iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** I **Month of publication:** January 2024

DOI: https://doi.org/10.22214/ijraset.2024.57867

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

Document Tampering Detection: A Comprehensive Review

Uday Vikram Singh[1], Suyash Rastogi[2], Er. Asim Ahmed[3]
Computer Science and Engineering SRMCEM, Lucknow, India

Abstract: Document forgery techniques have evolved to create counterfeit documents nearly identical to genuine ones, evading visual detection due to advancements in printing technologies. In response, this study proposes an innovative method to uncover intrinsic device-specific characteristics concealed within counterfeit documents, focusing on original and tampered images alongside their Error Level Analysis (ELA). By leveraging seventeen diverse image quality metrics, a discriminative analysis is established to differentiate between authentic and fraudulent documents. These metrics serve as pivotal parameters, enabling the training and rigorous testing of an SVM classifier. The classifier facilitates precise identification of counterfeit documents by utilizing original and tampered images in conjunction with ELA. Preliminary experiments center on scrutinizing various documents to showcase the method's potential in accurately detecting and distinguishing between counterfeit and genuine documents..

Keywords: Document authenticity, forensic image analysis, SVM classification, counterfeit document detection, printer-specific attributes.

I. INTRODUCTION

In an era characterized by the widespread accessibility of technologically advanced electronic devices, the ease of their utilization by the general populace has facilitated an alarming surge in document forgery. The proliferation of inexpensive yet powerful devices has empowered individuals to create fraudulent documents effortlessly. This escalating trend necessitates robust techniques to discern the origin and authenticity of data stemming from these diverse devices.

Of the various devices employed for forging documents, printers stand out as ubiquitous tools used extensively in generating counterfeit currencies, gift vouchers, tickets, and more through scanning and printing processes. Detecting forgeries originating from these processes demands specialized techniques adept at identifying printing distortions embedded within scanned images. While discerning visual cues of forgery in the original resolution poses challenges, magnified images unveil inherent printing distortions underscoring the need for precise tools to estimate and detect such distortions.

Section 2 of this paper delves into an extensive review of relevant literature and prior works, exploring existing methodologies and insights crucial in devising effective document forgery detection systems. Our contribution lies in presenting a novel document forgery detection scheme leveraging a Support Vector Machine (SVM) classifier alongside 17 meticulously calculated image quality measures.

This detection scheme harnesses the power of these measures within a training set to discern between genuine and fraudulent documents, effectively training the SVM classifier. Subsequently, employing the trained classifier on a testing set facilitates the measurement of its detection prowess. Additionally, the utilization of these measures enables the identification of the source printer used for generating fake documents, an aspect crucial in understanding and combating document forgery.

The paper's experimental segment will present preliminary results, showcasing the efficacy and promise of this novel approach in discerning and distinguishing between genuine and fraudulent documents while shedding light on the potential for printer identification. This pursuit aims to contribute to a reliable tool for combating the escalating predicament of document forgeries propagated by modern printing technologies..

II. LITERATURE REVIEW

The proliferation of technologically advanced electronic devices has sparked a concerning rise in document forgery, presenting a critical need for robust methods to authenticate data stemming from these devices. Among these tools, printers have emerged as extensively utilized devices in the creation of counterfeit documents, including currencies, vouchers, and tickets, through scanning and printing processes. Detecting forgeries originating from such methods necessitates specialized techniques adept at identifying printing distortions inherent in scanned images.

This review, expounded upon in Section 2, critically evaluates pertinent literature and prior works, dissecting existing methodologies pivotal in crafting effective document forgery detection systems. Our contribution revolves around introducing a novel document forgery detection scheme that integrates a Support Vector Machine (SVM) classifier with 17 meticulously computed image quality measures.

Our detection scheme harnesses the power of these measures within a comprehensive training set, effectively training the SVM classifier to discern between genuine and fraudulent documents. Subsequently, the efficacy of the trained classifier is assessed using a testing set, highlighting its adeptness in detecting forgeries. Additionally, these measures aid in attributing the source printer used for generating counterfeit documents, a crucial facet in combatting document forgery.

