

Digital Image Processing

Daniel T. McGuiness, PhD

Version: γ.2024.2



MCI



Table of Contents

1. Lecture Structure
2. Mathematical Fundamentals
3. Perception
4. Image Formats
5. Camera
6. Appendix

Lecture Structure



Table of Contents

First Steps

Introduction

Individual Assignment

Group Assignment

Point Distribution

Point Distribution

Resources



- 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.



- 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.



- 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.



- 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.



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.



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.



Mathematical Fundamentals

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





Perception

- Colour Blindness,
- Colour Standards,
- Colour Models





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.





Edge 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 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



Learning Outcomes

Convolution

Introduction

2D Convolution Example

Signal Sampling

Nyquist Sampling Theorem

Statistical Properties

Information Theory

Information and Entropy



Learning Outcomes

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





- 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



- Convolution, mathematically is defined as:

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

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:

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

$$3 \times [1, 2, 3, 4, 5] = [3, 6, 9, 12, 15]$$

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:

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



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	-----
Patients	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:

$$A = [1, 1, 2, 2, 1]$$

$$B = [1, 1, 1, 3]$$



Solution

```
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
```

```
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**):

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

- 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 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:

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

Distributive

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

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



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



- 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 $\delta(x, y)$:

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

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



- 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, not a practical one.



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

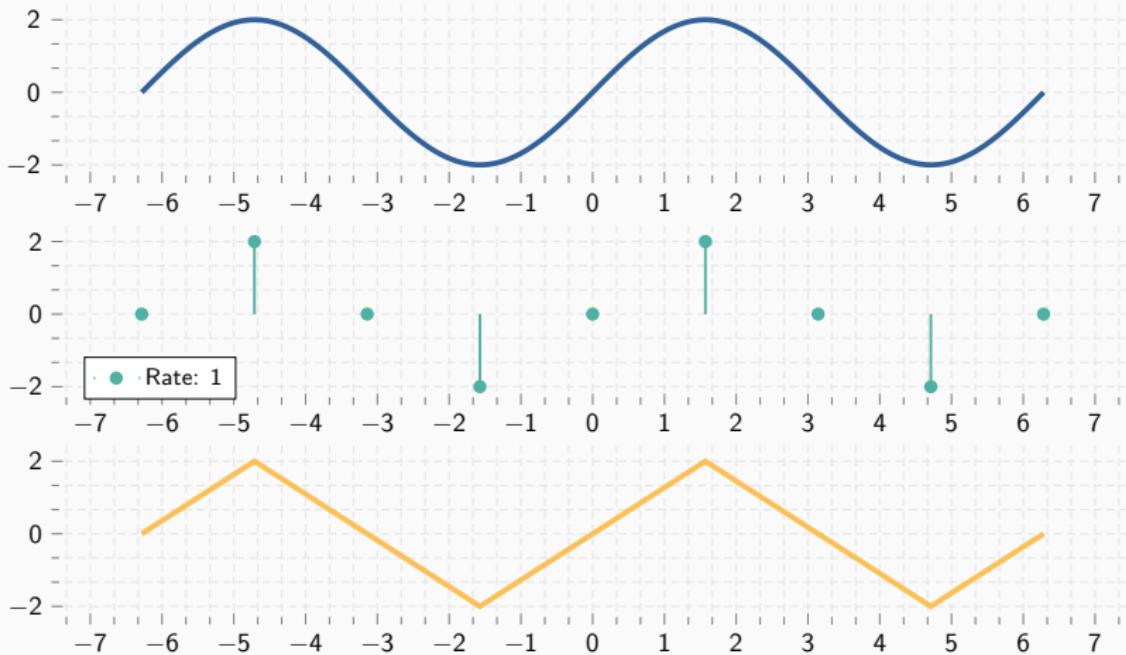


Figure 3: Reconstruction of the signal with 1 times the signal frequency.

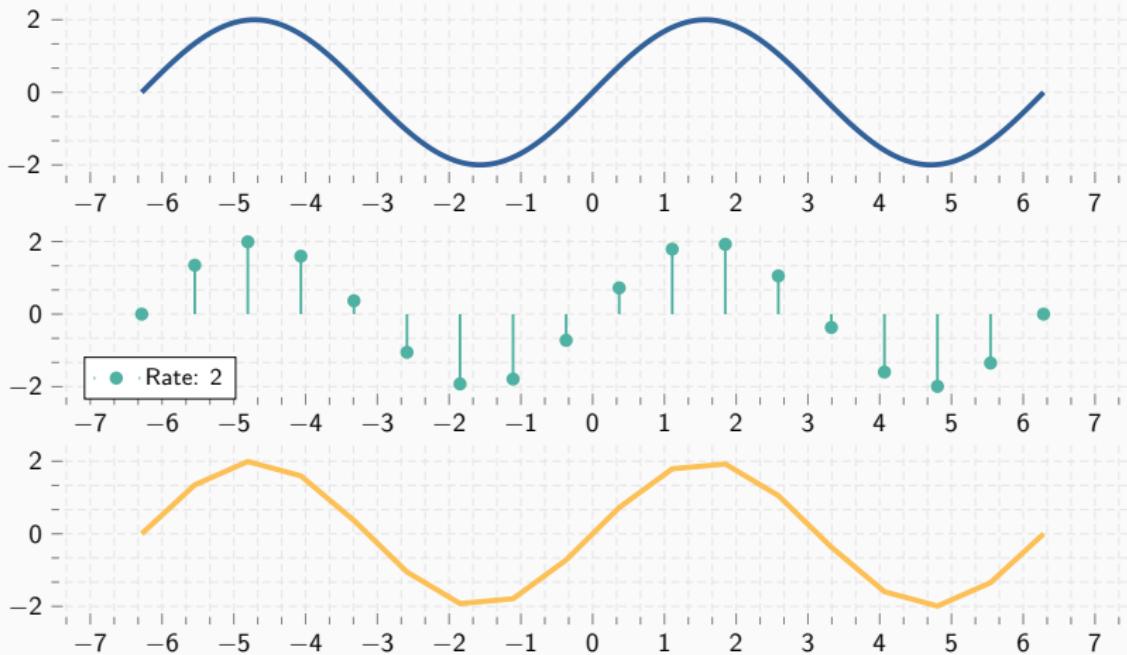


Figure 4: Reconstruction of the signal with 2 times the signal frequency.

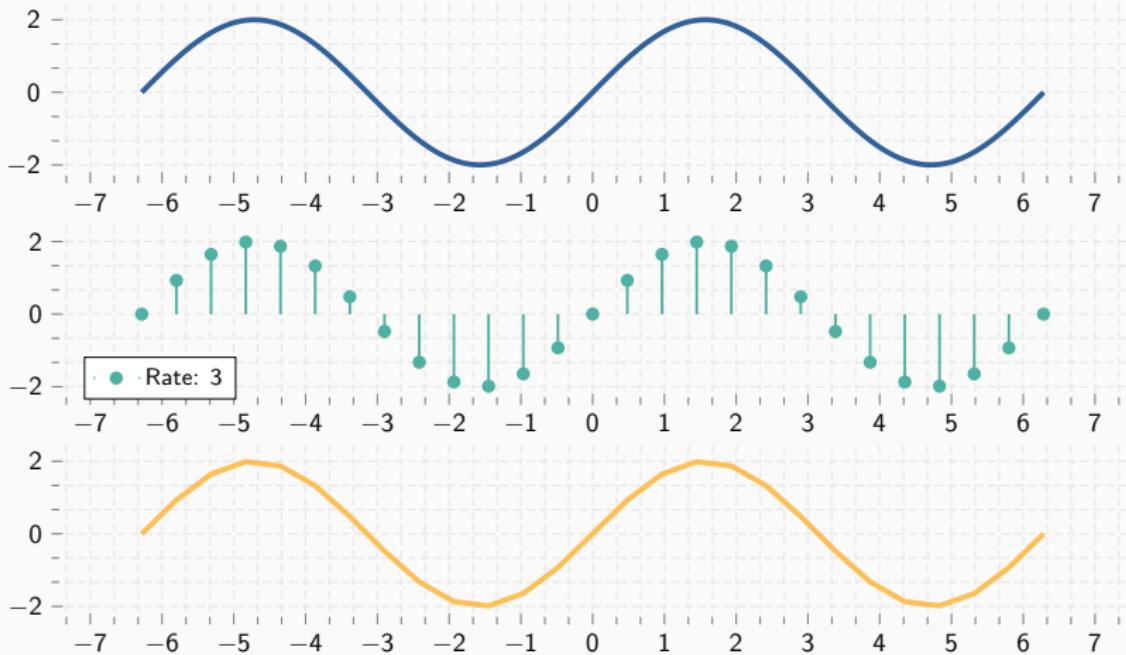


Figure 5: Reconstruction of the signal with 3 times the signal frequency.

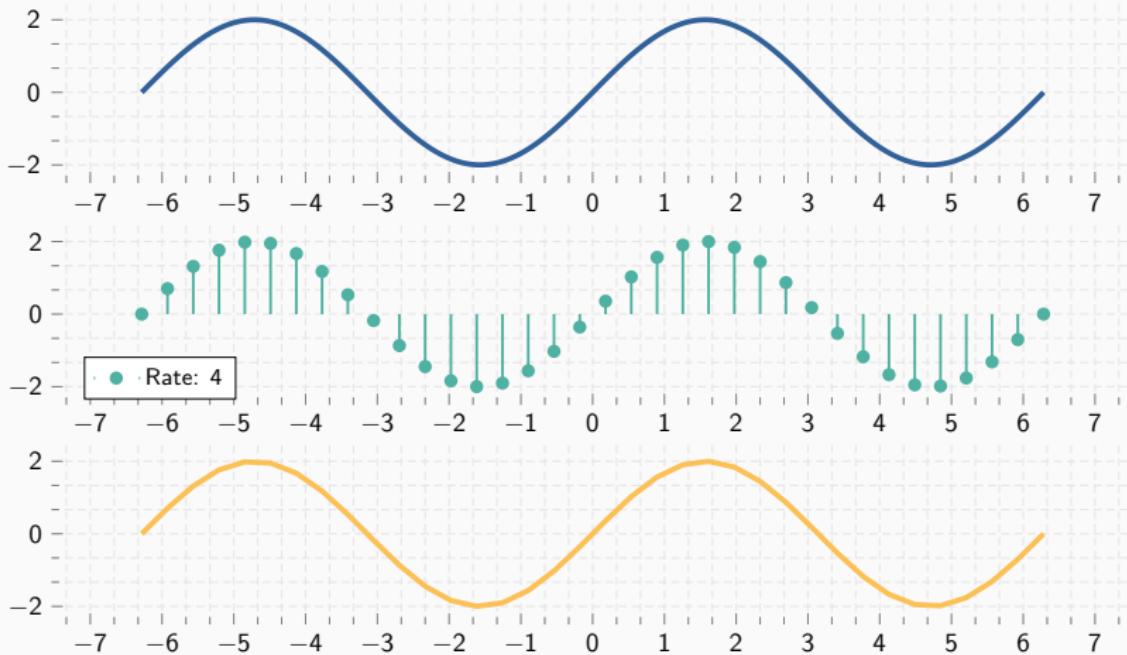


Figure 6: Reconstruction of the signal with 4 times the signal frequency.

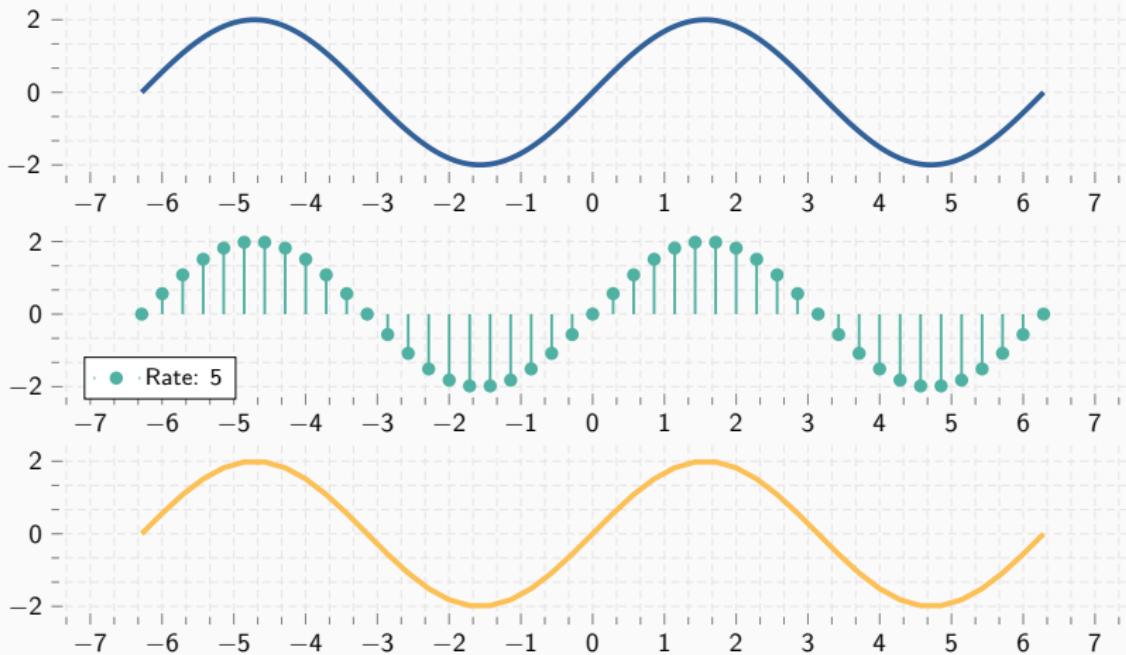


Figure 7: Reconstruction of the signal with 5 times the signal frequency.



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

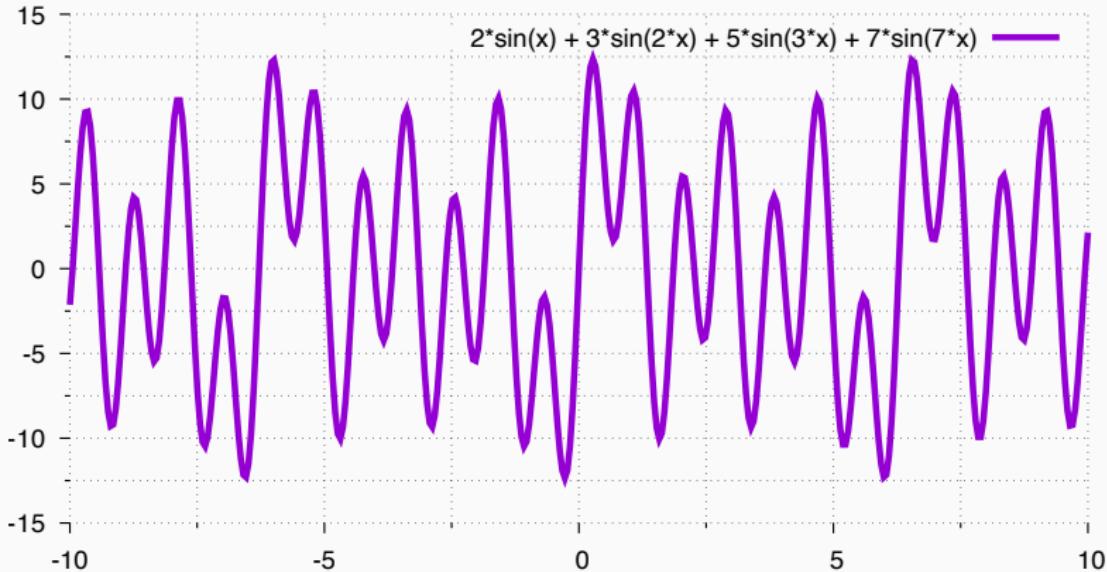


Figure 8: A sample signal with containing sample sine waves.



Figure 9: The FFT of the previous complex signal.



- 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 $= \pi$, 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.

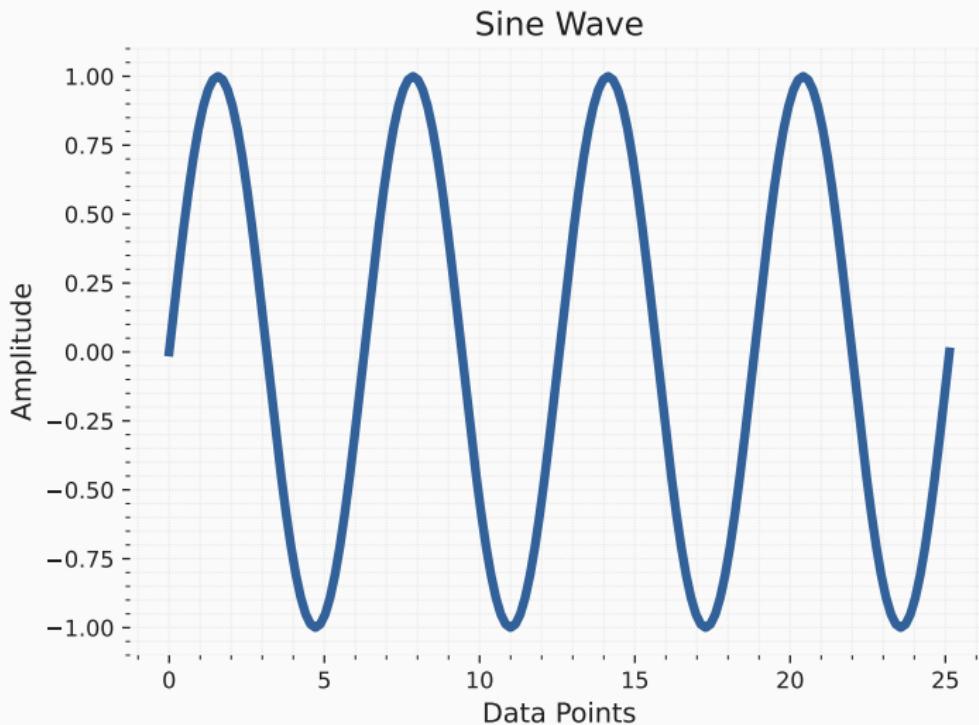


Figure 10: An example of a sine wave with four (4) complete cycles.



- Computing the discrete power spectrum gives:

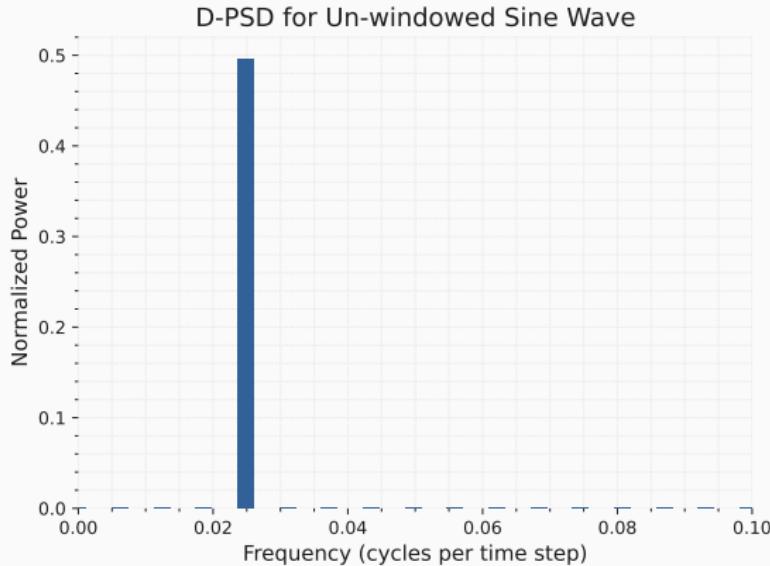


Figure 11: The PSD of an un-windowed sine wave.



- 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.

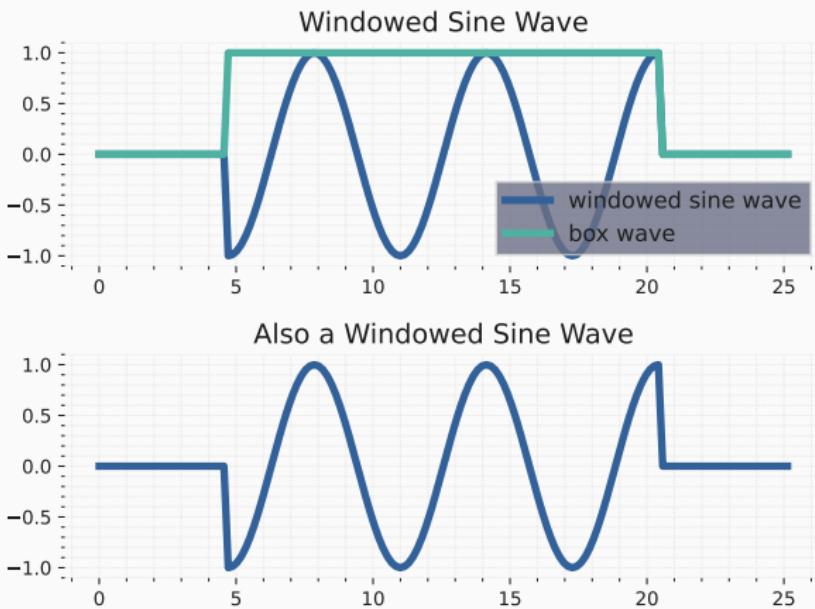


Figure 12: Windowed sine wave.



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

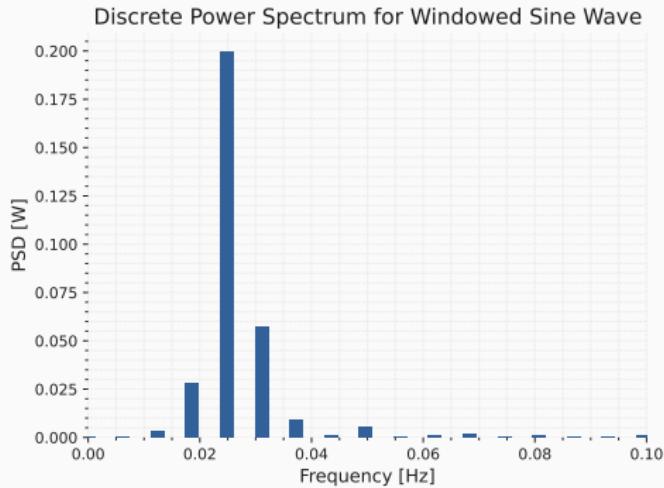


Figure 13: PSD of a windowed sine wave.



Example

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

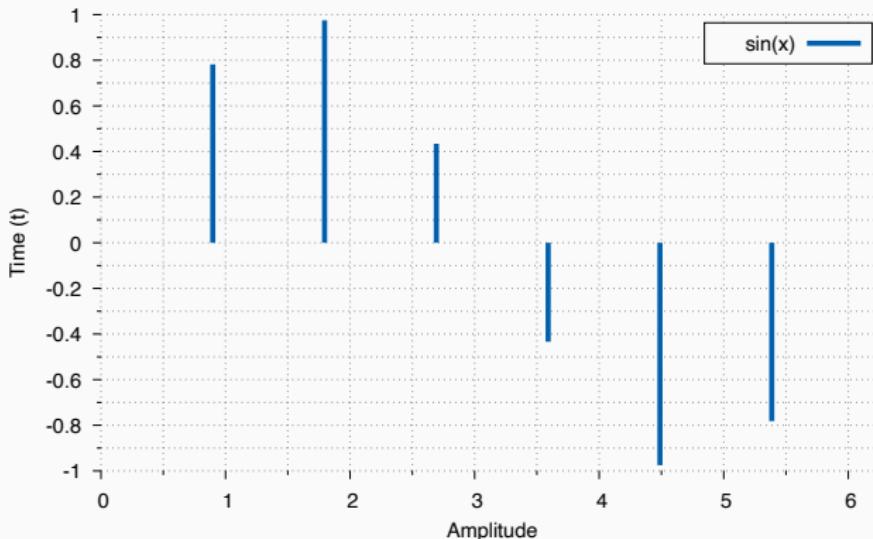


Figure 14: A Sampled Sine wave.



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:

$$x_k = [0, 0.707, 1, 0.707, 0, -0.707, -1, -0.707]$$

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

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-(j2\pi kn)/N}$$

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 :

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

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



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 :

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

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



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_4 :

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

