

Concept Paper

Not peer-reviewed version

Hyperspectral Ink Classification and Forgery Detection Using Dimensionality Reduction and Clustering

[Shumaila Idrees](#)^{*} and Ahsan Inam

Posted Date: 3 July 2025

doi: 10.20944/preprints202507.0327.v1

Keywords: PCA; K-means; HSI



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Concept Paper

Hyperspectral Ink Classification and Forgery Detection Using Dimensionality Reduction and Clustering

Shumaila Idrees * and Ahsan Inam

Department of Computer Science, Institute of Space Technology, Islamabad, Pakistan; ahsan.inam@ist.edu.pk

* Correspondence: shumaila.idrees@ist.edu.pk

Abstract

This research presents a pattern recognition framework for detecting and classifying inks in handwritten documents using hyperspectral imaging (HSI). We utilize the iVision HHID dataset, applying Principal Component Analysis (PCA) for dimensionality reduction and k-means clustering for ink segmentation. Spectral signatures are extracted from individual text lines to detect ink mismatches, aiding forgery detection and document authentication. Experiments demonstrate the ability of PCA-reduced data to separate inks based on spectral variance. This paper evaluates the strengths and limitations of unsupervised clustering methods in document forensics and suggests future directions involving deep learning and hybrid spectral-spatial models.

Keywords: PCA; K-means; HSI

1. Introduction

Document forensics plays a crucial role in legal and security domains. Detecting ink mismatches using Hyperspectral Imaging (HSI) has emerged as a reliable approach to identifying forgeries and tampered text. HSI captures reflectance across hundreds of narrow spectral bands, enabling the detection of subtle ink variations invisible to the naked eye. While earlier methods relied on manual inspection or RGB differences, HSI-based systems enable non-invasive and precise analysis [1,2].

Recent literature suggests a surge in research leveraging HSI for tasks like forgery detection, ink classification, and writer identification [3,4]. Challenges include high data dimensionality, spectral redundancy, and inter-ink spectral similarity. This work aims to design an unsupervised pattern recognition approach using dimensionality reduction and clustering for ink classification using the iVision HHID dataset.

2. Literature Review

Hyperspectral imaging (HSI) has gained substantial attention in document forensics due to its ability to detect subtle spectral differences in inks that are visually indistinguishable. Bajwa et al. [1] provided a comprehensive review outlining the major stages of HSI pipelines, including preprocessing, dimensionality reduction (DR), and classification, emphasizing the critical role of each in effective hyperspectral analysis.

Bulanon et al. [2] were among the early adopters who demonstrated the feasibility of detecting ink mismatches using hyperspectral reflectance data, highlighting the potential of HSI for non-invasive forensic analysis. Extending this work, Khan et al. [3] proposed automated band selection methods coupled with classification algorithms to optimize ink mismatch detection performance.

In the realm of unsupervised learning, Siddiqui et al. [4] applied fuzzy clustering for forgery detection, which eliminated the dependence on labeled data—a crucial advantage when ground truth annotations are scarce. Moving towards deep learning approaches, Gupta et al. [5] and Raza et al. [6] introduced convolutional neural networks (CNNs) that capture complex spatio-spectral relationships

for ink authentication tasks, demonstrating superior performance compared to traditional clustering methods.

Writer identification through HSI was explored by Islam et al. [7], leveraging the distinctive spectral profiles of individual handwriting styles. The introduction of hybrid CNN architectures by Islam et al. [8] further improved the robustness of document authentication models by integrating both spatial and spectral features. To support standardized benchmarking in this domain, Islam et al. [9] released the iVision HHID dataset, which has become a valuable resource for evaluating hyperspectral document analysis algorithms.

Finally, Qureshi et al. [10] reviewed the current challenges faced by hyperspectral document image processing, such as high dimensionality, spectral redundancy, and variability introduced by environmental factors, while also identifying promising future directions including hybrid models and real-world deployment scenarios.

These studies collectively establish a solid foundation for hyperspectral ink classification and forgery detection, motivating the development of unsupervised pattern recognition frameworks such as the one proposed in this research.