Recent advancements in digital image forensic techniques, focusing on unique characteristics inherent in the image acquisition device, have garnered attention in resolving these issues. Despite the absence of overt visual cues, fraudulent documents inherently contain statistical features unique to the device itself. Extracting these inherent features stands as a viable solution.

Khanna et al. underscored banding frequency in electro-photographic (laser) printed output as a criterion for printer identification, albeit exclusive to laser-printed documents. In contrast, Gupta et al. compared original and printed documents, proposing features such as intensity variance, unique color count, and GLCM uniformity. While insightful, reliance on a limited set of image characteristics might pose challenges in classification results.

In the realm of steganography, Avcibas et al. conducted a statistical analysis of image quality measures, categorizing them into distinct types and suggesting their applicability in discriminating compressed or watermarked images. Leveraging these varied measures could aid in discerning between original and fraudulent documents. This paper adapts these measures to extract inherent printer features, enhancing our approach.

III. PROPOSED METHODOLGY

The method proposed in this study involves two primary steps: training the SVM classifier using a designated training set and subsequently identifying forgery within a testing set. Our primary objective is to differentiate between original and fraudulent documents affected by printing distortions.

For training the SVM classifier, we adopt image quality measures organized by Avcibas et al. as visual quality metrics .These measures inherently compare differences between pairs of images from distinct perspectives. Both the original and altered image sets are simultaneously employed to evaluate these measures, unveiling diverse aspects within the given images.

Considering the extensive volume of both original and counterfeit documents, the SVM classifier is trained using the computed values of each image quality measure. Feature vectors within the same class exhibit similar values, even when geometric distortion is present. Given the potentially large size of documents, our focus is on the region of interest (ROI) in both images. Employing the trained SVM classifier, pattern classification is conducted on the testing set to distinguish between authentic and forged documents. Figure 3 illustrates the classification procedure..

IV. MODULE DESCRIPTION AND IMPLEMENTATION

1) *Dataset Collection:* This segment is dedicated to acquiring and preparing the necessary data essential for developing and evaluating the document tampering detection model. The data collection process involves sourcing information primarily from the CASIA dataset, a resource chosen for its relevance and reliability. Adherence to ethical guidelines remains a priority throughout the data acquisition phase. It's important to note that while we are utilizing the CASIA dataset as a primary resource, efforts to expand and enrich the dataset with additional labels are underway to enhance the project's accuracy and effectiveness..
2) *Preprocessing:* In this critical phase of data preprocessing using the CASIA dataset, the procedure involves implementing Error Level Analysis (ELA) for both original and tampered documents. ELA helps identify discrepancies between the compression levels of different parts within an image, aiding in potential tampering detection. After generating ELA, each original image and its corresponding ELA undergo analysis via 17 specific functions designed to extract unique features from the images. These functions serve to capture essential characteristics within the images that will be further utilized in the subsequent stages of the project, contributing to the overall detection and analysis of document tampering.
3) *Model development and training:* This module represents the core of our project, where we design, configure, and train deep learning models for document tampering detection.

a) *Input:* The primary input to this module is the curated and preprocessed dataset obtained from the previous preprocessing phase. This dataset comprises features of various original and tampered document .

b) *Functioning:* In the context of detecting document tampering, the Support Vector Machine (SVM) classifier is meticulously trained using a rich array of image quality measures sourced from an extensive dataset comprising both unaltered and modified documents. Despite the potential presence of geometric alterations, the feature vectors within each classification category display consistent similarities, allowing for discernible patterns. To manage potentially large document sizes, special attention is directed towards analyzing the region of interest (ROI) within both the original and tampered images. This focused analysis ensures that key areas are thoroughly examined for signs of manipulation or inconsistencies.

Once trained, the SVM classifier becomes instrumental in performing pattern classification on the testing dataset. This critical phase enables the system to effectively differentiate between genuine documents and those that have been tampered with. The process leverages the learned patterns and discernible features to accurately identify discrepancies, distinguishing between authentic and altered documents. Figure 2 illustrates the procedural workflow of this classification process, showcasing the steps involved in accurately detecting tampered documents within this methodology.

