CONTRAST ENHANCEMENT BASED JPEG FORENSIC DETECTABILITY

A PROJECT REPORT

Submitted By

AMAL A S

TVE18MCA006

to

the APJ Abdul Kalam Technological University

in partial fulfillment of the requirements for the award of the degree

of

Master of Computer Applications

Department of Computer Applications

College of Engineering

Trivandrum-695016

JULY 2021

Declaration

I undersigned hereby declare that the project report titled **"Contrast Enhancement Based JPEG Forensic Detectability"** submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Applications of the APJ Abdul Kalam Technological University,Kerala is a bonafide work done by me under supervision of Smt.Divya S K,Asst.Professor.This submission represents my ideas in my words and where ideas or words of others have been included.I have adequately and accurately cited and referenced the original sources.I also declare that I have adhered to ethics of academic honesty and integrity as directed in the ethics policy of the college and have not misrepresented or fabricated any data or idea or fact or source in my submission.I understand that any violation of the above will be a cause for disciplinary action by the Institute and/or University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained.This report has not been previously formed the basis for the award of any degree,diploma or similar title.

Place : Trivandrum **Amal A S**

Date : 27/06/2021

DEPARTMENT OF COMPUTER APPLICATIONS

COLLEGE OF ENGINEERING

TRIVANDRUM

CERTIFICATE

This is to certify that the report entitled **Contrast Enhancement Based JPEG Forensic Detectability** submitted by **Amal A S** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of the project work carried out by him under my guidance and supervision.This report in any form has not been submitted to any University or Institute for any purpose.

Internal Supervisor External Supervisor

Head of the Dept

Acknowledgement

First and for most I thank **GOD** almighty and to my parents for the success of this project.I owe a sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my project.

I am extremely thankful to **Dr Jiji C V**, Principal,College of Engineering Trivandrum for providing me with the best facilities and atmosphere which was necessary for the successful completion of this project.

I am extremely grateful to **Dr.Sabitha S**,HOD,Dept of Computer Applications,for providing me with best facilities and atmosphere for the creative work guidance and encouragement.

I express our sincere thanks to **Smt.Divya S K**,Asst. Professor,Department of Computer Applications,College of Engineering Trivandrum for her valuable guidance,support and advice that aided in the successful completion of my project.

I profusely thank other Asst.Professors in the department and all other staffs of CET,for their guidance and inspirations throughout my course of study.

I owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this project.No words can express my humble gratitude to my beloved parents and relatives who have been guiding me in all walks of my journey.

Amal A S

Abstract

Rapid advancement in digital image processing tools and software has made it extremely simple to manipulate the digital images without leaving any footprints.With the rise of social media,it has become a heated topic about society's security and threat.Image steganography hides secret information in an image called cover image so naturally that the other users cannot recognize the existence of information in the revealed image.

The project deals with the forensic department detecting compressed image even if it contains the presence of noise signal.The project investigates that the decoded hidden image information is a normal image or not.For detection two contrast enhancements based forensic algorithms via histogram peak/gap artifacts analysis are used.A deep neural network model and entropy features were used to train the normal and improperly decoded aberrant images.Because the information may be partially embedded in the cover image,image patches are used to process the discrimination.The proposed approach discriminates and recovers the hidden image information automatically from a tremendously large number of steganography encoding methods.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Growth in technology has made our life so much easier but also it has some drawbacks.With the rapid development of digital media editing techniques,digital image manipulation becomes rather convenient and easy.Although these techniques benefits the legal image processing but malicious users might use such innocent manipulations to tamper digital photograph images.These anti-forensic activities include compression or embedding with a file.Forensic department mainly focuses on detecting any kind of malpractices done in the image,whereas the anti-forensic department tries to fool the forensic analyst by hiding the traces of compression.The most commonly used image standard is JPEG.

JPEG has a property of following lossy compression which did not preserve the bit values.So it leaves traces after compression.This makes the forensic analyst to easily identify whether the file is compressed or not by analysing the histograms of both original and suspected file.Anti-forensic department further works to make the histograms same by adding a noise signal.So this makes very difficult to identify whether an image is manipulated or not.Image forgeries are widespread on Internet and other security related applications which utilize images are severely impacted.This problem can be easily solved by a Contrast enhancement based JPEG forensic detectability method to ensure the originality and authenticity of digital images.

Chapter 2

Problem Definition and Motivation

Contrast Enhancement Based Jpeg forensic detectability aims with detecting whether the image has undergone any compression or if it is embedded with a file.This method will increase the authenticity of the images that we have been seeing over the internet.This process could be very much useful for applications like law enforcement and news recordings where it is necessary to verify the integrity of the image.In this project I'm trying to develop a machine learning model which can be used to detect compressed and embedded images.

The major motivation behind choosing the project was the drawbacks of the existing system.Currently there are systems and software to find whether an image is tampered or morphed.The major drawbacks of the existing system are mentioned below,

- Difficulty in detecting compression if anti-forensics add noise to the compressed image.
- Detection accuracy and efficiency is very less.
- Large size images consume more time for detection.

The major objective the project is to find compressed image even if it contains the presence of noise signal or an image is embedded with a file.And through this project I can assure that the project can achieve it.

Chapter 3

Literature Review

Anti-forensic digital image manipulation is there from the beginning of technological growth.Different techniques are also used for detecting image manipulation during these period.As part of literature research,looked at a number of publications and talks on the subject.The quick summary of findings are specified in this chapter.

3.1 Using Intrinsic Fingerprints

Swaminathan proposed the intrinsic fingerprints technique for image forensic detection.This propose a new methodology for the forensic analysis of digital camera images.The intrinsic fingerprints of the various in-camera processing operations can be estimated through a detailed imaging model and its component analysis.To establish a linear time-invariant approximation and estimate the intrinsic fingerprints associated with these post camera activities,a blind deconvolution technique is used.Any changes in the estimated camera-imposed fingerprints,or the appearance of new fingerprint types,indicate that the image has been tampered with or steganographically embedded.

3.2 Using Fusion Boost

H. Cao and A. C. Kot proposed the method of fusion boost to detect manipulation on image patches.In this paper,designed a new ensemble manipulation detector to simultaneously detect a wide range of manipulation types on local image patches.

3.3 Using Overshoot Artifacts Analysis

G. Cao, Y. Zhao, R. Ni, and A. C. Kot presented the overshoot artefacts analysis approach in their paper "Unsharp masking sharpening detection by overshoot artefacts analysis" in 2011.In sharpened images,overshoot artefacts are prevalent at side-planar edges.When detected by a sharpening detector,such artefacts can serve as a relatively unique feature for distinguishing the preceding sharpening operation's performance.

