, 72\pi)/\ N} \\ e^{-(\mathbf{j}\, 84\pi)/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

For the case of $x_0 = 0$ the exponential part is removed and we are left with $X_0 = 0$.
For the cases of X_7:

$$X_7 = \sum_{n=0}^{7} x_n \cdot e^{-(j\,2\pi\,(7)\,n)/\ N}$$

$$= x \cdot \begin{bmatrix} 0 \\ e^{-(j\,14\pi)/\ N} \\ e^{-(j\,28\pi)/\ N} \\ e^{-(j\,42\pi)/\ N} \\ e^{-(j\,56\pi)/\ N} \\ e^{-(j\,70\pi)/\ N} \\ e^{-(j\,84\pi)/\ N} \\ e^{-(j\,98\pi)/\ N} \end{bmatrix} = 0 + j\,4 \quad \blacksquare$$

- Therefore the values are of the transform are:

$$X_k = \left[0, 0 - \mathbf{j}\,4, 0, 0, 0, 0, 0, 0 + \mathbf{j}\,4\right]$$

- We can see only the first and the seventh bins have values other than zero.
- Calculating the magnitudes of the bins, we arrive at 4.

$$|X_k| = \left[0, 4, 0, 0, 0, 0, 0, 4\right]$$

- The frequency resolution of the plot is the sampling frequency divided by the number of samples (i.e., f S/N).
- This means we can get values for every integer frequency values.

Solution

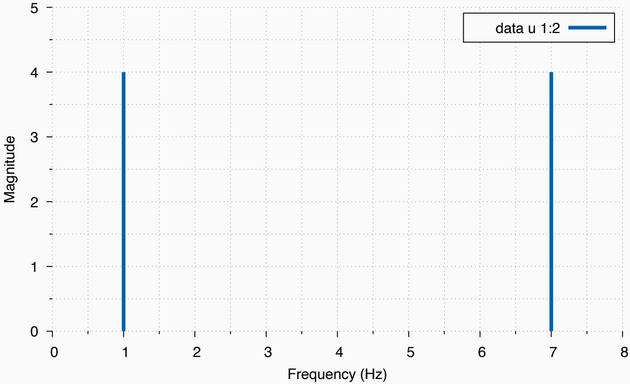

Figure 15: Sampled dataset of the original signal. There is still another step.

Solution

- We can see we get a value for the first frequency bin (1 Hz) and it makes sense.
- The reason we get a frequency bin is due to the plot being a two-sided frequency plot where it shows the energy in both the positive and negative domains.
- The negative frequencies are always complex conjugate to the positive frequencies, so there is no additional information in the negative frequencies.

Solution

- Therefore, to convert from a two-sided spectrum to a single-sided spectrum, discard the second half of the array and multiply every point except for DC by two.
- The last operation is to divide the magnitudes of the lower frequencies by the number of samples used in deriving these bins

$$|X_\mathrm{k}| = \left[0, 8, 0, 0\right] \rightarrow |X_\mathrm{k}| \, / N = \left[0, 1, 0, 0\right] \quad \blacksquare$$

Parseval's Theorem

- The sum (or integral) of the square of a function is equal to the sum (or integral) of the square of its transform.
- For continuous signals:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 \, \mathrm{d}\omega = \int_{-\infty}^{\infty} |F(2\pi f)|^2$$

- This signal energy is not to be confused with physical energy.

Average Value (μ)

- Defined as the sample mean of a given region.
- The equation is defined as below (it is also known as expected value):

$$\mu = \frac{1}{N}\sum_{i=0}^{N} x_{\mathrm{N}}$$

Standard Deviation (σ)

- The standard deviation is a measure of the amount of variation of the values of a variable about its mean (μ):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Mode

- The mode is the value appears most often in a set of data values. i.e., in an data pool of:

$$X = [1, 2, 2, 3, 4, 7, 9]$$

- The mode of is 2 as it is the most frequent value of the data set.
- whereas in:

$$X = [2, 4, 9, 6, 4, 6, 6, 2, 8, 2]$$

the mode is 2, 6 as there are two values with same frequency.

Median

- The median is the middle value separating the greater and lesser halves of the data set.
- For a ordered data set X with n elements,
  - if n is odd, med (x) = x (n+1)/2,
  - if n is even, med (x) = x (n/2) + x ((n/2)+1)/2
- i.e., in a ordered data set of:

$$X = [1, 2, 2, 3, 4, 7, 9]\,,$$

- the median is 3 and in:

$$Y = [1, 2, 3, 4, 5, 6, 8, 9]\,,$$

  the median is 4.5.

Mathematical Fundamentals

- The signal-to-noise ratio (SNR) can have several definitions depending on the study field.
- Noise is characterised by its standard deviation, σ.
- The characterisation of the signal can differ.
- If the signal is known to lie between two boundaries, $a_{\text{min}} \leq a \leq a_{\text{max}}$, then the SNR is defined as:

$$\text{SNR} = 20 \log_{10} \left(\frac{a_{\text{max}} - a_{\text{min}}}{s_{\text{n}}} \right) \text{ dB.}$$

- If the signal is not bounded but has a statistical distribution then two other definitions are known:

$$\text{SNR} = 20 \log_{10} \left(\frac{\mu}{\sigma} \right) \text{ dB.}$$

Mathematical Fundamentals

- In 1948, Claude Shannon published a paper called A Mathematical Theory of Communication.
- This paper heralded a transformation in our understanding of information.
- Before Shannon's paper, information had been viewed as a kind of poorly defined ethereal concept.
- But after Shannon's paper, it became apparent that information is a well-defined and, above all, measurable quantity.

Mathematical Fundamentals

- Information theory defines definite, unbreachable limits on precisely how much information can be communicated between any two components of any system, whether this system is man-made or natural.
- The basic laws of information can be summarised as follows.

1. there is a definite upper limit, the channel capacity, to the amount of information that can be communicated through that channel,
2. this limit shrinks the amount of noise in the channel increases,
3. this limit can very nearly be reached by judicious packaging, or encoding, of data.

Bits are Not Binary Digits

- The word bit is derived from binary digit,
  - but a bit and a binary digit are fundamentally different types of quantities.
- A binary digit is the value of a binary variable, whereas a bit is an amount of information.

Mathematical Fundamentals

- Consider a coin which lands heads up 90% of the time:

$$p\left(x_{\mathrm{h}}\right) = 0.9.$$

- When this coin is flipped, we expect it to land heads up $(x = x_{\mathrm{h}})$,
- When it does, we are less surprised than when it lands tails $(x = x_{\mathrm{t}})$.

> The more improbable a particular outcome is, the more surprised we are to observe it.

- If we use $\log_2$ then the Shannon information or surprisal of each outcome is measured in bits.

$$\text{Shannon Information} = \log_2 \frac{1}{p\left(x_{\mathrm{h}}\right)}$$

Entropy is Average Shannon Information

- We can represent the outcome of a coin flip as the random variable x, such that a head is $x = x_{\mathrm{h}}$ and a tail is $x = x_{\mathrm{t}}$.
- In practice, we are not usually interested in the surprise of a particular value of a random variable, but we are interested in how much surprise, on average, is associated with the entire set of possible values.
- The average surprise of a variable x is defined by its probability distribution $p(x)$, and is called the entropy of $p(x)$, represented as $H(x)$.

