Classification of Liquid and Viscous Inks using HSV Colour Space

Chakravarthy Bhagvati Dasari Haritha
Dept. of Computer and Information Sciences
University of Hyderabad
Hyderabad India 500046
Email: chakcs@uohyd.ernet.in, harithadasari@rediff.com

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. Ink types Liquid ink or Viscous ink will be derived from the color properties of ink by extracting its HSV characteristics. This classification helps in distinguishing Gel and Roller pens versus Ball pens, Offset and Ink jet printers versus Laser printers. Different types of inks exhibit different absorption characteristics that causes color and distribution of color pixels to change. So we have done a detailed analysis of all color spaces and in particular HSV color space as it matches human vision. We observed that the saturation histograms of writings or printings reveals difference in absorption characteristics of ink, while hue histograms do not. We found that the saturation histograms can be modeled as Gaussian distribution that resembles process of diffusion of ink into paper. The measures like F ratio, σ ratio, Saturation weighted hue variance(ANOVA) are used to classify inks.

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. Obliteration refers to hiding a portion of the contents by applying white or correcting fluids, or by overstriking. Erasures refer to removal of a portion of the contents and may be done by physical or chemical means.

The conventional methods used by expert questioned document examiners are varied and sophisticated involving both destructive and non-destructive means. References [3, 4] and [9] offer excellent overviews. Microscopy, study of colour variations under changing illumination[9], filters for improving colour contrasts[2], ultraviolet fluorescence, infrared absorbance[5], infrared luminescence[6] and Fourier transform infrared spectrometry[3] 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[10] and high performance liquid chromatography (HPLC)[13].

Of particular importance in document examination is identifying the writing instrument and the type of ink used in writing. Printed documents are frequently involved in crime and it is important to identify method of printing. Printing inks are of three main types: solid ink used in pencils, typewriters, many photocopying machines and laser printers; viscous ink used in ball pens; and, liquid ink used in fountain, gel and roller pens, inkjet printers, offset printers and older photocopying machines. Most modern solid inks are in the form of fine powders which are fused into the paper by applying heat (this is the basic principle behind photocopiers and laser printers). Viscous ink is a thick pastelike material using organic chemical solvents rather than water. Mixture of dyes provide the coloring matter, and an important constituent is the resinous material which serves to bind the ink to the paper. Liquid inks by contrast are usually water soluble and do not have a paste-like consistency but are otherwise similar to viscous inks. Offset printing differs in that a plate that consists of *letter areas*, which absorb water, and others, which do not, is used for selectively applying liquid ink.

Printing inks may be distinguished chemically to some extent using mass spectroscopy and chemical analysis. The binding agents are subjected to Pyrolysis Mass Spectroscopy and infrared absorption spectroscopy to isolate that the inorganic components which are then identified by emission spectroscopy and microprobe electron microscopy[4]. In 1993 Lofgen[7] worked on HPLC analysis of printing inks as a non-destructive technique.

The use of image processing techniques in forensic document examination is relatively new[4]. Image processing techniques offer significant cost benefits by eliminating or

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

at least minimizing the need for expensive instruments and destructive testing methods. An examination of recent issues of the *Journal of Forensic Sciences* or *Forensic Science International* reveals that forensic experts resort to the use of general purpose software packages such as CorelDraw, Photoshop, etc. for analysis. A majority of techniques applied to forensic examination of documents thus far are quite primitive by image processing standards and usually imply some form of contrast enhancement. Thus, questioned document examiners can greatly benefit from the image processing community. For the image processing community, the forensic domain opens a new area of research where new operations need to be defined and developed.

In this paper, we study the problem of identifying different ink types using colour image processing techniques. The proposed method helps in automated ink type recognition quickly thus avoiding tedious human labor.In printing inks classfication, this method can perform better than human perception.The underlying principle is to detect colour differences that correspond to the varying absorption characteristics of different ink types. For example, a paper is more likely to absorb liquid than viscous inks. In our analysis, we show that the HSV colour space is especially useful as absorption is revealed in the saturation values. The HSV,HSI and HLS spaces differ slightly and their precise definitions are given in [11]. HSV space offers several advantages in representing colour for image processing. Colour, given by H and S, is decoupled from intensity and operations may be defined that manipulate colour independent of intensity and vice-versa. The better perceptual uniformity of HSV and its closeness to human description make it a good choice for forensic document examination where alterations are almost always done using similar coloured inks and differ subtly based on ink type, pen pressure and writing instrument characteristics.

