FORENSIC SCIENCE JOURNAL SINCE 2002

Forensic Science Journal
2014;13(1):15-22
Available online at:fsjournal.cpu.edu.tw

Investigating Between-Model Variation of the PaperMate® Pen Using Infrared Spectrum and Multivariate Statistical Techniques

Loong Chuen Lee [1]* , M.Sc.

[1] Forensic Science Program, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 Kuala Lumpur, Malaysia

Received: February 02, 2014 / Accepted: June 27, 2014

Abstract

Pen ink analysis is an important area of forensic document examination field. In this study, the feasibility of using micro-ATR-FTIR spectroscopy to discriminate different models of pen within a particular brand was investigated. A total of 37 black ballpoint pens (including replicates) that represented nine models of pen from Papermate® brand were sampled from the Mydin shopping mall at Subang Jaya, Malaysia. Organic information of pen inks was collected with micro-ATR-FTIR. Preliminary examination on IR spectra revealed slight differences between pen models. Subsequently, the spectral data were analyzed with multivariate statistical techniques using SPSS software. Results showed that the proposed approach could be a powerful tool for differentiating pen models within the Papermate® brand.

Keywords: forensic science; inks analysis; multivariate statistical techniques; FTIR spectroscopy; within-model variation

Introduction

The analysis of ink, especially ballpoint pen inks, is an important aspect in forensic documents examinations. One of the main focuses of a forensic questioned-document examiner is to identify a suspicious ink entry. Pen ink can be discriminated based on the organic and inorganic composition. Most of the time, different analytical techniques would be applied to assist in identification of the source of questioned ink. Identification refers to the process of determination of the brand, model, manufacturer and year of production of the instrument(s) that could have produced the ink specimen [1].

Many instrumental techniques are often available for the analysis of inks. They can be of destructive or non-destructive in nature. Destructive techniques included chromatographic and mass spectrometry techniques that used to give better results in term of resolution [2-4]. But they required extraction of ink sample and thus disturbed the integrity of evidence. During court procedures, integrity of evidence is the main issue to be discussed. On the other hand, non-destructive techniques such as RAMAN and FTIR spectroscopy usually produce little information about inks as well as the interpretation of the spectrum tended to be subjective as it depends on human being eye to do the evaluation. Until now none of these is wholly satisfactory to provide an efficient and effective analytical method for identification of inks that are precise and accurate while preserving the integrity of ink samples.

The feasibility of micro-ATR-FTIR spectroscopy coupled with multivariate analysis has been studied for classification and identification of black ballpoint pen ink at brand [5] and model [6] levels, separately. This study aimed predominantly at examining part of the samples from [5, 6] in further detail where the main focus was on investigating within-pen-model variation by analyzing nine different models of Papermate® pens.

* Corresponding author: lc_lee@ukm.edu.my

Experimental

Sampling

A total of 37 Papermate® black ballpoint pens from nine different models were sampled from the Mydin shopping mall at Subang Jaya, Malaysia. All of these were obtained in multiple packs of the same product to ensure all of them were from the same batch of production as the variation between batches will not be considered in this study. All pen models were allocated a reference number for the purpose of this study. Detailed descriptions of collected samples used in this study are presented in Table 1. Double ATM, from Thailand3 was the white paper used throughout the study. Each of the 37 pens was used to draw three different small circles to ensure complete coverage of paper by the pens. Experimental work was carried out on the ink deposited on the white paper not later than one day after drawing to minimize changes due to ink aging.

Micro-ATR-FTIR Analysis

All experimental spectroscopy was carried out on a Thermo Scientific Nicolet iN10 MX FT-IR microscope with mercury cadmium telluride (MCT) detector. A Ge crystal tip ATR objective was used as micro-sampling accessory. The background spectrum was reacquired after every analysis to reduce variation in the spectra due to instrumental drift. Each spectrum was the result of an average 16 scans at 4 cm-1 resolution over a spectral range of 4000 to 675 wavenumbers (cm-1). Three spectrums were collected from each of 37 pens. The IR spectral data was stored in a data matrix in Microsoft Excel spreadsheet.

Table 1 Sample ID for nine different models of studied Papermate® black ballpoint pens.

