**CS691 Project Report**

Detection and decoding of LSB steganographic embedded photographs

*Submitted in partial fulfilment of the requirements*

*for the award of the degree of*

**Bachelor of Technology**

**in**

**Computer Science and Engineering**

**Submitted by**

Devadi Yekaditya (CSE/20039/559)

Mislah Rahman CP (CSE/20055/575)

Sriramsetty Bhanu Teja (CSE/20088/608)

**Under the guidance of**

**Dr. Amit Ranjan Azad**

****

Indian Institute of Information Technology Kalyani West Bengal, India, 741235

Autumn 2022

Contents

[Certificate 3](#_Toc134779173)

[Acknowledgement 5](#_Toc134779174)

[Abstract 6](#_Toc134779175)

[Introduction 7](#_Toc134779176)

[Steganography 7](#_Toc134779177)

[Technologies Used 7](#_Toc134779178)

[Observation 9](#_Toc134779179)

[Detection of steganographic images 11](#_Toc134779180)

[Statistical Analysis 11](#_Toc134779181)

[Discrete Cosine Transform 11](#_Toc134779182)

[Decoding of the steganographic images 14](#_Toc134779183)

[Results 16](#_Toc134779184)

[Work Flow 17](#_Toc134779185)

[Conclusion 18](#_Toc134779186)

[Applications 19](#_Toc134779187)

[Future Scope 19](#_Toc134779188)

[Bibliography & References: 20](#_Toc134779189)

Indian Institute of Information Technology Kalyani

# Certificate

This is to certify that the thesis entitled **“****Detection and decoding of LSB steganographic embedded photographs”** being submitted by Devadi Yekaditya (CSE/20039/559), Mislah Rahman CP (CSE/20055/575), Sriramsetty Bhanu Teja (CSE/20088/608), undergraduate students of Indian Institute of Information Technology Kalyani, West Bengal, 741235, India, for the award of Bachelors of Technology in Computer Science and Engineering is an original research work carried by them under my supervision and guidance. The thesis has fulfilled all the requirements as per the regulations of Indian Institute of Information Technology Kalyani and in my opinion, has reached the standards needed for submission. The works, techniques and the results presented have not been submitted to any other University or Institute for the award of any other degree or diploma.

Dr. Amit Ranjan Azad

Assistant Professor

Date:

Declaration

We hereby declare that the work which is being presented in the thesis entitled **“Detection and decoding of LSB steganographic embedded photographs”** is submitted to the Indian Institute of Information Technology Kalyani in partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science and Engineering during the period from January to May, 2023 under the supervision of Dr. Amit Ranjan Azad, Indian Institute of Information Technology Kalyani, West Bengal 741235, India, does not contain any classified information.

Devadi Yekaditya (CSE/20039/559)

Mislah Rahman CP (CSE/20055/575)

Sriramsetty Bhanu Teja (CSE/20088/608)

Computer Science and Engineering

Indian Institute of Information Technology Kalyani

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

-------------------

Dr. Amit Ranjan Azad

Assistant Professor

Indian Institute of Information Technology Kalyani

-------------------

Place: Kalyani, West Bengal

Date:

# Acknowledgement

We would like to thank our supervisor Dr. Amit Ranjan Azad for giving us the opportunity for working on this project entitled **“Detection and decoding of LSB steganographic embedded photographs”.** His constant guidance and useful insights led the development of this work and is due to his support that the work can be submitted in present state.

Devadi Yekaditya (CSE/20039/559)

Mislah Rahman CP (CSE/20055/575)

Sriramsetty Bhanu Teja (CSE/20088/608)

Computer Science and Engineering

Indian Institute of Information Technology Kalyani

Place: Kalyani, West Bengal

Date:

# Abstract

This report summarizes the usage of image and signal processing techniques to detect and decode least significant bit (LSB) steganographic content in photographs. This project focuses on identification and extraction of hidden image content encoded within the least significant bits of photographs. The outcomes of this research hold potential significance for network security and intelligence applications.

# Introduction

The project focuses on the detection and decoding of steganographic image content in photographs. Steganography is the practice of hiding information within digital media, such as images, in a way that is imperceptible to the human eye. It has become an increasingly popular technique for convert communication and data transmission.

The primary objective of this project is to develop robust techniques that can identify the presence of steganographic content in photographs and subsequently extract the hidden information. By leveraging image and signal processing methods, the project aims to detect the presence of payload embedded using least significant bit (LSB) steganography, where the payload itself is an image embedded within the cover image. The project's objective further extends to the extraction of this embedded payload image.

The successful detection and decoding of steganographic content hold significant implications for various fields, particularly in network security and intelligence applications. By uncovering hidden information, this project aims to contribute to the enhancement of digital forensics, data privacy protection, and counter intelligence efforts.

Throughout this report, we will explore and analyze various image and signal processing techniques utilized in the detection and decoding process. By leveraging statistical analysis, image processing, and signal processing, we aim to develop efficient and effective methodologies to reveal hidden steganographic content in photographs. The outcomes of this research will pave the way for improved network security measures and intelligence gathering techniques in an increasingly digitized world.

## Steganography

Steganography is an art of hiding information in plain sight. Unlike in cryptography, where the primary objective is to make the information unintelligible by encoding it, in steganography the primary objective is to conceal sensitive data within various forms of media without arousing suspicion. Throughout history, steganography has been employed as a clandestine communication technique, enabling the covert transmission of messages without the knowledge of unintended recipients.

Steganography utilizes various methods to achieve its concealment goals. One common technique is least significant bit (LSB) steganography, which involves replacing or manipulating the least significant bits of data in a carrier medium, such as an image or audio file. By altering these insignificant bits, the hidden information can be seamlessly integrated into the carrier without significantly altering its perceptible features. The human eye typically fails to detect these subtle modifications, allowing the concealed data to remain covert.

## Technologies Used

Python was chosen as the programming language for the testing and analysis process of the project due to its user-friendly nature, intuitive syntax, high level abstraction and the extensive libraries it offers for numerical computation, statistical analysis, image processing, and signal processing.

The representation of images in memory often involves large multidimensional arrays. Specifically, a standard high-definition 24-bit image with dimensions 1280 x 720 is represented as a 1280 x 720 x 3 array, containing a total of 2,764,800 elements. While Python does not inherently support C-like arrays, the NumPy library is utilized in this project. NumPy provides efficient C-like arrays and a collection of tools for complex matrix manipulations, arithmetic, statistical analysis, and other high-level mathematical operations.

