Localized Forgery Detection in Hyperspectral Document Images

Zhipei Luo[1], Faisal Shafait[2,3], Ajmal Mian[3]

[1]University of Science and Technology of China, Hefei, China
[2]National University of Sciences and Technology (NUST), Islamabad, Pakistan
[3]The University of Western Australia, Perth, WA, Australia
luozp@mail.ustc.edu.cn, faisal.shafait@seecs.nust.edu.pk, ajmal.mian@uwa.edu.au

Abstract—**Hyperspectral imaging is emerging as a promising technology to discover patterns that are otherwise hard to identify with regular cameras. Recent research has shown the potential of hyperspectral image analysis to automatically distinguish visually similar inks. However, a major limitation of prior work is that automatic distinction only works when the number of inks to be distinguished is known *a priori* and their relative proportions in the inspected image are roughly equal. This research work aims at addressing these two problems. We show how anomaly detection combined with unsupervised clustering can be used to handle cases where the proportions of pixels belonging to the two inks are highly unbalanced. We have performed experiments on the publicly available UWA Hyperspectral Documents dataset. Our results show that INFLO anomaly detection algorithm is able to best distinguish inks for highly unbalanced ink proportions.**

I. INTRODUCTION

There has been a significant growth in the need of fraud identification, especially to assess whether a certain document is a forgery. Forensic experts investigate the types of ink in a questioned document, if a handwritten note consists of more than one type of ink, it indicates that some manipulation towards the document has taken place. To do so, investigators need to distinguish the colors of different inks. Human eyes are sensitive to visible colors and are able to discriminate between different material spectral responses [1]. However, when different types of the same color inks are similar in appearance, naked eyes are not able to recognize the difference, although these inks share distinctive spectral information.

Current ink distinguishing methods can be divided into two approaches, destructive and non-destructive examinations. Thin Layer Chromatography (TLC) [2] is a destructive chemical analysis method, it has been applied to separate constituents of inks mixture. This approach shows effectiveness, yet it has a few shortcomings. For instance, the examination is destructive, time consuming and temperature sensitive.

On the other hand, non-destructive methods using spectral imaging techniques, such as the gel ink discriminating method proposed in [23], show more potential. Hyperspectral imaging is a novel and efficient non-destructive tool for detection, enhancement [3] and comparison of document forensic evidence [4]. Pioneering work in multi-spectral imaging of documents was done by Easton et al. [9]. To gather the multi-spectral images, Christens-Barry et al. [5] developed an imaging system called EurekaVision. Hedjam et al. [6] introduced algorithms for historical documents restoration based on multi-spectral imaging, the same group later focused on scenarios where the ink is invisible to human eyes and proposed a new method in [7]. Hollaus et al. [8] employed not only spectral but also spatial information in their method of enhancing degraded images.

The National Archives of Netherlands developed a hyper spectral imaging system to analyze historical or fragile documents in libraries [10]. The system was able to acquire high resolution spectral responses from near-UV to near-IR range. Its drawback was that the acquisition time was about 15 minutes, which is too slow for practical use when a large number of images need to be captured [11].

Currently investigators manually analyze the captured hyperspectral images on a band-by-band basis and try to identify the differences through visual inspection. This process is not only laborious and time consuming, but also rather subjective. Therefore, an automatic ink analysis system would be significantly helpful for document examiners.

Silva et al. [25] focused on inks of different ages and introduced a chemometricbased method for forgery identification in the support of hyperspectral imaging techniques. On the other hand, Morales et al. [24] developed a hardware system for ink analysis based on hyperspectral analysis and Least Square SVM classification.

Khan et al. [12], [13] presented a hyperspectral document analysis system to distinguish different metameric inks. By collecting the spectral responses in 33 different bands of a hyperspectral image, the method can distinguish between different inks used to write a handwritten note. This method can be holistically applied to the handwritten note and assumes that the same amount of text has been written with each of the two inks to be distinguished. However, in most of the practical applications, the modification to an original note is usually minor and localized to a small region of the document. These important scenarios cannot be handled by [12], [13], thereby limiting the applicability of the method.

In this paper, we introduce a method that is able to identify whether a local image region contains pixels of one ink or

two inks, and distinguish the pixels belonging to each ink type. Using outlier estimation methods in combination with traditional clustering approaches, we are able to correctly make the decision even if there are a few strokes of one ink compared to the other ink. Unsupervised ink discrimination is casted as a clustering problem, where each pixel serves as a 33-dimensional data point. Hence we can draw similarities to unbalanced clustering, which is essentially an outlier detection problem. To simplify the scenario, we assume that there are a maximum of two inks in a localized region of the questioned document, which is often the case in practice. We use different criteria to establish whether pixels in the local region come from a single ink or two inks. These criteria are then used in combination with a number of outlier detection methods including LOF (Local Outlier Factor) [15], COF (Connectivity-based Outlier Factor) [16], and INFLO (Influenced Outlierness) [17]. Note that, our work concentrates on the analysis of inks spectral responses, instead of distinguishing handwritings by the texture [18] or ink-deposition traces [19].

In conclusion, we present the following three main contributions to the study of forgery detection in this paper:

- Extension of the system developed in [12] to scenarios where the number of inks is unknown and the inks proportion in the handwritten note is biased.
- Two criteria to determine the number of inks in a local image region are introduced and compared.
- Three anomaly detection methods are applied to handle extremely unbalanced inks distribution scenarios.

II. UWA HYPERSPECTRAL DOCUMENTS DATASET

We use the UWA Hyperspectral Documents Dataset [12] in our work. This database consists of 70 hyperspectral images of hand-written notes, which were written by seven subjects with ten different inks, including five blue inks and five black inks. All seven subjects wrote the same sentence *'The quick brown fox jumps over the lazy dog'*, once in each ink on a white paper. The inks appear visually similar, but they were made by different manufactures and so each ink has a unique spectral response. Based on the images of individual sentences, mixed ink images are produced as in [12] by merging the left half image of the first ink note with the right half image of the second note. A single hyperspectral image created in this way consists of text written with two inks such that the amount of text written with both inks is roughly equal.

The local image regions used in our experiments were all created from the mixed ink images produced as outlined above. A 100 pixels wide scanning window with the height equal to that of one text-line in the handwritten notes is used to scan the mixed ink image. When the window is near the beginning of the image, pixels belonging to only one ink are present in the local image region. However, as the window moves towards the center of the image, the local image captures some pixels of the second ink. The proportion of the pixels belonging to the two inks changes from highly unbalanced to balanced as the window moves towards the center of the image. When the window passes the center point, the proportion of inks again

changes and eventually only one ink pixels remain in the local image. This setting allows us not only to test whether we can find the presence of two inks in a local image region on not, but also to test this for different proportions of inks.

III. DETERMINING THE NUMBER OF INKS IN A LOCAL IMAGE REGION

Determining the number of different inks in a given local image region is, in principle, the same as recognizing the number of different clusters in a given dataset. Thus, the key to solve this problem is to identify a proper indicator for determining the number of clusters.

A. Schwartz's criterion

Schwartz's criterion [14] is often used with k-means clustering to determine the number of centers of a given dataset in an unsupervised manner [20]. The criterion determines k, the number of clusters, by measuring the distance between data points and their possible cluster centers. This is done by choosing a value of k that minimizes the sum of average center-to-point distances within a cluster and a factor concerning the number of clusters:

$$k = \arg\min_k (E_{k-\text{means}} + \lambda m k \log(N)) \qquad (1)$$

where $E_{k-\text{means}}$ is the average point-to-center distance in k-means, λ is a weighting factor, m is the length of each data point x, and N is the total number of data points. The weighting factor λ needs to be manually estimated. A smaller value of λ may result in a higher value of k. The output gives the predicted number of clusters (ink types) in a given dataset.

Its validity is based on the assumption that the average distance in a data of two clusters declines sharper as k grows compared to a data of single cluster. However, it may not be feasible in cases where data in different clusters are highly similar. As shown in Fig. 1(d), the Schwartz's criterion fails to show promising results in our experiment. The horizontal axes of Fig. 1 are the starting position of the local image of interest in the original note, while the vertical axes are the estimates of the number of inks in that local image. Comparing 1(d) with 1(c), we can see that for a large value of λ, the portion of note estimated as $k = 2$ is too narrow compared to the ground truth, whereas for a small value of λ, the criterion results in a large area of ambiguity.

B. Point-to-Point distance criterion

As an alternate to Schwartz criteria, we introduce point-to-point (P2P) distance. This criterion is also based on k-means, however, point-to-point (P2P) distances are calculated instead of point-to-center distances. In this criterion, we presume that the given region consists of two different inks and apply k-means algorithm on the dataset. We use k-means to generate two clusters of data, and calculate the average value of the minimum distances between each data point of one cluster to all points in the second cluster:

(a) Actual mixed ink image

(b) Ink Segmentation

(c) Ground Truth

(d) Schwartz's result

(e) P2P distances (E_{P2P})

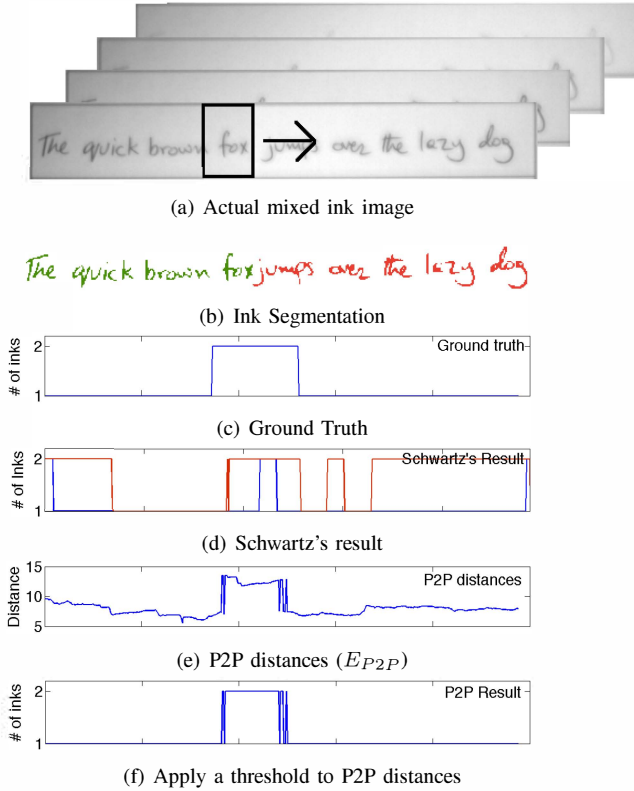(f) Apply a threshold to P2P distances

Fig. 1. From top to bottom: The handwritten note captured by the camera, local images are obtained by sliding a rectangular region window over the image. Ink segmentation of the note. The correct number of inks for each local image. The estimated number of inks using Schwartz criterion; the red and blue lines are respectively the results when λ is too small and too large. The P2P distance value of each local image; after applying a threshold to the previous image we get the estimated ink numbers in the bottom plot.

$$E_{P2P} = \frac{\sum_{p_1} \min D_{p_1 - C_2} + \sum_{p_2} \min D_{p_2 - C_1}}{\text{count}(C_1) + \text{count}(C_2)} \quad (2)$$

where p_1 and p_2 are data points of clusters C_1 and C_2 respectively, D calculates the distance, count returns the number of data points of a given cluster and the outcome of this equation E_{P2P} is the average minimum distance between the two clusters produced by k-means.

The value of E_{P2P} calculated from Eq. 2 is used in Eq. (3) to determine the value of ke, instead of directly predicting the number of ink types in the handwritten note.

$$k = \begin{cases} 2 & : \quad E_{P2P} > T \\ 1 & : \quad E_{P2P} \leq T \end{cases} \quad (3)$$

where E_{P2P} is the outcome of Eq. 2 and T is a value set manually according to the characteristics of the inks.

This criterion shows promising results for most ink combinations, as shown in Fig. 1(e) and 1(f), in which the values of E_{P2P} reflect the number of inks present in the local image region. However, the results are not stable in the regions where only a small fraction of the local image belongs to the other type of ink. Anomaly detection algorithms can be more suitable in these scenarios. Three different anomaly detection

algorithms, including LOF, COF and INFLO, were combined with the previous methods to test their ability to identify the inks in the case of unbalanced mixtures.

C. LOF anomaly detection

Local Outlier Factor (LOF) defines a local area as the k nearest neighbors of a given data point. It compares the local density of the data point to that of its neighbors, to accomplish this goal, the following definitions are used; ($N_k(p)$ denotes the k-neighborhood of a data point p), **reach-dist(p,o)** is the reachability distance. It is the maximum of $d(p, o)$ and $k -$ distance(o), in which $d(p, o)$ denotes the distance between p and o, $k-$distance(o) denotes the radius of its k neighborhood.

Local Reachability Density (lrd) is the inverse of the mean reachability distance over the neighborhood of the object.

$$\text{lrd}_{N_k(p)}(p) = \frac{|N_k(p)|}{\sum_{o \in N_k(p)} \text{reach-dist}(p,o)}$$

LOF is the ratio between the average local reachability density of the neighborhood to that of the object.

$$\text{LOF}_{N_k(p)}(p) = \frac{\sum_{o \in N_k(p)} lrd_{N_k(o)}(o)}{|N_k(p)| \cdot lrd_{N_k(p)}(p)}$$

Normal data would have LOF scores of approximately equal to 1, while outliers will have scores greater than 1.

D. COF anomaly detection

COF (Connectivity-based Outlier Factor) is a variant of LOF that uses the k-neighborhood and is also based on local density. In contrast to LOF, COF can handle outliers deviating from low density patterns. To calculate the COF score of object o, the following definitions are used (P and Q are two sets, object $p \in P$ and $q \in Q$):

d(P,Q) denotes the minimum distance between the elements of set P and Q.

nearest-neighbor(P,Q) If $\exists p$ that $d(p, q) = d(P, Q)$, then q is the nearest neighbor of P in Q.

Cost Description Sequence (CDS) defines an order to calculate the average chaining distance, [21] **Average Chaining Distance** $(n = |N_k(o)|)$

$$\text{ac-Dist}_{N_k(o)}(o) = \sum_{i=1}^{n} \frac{2(n-i+1)}{n(n+1)} \cdot \text{CDS}[i]$$

COF is the ratio of the ac-Dist of o to the mean of that over the k-neighborhood of o.

$$\text{COF}_k(o) = \frac{|N_k(o) \cdot \text{ac-Dist}_{N_k(\bullet)}(o)|}{\sum_{p \in N_k(\bullet)} \text{ac-Dist}_{N_k(\bullet)}(\bullet)}$$

Similar to LOF, a COF score of over 1 indicates that the point is an outlier.

E. INFLO anomaly detection

INFLO (INFLuenced Outlierness) is an anomaly detection algorithm introduced in order to handle the case where clusters with varying densities are in close proximity. The INFLO score of a data point can be calculated with Eq. 4.

$$\text{INFLO}_{k(p)} = \frac{\text{den}_{avg,k}(\text{IS}_k(p))}{\text{den}_k(p)} \quad (4)$$

(a) Actual mixed ink image

(b) Ink Segmentation

(c) Ground Truth

(d) Respectively shows the maximum values of the first 20 INFLO scores and that of the last 20 scores

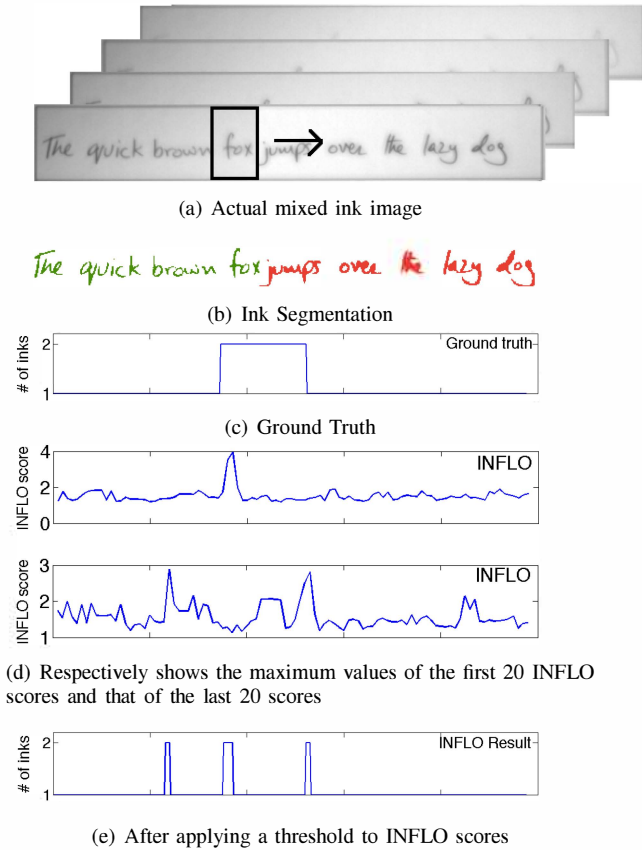(e) After applying a threshold to INFLO scores

Fig. 2. From top to bottom: The handwritten note captured by the camera, local images are obtained by sliding a rectangular region window over the image. Ink segmentation of the note. The correct number of inks for each local image. The maximum values of the first 20 INFLO scores of each local image. The maximum values of the last 20 INFLO scores. After applying a threshold to the previous figures, we gather the peaks in the bottom plot.

where $\text{den}_k(p)$ is the local density, calculated as the inverse of the k-distance of object p: $\text{den}_k(p) = 1/k - \text{distance}(p)$. $\text{den}_{avg,k}(S)$ is the average local density over set S,

$$\text{den}_{avg,k}(S) = \frac{\sum_{i \in S} \text{den}_k(i)}{|S|} \qquad (5)$$

IS is the influence space of the given data point p

$$\text{IS}_k(p) = \text{RNN}_k(p) \cup \text{N}_k(p) \qquad (6)$$

where RNN_k is the Reverse k Nearest Neighbor set.

$$\text{RNN}_k(p) = \{q | q \in D \wedge p \in \text{N}_k(q)\} \qquad (7)$$

The INFLO score of an anomaly would be higher than that of a normal point, therefore, a peak of INFLO score is expected around the border area, where the small amount of data points belonging to the other type of ink would be considered as anomaly (see Fig. 2(e)). Plots of the maximum values of the last 20 INFLO scores and the first 20 INFLO scores of each local image region are shown in 2(d). It can be observed that the peaks in each image in 2(e) roughly correspond to the cases of region boundaries in 2(c) obtained by thresholding INFLO scores. Though a false alarm occurs,
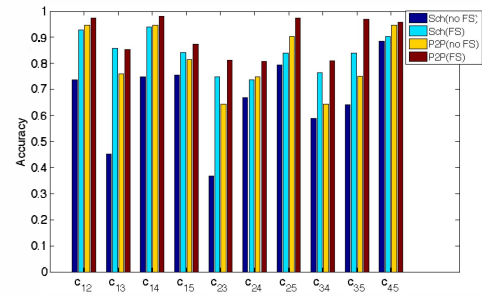

Fig. 3. Comparison of Schwartz and P2P criterion with and without feature selection (Blue inks)
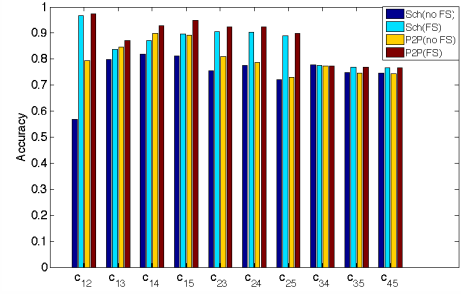

Fig. 4. Comparison of Schwartz and P2P criterion with and without feature selection (Black inks)

the other two peaks accurately match the two boundary cases in the ground-truth.

IV. COMPARATIVE ANALYSIS

To assess the effectiveness of different methods, we calculate the accuracy of each individual criterion. The accuracy measures the percentage of local images where the number of inks is correctly predicted, it ranges from 0 to 1 and is defined as

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Windows}} \qquad (8)$$

Since hyperspectral images contain a lot of redundant information, especially in nearby bands, removal of redundant information has previously shown to increase ink mismatch detection accuracy [12]. Therefore, we perform the first experiment to evaluate the suitability of the forward band / feature selection scheme [12] for use with Schwartz and P2P distance criteria. Figs. 3 and 4 compare the accuracy of Schwartz's criterion and P2P criterion with and without feature selection. The results show that P2P criterion with feature selection achieves the highest accuracy. Hence for the next experiments, we use P2P criterion with feature selection as the baseline method.

Fig. 5 and 6 compare different anomaly detection algorithms used to spot the presence of two inks when their relative proportion is highly unbalanced. The four methods are respectively the baseline method, and its combinations with LOF, COF and INFLO anomaly detection algorithms. As shown in Fig. 5 and 6, for most of the ink combinations, INFLO
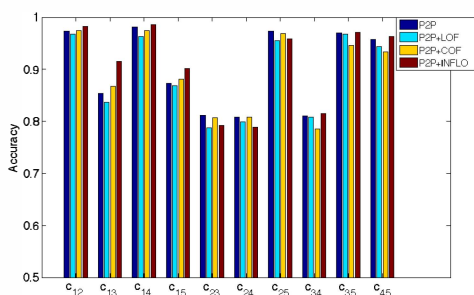
Fig. 5. Accuracy of different Anomaly Detection methods in combination with P2P distance with feature selection (Blue inks)
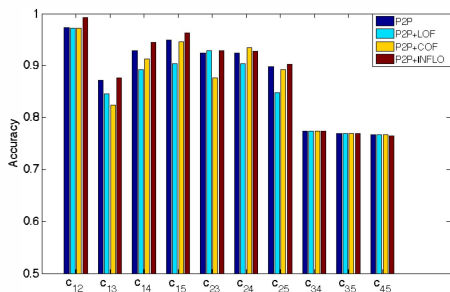

Fig. 6. Accuracy of different Anomaly Detection methods in combination with P2P distance with feature selection (Black inks)

method improves the performance, while LOF and COF have not shown significant enhancements.

For blue ink combinations C_{13}, C_{15} and black ink combinations C_{12}, C_{15}, the improvement of combining INFLO is significant. However for some combinations, for instance C_{23} and C_{24} in blue inks, instead of increasing, the accuracy declines after implementing INFLO. A possible reason could be that the characteristics of C_{23} and C_{24} were confusing to INFLO. On the other hand, for most ink combinations, the INFLO outperforms LOF and COF methods for unbalanced mixture identification. In conclusion, combining INFLO anomaly detection algorithm with the P2P criterion with feature selection shows the most promising results.

V. Conclusions and Future Improvements

In this research, we studied two main criteria (Schwartz and P2P), to predict the number of inks used in a local region of a handwritten note. We assume that for each local region, the maximum number of inks is two. Three anomaly detection methods were evaluated to distinguish inks when their relative proportions were highly unbalanced, which is a practically significant problem. The presented methods were able to reliably distinguish most of the ink combinations. The results demonstrate that the criterion would be very helpful for investigators to detect if there is an ink mismatch in a handwritten note, which may lead to determining whether the document is forged.

Acknowledgment

References

[1] E. H. Land, J. J. McCann et al., Lightness and retinex theory, Journal of the Optical society of America, vol. 61, no. 1, pp. 111, 1971.

[2] V. Aginsky, Forensic examination of slightly soluble ink pigments using thin-layer chromatography, Journal of Forensic Sciences, vol. 38, 1993.

[3] S. Joo Kim, F. Deng, and M. S. Brown, Visual enhancement of old documents with hyperspectral imaging, Pattern Recognition, vol. 44, no. 7, pp. 14611469, 2011.

[4] G. Edelman, E. Gaston, T. van Leeuwen, P. Cullen, and M. Aalders, Hyperspectral imaging for non-contact analysis of forensic traces, Forensic Science International, vol. 223, pp. 2839, 2012.

[5] W. A. Christens-Barry, K. Boydston, F. G. France, K. T. Knox, R. L. Easton Jr, & M. B. Toth, "Camera system for multispectral imaging of documents". Proc SPIE Sensors, Cameras, and Systems for Industrial/Scientific Applications X, pp. 724908-724908, 2009.

[6] R. Hedjam, and M. Cheriet, "Historical document image restoration using multispectral imaging system". Pattern Recognition, vol. 46, no. 8, pp. 2297-2312, 2013.

[7] R. Hedjam, M. Cheriet, and M. Kalacska, "Constrained Energy Maximization and Self-Referencing Method for Invisible Ink Detection from Multispectral Historical Document Images". 22nd Int Conf on Pattern Recognition (ICPR), Aug 2014, pp. 3026-3031.

[8] F. Hollaus, M. Gau, and R. Sablatnig, "Enhancement of multispectral images of degraded documents by employing spatial information", 12th Int Conf on Document Analysis and Recognition (ICDAR), Aug 2013, pp. 145-149.

[9] R. L. Easton Jr, K.T. Knox, and W.A. Christens-Barry, "Multispectral imaging of the Archimedes palimpsest". 32nd IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Oct 2003, pp. 111-116.

[10] R. Padoan, T. A. Steemers, M. Klein, B. Aalderink, and G. de Bruin, Quantitative hyperspectral imaging of historical documents: technique and applications, ART Proceedings, 2008.

[11] M. E. Klein, B. J. Aalderink, R. Padoan, G. De Bruin, and T. A. Steemers, Quantitative hyperspectral reflectance imaging, Sensors, vol. 8, no. 9, pp. 55765618, 2008.

[12] Z. Khan, F. Shafait, A. Mian, "Hyperspectral Imaging for Ink Mismatch Detection", 12th Int Conf on Document Analysis and Recognition (ICDAR), Aug 2013, pp. 877-881.

[13] Z. Khan, F. Shafait, A. Mian, "Automatic ink mismatch detection for forensic document analysis", Pattern Recognition, 2015.

[14] A. W. Moore. "K-means and hierarchical clustering". Lectures in datamining, CMU, 2001.

[15] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-Based Local Outliers". Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, May 2000, pp. 93104.

[16] J. Tang, Z. Chen, A. W. Fu, and D. W. Cheung, "Enhancing effectiveness of outlier detections for low density patterns". Advances in Knowledge Discovery and Data Mining, pp. 535548, 2002.

[17] W. Jin, A. Tung, J. Han, and W. Wang. "Ranking outliers using symmetric neighborhood relationship". Advances in Knowledge Discovery and Data Mining, pp. 577593, 2006.

[18] K. Franke, O. Bunnemeyer, and T. Sy, Ink texture analysis for writer identification, Proc. IEEE Workshop on Frontiers in Handwriting Recognition, 2002, pp. 268273.

[19] K. Franke and S. Rose, Ink-deposition model: The relation of writing and ink deposition processes, Proc. IEEE Workshop on Frontiers in Handwriting Recognition, 2004, pp. 173178.

[20] A. K. Jain, M. N. Murty, and P. J. Flynn, Data clustering: a review, ACM computing surveys, vol. 31, no. 3, pp. 264323, 1999.

[21] M. Amer, "Comparison of unsupervised anomaly detection techniques", Bachelors Thesis, German University in Cairo, Egypt, 2011.

[22] L. C. Molina, L. Belanche, and A. Nebot, Feature selection algorithms: A survey and experimental evaluation, in Proc. International Conference on Data Mining, 2002, pp. 306313.

[23] G. Reed, K. Savage, D. Edwards, and N. Nic Daeid, Hyperspectral imaging of gel pen inks: An emerging tool in document analysis, Science & Justice, vol. 54, no. 1, pp. 7180, 2014.

[24] A. Morales, M. A. Ferrer, M. Diaz-Cabrera, C. Carmona, and G. L. Thomas, The use of hyperspectral analysis for ink identification in handwritten documents, in ICCST. IEEE, 2014, pp. 15.

[25] C. S. Silva, M. F. Pimentel, R. S. Honorato, C. Pasquini, J. M. Prats-Montalban, and A. Ferrer, Near infrared hyperspectral imaging for forensic analysis of document forgery, Analyst, vol. 139, no. 20, pp. 51765184, 2014.