Sample ID	Model	Quantity
25	KM 100 Med	4
26	KM 100 Fine	4
27	KM 100 S/F	4
28	KV 2 Med	4
29	KV 2 Fine	4
30	KV 2 S/F	4
31	KM Med PT	4
32	KM Fine PT	5
33	KM F	4

Software

Data collected was processed and analyzed using statistical package SPSS (Statistical Package for the Social Sciences, Windows version 12.0, SPSS Inc., Chicago, USA).

Statistical Pretreatment

Normalization of selected IR region

All three spectra of each sample were included to assess the material reproducibility and homogeneity while ensuring representativeness of each sample due to

the minimal size of inks contacted by ATR Ge tip and the potential heterogeneity of the inks [7]. Triplicate analysis of thirty-seven pens gave 111 spectra. As such, full data set comprised 111 spectra, consisting of absorbance values as a function of wavenumber, were normalized prior to carrying out the multivariate analysis to reduce variation in the data due to different thickness of the pen inks deposited on paper samples. The absorbance values for each spectrum were divided by the total absorbance resulting from each spectrum across wavenumbers 2000-675 cm^{-1}.

Transformation of raw data to new set of variables

The selected IR region (2000-675 cm-1) was subjected to three different types of data-reduction techniques to produce new set of variables. The first approach is simply choosing few discrete representative variables without involving PCA. Absorbance values of wavenumbers that make up the three characteristic IR peaks of crystal violet: 1171, 1173, 1174, 1176, 1178, 1360, 1362, 1363, 1367, 1581, 1583, 1587 and 1589 cm-1 were selected as a new set of variables. This approach labeled as "discrete data-DA". The second approach named " PCA-DA" involved the decomposition of the whole 688 data by correlation PCA. The number of components to be extracted was decided based on the Kaiser criterion [8]. The resulting principal components act as another new set of variables. The third feature reduction technique which labeled as "ZV-PCA-DA" was achieved by conducting cluster analysis and principal component analysis sequentially. Initially, cluster analysis was performed on the standardized raw variables (688 data points) over the objects (ink samples) by using Ward's method and squared Euclidean distance to form eight different clusters. Each cluster is expected to contain variables that carry similar information about the objects. Subsequently, correlation PCA was performed on each of the eight clusters, separately. The number of components to be extracted was decided based on the Kaiser criterion [8]. The new set of variables comprised all selected principal components that were labeled accordingly, e.g. PCA 1_2 refers to the first component produced by PCA conducted on the second cluster of data points.

Discriminant analysis

After the feature reduction step, stepwise DA was conducted on each of the three new sets of variables to see whether all 111 IR spectral could be discriminated according to their pen models (category variables). DA was used with leave-one-out classification in the training set. Leave-one-out is a type of cross-validation techniques that enables all of the available data to be utilized for training while still giving an unbiased estimate of the generalization capabilities of the resulting classifier [9]. Stepwise DA was performed by using the Mahalanobis distance method and probability-F as the criterion. Prior probabilities were equal for all groups (models) as the representativeness of population was unknown [10].

Results and Discussion

The black ballpoint pen inks used in this study were actually part of the samples in [5, 6] which has been evaluated systematically at brand and then at model levels. The study presented here is linked to [5, 6] but was different in term of the level of variations that being focused on, i.e. to evaluate the variations within a single manufacturer (brand).

The identification of pen inks is one of the most difficult tasks in forensic document examination because the variability associated with the production of pen inks. Even within a single brand/manufacturer, two models may have different composition of inks in quality and also quantity. Most studies found across literatures draw only one sample to represent a particular group and thus only concerned with variation between-model/brand. This study has been focused on the variation within-brand or between-model of black ballpoint Papermate® pen. Nine models of pen from the brand of Papermate® were drawn to encounter all possible variations within a brand (intra-brand variability).

In this study, only the regions from 2000-675 cm-1 were analyzed because the IR absorption peaks at the region between 4000 and 2000 cm-1 were mainly due to water vapor and carbon dioxide from the atmosphere and only a few weak and broad bands attributable to ink [2]. Representative IR spectra of nine different models of black ballpoint Papermate® pen inks analyzed by micro-ATR-FTIR spectroscopy in the region of 2000-675 cm-1 are shown in Fig. 1.