For this project, another library called OpenCV (Open-Source Computer Vision Library) is employed. OpenCV is a comprehensive collection of algorithms and techniques used for image processing and analysis. Although primarily written in C++, OpenCV supports various programming languages such as C, Python, Java, and MATLAB, and provides GPU-based hardware acceleration capabilities.

Matplotlib, a popular data visualization library in Python, is extensively utilized throughout the project. It offers a comprehensive set of tools for generating static, animated, and interactive visualizations, including various types of plots, charts, and graphs. In this project, Matplotlib is used for plotting histograms and visualizing graphs.

While NumPy provides a vast array of mathematical tools, its functionality is still limited in certain aspects. To overcome these limitations, the SciPy library is employed. SciPy is a comprehensive scientific computing library in Python that extends the capabilities of NumPy. It offers a wide range of functions and tools for mathematical, scientific, and engineering applications. In this project, SciPy is used for some statistical analysis techniques and operations, such as the Discrete Cosine Transform.

# Observation

Even though the least significant bits have been altered in an image to embed a payload, normal eyes can’t detect or realize the difference between an embedded image and the original image. Here despite the payload image *(Figure 1A)* being embedded in the cover image *(Figure 2A)* to form the embedded image *(Figure 3A)*, both the cover image and embedded image looks exactly same.

Since the last significant bits contains embedded data in LSB steganography, If the least significant bits (LSB) of a few subpixels are joined together and plotted its intensity, it shows significant difference between the embedded images and non-steganographic images. Here *Figure 1B* shows the plot of the intensity values of the first 3500 pixels of the payload image *(Figure 1A)* formed by merging the last two bits of four consecutive subpixels. Similarly, *Figure 2B* and *Figure 3B* shows merged LSB plot of cover imageand the embedded image.



Figure 2A: Cover Image



Figure 1A: Payload Image



Figure 3A: Embedded Image

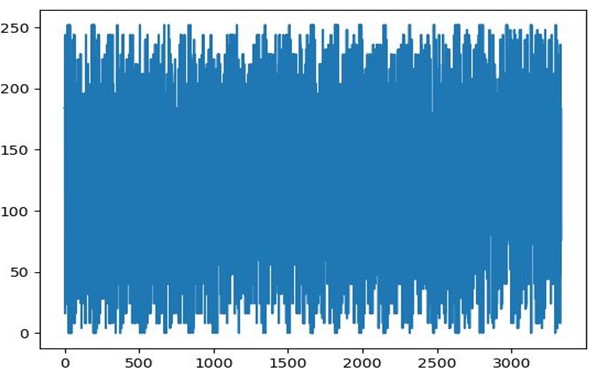


Figure 1B: Plot of intensity of merged subpixels of fig.1

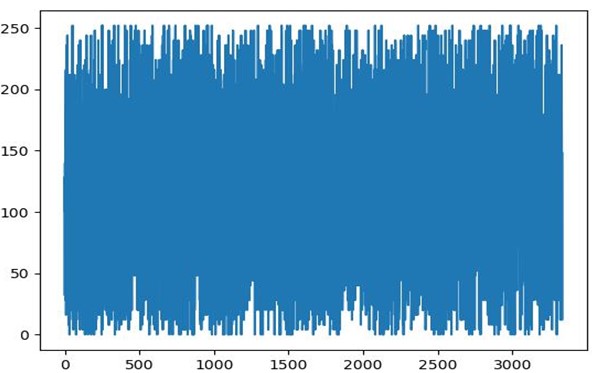


Figure 2B: Plot of intensity of merged subpixels of fig.2

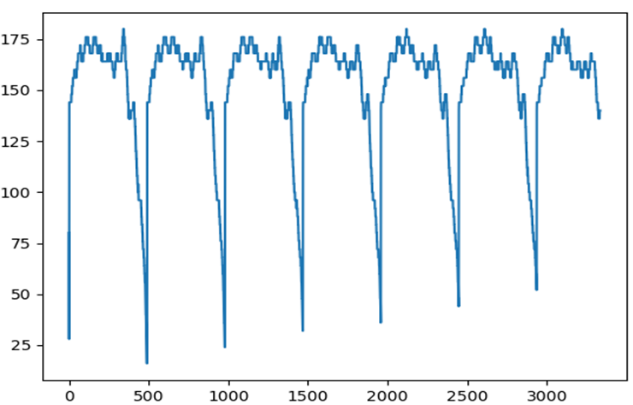


Figure 3B: Plot of intensity of merged subpixels of fig.3

Here, it is observed that for the cover image and the payload image, the intensity of the merged subpixels is completely random and its plot follows a uniform distribution. But for the embedded image, the intensity values are not random and it has repeating patterns after a certain interval. To confirm the observation, merged LSB plots of a greater number of non-steganographic *(Figure 4)* and embedded images *(Figure 5)* are obtained. The red line shows the mean intensity of the pixels.