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



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_5 :

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

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



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 :

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

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



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:

$$X_k = [0, 0 - j4, 0, 0, 0, 0, 0, 0 + j4]$$

- 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.

$$|X_k| = [0, 4, 0, 0, 0, 0, 0, 4]$$

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



Solution

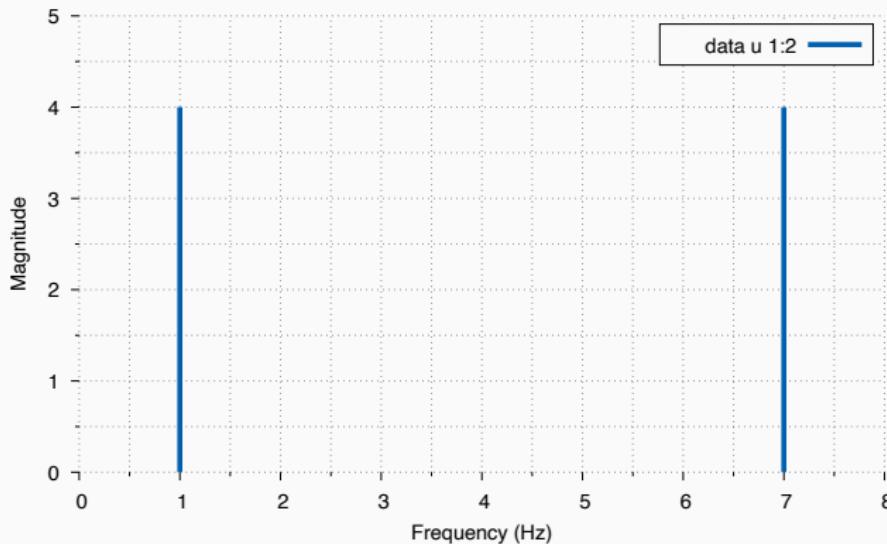


Figure 15: Sampled dataset of the original signal. There is still another step.



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

$$|X_k| = [0, 8, 0, 0] \rightarrow |X_k| / N = [0, 1, 0, 0] \blacksquare$$



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:

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

- 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**):

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

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 a 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, $\text{med}(x) = x(n+1)/2$,
 - if n is even, $\text{med}(x) = x(n/2) + x((n/2)+1)/2$
- 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.



- 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, $a_{\min} \leq a \leq a_{\max}$, then the SNR is defined as:

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

- 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 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.



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

$$p(x_h) = 0.9.$$

- When this coin is flipped, we expect it to land heads up ($x = x_h$),
- When it does, we are less surprised than when it lands tails ($x = x_t$).

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.

$$\text{Shannon Information} = \log_2 \frac{1}{p(x_h)}$$



Entropy is Average Shannon Information

- We can represent the outcome of a coin flip as the random variable x , such that a head is $x = x_h$ and a tail is $x = x_t$.
- 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 $H(x)$.

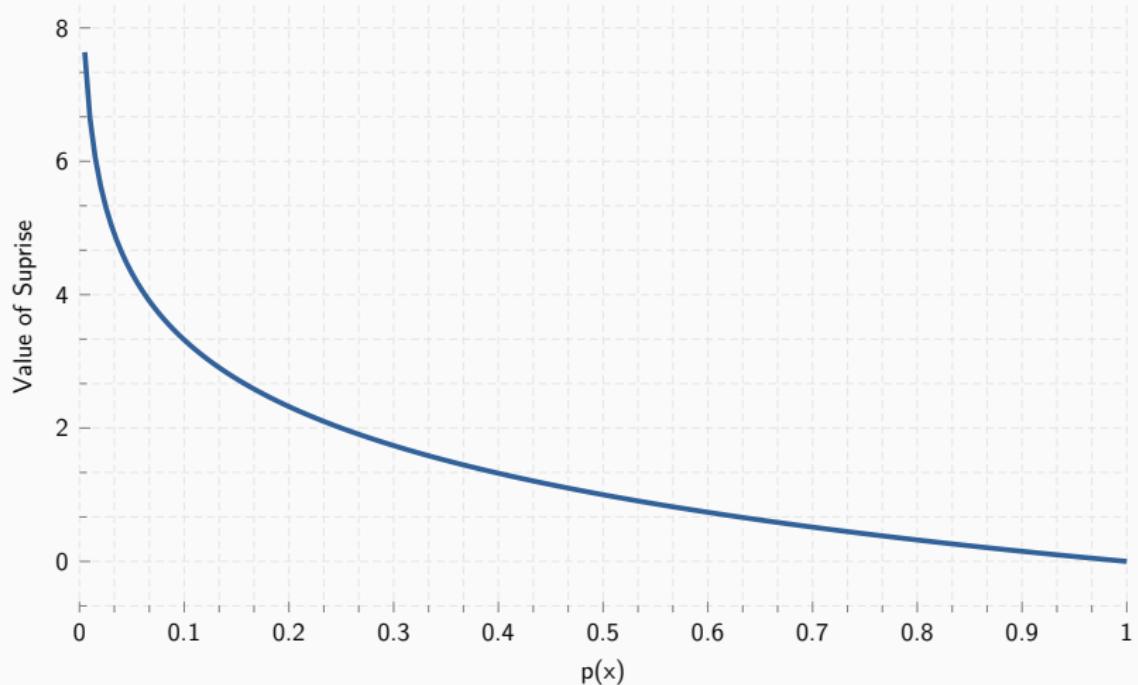


Figure 16: The quantifiable surprise with respect to increasing probability.



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 $p(x_h) = 0.9$.
 - 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 $p(x_h)$, and that the proportion of flips that yield a tail is $p(x_t)$ (where $p(x_h) + p(x_t) = 1$), 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:

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



- More generally, a random variable x with a probability distribution $p(x) = p(x_1), \dots, p(x_m)$ 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.

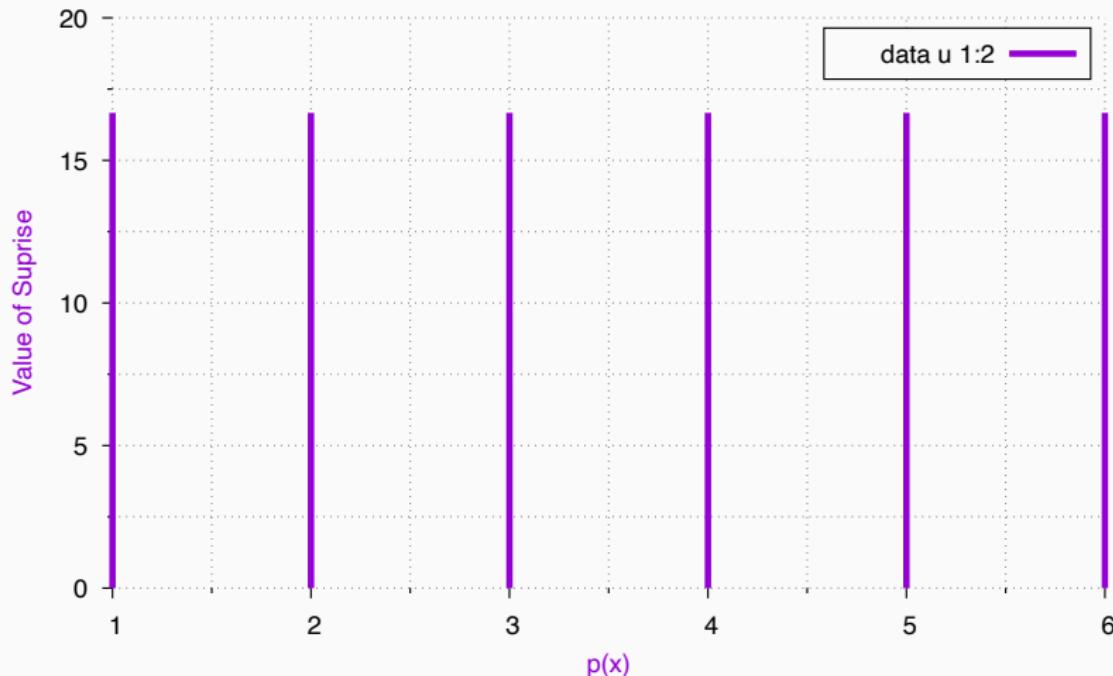


Figure 17: The probability distribution of 1 dice(s).

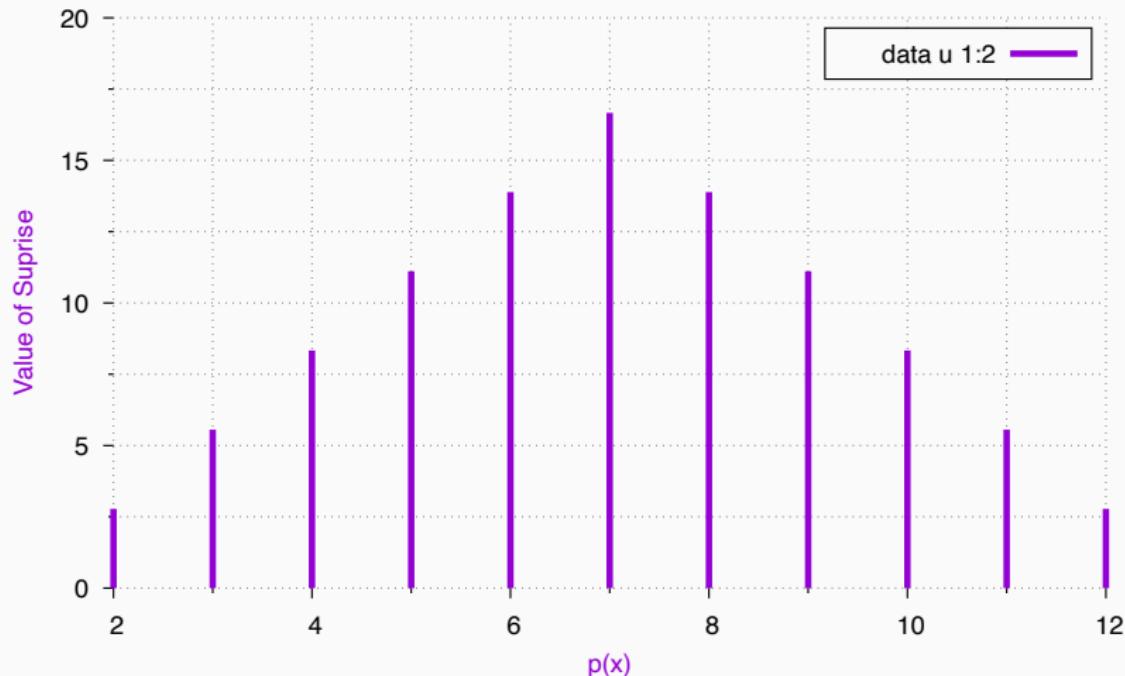


Figure 18: The probability distribution of 2 dice(s).

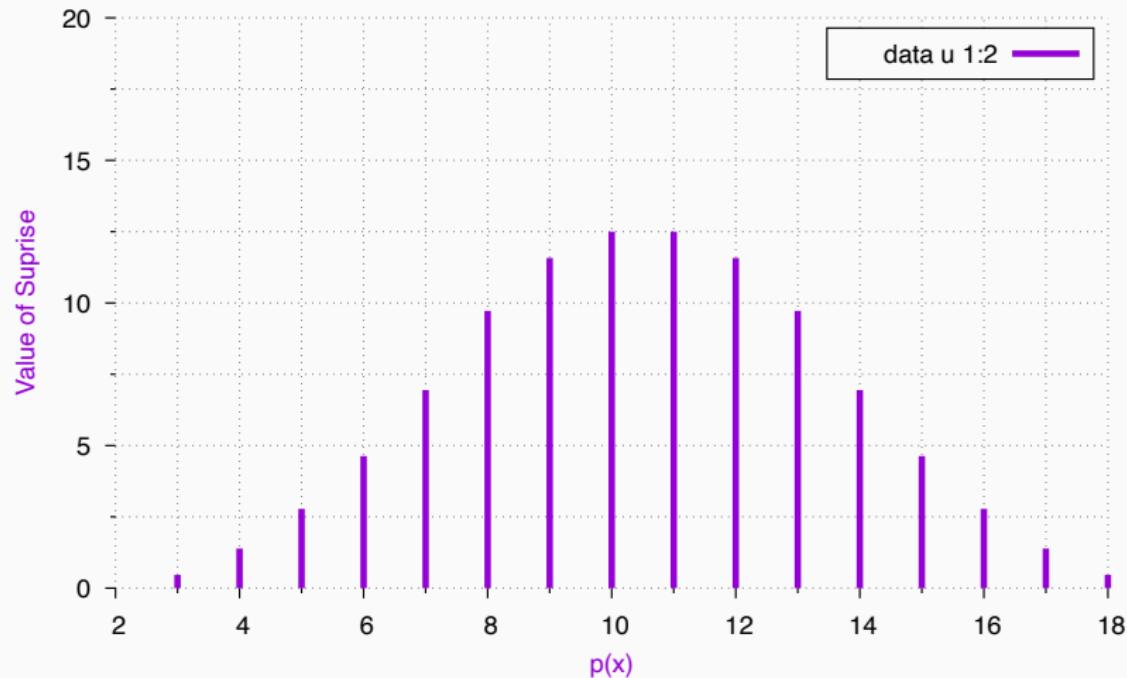


Figure 19: The probability distribution of 3 dice(s).

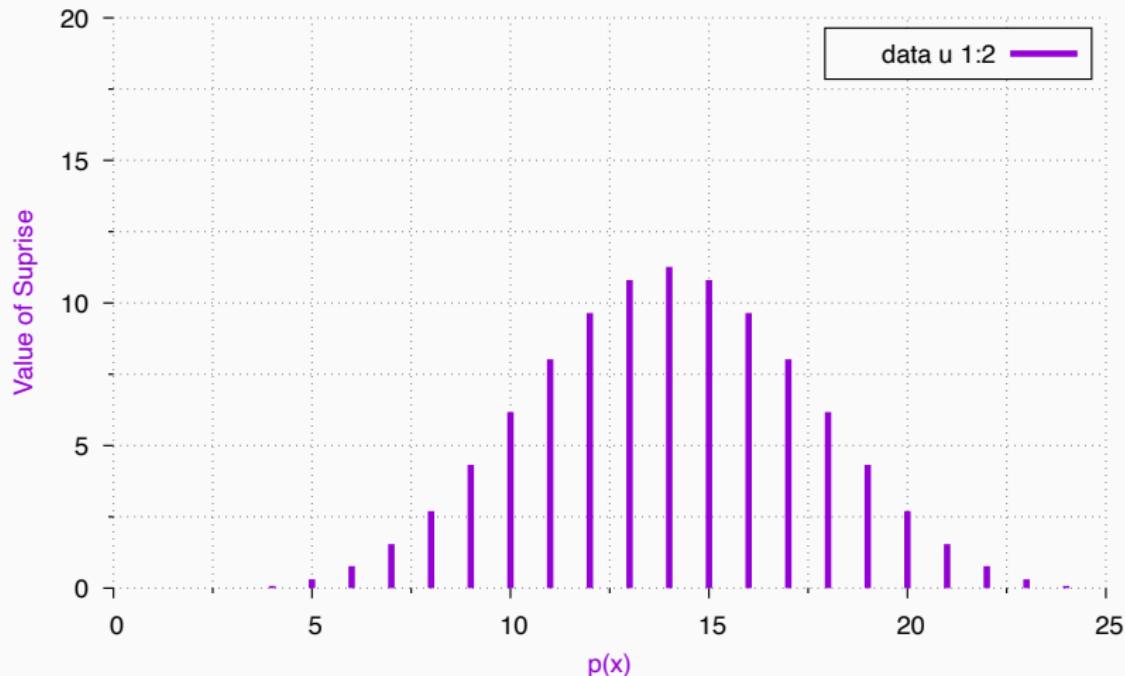


Figure 20: The probability distribution of 4 dice(s).



- 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.



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

$$\begin{aligned}H(x) &= p(x_1) \log \frac{1}{p(x_1)} + p(x_2) \log \frac{1}{p(x_2)} + \cdots + p(x_{11}) \log \frac{1}{p(x_{11})} \\&= 3.27 \text{ bits.}\end{aligned}$$

Perception



Learning Outcomes	Wide Gamut RGB
Introduction	Prophoto RGB
Human Vision	Adobe RGB
Brightness Sensitivity	CIE Chromaticity Coordinates
Stimulus Sensitivity	Chromaticity
Colour Sensitivity	
Colour Standards	Colour Models
sRGB	CYMK Colour Model
	HSL and HLV Colour Model
	YCbCr



Learning Outcomes

- (LO1) A Look into Human Vision,
- (LO2) Definition of Colour and Standardisation,
- (LO3) How Vision is Perceived,
- (LO4) Types of Colour-spaces.





- 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 [11].
- At the outset it is important to realise:
 1. The human visual system is **not well understood** [11].
 - 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**.

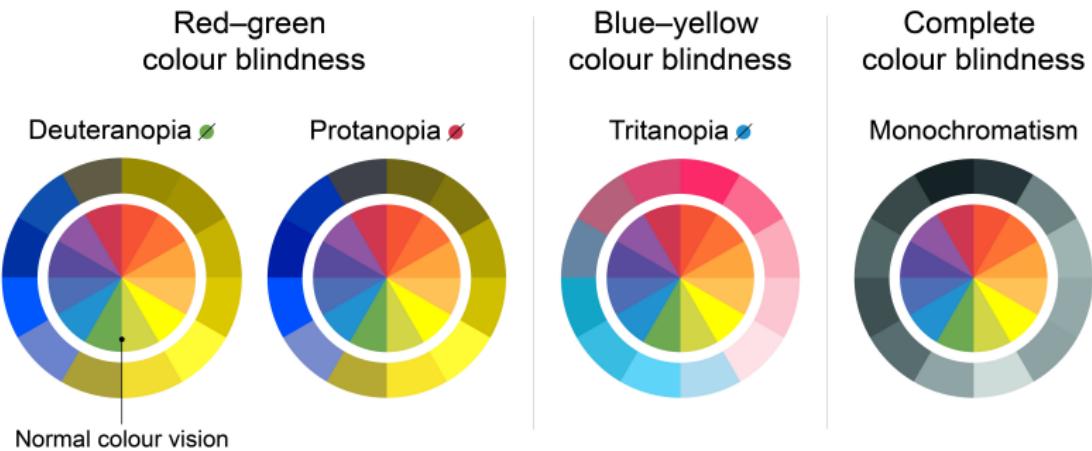
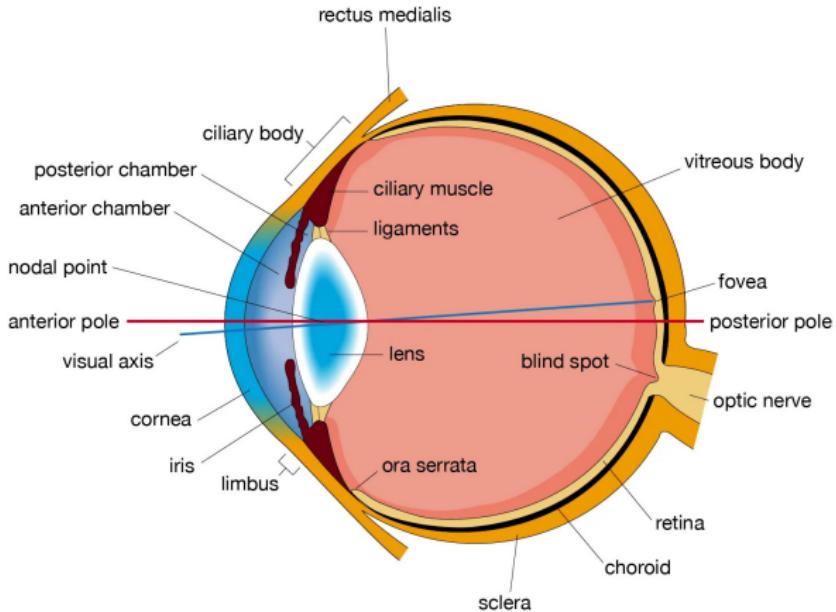


Figure 21: Colour-wheel of different kinds of blindness [24].



- Statistically 1 in 8 men and 1 in 200 women have a form of colour blindness [5].
- Men have one X chromosome and one Y chromosome, while women have two X chromosomes [4].
- To experience color blindness, the genetic mutation for colorblindness must be present on the X chromosome, but for women, this means it must be present on both X chromosomes.
- Men only need a mutation to be present on their singular X chromosome, making it much easier for them to inherit color blindness.



© 2013 Encyclopædia Britannica, Inc.

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 [5].
