**A Minor Project Abstract on**

**CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING**

**Submitted to the Dept. of Information Technology, SNIST**

**in the partial fulfillment of the academic requirements for the award of**

**B.Tech (Information Technology)**

**under JNTUH**

**by**

**B.Rahul Teja (18311A12J8)**

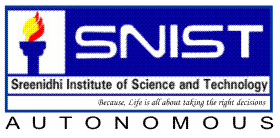
**V.T.Cavin (18311A12P7)**

**K.Sathvik Ram (18311AL3)**

**Under the Guidance of**

**G. Nutan Kumar**

**Associate Professor**



**Department of Information Technology**

**School of Computer Science and Informatics**

**Sreenidhi Institute of Science and Technology (An Autonomous Institution)**

**Yamnampet, Ghatkesar Mandal, R. R. Dist., Hyderabad – 501301**

**affiliated to**

**Jawaharlal Nehru Technological University Hyderabad**

**Hyderabad – 500 085**

**2020–2021**

Department of Information Technology



**CERTIFICATE**

This is to certify that this Internship Project report on“**CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING”**,submitted by **B.RAHUL TEJA(18311A12J8), V.T.CAVIN(1831A12P7), K.SATHVIK RAM(18311A12L3)**

in the year 2021 in partial fulfillment of the academic requirements of Jawaharlal Nehru Technological University for the award of the degree of Bachelor of Technology in Information Technology, is a bonafide work that has been carried out by them as part of their Minor Project during summer (2021), under our guidance. This report has not been submitted to any other institute or university for the award of any degree.

Dr. V.V.S.S.S. Balaram

Prof & Head,

Department of IT

**G. Nutan Kumar**,

Associate Professor,

Department of IT

Project coordinator

Internship coordinator

ernship coordinator

Internship coordinator

coordinator

Project Coordinator

External Examiner

**ACKNOWLEDGEMENT**

I would like to express my gratitude to all the people behind the screen who helped me to transform an idea into a real application.

I would like to express my heart-felt gratitude to my parents without whom I would not have been privileged to achieve and fulfill my dreams. I am grateful to our principal, **Dr.T.Ch.Siva Reddy,** who most ably run the institution and has had the major hand in enabling me to do my project.

I profoundly thank **Dr. V.V.S.S.S.Balaram,** Head of the Department of Information Technology who has been an excellent guide and also a great source of inspiration to my work.

I would like to thank my internal guide **G. Nutan Kumar** technical guidance, constant encouragement and support in carrying out my project at college.

The satisfaction and euphoria that accompany the successful completion of the task would be great but incomplete without the mention of the people who made it possible with their constant guidance and encouragement crowns all the efforts with success. In this context, I would like thank all the other staff members, both teaching and non-teaching, who have extended their timely help and eased my task.

* 1. **B.RAHUL TEJA (18311A12J8)**
  2. **V.T.CAVIN (1831A12P7)**
  3. **K.SATHVIK RAM (18311A12L3)**

**INDEX**

|  |  |
| --- | --- |
| **Details of Contents** | **Pg No.** |
|  | |
| **Abstract** | **1** |
|  | |
| **1. Introduction** | **2-4** |
| 1.1Purpose of the Project | **2** |
| 1.2Existing Methodology and its Disadvantages | **2-3** |
| 1.3Proposed System and its Advantages | **3** |
| 1.4Objective | **4** |
| **2. Analysis** | **5-8** |
| 2.1 Approach of the Project | **5** |
| 2.2 Fraud Detection Techniques | **5-6** |
| 2.3 Reasons why Frauds increased in 2020 | **6-8** |
| 2.4 Main Challenges Involved | **8** |
| **3. Algorithm** | **9-11** |
| 3.1 Structural Similarity Index Measure | **9** |
| 3.2 Formula Components  3.3 General Applications  3.4 Structural Similarity based metrics  3.5 SSIM properties | **10**  **10**  **11**  **11** |
| **4. Methodology** | **12-14** |
| 4.1 Image Processing | **12** |
| 4.2 Processed Images | **13** |
| 4.3 Packages Used | **14** |
| **5. Implementation** | **15-21** |
| 5.1Proposed Approach | **15-19** |
| **6. Conclusion** | **20** |
| **7. References** | **21** |

**ABSTRACT**

Credit card fraud refers to the physical loss of credit card or loss of sensitive credit card information. Many machine-learning algorithms can be used for detection. Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. Increase in fraud rates, researchers started using different machine learning methods to detect and analyse frauds in online transactions. One example is *phishing*. Attackers can manipulate images ever-so-slightly to trick unsuspecting users who don’t validate the URL into thinking they are logging into their banking website — only to later find out that it was a scam.

