

A Novel Approach for Detection of Alteration in Ball Pen Writings

Rajesh Kumar¹, Nikhil R. Pal², J.D. Sharma³, and Bhabatosh Chanda²

¹ Directorate of Forensic Science, MHA, GOI, New Delhi, India

² ECSU, Indian Statistical Institute, Kolkata, India

³ Dr. HSG University, Sagar (M. P.), India

Abstract. Addition or alteration to documents that have profound implication is very common. The technique that Forensic Document Examiners (FDEs) use for the examination of such documents is basically a physical examination. In this paper we consider the alteration detection as a two-class pattern recognition problem. Image processing techniques are used for feature extraction and a neural network based feature analysis technique is used for finding a set of discriminatory features. The results using a nearest neighbor classifier are very encouraging. The results also demonstrate the effectiveness of feature analysis.

Keywords: Alteration, ball pen, feature analysis, image processing.

1 Introduction

Although world is moving from the paper age to paperless age, billions of ball point pens are sold every year and people are frequently using these pens for writing and signing on various documents. Hence, number of forgeries involving these materials is also high. A little alteration by similar color ink in the amount written on a cheque can do havoc. Everyday at some part of the world people are victimized by such a white collar crime.

The techniques that FDEs are using in routine cases is basically a physical examination through different light sources ranging from ultraviolet to infrared and some kind of conventional optical filters. Osborne [1], the father of forensic document examination, himself suggested the physical examination of altered documents in various lighting conditions and also the use of some chemical examination. The work on alteration detection using ink analysis is being done in two major pathways. First is non-destructive kind of examination using techniques like Fourier-Transform Infrared (FTIR) microscopy, IR luminescence and laser examination. The second approach is based on destructive techniques ranging from basic and modern chromatographic techniques to Neutron activation analysis [2]. Application of image processing and pattern recognition for this problem is still limited to the extent of using some image enhancement techniques as an aid to the FDEs.

In this paper we have taken the alteration detection as a two-class problem. A machine learning technique is used for feature analysis and the nearest neighbor classifier is used for detection.

2 Proposed Methodology

The main type of cases that the FDEs get is insertion of some strokes on pre-existing strokes. To simulate this problem 10 different ball point pens of different brands are used. As quality of paper also has some influence on writing stroke and we are interested in detecting whether pens are same or different, the same brand of paper is used. Using combinations of these 10 pens we have prepared $^{10}C_2 = 45$ combinations of two intersecting strokes on the paper [see Fig.1(a)]. To eliminate bias on the order of strokes, the order of strokes is interchanged to create another 45 combinations. Two copies of each of these 90 combinations are then created. The pictures of intersecting strokes are enlarged 80 times and corresponding images are captured using VSC 5000.

The background area due to paper is removed by simple graylevel thresholding method. Each intersecting stroke in the image is then divided into two non-overlapping segments which are used for feature extraction.

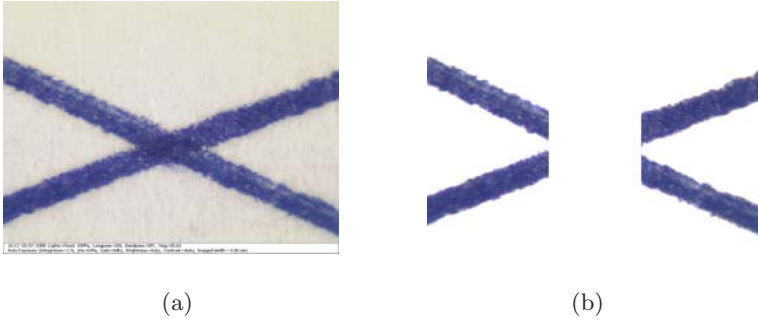


Fig. 1. A typical sample image data (a): Image of intersecting strokes (b): segmented strokes used for feature extraction

2.1 Feature Extraction

Our aim is to extract some features that are close to what can be perceived by human and can be used to distinguish between different pens. Keeping this in mind we have chosen YC_bC_r and opponent chromaticity space (rg, yb) to compute various features. The selection of two color spaces simultaneously can be justified through opponent process theory of color vision. According to that theory, the trichromatic theory of color vision does not explain all aspects of color vision. The reason for not seeing combinations like reddish-green or yellowish-blue, though that can be formed in YC_bC_r space, is that the opponent responses are controlled by the opponent neurons. So, considering both the spaces one can get complete color description of an object. A survey of trichromatic color spaces can be found in [3,4], while opponent chromaticity space is described in [4,5].

We compute various moments of joint distribution of various color channels as color features and various moments of graylevel co-occurrence matrix (gldcm)

as texture features. In particular, we calculate Legendre and geometric moments [6,7] up to third order for each of rg-yb and C_b-C_r spaces. Thus 18 moment features from each of the two spaces in addition to mean and standard deviation for each channel are the color features of each channel. We also compute three sets of texture features namely contrast, homogeneity and energy from glcms [8] defined for four distances (10, 20, 30 and 40) and involving all eight directions. Thus a total of 106 features are extracted for each stroke segment.