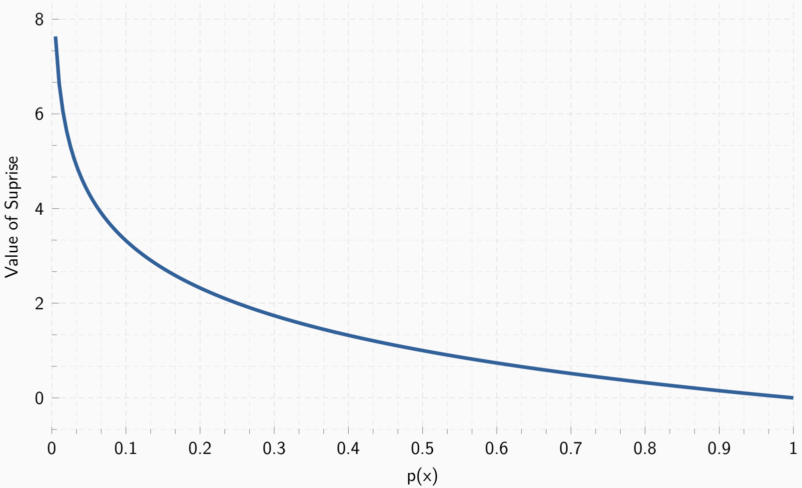

Figure 16: The quantifiable surprise with respect to increasing probability.

Entropy of a Fair Coin

- If a coin is fair or unbiased then:

$$p_{x_h} = p_{x_t} = 0.5$$

- the Shannon information gained when a head or a tail is observed is:

$$\log 1/0.5 = 1 \text{ bit}$$

- The average Shannon information gained after each coin flip is also 1 bit.
- Because entropy is defined as average Shannon information, the entropy of a fair coin is $H(x) = 1$ bit.

Entropy of an Unfair Coin

- If a coin is biased such that the probability of a head is $p(x_{\mathrm{h}}) = 0.9$.
  - it is easy to predict the result of each coin flip (i.e. with 90% accuracy if we predict a head for each flip)
- If the outcome is a head then the amount of Shannon information gained is $\log(1/0.9) = 0.15$ bits.
- But if the outcome is a tail then the amount of Shannon information gained is $\log(1/0.1) = 3.32$ bits.
- Notice that more information is associated with the more surprising outcome.

Entropy of an Unfair Coin

- Given that the proportion of flips that yield a head is p(xh), and that the proportion of flips that yield a tail is p(xt) (where p(xh) + p(xt) = 1), the average surprise is

$$H(x) = p(x_h) \log \frac{1}{p(x_h)} + p(x_t) \log \frac{1}{p(x_t)},$$

- Which comes to 0.469bits.
- If we define a tail as $x_1 = x_t$ and a head as $x_2 = x_h$ then the above equation is written as:

$$H(x) = \sum_{i=1}^{2} p(x_i) \log \frac{1}{p(x_i)} \text{ bits.}$$

Mathematical Fundamentals

- More generally, a random variable x with a probability distribution p(x) = p(x1), . . . , p(xm) has an entropy of

$$H(x) = \sum_{i=1}^{m} p(x_i) \log \frac{1}{p(x_i)} \text{ bits.}$$

Mathematical Fundamentals

- Entropy is a measure of uncertainty.
- When our uncertainty is reduced, we gain information,
  - so information and entropy are two sides of the same coin.
- However, information has a rather subtle interpretation, which can easily lead to confusion.
- Average information shares the same definition as entropy,
  - but whether we call a given quantity information or entropy depends on whether it is being given to us or taken away.
- For example, if a variable has high entropy the initial uncertainty of the variable is large and is, by definition, exactly equal to its entropy.
- If we are told the variable value, on average, we have been given information equal to the uncertainty (entropy) we had about its value.
- Thus, receiving an amount of information is equivalent to having exactly the same amount of entropy (uncertainty) taken away.

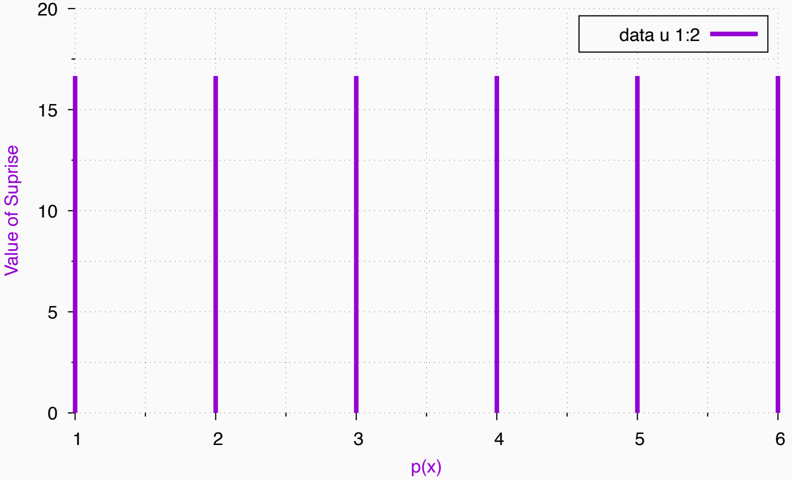

Figure 17: The probability distribution of 1 dice(s).

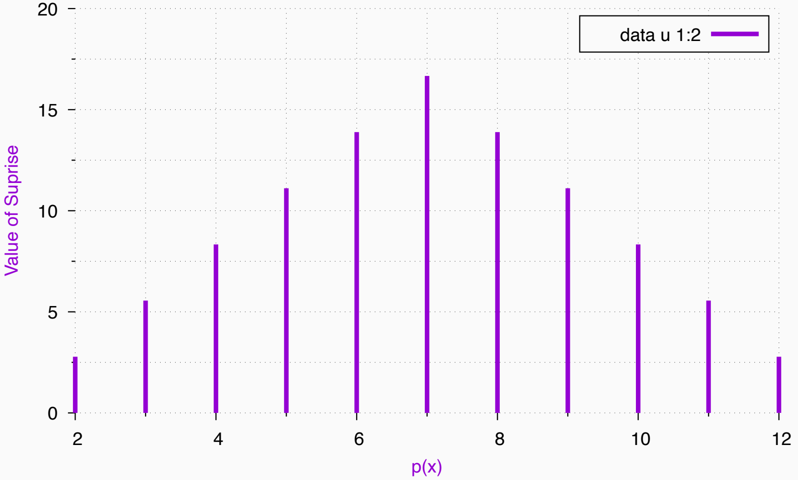

Figure 18: The probability distribution of 2 dice(s).

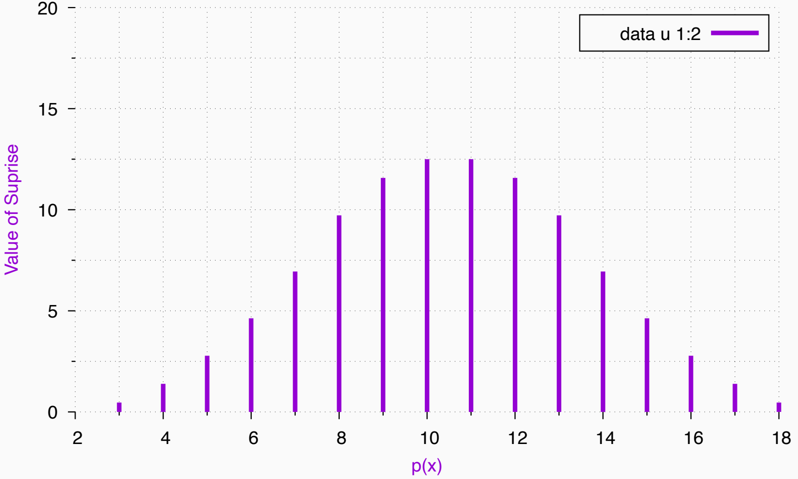

Figure 19: The probability distribution of 3 dice(s).