Developing a phishing detection system is obviously much more complicated than simple image differences, but we can still apply these techniques to determine if a given image has been manipulated.Comparing logos and known User Interface (UI) elements on a web page to an existing data-set could help reduce phishing attacks .In this project we compute the difference between two images, and view the differences side by side using OpenCV, scikit-image, and Python.

Using the **Structural Similarity Index (SSIM)**,Using this method, we were able to easily determine if two images were identical or had differences due to slight image manipulations, compression artifacts, or purposeful tampering.

1. **INTRODUCTION**

**1.1 Purpose of the Project:**

Our [fraud management solutions](https://www.softwareag.com/resources/Fraud-management-solutions) enable your system to:

* Detect fraud quickly to prevent revenue loss
* Monitor millions of transactions from multiple channels in real time
* Identify unusual patterns in fraudulent activity by combining real-time transactional data and historical analysis of customer behavior
* Adapt quickly to new fraud techniques by rapidly applying new rules
* Update and refine fraud rules in real time
* Balance and prioritize workloads to ensure maximum recovery
* Get complete auditing transparency and traceability

**1.2**  **Existing Methodology and its Disadvantages:**

Fraud detection is a set of activities undertaken to prevent money or property from being obtained through false pretenses. Fraud detection is applied to many industries such as banking or insurance. In banking, fraud may include forging checks or using stolen credit cards. Other forms of fraud may involve exaggerating losses or causing an accident with the sole intent for the payout.

With an unlimited and rising number of ways someone can commit fraud, detection can be difficult to accomplish. Activities such as reorganization, downsizing, moving to new information systems or encountering a cyber security breach could weaken an organization's ability to detect fraud.

This means techniques such as real-time monitoring for frauds is recommended. Organizations should look for fraud in financial transactions, location, devices used, initiated sessions and authentication systems.

Fraud can be committed in a number of different ways and in a number of different settings. For example, fraud can be committed in banking, insurance, government and in healthcare sectors.

One common type of fraud in banking is customer account takeover, where someone illegally gains access to a victim’s bank account using [bots](https://whatis.techtarget.com/definition/bot-robot). Other examples of fraud in banking include the use of malicious applications, the use of false identities, money laundering, credit card fraud and mobile fraud.

Fraud in insurance can include premium diversion fraud, which is the embezzlement of insurance premiums; or frees churning, which is excessive trading by a stockbroker to maximize commissions. Other forms of insurance fraud include asset diversion, workers compensation, car accident, stolen or damaged car, and house fire fraud. The motive behind all insurance fraud is financial profits.

Government fraud is committing fraud against federal agencies such as the departments of Health and Human Services, Transportation, Education, or Energy. Types of government fraud include billing for unnecessary procedures, overcharging for items that cost much less, providing old equipment when billing for new or reporting hours worked for a worker that does not exist.

Healthcare fraud includes drug fraud and medical fraud, as well as encompassing some insurance fraud. Healthcare fraud is committed when someone defrauds an insurer or government health care program.

**1.3 Proposed System and its Advantages:**

Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. Increase in fraud rates, researchers started using different machine learning methods to detect and analyse frauds in online transactions.

The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. Where cardholders are clustered into different groups based on their transaction amount. Then using sliding window strategy , to aggregate the transaction made by the cardholders from different groups so that the behavioural pattern of the groups can be extracted respectively. Later different classifiers are trained over the groups separately. And then the classifier with better rating score can be chosen to be one of the best methods to predict frauds. Thus, followed by a feedback mechanism to solve the problem of concept drift. In this paper, we worked with European credit card fraud data set.

It fraud is an ongoing problem for almost all industries in the world, and it raises millions of dollars to the global economy each year. Therefore, there is a number of research either completed or proceeding in order to detect these kinds of frauds in the industry. These researches generally use rule-based or novel artificial intelligence approaches to find eligible solutions. The ultimate goal of this paper is to summarize state-of-the-art approaches to fraud detection using artificial intelligence and machine learning techniques.

