Topics on Robotics & Vision

B.Sc

Digital Image Processing

LectureSlide

D. T. McGuiness, Ph.D

WS.2025

Table of Contents

Introduction

- The goal of this lecture is to introduce you to image processing and its wide applications in industry.
- We shall have a wide focus on the technologies and methods which make image processing an essential discipline for engineers.
- This lecture is a total of 4 SWS with a total of sixty (60) UE.
- A unit (UE) is defined as 45 min lecture.

- Lecture materials and all possible supplements will be present in its Github Repo.
    - You can easily access the link to the web-page from here.

> Github is chosen for easy access to material management and CI/CD capabilities and allowing hosting websites.

- In the lecture content is also distributed as a WebBook which can be accessed from the Repo website.

- The student should be comfortable with working with physical problems and have a basic understanding of material science along with calculus.

Requirements	Taught Lecture	Code	Degree	Outcome
Python	Programming I	PRG I	B.Sc	Python Programming
Linear Algebra	Mathematics I	MAT I	B.Sc	Signal Processing
				Image Processing
-				Camera Technology
-				Statistical Analysis

Table 1: Distribution of materials across the semester.

Description	Value
Official Name	Image Processing
Lecture Code	IMP
Module Code	MECH-B-5-MRV-IMP-ILV
Lecture Name	Digital Image Processing
Semester	5
Season	WS
Lecturer	Daniel T. McGuiness, Ph.D
Module Responsible	BnM
Software	Python
SWS Total	4
UE Total	60
ECTS	5
Working Language	English

- The lecture will have a single personal assignment comprising of a set list of questions which you can use programming languages to solve on your own.
- There will also be a group assignment where you will team up with your classmates to come up with ideas for applying image processing concepts to problems.

Assignment Type	Value
Personal Assignment	40
Group Project	60
Sum	100

Title
Fundamentals of Image Processing
Computer Vision: Algorithms and Applications
Feature Extraction and Image Processing for Computer Vision
Digital Image Processing
Types Of Camera Sensor
Introduction To Quantum Efficiency
Dark Current
Linearity - Imaging Topics

Table 2: Lecture sources which can be useful during the course of the lecture. For more information on sources, please consult the repo.

Introduction i

Topic	Units	Self Study
Mathematical Fundamentals	4	8
Perception	4	8
Image Formats	4	8
Camera	4	8
Display	4	8
Noise	4	8
Histogram Operations	4	8
Morphological Opeations	4	8
Blurring Filters	4	8

Introduction ii

Topic	Units	Self Study
Feature Analysis	4	8
Edge Detection	4	8
Introduction to Artificial Neural Networks	4	8
Computer Vision using Convolutional Neural Networks	4	8
SUM	52	104

Mathematical Fundamentals

Table of Contents

(LO1) An Overview of mathematical methods,

(LO2) A revisit on convolution.

(LO3) Definitions of analogue and digital,

(LO4) A mathematical look into Discrete Fourier Transform (DFT).

Mathematical Fundamentals

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

Concepts and Principles

- Principle of convolution,
- Discrete Fourier analysis,
- Shannon-Nyquist Sampling Theorem,
- A brief introduction to Information Theory,
- The concept of information entropy.

- Convolution, mathematically is defined as:

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

- Where f and g are arbitrary function and $*$ is the convolution operator.

> To put it simply, convolution is just fancy multiplication. But in principle, it has a strong relationship with Laplace transform and is used significantly in image processing.

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:

$$\text{Plan} \times \text{Patients} = \text{Daily Usage}$$
$$3 \times [1, 2, 3, 4, 5] = [3, 6, 9, 12, 15]$$

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:

$$[3] \times [4] = [12]$$

Solution

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

$$\text{Plan:}[3, 2, 1]$$

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:

$$\text{Patient:}[1, 2, 3, 4, 5]$$

what's our medicine usage each day? Let's see

Solution

- 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:

$$\text{New Patient List:} [5, 4, 3, 2, 1]$$

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

$$\text{Rooms:} [3, 2, 1]$$

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:

C.R. 1

text

```
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:

C.R. 2

text

```
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

C.R. 3

text

```
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):

C.R. 4

python