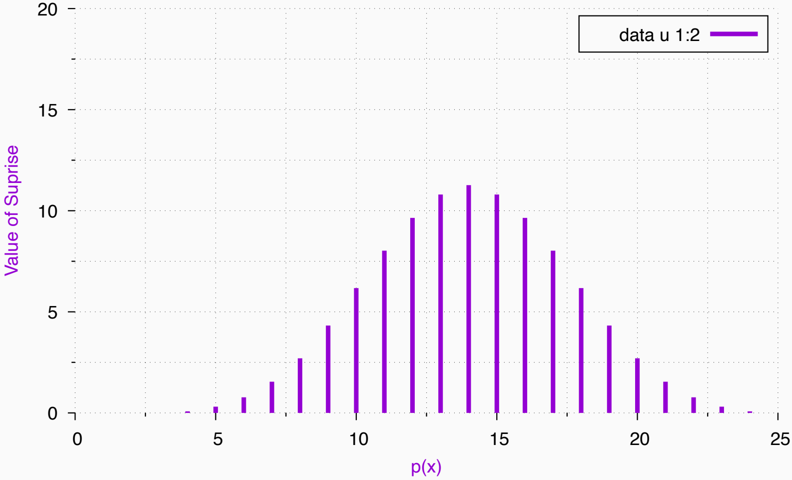

Figure 20: The probability distribution of 4 dice(s).

- Throwing a pair of 6-sided dice produces an outcome in the form of an ordered pair of numbers.
  - There are a total of 36 equiprobable outcomes,
- If we define an outcome value as the sum of this pair of numbers then there are $m = 11$ possible outcome values:

$$A_x = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}\,.$$

Dividing the frequency of each outcome value by 36 yields the probability p of each outcome value.

Mathematical Fundamentals

- We can use these 11 probabilities to find the entropy.

$$
\begin{aligned}
H(x) &= p(x_1)\log\frac{1}{p(x_1)} + p(x_1)\log\frac{1}{p(x_1)} + \cdots + p(x_{11})\log\frac{1}{p(x_{11})} \\
&= 3.27 \text{ bits.}
\end{aligned}
$$

Perception

Table of Contents

Learning Outcomes

(LO1) A Look into Human Vision,

(LO2) Definition of Colour and Standardisation,

(LO3) How Vision is Perceived,

(LO4) Types of Colour-spaces.

Perception

- Many image processing applications are intended to produce images to be viewed by humans.
  - This is in contrast to automated industrial robots.
- It is important to understand the characteristics and limitations of the human visual system [5].
- At the outset it is important to realise:
  1. The human visual system is not well understood [5].
     - It is not easy to study the human visual system without directly measuring it.
  2. No objective measure exists for judging the quality of an image that corresponds to human assessment of image quality,
     - A colour you find fitting might be repugnant to someone.
  3. A typical human observer does not exist.

Perception

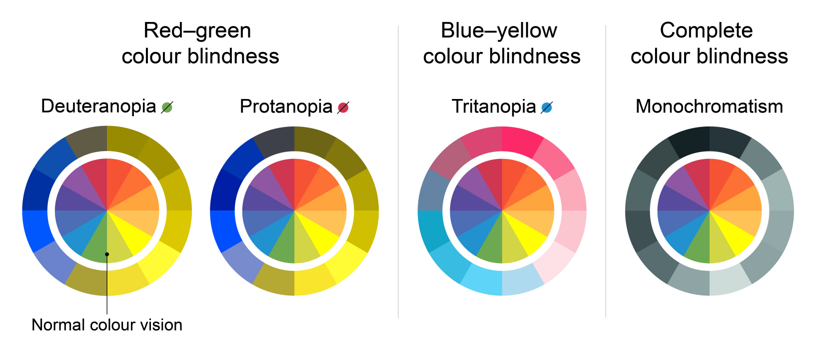

Figure 21: Colour-wheel of different kinds of blindness [10].

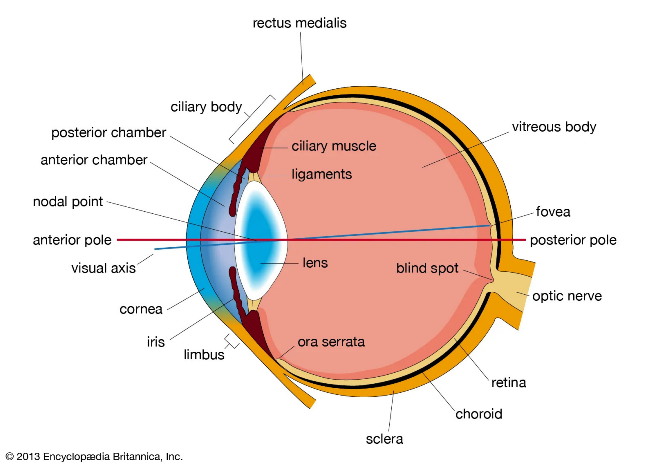

Figure 22: Horizontal section of the eye.

Trichromacy

- Normal colour vision uses all three (3) types of cone cells which function correctly.
- Another term for normal colour vision is trichromacy.
- People with normal colour vision are known as trichromats.

Anomalous Trichromacy

- People with faulty trichromatic vision will be colour blind to some extent and are known as anomalous trichromats [4].
- In people with this condition all of their three cone cell types are used to perceive light wavelengths but one type of cone cell perceives light slightly out of alignment.
- There are three (2) different types of effect produced depending upon which cone cell type is faulty and there are also different severities.

- The different anomalous condition types are [12]:

  protanomaly reduced sensitivity to red light,
  deuteranomaly reduced sensitivity to green light (most common),
  tritanomaly reduced sensitivity to blue light (most uncommon).

Achromatopsia

- Can see no colour at all and their world consists of different shades of grey ranging from black to white, rather like seeing the world on an old black and white television set [13].
- Achromatopsia is a specific eye condition in which people see in greyscale.
- In rare cases, partial Achromatopsia can happen which is a reduced sensitivity to all three (3) cones [13].

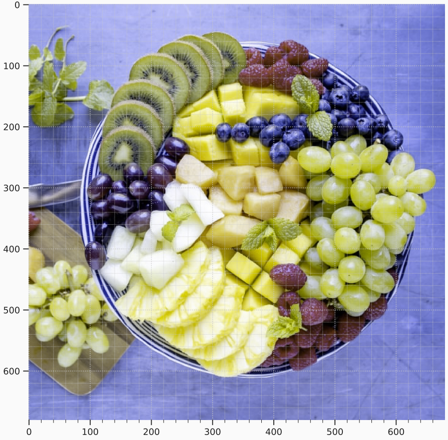

Figure 23: Image viewed by someone who has protanomaly.

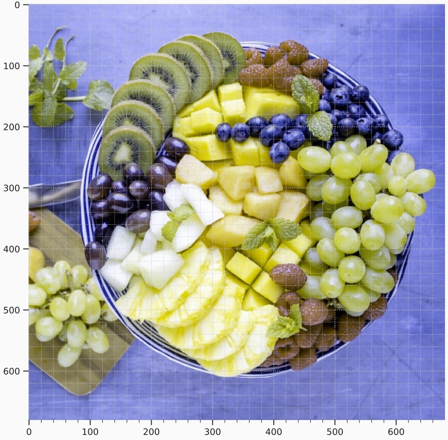

Figure 24: Image viewed by someone who has deuteranomaly.

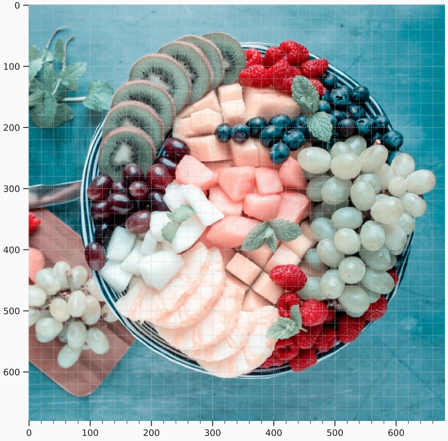

Figure 25: Image viewed by someone who has tritanomaly.

Perception

Figure 26: A colour blindness test issued to generally test before taking the driving license [3].

- There are ways to describe the sensitivity of human vision.
- Assume a homogeneous region in an image has an intensity as a function of wavelength (colour) given by $I(\lambda)$ and assume that $I(\lambda) = I_0$, a constant.

Wavelength Sensitivity

- The sensitivity of the human eye to light of a certain intensity varies strongly over wavelengths between 380 nm and 800 nm [9].
- Under daylight conditions, human eye is most sensitive at a wavelength of 555 nm, resulting in the fact that green light at produces the impression of highest “brightness” when compared to light at other wavelengths [9].