3.4 Using Statistical Intrinsic Fingerprints

M. C. Stamm and K. J. R. Liu proposed this method in "Forensic detection of image manipulation using statistical intrinsic fingerprints,"in 2010.They offered various methods for detecting global and local contrast augmentation,as well as histogram equalisation and the global addition of noise to a previously JPEG-compressed image.

By considering all the data,can say that detecting digital image manipulations like compression and embedding file can be done using various techniques.One of the greatest methods is to analyse histogram peak/gap artefacts using two contrast enhancements-based forensic algorithms.And got to know that the result of this model will be quite encouraging and the margin of error will be minimum.

Chapter 4

Requirement Analysis

4.1 Purpose

Verifying the integrity and authenticity of images on internet is very much necessary in current situation where digital image manipulation becomes rather convenient and easy.The purpose of this project is to develop a system to detect image manipulations like compression or embedding in an image given by the user.The "Contrast Enhancement Based JPEG Forensic Detectability " is build by using Machine Learning and Image Processing.The system helps to verify the authenticity of images available on internet.

4.2 Overall Description

Getting correct and unmanipulated information is everyone's right.However,due to the prevalence of anti-forensics image forgeries,the event and scene information conveyed in images may no longer be credible.Contrast Enhancement Based JPEG Forensic Detectability is a major forensic tool for verifying the integrity of the images.This technology is beneficial for applications such as law enforcement and news recording;nevertheless,it is also important to evaluate the originality and authenticity of digital photos,as well as to make the image altering history clear in order to obtain further information.In the current system two contrast enhancement based forensic algorithms via histogram peak/gap artifacts analysis is proposed.DCT algorithm is used for image pre-processing and LSB is used for finding the embedded image.

4.2.1 Hardware Requirements

- Processor : Intel Core i3
- Storage : 512 GB Hard Disk space
- Memory : 4 GB RAM

4.2.2 Software Requirements

- Operating System : Linux/Windows
- Platform : Visual Studio
- Front end : Visual C#

4.3 Functional Requirements

The functional requirements represents the behaviour of the system.The proposed system consist of 2 parts.It includes

4.3.1 Anti Forensic

Goal in the anti-forensic part is to enhance the contrast of original image and then save it as a new enhanced image for further analysis.It consist of

- **image pre-processing:** Here the input is image so first it needs to be pre-processed.Make sure that of image's extension is JPEG.Image matrix can be generated on the basis of pixel values of the original image.After building the matrix,draw the histogram of the original image.Histogram is a graph showing the number of pixels in an image at each different intensity value found in that image.The histogram of an original image typically conforms to a smooth curve.
- **Contrast enhancement generation:** Here contrast of the original image is enhanced.After enhancing the original image,generate its matrix and based on that draw the histogram.Then observe the differences in histogram of both original image and contrast enhanced image.The histogram of the original image has a smooth envelope,whereas the histogram of the enhanced image has peak/gap artefacts.After enhancing the contrast of original image the resulting new image can be saved in a new folder for further use.

4.3.2 Forensic

Forensic is the main part of this proposed work.It involves two novel algorithms.First,global contrast enhancement detection algorithm and second,identify source enhanced composite image algorithm.It involves

- **Global contrast enhancement detection algorithm:** This algorithm consist of Peaks/pits detection and Zero-height gap bin detection.The two factors affect the presence of histogram peak bins in a JPEG image are the flatness and JPEG quality factor.More apparent peak bins would result from the bigger flat zones and greater DC quantization step.Even for the block with sparse non-zero quantized AC coefficients,after decompression the number of pixel gray levels still decreases to some extent.As a result,there is still a risk of discontinuity in the global histogram.

4.4 Non Functional Requirements

4.4.1 Performance Requirements

- Accuracy : Accuracy in functioning and the nature of user-friendly should be maintained by the system.
- Speed : The system should be able to provide speed.
- Low cost: This system is very cheap to implement and is also user-friendly.
- Less Time consuming: It uses very less time comparing to the existing system .
- User Friendly: This proposed system is highly user friendly they enables to create a good environment.

4.4.2 Quality Requirements

- Scalability : All of the functional requirements will be met by the program.
- Maintainability : The system should be maintainable.It should keep backups to atone for system failures,and should log its activities periodically.
- Reliability : The acceptable threshold for down-time should be large as possible.i.e. mean time between failures should be large as possible.And if the system is broken,time required to get the system backup again should be minimum.
- Availability: This system is easily available as the core equipment in building the software is easily obtained.
- High- Functionality: This system is highly functional in all environment since,They are highly adaptable.

Chapter 5

Design And Implementation

The proposed system is used to detect whether an image has undergone any compression or is embedded with a file.The software uses two contrast enhancement based forensic algorithms via histogram peak/gap artifacts analysis.

5.1 Overall Design

The proposed system follows client server architecture.Contrast Enhancement Based JPEG forensic detectability system has a client part and a server part as well.The client part is used by the user to input the original and suspected images which is to be checked.The input is evaluated by the software and result is shown to the client.It is developed using visual C#

5.1.1 System Design

The system is software based.The input is taken from the user through the software and the input is passed to the software running in the server side.The server program perform tasks such as check image properties,check histograms,check DCT matrix,check presence of noise signal and find noise percentage on the input data for compression.It performs tasks such as find DCT of images,quantize each image,comparing quantized matrix with original matrix,check for difference in LSB bit and extract embedded data for embedding part.

5.1.2 Methodology

There are two parts in this project.The first part is the Anti-forensic side and the second one is the Forensic side.

The main process of this software is the detection of compression or embedding done in an jpeg image format.The major steps in the software creation are image compression,data embedding,detect image compression and detect data embedding.The major steps in the model creation are mentioned below.