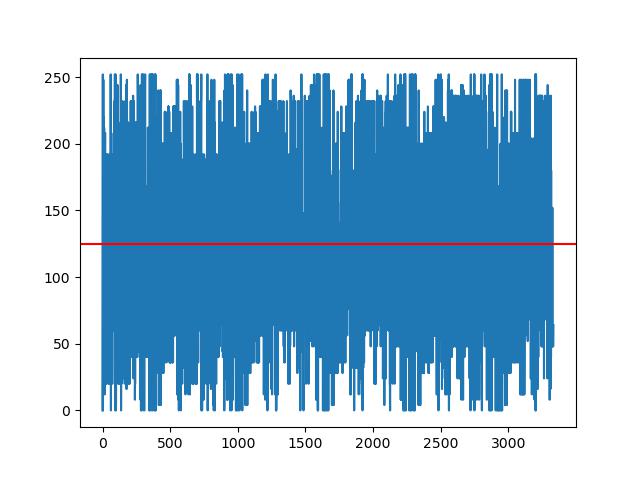
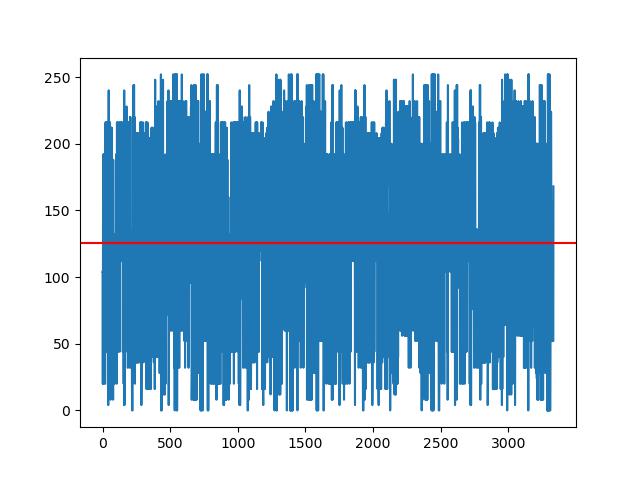
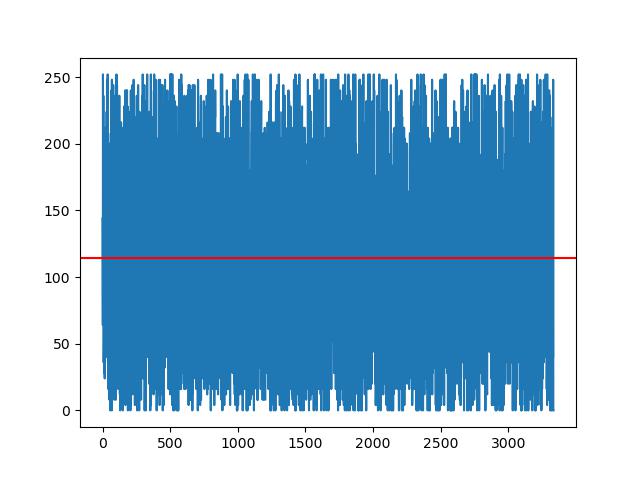
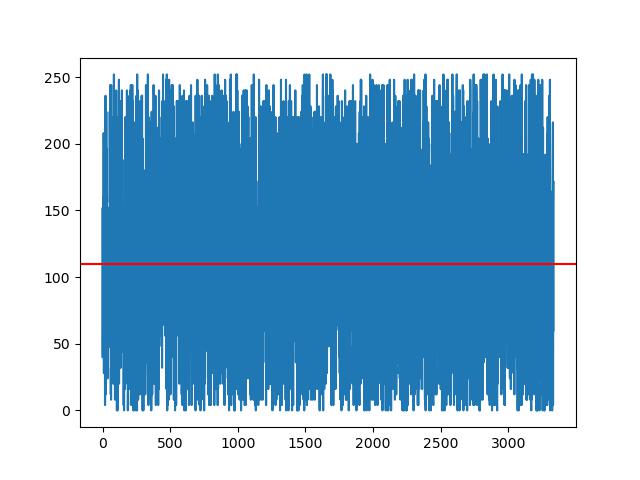
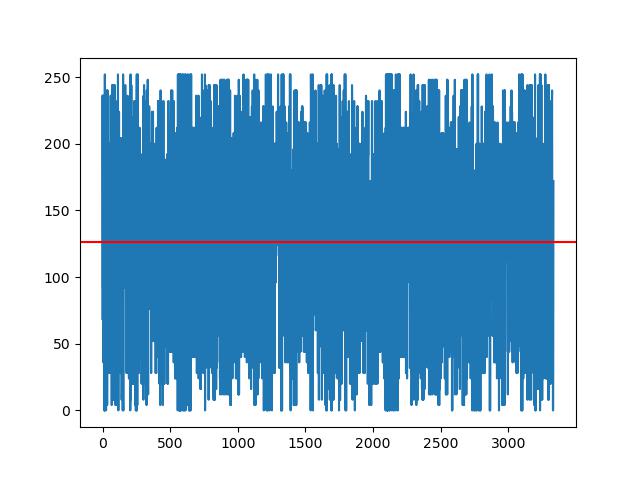
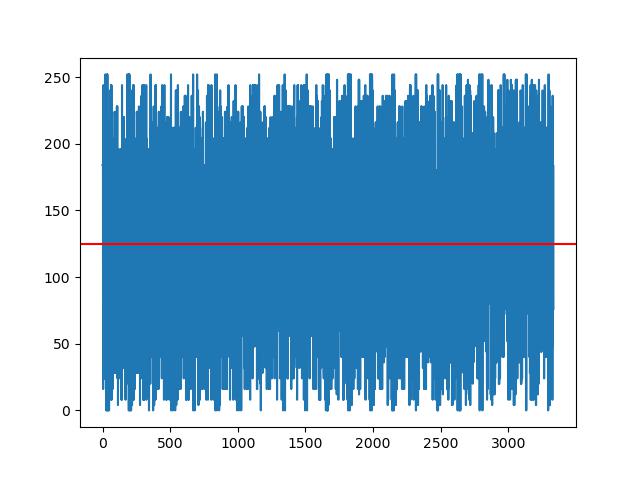