- The perceived intensity as a function of λ, the spectral sensitivity, for the typical observer is shown below.

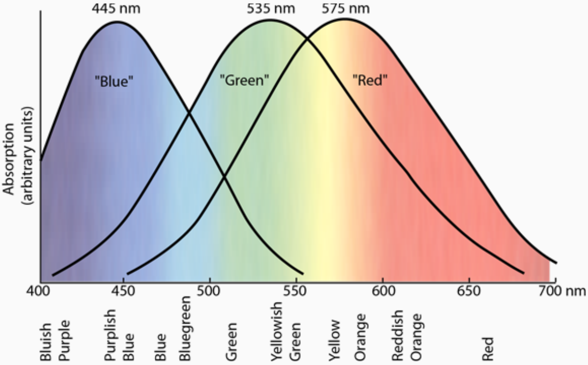

Figure 27: The colour sensitivity of the human eye [11].

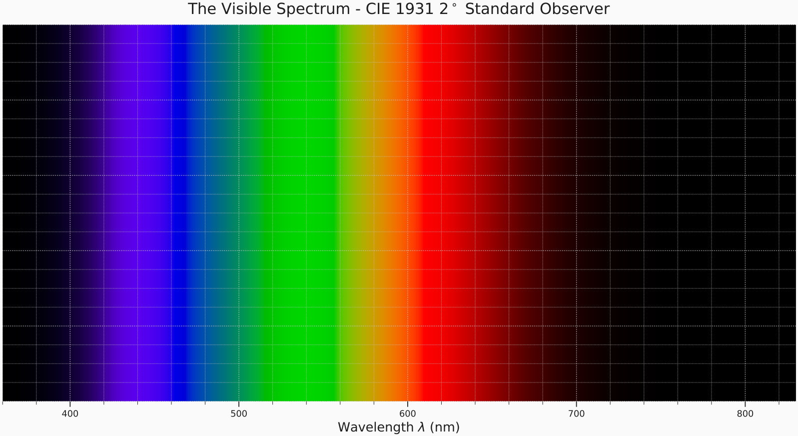

Figure 28: The visible colour spectrum visible with the human eye.

Perception

- If the constant intensity (i.e., brightness) I_0 is allowed to vary, then, to a good approximation, the visual response, R, is proportional to the logarithm of the intensity.
- This is known as the Weber-Fechner law [6].

> The law relates to human perception, more specifically the relation between the actual change in a physical stimulus and the perceived change.

$$R = \log\left(I_0\right)$$

- This means, equal perceived steps in brightness, $\Delta R = k$, require the physical brightness (i.e., the stimulus) to increase exponentially.

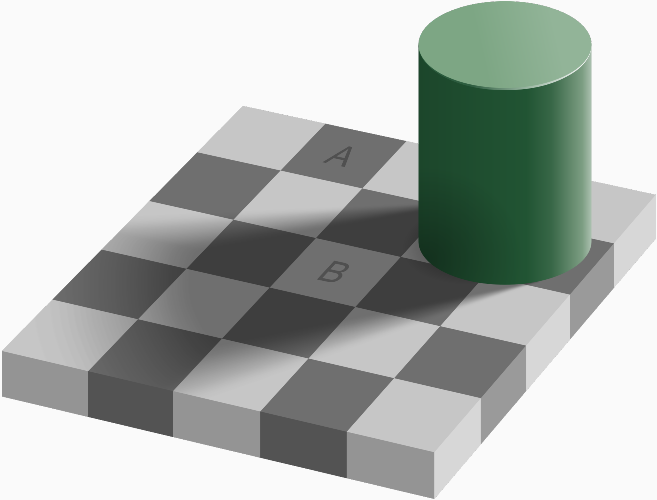

Figure 29: The checker shadow illusion [1].

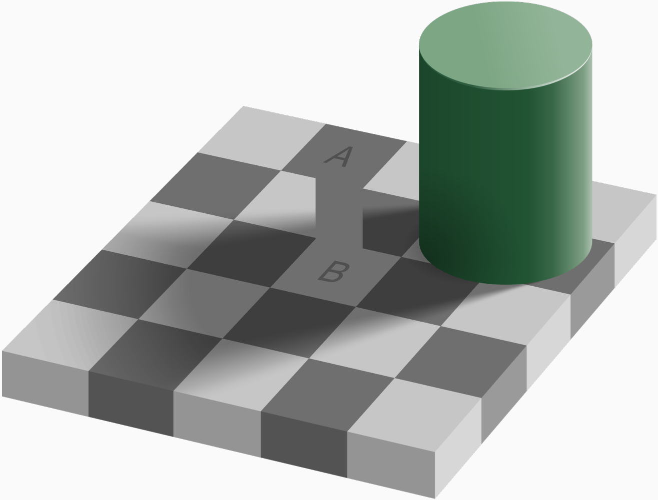

Figure 30: A region of the same shade has been drawn connecting A and B.

- Human colour perception is complex,
  - therefore we can only present a brief introduction.
- **Standard Observer:** Based on psychophysical measurements, standard curves have been adopted by the CIE [1] as sensitivity curves for the typical observer for three pigments $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$.

These are not the actual pigment absorption characteristics found in the standard human retina but rather sensitivity curves derived from actual data.

This standard is used by companies to produce monitors and software that are compatible with each other.

[1] Commission Internationale de l'Eclairage

Perception

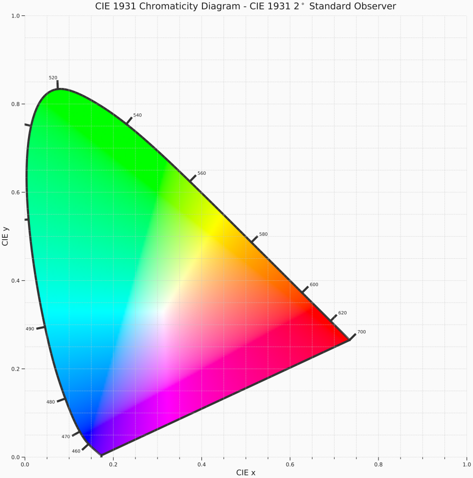

Figure 31: The colour gamut visible to the human eye, standardised by the CIE.

- sRGB is a standard RGB colour-space HP and Microsoft created cooperatively in 1996 to use on monitors, printers, and the World Wide Web [18].
- It was subsequently standardised by IEC as IEC 61966-2-1:1999 [14].
- sRGB is the current defined standard colour-space for the web, and it is usually the assumed colour-space for images that are neither tagged for a colour-space nor have an embedded color profile.
- It codifies the display specifications for the computer monitors in use at the time, which greatly aided its acceptance.
- sRGB uses the same colour primaries and white point as ITU-R BT.709 standard for HDTV, designed to match typical home and office viewing conditions.

Figure 32: The sRGB colour-space superimposed to the CIE colour-gamut.

Perception

- Due to the standardisation of sRGB on the digital-space, and on printers, many low- to medium-end consumer digital cameras and scanners use sRGB as the default working colour-space [19].
- However, consumer-level CCDs[1] are typically uncalibrated, meaning that even though the image is being labeled as sRGB, one can not conclude that the image is color-accurate sRGB.

[1] Charge-Coupled Device

Perception

- The wide-gamut RGB colour-space (Adobe Wide Gamut RGB) is developed by Adobe, which offers a large gamut by using pure spectral primary colours [16].
- It is able to store a wider range of colour than sRGB or Adobe RGB.

> For comparison, the wide-gamut RGB colour-space encompasses 77.6% of the visible colours, while Adobe RGB covers 52.1% and sRGB only 35.9% [21].