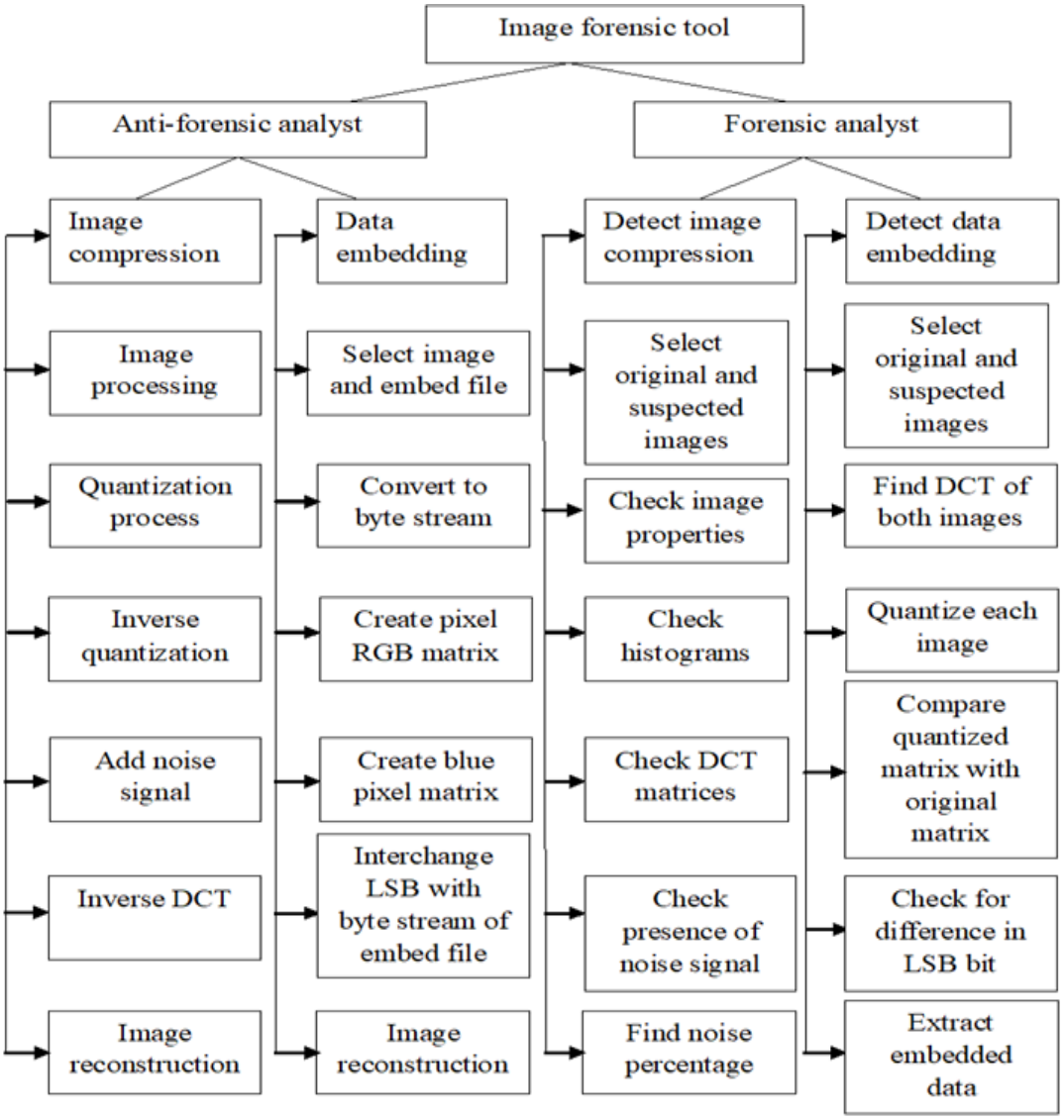

Figure 5.1: Architecture of the system

- **image compression:**For image compression the image is first processed and then using DCT algorithm,quantization matrix,adding noise and applying inverse DCT we obtain compressed image.

- **data embedding:**By creating a RGB matrix and finding the blue region in them we can embed the data by using LSB algorithm.

- **detect image compression:**Detecting compression of image includes checking image properties,histogram,checking presence of noise signal and also finding the noise percentage.

- **detect data embedding:** Data embedding can be identified by finding DCT of both images then quantize each image.After that quantized matrix are checked against the original matrix.Difference in LSB bit is also checked and finally embedded data is extracted.

5.2 Data Flow Diagram

DFD is one of the graphical representation techniques used in a project to show the flow of the data through a project.DFD helps us to obtain an idea about the input,output,and process involved.The things absent in a DFD are control flow,decision rules,and loops.It can be described as a representation of functions,processes that capture,manipulate,store,and distribute data between a system and the surrounding and between the components of the system.The visual representation helps for good communication.

It shows the journey of the data and how will it be stored in the last.It does not provide details about the process timings or if the process shall have a parallel or sequential operation.It is very different from a traditional flow chart or a UML that shows the control flow or the data flow.

In level 0 the basic data flow of the application is showcased.It does not show the flow of data much deeper.It will be evaluated in the higher levels of Data Flow Diagram.The Data Flow Diagram of Contrast Enhancement Based JPEG forensic detectability is shown below.