While summarizing, we will categorize the common problems such as imbalanced dataset, real time working scenarios, and feature engineering challenges that almost all research works encounter, and identify general approaches to solve them.

The imbalanced dataset problem occurs because the number of legitimate transactions is much higher than the fraudulent ones whereas applying the right feature engineering is substantial as the features obtained from the industries are limited, and applying feature engineering methods and reforming the dataset is crucial. Also, adapting the detection system to real time scenarios is a challenge since the number of credit card transactions in a limited time period is very high. In addition, we will discuss how evaluation metrics and machine learning methods differentiate among each research.

**1.4 Objectives:**

Developing a phishing detection system is obviously much more complicated than simple image differences, but we can still apply these techniques to determine if a given image has been manipulated.Comparing logos and known User Interface (UI) elements on a webpage to an existing dataset could help reduce phishing attacks.

The key objective of any credit card fraud detection system is to identify suspicious events and report them to an analyst while letting normal transactions be automatically processed. For years, financial institutions have been entrusting this task to rule-based systems that employ rule sets written by experts.

1. **ANALYSIS**

**2.1 Approach of the Project:**

Our project is about analyzing difference between pair of images. We have used Visualization techniques with Machine Learning concepts to build and train a model which is capable of predicting the difference between pair of images on input as processed pair of images.

Using python libraries such as :

1. OpenCV
2. Numpy
3. Pillow
4. Skicit-Image
5. Imutils
6. Argparse

We are able to take data of processed images and analyze it precisely with depth.In order to compute the difference between two images we’ll be utilizing the Structural Similarity Index. This method is already implemented in the [scikit-image](http://scikit-image.org/docs/stable/api/skimage.measure.html#skimage.measure.compare_ssim) library for image processing.The trick is to learn how we can determine exactly where, in terms of *(x, y)*-coordinate location, the image differences are.

Computing image difference is so important ,one example is phishing. Attackers can manipulate images ever-so-slightly to trick unsuspecting users who don’t validate the URL into thinking they are logging into their banking website only to later find out that it was a scam.

Comparing logos and known User Interface (UI) elements on a webpage to an existing dataset could help reduce phishing attacks (a big thanks to Chris Cleveland for passing along [PhishZoo: Detecting Phishing Websites By Looking at Them](http://www1.icsi.berkeley.edu/~sadia/papers/phishzoo-icsc_final.pdf) as an example of applying computer vision to prevent phishing).

Developing a phishing detection system is obviously much more complicated than simple image differences, but we can still apply these techniques to determine if a given image has been manipulated.

**2.2 Fraud Detection Techniques:**

Fraud is typically an act which involves many repeated methods; making searching for patterns a general focus for fraud detection. For example, data analysts can prevent insurance fraud by making algorithms to detect patterns and anomalies.

Fraud detection can be separated by the use of statistical data analysis techniques or artificial intelligence ([AI](https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence)).

Statistical data analysis techniques include the use of:

* Regression analysis
* Probability distributions and models.
* Data matching
* Calculating statistical parameters

AI techniques used to detect fraud include the use of:

* [Data mining](https://searchsqlserver.techtarget.com/definition/data-mining)- Which can classify, group and segment data to search through up to millions of transactions to find patterns and detect fraud.
* [Neural networks](https://searchenterpriseai.techtarget.com/definition/neural-network)- Which can learn suspicious looking patterns, and use those patterns to detect them further.
* [Machine learning](https://searchenterpriseai.techtarget.com/definition/machine-learning-ML)- Which can automatically identify characteristics found in fraud.
* [Pattern recognition](https://whatis.techtarget.com/definition/pattern-recognition)- Which can detect classes, clusters and patterns of suspicious behavior.