The rest of the paper is organized as follows. Section 2 describes characteristics of different inks in HSV colour space. Section 3 presents our algorithm based on modeling saturation variations using normal distributions and the use of saturation weighted hue variance measure to classify inks. Results are presented and discussed in Section 4 with conclusions in Section 5.

2. HSV color characteristics of inks

Colour and intensity of the written contents do not remain constant but are affected by a number of factors such as pen pressure, friction between pen and paper, ink and paper types, the angle at which the pen tip touches the paper, make of the pen, etc. Printed documents are also affected by the rate of ink flow, ink and paper types, the printing process, and many other factors.

On experimentation, it was observed that the Hue, Saturation and Value or Intensity histograms are similar for the sample words written by the same writing instrument(pen or printer). We modeled the Saturation histograms as Gaussian distribution as follows.

$$f(s) = h \quad e^{\frac{-(s-\mu)^2}{2\sigma^2}}$$

where μ is the peak saturation(saturation with maximum number of pixels), σ is the variation in the saturation of color, h is no. of pixels at the peak saturation / total no.of pixels, s is the saturation, f(s) is the saturation histogram function.

The first order statistics like mean, peak, variance of the distribution can be used to classify inks but are not sufficient parameters. The actual deviations from the estimated model curve can be used to classify the inks. Using the ANOVA(analysis of variance) [12], We are using F-Ratio as another helpful characteristic in classification of inks. If S_1^2 and S_2^2 are the variances of independent random samples of size n_1 and n_2 respectively, assumed to have the same variance, then

$$F\ ratio \ = \frac{S_1^2}{S_2^2}$$

is a value of a random variable having the F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. Here when we modeled the saturation of the ink as gaussian distribution, the variance of the estimated gaussian curve and variance of the actual saturation are to be the same. Therefore F ratio can be difined as

$$F\ ratio \ = \frac{variance\ of\ the\ actual\ saturation\ histogram}{variance\ of\ the\ estimated\ gaussian\ curve}.$$

This measure is used to estimate the deviation of the histogram from the estimated model and is used to classify inks. Another measure we are defining is the σ ratio. We know that from the properties of gaussian curve, when s-μ is equal to 2σ, the height of distribution is equal to 0.135 of the height at the peak value and when $s - \mu$ is equal to σ, the height of distribution is equal to the 0.607 of the height at the peak. we are defining σ ratio as the ratio of σ value obtained from the lower portion of the actual saturation histogram to the σ value obtained from the upper portion of the actual saturation histogram. For good approximation σ ratio is 1. Another statistical measure Saturation weighted hue variance defined by Allan Hanbury [1] is used to classify inks. As hue is an angular value circular statistical descriptors should be calculated for it. The calculation of statistics based only on the hue, has the disadvantage of ignoring the close relationship between the chrominance coordinates (hue and saturation). For weakly saturated colors, hue is unimportant. As saturation is more, the perceiveness

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

of color is more. Given n pairs of values, the hue H_i and its associated saturation S_i, find the resultant of unit vectors, the vectors with direction H_i and length S_i. The hues associated with small saturation values have less influence on the direction of the resultant vector. This weighing is given by

$$A_s \;=\; \sum_{i=1}^{n} S_i \cos H_i; \;\; B_s \;=\; \sum_{i=1}^{n} S_i \sin H_i$$

Saturation weighted hue mean= $H_s \;= \arctan(\frac{B_s}{A_s})$.

Mean length of the resultant vector is $R_n = \frac{\sqrt{A_s^2+B_s^2}}{n}$.

Here the length of the resultant vector is compared to the length obtained if all the vectors had the same direction and maximum saturation. The value of R_n gives an indication of the saturations of the vectors which gave rise to the hue mean as well as an indication of the angular dispersion of the vectors. The saturation weighted hue variance is defined as V = $1 - R_n$. This variance indicates variation in hue and saturation together. This variance ranges from 0 to 1 and lower values of V represent less dispersed data.

3. Methodology

Each document containing 100 words is scanned at a high 1200 dpi resolution and split into 100 images each containing one word.The RGB histograms are found to be bimodal and a threshold value of 235 is used to seperate background white pixels from the foreground or ink pixels.The images are converted into HSV by using the transformations in Equations given in [11]. Saturation values are scaled to the range 0–255 and all the pixels whose saturation is greater than 25 are taken as foreground or ink pixels.It was observed that the pixels having saturation less than 25 are the pixels that are spread out from the strokes. Such a fixed thresholding scheme works well as the colour of the paper used in testing is white. Find μ,the saturation value at which the number of pixels having Saturation as maximum and also h, the normalized number of pixels at the peak saturation.Find σthat represents variance in saturation histogram from the points at which the height of the distribution is greater than or equal to the 0.607 of the height at the peak.Model the Saturation histogram as Gaussian distribution using the equation given below.

$$f(s) \;=\; h \quad e^{\frac{-(s-\mu)^2}{2\sigma^2}}$$

Find F ratio, σ ratio and Saturation weighted hue variance.