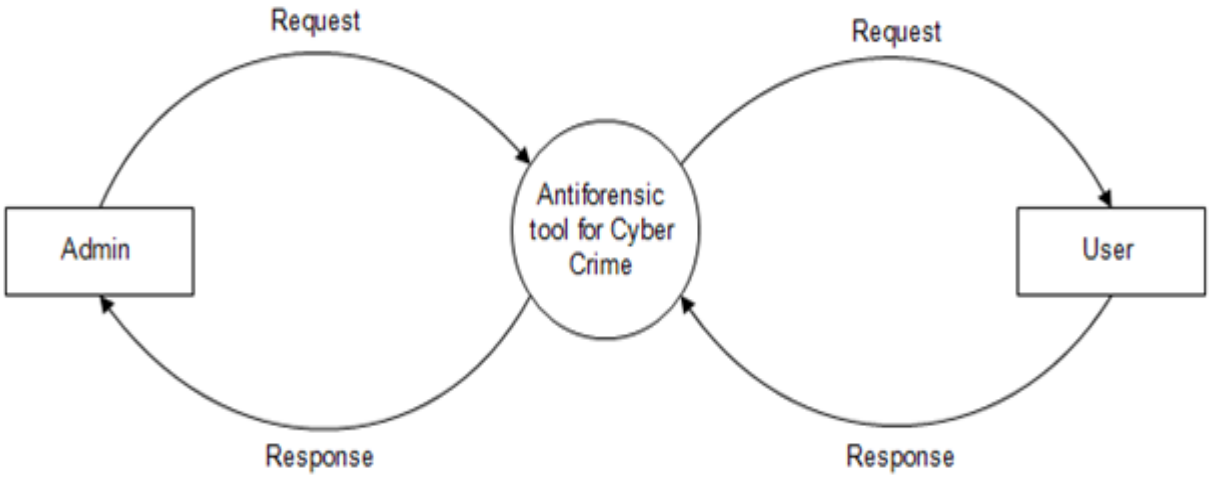

Figure 5.2: Level 0 DFD

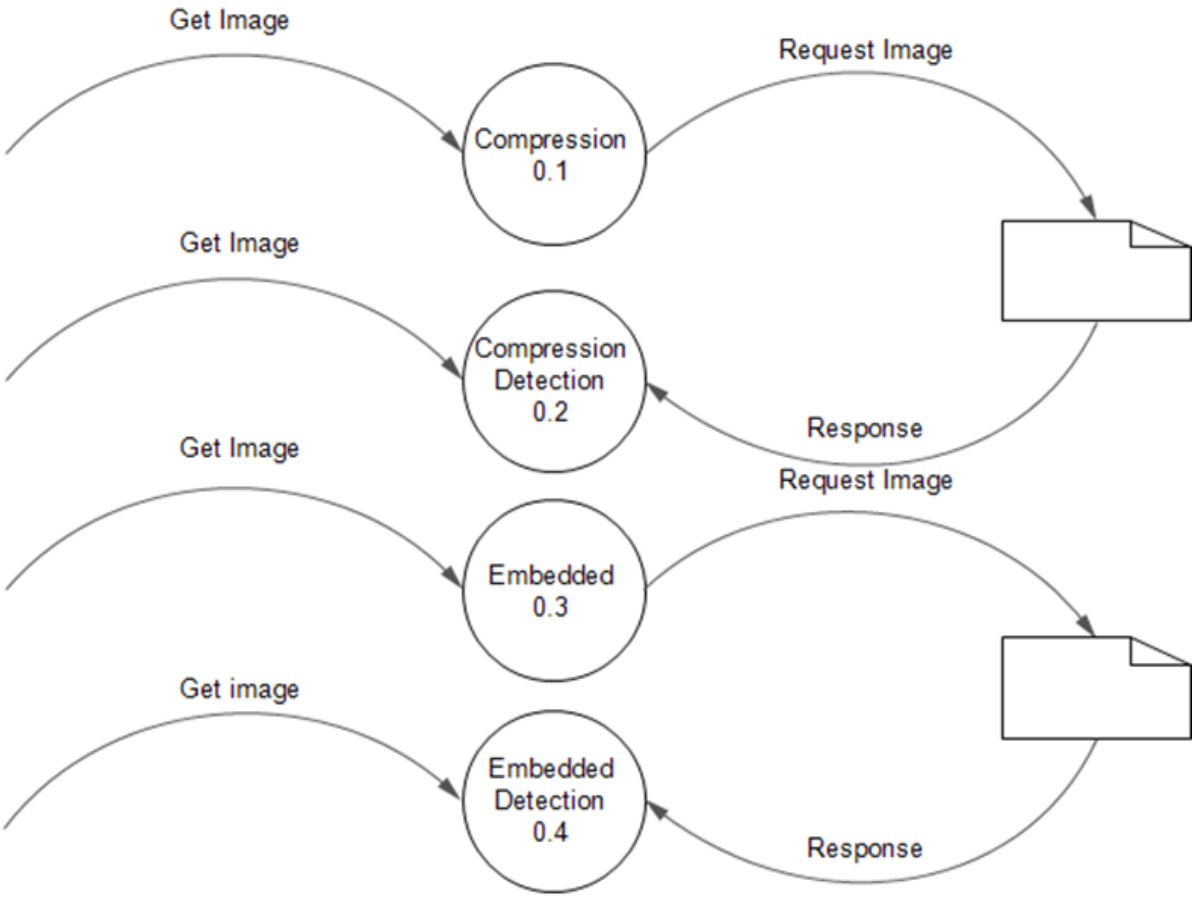

Figure 5.3: Level 1 DFD

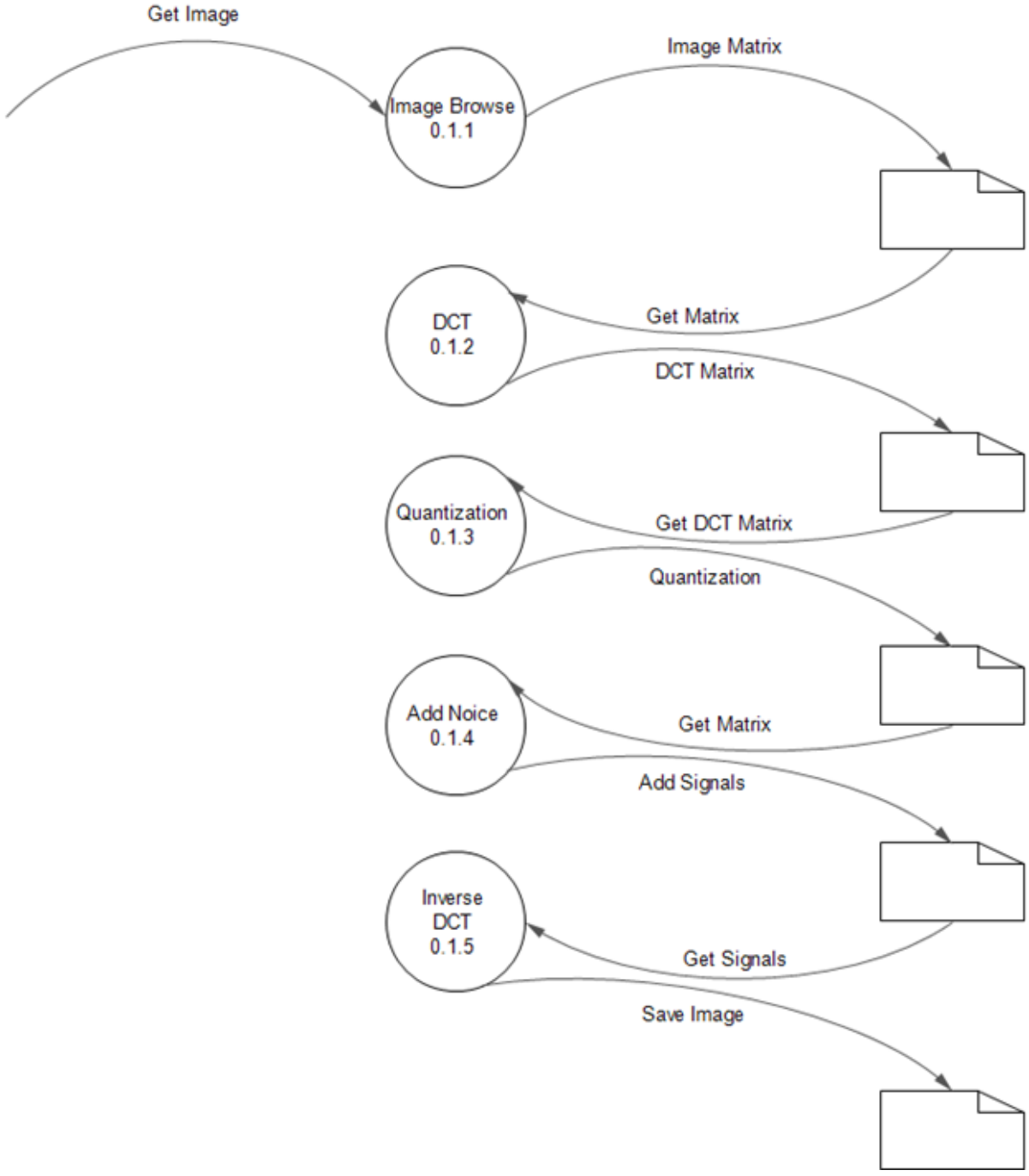

Figure 5.4: Level 2 DFD of Compression

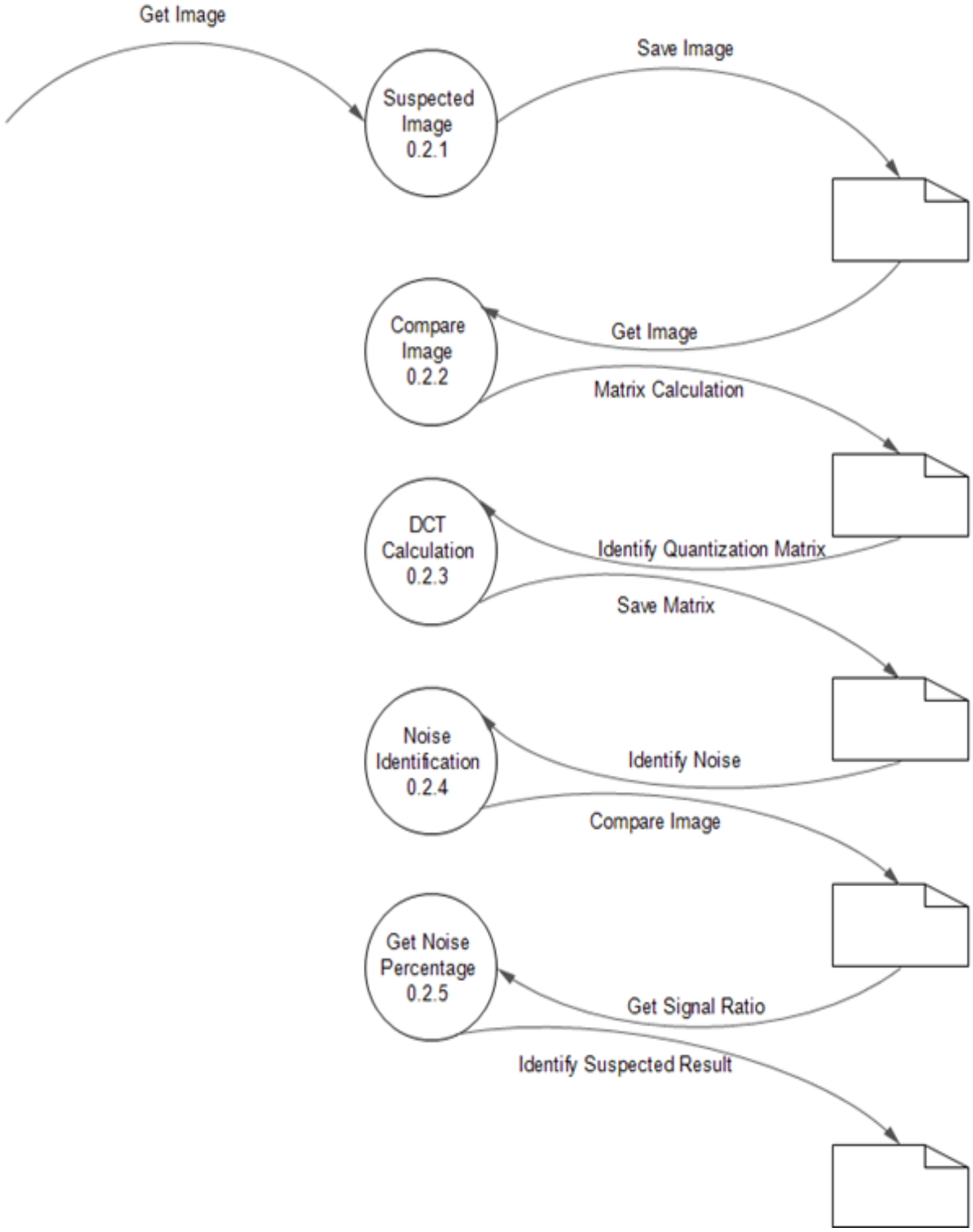

Figure 5.5: Level 2 DFD of Compression Detection

