DISTINGUISHING LIQUID AND VISCOUS BLACK INKS USING RGB COLOUR SPACE

Dasari Haritha and Chakravarthy Bhagvati
Dept. of Computer and Information Sciences, University of Hyderabad, Hyderabad, India 500046

Keywords: Questioned Document Examination, Black inks, Image processing, MLR.

Abstract: Analysis of inks on Questioned documents is often required in the field of document examination. This paper provides a novel approach for ink type recognition for black inks. Ink types Liquid ink or Viscous ink will be derived from the colour properties of ink, by extracting its amount of blackness. This classification helps in distinguishing Gel and Roller pens versus Ball pens. Different types of inks exhibit different absorption characteristics that causes colour and distribution of colour pixels to change. We observed that the RGB colour space is useful to reveal the differences in ink types. We used multiple linear regression to model the RGB data points of the writings to a plane. The distance from the origin (pure black) to that plane is calculated to classify inks i.e. liquid inks and viscous inks. The distance measures in RGB and HSV colour spaces are used to identify the particular ink. The accuracy of identification is analysed using Type I and Type II errors.

1 INTRODUCTION

Forensic examination of documents is fast emerging as a challenging field of research with the proliferation of fake documents using computers or computer-based technologies. A document is labeled *questioned document* if its authenticity is in doubt. The main goal of examining questioned documents is to determine whether they are genuine or fake by detecting alterations if any. Alterations occur primarily in three forms: addition of new text, obliterations, and erasures.

The conventional methods used by expert questioned document examiners are varied and sophisticated involving both destructive and non-destructive means. References (Brunelle, 1982), (Ellen, 1997) and (Hilton, 1993) offer excellent overviews. Microscopy, study of colour variations under changing illumination (Hilton, 1993), filters for improving colour contrasts (Bauer, 1966), ultraviolet fluorescence, infrared absorbance (Godown, 1964), infrared luminescence (Hardcastle, 1978) and Fourier transform infrared spectrometry (Brunelle, 1982) have all been used for examining questioned documents. A special instrument, called the video spectral comparator, is often used for the above non-destructive testing methods. The destructive chemical procedures include solubility tests, thin layer chromatography and high performance liquid chromatography (HPLC) (Tappolet, 1983).

The use of image processing techniques in forensic document examination is relatively new (Ellen, 1997). Image processing techniques offer significant cost benefits by eliminating or at least minimizing the need for expensive instruments and destructive testing methods. A majority of image processing techniques applied to forensic examination of documents thus far are quite primitive and usually imply some form of contrast enhancement. For the image processing community, the forensic domain opens a new area of research where new operations need to be defined and developed.

Of particular importance in document examination is identifying the writing instrument and the type of ink used in writing. Viscous ink (a thick pastelike material) is used in ball pens and liquid ink (water soluble) is used in fountain, gel and roller pens. In this paper, we study the problem of identifying different ink types using colour image processing techniques. The underlying principle is to detect colour

Haritha D. and Bhagvati C. (2007).
DISTINGUISHING LIQUID AND VISCOUS BLACK INKS USING RGB COLOUR SPACE.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IFP/IA*, pages 394-397

differences that correspond to the varying absorption and colour characteristics of different ink types. For example, a gel black ink appears more blacker than black viscous ink. In our previous analysis, we show that the HSV colour space is especially useful as absorption is revealed in the saturation values in bringing out the differences in blue and blueblack pens (Haritha, 2005). For black writings, hue is distributed over the entire range (0, 360°) and hence HSV colour space is not suitable. In this paper, we use RGB colour space to reveal the differences in colour characteristics of black inks.

The rest of the paper is organized as follows. Section 2 is a brief description of Multiple Linear regression and modeling black writings using MLR. Section 3 describes different distance measures in RGB and HSV colour spaces. Section 4 explains our methodology used for fitting the RGB data into a plane in cubical space and use of distance measures to distinguish different inks. Results are presented and discussed in Section 5 with conclusions in Section 6.

2 MODELING BLACK WRITINGS USING MLR

Multiple Linear Regression model (MLR) is the most commonly applied statistical technique for relating a dependent variable and a set of two or more independent variables (Siegel, 1996). Using MLR, the RGB data of black writings is fitted into the plane given by

$$b_0 + b_1 \times red + b_2 \times blue + b_3 \times green = 0$$

It is very interesting to know how much black is a black writing. To quantify the blackness of a black writing, the distance of the fitted plane of the data of a black writing to pure black point (0,0,0) i.e. the center of RGB cube is calculated. The distance is $\frac{b_0}{\sqrt{b_1^2+b_2^2+b_3^2}}$. It is observed that this distance of the fitted plane from the pure black (0,0,0) can be used to classify the viscous black inks versus liquid black inks. We know that the ball pens use viscous ink where as gel/roller pens use liquid ink. Several regression statistics are computed as functions of the sums-of-squares terms. The explanatory power of regression is summarized by its "$R - squared$" value, also called the coefficient of determination, is often described as the proportion of variance "described" by regression. If the regression is "perfect", R^2 is 1. If the regression is failure, R^2 equals zero.

3 DISTANCE MEASURES IN RGB AND HSV COLOUR SPACES

Identification of the particular ink/writing instrument used is the next step to classification in forensic examination of documents. We use distance measures to identify the particular ink/writing instrument used. Black colour can be treated as a dark shade of any colour. Two black colours can be distinguished by their mean R, G, B colour vectors. This section describes various distance measures existing for discriminating two colours and the two new derived distance measures namely WeightedL1 HSV distance measure and Geodesic HSV distance measure (Chakravarthy, 2005).

In Table 1 the expressions for RGB distance measures and new HSV distance measures between two colour vectors are given. The expressions for L1 HSV, L2 HSV, Canberra HSV, Cosine Angle HSV and Czekanowski HSV distances are similar to that of RGB distances except that r is to be replaced by h, g by s and b by v.

Table 1: Distance Measures in RGB and HSV colour spaces.

DISTANCE MEASURE	EXPRESSION FOR D(X,Y)
L1 RGB	$((\lvert r_1 - r_2\rvert) + (\lvert g_1 - g_2\rvert) + (\lvert b_1 - b_2\rvert))$
L2 RGB	$((r_1 - r_2)^2 + (g_1 - g_2)^2 + (b_1 - b_2)^2)^{\frac{1}{2}}$
Canberra RGB	$\frac{\lvert r_1-r_2\rvert}{r_1+r_2} + \frac{\lvert g_1-g_2\rvert}{g_1+g_2} + \frac{\lvert b_1-b_2\rvert}{b_1+b_2}$
Cosine Angle RGB	$\cos^{-1}(\frac{r_1r_2+g_1g_2+b_1b_2}{\sqrt{r_1^2+g_1^2+b_1^2}\sqrt{r_2^2+g_2^2+b_2^2}})$
Czeknowski RGB	$1 - \frac{2(min(r_1,r_2)+min(g_1,g_2)+min(b_1,b_2))}{r_1+r_2+g_1+g_2+b_1+b_2}$
Weighted Euclidean HSV	$\sqrt{(v_1 - v_2)^2 + s_1^2 + s_2^2 - 2s_1s_2cos(h_{diff})}$
Weighted L1 HSV	$S \times h_{diff} + \lvert s_2 - s_1\rvert + \lvert v_2 - v_1\rvert$
Geodesic HSV	$\sqrt{(s_2 - s_1)^2 + s_1^2(h_2 - h_1)^2 + (v_2 - v_1)^2}$

Statistical evaluation The validity of performance of the distance measures in identification of writing instrument is analyzed on statistical basis. We take the distance of feature vectors between writings of the same writing instrument and call it as a *within* writing instrument distance denoted by d_w. The *between* writing instrument distance d_b is obtained by measuring the distance between two different writing instruments.

