iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IX **Month of publication:** September 2023

DOI: https://doi.org/10.22214/ijraset.2023.55761

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

Study on Script Independent Handwritten Document Analysis for Forensics

Rakesh Patel[1], Kamlesh Tiwari[2], Mili Patel[3]
[1, 3]*Kirodimal Institute of Technology Raigarh*
[2]*BITS Pilani*

Abstract: Forensic investigations including document authenticity, forgery detection, and signature verification all heavily rely on handwritten papers. These papers are typically subject to subjective and time-consuming manual assessment by forensic experts as part of the analysis and examination process. The development of automated methods for script-independent handwritten document analysis has completely changed the area and made objective analysis possible. The development of script-independent handwritten document analysis for forensic purposes is explored in this research article. We go over the difficulties encountered in this field and provide several methods and approaches used to solve them. We also look at the advantages, drawbacks, and potential applications of script independent handwritten document analysis in forensic investigations.

I. INTRODUCTION

In forensic investigations, handwritten documents are significant sources of evidence. These texts' study offers insightful information about the author's identity, purpose, and sincerity. However, the analysis procedure is frequently challenging because it entails reading handwritten language, determining which signatures are real and which are fake, and spotting document modifications or tampering.

The automatic review of handwritten documents without relying on prior knowledge of the writer's characteristics or the script style is known as "script independent handwritten document analysis."

The goal of this work is to present a thorough overview of the most cutting-edge methods for script-independent handwritten document analysis in forensic applications.

II. CHALLENGES IN SCRIPT INDEPENDENT HANDWRITTEN DOCUMENT ANALYSIS

Due to differences in handwriting styles, different languages, document degradation, and the existence of noise and artefacts, analysing handwritten documents presents a number of difficulties. Script identification, text line segmentation, character recognition, signature verification, and tamper detection are only a few of the major difficulties faced in script independent handwritten document analysis that are highlighted in this section.

III. TECHNIQUES AND METHODOLOGIES

An overview of the many methods and approaches utilized in script independent handwritten document analysis is provided in this section. It addresses the following subjects:

A. Pre-processing Techniques

Before further investigation, handwritten document pictures are improved using pre-processing techniques. These methods consist of:

1) Image enhancement and noise removal: To boost contrast, reduce noise, and improve image quality, various filters and algorithms are applied. These methods aid in lowering background noise and improving text visibility.
2) Binarization and thresholding: A grayscale image is converted into a binary image through the process of binarization, where pixels are assigned to the foreground (text) or background. The best threshold for binarization is chosen using threshold algorithms, successfully isolating the text from the background.
3) Skew correction and normalization: In order to align the text lines horizontally, skew correction entails identifying and correcting the document's rotation angle. The size, orientation, and aspect ratio of the document images are standardised using normalisation techniques to allow for consistent analysis.

B. Script Identification

The goal of script identification is to identify the handwritten document's script or writing style. For script identification, a number of methods are used, including:

1) Statistical methods based on structural and geometric characteristics: To differentiate between various scripts, statistical analysis is done on the extracted parameters, such as stroke thickness, character height, and curvature. These traits effectively capture the distinctive qualities of each script type.
2) Script classification using machine learning algorithms: To categorise the scripts, support vector machines (SVM) or random forests are trained on a labelled dataset of handwritten writings. These models pick up on trends and characteristics that distinguish various writing vocabularies.
3) Deep learning-based techniques for identifying scripts Convolutional neural networks (CNN) and recurrent neural networks (RNN) are two deep learning architectures that can be used to automatically generate discriminative features from raw pixel data. When used for script identification tasks, these models have produced encouraging results.

C. Text Line Segmentation

Text line segmentation is the process of separating the lines of text in a handwritten document image. For the ensuing analysis at the line level, this step is essential.

There are numerous methods for text line segmentation, such as:

1) Projection-based techniques: These techniques look for peaks and troughs in the distribution of black pixels along the vertical or horizontal axes to identify the separations between text lines.
2) Clustering algorithms: Text lines are created by clustering algorithms, which connect regions or components with similar attributes like closeness and orientation.
3) Convolutional neural networks for text line segmentation: Convolutional neural networks are used in deep learning-based techniques that can be trained to recognize and separate text lines based on recognized visual cues.