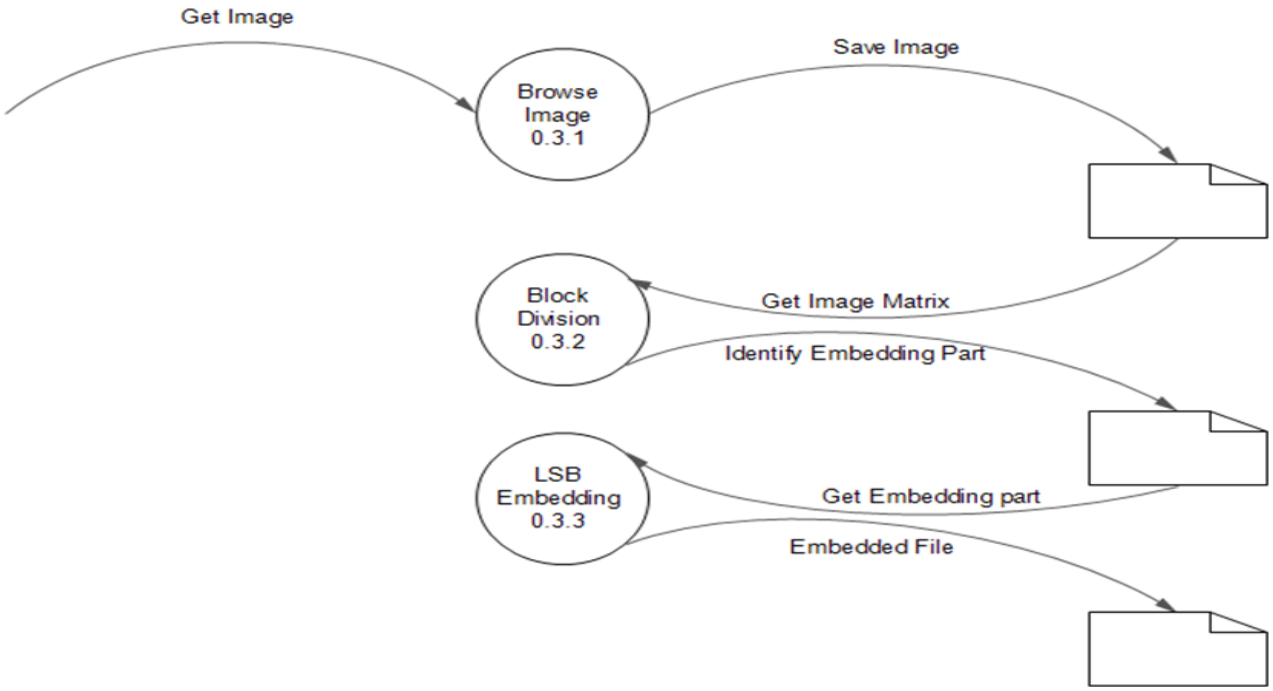

Figure 5.6: Level 2 DFD of Embedded

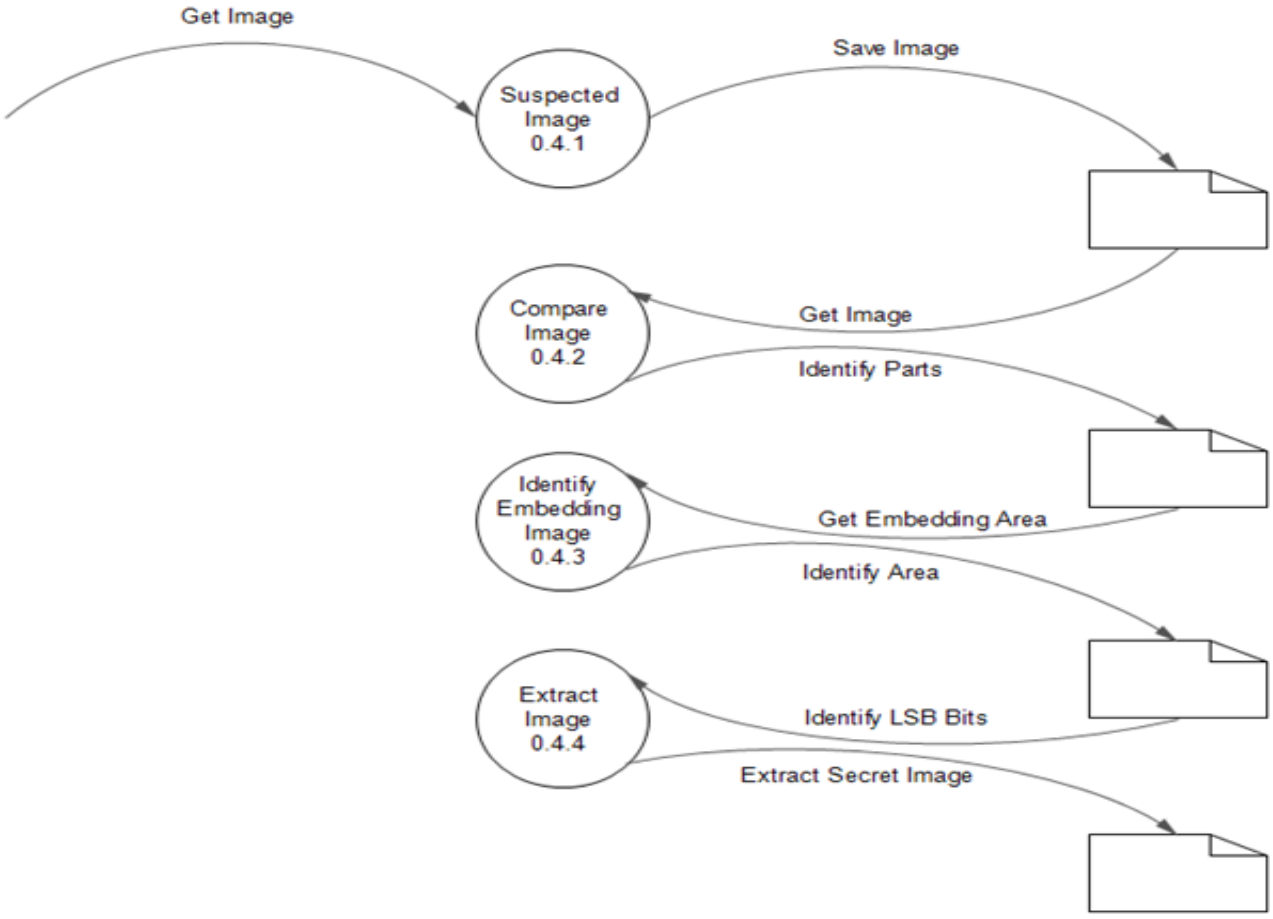

Figure 5.7: Level 2 DFD of Embedded Detection

5.3 Screenshots of user interface

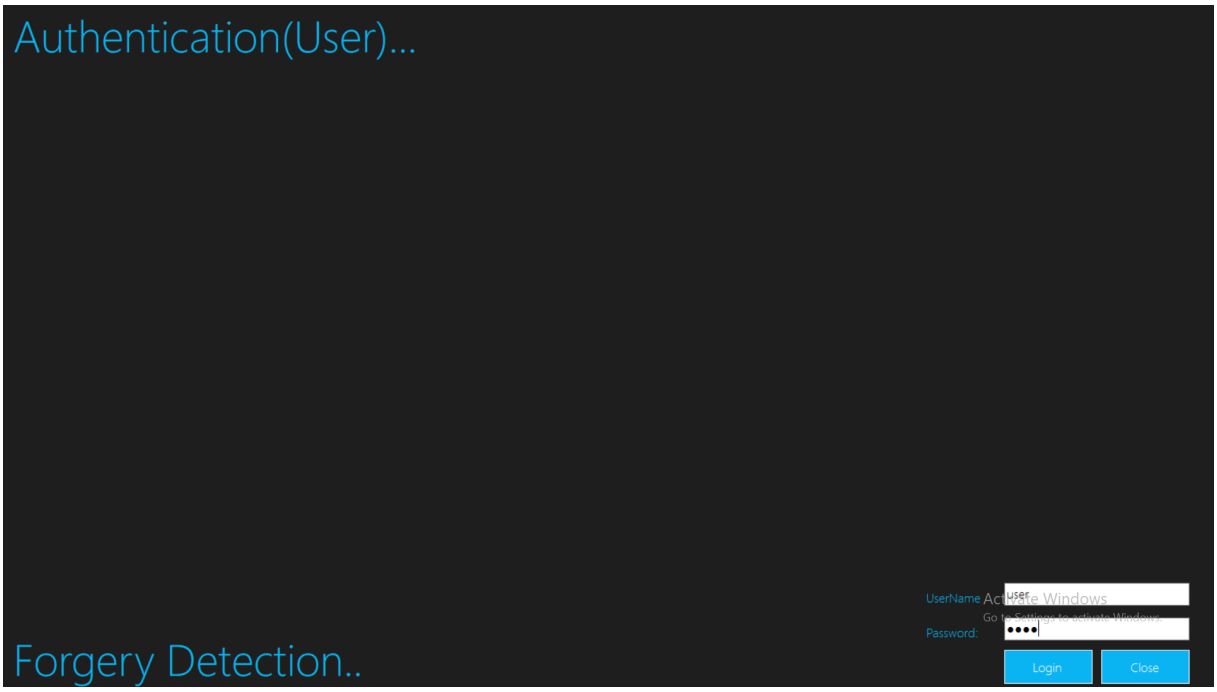

Figure 5.8: Login

Figure 5.9: RGB format

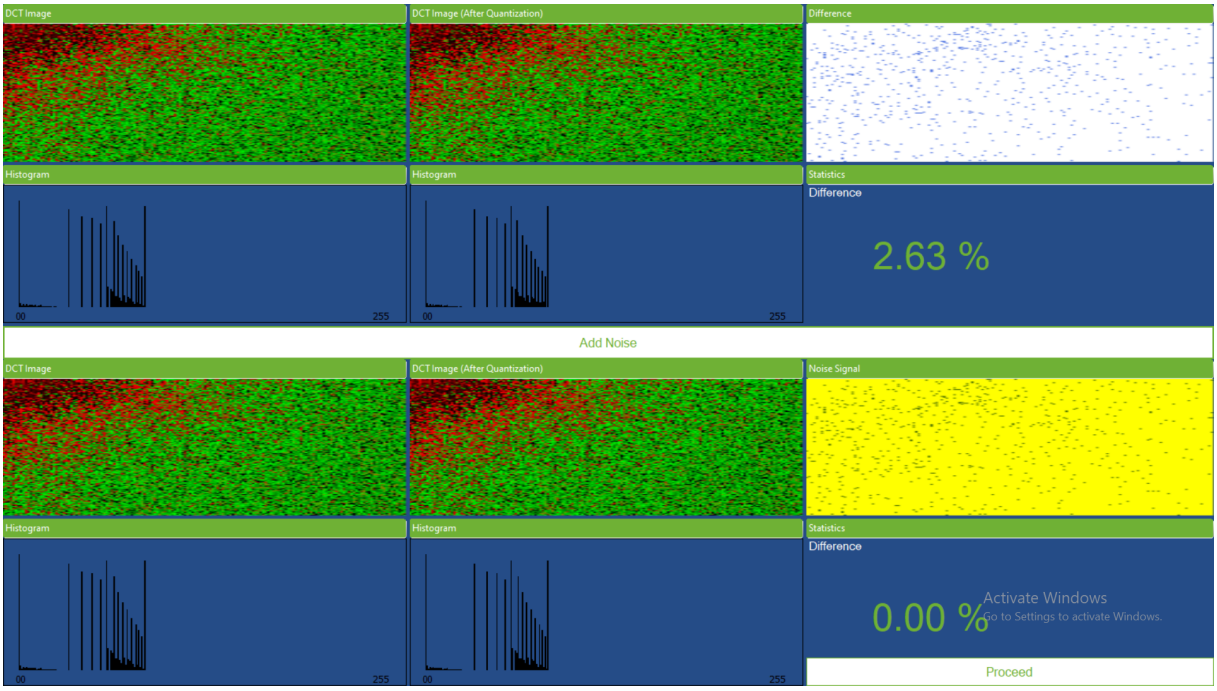