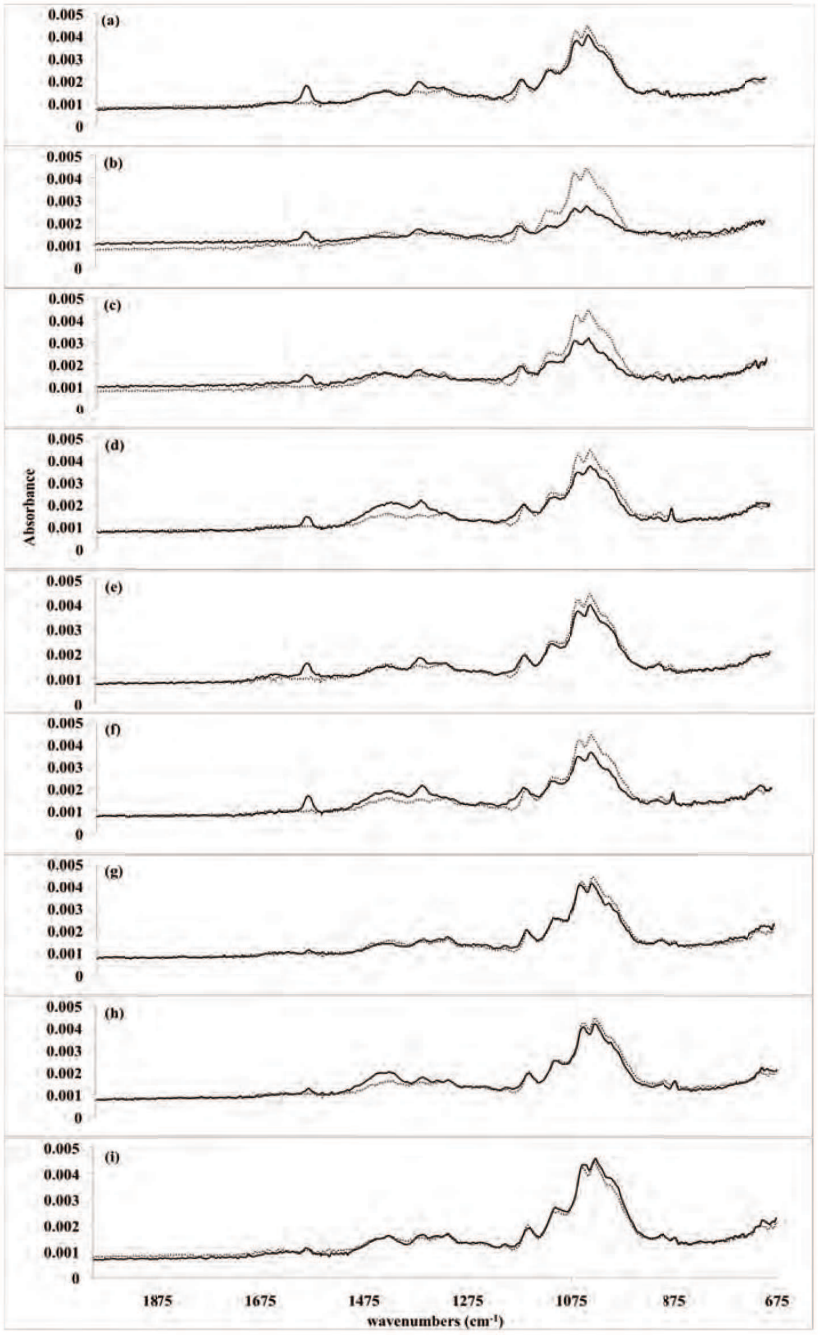

Fig. 1 The respective spectra of nine different models of black ballpoint Papermate® pen inks (-) and blank paper (….) analyzed by micro-ATR-FTIR spectroscopy in the region of 2000-675 cm-1: (a) KM 100 med; (b) KM 100 fine; (c) KM 100S/F; (d) KV2 med; (e) KV2 fine; (f) KV2 S/F; (g) KM med PT; (h) KM fine PT; (i) KM F

In general, all spectra were showing similar spectrum patterns in which two prominent peaks were observed at approximately 1584 and 1360 cm-1 but with differences in the form of relative peak height and its shape (Fig. 1). Model 31, 32 and 33 gave lowest absorbance value at 1584 cm-1. While model 30 gave the highest absorbance value at 1360 cm-1. In fact, both of the peaks are characteristic peaks of ballpoint pen inks, and have previously been assigned to a skeletal vibration of triarylmethane dye and the C=C stretch vibration of epoxy resin (about 1581 cm-1)[11]. The region between wavenumbers 1100-1000 cm-1 is of less concern as it contains peaks mostly from paper substrate [12]. The vibrational assignment for other peaks has been explained elsewhere [11]. As such, the quantity of crystal violet may play an important role in the differentiation of IR spectra from different models.