```
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

C.R. 5

python

```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
```

Solution

C.R. 6

python

```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:

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

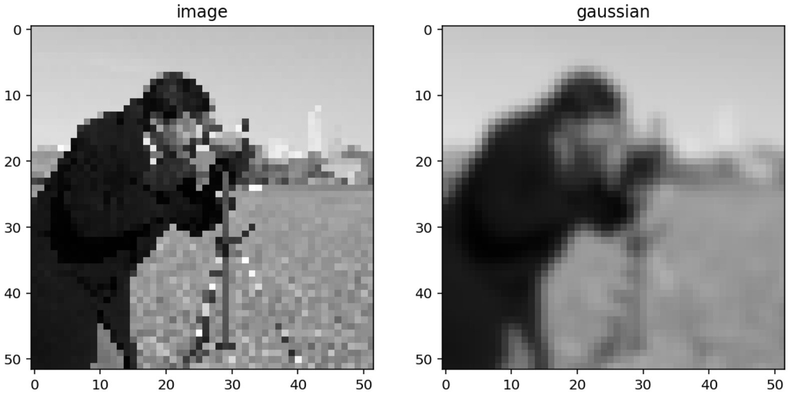

Figure 1: An example of convolution used in image processing. Here a pixellated image is smoothened out using Gaussian Blur which relies on convolution.

- 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):

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

- 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 as it relates the three (3) signals of interest:
  1. Input signal,
  2. Output signal,
  3. Impulse response.
- But now, let's look at some of it's properties:

Commutative

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

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

Distributive

- if there are three signals $f(t), g(t), h(t)$, then the convolution of $f(t)$ is said to be distributive:

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

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:

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

- 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 $\delta\,(x,\ y)$:

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

  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

- 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.

$$f_s \geq 2f_m$$

where f_s is the signal frequency, f_m is the maximum sample frequency.

> This is only a theoretical limit and **NOT** a practical one.

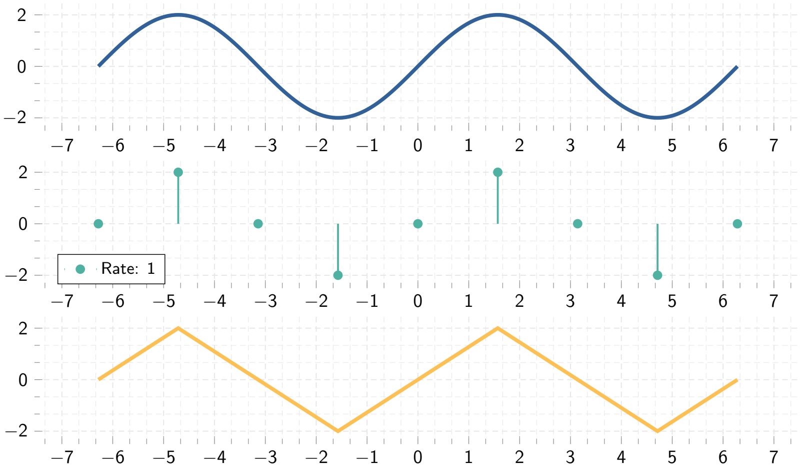

Figure 2: Reconstruction of the signal with 1 times the signal frequency.

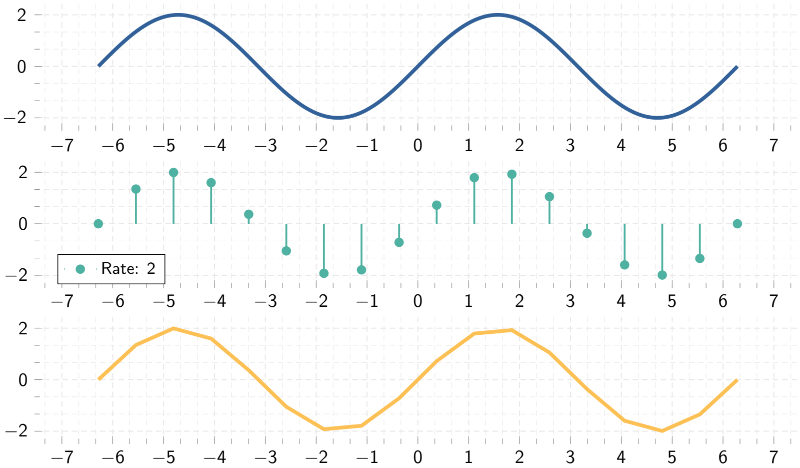

Figure 3: Reconstruction of the signal with 2 times the signal frequency.

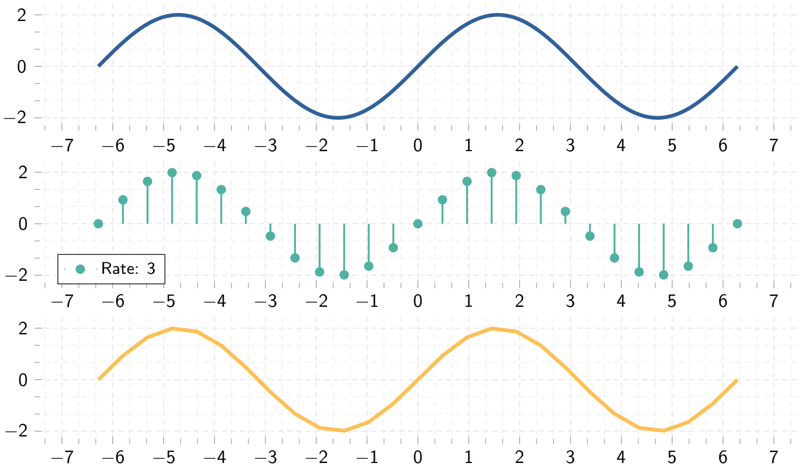

Figure 4: Reconstruction of the signal with 3 times the signal frequency.

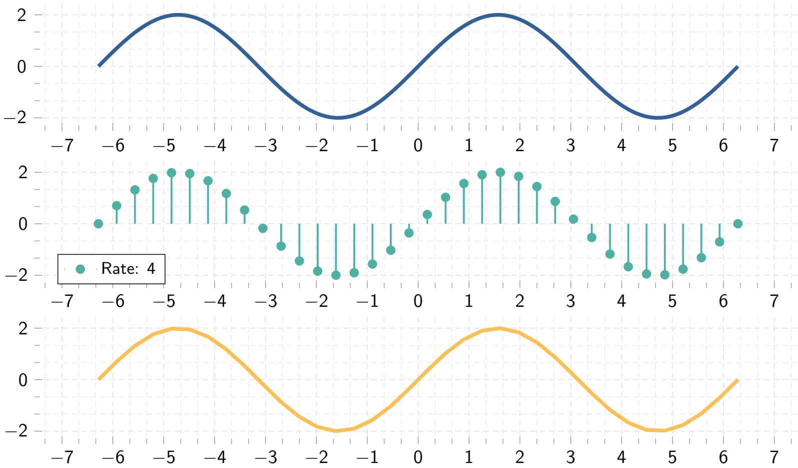

Figure 5: Reconstruction of the signal with 4 times the signal frequency.

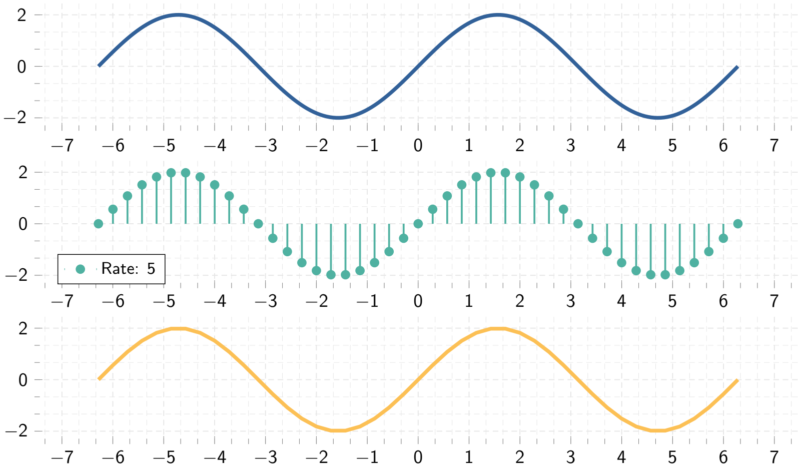

Figure 6: Reconstruction of the signal with 5 times the signal frequency.

- 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 which will vary based on phase.
- Even 4 times 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.

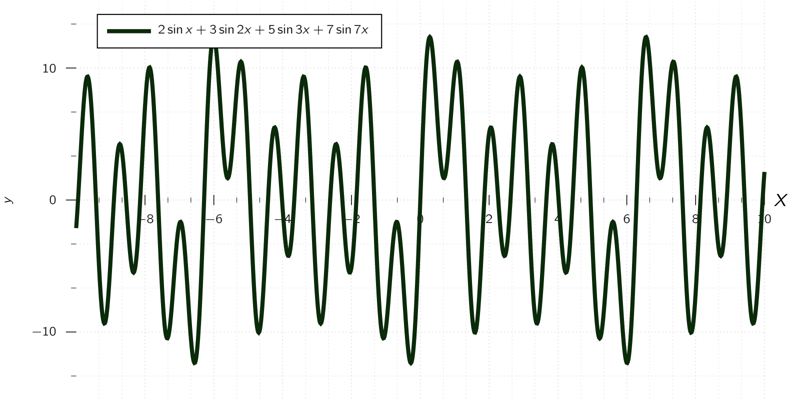

Figure 7: A sample signal with containing sample sine waves.

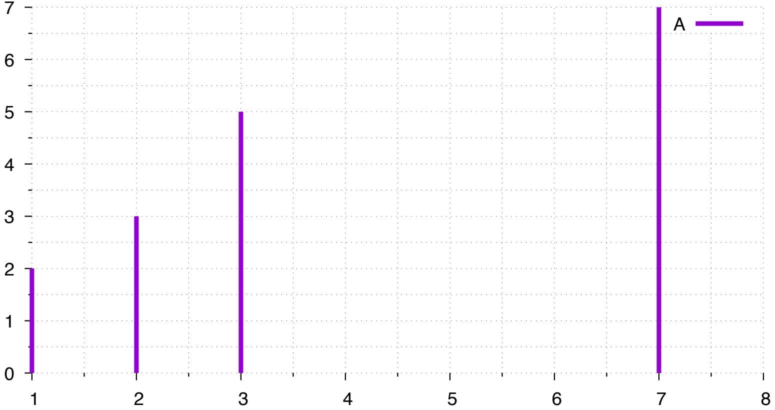

Figure 8: The FFT of the previous complex signal.

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

Aliasing Only resolving frequencies lower than the *Nyquist frequency* and higher frequencies get aliased to lower frequencies.

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

- 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.

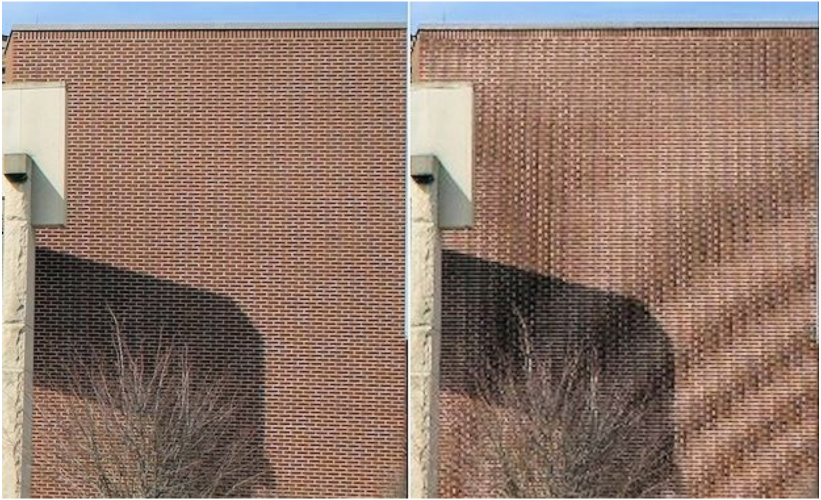

Figure 9: An example of under-sampling an image. Here aliasing produces non-real distortions of digitized images.

- 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, might cause the transform to be corrupted in some way.
- Let's start with looking at a simple sine-wave