- 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 (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 [26]:
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 [27].
- 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 [27].



Figure 23: Image viewed by someone who has 3 cones.

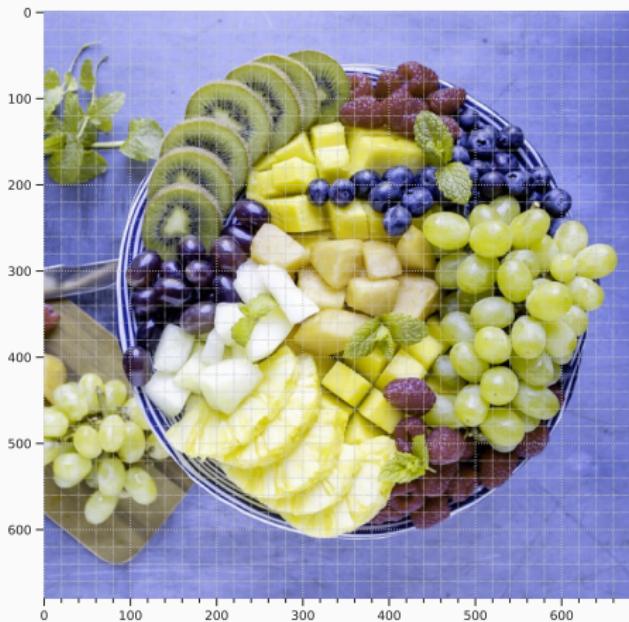


Figure 24: Image viewed by someone who has protanomaly.

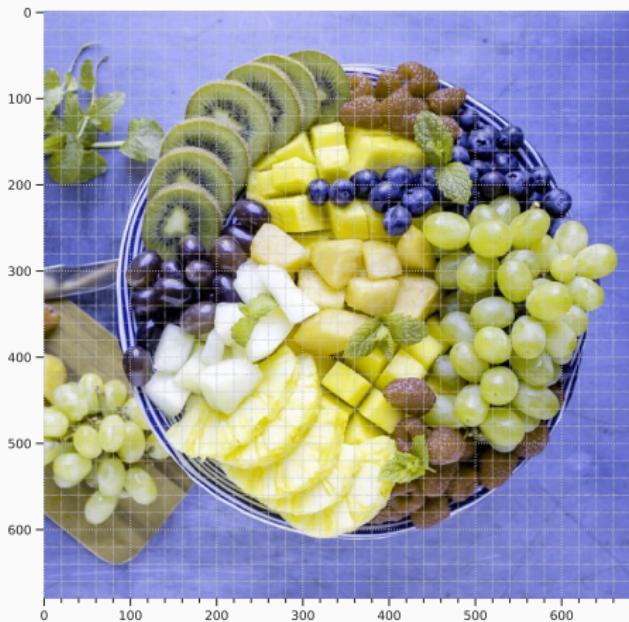


Figure 25: Image viewed by someone who has deuteranomaly.

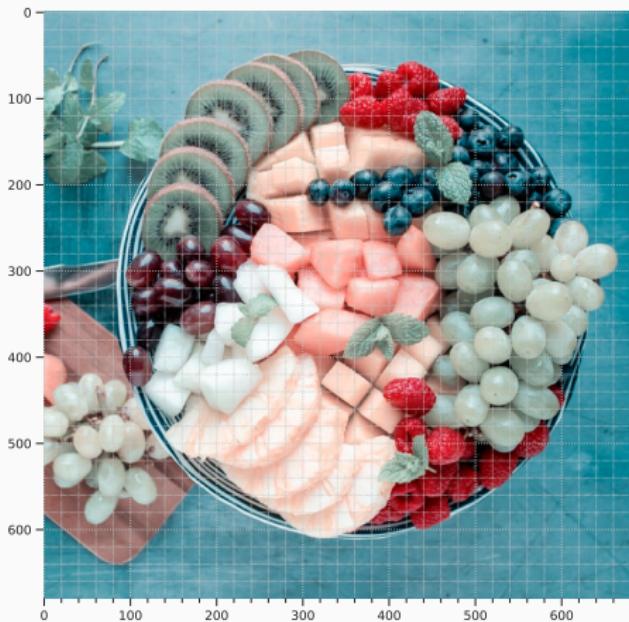


Figure 26: Image viewed by someone who has tritanomaly.

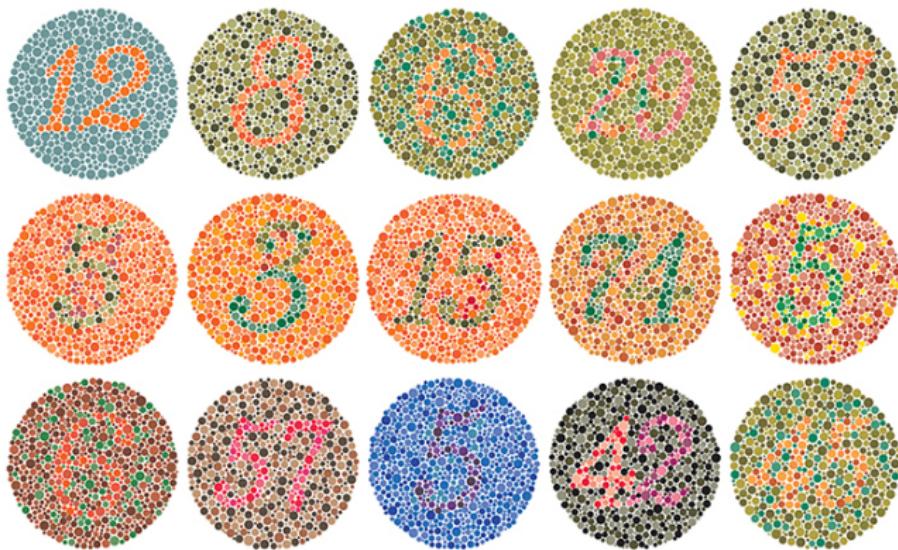


Figure 27: A colour blindness test issued to generally test before taking the driving license [3].



- 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 [23].
- 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 [23].



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

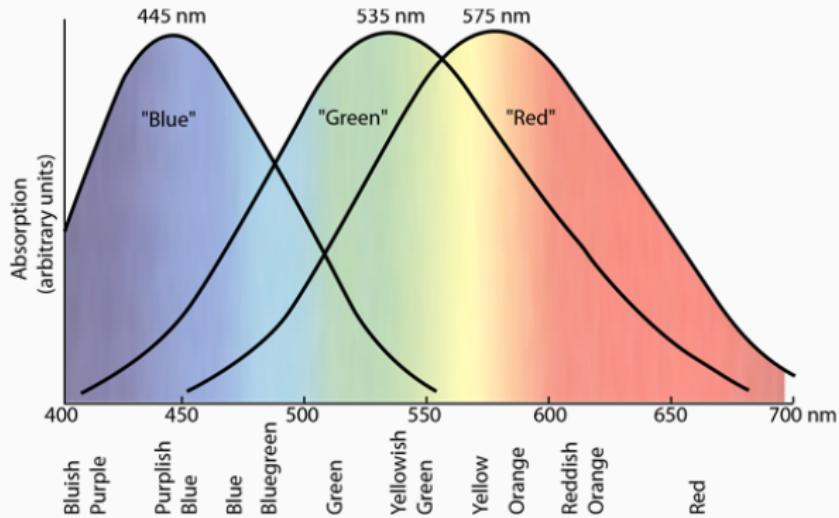


Figure 28: The colour sensitivity of the human eye [25].



The Visible Spectrum - CIE 1931 2° Standard Observer

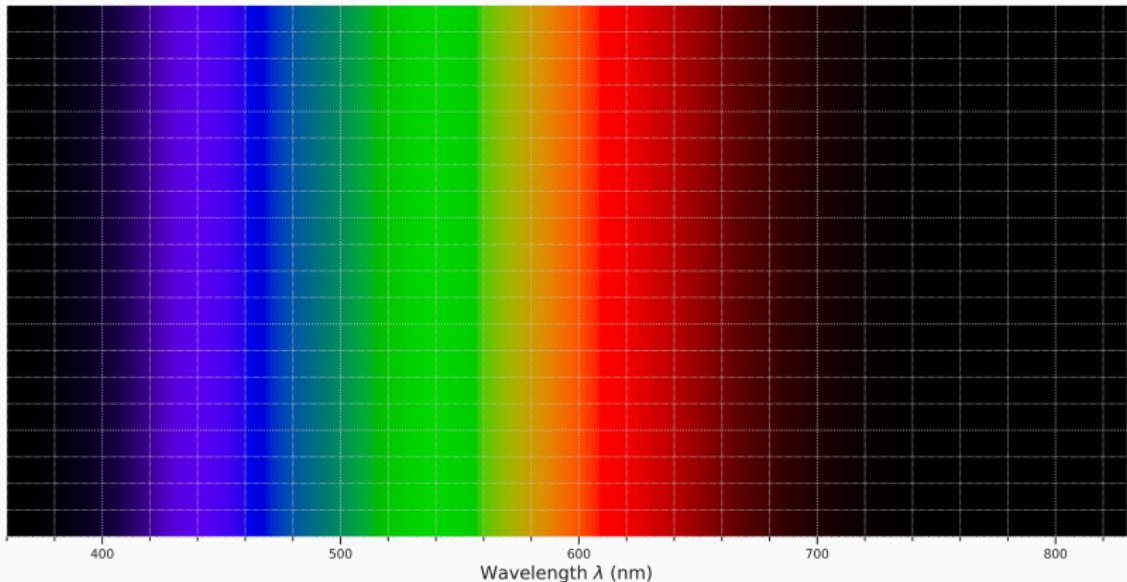


Figure 29: The visible colour spectrum visible with the human eye.



- 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** [12].

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

$$R = \log(I_0)$$

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

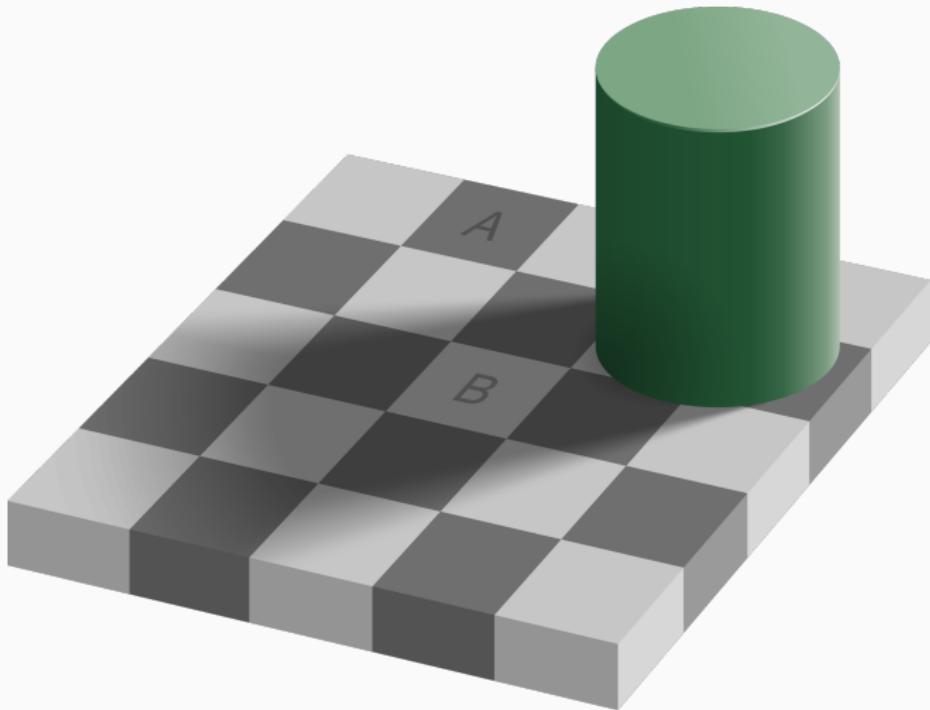


Figure 30: The checker shadow illusion [1].

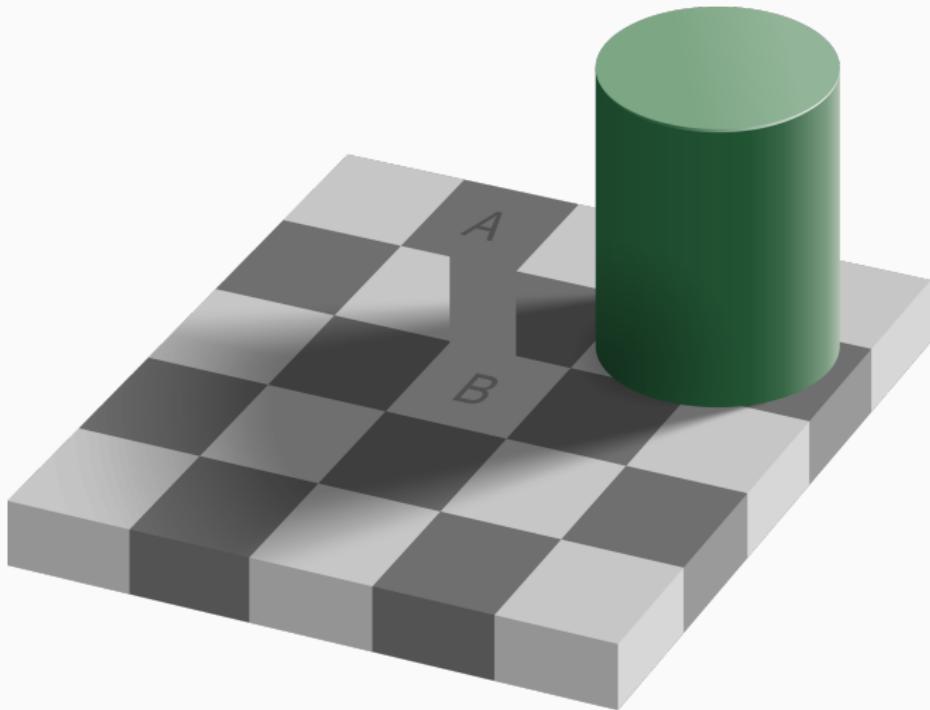


Figure 31: A region of the same shade has been drawn connecting A and B.



Figure 32: A simple picture of a dress divided the internet.



Figure 33: A simple picture of a dress divided the internet [6].



- 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 CIE¹ 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.

¹Commission Internationale de l'Eclairage

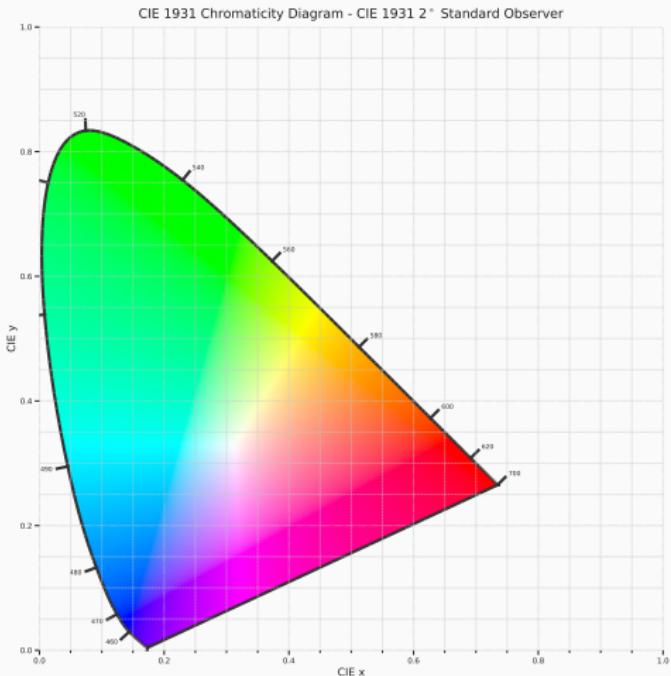


Figure 34: The colour gamut visible to the human eye, standardised by the CIE.



- sRGB is a standard RGB colour-space HP and Microsoft created cooperatively in 1996 to use on monitors, printers, and the World Wide Web [37].
- It was subsequently standardised by IEC as IEC 61966-2-1:1999 [31].
- 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 35: 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 [39].
- However, consumer-level CCDs¹ 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.

¹Charge-Coupled Device



- 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 [35].
- 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% [42].



Figure 36: 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 [36].
- 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 [46].
- 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 visible colours are imaginary colors that do not exist and are **impossible colour**.

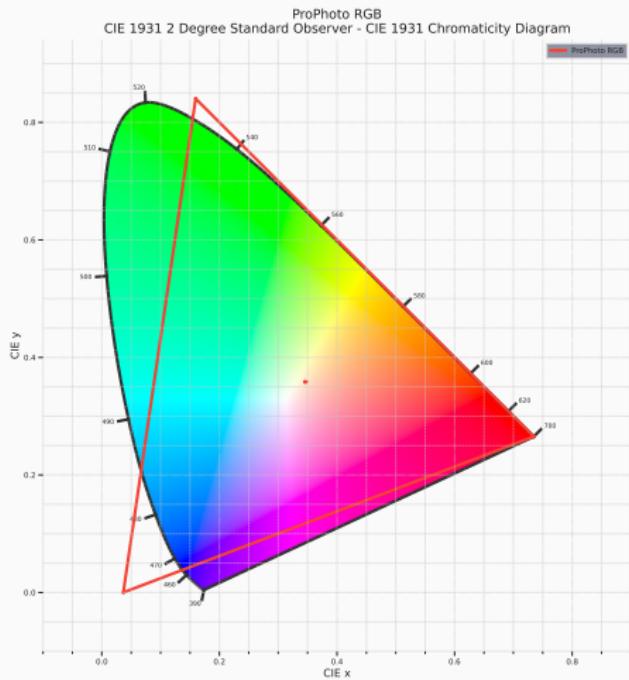


Figure 37: The ProPhoto colour-space superimposed to the CIE colour-gamut.



- 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 subsequently standardised by the IEC as IEC 61966-2-5:1999 with a name opRGB (optional RGB color space) and is used in HDMI [32].



- 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**:

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

- 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:

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

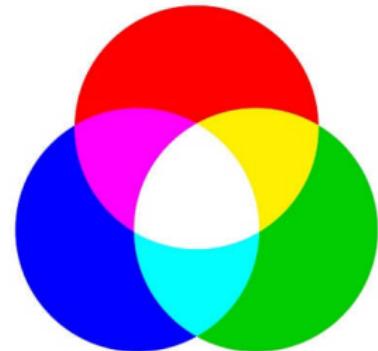
- Before we end our look on to these colour spaces, lets have a look at two (2) more standards.



- 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.



RGB



CMYK

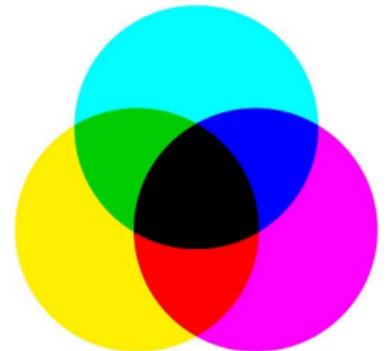


Figure 38: The differences between RGB and CMYK colours [40].

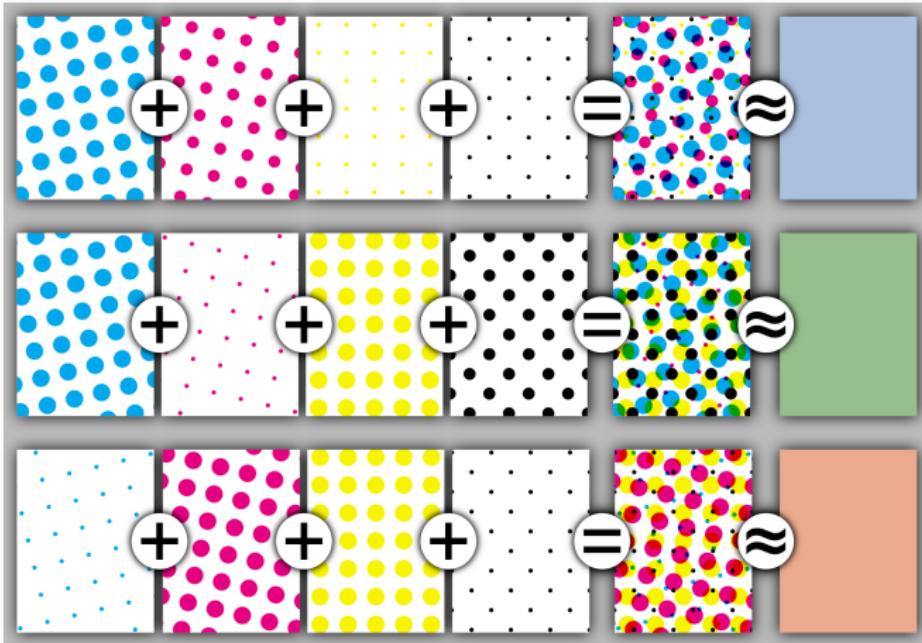


Figure 39: Three examples of color halftoning with CMYK separations, as well as the combined halftone pattern and how the human eye would observe the combined halftone pattern from a sufficient distance [44].

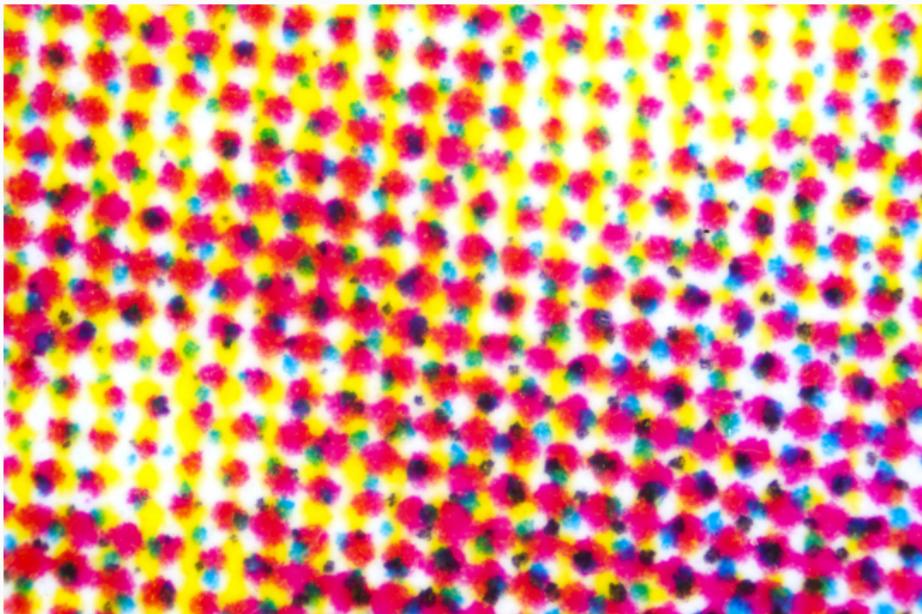


Figure 40: A printer creates any colour by combining dots in particular places relative to the other dots.



- 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.

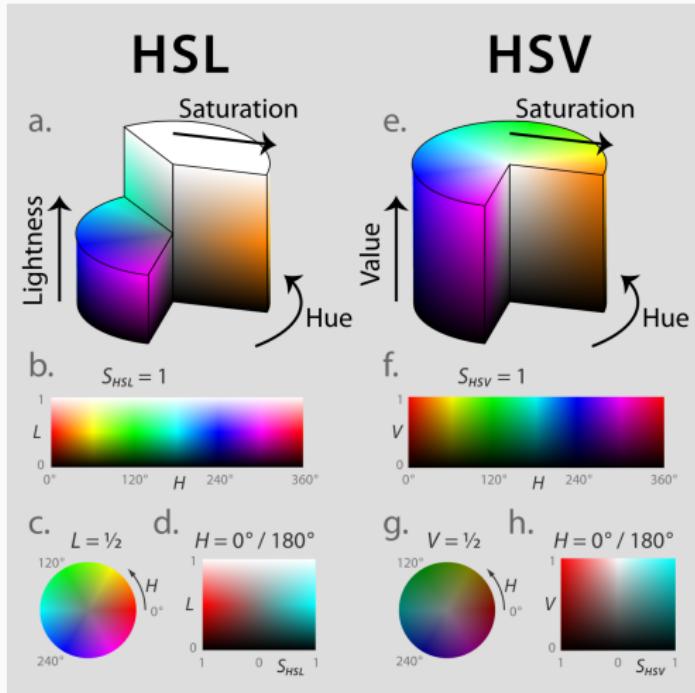


Figure 41: HSL and HSV models.



- 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.



- CRT¹ 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.

¹Cathode Ray Tube

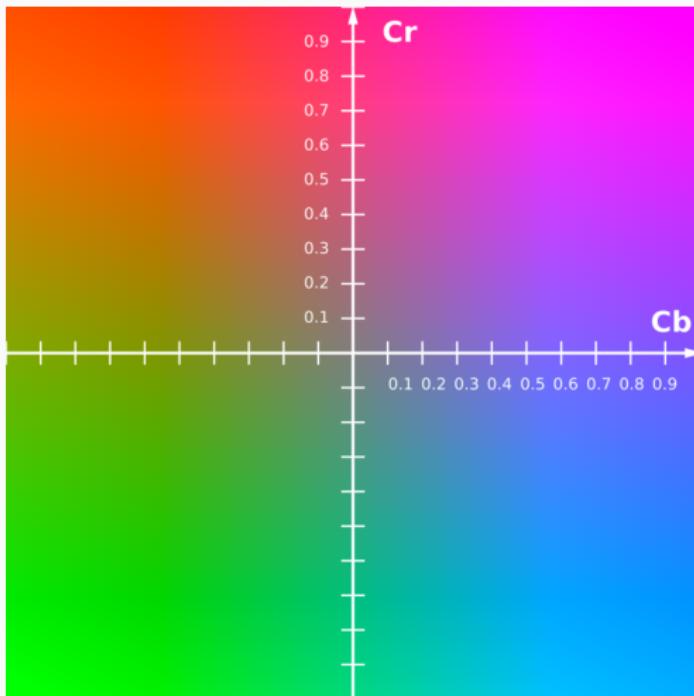


Figure 42: The $Y'C_bC_r$ plane at constant luma $Y=0.5$ [16].

Image Formats



Table of Contents

Learning Outcomes	JPEG
Introduction	JPEG Compression
File Formats	Down-sampling
A General Overview	Isolate the Colour Information
Compression Methods	Throw Away some Colour Information
Lossy Compression	Convert Image to Frequency Domain
Important File Formats	Quantisation
RAW File	Lossless Data Compression



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.



Image Formats

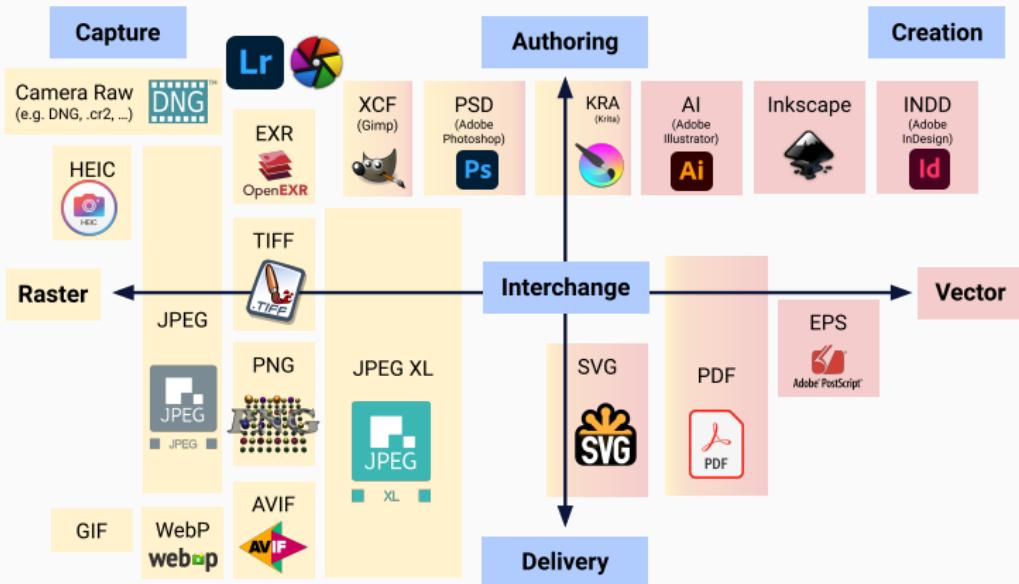


Figure 43: Categorization of image formats by scope which are used in commercial, industrial and personal use [45].