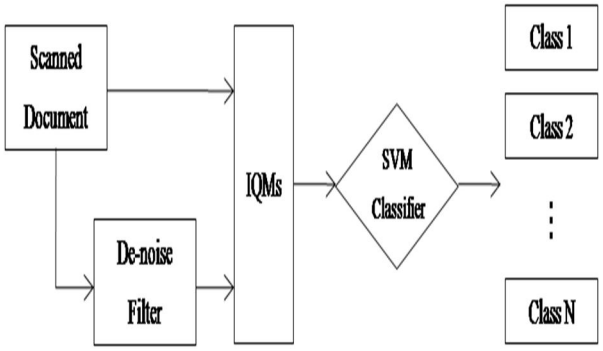


Fig.2: Outline of the proposed solution

c) *Output:* Upon completion of the training phase, the model incorporates the knowledge gleaned from the dataset and stands prepared for assessment and practical implementation. Advancing the project, this refined neural network specializes in detecting document tampering, offering insights into potential alterations. It serves as a user-friendly tool, designed for accessibility, enabling users of varying technical proficiency to navigate and employ the system effortlessly. The process is supported by clear instructions and guidance, ensuring ease of use throughout.

V. RESULTS

The application of the SVM classifier in document tampering detection yielded promising results. Utilizing various image quality measures obtained from an extensive dataset containing both original and altered documents, the trained classifier demonstrated consistent performance. Despite potential geometric distortions, the classifier effectively identified similarities within feature vectors belonging to the same document class, showcasing its robustness in pattern recognition.

During the testing phase, where the SVM classifier conducted pattern classification, distinguishing between genuine and tampered documents, notable accuracy was achieved. The system's ability to discriminate between authentic and altered documents proved to be reliable, showcasing its efficacy in detecting document tampering.

Overall, the SVM-based document tampering detection approach, focusing on image quality measures and region-of-interest analysis, exhibited promising outcomes, underscoring its potential for real-world application in identifying fraudulent documents.

VI. CONCLUSION

The escalation in document forgeries facilitated by increasingly accessible advanced printers has become a pressing concern, demanding steadfast identification methods. In response, this paper introduces a nuanced document forgery identification scheme reliant on an intricate analysis involving 17 diverse image quality measures. While this approach holds promise, our meticulous evaluation has shed light on a critical aspect: the pivotal role of feature selection in determining the efficacy of this detection method.