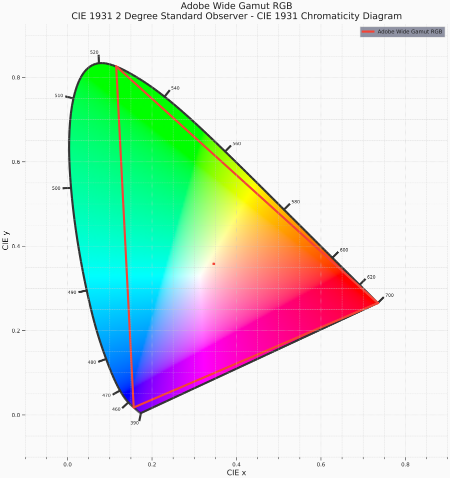

Figure 33: The wide gamut RGB colour-space superimposed to the CIE colour-gamut.

Perception

- The ProPhoto RGB colour space, a.k.a. ROMM RGB (Reference Output Medium Metric), is an output referred RGB color space developed by Kodak [17].
- Offers an especially large gamut designed for use with photographic output in mind.
- The gamut encompasses over 90% of possible color space, and 100% of likely occurring real-world surface colours making ProPhoto even larger than the Wide-gamut RGB color space [24].
- The ProPhoto RGB primaries were also chosen in order to minimise hue rotations associated with non-linear tone scale operations.

> One of the downsides is that approximately 13% of the representable colours are imaginary colors that do not exist and are impossible colour.

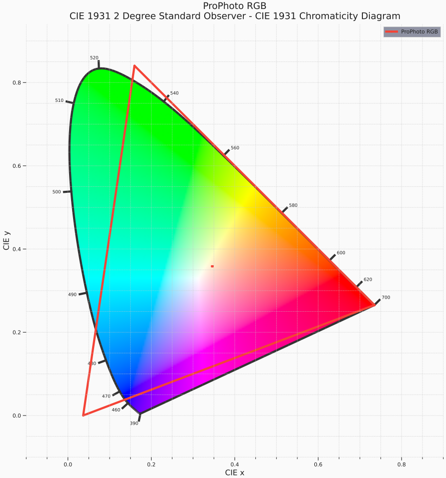

Figure 34: The ProPhoto colour-space superimposed to the CIE colour-gamut.

Perception

- The **Adobe RGB (1998)** or **opRGB** is a color space developed by Adobe Inc. in 1998.
- It was designed to encompass most of the colors achievable on CMYK color printers, but by using RGB primary colors on a device such as a computer display.
- The Adobe RGB (1998) color space encompasses roughly 30% of the visible colors specified by the CIE – improving upon the gamut of the sRGB color space, primarily in cyan-green hues.
- It was subsequently standardised by the IEC as IEC 61966-2-5:1999 with a name opRGB (optional RGB color space) and is used in HDMI [15].

- For an arbitrary homogeneous region in an image that has an intensity as a function of wavelength (colour) given by $I(\lambda)$, the three responses are called the tristimulus values:

$$A = \int_{-\infty}^{+\infty} I(\lambda)\ \overline{a}(\lambda)\ d\lambda \quad \text{where} \quad A = \{X,\ Y,\ Z\}\ ,\ a = \{x,\ y,\ z\}\ .$$

- The chromaticity coordinates which describe the perceived colour information are defined as:

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{y}{X+Y+Z}, \quad z = 1-(x+y)\ .$$

- The tristimulus values are linear in $I(\lambda)$ and thus the absolute intensity information has been lost in the calculation of the chromaticity coordinates $\{x,\ y\}$.
- All colour distributions, $I(\lambda)$, that appear to an observer as having the same colour will have the same chromaticity coordinates.

- The formulas for converting from the tristimulus values (X, Y, Z) to **RGB** colours (R, G, B) and back are given by:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.19107 & -0.5326 & -0.2883 \\ -0.9843 & 1.9984 & -0.0283 \\ 0.0583 & -0.1185 & 0.8986 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

and:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.6067 & 0.1736 & 0.2001 \\ 0.2988 & 0.5868 & 0.1143 \\ 0.0000 & 0.0661 & 1.1149 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- Before we end our look on to these colour spaces, lets have a look at two (2) more standards.

Perception

- The CMYK model is a subtractive model used in colour printing, and describing the printing process itself.
- The abbreviation CMYK refers to the four inks used:

  cyan, magenta, yellow, and key (black).
- Works by partially or entirely masking colours on a lighter, usually white, background.
- The ink limits the reflected light.
- Such a model is called subtractive because inks subtract the colours red, green and blue from white light.
- White light minus red leaves cyan, white light minus green leaves magenta, and white light minus blue leaves yellow.

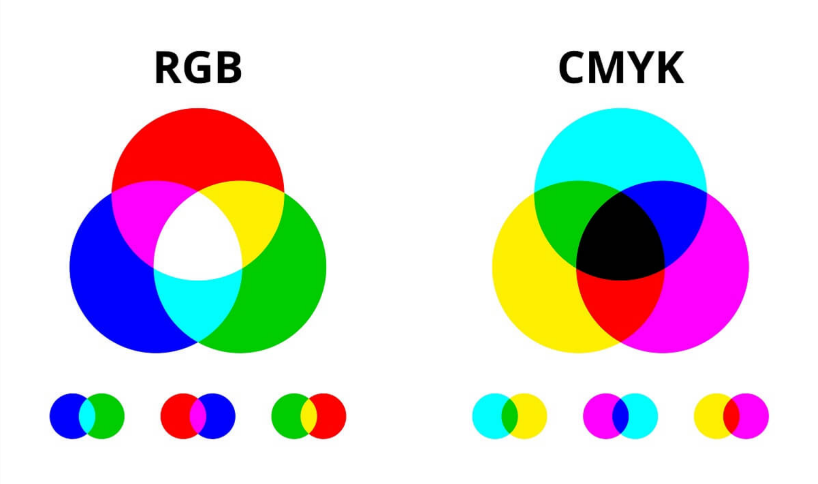

Figure 35: The differences between RGB and CYMK colours [20].

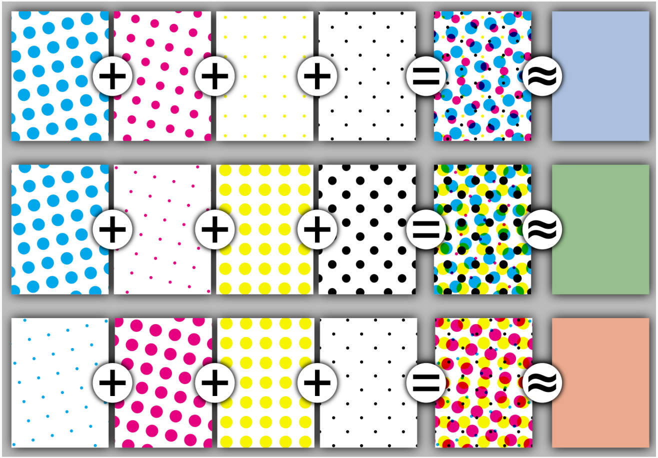

Figure 36: Three examples of color halftoning with CMYK separations, as well as the combined halftone pattern and how the human eye would observe the combined halftone pattern from a sufficient distance [22].

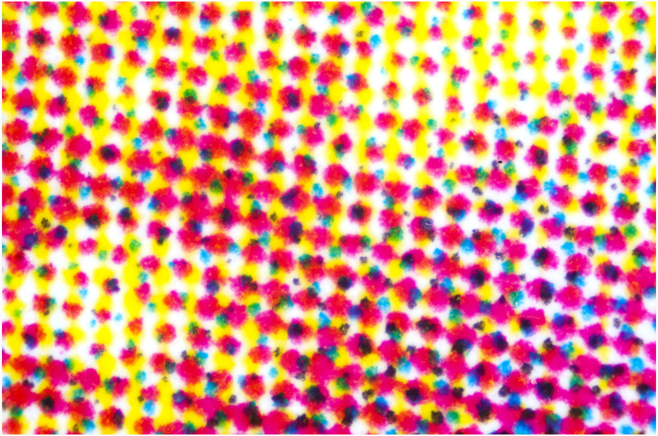

Figure 37: A printer creates any colour by combining dots in particular places relative to the other dots.