Multivariate analysis such as cluster analysis (CA), principal component analysis (PCA) and discriminant analysis (DA) have been used for analyzing spectrum of forensic samples included soil [6], document paper [13], ivory [7] and ballpoint pen inks [2,14]. If DA was used to classify or discriminate the sample, number of data points to be used must be reduced prior to DA. This is an important step as discriminant analysis (DA) is restrictive with regard to the number of predictor variables versus the number of samples. In addition, a smaller number of input variables also means results can be obtained with reduced computational and economical expense as well as utilization times while improving the accuracy and efficiency of the classification tasks [9]. Different feature selection/extraction methods have been applied onto IR or UV-visible spectra data prior to DA. Brody et al. [7] used eleven integrated areas calculated from Raman spectrum as input variables for DA. Thanasoulias et al. [14] used K-means cluster analysis to assist in selection of few raw variables that will then be further reduced by PCA. On the other hand, Kher et al. [2,13] decomposed the complete spectrum by PCA without any prior selection. The resulting PCA scores were used as DA variables. In addition, Kher et al. [2] also suggested selecting discrete variables based on the ratio of the between-to-within-variability. While in this study, the selected IR region (2000-675 cm-1) composed of 688 raw data. Three different data extraction techniques were proposed to transform the normalized absorbance values into a new set of variables.

The use of selected discrete variables (the first approach) resulted in the lowest correct classification (24.30 % correct classification) while the decomposition of the whole spectrum by once PCA (the second approach) resulted in 35.10 % correct classification, higher than the first approach. A total of 47.70 % of ink samples were correctly classified when the same dataset was analyzed by several PCA according to eight different clusters formed from it. Therefore, the ZV-PCA-DA approach (third approach) was selected as the best approach. There are a few advantages behind the third approach. Firstly, the newly formed components will be more informative than a single wavenumber as principal component is a linear combination of several wavenumbers. PCA conducted on each of the clusters of data points of IR spectrum grouped according to similarity is expected to be more effective in extracting important information from IR spectrum than PCA conducted on whole data points of spectrum. This also helps in reducing the risk of losing important information. In addition, PCA running on a smaller number of data points would result in a more stable matrix than those running on whole 688 data points all at once. On the other hand, PCA scores are proven to be more effective than raw variables for DA [14]. In addition, by using extracted components as new latent variables, the problem of redundancy is eliminated as the components are orthogonal and uncorrelated.

The ZV-PCA-DA approach reduced 688 data points to 37 principal components. Stepwise DA selected 13 components from 37 components as predictor variables. Fig. 2 shows the projections of the 111 spectra into the space of the first two discriminant functions, displaying 85.70% of the between-to-within group variation in the data. Based on Fig. 2, discriminant function 1 was found to be responsible for separation of models 31, 32 and 33 from the others while models 25, 26 and 27 were separated from the others by discriminant function 2. As such, all the pen models may be categorised into three groups according to the LDF 1-2 scores. The first, second and third clusters were composed of model 25, 26 and 27; model 28, 29 and 30; model 31, 32 and 33, respectively. In fact, each cluster composed of models of the same model name except their point size, i.e. Med, Fine. This indicates that pens of same model name except point size do contain highly similar or exactly same composition of inks. In another words, approximately

100% correct classification would be expected if DA was used to classify the nine models according to their model without taking into account of the point size, i.e., KM100, KV2 and KM.