d_w = d($f_{ij} - f_{ik}$) where i = 1 to n and j, k=1 to m.

d_b = d($f_{ij} - f_{kl}$) where i, k= 1 to n and j, l=1 to m.

where n is the number of writing instruments, m is the number of sample images written by each writing instrument, f_{ij} etc. are the feature vectors of the corresponding images, and d is the distance between two feature vectors of an image. Let n_w and n_b are the sizes of *within* and *between* writing instrument distance classes respectively. If n writing instruments provide m writings, there are $n_w = m_{C_2} \times n$ within writing instrument distance data and $n_b = m \times m \times n_{C_2}$. In our data collection we have taken 15 ball pens and 15 gel and roller pens. For each pen 15 images are taken.

n_w = 30 x 15_{C_2} = 3150 data.

n_b = 15 x 15 x 30_{C_2} = 97,875 data.

A good descriptive way to represent the relationship between two classes is calculating overlaps between two distributions. It can be done with two types of errors. Type I error occurs when the images of same writing instrument are identified as of different writing instruments. The type II error occurs when the images of different writing instruments is classified as of same writing instrument.

4 METHODOLOGY

Twenty Five black pens including ball,roller,gel and fountain pens with different manufactures were taken. A page containing 100 words was written by each of these pens on A4 size white xerox paper and from that page selected samples (20) were scanned at high optical resolution i.e,1200 dpi. Algorithm for classifying inks is comprising of the following steps.

1. Select suitable threshold values of R, G and B from the RGB histograms to seperate background and foreground pixels.
2. Fit the foreground data into a plane in the RGB cube.
3. Find the distance d from the fitted plane to the pure black point (0,0,0). Find R^2 coefficient of determination and MSE.
4. Classify liquid and Viscous Inks using the distance from the pure black (if d is $\geq$ 15, it is viscous ink otherwise it is liquid ink) .

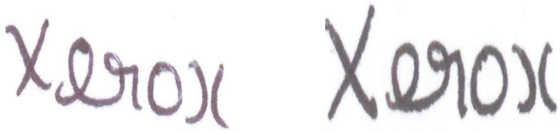

Figure 1: Writings of sample image of cello ball and Add roller pens.

Algorithm for identification of inks comprises the following steps.

1. Seperate background and foreground pixels.
2. Find the feature vector (mean colour) (r,g,b) of the foreground pixels. Find its equivalent h, s, v values. Find the distance d between the mean colour vectors of two images.
3. Label the distance as *within writing instrument distance*, if the two images belong to the same pen or as *between writing instrument distance*, if two images belong to different pens.
4. Repeat the above step for all the images of the database. Find the Type I and II errors from the distributions of *within* and *between* writing instrument distances.

5 RESULTS AND DISCUSSION

Figure 1 shows the sample images taken using Cello ball, Add roller pen. The data from the sample writings of different pens is fitted into a corresponding plane in the RGB cube. Figure 2 shows the fitted data of ball pen, actual data in to the corresponding planes in the RGB cubic space. The green coloured plane indicates the fitted plane of the data. The red coloured pixels indicate actual data of the scanned image. The blue coloured pixels indicated the estimated pixels. The coefficient of determination ”R-Squared ratio” in fitting the data of Cello ballpen image using regression is 0.999268 and MSE is 14.7248. The R^2 ratio in fitting the data of Add roller pen image using regression is 0.999586 and MSE is 7.6987. We can observe that R^2 ratio is closure to 1 indicating that regression is ”good”.