**2.3 Reasons why Frauds increased in 2020:**

Digital fraud is a problem businesses have been facing since the advent of e-commerce in the 1990s, and its threat only increases with each passing year. In fact, [Experian found](https://www.experian.com/content/dam/marketing/na/assets/im/decision-analytics/reports/global-identity-and-fraud-report-2020.pdf) that losses from fraudulent identities increased from 51% in 2017 to 57% in 2019. And [according to PwC](https://www.pwc.com/gx/en/services/advisory/forensics/economic-crime-survey.html), these crimes cost companies $42 billion in the last 24 months. But what is causing digital fraud to rise year over year? From current trends and consumer attitudes to technological enhancements and more sophisticated tactics, let’s take a look at the top nine reasons digital fraud is rapidly increasing:

1. **Chaos caused by the global COVID-19 crisis.**Opportunistic hackers are taking advantage of the chaotic, global crisis to commit even more fraudulent activity. Tactics include stealing stimulus checks and unemployment benefits, collecting payments for fake COVID-19 treatments, tricking Americans into donating to fraudulent charities, and more. In fact, there were [1.1 billion fraud attacks](https://www.securitymagazine.com/articles/93008-1-billion-fraud-attacks-detected-since-the-beginning-of-2020#:~:text=Arkose%20Labs%20revealed%20new%20data,and%20stopped%201.1%20billion%20attacks.) in the first half of 2020, which is double the attack volume compared to the second half of 2019. And according to the [Federal Trade Commission](https://www.ftc.gov/enforcement/data-visualizations/explore-data), Americans have lost $145 million to fraud related to COVID-19.
2. **A changing e-commerce landscape.**Another trend impacting the rise in fraud is more retail purchases shifting online. In particular, card not present (CNP) transactions have increased dramatically in recent years, with these transactions accounting for [27% of all debit transactions](https://www.pulsenetwork.com/content/dam/pulse/assets/2020%20Debit%20Issuer%20Study.pdf) in 2019 and increasing 10 times faster than card present transactions. Now, as more consumers are staying home as a result of COVID-19, even more commerce has moved online. This trend makes it even easier for bad actors to make fraudulent transactions. Point of sale (POS) lending has also become more common, allowing customers to make payments in installments or take out loans for purchases both large and small. While POS lending makes it easy for consumers to gain approval and make a purchase in a matter of minutes, it also opens the door to fraudsters.
3. **The advent of new marketplace platforms.**From social networks and dating apps to food delivery, alternative transportation, and vacation rentals, digital channels have revolutionized almost every industry. Throughout this year, country-wide quarantines have caused an even greater spike in mobile application use, with consumers ordering the delivery of everything from groceries to automobiles. With the growing number of marketplace platforms and services available and their widespread popularity -- especially in recent months -- fraudsters have shifted their tactics to take advantage of rising in-app and online marketplace purchases.
4. **Payments moving online.**In addition to consumers transacting more in online marketplaces, they are also using peer-to-peer payment (P2P) and eWallet apps more often. These apps are most popular in Europe and Asia but are becoming increasingly popular in the U.S. too, with [71% of Americans](https://www.aarp.org/content/dam/aarp/research/surveys_statistics/econ/2020/peer-to-peer-payment-platforms-national-report.doi.10.26419-2Fres.00383.001.pdf) saying they have used a P2P payments platform. Users turn to these platforms to digitally split dinner checks with friends, send money to family members in other parts of the world, pay for services from a local vendor, and more. But with more than half of P2P transactions taking place between consumers and an unknown entity, the fraud risk is high.
5. **Increasingly digital banking services.**Today’s consumers demand more online and mobile services from their financial institutions.As a result, legacy banks are going digital. They are doing more account onboarding and transaction approvals online and deemphasizing in-person transactions, which makes it harder to verify identities. Also in response to consumer demands, a new breed of “challenger banks” – born and doing business entirely in the online world – have emerged and are differentiating themselves by providing easy-to-use and digital-native experiences. A majority of these institutions’ customers are those who have “thin file” credit histories (i.e., don’t have much credit data). Less data means a greater risk of fraud.
6. **New consumer expectations.**Today’s consumers also expect their data to be secure. Yet they will abandon any transaction that takes too long, requires too much data, or is too complex. In fact, [92% of consumers](https://www.vansonbourne.com/client-research/29081801ep) expect a fast, frictionless experience while also getting one that is as trustworthy and secure as possible. These steep expectations are causing banks and retailers to juggle preventing losses with keeping fraud prevention measures from rejecting good customers and transactions. Cyber criminals understand the struggle these organizations face and take advantage of those that fail to strike the right balance of secure, yet frictionless customer experiences.
7. **More sophisticated fraud tactics.**Due to an increasing number of data breaches over recent years, fraudsters can more easily access PII (personally identifiable information) and use it against consumers. For instance, fraudsters combine real and fake data (such as an address from one person mixed with another’s social security number) to create new, synthetic identities that are harder to detect. Then, they establish open bank accounts and cards, acting like legitimate customers. Once they’ve established strong credit scores, the fraudsters ask for higher credit limits or larger loans and simply stop paying. Synthetic identity fraud is damaging for consumers, but also expensive for lenders too, costing them [$6 billion annually](https://www.cnbc.com/2020/01/16/criminals-using-frankenstein-identities-to-steal-from-banks.html)**.** Fraudsters also leverage PII for account takeover. By using passwords and credentials obtained via data breaches or social engineering, they can gain control over accounts and make fraudulent online purchases. These transactions can be as minor as buying groceries on a debit card or as severe as using someone else’s account to take out a mortgage. Account takeover fraud is a serious threat for consumers that [Juniper Research predicts](https://www.juniperresearch.com/researchstore/fintech-payments/online-payment-fraud) will result in losses exceeding $200 billion between 2020 and 2024.
8. **Unclear legal jurisdiction of cross-border fraud.**Global commerce gives today’s online retailers and marketplaces an opportunity to reach even more customers. [Forrester estimates](https://go.forrester.com/press-newsroom/cross-border-ecommerce-will-reach-627-billion-by-2022/) cross-border e-commerce sales will reach $627 billion in 2022, which would represent 20% of all e-commerce. However, cross-border transactions don’t come without some risk. Because they typically encompass multiple countries, it is difficult for individual jurisdictions to properly monitor for fraud risk. Further, data privacy and protection regulations vary across regions -- if they exist at all -- making it even easier for fraudsters to commit cross-border transaction crimes.
9. **Technological advancements.**Today, fraud has also accelerated and grown even more sophisticated due to the rise of e-commerce, mobile payments, and computing power. Many of the same technologies that companies rely on to innovate and rapidly introduce new products and services are also being adopted by fraudsters. Criminals can more easily commit fraud using cheap, on demand compute power or deploy algorithms using machine learning that are more subtle and traditional rules-based fraud prevention systems that organizations have relied on for years now struggle to keep up.