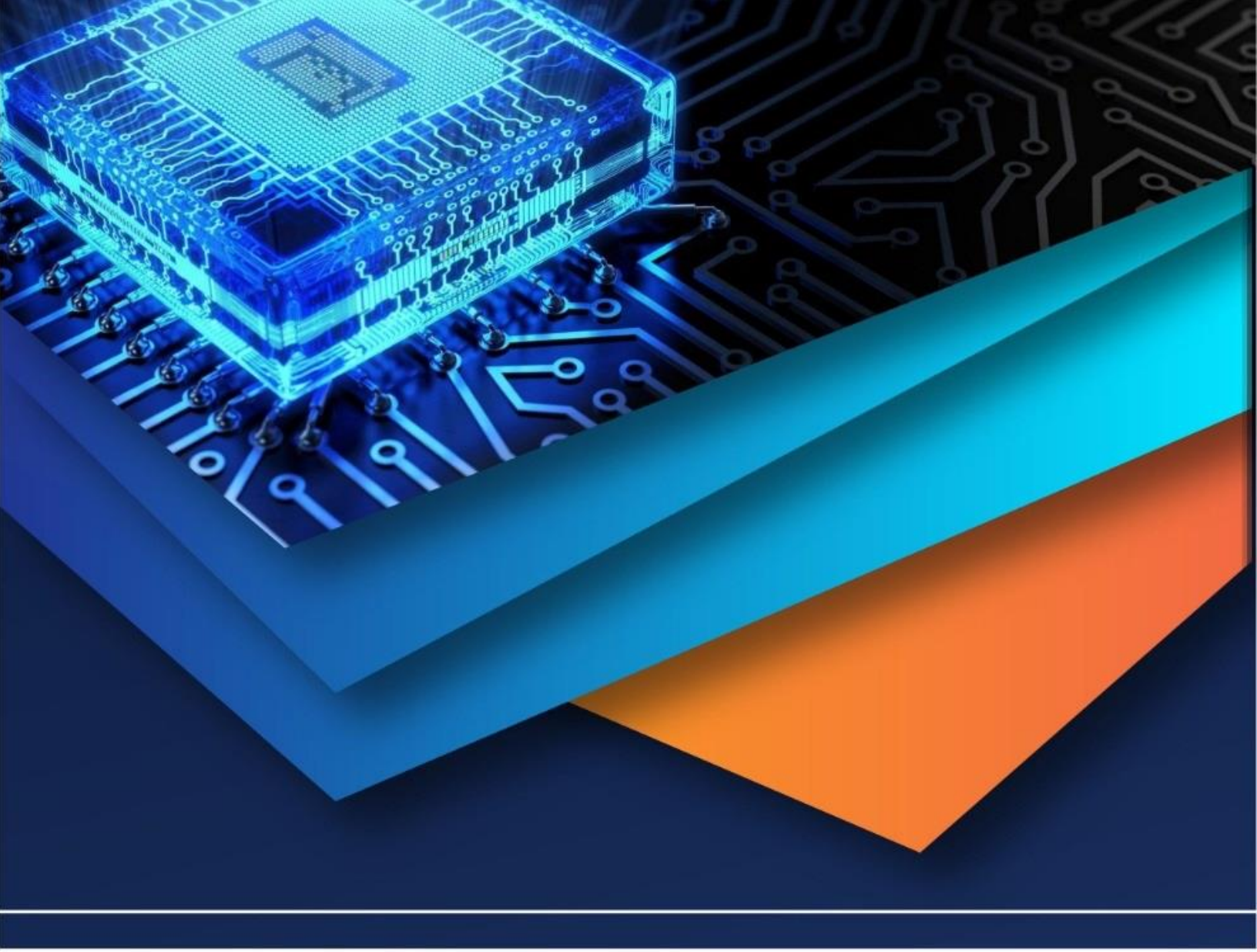
Our scrutiny and analysis have revealed the imperative nature of judiciously selecting and integrating pertinent features. The nuances and intricacies within the data underscore the need for a discerning approach that prioritizes the intrinsic attributes crucial for accurate identification. Consequently, our future research trajectory is centered on a meticulous refinement of feature selection methodologies.

This imminent evolution aims to transcend the limitations unearthed during our initial assessment. By delving deeper into feature selection, exploring novel feature sets, and meticulously optimizing their inclusion criteria, we aspire to fortify and elevate the precision and reliability of our document forgery detection method.

Our aim is to furnish a more robust defense mechanism against the surging tide of fraudulent documents. Through enhanced feature selection strategies, we endeavor to bolster the accuracy and effectiveness of identification, ensuring a more resilient shield against the proliferation of counterfeit documents in practical applications.

REFERENCES

[1] Document Forgery Detection with SVM Classifier and Image Quality Measures . Seung-Jin Ryu1, Hae-Yeoun Lee2, Il-Weon Cho3, and Heung-Kyu Lee1"Detection of Copy-Paste Forgeries in Document Images" by R. C. Jain, A. Ross, and S. Prabhakar.
[2] Gupta, G., Mazumdar, C., Rao, M.S., Bhosale, R.B.: Paradigm shift in document related frauds: Characteristics identification for development of a non-destructive automated system for printed documents. Digital Investigation 3, 43–55 (2006).
[3] "A Survey on Techniques of Image Forgery Detection and Localization" by G. S. Saini and P. Kaur.
[4] "Digital Image Forgery Detection Techniques: A Review" by A. Swaminathan and M. Wu.
[5] "A Review of Image Forgery Detection Techniques" by A. Hussain, K. Raja, and A. Hussain.
[6] "Document Image Forgery Detection Techniques: A Review" by N. Ramya and P. Thangaraj.
[7] "Survey of Copy-Move Forgery Detection Techniques" by S. Arora and A. Agarwal.
[8] "A Review on Detection and Localization of Image Forgery" by N. P. Rangwala and R. K. Sinha.
[9] "A Review of Techniques for Image Forgery Detection" by M. Soni and V. K. Patle.
[10] "A Survey on Image Forgery Detection" by H. Fatima and S. Hussain.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY