Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

IMAGE FORGERY DETECTION USING OPEN-CV AND MD5

P.AppalaNaidu Professor, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India.

appalanaidu.p@raghuenggcollege.in

P.S.J.Sanjana, S.Jyothika, T.Vikash Students, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India.

19981a05c8@raghuenggcollege.in, 19981a05e6@raghuenggcollege.in, 19981a05g7@raghuenggcollege.in

Abstract*— The digital industry heavily relies on digital images, but image forgery has become a growing concern, posing a threat to data authenticity. Although digital photos were designed to improve efficiency and accessibility, some people have abused the technology. Traditional methods for detecting fake images exist, but with the increasing rate of image forgery, more advanced approaches are required. While OpenCV and MD5-based methods have had an impact on image fraud detection, they have limitations in detecting specific types of fraud. As a result, it is necessary to develop and improve efficiency.

Keywords— ***Digital images, Image forgery, Authenticity, Traditional methods, OpenCV, Fraud detection, MD5.***

I. INTRODUCTION

Advancements in technology and globalization have made digital cameras widely accessible and affordable. As a result, people capture and collect numerous images using various camera sensors, often in soft copy for online documents and sharing on social media. Images play an essential role in the digital world for storing and distributing data because they are easily accessible and can convey information even to those who are illiterate. There are various tools available for editing images that were initially developed to enhance the quality of the pictures. However, some individuals misuse these tools to create fake images and spread false information. This is a significant concern because the consequences of falsified images can be severe and often irreversible.
The idea behind this concept is to reduce the forgery that is happening and help determine the difference between an actual authentic image and a fake image. This will help to identify more easily and more accurately giving better efficiency.
Image forgeries can be classified into two types: image splicing and copy-move.

Image Splicing: Image splicing is a type of digital image forgery where one or more regions from different images are combined to form a new image. This is often done to create a composite image that appears to be authentic but contains elements that were taken from different sources. Detecting image splicing requires sophisticated techniques such as analyzing inconsistencies in the lighting, color, and texture of different parts of the image [1][5].
Copy-Move: Copy-move is a type of image forgery where a part of an image is copied and pasted into another location within the same image with the intent of hiding or duplicating content. This technique is often used to create multiple instances of a person or object in an image or to cover up unwanted elements. Advanced digital image processing techniques, such as content-based picture retrieval, feature extraction, and pattern recognition, can be used to detect copy-move forgery and identify duplicated parts in the image [6].

The following is how the rest of the essay is structured: Section 2 discusses significant studies on image forgery detection using image-splicing convolutional neural networks. Section 3 describes the study's approach, as well as the research, and explains the mathematical principles used in this work. Finally, we discussed the analysis of our project's results in Section 4.

II. LITERATURE SURVEY

Xiao et al [1]
They presented a two-part approach to detect splicing forgery in images. It uses a C2RNet and adaptive clustering to extract differences in image properties between tampered and un-tampered regions. The proposed method effectively detects splicing forgeries and achieves promising results compared to state-of-the-art approaches.
Additionally, the proposed approach is computationally efficient and achieves promising results even under different attack conditions. The combination of C2RNet and adaptive clustering allows for the accurate detection of splicing forgery by learning the differences in image properties between tampered and un-tampered regions. Overall, the method provides a robust solution for detecting splicing forgery in images.

Kwon et al [2]
CAT-Net is a fully convolutional neural network intended for picture splicing localization. The network incorporates RGB and DCT streams to learn the forensic aspects of compression artifacts in both domains. The RGB stream considers several

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

resolutions to deal with the spliced object's shapes and sizes, whereas the DCT stream is pre-trained on double JPEG detection to make use of JPEG artifacts. The suggested method outperforms state-of-the-art neural networks in both JPEG and non-JPEG image localization, making it a useful tool for combating malicious image forgeries.

Zheng et al[3].
A survey on picture tampering and its detection in real-world photos, changing images using software is now relatively easy, but if someone has concealed an object or altered someone's face, it seems suspicious. It is vital to determine which component of the image has been modified before questioning their motives. This necessitates the development of automatic technologies capable of distinguishing between genuine and manipulated photos. This review looks at typical picture manipulation methods, previously available manipulated image datasets, and new tampering detection approaches. It also provides a new viewpoint on reconsidering the assumptions of tampering clues underlying distinct detection systems, urging the research community to build generic tampering localization methods rather than depending on single-type tampering detection.