Perception

- Two most common cylindrical-coordinate representations of points in an RGB color model.
- The two representations rearrange the geometry of RGB in an attempt to be more intuitive and perceptually relevant than the cartesian (cube) representation.
- Developed in the 1970s for computer graphics applications, are used in color pickers, in image editing software, and less commonly in image analysis and computer vision.

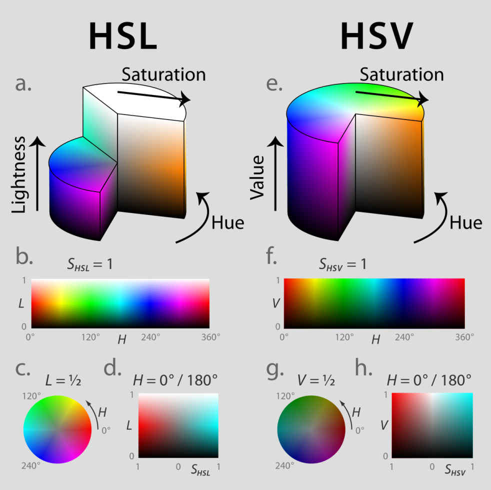

Figure 38: HSL and HSV models.

Perception

- $\mathbf{YC_bC_r}$ is a family of colour spaces used as a part of the color image pipeline in video and digital photography systems.
- **Y** is the luma (i.e., brightness) component and CB and CR are the blue-difference and red-difference chroma components.
- Y (with prime) is distinguished from Y, which is luminance, meaning that light intensity is nonlinearly encoded based on gamma corrected RGB primaries.

Perception

- CRT[1] uses RGB signals, but they are not the best solution for storing information as they have a lot of redundancy.
- **YC_bC_r** is a practical approximation, where the primary colours corresponding roughly to red, green and blue are processed into perceptually meaningful information.
- **$Y'C_bC_r$** is used to separate out a luma signal (Y) that can be stored with high resolution or transmitted at high bandwidth, and two chroma components (CB and CR) that can be bandwidth-reduced, subsampled, compressed, or otherwise treated separately for improved system efficiency.

> One practical example would be decreasing the bandwidth or resolution allocated to "color" compared to "black and white", since humans are more sensitive to the black-and-white information (see image example to the right). This is called chroma subsampling.

[1]Cathode Ray Tube

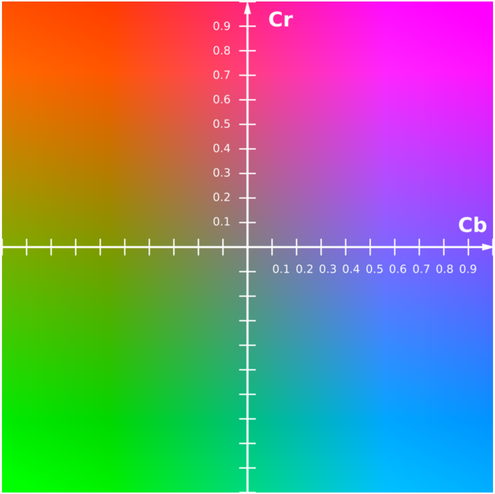

Figure 39: The $\mathbf{Y'C_bC_r}$ plane at constant luma Y =0.5 [7].

Image Formats

Table of Contents

Learning Outcomes

(LO1) Types of File Formats used,

(LO2) Defining Compression and its Types,

(LO3) A Look into Selective File Types.

(LO4) A Deep Dive into JPEG Compression.

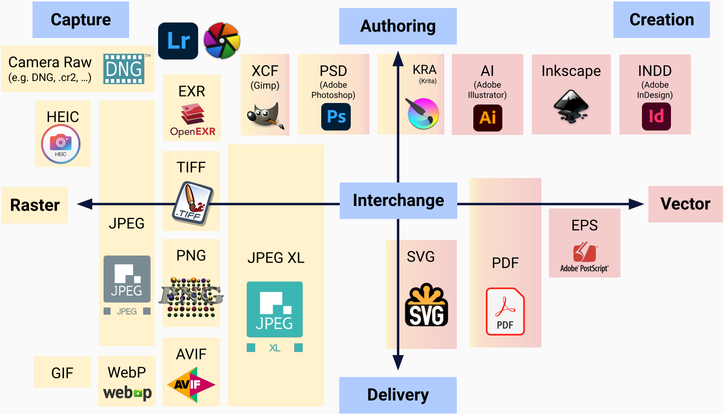

Figure 40: Categorization of image formats by scope which are used in commercial, industrial and personal use [23].

Image Formats

TIFF(`.tif`, `.tiff`) Tagged Image File Format

- Store image data without losing any data.
  - It does not perform any compression on images,
  - High-quality image is obtained however the image size is large,
- Good for printing, and professional work.

JPEG (`.jpg`, `.jpeg`) Joint Photographic Experts Group

- It is a loss-prone (lossy) format.
  - Some data is lost to reduce the image size.
  - Due to compression, some data is lost but that loss is very less
- Used in digital cameras, non-professional prints, e-Mail, etc.

Image Formats

GIF (`.gif`) GIF or Graphics Interchange Format

- Primarily used in web graphics (i.e., Reddit, WhatsApp.)
  - They can be animated and are limited to only 256 colours.
  - These images can also allow transparency.
- GIF files are typically small in size compared to other formats.

PNG (`.png`) PNG or Portable Network Graphics

- It is a lossless image format.
  - It was designed to replace GIF as it supported 256 colours.
  - PNG, whereas, supports 16 million colours.

Image Formats

Bitmap (`.bmp`) Bit Map Image

- It was developed by Microsoft for windows.
- It is same as TIFF due to lossless, no compression property.
- As it is a proprietary format, it is generally recommended to use TIFF.

EPS (`.eps`) Encapsulated PostScript

- It is a vector file.
- EPS files can be opened in applications such as Adobe Illustrator.

Image Formats

RAW Image Files (`.raw`, `.cr2`, `.nef`, `.orf`, `.sr2`)

- Unprocessed and created by a camera or scanner.
- Many DSLR cameras can shoot in RAW.
- These images are the equivalent of a digital negative, meaning that they hold a lot of image information.
- It saves metadata and is used for photography.

Image Formats

- The class of data compression methods using inexact approximations and partial data discarding to represent the content.
- These techniques are used to reduce data size for storing, handling, and transmitting content.
- It can also remove metadata to save up on space [2].
- An example would be JPEG compression.

> Data is permanently removed from the file.

> Well-designed lossy compression technology often reduces file sizes significantly before degradation is noticed by the end-user.

Image Formats

- A camera RAW image file contains unprocessed or minimally processed data from the image sensor of either a digital camera, a motion picture film scanner, or other image scanner.
- Named RAW as they are not yet processed, and contain large amounts of potentially redundant data.
- Normally, the image is processed by a RAW converter, in a wide-gamut internal color space where precise adjustments can be made before conversion to a viewable format.
- There are dozens of RAW formats in use by different manufacturers of digital image capture equipment.
  - i.e., Nikon uses `.nef` format, and Canon uses `.cr2`.

Image Formats

- It is a lossy compression method.
- Usually stored in the `.jfif` (JPEG File Interchange Format) or the `.exif` (Exchangeable image file format) file format.
- The JPEG filename extension is `.jpg` or `.jpeg`.
- Nearly every camera can save images in the `.jpeg`, which supports eight-bit grayscale images and 24-bit color images
  - Eight bits each for `red`, `green`, `blue`.
- Compression can result in a significant reduction of the file size.
- Applications can determine the degree of compression to apply.
- When not too high, compression does not affect or detract from the image's quality, but JPEG files suffer generation loss when repeatedly edited and saved.

Figure 41: A photo of a European wildcat with the compression rate, and associated losses, decreasing from left to right [8].