Figure 4A: Merged LSB plots of non-steganographic images

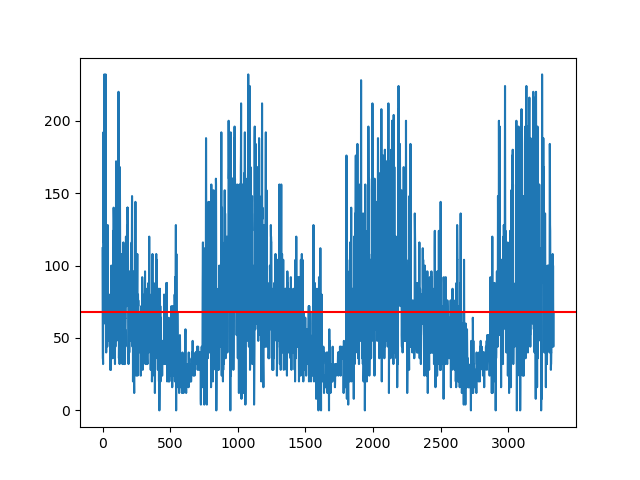
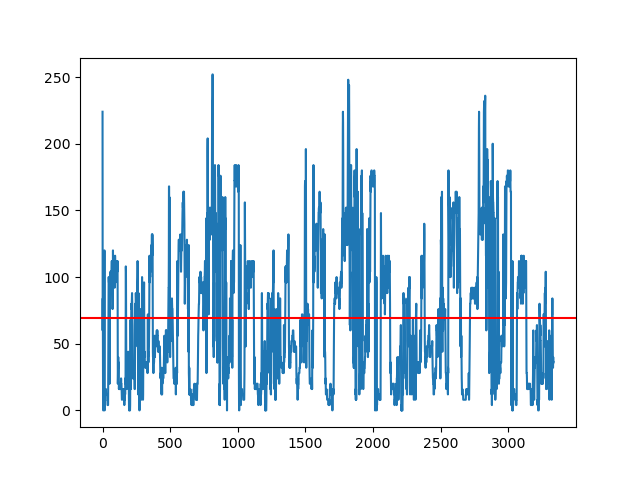
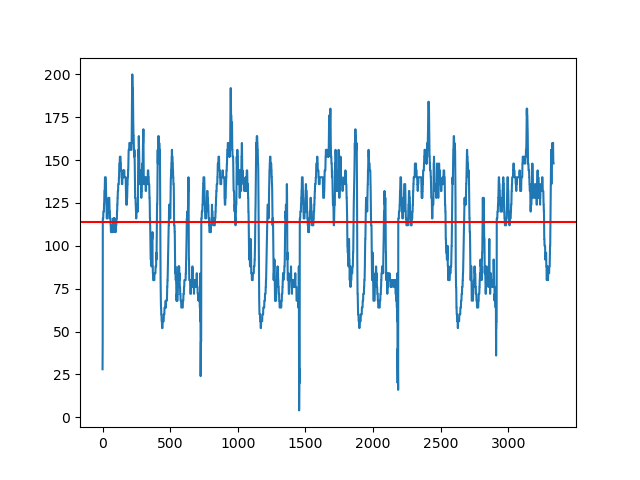
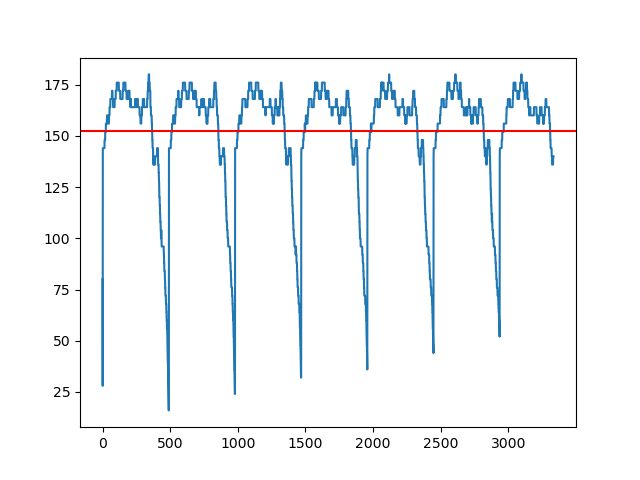
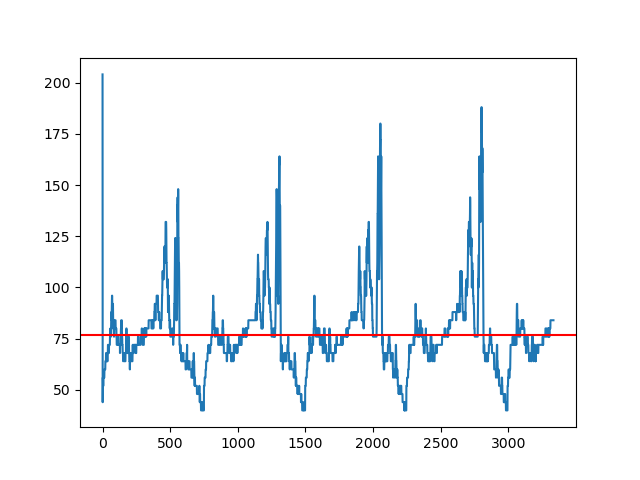
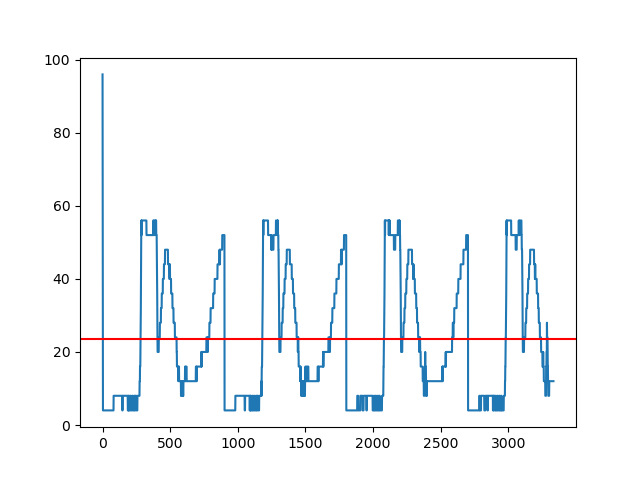


Figure 4B: Merged LSB plots of steganographic images

Here it is observed that the for all non-steganographic images, the plot of merged least significant bits follows uniform distribution. This is due to the fact that in photographs, the least significant bits can contain noises. For steganographic images, it is clear that it doesn’t follow any common probability distribution but follows repeating patterns almost after same interval.

In photographs, the LSB usually contains a lot of noise from various sources. The noise is random and follows almost a uniform distribution in a sufficiently large sample space. But as the transition between pixels are smooth most of the time and likelihood of MSB getting corrupted by random noises is very low, the MSBs changes less often than the LSB of adjacent pixels. Since the steganographic images contains the most significant bits (MSB) of the payload image is concealed within the least significant bits (LSB) of the embedded image, the change in LSB of the embedded image is significantly less. Hence, the plot of embedded images doesn’t follow uniform distribution.

As the transition between pixels in photographs are very smooth, the whole neighborhood of most of the pixels have almost same colors. This cause variation of more significant bits slower than that of lesser significant bits. Also, the consecutive rows follow same patterns of intensity levels with minute difference between each row. These patterns are clearly visible when the intensity values of each pixel are plotted. Since LSB steganographic images stores all information about each pixel in the LSB, upon merging LSBs of consecutive pixels, the patterns are visible.

# Detection of steganographic images

## Statistical Analysis

We can safely assume that the merged LSB of non-steganographic images follows a uniform distribution. As LSB already contains noise it gets multiplied upon merging multiple LSBs together. Also, from the observation it is clear that the merged LSB of non-steganographic images follows a non-uniform distribution. In order to detect the presence of LSB steganographic content in photographs, the ideal method is to compare the deviation of merged LSB with that of the expected distribution (i.e., uniform distribution).

One method is to check if the difference between mean and standard deviation of the obtained values and that of uniform distribution. There are statistical tests like Kolmogorov-Smirnov Test, Chi-Square Test, Anderson-Darling Test to check how perfectly the dataset fits to uniform distribution. Also, the entropy of the uniform distribution is maximum as the data is random, so by comparing the entropy of the merged LSB with an expected entropy of uniform distribution of the same size, we can guess whether any steganographic content is present or not.

## Discrete Cosine Transform

Discrete cosine transform transforms the signals from spatial domain to frequency domain. As the non-steganographic images follows uniform distribution, a single peak near to zero is observed. This behavior is a consequence of the characteristics of the DCT and the nature of the uniform distribution.