3. Methodology

We use a hyperspectral cube from the iVision HHID dataset containing handwritten lines using multiple pens. The analysis follows these stages:

- Extract 1st, 30th, 60th, and last band from the HSI for grayscale visualization.
- Segment foreground text pixels using thresholding and morphological operations.
- Plot spectral reflectance for each text line across bands.
- Apply Principal Component Analysis (PCA) for dimensionality reduction.
- Use k-means clustering to classify ink types.
- Reapply clustering on PCA output and compare ink detection accuracy.

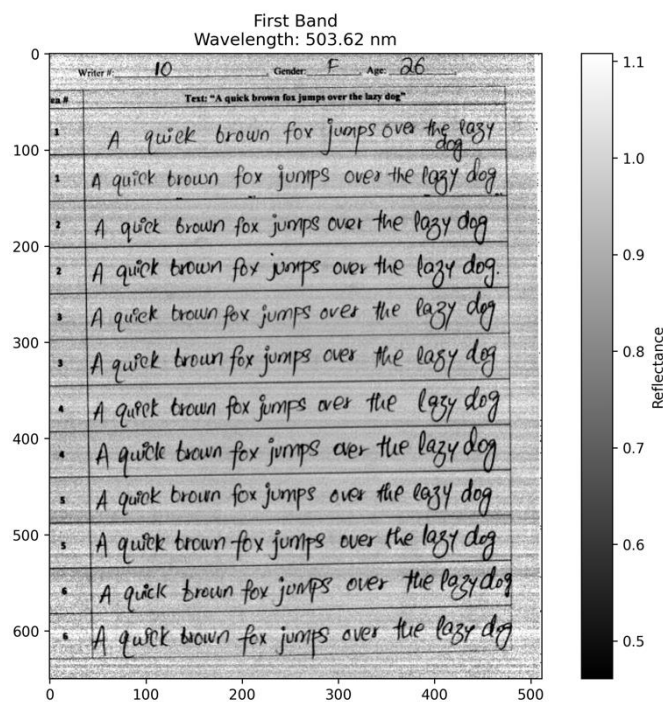


Figure 1. First Band.

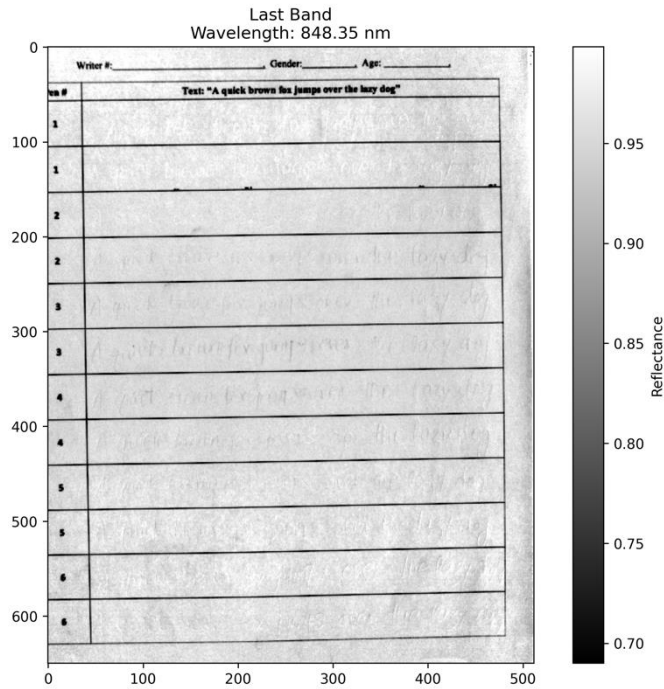


Figure 2. Last Band.