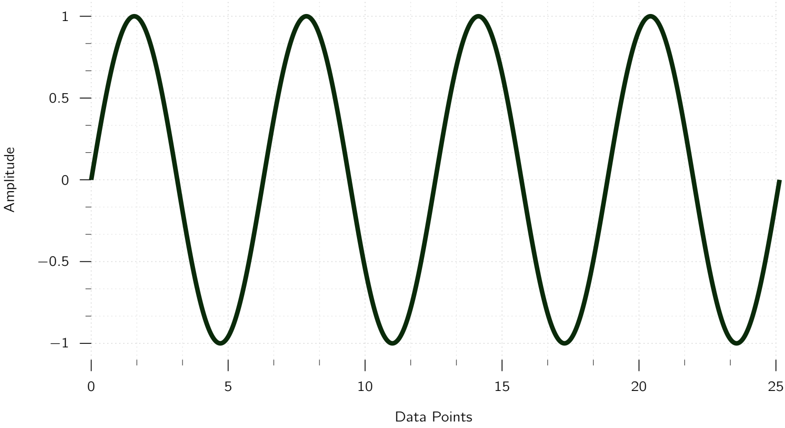

Figure 10: An example of a sine wave with four (4) complete cycles.

- Computing the discrete power spectrum gives:

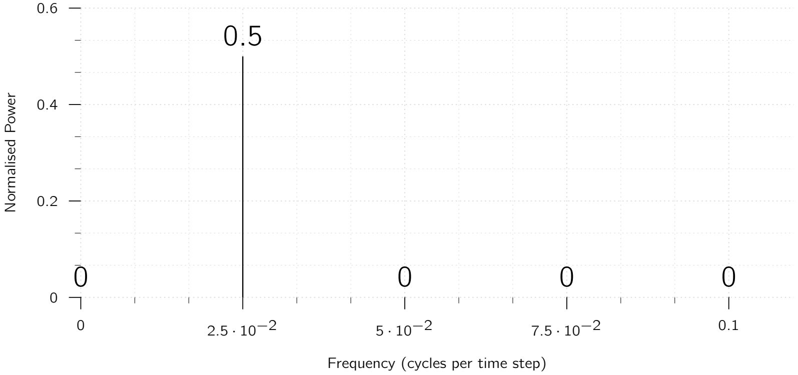

Figure 11: The PSD of an un-windowed sine wave.

- 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 which **cuts off** the sine wave such that the sine function does not complete an integer number of cycles within the time domain.

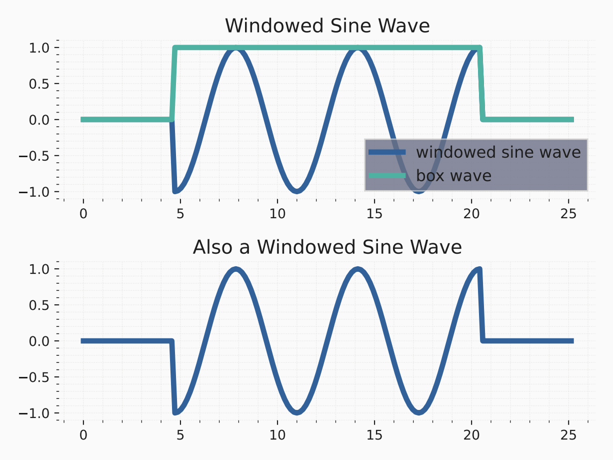

Figure 12: Windowed sine wave.

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

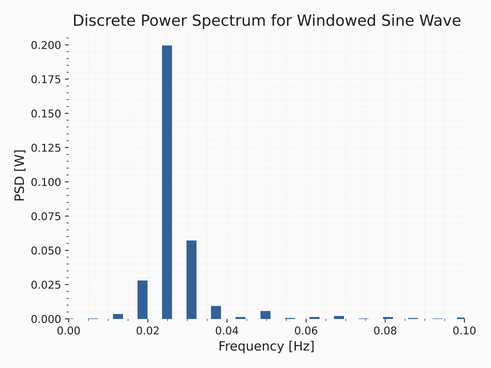

Figure 13: PSD of a windowed sine wave.

Example

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

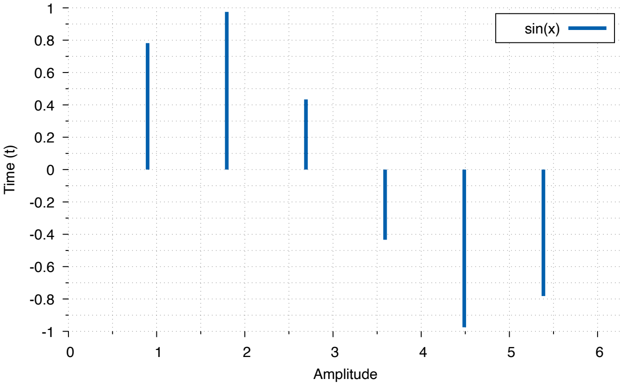

Figure 14: A Sampled Sine wave.

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:

$$x_k = \begin{bmatrix} 0 & 0.707 & 1 & 0.707 & 0 & -0.707 & -1 & -0.707 \end{bmatrix}$$

Solution

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

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

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

For $x_0 = 0$ the exponential is removed and are left with $X_0 = 0$.

Solution

For the cases of X_1:

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

$$= \begin{bmatrix} 0 \\ 0.707 \\ 1 \\ 0.707 \\ 0 \\ -0.707 \\ -1 \\ -0.707 \end{bmatrix}^{\mathsf{T}} \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,2\pi\,(1\times 1))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(2\times 1))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(3\times 1))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(4\times 1))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(5\times 1))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(6\times 1))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(7\times 1))/\;N} \end{bmatrix} = 0 - \mathbf{j}\,4 \quad \blacksquare$$

Solution

For the cases of X_2:

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

$$= \begin{bmatrix} 0 \\ 0.707 \\ 1 \\ 0.707 \\ 0 \\ -0.707 \\ -1 \\ -0.707 \end{bmatrix}^{\mathrm{T}} \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\, 2\pi\, (1 \times 2))/\, N} \\ e^{-(\mathbf{j}\, 2\pi\, (2 \times 2))/\, N} \\ e^{-(\mathbf{j}\, 2\pi\, (3 \times 2))/\, N} \\ e^{-(\mathbf{j}\, 2\pi\, (4 \times 2))/\, N} \\ e^{-(\mathbf{j}\, 2\pi\, (5 \times 2))/\, N} \\ e^{-(\mathbf{j}\, 2\pi\, (6 \times 2))/\, N} \\ e^{-(\mathbf{j}\, 2\pi\, (7 \times 2))/\, N} \end{bmatrix} = 0 \quad \blacksquare$$

Solution

For the cases of X_3:

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

$$= \begin{bmatrix} 0 \\ 0.707 \\ 1 \\ 0.707 \\ 0 \\ -0.707 \\ -1 \\ -0.707 \end{bmatrix}^{\mathsf{T}} \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,2\pi\,(1\times 3))/\,N} \\ e^{-(\mathbf{j}\,2\pi\,(2\times 3))/\,N} \\ e^{-(\mathbf{j}\,2\pi\,(3\times 3))/\,N} \\ e^{-(\mathbf{j}\,2\pi\,(4\times 3))/\,N} \\ e^{-(\mathbf{j}\,2\pi\,(5\times 3))/\,N} \\ e^{-(\mathbf{j}\,2\pi\,(6\times 3))/\,N} \\ e^{-(\mathbf{j}\,2\pi\,(7\times 3))/\,N} \end{bmatrix} = 0 \quad \blacksquare$$

Solution

For the cases of X_4:

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