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.



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.



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.



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.



- 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.



- A class of data compression that allows the original data to be **perfectly** reconstructed from the compressed data with no loss of information.
- Lossless compression is possible because most real-world data exhibits statistical redundancy [43].
- In essence, lossless compression algorithms are needed in cases that require compression where we want the reconstruction to be identical to the original.
- Huffman Coding is a perfect example of lossless compression [29].



- 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` .



- 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 44: A photo of a European wildcat with the compression rate, and associated losses, decreasing from left to right [22].



- 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.



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.



- 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 is 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 2×0.25 channels, for a total of 1.5.
- We are already down to half of the information we started with.



- 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.



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



- 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.



- 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.



- 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.



- 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.

Camera



Learning Outcomes	Signal-to-Noise Ratio
Camera Sensors	Thermal Noise (Dark Current)
Charge Coupled Devices	Quantum Efficiency
Complementary MOS	ISO
CCD v. CMOS	Camera Lenses
Colour Filters	Introduction
Camera Properties	Aperture
Linearity	Types of Lenses
Sensitivity	Prime Lens



Learning Outcomes

- (LO1) An Overview of Camera Sensors,
- (LO2) Camera as a Scientific Instrument,
- (LO3) Working with Colour Filters,
- (LO4) Lens Technologies.





- The cameras and recording media available for modern digital image processing applications are constantly changing.
- In this section we will focus on two (2) technologies:
 - CCD** Charge Coupled Device,
 - CMOS** Complementary Metal-Oxide-Semiconductor,
- The techniques that are used in these technologies are **universal**.

- There are also variations of this technologies which are more recent:
 - EMCCD** Electron-multiplying charge-coupled device [41],
 - sCMOS** Back-illuminated CMOS [34].



- First sensors used in digital cameras, available since 1970s [48].
- A Charge Coupled Device (CCD) is a **highly sensitive photon detector**.
- Divided up into a large number of light-sensitive small areas (known as pixels) which can be used to build up an image.
- A photon of light which falls within the area defined by one of the pixels will be converted into one (or more) electrons and the number of electrons collected will be directly proportional to the intensity of the scene at each pixel.

This allows the camera to be quantitative.

- When the CCD is clocked out, the number of electrons in each pixel are measured and the scene can be reconstructed.

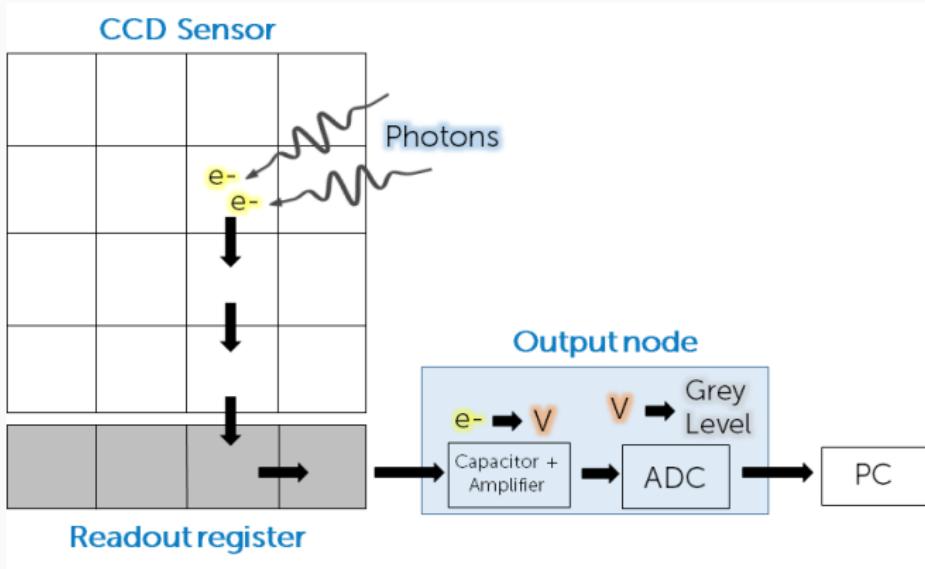


Figure 46: How a CCD sensor works. Photons hit a pixel and are converted to electrons, which are then shuttled down the sensor to the readout register, and then to the output node, where they are converted to a voltage, then grey levels, and then displayed with a PC.
[48]



Limitations

- The main issues with CCDs are their lack of speed and sensitivity, making it a challenge to perform low-light imaging or to capture dynamic moving samples.

Sensor millions of pixels of signal have to be shuttled through one node, creating a bottleneck and slowing the camera.

Noise to control the read noise from fast moving electrons, CCDs slows down the electrons.

Refresh Whole sensor needs to be cleared of the electron signal before the next frame can be exposed

- Small full-well capacity limits electron storage in each pixel.
- If a pixel becomes full and displayed the brightest signal, and blooming, where the pixel overflows and the excess signal is smeared down the sensor as the electrons are moved to the readout register.

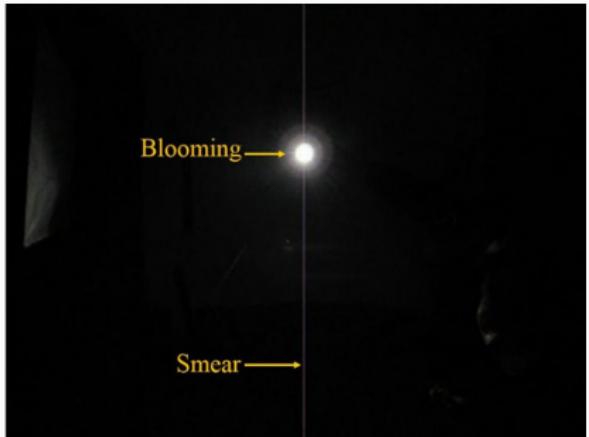


Figure 47: Examples of blooming caused by saturation of a CCD sensor pixel. Left) Picture of a sunset. The sun is so bright in the image that there is blooming on the sun itself, leaking into the surrounding pixels, and a vertical smear across the whole image. Right) A similar situation with the blooming and smear labeled [48].



- The CMOS image sensor is made up of an array of tiny light-sensitive cells also known as pixels, each of which is connected to a transistor that acts as a switch [20].
- When light strikes the pixels, it is converted into an electrical charge which is amplified and read out by the sensor's readout electronics.