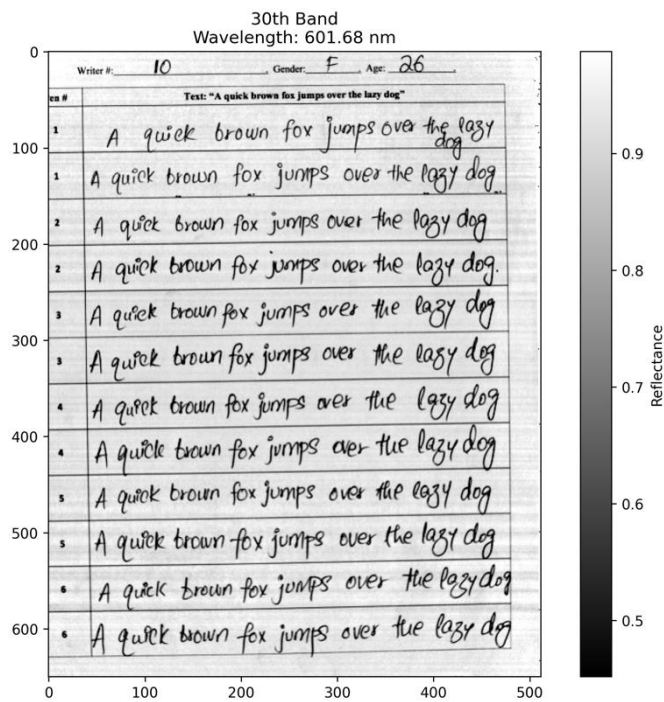


Figure 3. 30th Band.

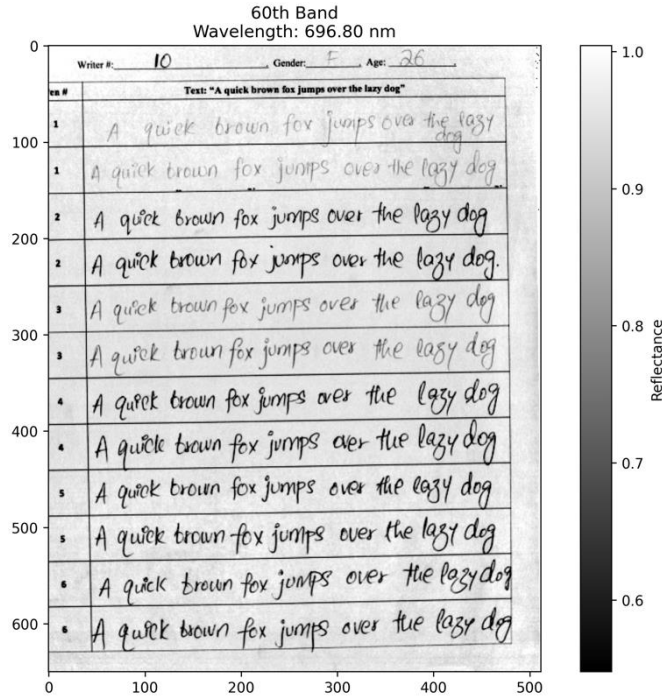


Figure 4. *th Band.*

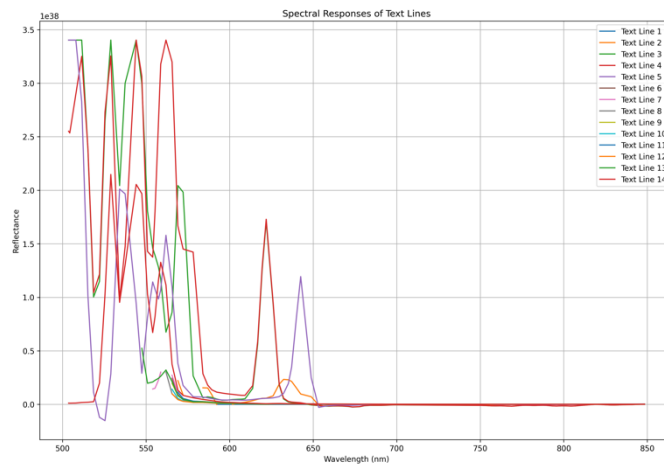


Figure 5. Spectral Responses of Test Lines.

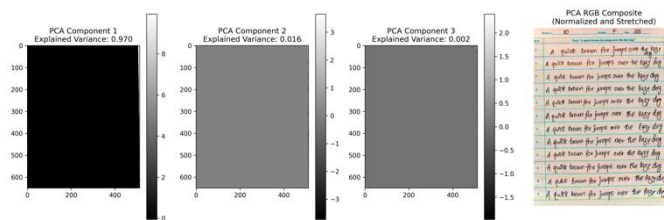


Figure 6. PCA for dimensionality reduction.

4. Results and Discussion

The Grayscale band images revealed intensity variation with wavelength, supporting spectral diversity of inks. Spectral response plots illustrated distinct reflectance curves for each text line, confirming unique ink signatures. PCA retained over 98% of variance in the first 3 components.

Clustering on PCA-reduced data segmented the document into three ink types accurately, matching visual inspection.