2.2 Detection of Alteration

We are considering alteration detection as a two-class problem: class-I, when the same pen is used to create two intersecting strokes, and class-II, when two different pens are used for the creation of the two strokes. Each image in the data set is having two intersecting strokes and each stroke is segmented into two segments. Thus, ${}^4C_2=6$ feature vectors are created from each image combining the features of 2 stroke segments at a time. By toggling the feature vectors of each stroke segment, another 6 feature vectors are obtained from the same image. Thus a total of 12 feature vectors each with 212 features are generated from each image. These 12 feature vectors are included into class-I (same pen stroke) or class-II (different pen stroke) according to their combination.

The nearest neighbor classifier is used for the detection of alteration due to its simplicity and robustness. One particular pen P_k is kept out and the data for remaining 9 pens are used as the training data. The process is repeated for each pen $P_k; k = 1, \dots, 10$. By doing so, one can avoid bias towards any pen.

2.3 Feature Analysis

We shall see later that the detection results with the nearest neighbor classifier using all features exhibit some undesirable behavior when a particular pen is left out for testing. This could be because of presence of some poor feature(s) that may be affecting the performance noticeably when a particular pen is left out. Moreover, use of all 212 features may not be needed for this problem. For an easy identification of the decision making system as well as for better generalization it is desirable to use just an adequate set of features, possibly with limited redundancy. Hence, with a view to improve the overall classification performance and with a hope to realize a more uniform performance when different pens are left out, we do some feature analysis.

There are many feature selection techniques available in the literature [9]. Here we use a Multilayered Perceptron (MLP) based feature selection method, that we call FSMLP, in short [10]. In this method we associate an adaptive gate to each input node (hence each feature) of the network. The gate is modeled by a monotonic differentiable function $g()$ whose range is $[0,1]$. Each gate has a tunable parameter. The degree to which a gate is opened determines the goodness of the feature. Unlike conventional MLP, here each input node modulates the input feature value by computing the product of the input feature and the gate function value. This modulated input is then passed into the next layer of the

network. The FSMLP begins its training assuming that all features are bad features. In other words, at the beginning, each modulated feature takes a value of nearly zero and that is passed into the higher layers of the network. We want to emphasize on the word *almost zero*. If the value is equal to zero then the learning cannot proceed. The learning process uses gradient descent to minimize the classification error. Therefore, the features which can reduce the error faster are likely to get their associated gates opened faster; while the gates associated with bad features (that cannot reduce the error) are likely to get closed more tightly. In all our experiments, we train the FSMLP only for 1000 iterations. There is no need to train FSMLP till error becomes very low. We can stop when the training error is just satisfactory as our objective is only to pick up the good features.

An important advantage of this method over many others is that it can account for the non-linear interaction between the features as well as that between features and the tool (here neural network). For details, please see [10]. Use of MLP type networks raises many issues like choice of network size etc. Here we have made a few experiments and found that a choice of 8 or 9 hidden nodes is good enough. Hence, in all our experiments, we have used 8 hidden nodes. Since such a network uses gradient search, depending on the initial conditions, different runs can result in different sets of features, each of which may be equally good. This can particularly happen when there are many correlated features. Hence, we proceed as follows. A pen, P_k , is kept out for testing. The remaining nine pens are then used for feature selection and classifier design. Using these nine pens, the FSMLP is run 10 times. Each run R generates a gate opening value for each feature, f as g_f^R ; $R = 1, \dots, 10$; $f = 1, \dots, 212$. Now we compute $g_f = \sum_{R=1}^{10} g_f^R$ as the composite importance for feature f and use these importance value to select a set of features for any experiment with pen P_k left out. Using these selected features we compute the performance of a nearest neighbor classifier. This process is repeated for each of the 10 pens, P_k ; $k = 1, \dots, 10$.

3 Results and Discussion

The second column of Table 1 shows the performance of a nearest neighbor classifier using all 212 features. Training is performed by keeping a particular pen P_k out and testing on pen P_k . One can see that when the pen P_1 is kept out the performance decreased to 69.41% while overall performance is around 78%. The declined performance of the pen P_1 with all 212 features motivated us to do feature analysis.

Figure. 2 shows the overall performance of the nearest neighbor classifier using different sets of features selected by FSMLP, the detailed performance is included in Table 1. Figure. 2 reveals that the performance of the classifier decreases in either side of 60 features. Thus only 60 ranked features are sufficient and good to discriminate strokes using the same pen or different pens. In the set of top 60 features, contrast in each of the five channels except in C_b , Homogeneity in C_b , mean of intensity in rg and Y, energy in rg and yb, third order geometric

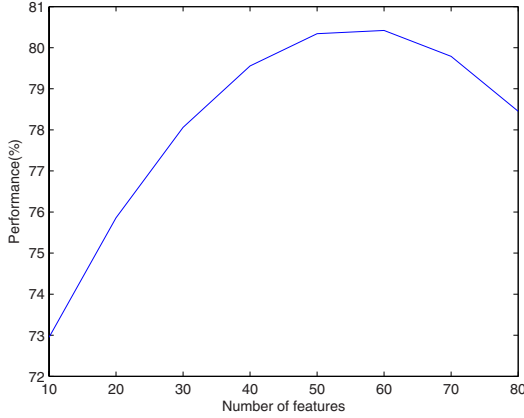


Fig. 2. A sketch of performance of a NN classifier with different number of features

Table 1. Performance(%) of a nearest neighbor classifier

No.of features	212	80	70	65	60	55	50	40
P_1	69.41	69.71	70.59	72.06	72.65	72.06	72.65	71.76
P_2	85.71	81.71	81.71	80.57	82.57	81.71	80.86	78.29
P_3	73.89	75.00	75.83	77.22	79.17	78.33	78.89	76.11
P_4	74.55	76.06	76.97	78.48	77.27	77.58	76.67	74.85
P_5	93.13	95.94	96.56	96.56	96.25	96.25	97.19	96.88
P_6	82.35	81.18	79.71	79.41	78.24	78.53	77.94	77.06
P_7	79.44	88.89	89.72	92.50	92.50	91.94	92.78	94.72
P_8	74.44	74.72	75.00	75.00	75.83	77.50	74.42	75.00
P_9	74.84	74.52	74.52	75.48	75.16	76.45	77.10	77.42
P_{10}	72.00	71.14	73.43	73.71	73.14	70.57	72.29	70.86
Total accuracy	77.98	78.89	79.40	80.10	80.28	80.09	80.11	79.29

Table 2. False Negative and False Positive rates(%) of a NN classifier

No.of features	212	80	70	65	60	55	50	40
False Negative	26.22	24.87	24.23	23.48	23.28	23.49	23.17	25.29
False Positive	05.22	06.08	06.06	05.59	05.48	05.59	06.78	06.29

and Legendre moments in rg-yb and one particular geometric moment in C_b-C_r appeared as the most consistent candidates for all pens.

One can easily compare the performance before and after the feature analysis as shown in Table 1. As expected, not only the overall performance increased to 80.28% but also the performance of the first pen has improved to 72.65%. Beside this we can see the uniformity in performance over all the pens.

Moreover, the best part of the classification is its low false positive rate (Table 2), that is very important for making legal decisions, where the motto is, *no innocent person should be convicted*. From Table 2, the impact of feature analysis can also be seen on false positive and false negative rates. We find that with 60 selected features the false negative rate goes down and the false positive rate is comparable to that of the case with all features.

4 Conclusion and Future Work

In this study we have attempted to solve an important forensic problem using image processing and pattern recognition techniques and achieved an accuracy of more than 80%. Moreover, we have obtained a low false positive rate of 5.48. which is good for making legal decisions. Our approach is quite novel and reliable as far as forensic document examination is concerned. The proposed algorithm can be an aid to FDEs for the alteration detection in combination with other conventional methods.

We shall continue our investigation to improve the performance as well as to extract and select robust features. Further, we are planning to assign a confidence value to each decision so that in case of low confidence additional information can be sought. Such a system would be of great help to forensic community.

References

1. Osborne, A.S.: Questioned Documents. Boyd Printing Co., New York (1929)
2. Ellen, D.: The Scientific Examination of Documents Methods and Techniques, 2nd edn. Taylor and Francis, London (2003)
3. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Printice Hall, New Jersey (1996)
4. Koschan, A., Abidi, M.: Digital Color Image Processing. Wiley Interscience, New Jersey (2008)
5. Berens, J., Finlayson, G.D., Qiu, G.: A Statistical Image of Color Space. In: IEE Proc. International conference on Image Processing and its Application (1), pp. 348–352 (1999)
6. Yap, P.T., Parmesaran, R.: Content Based Image Retrieval using Legendre Chromaticity Distribution moments. In: IEE Proc. Visual Image Signal Processing, pp. 17–24 (2006)
7. Teh, C.H., Chin, R.T.: On Image Analysis by the Methods of Moments. IEEE Transaction on Pattern Analysis and Machine Intelligence 10(4), 496–513 (1988)
8. Haralick, R.M., Bosley, R.: Texture Features for Image Classification. In: Proc. Third ERTS symposium, NASA, pp. 1219–1228 (1973)
9. Liu, H., Motoda, H.: Computational Methods of Feature Selection. Chapman and Hall/CRC, Taylor and Francis (2007)
10. Pal, N.R., Chintalapudi, K.K.: A Connectionist System for Feature Selection. Neural, Parallel and Scientific Computations (5), 359–382 (1997)