Image Formats

- For good lossy compression, throw away the unnecessary information.
- In the case of image compression, Start by understanding which parts of an image are important to human perception, and which aren't.
- Then you find a way to keep the important qualities and trash the rest.
- JPEG, uses two (2) psychovisual principles to compress the image.

Image Formats

1. Changes in brightness are more important than changes in colour:

> Human retina contains about 120 million brightness-sensitive rod cells, but only about 6 million colour-sensitive cone cells.

2. Low-frequency changes are more important than high-frequency changes.

> The human eye is good at judging low-frequency light changes, like the edges of objects. It is less accurate at judging high-frequency light changes, like the fine detail in a busy pattern or texture. Camouflage works in part because higher-frequency patterns disrupt the lower-frequency edges of the thing camouflaged.

Image Formats

- Each pixel is stored as three numbers:
  - Representing red, green and blue.

> The problem, is that the image's brightness information is spread evenly across RGB.

> Remember, brightness is more important than colour.

- We want to isolate brightness from the colour information so we can deal with it separately.
- To do this, JPEG converts the image from RGB to YCbCr.
- In YCbCr all brightness information in one channel (Y) while splitting the colour information between the other two (Cb and Cr).

- Before doing processing, JPEG throws away some of the colour information by scaling down just the Cb and Cr (colour) channels while keeping the important Y (brightness) channel full size.

> This step is optional.

- You can keep all of the colour information, half of it, or a quarter.
- For images, most software will keep half of the colour information;
- for video it is usually a quarter and for this lets do quarter.
- We started with 3 full channels and now we have 1 full channel and 2 x 0.25 channels, for a total of 1.5.
- We are already down to half of the information we started with.

Image Formats

- To use the second observation about perception, start by dividing each of the Y, Cb, and Cr channels up into 8-by-8 blocks of pixels.
- Transform each of these blocks from the spatial domain to the frequency domain.
- Consider just one of these 8-by-8 blocks from the Y channel.
- The spatial domain is what we have now:
  - the value in the upper-left corner represents the brightness (Y-value) of the pixel in the upper-left corner of that block. Likewise, the value in the lower-right corner represents the brightness of the pixel in the lower-right corner of that block. Hence the term spatial: position in the block represents position in the image.

Image Formats

- When we transform this block to the frequency domain, position in the block will instead represent a frequency band in that block of the image. The value in the upper-left corner of the block will represent the lowest-frequency information and the value in the lower-right corner of the block will represent the highest-frequency information.

Image Formats

11	11	10	10	10	10	9	9
11	11	10	10	10	10	9	9
11	11	12	12	9	9	8	8
11	11	12	12	9	9	8	8
10	10	10	10	11	11	9	9
10	10	10	10	11	11	9	9
10	10	8	8	11	11	10	10
10	10	8	8	11	11	10	10

79.5	3.58	-1.39	1.89	0.0	-1.26	0.57	-0.71
0.64	3.91	-0.59	-2.43	0.0	1.62	0.25	-0.78
-0.46	-2.41	2.13	2.09	0.0	-1.4	-0.88	0.48
0.22	-1.61	-0.21	1.34	0.0	-0.9	0.09	0.32
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-0.15	1.08	0.14	-0.9	0.0	0.6	-0.06	-0.21
0.19	1.0	-0.88	-0.87	0.0	0.58	0.37	-0.2
-0.13	-0.78	0.12	0.48	0.0	-0.32	-0.05	0.15

Figure 42: On the left we have an input of 8x8 pixels, on the right we have DCT coefficients rounded to two decimal places.

Image Formats

- The next step is to selectively throw away some of the frequency information.
- If you have ever saved a JPEG image and chosen a quality value, this is where that choice comes into play.
- Start with two 8-by-8 tables of whole numbers, called the quantization tables.
- One table is for brightness information, and one is for colour information.
- You will use these numbers on each of the 8-by-8 blocks in the image data by dividing the frequency value in the image data by the corresponding number in its quantization table.

Image Formats

- So the upper-left corner of each 8×8 block in the Y frequency channel will be divided by the number in the upper-left corner of the brightness quantization table, and so on.
- The result of each division is rounded to the nearest whole number and the fractional parts are thrown away.

> The quantization tables are saved along with the image data in the JPEG file. They'll be needed to decode the image correctly.

Image Formats

- If you think carefully about what just happened, you will realize that even though we threw away some frequency information by tossing the decimal parts after division, we still have the same amount of data:
- one number for each pixel from each of the three channels.
- However, this data is now going to be compressed using traditional lossless compression.

Image Formats

- But wait, wasn't the whole reason we used lossy compression in the first place that lossless compression doesn't work well for images? Yes, but that quantization we just did is going to make the data more compressible by making it less noisy. To see why, compare these three number sequences:
- JPEG has one last trick for making the data more compressible: it lists the values for each 8×8 block in a zig-zag pattern that puts the numbers in order from lowest frequency to highest. That means that the most heavily quantized parts (with the largest divisors) are next to each other to make nice, repetitive patterns of small numbers.

Appendix

Go Back

- Univariate object is an expression, equation, function or polynomial involving only **one variable**.
- Objects involving more than one variable are multivariate.
  - In statistics, a univariate distribution characterizes one variable, although it can be applied in other ways as well.
    - i.e., in time series analysis, the whole time series is the "variable": a univariate time series is the series of values over time of a single quantity.

Go Back

In probability theory, the sample space (also called sample description space,[1] possibility space,[2] or outcome space[3]) of an experiment or random trial is the set of all possible outcomes or results of that experiment.[4] A sample space is usually denoted using set notation, and the possible ordered outcomes, or sample points,[5] are listed as elements in the set. It is common to refer to a sample space by the labels S, Ω, or U (for "universal set"). The elements of a sample space may be numbers, words, letters, or symbols. They can also be finite, countably infinite, or uncountably infinite.[6]

Go Back

Stands for red green blue alpha.

While it is sometimes described as a color space, it is actually a three-channel RGB color model supplemented with a fourth alpha channel.

Alpha indicates how opaque each pixel is and allows an image to be combined over others using alpha compositing, with transparent areas and anti-aliasing of the edges of opaque regions. Each pixel is a 4D vector.

Appendix

Go Back

There are a wide variety of data types used in `opencv`. The most common are:

Go Back

Aperture of an optical system is a hole or an opening that primarily limits light propagated through the system.

Figure 43: Different apertures of a lens.

Go Back

The Gaussian function $g\left(x\right)$, is defined as:

$$f\left(x\right) = \exp\left(-ax^2\right)$$

The Fourier transform $\left(\mathcal{F}_\mathrm{x}\right)$ if given by

$$\begin{aligned}
\mathcal{F}_\mathrm{x}\left(k\right) &= \int_{-\infty}^{\infty} \exp\left(-ax^2\right) \exp\left(-2\pi\mathbf{j}\,kx\right)\,dx,\\
&= \int_{-\infty}^{\infty} \exp\left(-ax^2\right) \left[\cos 2\pi kx - \mathbf{j}\,\sin 2\pi kx\right]\,dx\\
&= \int_{-\infty}^{+\infty} \exp\left(-ax^2\right) \cos 2\pi kx\,dx - \mathbf{j}\int_{-\infty}^{+\infty} \exp\left(-ax^2\right) \sin 2\pi kx\,dx.
\end{aligned}$$

Go Back

The second integrand is odd, so integration over a symmetrical range gives 0.

The value of the first integral is given by Abramowitz and Stegun, so:

$$\mathcal{F}_x(k) = \sqrt{\frac{\pi}{a}} \exp\left(\frac{-\pi^2 k^2}{a}\right),$$

therefore a Gaussian transforms to another Gaussian ■

Go Back

The Weierstrass transform of a function $f : \mathbf{R} \to \mathbf{R}$, is a smoothed version of f(x) obtained by averaging the values of f, weighted with a Gaussian centred at x.