$$= \begin{bmatrix} 0 \\ 0.707 \\ 1 \\ 0.707 \\ 0 \\ -0.707 \\ -1 \\ -0.707 \end{bmatrix}^{\mathrm{T}} \cdot \begin{bmatrix} 0 \\ e^{\text{-}(\mathbf{j}\, 2\pi\, (1 \times 4))/\; N} \\ e^{\text{-}(\mathbf{j}\, 2\pi\, (2 \times 4))/\; N} \\ e^{\text{-}(\mathbf{j}\, 2\pi\, (3 \times 4))/\; N} \\ e^{\text{-}(\mathbf{j}\, 2\pi\, (4 \times 4))/\; N} \\ e^{\text{-}(\mathbf{j}\, 2\pi\, (5 \times 4))/\; N} \\ e^{\text{-}(\mathbf{j}\, 2\pi\, (6 \times 4))/\; N} \\ e^{\text{-}(\mathbf{j}\, 2\pi\, (7 \times 4))/\; N} \end{bmatrix} = 0 \quad \blacksquare$$

Solution

For the cases of X_5:

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

$$= \begin{bmatrix} 0 \\ 0.707 \\ 1 \\ 0.707 \\ 0 \\ -0.707 \\ -1 \\ -0.707 \end{bmatrix}^{\mathrm{T}} \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\,2\pi\,(1\times 5))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(2\times 5))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(3\times 5))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(4\times 5))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(5\times 5))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(6\times 5))/\;N} \\ e^{-(\mathbf{j}\,2\pi\,(7\times 5))/\;N} \end{bmatrix} = 0 \quad \blacksquare$$

Solution

For the cases of X_6:

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

$$= \begin{bmatrix} 0 \\ 0.707 \\ 1 \\ 0.707 \\ 0 \\ -0.707 \\ -1 \\ -0.707 \end{bmatrix}^{\mathsf{T}} \cdot \begin{bmatrix} 0 \\ e^{-(\mathbf{j}\, 2\pi\, (1 \times 6))/\ N} \\ e^{-(\mathbf{j}\, 2\pi\, (2 \times 6))/\ N} \\ e^{-(\mathbf{j}\, 2\pi\, (3 \times 6))/\ N} \\ e^{-(\mathbf{j}\, 2\pi\, (4 \times 6))/\ N} \\ e^{-(\mathbf{j}\, 2\pi\, (5 \times 6))/\ N} \\ e^{-(\mathbf{j}\, 2\pi\, (6 \times 6))/\ N} \\ e^{-(\mathbf{j}\, 2\pi\, (7 \times 6))/\ N} \end{bmatrix} = 0 \quad \blacksquare$$

Solution

For the cases of X_7:

$$X_7 = \sum_{n=0}^{7} x_\mathrm{n} \cdot e^{\text{-}(\mathbf{j}\,2\pi\,(7)\,n)/\;N}$$

$$= \begin{bmatrix} 0 \\ 0.707 \\ 1 \\ 0.707 \\ 0 \\ -0.707 \\ -1 \\ -0.707 \end{bmatrix}^\mathrm{T} \cdot \begin{bmatrix} 0 \\ e^{\text{-}(\mathbf{j}\,2\pi\,(1\times 7))/\;N} \\ e^{\text{-}(\mathbf{j}\,2\pi\,(2\times 7))/\;N} \\ e^{\text{-}(\mathbf{j}\,2\pi\,(3\times 7))/\;N} \\ e^{\text{-}(\mathbf{j}\,2\pi\,(4\times 7))/\;N} \\ e^{\text{-}(\mathbf{j}\,2\pi\,(5\times 7))/\;N} \\ e^{\text{-}(\mathbf{j}\,2\pi\,(6\times 7))/\;N} \\ e^{\text{-}(\mathbf{j}\,2\pi\,(7\times 7))/\;N} \end{bmatrix} = 0 + \mathbf{j}\,4 \quad \blacksquare$$

Solution

- Therefore the values are of the transform are:

$$X_\mathrm{k} = \begin{bmatrix} 0 & 0 - \mathbf{j}4 & 0 & 0 & 0 & 0 & 0 & 0 + \mathbf{j}4 \end{bmatrix}$$

- 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.

$$|X_\mathrm{k}| = \begin{bmatrix} 0 & 4 & 0 & 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Solution

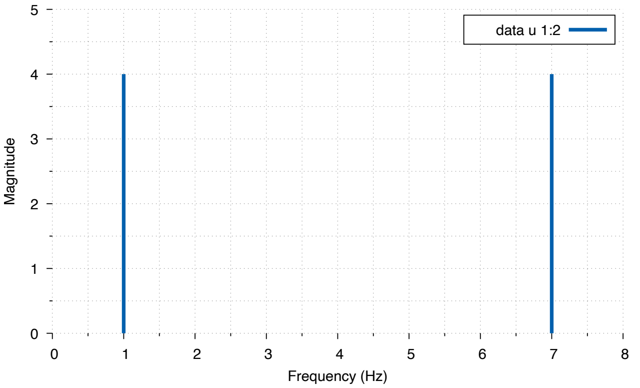

Figure 15: Sampled dataset of the original signal. There is still another step.

Solution

- The frequency resolution of the plot is the sampling frequency divided by the number of samples:

$$\text{Resolution} = \frac{\text{Sampling Frequency}}{\text{Number of Samples}}$$

- This means we can get values for every integer frequency values.

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 frequency.