- The output is then converted into a digital signal, ready for further image processing or storage.

Emerged as an alternative to charge-coupled device (CCD) image sensors and eventually outsold them by the mid-2000s [19].

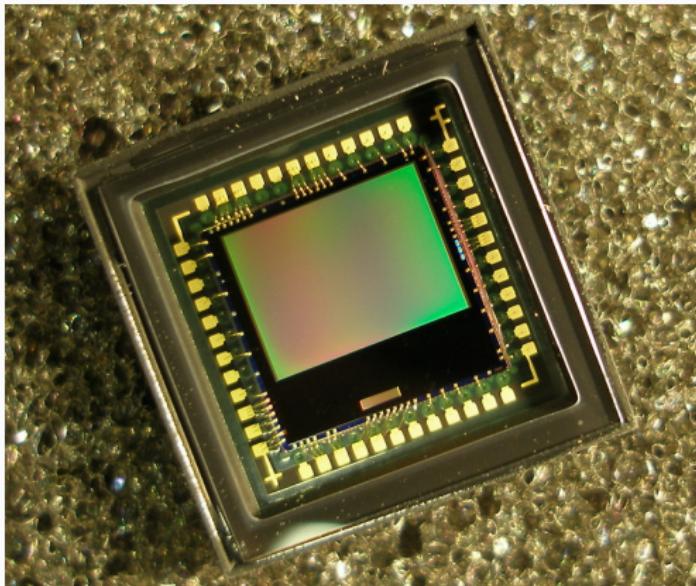


Figure 48: CMOS image sensor [17].

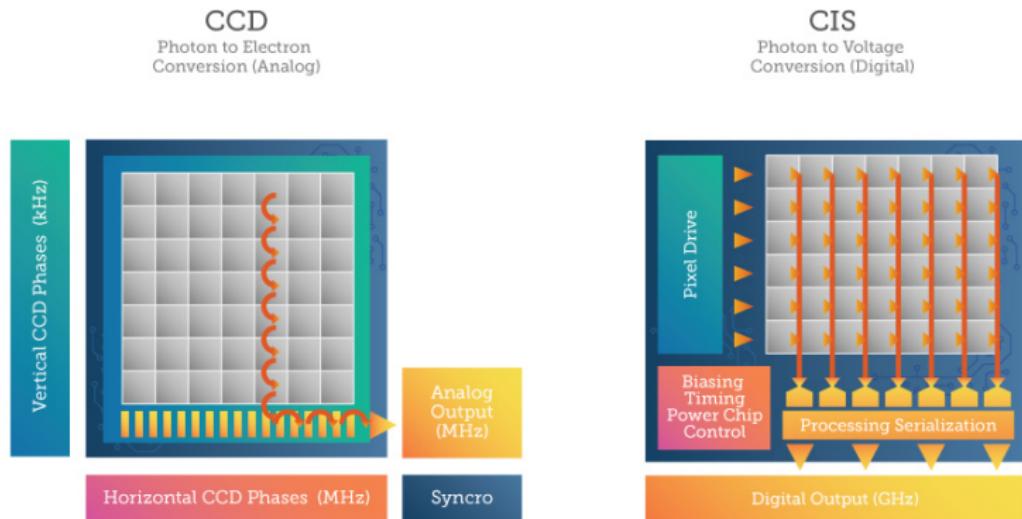


Figure 49: A comparison on the working principle of two types of sensors: CCD and CMOS [21].



The first difference lies in the way the image data is captured/read.

- A CCD captures image data by moving charge packets in a linear sequence from one pixel to the next, which is referred to as **charge transfer**.
- This method results in improved image quality, but it also has drawbacks such as a slower readout speed (i.e., pixels per second).
- On the other hand, CMOS capture image data by reading out each pixel individually,
 - a process known as **parallel readout**.

This results in a faster readout time but generally lower image quality compared to CCD sensors.



- CMOS is cheaper to produce than CCD due to their more complex electronic circuitry and manufacturing process.
- CCD is known to be more power-hungry as compared to CMOS.
 - CCD needs active power as opposed to the CMOS which don't.
- CMOS has a distinct advantage as readout circuits for each pixel is integrated on the same chip as the pixel.
 - This allows for a wide range of enhanced functions such as image processing and signal amplification.

CCDs are known for their high image quality, lower noise, and greater light sensitivity while CMOSs have the advantage of being faster, less power-hungry and more cost-effective.



- A color filter array (CFA), or color filter mosaic (CFM), is a mosaic of tiny color filters placed over the pixel sensors of an image sensor to capture color information.
- Color filters are needed because the typical photosensors detect light intensity with little or no wavelength specificity and therefore cannot separate color information [33].

Diazonaphthoquinone (DNQ)-novolac photoresist is one material used as the carrier for making color filters from color dyes or pigments [28, 9].

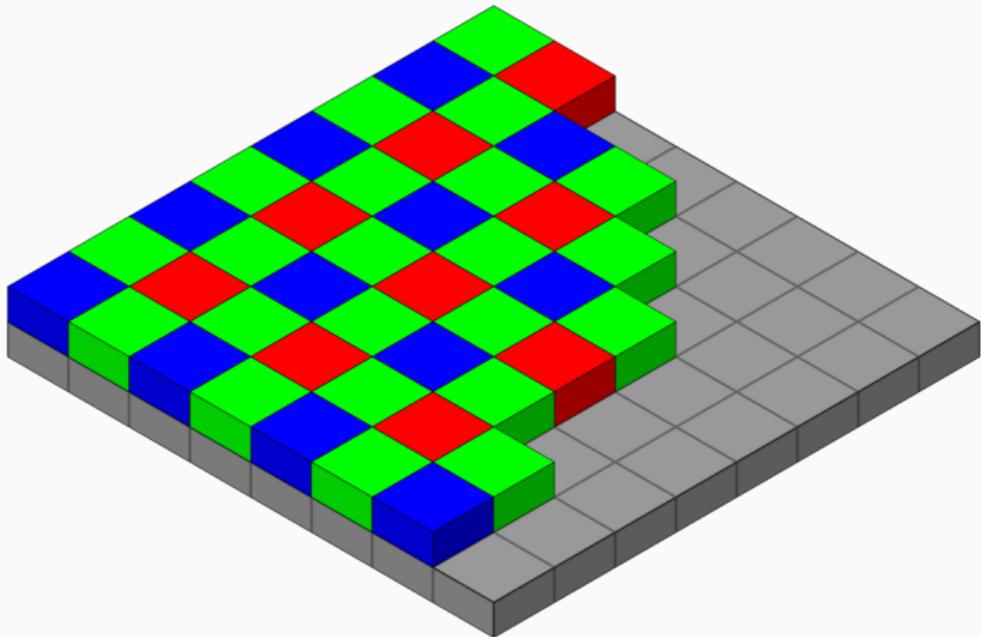
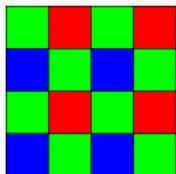
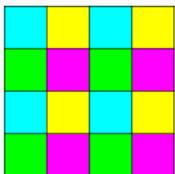


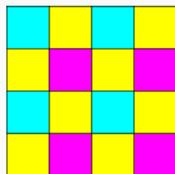
Figure 50: The Bayer color filter mosaic. Each two-by-two submosaic contains 2 green, 1 blue, and 1 red filter, each filter covering one pixel sensor. The reason as to why there are two greens to red and blue is to mimic the physiology of the human eye [7].



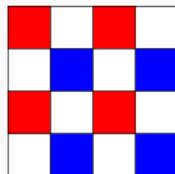
(a) Bayer Filter



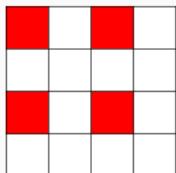
(b) CYGM Filter



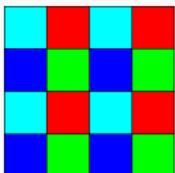
(c) CYYM Filter



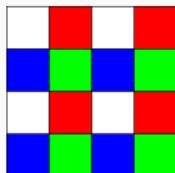
(d) RCCB Filter



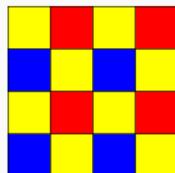
(e) RCCC Filter



(f) RGBE Filter



(g) RGBW Filter



(h) RYYB Filter

Figure 51: Types of colour filter used in commercial and industrial applications



Figure 52: Sony F828, produced in 2003 which uses an RGBE filter. Once the image is taken a demosaicing algorithm is used to reconstruct an image [18].



- It is generally desirable the relationship between the input physical signal and the output signal be linear [50].
- This means that if we have two images, a and b :

$$\mathcal{C} = \mathcal{R}\{w_1a + w_2b\} = w_1\mathcal{R}\{a\} + w_2\mathcal{R}\{b\}.$$

where $\mathcal{R}(\cdot)$ is the camera response and \mathcal{C} is the camera output.

- In practice the relationship between a and \mathcal{C} is given by:

$$\mathcal{C} = \text{Gain} \cdot a^\gamma + \text{Offset}.$$

where γ is the gamma of the recording medium.

- A linear system must have $\gamma = 1$ and Offset = 0.
- Unfortunately, Offset is never zero and must be compensated for if the intention is to extract intensity measurements.



- There are two (2) ways to describe the **sensitivity** of a camera.
 1. We can determine the minimum number of detectable photoelectrons called **absolute sensitivity**.
 2. We can describe the number of photoelectrons necessary to change from one digital brightness level to the next called **relative sensitivity**.



Figure 53: Types of camera lenses.



Absolute Sensitivity

- We need a characterisation of the camera in terms of its **noise**.
- If total noise has a σ of, 100 photoelectrons, then to ensure detectability of a signal we could say that, at the 3σ level (% 99.7), the minimum detectable signal (or absolute sensitivity) would be 300 photoelectrons.
- If all the noise sources, with the exception of photon noise, can be reduced to negligible levels, this means that an absolute sensitivity of less than 10 photoelectrons.
- This can be rephrase to the **number of photons needed to have signal equal to noise** [47].
- This quantity plays an important role in very-low light applications.



Relative Sensitivity

- When coupled to the linear case, with $\gamma = 1$, leads to the result:

$$\text{Relative Sensitivity} = 1/\text{gain}.$$

The sensitivity or gain can be deduced in two (2) distinct ways.

- If a can be precisely controlled by either **shutter time** or intensity, the gain can be estimated by estimating the slope of the resulting straight-line curve.
 - To translate this to units, a standard source must be used emitting photons to the camera sensor and the quantum efficiency (η) of the sensor must be known.
- If the limiting effect of the camera is only the **photon noise**, then:

$$S = \frac{\mu}{\sigma^2} = \frac{m_c}{s_c}$$



Camera	Pixels	Pixel Size	Temp.	S	Bits
Label		$\mu\text{m} \times \mu\text{m}$	K	e^-/ADU	
C-1	1320×1035	6.8×6.8	231	7.9	12
C-2	578×385	22.0×22.0	227	9.7	16
C-3	1320×1035	6.8×6.8	293	48.1	10
C-4	576×384	23.0×23.0	238	90.9	12
C-5	756×581	11.0×5.5	300	109.2	8

Table 3: Sensitivity measurements of sample cameras.

- In a scientific-grade CCD camera (C-1), only 8 photoelectrons (approximately 16 photons) separate two gray levels in the digital representation of the image.
- For a considerably less expensive video camera (C-5), only about 110 photoelectrons (approximately 220 photons) separate two gray levels.



- In a modern camera (CCD or CMOS), the noise is defined by:
 - Amplifier noise in the case of colour cameras.
 - This is more apparent in blue channel.
 - Thermal noise which, itself, is limited by the chip temperature K and the exposure time T ,
 - Photon noise which is limited by the photon production rate ρ and the exposure time T .

We will have a detailed look into these in our lecture called **Noise**.



- Caused by the thermal energy within the camera sensor [49].
- As sensor works, the heat from operation causes thermal energy generation.

These thermal electrons are independent of the photo-electrons.

- Cameras can't decide if the electrons are from the light or thermal.
- Any thermal electrons that accumulated in the sensor pixel wells are counted as signal upon readout, despite **not being part of the signal from the sample**.
- This is known as the dark current.

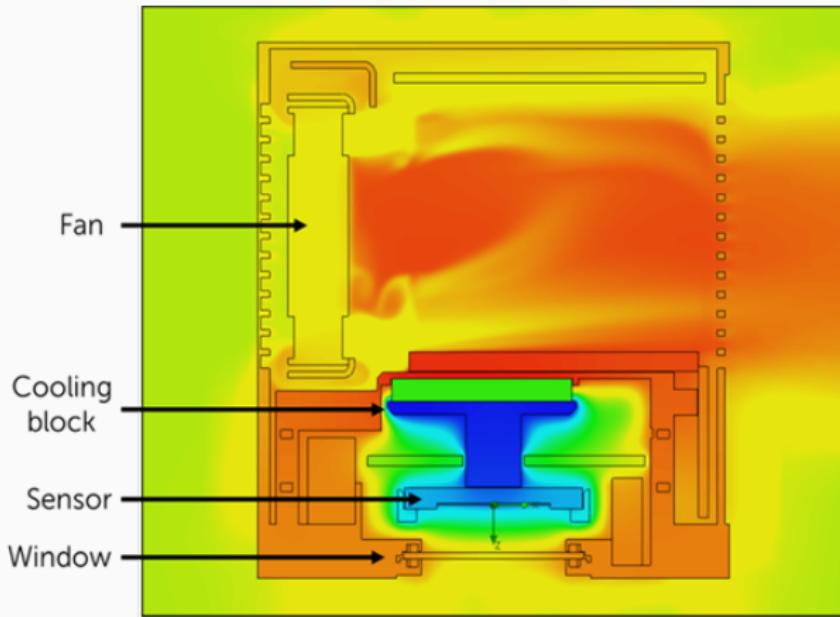


Figure 54: Scientific camera cooling and Citadel Chamber Technology. It shows typical temperature levels within a CMOS camera. The sensor is cooled absolutely and evenly, is an optimal distance from the window, and heat is effectively expelled (edited from [49]).



Camera	Dark Current	Data Sheet
Prime 95B sCMOS (Normal Operation)	0.55	
Prime 95B sCMOS (Air-Cooled)	0.3	
Retiga R6 CCD	0.00073	
Retiga E7 CMOS	0.001	
Nikon D5300 (ISO 200)	6.23	

Table 4: Values of Dark Current on Cameras.



- Quantum efficiency (QE) is the measure of the **effectiveness** to convert incident photons into electrons.

For example, if a sensor had a QE of 100% and was exposed to 100 photons, it would produce 100 electrons of signal.

- In practice, sensors are never 100% efficient, and different sensor technologies have different QE values.

The highest-end scientific cameras can achieve up to 95% QE but this is dependent on the wavelength of light being detected.

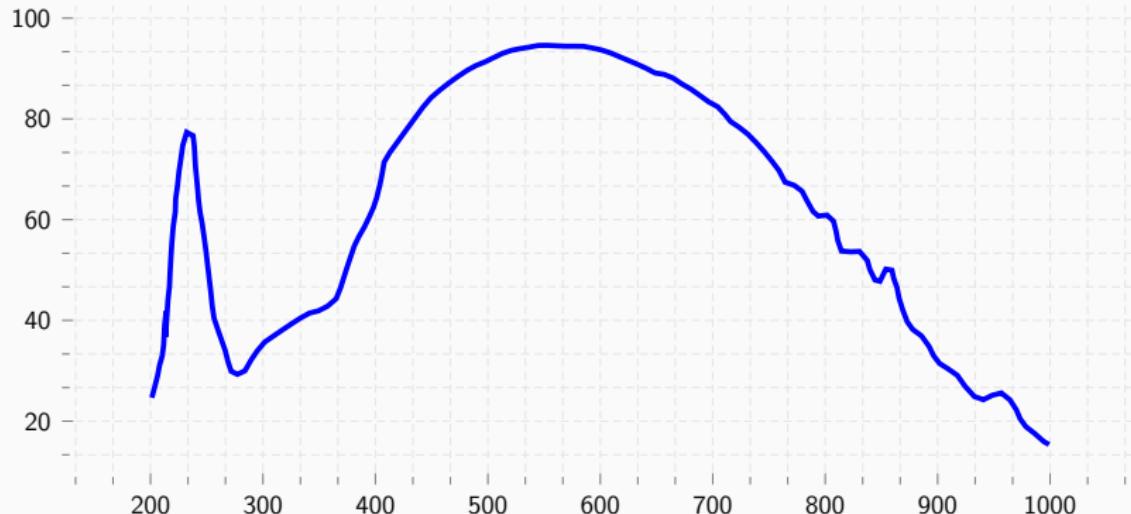


Figure 55: The quantum efficiency (QE) of a 95% quantum efficient sensor at different photon wavelengths. 95% QE is possible at 500-600 nm wavelengths (green/yellow) but it is less efficient at shorter (violet, 300-400 nm) and longer (infrared, 800-1000 nm) wavelengths. This particular sensor also has a peak in QE at near-UV wavelengths, around 220-250 nm.



- ISO is a camera setting which brightens or darkens a photo.
- Increasing the ISO number, makes photos progressively brighter.
- For that reason, ISO can help you capture images in darker environments, or be more flexible about your aperture and shutter speed settings.

Raising your ISO has consequences. A photo taken at too high of an ISO will show a lot of grain, also known as noise, and might not be usable.

- Brightening a photo via ISO is always a trade-off.
- Only raise your ISO when you are unable to brighten the photo via shutter speed or aperture instead.



Figure 56: A comparison of different ISO numbers. The ISO controls how sensitive the camera is to the light and therefore a higher ISO number would indicate a brighter image [8].



- A typical lens used with an industrial camera is actually a lens system made of multiple types of optical lenses within an enclosure.
- This type of camera lens is technically called a “compound lens.”
- However, a camera lens is the colloquial term for a lens system when it is built with a mounting ring that can fit a variety of cameras.
- Choosing the correct camera lens for a vision system is critical for achieving a specific imaging result.

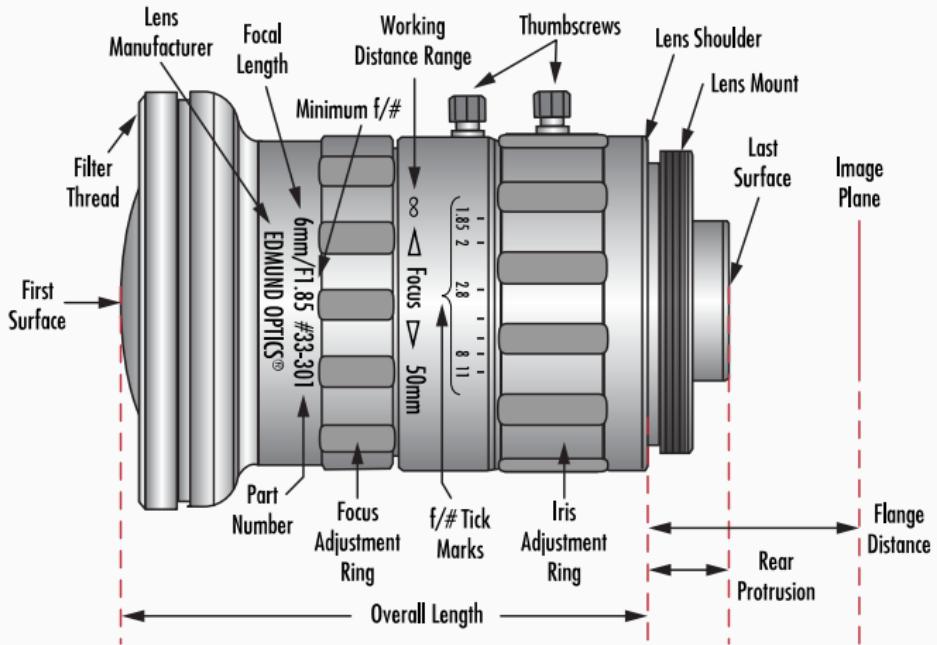


Figure 57: Anatomy of a lens [15].



$f/1.4$



$f/2.0$



$f/2.8$



$f/4.0$



$f/5.6$



$f/8.0$



Figure 58: Different apertures of a lens.



APERTURE

LESS LIGHT



f/22

f/14

f/8

f/5.6

f/4

f/2.8

*"SLOWER"
"SMALLER"*

*"FASTER"
"WIDER"*

MORE LIGHT

Figure 59: The effects of aperture on the image.