R. Shao et al [4]
The paper presents a system for detecting manipulated regions in scanned images using deep learning techniques. The system is trained on a dataset of over 3,800 scanned images from 169 different scanner models, using popular convolutional neural networks architectures like InceptionV3, Resnet34, and Xception Net. The system generates a reliability map that highlights any regions of the image that may have been manipulated. It uses advanced deep-learning techniques and a large dataset of scanned images to differentiate between features of different scanner models and identify any areas that may have been manipulated.

Several authors [7-11] contributed to the development of deep learning and machine learning models to anticipate forgeries using techniques such as Convolutional Neural Networks and Support Vector Machines. These algorithms have yielded encouraging results in detecting forgeries, highlighting the possibility of automated image forgery detection systems. However, further study is needed to investigate and compare the performance of various algorithms and their combinations to produce more accurate and dependable image forgery detection models.

III. METHODOLOGY

Image Forgery Detection is a technique used in digital forensics to determine if an imaged report or image is forged. This procedure involves the use of computer vision libraries such as OpenCV to analyze the attributes of the scanned document and extract unique features that can be used for identification. Furthermore, the MD5 hashing algorithm is used to generate a unique digital fingerprint of the scanned document that can be used for comparison.

The proposed method extracts attributes of an image using an open-source computer vision and machine learning software library, which are then passed through a hash function to generate a digital signature. The signature is compared to the original signature to detect any differences that may indicate image tampering. This was tested using a dataset of real-world photos that had been manipulated in various ways, including copy-move, splicing, and removal.

The findings from experiments indicate that the proposed method identifies image forgery quickly, outpacing existing methods in terms of image detection period and efficiency. This has potential applications in forensics, security, and media authentication, among many others, and reflects a promising approach to addressing the growing concern of image forgery.

Convert the image to grayscale, resize it to a standard size, and use any necessary filtering or enhancement techniques to remove noise and improve image quality.
OpenCV can be used to extract image features such as texture, color, and shape. This can be accomplished using techniques such as local binary patterns (LBP), scale-invariant feature transform (SIFT), or histograms of oriented gradients (HOG).

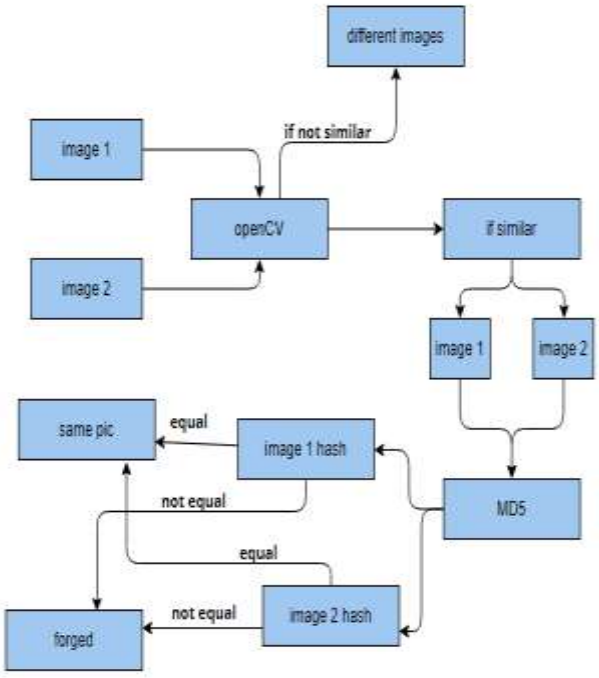


Figure 1. Architecture

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

In this study, we are calculating the grayscale of the images using the histogram technique.
Creating an MD5 hash value for the image. This will serve as a unique digital signature for the image and help us determine which one is better.
Taking the hash value of the original image and comparing it to the hash value of the potentially tampered image. If there is a significant difference between the two hash values, it indicates that the image has been tampered with.
Finally, report the analysis results, including the location and extent of any tampered regions, as well as the detection's confidence level. This information can be used for further investigation or legal purposes.

A. Dataset Explanation

For detection, we are using the CASIA public dataset, which contains both original and tampered images. Using the open cv library for image processing, we can compare any two images and determine whether they are forged or not. The CASIA dataset is a widely used benchmark in image forgery detection research. It includes many original images as well as tampered versions that have been altered using various image editing techniques.
The dataset's rows and columns refer to the individual photographs as well as the features or attributes that describe each image. Each row represents a single image, and each column represents a distinct aspect or attribute of that image. These characteristics include image resolution, color distribution, texture, and edge information.

For our project, we are selecting characteristics from the CASIA dataset, such as text and edge features, assuming that each image carries a unique hash value, and comparing them by calculating the grayscale from the text feature and histogram from the edge features. Because we are utilizing md5, a hashing function, we take an individual photo from the system itself rather than loading the dataset and comparing it for forgeries.
The methods employed in the CASIA dataset for image forgery detection using OpenCV and MD5 are meant to detect various sorts of image alteration, including copy-pasting, splicing, and retouching.
These techniques involve changing various aspects of an image, such as the color, texture, or edges, to make a plausible fake. It is feasible to detect these types of changes and identify picture forgeries by examining the visual elements of an image and comparing its hash to the hash of an authentic image.

B. Techniques Used

1. OpenCV (Open-Source Computer Vision Library):
It is an open-source computer vision and machine learning software library that is widely used in academic and industrial research and development for real-time image and video processing, object detection and tracking, face recognition, and many other applications. It was created by Intel in 1999 and is now maintained by the OpenCV.org community.

OpenCV is written in C++ and includes a C++ API as well as APIs for Python, Java, and other programming languages. It includes over 2,500 optimized algorithms for image and video capture, filtering, segmentation, feature detection and extraction, object detection and recognition, face detection and recognition, tracking, and 3D reconstruction [12].

OpenCV's main features include:
Cross-platform compatibility: OpenCV can run on Windows, Linux, Mac OS, iOS, and Android.
High performance: OpenCV is designed to take advantage of multi-core CPUs and GPUs to achieve real-time performance.
Simple APIs: OpenCV offers both high-level APIs for basic computer vision tasks and low-level APIs for more complex customization. OpenCV has a large and active community that contributes to its development and maintenance, as well as many third-party libraries and tools that integrate with OpenCV.
Many real-world applications rely on OpenCV, including self-driving cars, security and surveillance systems, medical imaging, robotics, augmented reality, and video game development.

2. **MD5:**
MD5 (Message Digest 5) is a widely used cryptographic hash function that generates a fixed-size, unique hash value from an arbitrary-length message. Ronald Rivest created it in 1991 as a replacement for the earlier MD4 algorithm.
The MD5 algorithm processes an input message through a series of mathematical operations to produce a fixed-size, 128-bit hash value. This hash value is unique to the input message, which means that any changes to the input message will result in a different hash value.
MD5 is widely used for integrity checking, authentication, and data validation in a variety of applications. For example, it is often used to hash passwords in online applications to ensure that the password is not stored in plaintext and cannot be easily recovered if the password database is stolen. It is also employed in digital signatures to ensure that a message has not been tampered with or altered during transmission [13].

3. Django Web Framework:
Django is a high-level Python web framework that adheres to the model-view-controller (MVC) architectural pattern. It has a sleek and practical architecture that allows you to construct web applications quickly and with less code. Django's key strengths are its stability, scalability, and adaptability. It includes several built-in capabilities such as an ORM (Object-Relational Mapping) for interfacing with databases, a sophisticated template engine, built-in authentication, and a complete admin interface.
Django is also highly extendable, with various third-party packages and plugins available to extend its capability. Many famous websites use it, including Instagram, Pinterest, and Mozilla.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Overall, Django is a robust and adaptable web framework that enables developers to create sophisticated web applications fast and efficiently [14].

4. Grayscale:

Grayscale refers to the process of transforming a color image to a black-and-white image, where each pixel value is represented by a single grayscale value between 0 and 255. This technique is frequently used to simplify a picture, reduce the quantity of data that must be processed, or focus on specific image elements that can be better viewed in grayscale. It aids in the simplification of algorithms and the elimination of complications associated with computational requirements.

It allows for easier learning for people who are new to image processing. This is because grayscale reduces an image to its basic minimum of pixels. It facilitates visualization. Because it is primarily in two spatial dimensions, it distinguishes between an image's shadow details and highlights 2d rather than others [15].

C. MATHEMATICAL CONCEPTS

Grayscale Conversion:

For a given pixel having red, green, and blue color values (R, G, B), the grayscale value is calculated (G). The cv2.cvtColor() method is used to convert an image to grayscale using the BGR to grayscale conversion formula:

$$Y = 0.299\ R + 0.587\ G + 0.114\ B.$$

Histogram Calculation:

The cv2.calcHist() method is used to compute the histogram of the grayscale image. The formula for histogram computation comprises counting the number of pixels in each intensity bin (from 0 to 255) and plotting a histogram with the counts on the y-axis and the intensity values on the x-axis.

$$H(i) = N(i) / (M * N)$$

Where H(i) is the frequency of occurrence of grayscale level I N(i) is the number of pixels of grayscale level I M is the number of rows in the image, and N is the number of columns in the image. To generate a probability density function, normalize the histogram by dividing each frequency by the total number of pixels (M * N).

Euclidean Distance Calculation:

The Euclidean distance between two histograms is obtained using the formula:

D = sqrt((h1[0]-h2[0])**2 + (h1[1]-h2[1]) (h1[1]-h2[1])**2 + … + (h1[n]-h2[n]) (h1[n]-h2[n])**2)
Where h1 and h2 are the histograms of the two photos and n is the number of intensity bins (256 in this case).

MD5 Hash Calculation:

The md5hash.scan() function is used to compute the MD5 hash of an image. The MD5 hash is a 128-bit value that is unique to the input data and is calculated using the MD5 message-digest method.

IV. RESULTS AND ANALYSIS

Image forgery detection with OpenCV and MD5 includes comparing two photos for similarity and identifying any evidence of tampering. If the Euclidean distance between the photos is less than a given threshold, the images are considered comparable. Furthermore, the md5hash library is used to generate unique hash values for the photos, which are then compared to detect any evidence of tampering.
The results of this project can be quite good at detecting manipulation and can give a solid platform for future inquiry. According to prior efforts, detecting forgeries takes time, however, our project takes less time and operates more efficiently. Overall, combining OpenCV and MD5 can be a useful addition to the arsenal of image forgery detection methods.

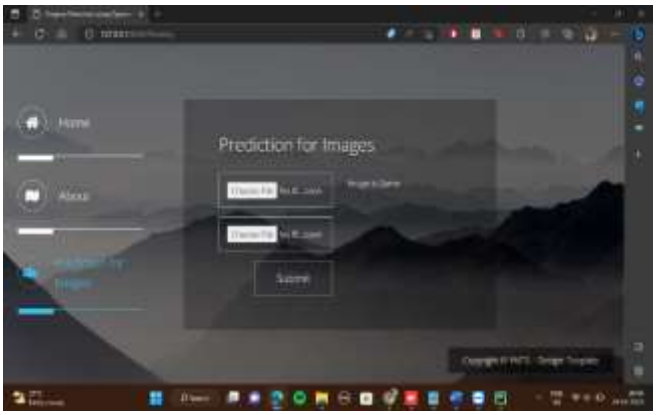
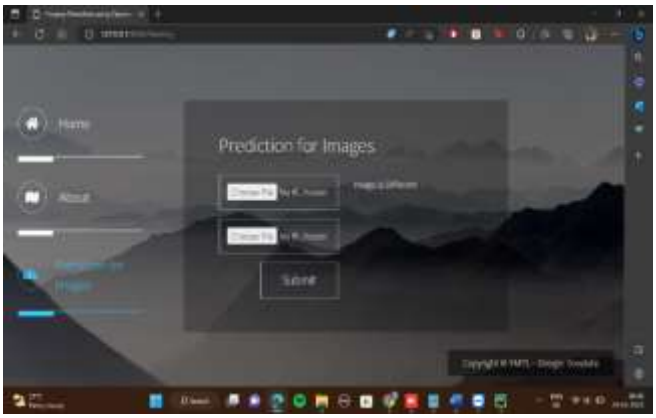


Figure 4. images are the same



Figures 5. Images are forged

The script first sets the DJANGO SETTINGS MODULE environment variable to the web app.settings module. This module contains Django project setup settings, such as database settings, installed apps, and middleware.

The idea behind this method is to compare the grayscale intensity and histograms of two photos to determine if they are similar or not.
The similar() function first converts the input photos to grayscale using cv2.cvtColor (). The histogram of the grayscale photos is then calculated with cv2.calcHist (). The histogram is a graph that depicts the distribution of pixel intensities in an image. We can tell whether two photos are similar or not by comparing their histograms. This is accomplished by computing the Euclidean distance between the histograms using the formula:

img src=<"https://latex.codecogs.com/svg.image?d=sqrtsum i=1n(x i-y i)2" title="d = sqrtsum i=1n(x i-y i)2" />
where x i and y i are the pixel values of the histograms of the two images at bin I and n is the number of bins.