**2.4 Main Challenges Involved:**

1. Enormous Data is processed every day and the model build must be fast enough to respond to the scam in time.

2. Imbalanced Data i.e most of the transactions (99.8%) are not fraudulent which makes it really hard for detecting the fraudulent ones.

3. Data availability as the data is mostly private.

4. Misclassified Data can be another major issue, as not every fraudulent transaction is caught and reported.

5. Adaptive techniques used against the model by the scammers.

**3. ALGORITHM**

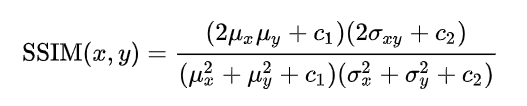
**3.1 Structural Similarity Index Measure:**

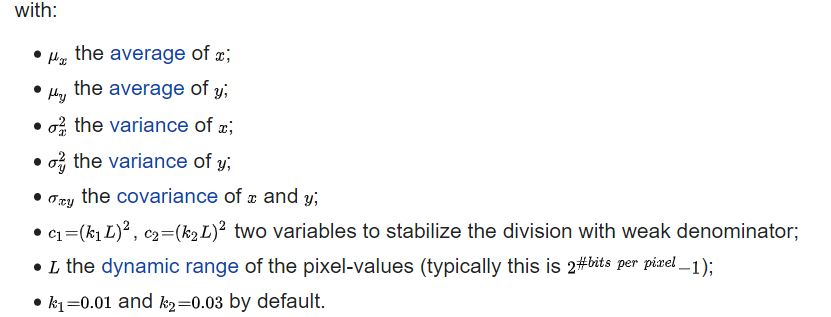
The **structural similarity** **index measure** (**SSIM**) is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. SSIM is used for measuring the similarity between two images. The SSIM index is a [full reference metric](https://en.wikipedia.org/wiki/Video_quality#Classification_of_objective_video_quality_metrics); in other words, the measurement or prediction of [image quality](https://en.wikipedia.org/wiki/Image_quality) is based on an initial uncompressed or distortion-free image as reference.

SSIM is a perception-based model that considers image degradation as perceived change in structural information, while also incorporating important perceptual phenomena, including both luminance masking and contrast masking terms. The difference with other techniques such as [MSE](https://en.wikipedia.org/wiki/Mean_squared_error) or [PSNR](https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio) is that these approaches estimate absolute errors. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close.

These dependencies carry important information about the structure of the objects in the visual scene. Luminance masking is a phenomenon whereby image distortions (in this context) tend to be less visible in bright regions, while contrast masking is a phenomenon whereby distortions become less visible where there is significant activity or "texture" in the image.

The SSIM index is calculated on various windows of an image. The measure between two windows and  of common size *N*×*N* is:

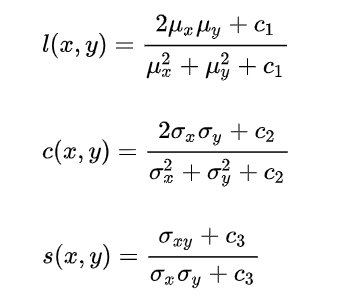




**3.2 Formula Components:**

The SSIM formula is based on three comparison measurements between the samples of  x and y : luminance(l),contrast( c) ,structure(s).

The individual comparison functions are:



with, in addition to above definitions:

C3=c2/2.

**3.3 General Applications:**

**Image Compression**: In lossy [image compression](https://en.wikipedia.org/wiki/Image_compression), information is deliberately discarded to decrease the storage space of images and video. The MSE is typically used in such compression schemes. According to its authors, using SSIM instead of MSE is suggested to produce better results for the decompressed images.

**Pattern Recognition**: Since SSIM mimics aspects of human perception, it could be used for recognizing patterns. When faced with issues like image scaling, translation and rotation, the algorithm's authors claim that it is better to use CW-SSIM.

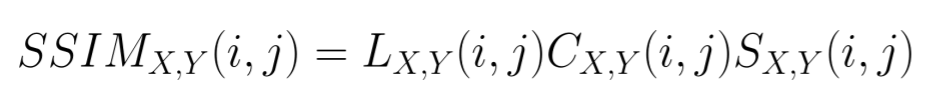
Which is insensitive to these variations and may be directly applied by template matching without using any training sample. Since data-driven pattern recognition approaches may produce better performance when a large amount of data is available for training, the authors suggest using CW-SSIM in data-driven approaches.

**3.4 Structural Similarity based metrics:**

Measure loss of structure in the image as opposed to just any deviation with respect to reference

• Loss of image structure measured locally through – Luminance similarity -Contrast similarity – Structural similarity   
• Perform average of local measure across the image.

SSIM has revolutionized QA – many applications in other domains as well

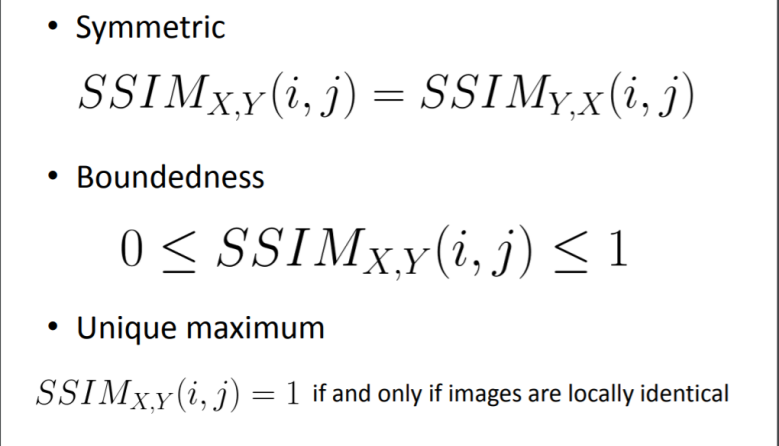


• L(i,j) – luminance similarity at location (i,j)

• C(i,j) – contrast similarity

• S(i,j) – structure similarity

**3.5 SSIM properties:**



**4. METHODOLOGY**

**4.1 Image Processing:**

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps:

* Importing the image via image acquisition tools;
* Analyzing and manipulating the image;
* Output in which result can be altered image or report that is based on image analysis.

There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques.