Figure 5.10: Adding noise

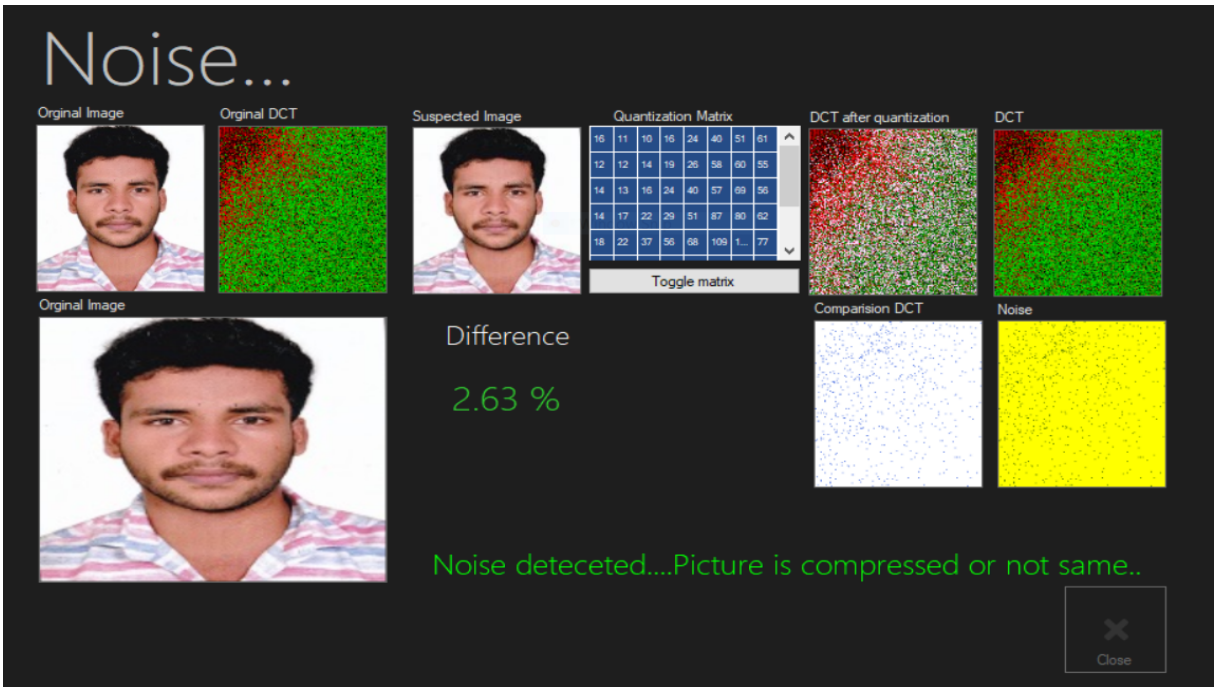

Figure 5.11: Compression detected

Chapter 6

Coding

Algorithm 1 Algorithm for Compression:

1: Select the image to be compressed.

2: Process the image by finding the image properties and forming the histograms (RGB format).

3: Form the RGB matrix (hex value matrix) based on the pixel values.

4: Find the DCT matrix (dividing into 8*8 matrix).

5: Quantize DCT matrix with the standard jpeg quantization matrix.

6: Round off the resultant quantization matrix and perform inverse quantization(multiplication with standard jpeg quantization matrix)to obtain the modified DCT.