4. Results and Discussion

Figure 1 shows the sample writings taken using Ball, Gel and Roller pen. Figure 2, 3 and 4 shows the saturation histograms and corresponding modeled Gaussian curves of the Ball, Gel and Roller pens respectively. It is observed that in case of liquid inks(Gel and Roller pens), the actual saturation histograms deviate more from the modeled Gaussian curve in the lower portion of the histogram. We know that color of writing depends on reflectance and absorbance properties of ink and paper. For liquid inks, absorbance of ink into paper is more when compared with viscous inks. Because of this the color of the ink gets desaturated. These desaturated pixels deviate more from the peak saturation value. Hence the saturation histogram of liquid ink writings deviate more from the Gaussian curve. Here these deviations from the estimated model can be figured out by using the measures F ratio and σ ratio. we observed that for viscous inks the F ratio is less than 2 and for liquid inks F ratio is more than 2. The reason for threshold of 2 can be explained from our initial experiments. While modeling we are keeping the height at the peak saturation of the saturation histogram and estimated curve equal. we know that Gaussian function can be written as

$$f(s) \;=\; h \quad e^{\frac{-(s-\mu)^2}{2\sigma^2}} = area \quad \frac{1}{\sqrt{2\pi}\sigma}\, e^{\frac{-(s-\mu)^2}{2\sigma^2}}$$

If f(p) is the function that represents saturation histogram and f(t) is the estimated gaussian function, f(t) and f(p) can be written as

$$f(p) \;= area_p \quad \frac{1}{\sqrt{2\pi}\sigma_p}\, e^{\frac{-(s-\mu)^2}{2\sigma_p^2}}$$

$$f(t) \;= area_t \quad \frac{1}{\sqrt{2\pi}\sigma_t}\, e^{\frac{-(s-\mu)^2}{2\sigma_t^2}}$$

we know that $area_p$ the area under the histogram is 1 and area is inversly proportional to σ. We observed that the gaussian curve for liquid inks occupies at the maximum of $\frac{1}{\sqrt{2}}$ area of the histogram. Therefore,

$$F\,ratio\,threshold \;=\; \frac{\sigma_p^2}{\sigma_t^2} \;=\; \frac{area_p^2}{area_t^2} \;=\; (\frac{\sqrt{2}}{1})^2 \;=\; 2$$

The widened lower portion of the saturation histogram results in lesser σ ratio for liquid inks (< 0.707) and greater σ ratio for viscous inks(> 0.707). The threshold may be derived using the above procedure in more general cases also.

Ball point inks are only partly absorbed into the paper, they have a characteristic glossy appearance. A ball at the end of a tube picks up ink from the reservoir above it and transfers it to the paper surface. In many cases striations caused by imperfect or dirty ball housings are apparent, and there is a tendency, just after the pen has turned a corner, for an extra amount of ink to be deposited in the line. This leads to more non uniformity in the color perception of ball pen writings. So our intuition is that there must be large

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

variation in the chrominance of color of ballpen writings. Instead of measuring hue variance and saturation variance seperately, we can think of measuring saturation weighted hue variance to be more helpful in bringing out the above factor. In contast, a liquid ink or water based ink, will color the paper by being absorbed in it. The ink itself will not be visible as a layer of added material but rather as colored area of an evenly textured surface. It results in lesser variation in perception of its color. Instead of measuring hue variance and saturation variance seperately, we can think of measuring saturation weighted hue variance to be more helpful in bringing out the above factor.The saturation weighted hue variance is also helpful in distingushing one pen from the other i.e.,this measure is of same value for all the writings of a pen.We observed for Gel and Roller pens F ratio greater than 2, σ ratio less than 0.707 and saturation weighted hue variance less than 0.80 and for ball pens the values are F ratio less than 2,σ ratio greater than 0.707and saturation weighted hue variance greter than 0.80.