Figure 60: The entrance pupil is typically about 4 mm in diameter, although it can range from as narrow as 2 mm ($f/8.3$) in diameter in a brightly lit place to 8 mm ($f/2.1$)



- Choosing the proper lens needs careful consideration **focal length**.
- Focal length is the distance between the camera's sensor and where the light in the camera lens is focused.
- It is often measured in millimetres (mm) and provides plenty of information about what type of image will be captured,
 - such as the angle of view (often referred to as the field of view or FoV)
 - the magnification of the target within the image.



FOCAL LENGTH AND ANGLE OF VIEW

THE DARK ROOM .com

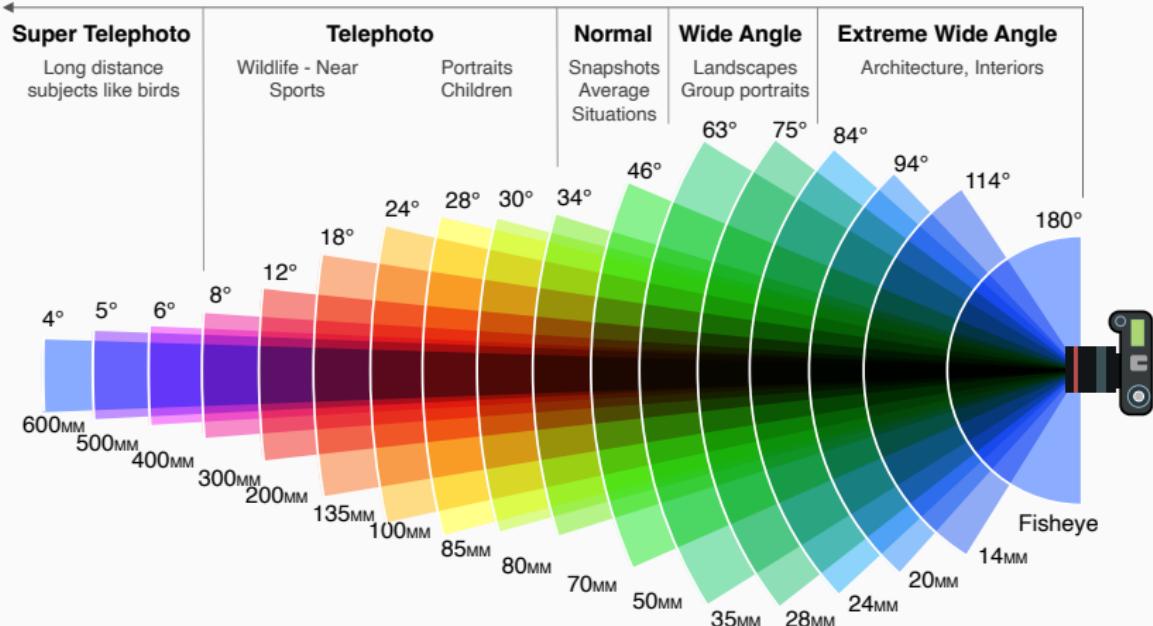


Figure 61: The effects of focal length and the angle of view [10].



- There are two (2) main types of camera lenses:
 1. Prime lens,
 2. Zoom lens (also called varifocal/kit)

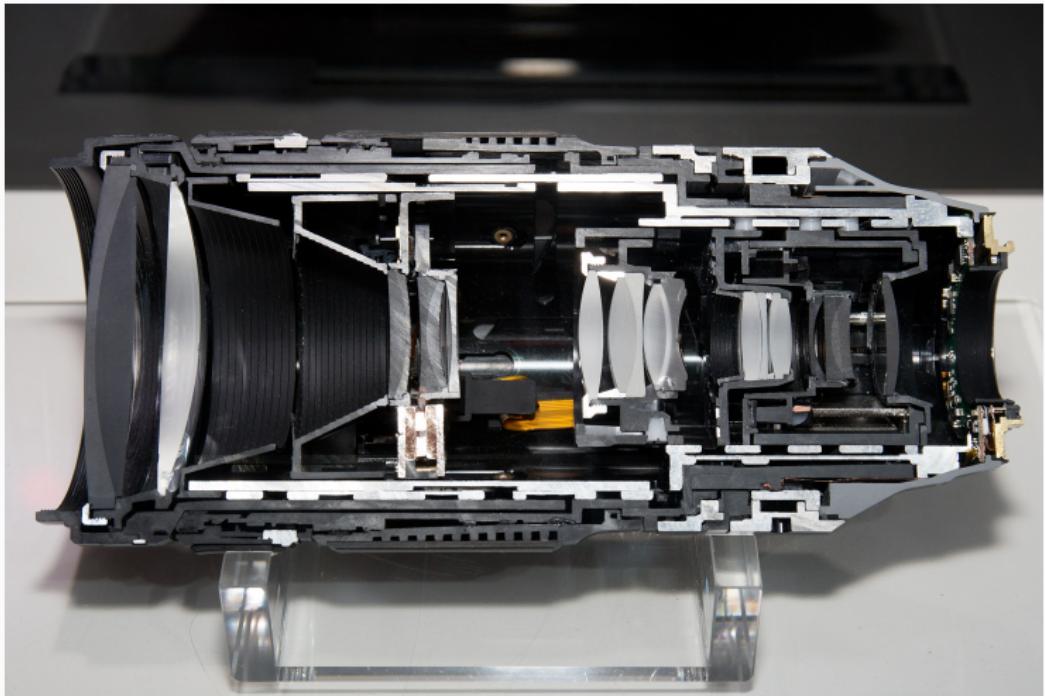


Figure 62: The cross-section of Fujinon XF100-400mm zoom lens [30].

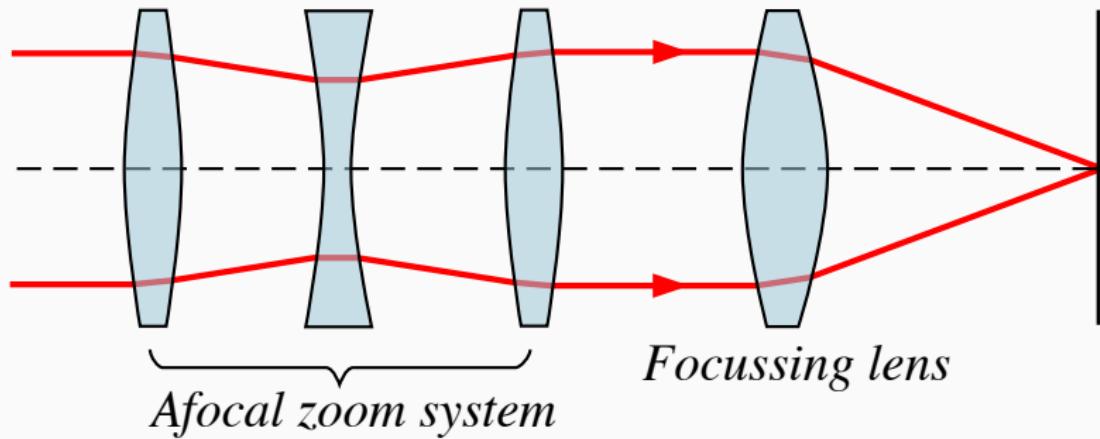


Figure 63: A simple zoom lens system. The three lenses of the afocal system are [1], [2], [3] (from left). [1] and [2] can move to the left and right, changing the overall focal length of the system [14]



- An important issue in zoom lens design is the correction of optical aberrations (such as chromatic aberration and, in particular, field curvature) across the whole operating range of the lens;
- This is considerably harder in a zoom lens than a fixed lens, which needs only to correct the aberrations for one focal length.



Figure 64: Photographic example showing a high quality lens (top) compared to a lower quality one exhibiting transverse chromatic aberration [51].



- Prime lens is a fixed focal length photographic lens (as opposed to a zoom lens), typically with a maximum aperture from f2.8 to f1.2.
- While a prime lens of a given focal length is less versatile than a zoom lens, it is often of superior optical quality, wider maximum aperture, lighter weight, and smaller size.
- These advantages stem from having fewer moving parts, optical elements optimized for one particular focal length, and a less complicated lens systems that creates fewer optical aberration issues.
- Larger maximum aperture (smaller f-number) facilitates photography in lower light, and a shallower depth of field.

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.



Appendix

[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.



[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 65: Different apertures of a lens.

[Go Back](#)

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

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

The Fourier transform (\mathcal{F}_x) if given by

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



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 ■



Appendix

Go Back

The Weierstrass transform of a function $f : \mathbf{R} \rightarrow \mathbf{R}$, 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).

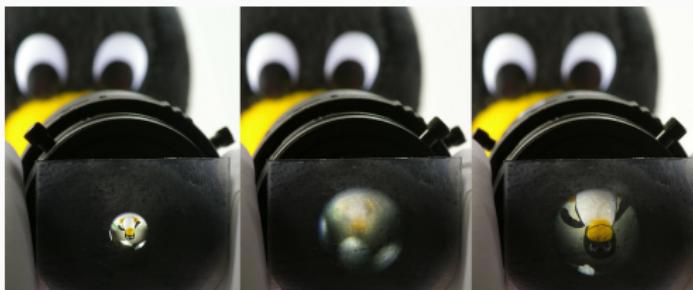


Figure 66: A varifocal lens. Left image is at 2.8 mm, in focus. Middle image is at 12 mm with the focus left alone from 2.8 mm. Right image is at 12 mm refocused. The close knob is focal length and the far knob is focus.



Appendix

[Go Back](#)

In photography and optics, vignetting (/vɪnɪtɪŋ/; 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.

[Go Back](#)

Shutter speed or exposure time is the length of time that the film or digital sensor inside the camera is exposed to light (that is, when the camera's shutter is open) when taking a photograph [38]. The amount of light that reaches the film or image sensor is proportional to the exposure time. $1/500$ of a second will let half as much light in as $1/250$.



Figure 67: Effects of shutter speed on image [13]. As shutter speed slows down the image gets blurrier.



- [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. *Inherited Colour Vision Deficiency*. 2024. URL: <https://www.colourblindawareness.org/colour-blindness/causes-of-colour-blindness/inherited-colour-vision-deficiency/>.
- [5] Colour Blind Awareness. *Types of Colour Blindness*. 2024. URL: <https://www.colourblindawareness.org/colour-blindness/types-of-colour-blindness/>.
- [6] BBC. *Optical illusion: Dress colour debate goes global*. 2015. URL: <https://www.bbc.com/news/uk-scotland-highlands-islands-31656935>.



- [7] Cmglee. *A Bayer pattern on a sensor in isometric perspective/projection*. 2020. URL: https://commons.wikimedia.org/wiki/File:Bayer_pattern_on_sensor.svg.
- [8] Crisp. *Getting out of Auto: Understanding ISO on your digital camera*. 2016. URL: <https://newatlas.com/photography-iso-range-setting-guide/44599/>.
- [9] Ralph Dammel. *Diazonaphthoquinone-based resists*. Vol. 11. SPIE press, 1993.
- [10] DarkRoom. *Effects of Focal Length and Angle*. 2024. URL: <https://thedarkroom.com/focal-length/>.
- [11] Michael F Deering. "The limits of human vision". In: *2nd international immersive projection technology workshop*. Vol. 2. 1. 1998.