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, we discard the second half of the array and multiply every point except for DC by two (2).
- The last operation is to divide the magnitudes of the lower frequencies by the number of samples used in deriving these bins:

$$\boldsymbol{X}_k = \begin{bmatrix} 0 & 8 & 0 & 0 \end{bmatrix}$$
$$\boldsymbol{X}_k / N = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \quad \blacksquare$$

- 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:

$$X = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 2 & 7 & 2 & 9 \end{bmatrix}$$

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

$$X = \begin{bmatrix} 2 & 4 & 9 & 6 & 4 & 6 & 6 & 2 & 8 & 2 \end{bmatrix}$$

the mode is (2, 6) as there are two (2) 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:

$$\mathrm{med}(x) = x\frac{n+1}{2},$$

  - if n is even,

$$\mathrm{med}(x) = x\left(\frac{n}{2}\right) + x\frac{\frac{n}{2}+1}{2}$$

- i.e., in a ordered data set of:

$$X = \begin{bmatrix} 1 & 2 & 2 & 3 & 4 & 7 & 9 \end{bmatrix},$$

- the median is 3 and in:

$$Y = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 8 & 9 \end{bmatrix}.$$

  the median is 4.5.

- The signal-to-noise ratio (SNR) can have several definitions depending on the field.
- Noise is characterised by its standard deviation, σ.
- The characterisation of the signal can differ.

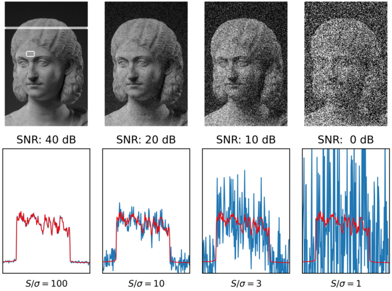

Figure 16: A gray-scale photography with different signal-to-noise ratios (SNRs).

- If the signal is known to lie between two (2) 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:

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

- 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.

- Information theory defines definite, unbreachable limits on precisely how much information can be communicated between any two (2) 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 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,

This limit can be approached by clever methods of encoding data.

- 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.

- 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.

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

- We can represent the coin flip outcome as 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 $H(x)$.

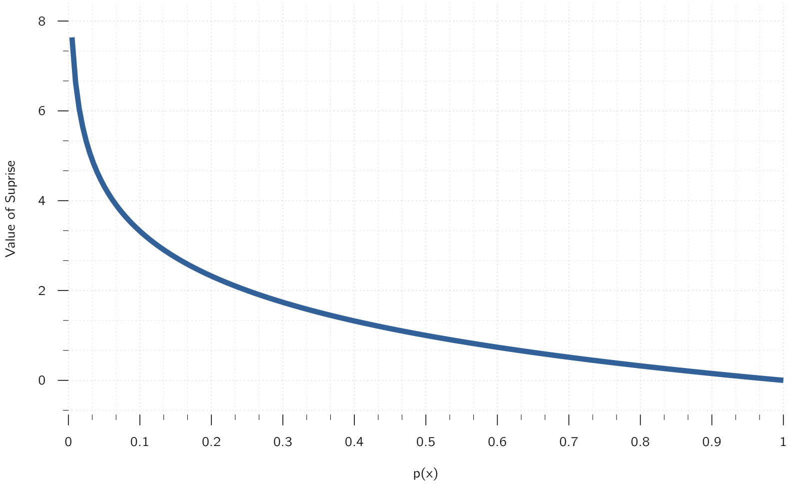

Figure 17: The quantifiable surprise with respect to increasing probability.

- 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.

- Let's look at a biased coin with a 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) = XX), the average surprise is

$$H(x) = p(x_{\mathrm{h}}) \log \frac{1}{p(x_{\mathrm{h}})} + p(x_{\mathrm{t}}) \log \frac{1}{p(x_{\mathrm{t}})},$$

- Which comes to XX bits.
- If we define a tail as $x_1 = x_{\mathrm{t}}$ and a head as $x_2 = x_{\mathrm{h}}$ then the above equation is written as:

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

- More generally, a random variable x with a probability distribution

$$p(x) = p(x_1), \ldots, p(x_m)$$

has an entropy of

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

- 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.