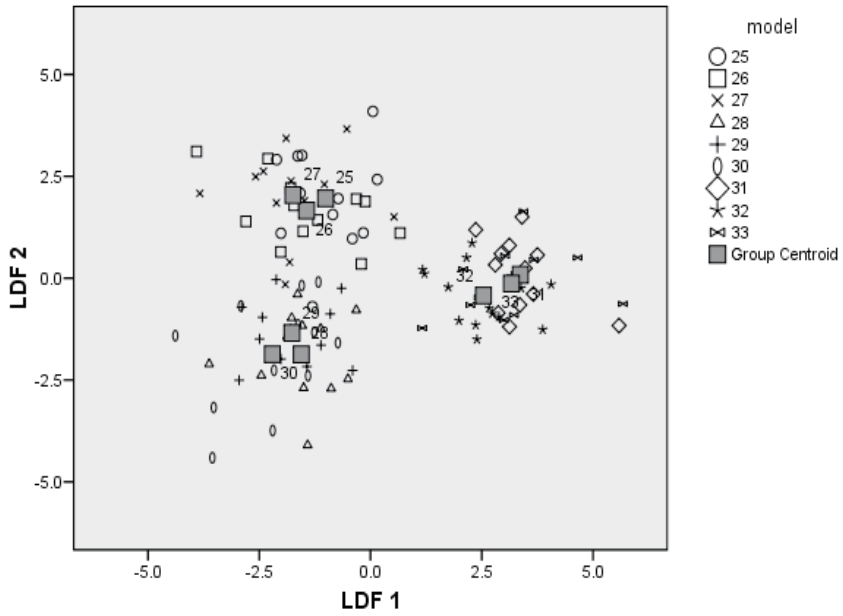

Fig. 2 Score plot of first two linear discriminant functions showing the separation of group centroids for nine different models of Papermate® pen inks.

Conclusions

The results of this study suggest that the proposed method can be of value in identifying different models of pen within the brand of Papermate®, non-destructively. Basically, for Papermate® pen, models differed only point size seem to contain similar inks The method can be further tested on more pen samples of different models as well as brands.

Acknowledgements

We would like to thank Mr.Fizol and Mrs. Rusikah bt. Minhad from Jabatan Kimia Malaysia for the assistance in the acquisition of infrared data. Many thanks also to all the staff of the forensic science program from UKM as well as Mrs. Hartinah Annuar.

References

1. Neumann C, Ramotowski R, Genessay T. Forensic examination of ink by high-performance thin layer chromatography--the United States Secret Service Digital Ink Library. J. Chromatogr. A. 2011; 1218: 2793-811.
2. Kher A, Mulholland M, Green E, Reedy B. Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis. Vib. Spectrosc. 2006; 40: 270-77.
3. Denman JA, Skinner WM, Kirkbride KP, Kempson IM. Organic and inorganic discrimination of ballpoint pen inks by TOF-SIMS and multivariate statistics. Appl. Surf. Sci. 2010; 256: 2155-63.
4. Gallidabino M, Weyermann C, Marquis R. Differentiation of blue ballpoint pen inks by positive and negative mode LDI-MS. Forensic Sci. Int. 2011;

204: 169-177.

5. Lee LC, Mohamed RO, Pua H, Systematic Assessment of attenuated total reflectance-Fourier transforms infrared spectroscopy coupled with multivariate analysis for forensic analysis of black ballpoint pen inks. Malaysian Journal of Analytical Sciences. 2012; 16: 262-272
6. Lee LC, Mohamed RO, Pua H, Ab AI, Seri MI, Classification and identification of black ballpoint pen inks based on multivariate analysis and infrared spectra. Problems of Forensic Sciences. 2012; 92: 253-264
7. Brody RH, Edwards HGM, Pollard AM. Chemometric method applied to the differentiation of Fourier-transform Raman spectra of ivories. Anal. Chim. Acta. 2001; 427: 223-32.
8. Kaiser HF. The application of electronic computers to factor analysis. Edu. Psychol. Meas. 1960; 20: 141-151.
9. Mallet Y, Coomans D, de Vel O. Recent developments in discriminant analysis on high dimensional spectral data. Chemom. Intell. Lab. Syst. 1996; 35: 157-173.
10. Chan YH. Biostatistics 303: Discriminant analysis. Singapore Med. J. 2005; 46: 54-61.
11. Wang J, Lou G, Sun S, Wang Z, Wang Y. Systematic analysis of bulk blue ballpoint pen ink by FTIR spectrometry. J. Forensic Sci. 2001; 46: 1093-1097.
12. Causin V, Marega C, Marigo A, Casamassima R, Peluso G, Ripani L. Forensic differentiation of paper by X-ray diffraction and infrared spectroscopy. Forensic Sci. Int. 2010; 197: 70-74.
13. Kher A; Mulholland M; Reedy B; Maynard P. Classification of Document Papers by Infrared Spectroscopy and Multivariate Statistical Techniques. Appl. Spectrosc., 2001, 55, 1192-1198.
14. Thanasoulias NC; Parisis NA; Evmiridis NP. Multivariate chemometrics for the forensic discrimination of blue ball-point pen inks based on their Vis spectra. Forensic Sci. Int., 2003, 138, 75-84.

中央警察大學
誠
CENTRAL POLICE UNIVERSITY