Strengths include label-free ink detection and efficient DR using PCA. However, k-means assumes spherical clusters, which may not model all spectra well. Inks with near-identical spectra may be clustered together. Deep models could improve separation by learning non-linear spectral embeddings.

5. Future Work

Future directions include replacing k-means with fuzzy c-means or DBSCAN for better cluster handling. Supervised classifiers like SVMs or CNNs could be explored using annotated datasets. CAEs or hybrid CNNs can learn spectral-spatial patterns, potentially improving classification. Testing under varying illumination or noise conditions could enhance robustness.

6. Conclusions

This paper presented an unsupervised framework for detecting and classifying inks in hyperspectral documents. PCA effectively reduced spectral dimensionality, and k-means clustering successfully separated different ink types. The approach is promising for non-destructive document forensics, with potential for improvement through supervised and deep learning methods.

References

1. M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern Trends in Hyperspectral Image Analysis: A Review," *IEEE Access*, vol. 6, pp. 14118-14119, 2018.
2. Z. Khan, F. Shafait, and A. Mian, "Hyperspectral Imaging for Ink Mismatch Detection," in *Proc. 12th Int. Conf. Document Analysis and Recognition (ICDAR)*, Washington DC, USA, 2013, pp. 877-881.
3. A. Abbas, K. Khurshid, and F. Shafait, "Towards Automated Ink Mismatch Detection in Hyperspectral Document Images," in *Proc. 14th IAPR Int. Conf. Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 2017, pp. 1229-1236.
4. M. J. Khan, A. Yousaf, K. Khurshid, A. Abbas, and F. Shafait, "Automated Forgery Detection in Multispectral Document Images Using Fuzzy Clustering," in *Proc. 13th IAPR Int. Workshop Document Analysis Systems (DAS)*, Vienna, Austria, 2018, pp. 393-398.
5. M. J. Khan, A. Yousaf, A. Abbas, and K. Khurshid, "Deep learning for automated forgery detection in hyperspectral document images," *Journal of Electronic Imaging*, vol. 27, no. 5, p. 053001, 2018.
6. M. J. Khan, K. Khurshid, and F. Shafait, "A Spatio-Spectral Hybrid Convolutional Architecture for Hyperspectral Document Authentication," in *Proc. Int. Conf. Document Analysis and Recognition (ICDAR)*, Sydney, Australia, 2019, pp. 1097-1102.
7. A. ul Islam, M. J. Khan, K. Khurshid, and F. Shafait, "Hyperspectral Image Analysis for Writer Identification using Deep Learning," in *Proc. Digital Image Computing: Techniques and Applications (DICTA)*, Australia, Dec. 2019.
8. A. ul Islam, M. J. Khan, K. Khurshid, and F. Shafait, "Hybrid CNNs for HSI-based Document Authentication," in *Proc. Int. Conf. Document Analysis and Recognition (ICDAR)*, 2019.
9. A. ul Islam, M. J. Khan, M. Asad, H. A. Khan, and K. Khurshid, "iVision HHID: Handwritten Hyperspectral Images Dataset for Benchmarking Hyperspectral Imaging-based Document Forensic Analysis," *Data in Brief*, vol. 42, p. 108306, 2022.
10. R. Qureshi, M. Uzair, K. Khurshid, and H. Yan, "Hyperspectral Document Image Processing: Applications, Challenges and Future Prospects," *Pattern Recognition*, vol. 90, pp. 12-22, 2019.
11. X. Huang, and L. Zhang, "Principal Component Analysis for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2656-2666, 2007.
12. H. Wang, Y. Liu, and F. Shafait, "Fuzzy C-Means Clustering for Document Forensics," *Elsevier Journal of Forensic Science International*, vol. 289, pp. 23-29, 2018.

13. Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Convolutional Autoencoders in Hyperspectral Imaging," *Sensors*, vol. 20, no. 1, p. 113, 2020.
14. J. Li, W. Liu, and H. Lu, "Deep Spectral Learning for Ink Classification," in *Proc. SPIE Defense + Commercial Sensing*, 2021, p. 11734.
15. K. Zhang, M. Sun, and J. Liu, "Evaluation of Spectral Variability in Ink Analysis," *Journal of Forensic Sciences*, vol. 67, no. 1, pp. 145-153, 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.