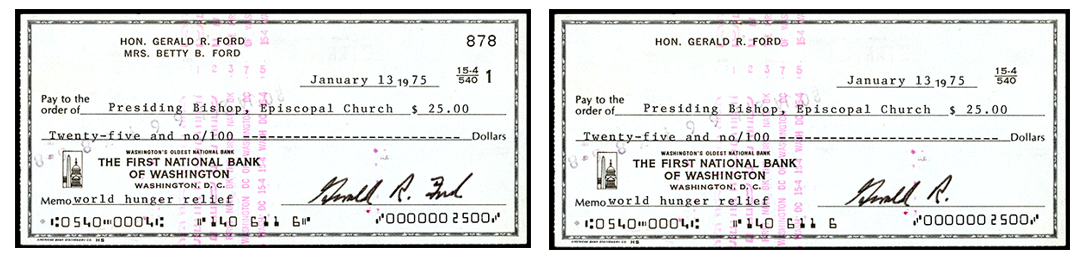
Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

In this lecture we will talk about a few fundamental definitions such as image, digital image, and digital image processing. Different sources of digital images will be discussed and examples for each source will be provided. The continuum from image processing to computer vision will be covered in this lecture. Finally we will talk about image acquisition and different types of image sensors.

**4.2 Processed Images:**

****





**4.3 Packages Used:**

1. **Scikit-image :** From this library we import **compare\_ssim**.

-- The structural similarity index measure (SSIM) is a method for predicting the perceived quality of digital images.

**--** SSIM is used for measuring the similarity between two images.

-- Using this method, we were able to easily determine if two images were identical or had differences.

1. **Argparse:** Parser for command-line options, arguments and sub-commands. The **argparse** module makes it easy to write user-friendly command-line interfaces.The **argparse** module also automatically generates help and **usage** messages and issues errors when users give the program invalid arguments.
2. **Imutils:** Imutils are a series of convenience functions to make basic image processing functions such as translation, rotation, resizing, skeletonization, and displaying Matplotlib images easier with OpenCV.
3. **Open CV:** open cv used for various operations on images like reading, color conversion,bounding rectangle,printing and finding contours etc.
4. **Pil:** The Python Imaging Library, or **PIL** for short, is an open source library for loading and manipulating images.
5. **Numpy:** Numpy is used , adding, support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

**5. IMPLEMENTATION**

**5.1 Proposed Approach:**

Reading input images:

-- we load each image from disk and converting them to gray scale.



**Analyzing the images using histogram** :

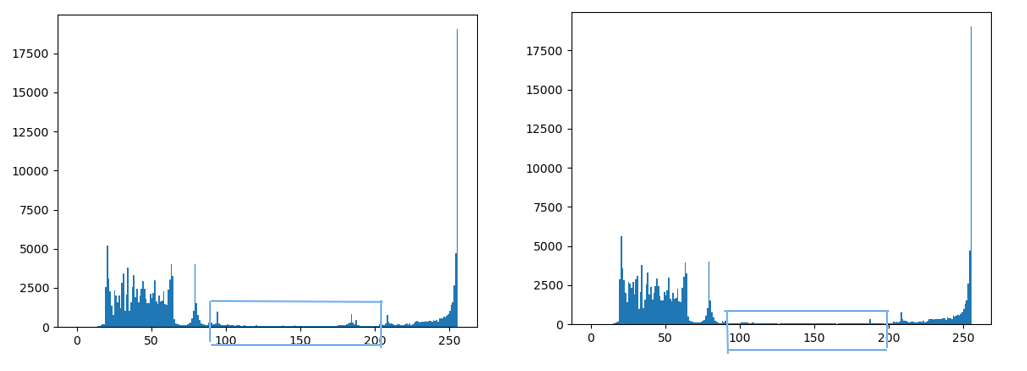
Histogram is considered as a graph or plot which is related to frequency of pixels in an Gray Scale Image.

With pixel values (ranging from 0 to 255). Grayscale image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information where pixel value varies from 0 to 255.

Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest where Pixel can be considered as a every point in an image.

X-axis:Pixel values

Y-axis:No.of Pixels



**Computing the SSIM between two gray scale images:**

Using the **compare\_ssim** function from **scikit-image**, we calculate a Score and difference image, diff The score represents the structural similarity index between the two input images.

This value can fall into the range *[-1, 1]* with a value of one being a “perfect match”.

The diff image contains the actual *image differences* between the two input images that we wish to visualize.

