

Differentiating Pen Inks in Handwritten Bank Cheques Using Multi-layer Perceptron

Prabhat Dansena¹  , Soumen Bag¹  , and Rajarshi Pal² 

¹ Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Jharkhand, India

p.dansena23@gmail.com, bagsoumen@gmail.com

² Institute for Development and Research in Banking Technology, Hyderabad, India
iamrajarshi@gmail.com

Abstract. In handwritten Bank cheques, addition of new words using similar color pen can cause huge loss. Hence, it is important to differentiate pen ink used in these types of documents. In this work, we propose a non-destructive pen ink differentiation method using statistical features of ink and multi-layer perceptron (MLP) classifier. Large sample of blue and black pen ink is acquired from 112 Bank cheque leaves, written by nine different volunteers using fourteen different blue and black pens. Handwritten words are extracted from scanned cheque images manually. Pen ink pixels are identified using K -means binarization. Fifteen statistical features from each color handwritten words are extracted and are used to formulate the problem as a binary classification problem. MLP classifier is used to train the model for differentiating pen ink in handwritten Bank cheques. The proposed method performs efficiently on both known and unknown pen samples with an average accuracy of 94.6% and 93.5% respectively. We have compared the proposed method with other existing method to show its efficiency.

Keywords: Bank cheques · Handwritten forensics · Ink analysis · K -means clustering · MLP · Statistical features

1 Introduction

Forensics in handwritten Bank cheques from the perspective of differentiating pen ink have great importance to the judicial system. In handwritten Bank cheque forensics, it is often important to establish a relation between the pen inks. It helps to identify whether a single pen has been used to write the Bank cheque or multiple pens. Numerous possibilities of fraud exist in handwritten Bank cheques. In this work, we focus only on pen ink differentiation in Bank cheques. Possibilities of fraud in any Bank cheque and its consequences helps to understand the importance of the work.

Example of new words addition in Bank cheque using a different pen is depicted in Fig. 1 which is elaborated as follows. The cheque was initially issued

to Mr. Ravi Kumar Singh, amounting to Seventy thousand only. Later, forger appended new words in pay name and amount section as marked by red circles in Fig. 1. This difference in pen ink can not be always perceived by naked eye. This type of case helps us to understand the possibility of addition of new words in handwritten Bank cheques. A number of handwritten document frauds are possible in bill, business agreement, educational documents, etc. This motivates us to differentiate pen ink in Bank cheques.



Fig. 1. Addition of handwritten words (marked by red circles) in a Bank cheque image. (Color figure online)

Pen ink analysis techniques can be categorized in two major pathways: destructive and non-destructive techniques. Merrill and Bartick [1] have used infrared spectrum to differentiate pen ink. Taylor [2] has proposed a method to analyze intersecting lines using stereo microscope, distilled water, and wax lift techniques. Taylor [3] has also proposed TLC plate analysis method using solvent and micro-dispenser for pen ink classification. The second category of technique includes non-destructive techniques, which include modern chromatographic, image processing, and pattern recognition techniques. Khan *et al.* [4] have used spectral response and *K*-means clustering algorithm for pen ink difference identification. Khan *et al.* [5] have also used Principal Component Analysis for spectral response feature reduction. Then *K*-means clustering has been done to differentiate pen ink. Dasari and Bhagvati [6] have proposed statistical features of ink pixels from HSV color channel and distance measure based classification is performed. Kumar *et al.* [7,8] have shared statistical features as gray-level co-occurrence (GLCM), geometric, and legendre moments from YC_bC_r and opponent color models. In these methods, nearest neighbor and Support Vector Machine with feature selection have been used as classifiers to differentiate pen ink. Gorai *et al.* [9] have extracted twelve feature images from color input image and corresponding gray version using local binary pattern and Gabor filters. In this method, histograms of pen ink pixels from feature images are calculated and histogram matching has been performed to identify the ink mismatch.

It is observed that most of the works in the area of non-destructive ink analysis ranges from hyper-spectral and microscopic imaging to chromatographic technique. This requires high configuration hardwares those are too costly as well as rarely available in market. In this paper, we have proposed a method that is capable of differentiating pen ink using simple standard scanning devices. Such devices are easily available and at the same time cost effective. In this method, pen ink samples are extracted manually from scanned Bank cheque leaves. K -means binarization has been used to identify ink pixels from each color channel of word images. Statistical features of ink pixels are extracted from each channel. Extracted feature set is used to train the MLP classifier for pen ink difference identification.

The rest of the paper is organized as follows. Section 2 discusses the proposed methodology for pen ink differentiation in handwritten Bank cheques. Experimental results and relevant discussion are presented in Sect. 3. The concluding remarks are given in Sect. 4.

2 Proposed Model

In this proposed method, pair of words have been analyzed to detect whether they have been written by same pen or not. Pen ink differentiation problem is formulated into a binary classification problem where two different pens are used to write on a particular Bank cheque. If two different pens are used to write word-pairs in a same Bank cheque, then it is labeled as class-I; otherwise it is labeled as class-II. The system architecture of the proposed method is depicted in Fig. 2.

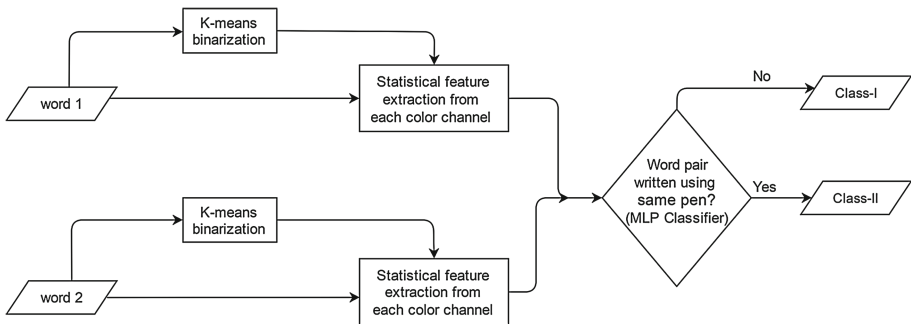


Fig. 2. System architecture of the proposed method.

2.1 K -means Algorithm Based Foreground Pixel Identification

Pen ink pixels (PI) identification is an important task in handwritten Bank cheque for differentiating pen ink. We have used K -means algorithm to binarize

the word images for this purpose. Basic idea behind K -means is to minimize the objective function (i.e., inter cluster Euclidean distance), where K is an user defined parameter. In our experiment, we have chosen $K = 2$ to identify PI as foreground pixels. Color handwritten word image extracted from Fig. 1 is taken as input (Fig. 3a) and corresponding gray image is obtained. Gray version of input is used to identify the PI in color handwritten word image. K -means binarization partitions n gray values into K clusters, which separates the foreground from the background. This binarization method is used to identify PI as foreground pixels as depicted in Fig. 3b. This method works well for ink pixels identification because foreground and background intensity profiles are not overlapping in handwritten word images.



Fig. 3. K -means image binarization: (a) Color input image; (b) Binarized output image.

2.2 Extraction of Statistical Features from Ink Pixels

Once coordinates of ink pixels (i, j) are identified using K -means binarization, following five statistical features are extracted from each color channel of ink pixels.

(a) **Mean:-** The Mean (\bar{m}) for ink pixels is defined by

$$\bar{m} = \frac{m_{xy}}{N}, \text{ where} \tag{1}$$

$$m_{xy} = \sum_{j=0}^y \sum_{i=0}^x w_k(i, j) \mid (i, j) \in PI \tag{2}$$

$$N = \sum_{j=0}^y \sum_{i=0}^x 1 \mid (i, j) \in PI \tag{3}$$

(b) **Variance:-** The Variance (Var) for ink pixels is defined by

$$Var = \frac{1}{N-1} \sum_{j=0}^y \sum_{i=0}^x [w_k(i, j) - \bar{m}]^2 \mid (i, j) \in PI \tag{4}$$

(c) **Skewness:-** The Skewness ($Skew$) for ink pixels is defined by

$$Skew = \frac{1}{N} \sum_{j=0}^y \sum_{i=0}^x \left[\frac{w_k(i, j) - \bar{m}}{\sqrt{Var}} \right]^3 \mid (i, j) \in PI \tag{5}$$

(d) **Kurtosis:-** The Kurtosis ($Kurt$) for ink pixels is defined by

$$Kurt = \left\{ \frac{1}{N} \sum_{j=0}^y \sum_{i=0}^x \left[\frac{w_k(i,j) - \bar{m}}{\sqrt{Var}} \right]^4 - 3 \right\} \mid (i, j) \in PI \tag{6}$$

(e) **Mean Absolute Deviation:-** The Mean Absolute Deviation (MAD) for ink pixels is defined by

$$MAD = \frac{1}{N} \sum_{j=0}^y \sum_{i=0}^x |w_k(i, j) - \bar{m}| \mid (i, j) \in PI \tag{7}$$

Where N is total number of foreground pixels, defined by the Eq. 3. Foreground pixels of handwritten word ($w(i, j)$) is defined by PI using K -means binarization. Handwritten word from each color channel $R, G,$ and B are denoted by $w_k(i, j)$, where $k = \{R, G, B\}$ and (i, j) is coordinates of ink pixels.

2.3 Differentiation of Pen Ink Using MLP Classifier

MLP classifier is used for differentiating pen inks. MLP architecture with input layer, output layer, and one hidden layer with seventeen computational nodes is considered for our experimental purpose. Sigmoid activation function is used in our MLP architecture. Features from two words under consideration are fed into the MLP network to identify whether same pen has been used or not. This MLP architecture is trained with 5000 iterations at learning rate $\alpha = 0.2$. Post training MLP architecture is used for classification of known and unknown pen samples.

3 Experimental Results and Discussion

3.1 Data Set Acquisition

Data is extracted from the IDRBT Cheque Image Dataset [10] with diverse texture and ink color. Total 112 cheque leaves from four different Indian Banks are used as source document. In order to simulate the pen ink difference in cheque leaves, seven blue and seven black pens are used. To avoid biasness due to writing, nine different volunteers have taken active participation to prepare data set. A total of $14 \times 9 = 126$ pen-volunteer combinations (fourteen pens and nine volunteers) are used for pen ink data generation. In practical scenario, similar color pens are used for addition of new words in source document. Each cheque is written by two volunteers using two different pens (either blue or black). Hence, data set is created with $2 \times 7C_2 = 42$ possible combinations of blue and black pens. All the cheque leaves are scanned in normal scanner at 300 dpi resolution. Handwritten words from each scanned cheque are cropped manually and grouped based on pen used to write the words.

Table 1. Proposed method accuracy for known and unknown pen.

S.N	Pen Combination kept out	Known Pen Accuracy(%)	Unknown Pen Accuracy(%)
1	$P_1 P_2$	93.34	93.85
2	$P_1 P_3$	95.37	93.07
3	$P_1 P_4$	94.02	93.52
4	$P_1 P_5$	94.86	93.65
5	$P_1 P_6$	94.42	94.13
6	$P_1 P_7$	94.80	93.10
7	$P_2 P_3$	95.68	93.29
8	$P_2 P_4$	93.96	93.54
9	$P_2 P_5$	95.33	93.46
10	$P_2 P_6$	95.39	92.95
11	$P_2 P_7$	95.57	93.63
12	$P_3 P_4$	95.50	92.49
13	$P_3 P_5$	96.11	93.10
14	$P_3 P_6$	95.77	93.74
15	$P_3 P_7$	96.75	92.83
16	$P_4 P_5$	94.86	93.56
17	$P_4 P_6$	94.96	92.71
18	$P_4 P_7$	95.19	93.25
19	$P_5 P_6$	96.94	92.69
20	$P_5 P_7$	96.01	92.22
21	$P_6 P_7$	94.89	93.09
22	$P_8 P_9$	94.53	93.61
23	$P_8 P_{10}$	93.81	94.40
24	$P_8 P_{11}$	92.90	93.91
25	$P_8 P_{12}$	94.32	94.11
26	$P_8 P_{13}$	93.57	93.68
27	$P_8 P_{14}$	93.32	94.41
28	$P_9 P_{10}$	94.80	93.73
29	$P_9 P_{11}$	94.42	92.72
30	$P_9 P_{12}$	95.66	93.25
31	$P_9 P_{13}$	93.48	93.74
32	$P_9 P_{14}$	94.13	94.39
33	$P_{10} P_{11}$	94.12	93.13
34	$P_{10} P_{12}$	94.47	93.97
35	$P_{10} P_{13}$	94.58	93.83
36	$P_{10} P_{14}$	93.65	93.23
37	$P_{11} P_{12}$	94.47	94.23
38	$P_{11} P_{13}$	93.39	93.46
39	$P_{11} P_{14}$	92.62	93.32
40	$P_{12} P_{13}$	94.25	93.52
41	$P_{12} P_{14}$	95.20	93.95
42	$P_{13} P_{14}$	91.86	94.58
Average Accuracy	-	94.60	93.50

3.2 Experimental Set-up

In each cheque, two pens P_i and P_j are used for writing m and n number of different words respectively. Set W_{p_i} and W_{p_j} contains words written by P_i and

P_j respectively, where $W_{P_i} = \{m_1, m_2, \dots, m_m\}$ and $W_{P_j} = \{n_1, n_2, \dots, n_n\}$. The word pairs written by different and same pens are considered in case-I and case-II respectively.

Case-I: Two different pens are used to write the word pairs. The Cartesian product of $W_{P_i} \times W_{P_j} + W_{P_j} \times W_{P_i}$ includes the total number of word-pairs written using different pens, where $W_{P_i} \times W_{P_j} = \{(m_i, n_j) \mid m_i \in W_{P_i} \wedge n_j \in W_{P_j}\}$ and $W_{P_j} \times W_{P_i} = \{(n_j, m_i) \mid n_j \in W_{P_j} \wedge m_i \in W_{P_i}\}$. Thus, total number of word-pairs for class-I will be $2 \times (m \times n)$.

Case-II: Same pen is used to write the word pairs. The Cartesian product of $W_{P_i} \times W_{P_i} + W_{P_j} \times W_{P_j}$ includes the total number of word-pairs written using same pen, where $W_{P_i} \times W_{P_i} = \{(m_i, m_i) \mid m_i \in W_{P_i}\}$ and $W_{P_j} \times W_{P_j} = \{(n_j, n_j) \mid n_j \in W_{P_j}\}$. Thus, total number of word-pairs for class-II will be $\{(m \times m) - m\} + \{(n \times n) - n\}$, after excluding the pairs of word with itself. For each cheque, total instances of class-I and class-II are calculated and stored. The number of word pairs for case-I and case-II in Fig. 1 can be calculated as follows. Set of words written using pens P_1 and P_2 are $W_{P_1} = \{J, Two, lakh\}$ and $W_{P_2} = \{Ravi, Kumar, Singh, Seventy, thousand\}$ respectively. The total number of word pairs for case-I are $(3 \times 5) + (5 \times 3) = 30$. The number of word pairs belongs to the case-II are $\{(3 \times 3) - 3\} + \{(5 \times 5) - 5\} = 26$. Thus, total instances including class-I (30) and class-II (26) are $30 + 26 = 56$.

To simulate pen ink difference identification, seven blue and seven black pens are used on Bank cheques. Each instance has thirty features and a class value. For each instance, $2 \times 15 = 30$ features are extracted from each handwritten word pair under consideration. The whole data set is divided into three subsets, namely training, validation, and test set using leave-k-out method. $K = 2$ is used to keep two unknown pen samples out for testing and performance evaluation of MLP classifier. Keeping two pens out, total possibilities are $2 \times 7C_2 = 42$ for both blue and black pen samples. Remaining data set after excluding the test subset is partitioned into ten approximately equal parts. One of ten data parts is kept as validation set remaining partitions are used as training set. The process of selecting validation set is repeated ten times, with each one of the ten data parts exactly once. Training set is used to train the MLP model inter and intra class difference. Validation is performed to check MLP classifier performance on known pen ink samples. Model testing is performed on the test set to check the performance of the MLP model on unknown pen ink samples.

3.3 Experimental Results and Comparison

We evaluate the performance of the binary classification problem for differentiating pen ink in handwritten Bank cheque. Both blue and black pen average accuracy of MLP classifier is presented in Table 1 for known and unknown pen samples, where P_1-P_7 and P_8-P_{14} are black and blue pens respectively. To show the efficiency of the proposed work, result analysis is performed using leave 2 pen out method. The average accuracy on both blue and black pen of MLP classifier is 94.60% and 93.50% for known and unknown pen samples respectively.

Table 2. Comparison in between proposed and existing method.

Method	Black Pen Accuracy(%)	Blue Pen Accuracy(%)	Average Accuracy(%)
Gorai <i>et al.</i> [9]	54.71	54.8	54.76
Proposed method	93.77	93.23	93.50

We have compared our result with Gorai *et al.* [9], which introduced technique for ink analysis and difference identification using simple scanning devices. Moreover, this method [9] did not take biasness due to writer into consideration. Our proposed method has taken this issue into consideration and provides better results than the previous one. A comparative analysis of proposed method with method in [9] is presented in Table 2.

4 Conclusion

In this paper, we have proposed pen ink difference identification method in handwritten Bank cheques. Differentiation of pen ink problem is formulated as a binary classification problem. Thirty features for each instance of word pair are extracted. These extracted features are used to train the MLP classifier on known pen ink pixels. Performance of MLP classifier is evaluated on both known and unknown pen ink pixels. Result analysis and comparison shows the superiority of the proposed method over the existing method on both black and blue pen samples.

Acknowledgment. A part of this work is sponsored by the project “Design and Implementation of Multiple Strategies to Identify Handwritten Forgery Activities in Legal Documents” (No. ECR/2016/001251, Dt.16.03.2017), SERB, Govt. of India.

References

1. Merrill, R.A., Bartick, E.G.: Analysis of ball pen inks by diffuse reflectance infrared spectrometry. *J. Forensic Sci.* **29**(1), 92–98 (1992)
2. Taylor, L.R.: Intersecting lines as a means of fraud detection. *J. Forensic Sci.* **37**(2), 528–541 (1984)
3. Taylor, L.R.: Developments in the analysis of writing inks on questioned documents. *J. Forensic Sci.* **37**(2), 612–619 (1992)
4. Khan, Z., Shafait, F., Mian, A.: Hyperspectral imaging for ink mismatch detection. In: *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 877–881 (2013)
5. Khan, Z., Shafait, F., Mian, A.: Automatic ink mismatch detection for forensic document analysis. *Pattern Recogn.* **48**(11), 3615–3626 (2015)
6. Dasari, H., Bhagvati, C.: Identification of non-black inks using HSV color spaces. In: *Proceedings of the International Conference on Document Analysis and Recognition*, pp. 486–490 (2007)

7. Kumar, R., Pal, N.R., Sharma, J.D., Chanda, B.: A novel approach for detection of alteration in ball pen writings. In: Proceedings of International Conference on Pattern Recognition and Machine Intelligence, pp. 400–405 (2009)
8. Kumar, R., Pal, N.R., Sharma, J.D., Chanda, B.: Forensic detection of fraudulent alteration in ball-point pen strokes. *IEEE Trans. Inf. Forensics Secur.* **7**(2), 809–820 (2012)
9. Gorai, A., Pal, R., Gupta, P.: Document fraud detection by ink analysis using texture features and histogram matching. In: International Joint Conference on Neural Networks, pp. 4512–4517 (2016)
10. IDRBT Cheque Image Dataset: <http://www.idrbt.ac.in/icid.html>