The function F is defined by:

$$F(x) = \frac{1}{\sqrt{4\pi}} \int_{-\infty}^{\infty} f(y)\, e^{-\frac{(x-y)^2}{4}}\, dy = \frac{1}{\sqrt{4\pi}} \int_{-\infty}^{\infty} f(x-y)\, e^{-\frac{y^2}{4}}\, dy$$

the convolution of f with the Gaussian function

$$\frac{1}{\sqrt{4\pi}} e^{-x^2/4}$$

The factor $\frac{1}{\sqrt{4\pi}}$ is chosen so that the Gaussian will have a total integral of 1, with the consequence that constant functions are not changed by the Weierstrass transform.

Go Back

Bivariate analysis is a form of quantitative analysis which involves the analysis of two variables, for the purpose of determining the empirical relationship between them.

Bivariate analysis can be helpful in testing simple hypotheses of association. Bivariate analysis can help determine to what extent it becomes easier to know and predict a value for one variable (possibly a dependent variable) if we know the value of the other variable (possibly the independent variable) (see also correlation and simple linear regression).

Go Back

The central limit theorem (CLT) states that, under appropriate conditions, the distribution of a normalized version of the sample mean converges to a standard normal distribution. This holds even if the original variables themselves are not normally distributed. There are several versions of the CLT, each applying in the context of different conditions.

The theorem is a key concept in probability theory because it implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions.

Go Back

A varifocal lens is a camera lens with variable focal length in which focus changes as focal length (and magnification) changes, as compared to a parfocal ("true") zoom lens, which remains in focus as the lens zooms (focal length and magnification change).

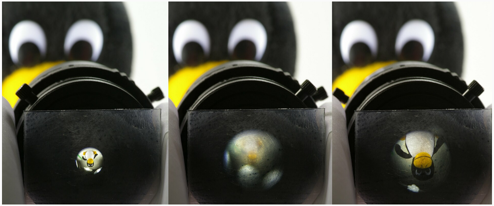

Figure 44: A varifocal lens. Left image is at 2.8 mm, in focus. Middle image is at 12 mm with the focus left alone from 2.8 mm. Right image is at 12 mm refocused. The close knob is focal length and the far knob is focus.

Appendix

Go Back

In photography and optics, vignetting (/v n j t ŋ/; vin-YET-ing) is a reduction of an image's brightness or saturation toward the periphery compared to the image center. The word vignette, from the same root as vine, originally referred to a decorative border in a book. Later, the word came to be used for a photographic portrait that is clear at the center and fades off toward the edges. A similar effect is visible in photographs of projected images or videos off a projection screen, resulting in a so-called "hotspot" effect.

Go Back

Impossible colors are colors that do not appear in ordinary vision.

Different color theories suggest different hypothetical colors that humans are incapable of perceiving for one reason or another, and fictional colors are routinely created in popular culture.

While some such colors have no basis in reality, phenomena such as cone cell fatigue enable colours to be perceived in certain circumstances that would not be otherwise.

Go Back

Generation loss is the loss of quality between subsequent copies or transcodes of data.

Anything that reduces the quality of the representation when copying, and would cause further reduction in quality on making a copy of the copy, can be considered a form of generation loss.

File size increases are a common result of generation loss, as the introduction of artifacts may actually increase the entropy of the data through each generation.

[1] Edward H Adelson. "Checkershadow illusion. 1995". In: *URL http://web. mit. edu/persci/people/adelson/checkershadow_illusion. html* (2005).

[2] Adobe. *Lossy vs Lossless Compression Differences and When to Use.* 2024. URL: https://www.adobe.com/uk/creativecloud/photography/discover/lossy-vs-lossless.html.

[3] American Optometric Association. *Color vision deficiency.* 2024. URL: https://www.aoa.org/healthy-eyes/eye-and-vision-conditions/color-vision-deficiency?sso=y.

[4] Colour Blind Awareness. *Types of Colour Blindness.* 2024. URL: https://www.colourblindawareness.org/colour-blindness/types-of-colour-blindness/.

[5] Michael F Deering. "The limits of human vision". In: *2nd international immersive projection technology workshop*. Vol. 2. 1. 1998.

[6] Zoltan Derzsi and Robert Volcic. "Not only perception but also grasping actions can obey Weber's law". In: *Cognition* 237 (2023), p. 105465.

[7] S. A. Eugster. *CbCr plane of the YCbCr color space, with a Y value of 0.5, Cb on the horizontal x axis going from -1 to 1, Cr on the y axis going from -1 to 1 as well.* 2010. URL: `https://commons.wikimedia.org/wiki/File:YCbCr-CbCr_Scaled_Y50.png`.

[8] Michael Gabler. *Various image formats are categorized by scope in this diagram, along two axes: Raster vs Vector, and Authoring vs Delivery.* 2011. URL: `https://commons.wikimedia.org/wiki/File:Felis_silvestris_silvestris_small_gradual_decrease_of_quality.png`.

[9] Gigahertz-Optik. *Spectral Sensitivity of the Human Eye.* 2002. URL: `https://www.gigahertz-optik.com/en-us/service-and-support/knowledge-base/basics-light-measurement/light-color/spectr-sens-eye/`.

[10] healthdirect. *Colour blindness.* 2024. URL: `https://www.healthdirect.gov.au/colour-blindness`.

[11] Hyperphysics. *The Color-Sensitive Cones.* 2024. URL: `http://hyperphysics.phy-astr.gsu.edu/hbase/vision/colcon.html`.

[12] Amanda Jefferies et al. "Do You See What I See?: Understanding the Challenges of Colour-Blindness in Online Learning". In: *Proceedings of 10th European Conference for E-Learning*. Academic Conferences Ltd. 2011.

[13] Susanne Kohl et al. "Achromatopsia". In: (2018).

[14] *Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB.* Standard. Geneva, CH: International Electrotechnical Commision, 1999.

[15] *Multimedia systems and equipment - Colour measurement and management - Part 2-5: Colour management - Optional RGB colour space - opRGB.* Standard. Geneva, CH: International Electrotechnical Commision, 2007.

[16] Danny Pascale. "A review of rgb color spaces... from xyy to r'g'b'". In: *Babel Color* 18 (2003), pp. 136–152.

[17] *Photography and graphic technology — Extended colour encodings for digital image storage, manipulation and interchange.* Standard. Geneva, CH: International Organization for Standardization, 2013.

[18] PolBr. *sRGB colors situated at calculated position in CIE 1931 chromaticity diagram (edited from* $CIExy1931_sRGB.svg).LuminanceYsetsothatR + G + B = 1toavoidbrightlinestowardprimaries'complementarycolours.$. 2021. URL: `https://commons.wikimedia.org/wiki/File:SRGB_chromaticity_CIE1931.svg`.

[19] R. Rehák, P. Bodrogi, and J. Schanda. "On the use of the sRGB colour space". In: *Displays* 20.4 (1999), pp. 165–170. ISSN: 0141-9382. DOI: https://doi.org/10.1016/S0141-9382(99)00019-0. URL: https://www.sciencedirect.com/science/article/pii/S0141938299000190.

[20] Inc. Reproductions. *Understanding the Differences Between RGB and CYMK Colors*. 2024. URL: https://reproductionsinc.com/understanding-the-differences-between-rgb-and-cymk-colors/.

[21] Andrew Rodney. "The role of working spaces in Adobe applications". In: *Technical Paper. Adobe. Retrieved* (2008), pp. 05–09.

[22] Slippens. *Three examples of color halftoning with CMYK separations*. 2009. URL: https://en.wikipedia.org/wiki/File:Halftoningcolor.svg.

[23] J. Sneyers. *Various image formats are categorized by scope in this diagram, along two axes: Raster vs Vector, and Authoring vs Delivery*. 2022. URL: https://commons.wikimedia.org/wiki/File:Image_formats_by_scope.svg.

[24] Kevin E Spaulding, Geoffrey J Woolfe, and Edward J Giorgianni. "Reference input/output medium metric rgb color encodings". In: *Color Imaging Conference*. Vol. 1. 2000, p. 2.