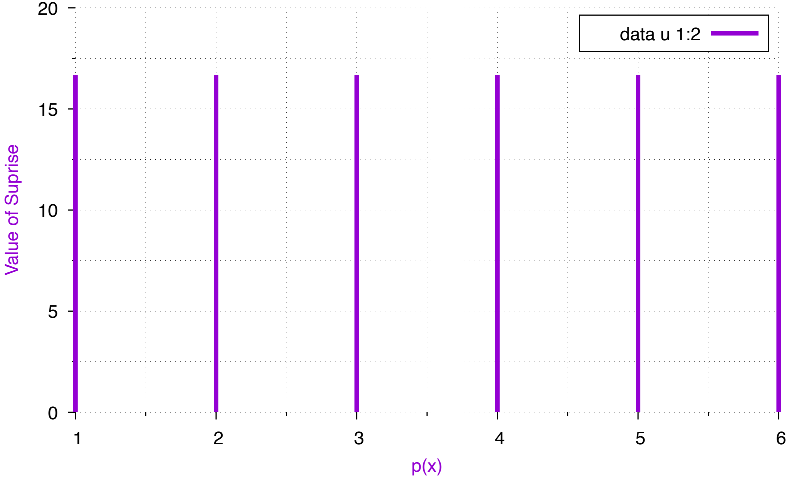

Figure 18: The probability distribution of 1 dice(s).

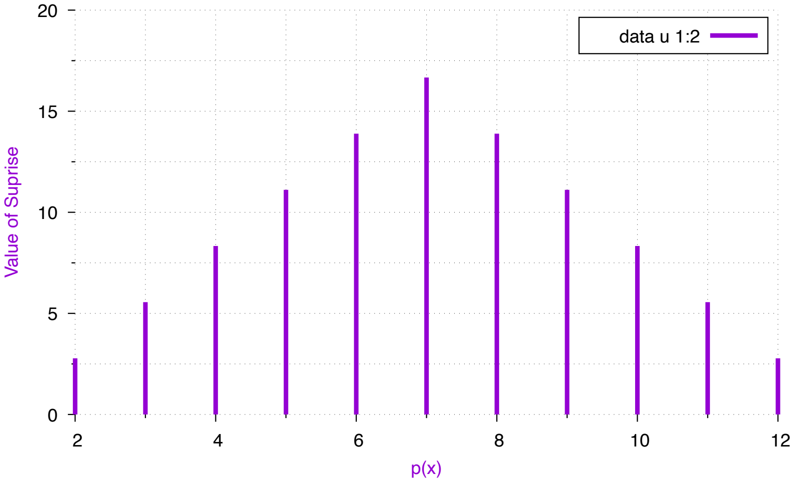

Figure 19: The probability distribution of 2 dice(s).

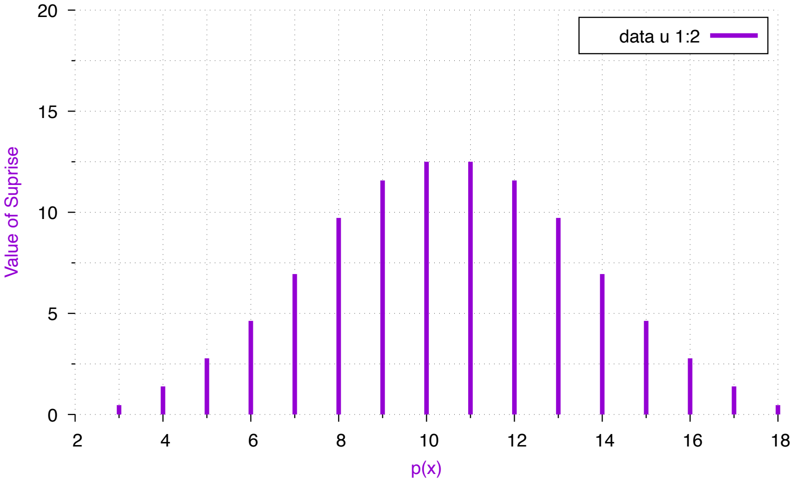

Figure 20: The probability distribution of 3 dice(s).

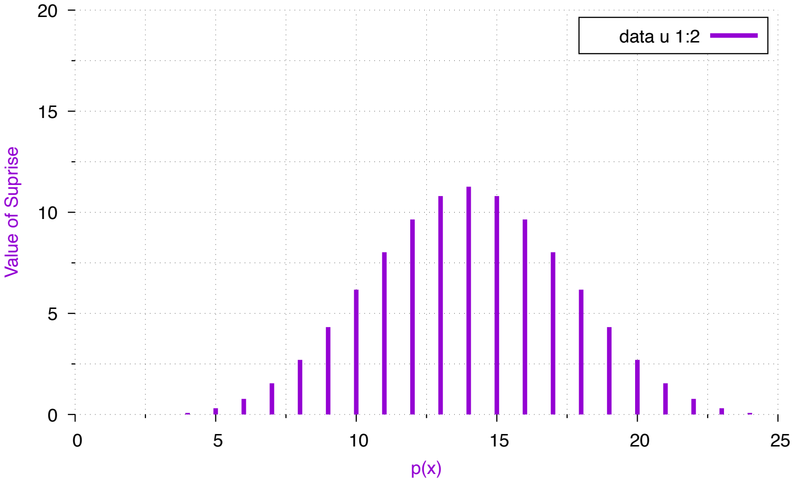

Figure 21: The probability distribution of 4 dice(s).

- 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.

- 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

(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 industrial robots.
- It is important to understand the characteristics and limitations of the human visual system [1].
- At the outset it is important to realise:
  1. The human visual system is not well understood [1].
     - 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.

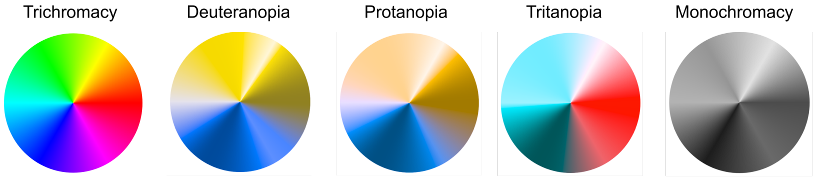

Figure 22: A color wheel depicted approximately as it would be seen by a person with different kinds of color vision or color blindness: trichromacy (normal colour vision), deuteranopia (red-green color blind), protanopia (red-green colour blind), tritanopia (blue-yellow color blind), and monochromacy (completely colour blind). [2].

- Statistically 1 in 8 men and 1 in 200 women have a form of colour blindness [3].
- Men have one $\boldsymbol{X}$ chromosome and one $\boldsymbol{Y}$ chromosome, while women have two (2) $\boldsymbol{X}$ chromosomes [4].
- To experience color blindness, the genetic mutation for color-blindness must be present on the $\boldsymbol{X}$ chromosome, but for women, this means it must be present on both $\boldsymbol{X}$ chromosomes.

> Men only need to mutation to be present on their singular $\boldsymbol{X}$ chromosome, making it much easier for them to inherit color blindness.

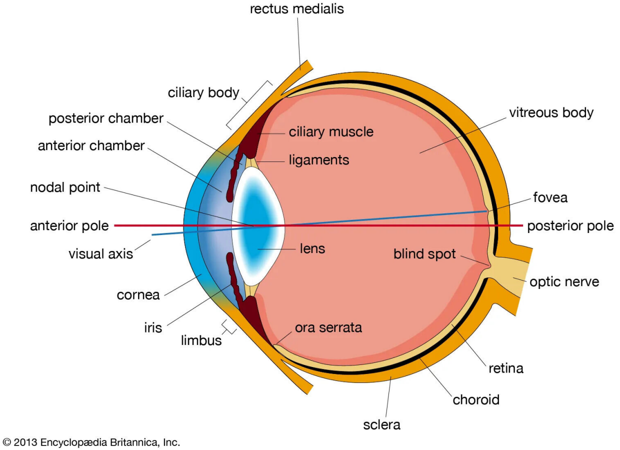

Figure 23: Horizontal section of the eye.

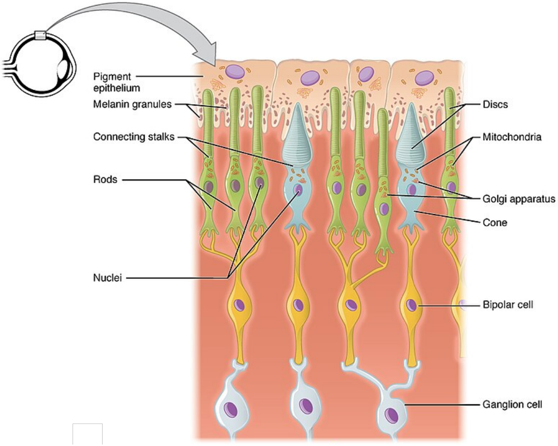

Figure 24: Functional parts of the rods and cones, which are two of the three types of photosensitive cells in the retina.

Trichromacy

- Normal colour vision uses all three (3) types of cone cells.
- 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 [3].
- In people with this condition all of their three (3) cone cell types are used to perceive light wavelengths but one type of cone cell perceives light slightly out of alignment.
- There are three (3) 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 [5]:

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 [6].
- Achromatopsia is a specific eye condition in which people see in grey-scale.
- In rare cases, partial Achromatopsia can happen which is a reduced sensitivity to all three (3) cones [6].

Figure 25: Image viewed by someone who has 3 cones.

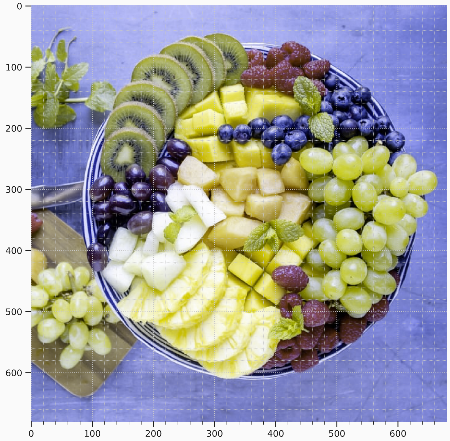

Figure 26: Image viewed by someone who has protanomaly.

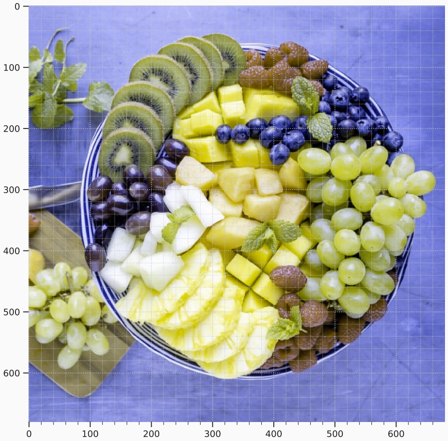

Figure 27: Image viewed by someone who has deuteranomaly.

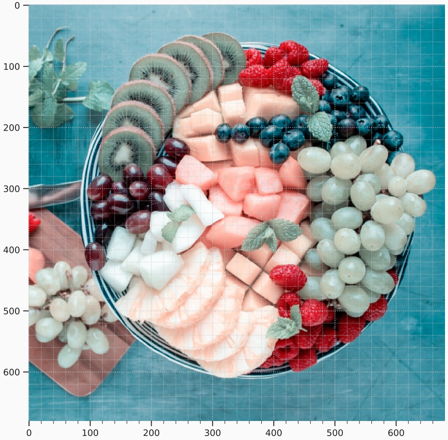

Figure 28: Image viewed by someone who has tritanomaly.

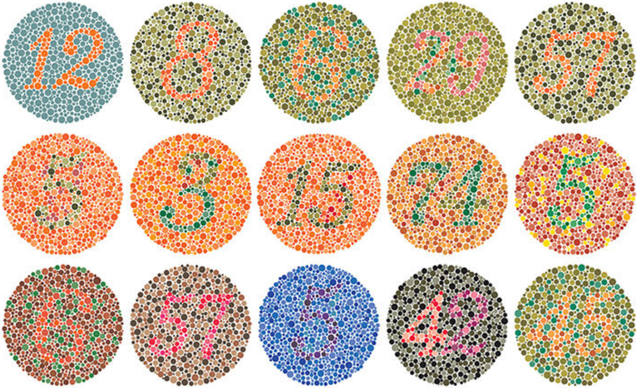

Figure 29: A colour blindness test issued to generally test before taking the driving license [7].