To generate a hash of each image, the createHash() function uses the MD5 hash technique. If the hashes of the two photos are similar, the method returns True, indicating that the images are the same. Otherwise, it returns False.

To summarize, our method detects image counterfeiting using two different techniques (grayscale intensity and histogram comparison and MD5 hashing). While this method is not perfect, it is a simple and effective way to detect image alteration.

V. CONCLUSION AND FUTURE SCOPE

In conclusion, the combination of OpenCV and MD5 can be a powerful tool for detecting image forgery. In this project, we developed an application for detecting image forgery. The application utilized OpenCV and MD5 methods for detecting tampered images with high accuracy. MD5 method performed better compared to other methods in image detection.
However, it is important to note that there are limitations to this approach. While OpenCV can detect many types of image tampering, it may not be able to detect more sophisticated techniques, such as deepfake or AI-generated images. Additionally, while MD5 is a secure hash function, it is not immune to attacks, and newer hash functions may be necessary for more sensitive applications.

To overcome the limitations of OpenCV and MD5 in image forgery detection, the future scope involves using advanced techniques such as deep learning models, cryptography methods like SHA-3 and BLAKE3, and multi-modal analysis systems. Collaboration with other fields like computer science, mathematics, and physics can also help in developing more advanced and effective image forgery detection systems.

OpenCV examines an image's visual attributes to discover discrepancies, whereas MD5 provides a hash that may be used to compare two images and determine if they are identical. OpenCV is more effective at detecting certain visual manipulation techniques, whereas MD5 is more effective at certifying an image's validity. A mix of approaches and instruments is generally required to detect image counterfeiting efficiently.
Overall, our project demonstrated the effectiveness of the use of OpenCV and MD5 methods. This has significant implications in various fields such as digital forensics, law enforcement, and online security.

REFERENCES

[1] Xiao, B.; Wei, Y.; Bi, X.; Li, W.; Ma, J. Image splicing forgery detection combining coarse to a refined convolutional neural network and adaptive clustering. Inf. Sci. 2020, 511, 172–191.

[2] Kwon, M.J.; Yu, I.J.; Nam, S.H.; Lee, H.K. CAT-Net: Compression Artifact Tracing Network for Detection and Localization of Image Splicing. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 5–9 January 2021; pp. 375–384.

[3] Zheng, L.; Zhang, Y.; Thing, V.L. A survey on image tampering and its detection in real-world photos. J. Vis. Commun. Image Represent. 2019, 58, 380–399.Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 5–9 January 2021; pp. 375–384.

[4] R. Shao and E. J. Delp, "Forensic Scanner Identification Using Machine Learning," 2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), Albuquerque, NM, USA, 2020, pp. 1-4, doi: 10.1109/SSIAI49293.2020.9094618.

[5] Y. Q. Shi, C. Chen, and W. Chen, “A natural image model approach to splicing detection,” Proceedings of the 9th workshop on Multimedia & Security, pp. 51–62, September 2007, Dallas, TX.

[6] Sevinc Bayram, Husrev Taha Sencar, and Nasir Memon, “An efficient and robust method for detecting copy-move forgery,” Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp.1053–1056, April 2009, Taipei, Taiwan.

[7] L. Bondi, L. Baroffio, D. G¨uera, P. Bestagini, E. J. Delp, and S. Tubaro, “First steps toward camera model identification with convolutional neural networks,” IEEE Signal Processing Letters, vol. 24, no. 3, pp. 259–263, March 2017.

[8] B. Bayar and M. C. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, pp. 5–10, June 2016, Vigo, Galicia, Spain.

[9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, June 2016, Las Vegas, NV.

[10] Reshma P.D and Arun Vinod. C “IMAGE FORGERY DETECTION USING SVM CLASSIFIER” 2015 IEEE Royal College Of Engineering And Technology Akkikavu Kerala ,INDIA 978-1-4799-6818-3/15 © 2015.

[11] S.L.Jothilakshmi and V.G.Ranjith “Automatic Machine Learning Forgery Detection Based On SVM Classifier” 2014 (IJCSIT) International Journal of Computer Science and Information Technologies *NI university, Tamilnadu India* 2014, 3384-3388

[12] Open-Source Computer Vision Library of March 28, 2023.

[13] Message Digest 5-in cryptographic hash function of March 28, 2023.

[14] Django Web Framework is an open-source, Python-based web framework of March 28, 2023.

[15] Grayscale-technique/March 28, 2023.