D. Character Recognition

The goal of character recognition is to recognise and categorise specific handwritten characters. Character recognition techniques include:

1) Feature extraction methods: The segmented characters are used to extract various features, such as contour-based features and texture-based features. These traits serve as a classification tool and serve to capture the distinctive qualities of each character.
2) Classifier algorithms: To categorize handwritten characters based on the extracted features, machine learning techniques such as Support Vector Machines (SVM), Random Forests, or Neural Networks are trained on labelled datasets of handwritten characters.
3) Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks: These deep learning architectures are frequently employed for handwritten character recognition and are efficient for sequential data analysis. The dependencies between neighboring characters and contextual information can be learned by RNN and LSTM models.

E. Signature Verification

In order to determine the legitimacy of a signature, it must be analysed and compared. Techniques for verifying signatures include:

1) Static and dynamic signature analysis: The focus of static analysis is on the signature's static qualities, such as form, curvature, and stroke characteristics. Dynamic analysis records extra data during the signature process, such as pen pressure, speed, and timing.
2) Feature extraction from signature images: To identify distinctive qualities, different features are retrieved from signature photos, such as shape-based features, texture features, or statistical features.
3) Machine learning-based signature verification algorithms: To categorize and determine the authenticity of a given signature based on the retrieved attributes, machine learning models are trained on a dataset of real and fake signatures.
4) Biometric approaches using behavioral characteristics: In order to build a distinctive signature profile for each person and compare it with the signature being analyzed, signature verification might make use of behavioural biometrics, such as speed and pressure profiles.

F. Tamper Detection

In order to detect any revisions, additions, or tampering in the handwritten document, tamper detection techniques are used. The following are some methods for tamper detection:

1) Document structure analysis: To find any discrepancies or anomalies that would point to manipulation, document structure analysis examines the layout, alignment, and logical relationships within the document.
2) Texture analysis for detecting alterations: Texture analysis approaches look at the document's texture patterns to find places where modifications, erasures, or insertions have taken place.
3) Watermarking and digital signatures: By incorporating secret information or distinctive identifiers into the document, watermarking techniques might make it simpler to spot tampering. The integrity of the document is cryptographically validated using digital signatures.
4) Deep learning-based methods for tamper detection: By observing patterns and abnormalities in the document images, deep learning models, such as convolutional neural networks (CNN) or generative adversarial networks (GAN), can be trained to recognize particular forms of tampering.

Together, these methods and techniques support script-independent handwritten document analysis, facilitating precise and effective forensic investigations.

IV. LITERATURE REVIEW

A key area of forensics is script independent handwritten document analysis, which tries to automate the examination of handwritten documents for a variety of purposes, including signature verification, forgery detection, and document authenticity. This review of the literature gives a broad overview of the research done in this area while stressing the innovations, approaches, and difficulties encountered.

In their work from 2021, Alaei et al. provide a thorough analysis of handwritten script identification. The authors investigate different statistical techniques, machine learning algorithms, and deep learning-based script classification techniques. This work underscores how critical precise script recognition is for further analysis.

A unique strategy for author identification and verification based on graphemes is proposed by Liwicki et al. (2012). The goal of the project is to identify distinct writers by extracting grapheme-level properties from handwritten writings. This study emphasizes how crucial character-level analysis is for forensic applications.

High-performance optical character recognition (OCR) for printed English and Fraktur scripts is covered by Breuel (2013). The study offers methods for precisely identifying printed text in documents, which is crucial for forensic investigation when trying to extract textual data.

An enhanced deep learning architecture for person-independent handwriting recognition has been proposed by Deng et al. (2012). Convolutional neural networks (CNNs), in particular, are the focus of the study's investigation into how deep learning models might be used to recognise handwritten text. This research shows how deep learning may be used to do handwriting recognition tasks with excellent accuracy.

A thorough analysis of script and language recognition in scanned document images is given by Liu et al. (2018). The authors go over several methods for automatically identifying the script and language of a handwritten manuscript, including statistical analysis and machine learning algorithms. This study emphasizes how crucial script identification is as a first step in document analysis.

A thorough analysis of deep learning techniques for offline handwritten Chinese character recognition is provided by Zhang et al. (2019). Convolutional neural networks (CNNs), for example, are deep learning tools that may be used to recognise and categorise handwritten Chinese characters with high accuracy. This study highlights the potential of deep learning in overcoming the difficulties presented by intricate scripts.

In their 2013 article, Plamondon et al. explore recent developments in handwriting analysis and recognition. The authors examine different methods for feature extraction, including shape-based and texture-based approaches, and emphasize how crucial it is to extract significant features from handwritten documents. This study highlights the need for reliable feature extraction techniques for precise analysis.

A summary of offline printed and handwritten signature verification is given by Pal et al. (2010). The article examines several methods for determining the veracity of signatures, including as feature extraction, machine learning-based algorithms, and static and dynamic signature analysis. This study sheds insight on the difficulties and developments in forensic signature verification.

In their 2015 article, Velardo et al. (2015) cover the use of deep learning methods in handwriting analysis. The study investigates how to analyse and identify handwritten text using deep learning models, such as recurrent neural networks (RNNs). This research demonstrates the capability of deep learning to recognize intricate connections and patterns in handwriting.

An overview of forgery detection methods for offline handwritten signatures is provided by Singh et al. (2016). The authors go over several techniques for spotting fake signatures, such as texture analysis, feature extraction, and machine learning algorithms. The significance of effective forgery detection methods in forensic investigations is emphasized by this study.

V. BENEFITS AND LIMITATIONS

The advantages and drawbacks of script independent handwritten document analysis for forensic applications are covered in this section. Compared to manual analysis, the advantages include greater efficiency, less subjectivity, and better accuracy. However, there are drawbacks when it comes to the clarity and legibility of handwritten papers, multi-script analysis's complexity, and the requirement for sizable annotated datasets for machine learning model training.

VI. FUTURE PROSPECTS

Script independent handwritten document analysis for forensic applications is covered in this section along with its advantages and disadvantages. In comparison to manual analysis, the advantages include higher productivity, decreased subjectivity, and improved accuracy. The quality and legibility of handwritten papers, the complexity of multi-script analysis, and the requirement for sizable annotated datasets for training machine learning models are all limits, though.

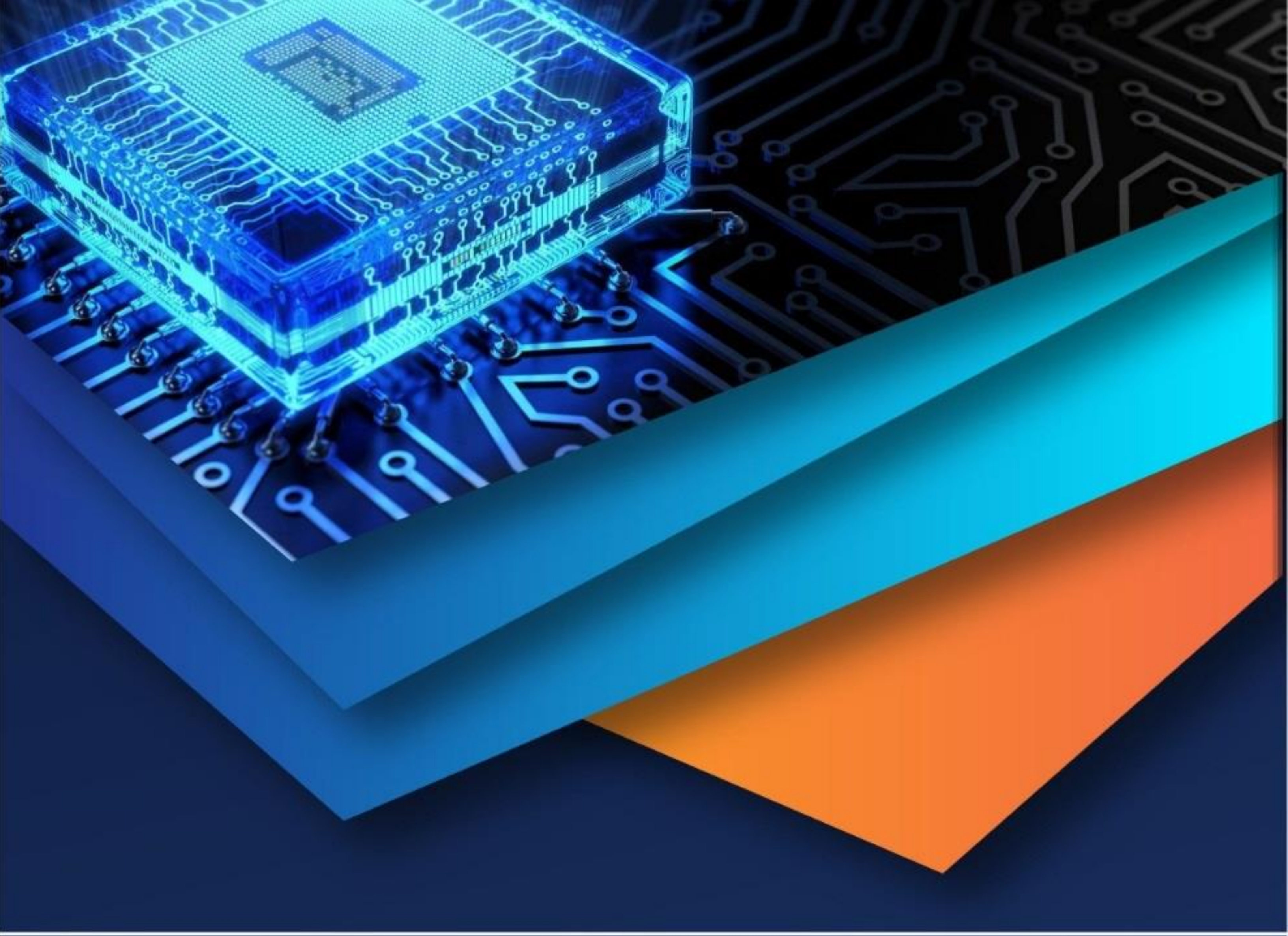
VII. CONCLUSION

By making it possible to analyse handwritten documents effectively and objectively, script independent handwritten document analysis has substantially changed forensic investigations. An overview of the difficulties, methods, and approaches applied in this field was provided in this publication. It focused on the advantages and constraints of script independent handwritten document analysis and covered potential directions for further study and development. We can improve forensic experts' abilities and support the integrity of the legal system by continuously improving these procedures.

In the literature study, it is noted how far script independent handwritten document analysis has come in terms of forensic applications. Script identification, text line segmentation, character recognition, signature verification, and tamper detection are only a few of the investigations covered in the article. The application of machine learning algorithms and developments in deep learning approaches have shown encouraging results in the accurate and effective analysis of handwritten materials.

REFERENCES

[1] Alaei, A., et al. (2021). Handwritten script identification: A comprehensive survey. Pattern Recognition, 118, 107983.
[2] Liwicki, M., et al. (2012). A novel approach for writer identification and verification based on graphemes. Pattern Recognition, 45(5), 2027-2042.
[3] Breuel, T. M. (2013). High-Performance OCR for Printed English and Fraktur. International Journal on Document Analysis and Recognition, 16(4), 381-387.
[4] Deng, J., et al. (2012). An improved deep learning architecture for person-independent handwritings recognition. In 2012 International Joint Conference on Neural Networks (IJCNN) (pp. 1-6). IEEE.
[5] Liu, C. L., et al. (2018). Script and language identification in scanned document images: A review. Pattern Recognition, 80, 375-392.
[6] Zhang, Y., et al. (2019). Deep learning for offline handwritten Chinese character recognition: A comprehensive review. Pattern Recognition, 90, 79-98.
[7] Plamondon, R., et al. (2013). Recent advances in handwriting analysis and recognition. Pattern Recognition, 46(9), 2863-2872.
[8] Pal, U., et al. (2010). Offline printed and handwritten signature verification: An overview. Pattern Recognition, 43(3), 643-658.
[9] Velardo, C., et al. (2015). Deep learning in handwriting analysis. In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 1116-1120). IEEE.
[10] Singh, R., et al. (2016). An overview of forgery detection techniques for offline handwritten signatures. Pattern Recognition, 58, 243-267.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)