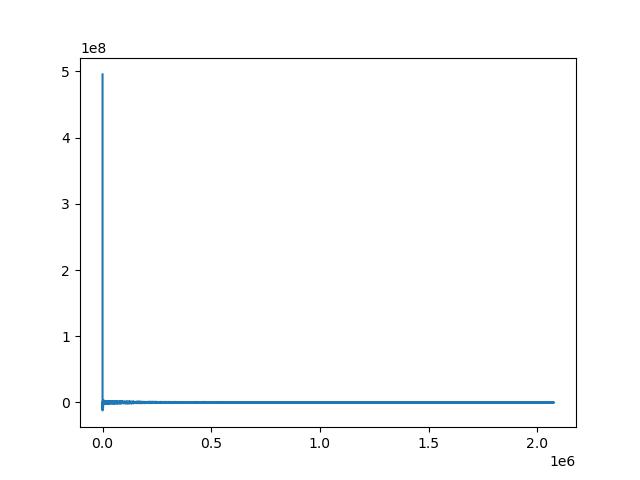
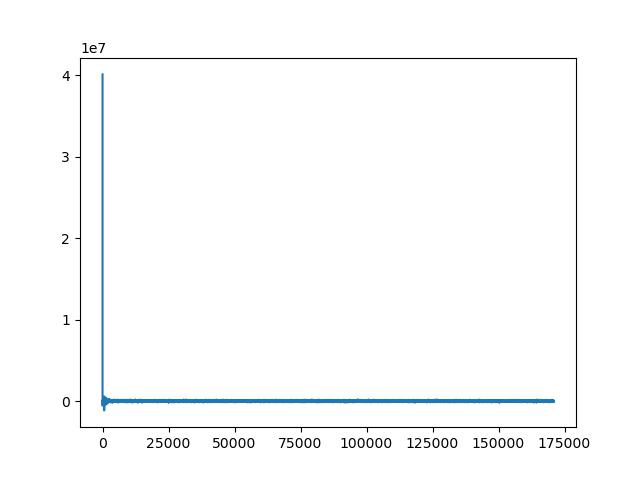
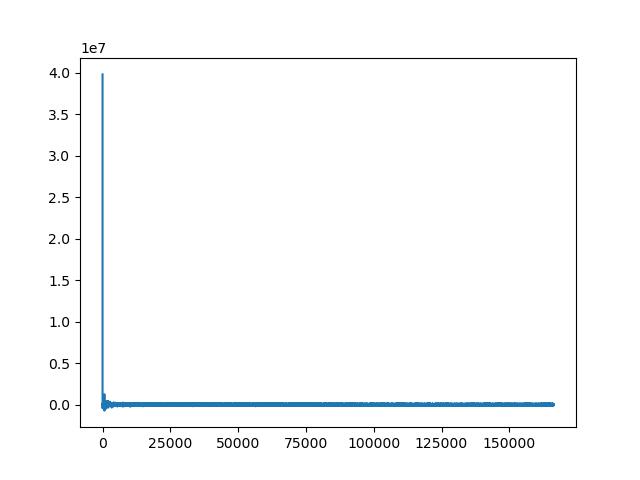
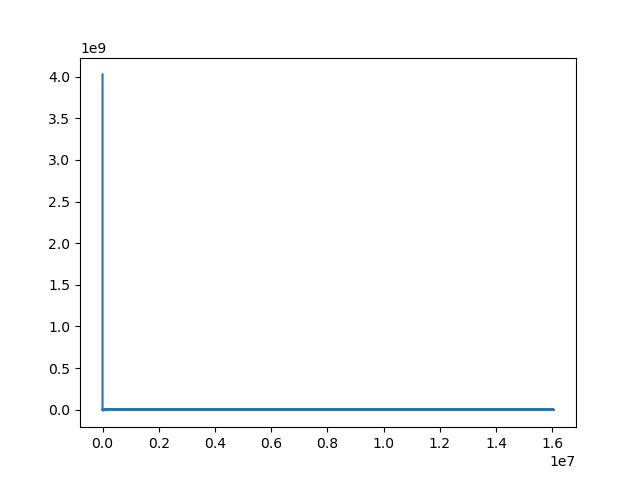
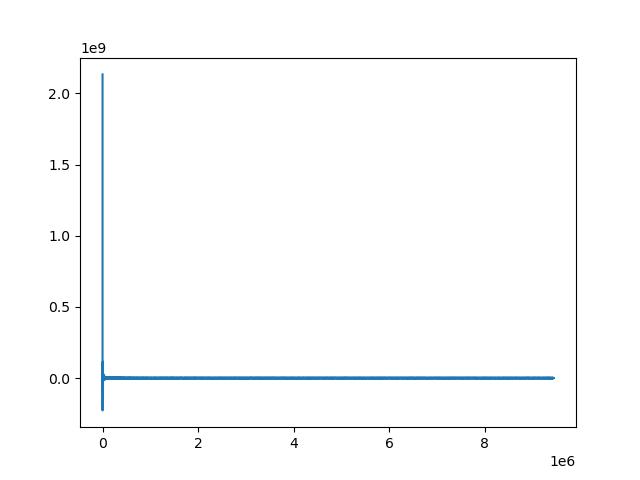
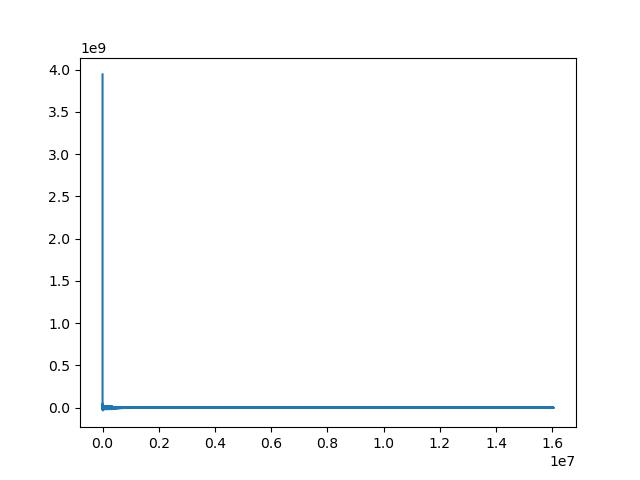
The DCT is designed to efficiently represent signals that have a strong correlation in the spatial domain. In the case of a uniform distribution, where all values are equally likely, there is no strong correlation or pattern in the signal. As a result, the DCT coefficients will have a relatively uniform distribution in the frequency domain.

Specifically, when you compute the DCT of a uniform distribution, the resulting DCT coefficients will have values close to zero for most of the coefficients, except for the first coefficient (C₀ or the DC coefficient). The DC coefficient represents the average value or the overall intensity of the signal. Since a uniform distribution has a constant value across its range, the DC coefficient will have a significant value, while the remaining coefficients will be close to zero. The peak corresponds to the DC coefficient, which represents the constant value or average intensity of the uniform distribution. The other coefficients, capturing higher-frequency components, have negligible values due to the lack of correlations or patterns in the uniform distribution.

But when the DCT of merged LSB of steganographic image is obtained, it shows multiple peaks in the coefficient values. It is due to the presence of repeated patterns in the image.

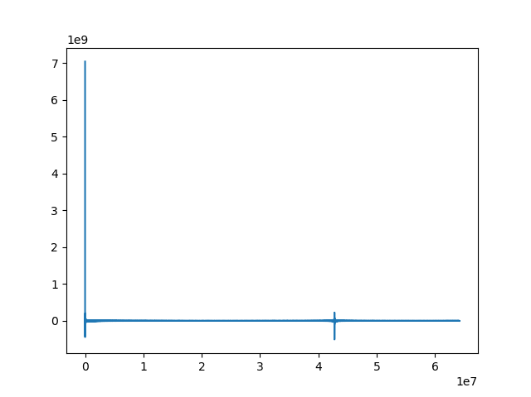
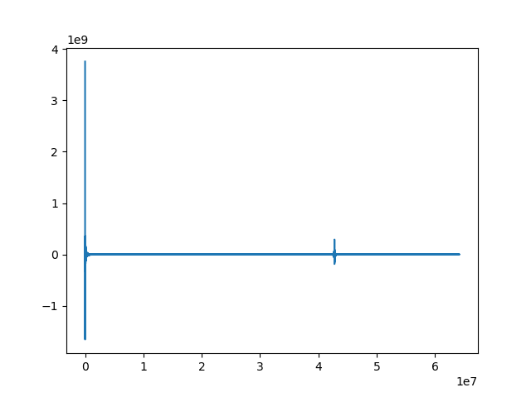
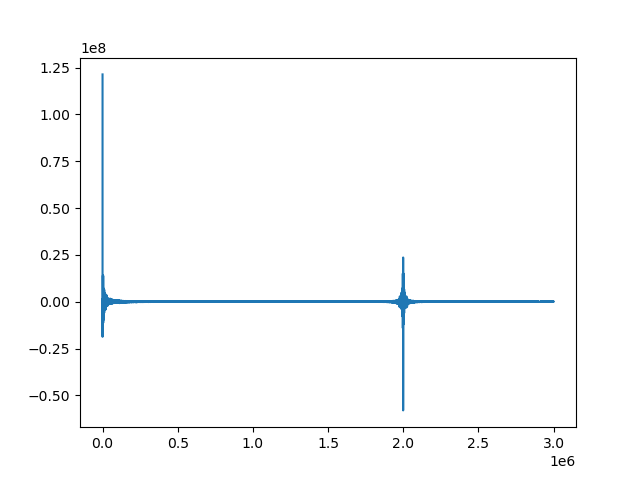
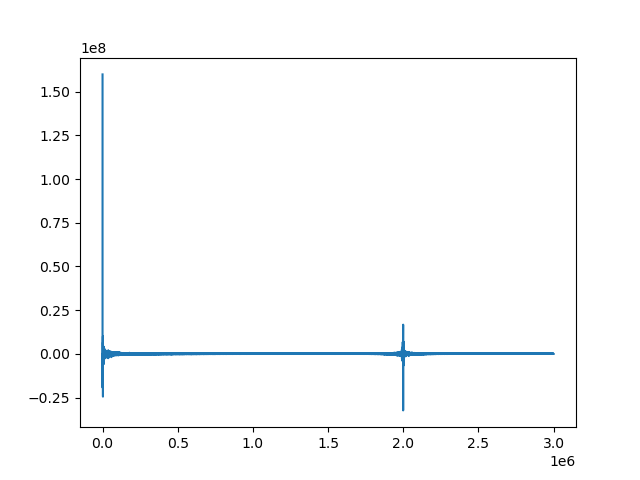
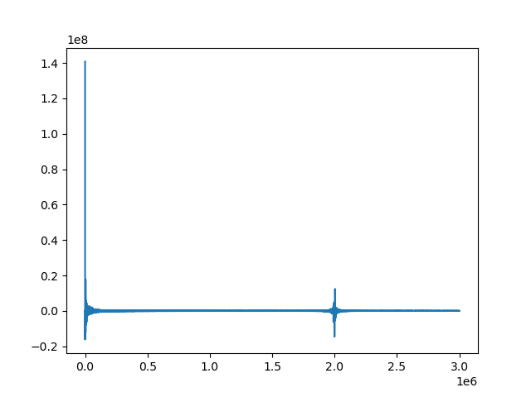
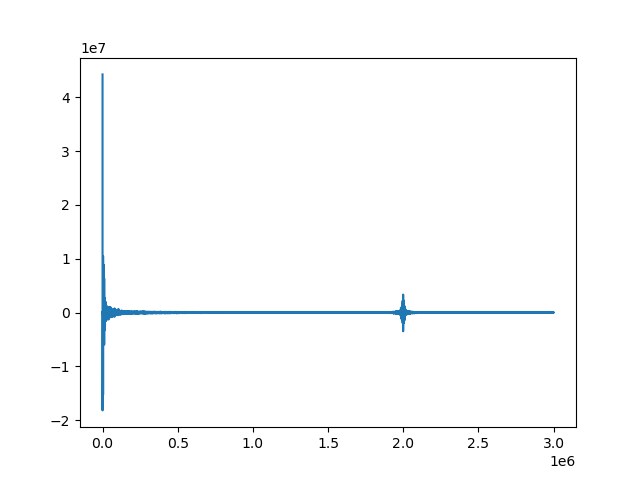
After the principal coefficient, atleast one more coefficient shows large spike which shows large spike which shows the presence of periodic repetition.

Figure 5: DCT of non-steganographic images



In Figure 5 (DCT of non-steganographic images), there is only one peak in the graph that is in the starting of the graph showing the changes in the LSBs of the pixel values in the image. Non-steganographic images do not have any hidden data embedded in them and therefore do not show any such deviations or anomalies in their DCT graphs. The DCT graphs of non-steganographic images typically display a distribution with a single spike near to zero.

Figure 6: DCT of Steganographic images



From Figure 6 we can see they are DCT graphs of steganographic images in which we can two peaks. One peak is in the starting which is same as a non-steganographic image graph plot.

From these observations we can say that steganographic images may exhibit changes or modifications in their frequency domain due to the presence of hidden data, whereas non-steganographic images do not have any hidden data embedded in them and therefore do not exhibit any such changes.

# Decoding of the steganographic images

As observed the patterns are repeating after a certain interval which is equal to the width of the image. If the period of the repetition of the pattern can be obtained the width of the payload image can be approximated. The width can be substituted in different LSB steganographic decoding algorithms to obtain payload image. The easiest approach is to find the period of the patterns is to identify and search for any distinct features of the repeating pattern. One approach is to search for local maxima or local minima in the plot. For demonstrating this, the payload *(Figure 7)* is embedded in the cover image *(Figure 6)*.



Figure 7A: Cover Image



Figure 7B: Payload Image

The plot of the merged LSB bits of the embedded image with a random range with local maxima marked is shown in the *Figure 8*.

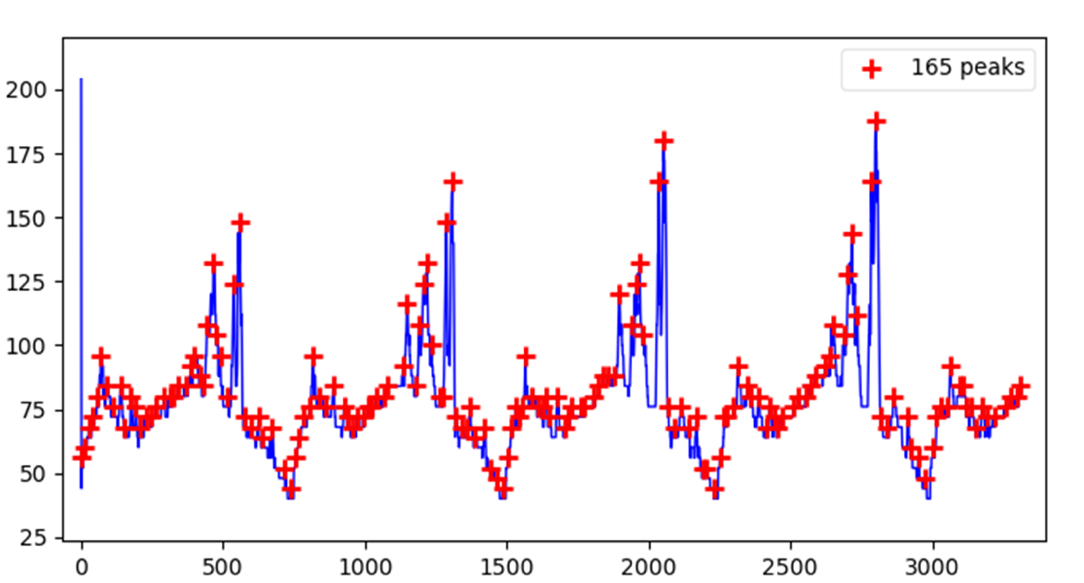


Figure 8A: Local maxima of merged LSB of random range

Here it is observed that all of the local maxima have been plotted. But we only need the largest peak in each of the pattern segment to find the period of the pattern. One approach is to apply a threshold for detecting the peak by setting a minimum peak height (minimum height of the peak to be marked) or minimum peak distance (minimum distance between two consecutive peaks to be marked). Since we don’t know the optimal threshold to find the global maxima of a pattern segment, we can iteratively increase the threshold until the distance between consecutive maxima to be of minimal standard deviation. By applying iteratively increasing the threshold, the following plot *(Figure 9)* is obtained and the distance between the peaks are marked.

Here the mean distance between three consecutive peaks is 747.33. As the width should be of an integer value, it is rounded to nearest integer, i.e., 747 and applying on various commonly used LSB steganography decoding algorithms *(Figure 8A)*.

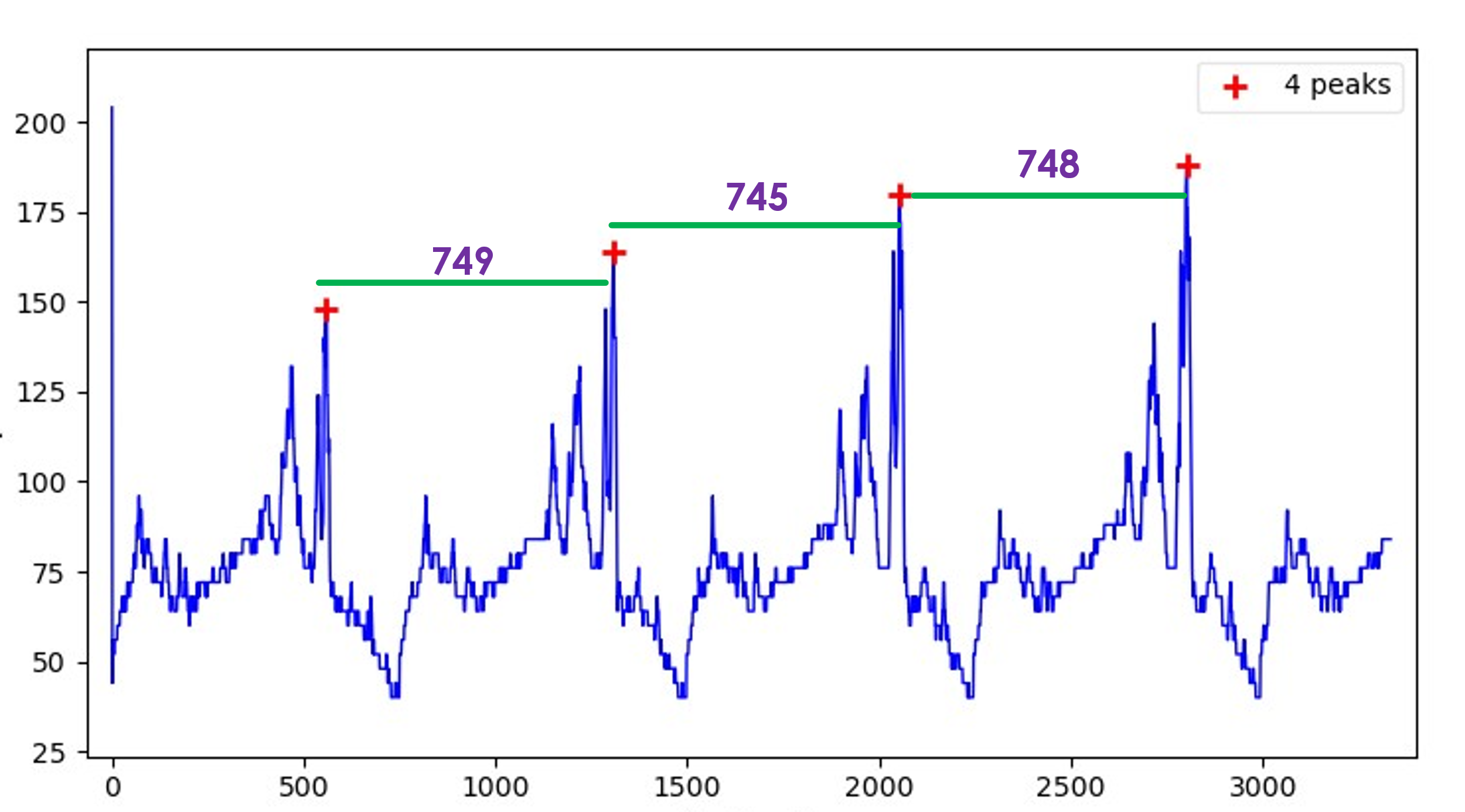


Figure 8B: Local maxima of merged LSB of random range after applying threshold

The obtained image is visible but the color is wrong. It is due to the color channels are out of order. To fix it, the channel offset is corrected *(Figure 8B)*.

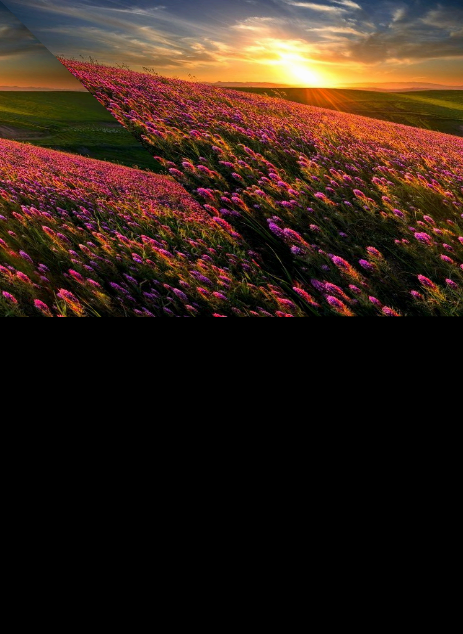
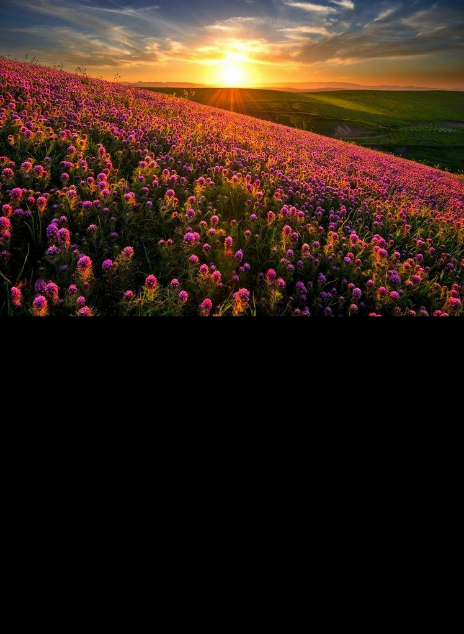


Figure 9C: After correcting the width of the image

Figure 9A: Decoding by applying mean as width

Figure 9B: Decoded image after channel correction

# Results

The techniques devised in this project is tested and verified over a collection of steganographic and non-steganographic images. The test data is documented quantitatively below.

Over 30 steganographic images have been tested and 27 were accurately detected. Out of those images, 15 payload images have been successfully extracted and 6 images partially.

2 false positives were detected upon testing over 20 non-steganographic images.

The accuracy is calculated as:

Accuracy of detection of steganographic images

Accuracy of detection of non-steganographic images

Accuracy of extraction of payload images from steganographic images

Accuracy of partial extraction of payload images from steganographic images

# Work Flow

# Conclusion

Here we successfully achieved a consistent result with accuracy over 95% for the tested images, which is good for an unoptimized technique when compared to the advanced techniques already available. Since there are chances for error, it can’t be yet used for a fully automated system without human assistance.

The drawing of circles and numbering object helps the user to easily identify any errors and manually fix it. Since the circles are superimposed on the original image, user find it easy to identify any missing circle or extra circle and can fix the count manually.

The report provides a promising approach to count descriptive, distinct, non-derogative and non-redundant objects in an image based on classical image processing and analysis techniques.

# Applications

This work can be used in various fields like information security and intelligence. It is used in information security to identify and prevent the use of steganography as a means of concealing malicious code or other threats. Steganalysis techniques can help security professionals detect and prevent the use of steganography in cyberattacks, such as the distribution of malware or the exfiltration of sensitive data.

Steganalysis helps in digital forensics which is an important part for investigation where it is used to uncover hidden information or messages that may be relevant to a criminal investigation. Digital forensics experts can use steganalysis techniques to identify and recover hidden data from digital media, which can be used as evidence in court.

It can also be used to verify the authenticity of digital media, such as images or videos, by detecting any hidden information or manipulation that may have been applied. This is particularly relevant in contexts where media is used as evidence, such as in journalistic reporting or legal proceedings.

In summary, steganalysis is an important tool in digital forensics, information security, media verification. By detecting and analyzing hidden information or messages that may have been embedded in digital media, steganalysis can help prevent cyberattacks, protect intellectual property, and ensure the authenticity and integrity of digital media.

# Future Scope

It can expand our research to cover various steganographic techniques beyond LSB .Can explore other popular methods such as frequency domain techniques (Discrete Fourier Transform), spread spectrum techniques, or other advanced LSB-based algorithms like random pixel selection, pixel-value differencing, etc. This will enhance the scope of your project and make it more comprehensive.

It can expand our research to cover various steganographic techniques beyond LSB. We can explore other popular methods such as frequency domain techniques (Discrete Fourier Transform), spread spectrum techniques, or other advanced LSB-based algorithms like random pixel selection, pixel-value differencing, etc. This will enhance the scope of our project and make it more comprehensive.

We can design a user-friendly interface for this project. Develop a software application or a web-based tool that allows users to easily upload images, detect steganographic content, and decode hidden information. Providing an accessible interface will make our project more usable for a wider audience.

# Bibliography & References:

* Digital Image Processing, 3rd edition - Rafael C. Gonzales, Richard E. Woods
* Digital Image Processing and Analysis, 2nd edition – B. Chanda and Dutta Majumder; PHI Learning.
* Chi-Kwong Chan ∗ , L.M. Cheng “Hiding data in image by simple LSb substitution”, the journal of the pattern recognition society, pattern recognition 37(2004) 469-474.
* Tirkey, P., Kudiyam, D., Dhruw, N., Markam, D. and Ghosh, M.R. Image Steganography Using LSB Along With IDEA Algorithm. International Journal of Engineering and Computer Science, 5(12) (2016).