We have taken two datasets, one for training and analysis and another for testing purpose.The first dataset comprises of 15 ball pens and 10 roller/gel pens each of 20 sample images. The second dataset testset comprises of 10 ball pens and 10 gel/roller pens each of 10 sample images. The results were analysed using False Acception Ratio (FAR) and False Rejection Ratio (FRR). The FAR and FRR are calculated for classification of liquid inks and viscous

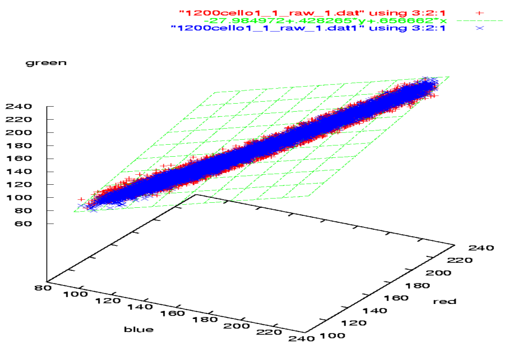

Figure 2: Data fitting of Cello ball pen using MLR.

inks based on distance of the fitted plane from pure black and is shown in Table 2.

Table 2: FAR and FRR errors for Ballpens and Gel/Roller pens training and test sets.

PEN TYPE	DATA SET	FAR %	FRR %
Ball pens	training	02.34	1.35
	test	1.87	1.09
Gel/Roller pens	training	01.35	02.34
	test	1.09	1.87

After classifying the inks, we must be able to identify the writing instrument. For each handwritten image, the mean vector of foreground pixels is obtained. The distance between these mean vectors of two images is calculated. For analysis the distributions of within the same writing instrument distance, between different writing instrument distances is plotted. We have taken 15 ball pens and 15 gel/roller pens and each pen with 15 images. Table 3 shows the TYPE I and TYPE II errors of identification. Type I error is minimum for our two new distance measures Weighted L1 HSV and Geodesic HSV distances. However, the overall accuracy is more for L1 HSV and L2 HSV distances.

6 CONCLUSIONS

In this paper, we have shown the use of RGB colour space in classification of black inks. An approach for fitting the black writings into a plane in RGB space, using MLR is given. This work is useful as the first step in identification of alterations in forged documents. The use of distance measures in identification of particular writing instrument is explored. This work can be further extended to distinguish gel versus roller pens and to classify printed documents. It can also be further extended to develop a system that can be used to classify and identify black writings or printings for forensic examination of documents.

Table 3: Type I and II errors for HSV and RGB distance measures.

DISTANCE MEASURE	TYPE I ERROR	TYPE II ERROR
L1 HSV	7.176	6.595
L2 HSV	6.82	08.33
Cosine HSV	12.16	11.94
Canberra HSV	14.04	7.15
CzeknowskiHSV	8.432	11.28
Euclidean HSV	10.608	10.067
WeightedL1 HSV	6.55	11.59
Geodesic HSV	5.46	16.767
L1 RGB	16.22	14.71
L2 RGB	12.16	15.27
Canberra RGB	9.929	21.47
Cosine RGB	16.802	17.39
CzeknowskiRGB	9.672	22.6

REFERENCES

Bauer B, 1966, The Use of Partially Transparent Plates to Increase the Contrast of images, *Applied Optical Journal*.

Brunelle R., 1982, Questioned Document Examination in Forensic Science Handbook, *Prentice Hall*.

Chakravarthy B., Haritha D., 2005, Distance Measures in RGB and HSV colour spaces, *ICATA*.

David Ellen, 1997, The Scientific Examination of Documents, *Taylor & Francis, 2nd Edition*.

Godown L., 1964, New Non Destructive Testing Methods, *Journal of Criminal Law, Criminology and Police Science*.

Hardcastle R., Hall M., 1978, A Technique for the Enhancement of the Infrared Luminescence of Inks, *Journal of Forensic Science Society*.

Haritha D., Chakravarthy B., 2005, Classification of Liquid and Viscous inks using HSV colour space, *ICDAR*.

Ordway Hilton, 1993, Scientific Examination of Questioned Documents, *CRC Press*.

Siegel F.A., Morgan C.J., 1996, Statistics and Data Analyis an Introduction, *2nd Edition, John Wiley & Sons, New York*.

Tappolet J.A, 1983, The High Performance Thin Layer Chromatography, its Application to the Examination of Writing Inks, *Forensic Science International*.