- 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 $I(\lambda)$.
  - assume $I(\lambda) = I_0$ as 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 [8].
- 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 [8].

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

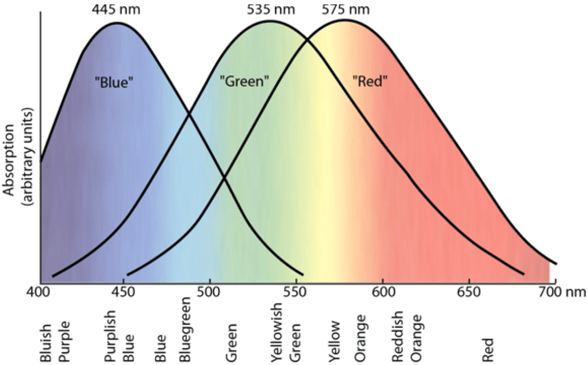

Figure 30: The colour sensitivity of the human eye [9].

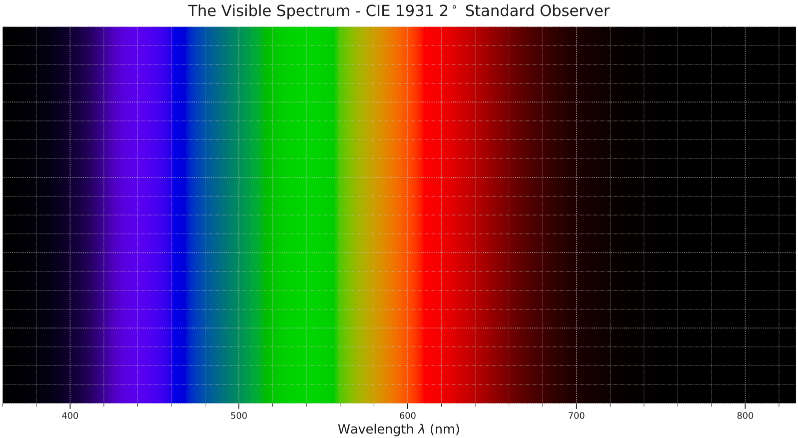

Figure 31: The visible colour spectrum visible with the human eye.

- 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 [10].

> Relates to human perception, specifically the relation between the actual change in a physical stimulus and 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.

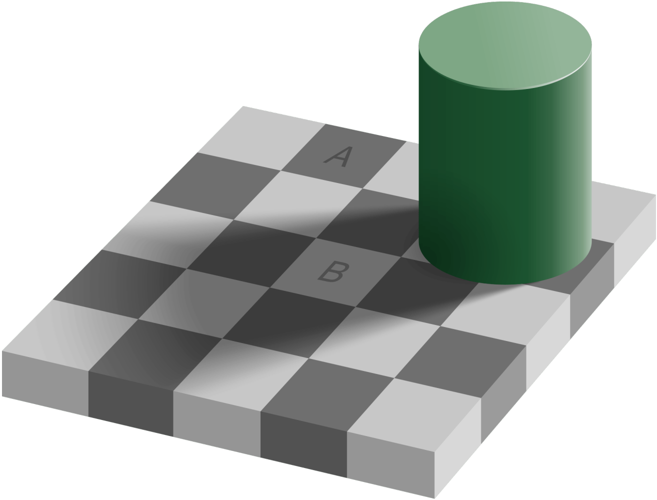

Figure 32: The checker shadow illusion [11].

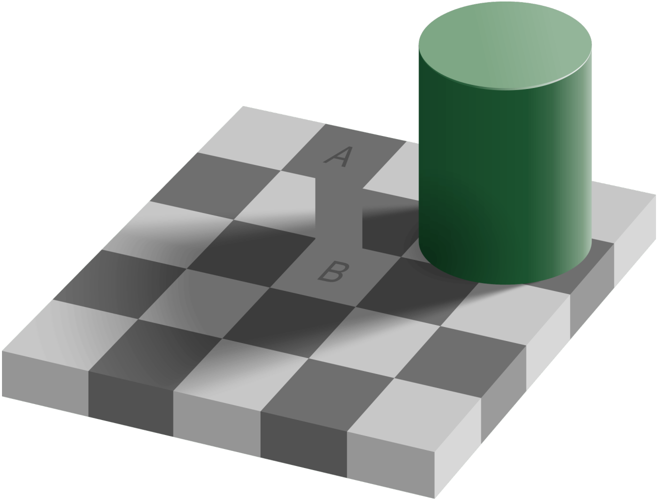

Figure 33: A region of the same shade has been drawn connecting A and B.

Figure 34: A simple picture of a dress divided the internet.

Figure 35: A simple picture of a dress divided the internet [12].

- Human colour perception is complex,
  - We can approximate its behaviour.
- **Standard Observer:** Based on psychophysical measurements, standard curves have been adopted by the CIE as sensitivity curves for the typical observer for three pigments: $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, and $\overline{z}(\lambda)$.

> These are not pigment absorption characteristics found in 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.

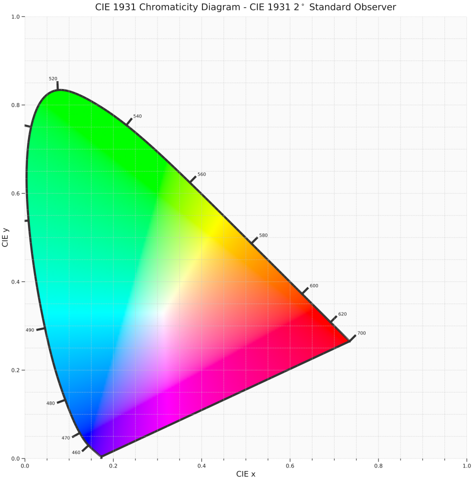

Figure 36: The colour gamut visible to the human eye, standardised by the CIE.

- A standard RGB colour-space defined by both HP and Microsoft in 1996 to use on monitors, printers, and the World Wide Web [13].
- It was subsequently standardised by IEC as IEC 61966-2-1:199 [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.

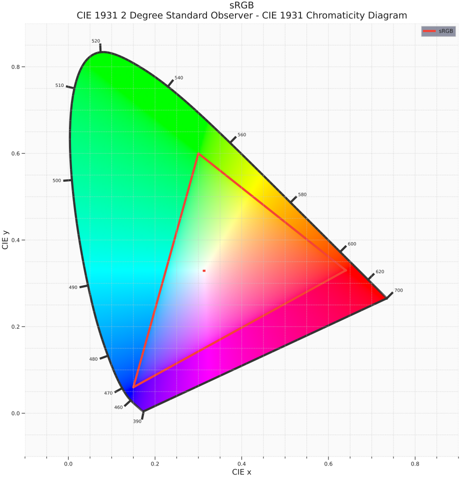

Figure 37: The sRGB colour-space superimposed to the CIE colour-gamut.

- 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 [15].
- However, consumer-level Charge Coupled Device (CCD)s 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.

- 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% [17].

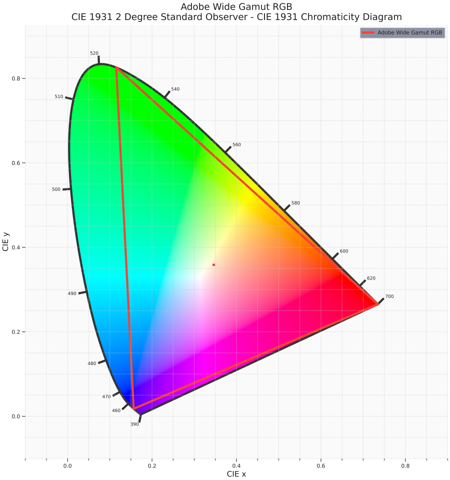

Figure 38: The wide gamut RGB colour-space superimposed to the CIE colour-gamut.

- The ProPhoto RGB colour space, a.k.a. ROMM RGB (Reference Output Medium Metric), is an output referred RGB color space developed by Kodak [18].
- 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 [19].
- The ProPhoto RGB primaries were also chosen in order to minimise hue rotations associated with non-linear tone scale operations.

> A downside is that approximately 13% of the visible colours are imaginary colors that do not exist and are impossible colour.

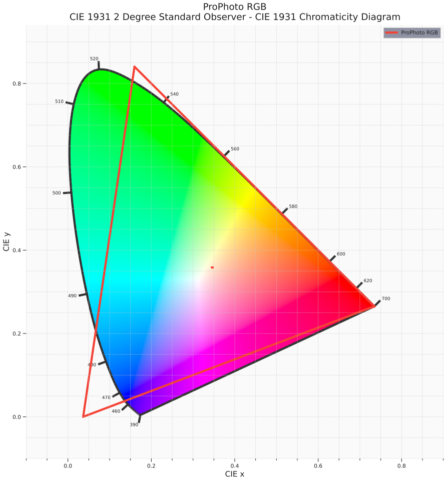

Figure 39: The ProPhoto colour-space superimposed to the CIE colour-gamut.

- 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 then standardised by the IEC as IEC 61966-2-5:1999 with a name opRGB (optional RGB color space) and is used in HDMI [20].

- 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:

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