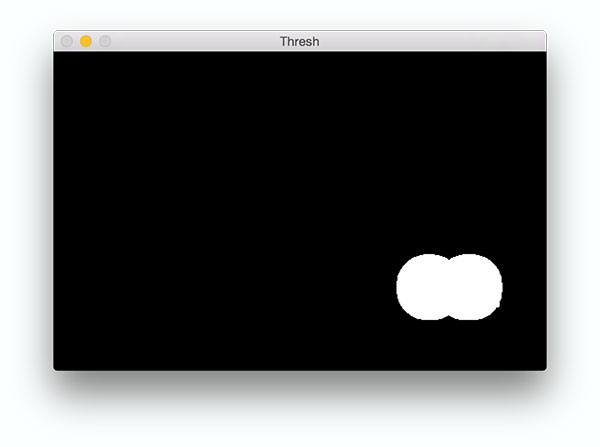
The difference image is currently represented as a floating point data type in the range *[0, 1]* so we first convert the array to 8-bit unsigned integers in the range *[0, 255]* , before we can further process it using OpenCV.

**Finding the contours so that we can place rectangles around the regions identified as “different”:**

We threshold our diff image using both **cv2.THRESH\_BINARY\_INV** and **cv2.THRESH\_OTSU** both of these settings are applied at the same time using the vertical bar ‘or’ symbol, ”|”.

Subsequently we find the contours of thresh.The ternary operator simply accommodates difference between the **[cv2.findContours return signature](https://pyimagesearch.com/2015/08/10/checking-your-opencv-version-using-python/)** in various versions of OpenCV.

The image below clearly reveals the ROIs of the image that have been manipulated:



Using thresholding to highlight the image differences using OpenCV and Python.

**Draw rectangles around the different regions on each image:**

We loop over our contours,cnts. First, we compute the bounding box around the contour using the **cv2.boundingRect** function.

We store relevant *(x, y)*-coordinates as **x** and **y** as well as the width/height of the rectangle as **w** and**h**.Then we use the values to draw a red rectangle on each image with **cv2.rectangle.**

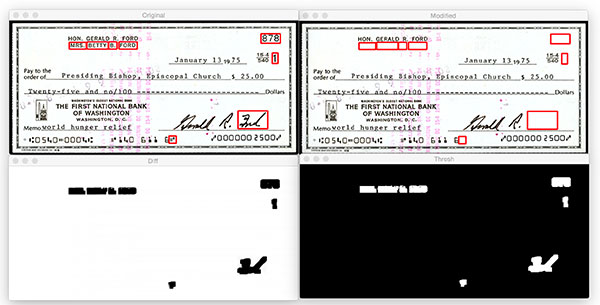
Finally, we show the comparison images with boxes around differences, the difference image, and the thresholded image.



Visualizing image differences using Python and OpenCV.

**Other Images Output:**





Computing image differences and highlighting the regions that are different.

**6. CONCLUSION**

In conclusion,our model is capable of predicting, the difference between given two pair of images.We have focused on the key parameters like sizing,processing of images,number of pixels etc. We have also calculated the number of pixels different in the given pair of images to understand the difference precisely by the end user.

Using this method we can also implement in several areas in real time like:

* This can be implemented in phishing to find the criminal frauds.
* In further Comparing logos and known User Interface (UI) elements on a web page to an existing data set could help reduce phishing attacks as an example of applying computer vision to prevent phishing.

Ultimately, the current predictor is robust and fairly accurate in its result derivation. It understands the data being fed to it and can train in a scalable, dynamic manner to include larger, more complex data sets in the future.

On a complex image like a check it is often difficult to find all the differences with the naked eye. Luckily for us, we can now easily compute the differences and visualize the results with this handy script made with Python, OpenCV, and scikit-image.

We have done about how to compute image differences using OpenCV, Python, and scikit-image’s Structural Similarity Index (SSIM). Based on the image difference we also learned how to mark and visualize the different regions in two images.

1. **REFERENCES**

1.<https://scikit-image.org/docs/dev/auto_examples/transform/plot_ssim.html>

2.<https://ieeexplore.ieee.org/abstract/document/6889510>

3.<https://answers.opencv.org/question/213095/visualize-differences-between-two-images/>

4.<https://codedeepai.com/finding-difference-between-multiple-images-using-opencv-and-python/>