- [12] Zoltan Derzsi and Robert Volcic. "Not only perception but also grasping actions can obey Weber's law". In: *Cognition* 237 (2023), p. 105465.
- [13] Dilmen. *Windflower and Shutter speed*. 2004. URL: <https://commons.wikimedia.org/wiki/File:Windflower-05237-nevit.JPG>.
- [14] DrBob. *Simple zoom lens*. 2005. URL: <https://commons.wikimedia.org/wiki/File:Zoomlens1.svg>.



- [15] EdmundOptics. *Anatomy of a Lens*. 2024. URL: <https://www.edmundoptics.com/knowledge-center/application-notes/imaging/the-anatomy-of-a-lens/>.
- [16] 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.
- [17] Filya1. *CMOS image sensor*. 2009. URL: <https://en.m.wikipedia.org/wiki/File:Matrixw.jpg>.
- [18] FlickreviewR. *Sony F828 Camera*. 2009. URL: https://commons.wikimedia.org/wiki/File:Sony_F828.jpg.
- [19] Eric R Fossum. “Active pixel sensors: Are CCDs dinosaurs?” In: *Charge-Coupled Devices and Solid State Optical Sensors III*. Vol. 1900. SPIE. 1993, pp. 2–14.
- [20] Eric R Fossum and Donald B Hondongwa. “A review of the pinned photodiode for CCD and CMOS image sensors”. In: *IEEE Journal of the electron devices society* (2014).
- [21] Eric Fox. *CMOS TDI: A Game Changer in High-Fidelity Imaging*. 2019. URL: <https://possibility.teledyneimaging.com/cmos-tdi-a-game-changer-in-high-fidelity-imaging/>.



Appendix iv

- [22] 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.
- [23] 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/>.
- [24] healthdirect. *Colour blindness.* 2024. URL: <https://www.healthdirect.gov.au/colour-blindness>.
- [25] Hyperphysics. *The Color-Sensitive Cones.* 2024. URL: <http://hyperphysics.phy-astr.gsu.edu/hbase/vision/colcon.html>.
- [26] 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.
- [27] Susanne Kohl et al. “Achromatopsia”. In: (2018).



- [28] Harris R Miller. "Color filter array for CCD and CMOS image sensors using a chemically amplified thermally cured pre-dyed positive-tone photoresist for 365-nm lithography". In: *Advances in Resist Technology and Processing XVI*. Vol. 3678. SPIE. 1999, pp. 1083–1090.
- [29] Alistair Moffat. "Huffman coding". In: *ACM Computing Surveys (CSUR)* 52.4 (2019), pp. 1–35.
- [30] Morio. *Fujinon XF100-400mm F4.5-5.6 R LM OIS WR cutaway*. 2016. URL: https://commons.wikimedia.org/wiki/File:Fujinon_XF100-400mm_F4.5-5.6_R_LM_OIS_WR_cutaway_2016_China_P%26E.jpg.
- [31] *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.
- [32] *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.
- [33] Junichi Nakamura. *Image sensors and signal processing for digital still cameras*. CRC press, 2017.
- [34] Bedabrata Pain et al. "A back-illuminated megapixel CMOS image sensor". In: (2005).



- [35] Danny Pascale. “A review of rgb color spaces... from x_{yy} to $r'g'b'$ ”. In: *Babel Color* 18 (2003), pp. 136–152.
- [36] *Photography and graphic technology — Extended colour encodings for digital image storage, manipulation and interchange*. Standard. Geneva, CH: International Organization for Standardization, 2013.
- [37] PolBr. *sRGB colors situated at calculated position in CIE 1931 chromaticity diagram (edited from CIExy1931_sRGB.svg). Luminance Y sets 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.
- [38] Sidney Ray. *Applied photographic optics*. Routledge, 2002.
- [39] 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](https://doi.org/10.1016/S0141-9382(99)00019-0). URL: <https://www.sciencedirect.com/science/article/pii/S0141938299000190>.



- [40] Inc. Reproductions. *Understanding the Differences Between RGB and CYMK Colors*. 2024. URL: <https://reproductionsinc.com/understanding-the-differences-between-rgb-and-cymk-colors/>.
- [41] Mark Stanford Robbins. “Electron-multiplying charge coupled devices–emccds”. In: *Single-Photon Imaging* (2011), pp. 103–121.
- [42] Andrew Rodney. “The role of working spaces in Adobe applications”. In: *Technical Paper. Adobe. Retrieved* (2008), pp. 05–09.
- [43] Khalid Sayood. *Introduction to data compression*. Morgan Kaufmann, 2017.
- [44] Slippens. *Three examples of color halftoning with CMYK separations*. 2009. URL: <https://en.wikipedia.org/wiki/File:Halftoningcolor.svg>.
- [45] 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.
- [46] 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.



Appendix viii

- [47] Teledyne. *How to Evaluate Camera Sensitivity*. 2021. URL: <https://www.flir.com/discover/iis/machine-vision/how-to-evaluate-camera-sensitivity/>.
- [48] Teledyne. *Types of Camera Sensor*. 2024. URL: <https://www.photometrics.com/learn/camera-basics/types-of-camera-sensor>.
- [49] TeledyneCurrent. *Dark Current*. 2024. URL: <https://www.photometrics.com/learn/advanced-imaging/dark-current>.
- [50] teledyneimage. *Linearity: Image Topics*. 2024. URL: <https://www.photometrics.com/learn/imaging-topics/linearity>.
- [51] Zurek. *Chromatic aberration*. 2006. URL: [https://commons.wikimedia.org/wiki/File:Chromatic_aberration_\(comparison\).jpg](https://commons.wikimedia.org/wiki/File:Chromatic_aberration_(comparison).jpg).