Perception

- 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.

- 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.

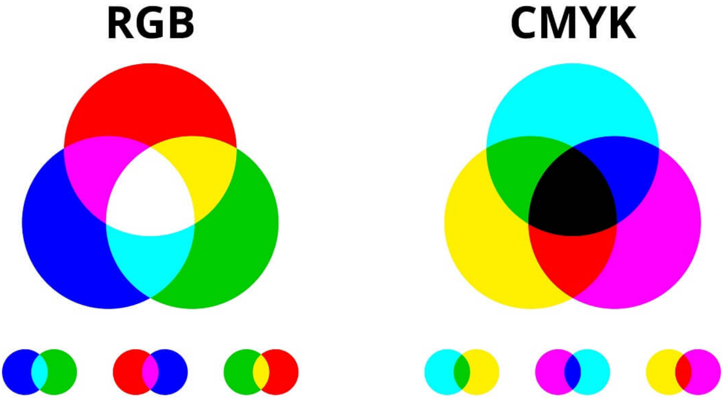

Figure 40: The differences between RGB and CYMK colours [21].

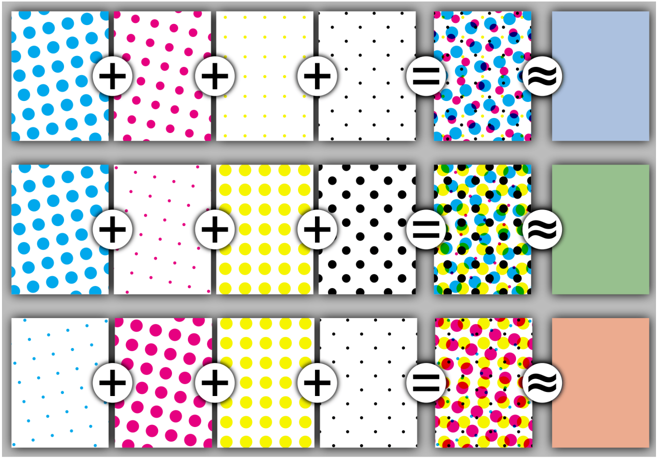

Figure 41: Three examples of color halftoning with CMYK separations, as well as the combined halftone pattern and how the human eye would observe the combined halftone pattern from a sufficient distance [22].

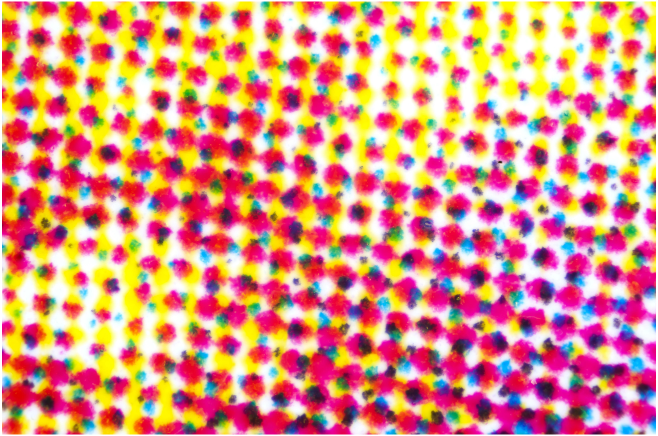

Figure 42: A printer creates any colour by combining dots in particular places relative to the other dots.

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.

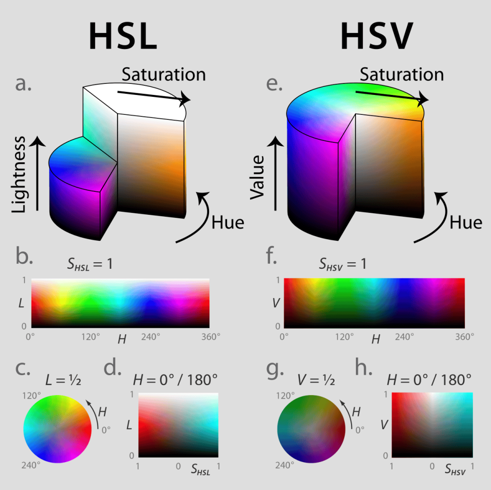

Figure 43: HSL and HSV models.

Perception

- $\boldsymbol{YC_bC_r}$ is a family of colour spaces used as a part of the color image pipeline in video and digital photography systems.
- $\boldsymbol{Y}$ 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 uses RGB signals, but they are not the best solution for storing information as they have a lot of redundancy.
- $\boldsymbol{YC}_{\mathrm{b}}\boldsymbol{C}_{\mathrm{r}}$ 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.

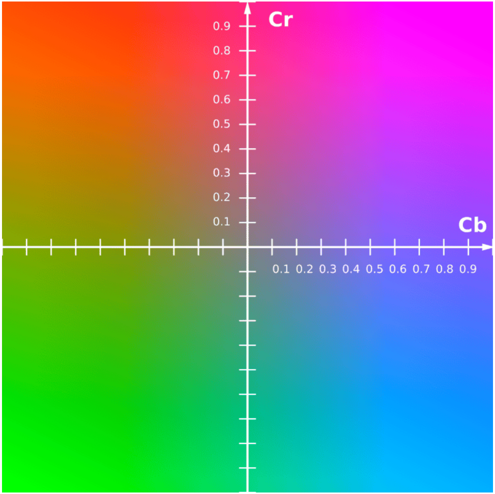

Figure 44: The $Y'C_bC_r$ plane at constant luma Y =0.5 [23].

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

[2] Hawesthoughts. ***Colorblindness Wheel.*** 2023. URL: https://commons.wikimedia.org/wiki/File:Color_blindness_wheels.svg.

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

[4] Colour Blind Awareness. ***Inherited Colour Vision Deficiency.*** 2024. URL: https://www.colourblindawareness.org/colour-blindness/causes-of-colour-blindness/inherited-colour-vision-deficiency/.

[5] 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.

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

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

[8] 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/.

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

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

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

[12] BBC. ***Optical illusion: Dress colour debate goes global***. 2015. URL: https://www.bbc.com/news/uk-scotland-highlands-islands-31656935.

[13] PolBr. ***sRGB colors situated at calculated position in CIE 1931 chromaticity diagram (edited from CIExy1931_sRGB.svg). Luminance Y set so that R+G+B = 1 to avoid bright lines toward primaries' complementary colours.*** 2021. URL: https://commons.wikimedia.org/wiki/File:SRGB_chromaticity_CIE1931.svg.

[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] 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.

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

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

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

[19] 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.

[20] ***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.

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

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

[23] 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.