7: Compare modified DCT matrix with the original DCT matrix in order to find the difference and then add a noise signal to nullify the difference.

8: Perform inverse DCT on the matrix to get modified RGB matrix.

9: Reconstruct the image from the modified RGB matrix and save the image.

Algorithm 2 Algorithm for Compression Detection:

1: Select original image and suspected image.
2: Check whether image properties are same and if so,proceed to next step or else,the images are different.
3: Check whether histograms are same and if so,proceed to next step or else,the images are different.
4: Check whether DCTs are same and if so,proceed to next step or else,the images are different.
5: Check presence of noise signal and if so,images are not same and has gone through an anti-forensic compression method and find out the percentage of noise signal added; or else the images are same.

Algorithm 3 Algorithm for Embedding data on image:

1: Load the image and select the file to be embedded.
2: Convert to byte stream.
3: Create pixel RGB matrix.
4: Retrieve and create the blue pixel matrix.
5: Interchange the LSB bit with the byte stream of the file to be embedded.
6: Reconstruct the image from the modified matrix.
7: Save the image.

Algorithm 4 Algorithm for Detecting data embedded on image:

1: Select original image and suspected image.
2: Find the DCT of both images.
3: Quantize each of the images.
4: Compare the original matrix with quantized matrix.
5: Check the difference in the LSB bits.
6: Find the difference and extract the data.

Chapter 7

Testing and Implementation

7.1 Testing and various types of testing used.

Once a software is developed,the major activity is to test whether the actual results match with the experimental results.This process is called testing.It's used to make sure that the developed system is defect free.The main aim of testing is to find the errors and missing operations by executing the program.It also ensure that all of the objectives of the project are met by the developer.The objective of testing is not only to evaluate the bugs in the created software but also finding the ways to improve the efficiency,usability and accuracy of it.It aims to measure the functionality,specification and performance of a software program.Tests are performed on the created software and their results are compared with the expected documentation.When there are too much errors occurred,debugging is performed.And the result after debugging is tested again to make sure that the software is error free.The major testing processes applied to this project are unit testing,integration testing and system testing.Unit testing is a procedure used to validate that individual units of source code are working properly.It makes sure that all of the units of the software works as it intended.Individual software modules are joined and assessed as a group during the integration testing phase of software testing.It helps us to find out the faults that may arise when the units are combined.Software or hardware system testing is testing done on a complete,integrated system to see if it complies with the system's requirements.The tables shown below describes the testing process occurred during the development of this project "Contrast Enhancement Based Jpeg forensic detectability".This defines the various steps took to create the project error free.

7.1.1 Unit Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	Upload image into Forensic and Anti forensic Part of the project	Load the image with image properties and histogram	Same as expected	Pass
2	Apply contrast method,DCT and Inverse DCT method for image compression	Successfully compress image with percentage of noise signal	same as expected	Pass
3	Apply LSB method in image embedding method	Successfully embedded image	same as expected	Pass
4	Compare Suspicious and original image	Compare the image with contrast image properties	same as expected	Pass

Table 7.1: Unit test cases and results

7.1.2 Integration Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	Process the input image from the source and suspected folder	Check image contrast properties from the image file and give result	Same as expected	Pass
2	Check for Noise signal Ratio in compression image	Detect the percentage of Noise Ratio with the help of DCT matrix in the suspicious image	Same as expected	Pass
3	Check and identify embedded image in the suspicious image	Detect and Extract Embedded image from the suspicious image using contrast and LSB method	Same as expected	Pass

Table 7.2: Integration cases and result

7.1.3 System Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	Overall testing of the project with Graphical User Interface	Successfully Detected embedded image and compression ratio	Same as expected	Pass

Table 7.3: System test cases and results

Chapter 8

Results and Discussion

The main aim of the project was to detect image manipulation like compression or embedding in jpeg format.And it is observed that the system performs all the functionalities as expected.By using this machine learning model the computer can detect whether an image is undergone any compression or embedded with any data.

8.1 Advantages and Limitations

The proposed system is a machine learning model to evaluate the input images and detect whether it is been compressed or embedded.The proposed system has more advantages over the existing system.The proposed system save a huge amount of time.Like every other system,this system also have it's own disadvantages.But they are negligible while comparing with the advantages and they can be overcame in future.

8.1.1 Advantages

- Can save the time needed for detecting image manipulation.
- Can detect if an image has been compressed even if noise has been added to it.
- Increases the authenticity and integrity of images available on internet.
- Can find out the embedded data in an image.

8.1.2 Limitations

- Anti forensics may conceal the abnormal histogram traces which makes the histogram as normal.Proposed software is not suitable for that type of attacks.
- Proposed algorithms are not strong against post-processing which affects the spotting of zero-height gap bins.

Chapter 9

Conclusion and Future Scope

Contrast Enhancement Based Jpeg forensic detectability is a very useful machine learning application to the current world.It helps us to detect whether the image has undergone any compression or if it is embedded with a file.Various methods are existing to find out digital image manipulations.Through this project,a machine learning model is developed using two contrast enhancement based forensic algorithms via histogram peak/gap artifacts analysis to evaluate and find out the hidden data or noise in a jpeg image format.

The results obtained by the created model seems encouraging and can be improved in future.The rate of errors in the machine learning model is very minimum.The majority of the project was built in C#.DCT algorithm is used to divide image into pixels and before it is converted into RGB matrix.Quantization matrix for jpeg format is used for compression.We expanded our detection to include both uncompressed and previously jpeg compressed photos.

The feature scope of this particular machine learning model can be extended to multiple dimensions.Future development on this topic may lead to decrease in the number of anti forensic cases registering in our country.In the future,works can be done on making these strategies more resistant to post-processing, such as JPEG compression.It is also important to enhance the security on opposing the existing and possible anti-forensic techniques.

Bibliography

[1] S. Bayram, I. Avcubas, B. Sankur, and N. Memon, Image manipulation detection,2006 Journal of Electronic Imaging.

[2] A. Swaminathan, M. Wu, and K. J. R. Liu, Digital image forensics via intrinsic fingerprints,2008 IEEE Transactions on Information Forensics and Security.

[3] H. Cao and A. C. Kot, Manipulation detection on image patches using FusionBoost,2012 June IEEE Transactions on Information Forensics and Security.

[4] M. C. Stamm and K. J. R. Liu, Forensic detection of image manipulation using statistical intrinsic fingerprints, 2010 sep IEEE Transactions on Information Forensics and Security.

[5] P. Ferrara, T. Bianchiy, A. De Rosaz, and A. Piva, Reverse engineering of double compressed images in the presence of contrast enhancement, 2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)