Figure 7 shows paper absorbance and roughness effect on Gel/Roller and Ball pen writings. Black and blue colored curves are the saturation histograms of the writings of the same ball pen on A4 white xerox sheet and on rough paper. Because of the roughness of the fibers of the paper, there is slight desaturation in all pixels, but the effect is very less. Light Blue and Green colored curves are the histograms of the same Gel pen on A4 white xerox sheet and on rough paper. We can observe more deviation in the lower portion of the histogram in case of rough paper as its absorbance is more. This result strengthens our reasoning for difference in absorbance reveals changes in saturation histograms.

Figure 5 shows sample printings taken from the Ink jet, Laser and Offset printers. Laser and Inkjet printigs are taken using the same file(same color settings). Figure 6 shows the saturation histograms and corresponding modeled gaussian curves for Laser, Ink jet and Offset printings. We can observe that for Laser printing characteristics resembles with that of the viscous ink. It uses powdered toner material that adhers to paper under high temperature but not absorbed into paper. Ink jet and Offset printers use ink that gets more absorbed into paper and hence resembles liquid ink characteristics explained above. If F ratio greater than 2, σ ratio less than 0.707 and saturation weighted hue variance less than 0.80 the printing can be either Ink jet or Offset otherwise it is Laser printing. This gives primary idea, further we have to extend the work to distinguish Ink jet versus Offset printing, Gel pen versus Roller pen.

5. Summary and Conclusions

In this paper, we have shown that the HSV colour space that match human perception is more useful in Forensic Document Examination.Saturation histograms are more appro-

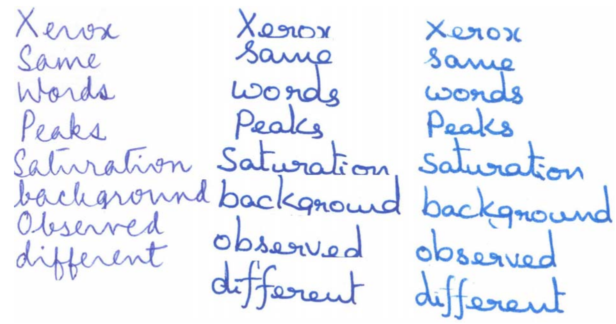

Figure 1: Sample Writings of Ball, Gel and Roller Pens.

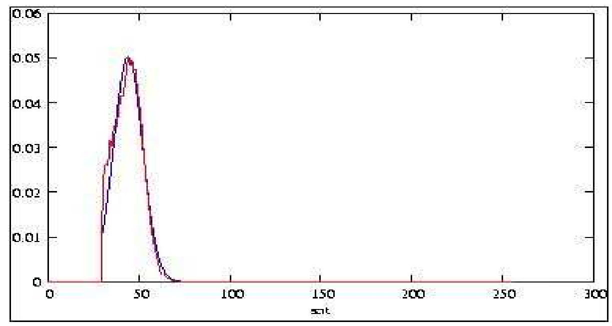

Figure 2: Saturation histogram and Modeled Gaussian Curves of Ball Pen.

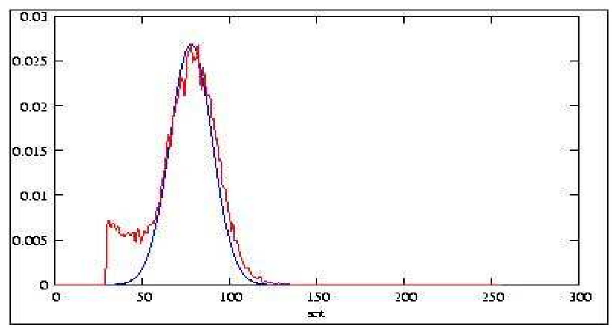

Figure 3: Saturation histogram and Modeled Gaussian curve of Gel Pen.

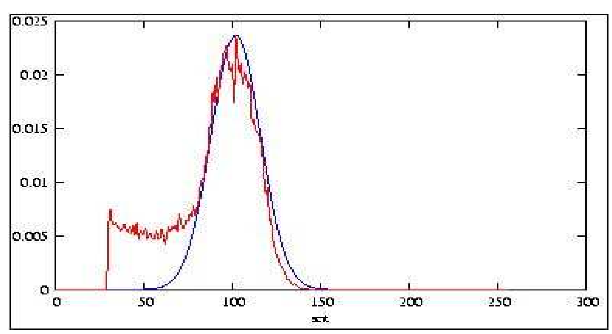

Figure 4: Saturation histogram and Modeled Gaussian curve of Roller Pen.

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

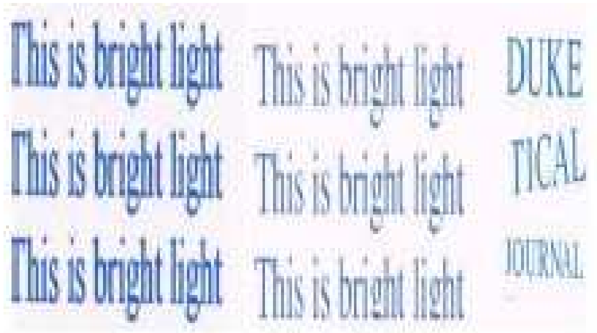

Figure 5: Sample Writings of Inkjet, Laser and Offset printers.

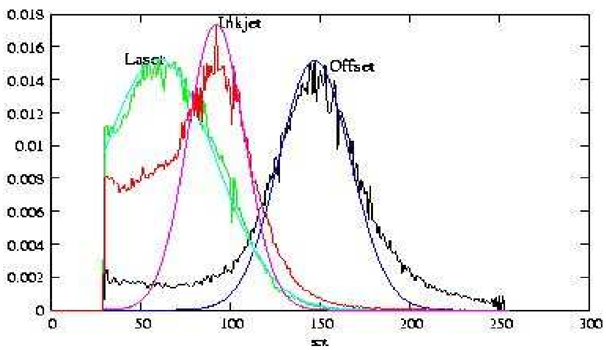

Figure 6: Saturation histograms and Modeled Gaussian curves for Laser, Inkjet and Offset printings

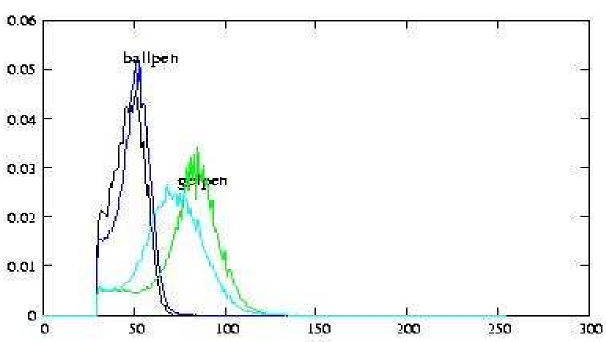

Figure 7: Effect of Paper Absorbance on Viscous and Liquid Ink Saturation Histograms.

priate to deal with, as they reveals changes in absorption characterisics of inks.We modeled Saturation Histograms as Gaussian distribution.Different statistical measures are used to classify Liquid inks versus Viscous inks.This classification is useful in identification of alterations in forged documents.This work can be further extended to distinguish Gel versus Roller pens and Inkjet versus Offset printers.

References

[1] Allan Hanbury and Jean Serra, "Circular Statistics Applied to Color Images," *Computer Vision-CVWW'03*, 2003.

[2] Bauer B, "The Use of Partially Transparent Plates to Increase the Contrast of images," *Applied Optical Journal*,Vol. 5,pp.1361-1364,1966.

[3] Brunelle R, "Questioned Document Examination in Forensic Science Handbook," *Prentice Hall*,pp.710,1982.

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

[5] Godown L, "New Non Destructive Testing Methods," *Journal of Criminal Law, Criminology and Police Science*,pp280,1964.

[6] Hardcastle R, Hall M, "A Technique for the Enhancement of the Infrared Luminescence of Inks," *Journal of Forensic Science Society*,Vol .18,pp.53-55,1978.

[7] Lofgen B, Andrasko J, "HPLC Analysis of Printing Inks," *Journal of Forensic Sciences*,Vol .38,pp.1151-1154,1993.

[8] Luong Q.T., "Color in Computer Vision," *Handbook of Pattern Recognition and Computer Vision*, pp. 311-368, 1993.

[9] Ordway Hilton, "Scientific Examination of Questioned Documents," *CRC Press*,1993.

[10] Schweppe H, "Thin Layer Chromatography," *Wiley International, Wiley Interscience, New York*,1977.

[11] Smith A.R., "Color Gamut Transform Pairs," *Computer Graphics*,Vol .12(3), pp. 12-19, 1978.

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

[13] Tappolet J.A, "The High Performance Thin Layer Chromatography,its Application to the Examination of Writing Inks," *Forensic Science International*,Vol .22,pp.99-109, 1983.

[14] Von Bremen U, "Invisible Ultraviolet Flourescence," *Journal of Forensic Science*,p368,1965.

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)