Forensic Image Analysis of Laser-Printed Documents

Williams D. Mazzella, MS and Raymond Marquis, MS

Ecole des Sciences Criminelles, Institut de Police Scientifique, University of Lausanne, Switzerland

A preliminary study was undertaken to determine if objective measurements, using a proprietary image analysis system, may be routinely applied for the analysis of electrophotographic-printed documents, in particular laser-printed documents. Text and dot-quality objective measurements were studied to differentiate printed outputs. This paper highlights the complementarity of forensic image analysis to classical methods employed for the analysis of laser-printed documents. Some practical tests were prepared and discussed. Discriminant analysis correctly classified the samples from the tests, thus demonstrating that forensic imaging analysis is realistically applicable in closed-set cases (where the number of potential laser printers can be clearly defined) and in page-substitution cases.

Introduction

Due to the extensive accessibility of business machines over the past 30 years, forensic document examiners (FDEs) are often requested to analyze documents printed by an electrophotographic process using a dry toner (e.g., laser printers and photocopiers). FDEs routinely examine laser-printed documents using microscopy, as well as analyzing organic and inorganic components to determine if 2 or more documents were printed by the same laser printer or the same cartridge unit. The forensic literature describing toner analysis or electrophotographic printing system identification is extensive (Merrill and others 1996, Trzcinska 2006). Tse, Forrest, and She (1995) were able to show that character width is strongly correlated to electrophotographic properties of the OPC drums and to the type of toner used. An efficient method based on image analysis and quantitative measurements was explored. In addition, complementary techniques that include the use of an optoelectronics camera and Fourier transform infrared spectroscopy (FTIR).

Methods and Materials

Sampling and Tests Preparation

Three blind tests were conducted and will be referred to as Tests A, B, and C. Test A, consisting of a 5-page contract, was given as an exercise during the 63rd American Society of Questioned Document Examiners Conference in Montréal, Quebec, Canada (2005). The aim of the task was to determine if any of the pages were inserted into the contract. In addition to this test, the 1st page of the same general contract was printed on A4 Xerox® Business paper (80g/m^2) using 13 laser printers obtained from the Institut de Police Scientifique (Table 1). Three exemplars were printed on the same day from each of the 13 laser printers using Microsoft® Word 2002 with Times New Roman 12 font. Thus, a total of 39 pages were printed for additional Tests B and C.

For Test B, a member of the Institut de Police Scientifique was asked to prepare a set of 6 pages from among the 39 total pages printed. The purpose was to establish if the 6 pages were printed by the same laser printer.

For Test C, the first 2 pages printed by each laser printer were used to establish a closed set database. Again, a member of the Institut de Police Scientifique was asked to prepare a blind test using 3 randomly chosen pages that were not included in the database. The goal was to determine which laser printer, out of the 13 printers available, was used to print the 3 questioned pages.

Optoelectronic Systems

Single-component toner powders contain magnetic material which is incorporated into the toner particles when affixed to paper. The magnetic properties were detected using a Mag optoelectronic system (Vildis, Russia) and can be used to differentiate magnetic single-component toners from dual-component toners (Neumann and Mazzella 2004).

IPS code	Brand and Model	Single or dual component toner	Type of Binder
1	Canon IR400	Single	1
2	HP Laserjet 4L	Single	2
3	HP Laserjet 1300	Single	3
4	IBM Infoprint C 1220	Dual	4
5	Lexmark Optra L Plus	Dual	4
6	Lexmark Optra T630	Dual	4
7	Lexmark Optra T630	Dual	4
8	Lexmark Optra T632	Dual	4
9	Lexmark Optra T632	Dual	4
10	Lexmark Optra T632	Dual	4
11	Lexmark Optra T632	Dual	4
12	Lexmark Optra T634	Dual	4
13	Lexmark C510	Dual	4

Table 1. List of the laser printers used for Tests B and C.

IAS Settings	Text Analysis	Dot Analysis
Color	Gray	Gray
Density Standard	ANSI A	ANSI A
Sample	Print	Print
Stochastic (FM) Screen	off	off
Measurement	text	dot
Dark or Light Text or Dot	Light dot	Dark dot
Minimum diameter	50	60
Maximum diameter	5000	5000
Screen Ruling	cm	cm

Table 2. Personal IAS settings for the text (letters "a" and "e") and dot analysis.

Stereomicroscopic Observation

All printed documents for Tests A, B, and C were examined using a stereomicroscope for general observation, font type, and potential print defects.

Fourier Transform Infrared Spectroscopy

Analyses of toners were performed by microscopic Attenuated Total Reflectance (ATR) with an Internal Reflection Element (Germanium crystal) using a Digilab® Excalibur spectrometer (Canton, USA). FTIR spectra were acquired from 4000 to 650 cm^{-1} with a resolution of 4 cm^{-1}, using 64 scans that were co-added for each spectrum.

Image Analysis

Image analyses were carried out using the Personal IAS™ made by Quality Engineering Associate Inc. (QEA®), Burlington, USA. The Personal IAS™ is a full-featured pocket PC coupled within a color CCD camera that objectively evaluates the printing material by measuring the quality of basic print attributes such as area, dot, lines, and text. The instrument is calibrated in the factory for dimensional accuracy (± 2 µm) and has a resolution of 5 µm /pixel (QEA 2004). Measurements were carried out on the internal loop of letters "a" and "e" and on the final dot (.), and the Personal

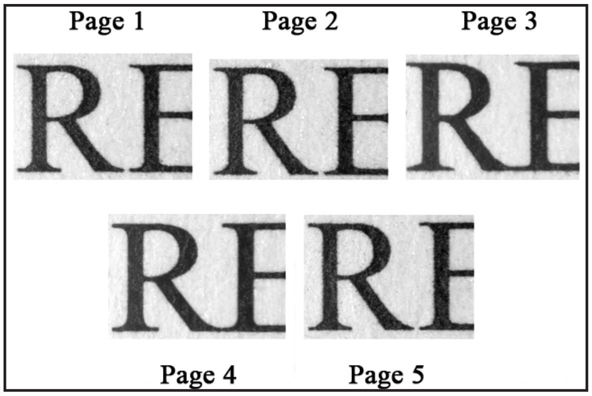

Figure 1. Illustration of connection between letters "R" and "E" on pages 3 and 4 that is not observed on pages 1, 2, and 5 of test A.

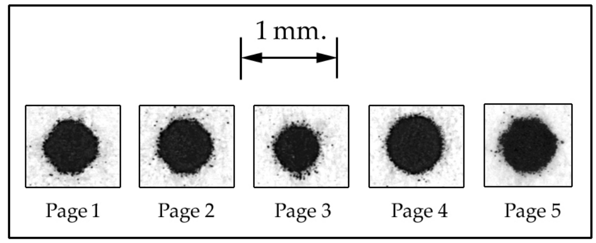

Figure 2. Illustration of size and shape of dot "." on the 5 pages of Test A: it is the smallest on page 3, the biggest on page 4, and medium and of similar shape on pages 1, 2, and 5.

IAS™ settings for text ("a" and "e") and dot analysis were set (Table 2).

For each selected feature, 10 measurements were carried out on each page. Although the Personal IAS™ provides many different variables for each feature, only area, perimeter, diameter, boxratio, and circularity were retained and transferred to a PC for statistical assessment. All measurements were independently carried out by 2 FDEs.

Statistical Treatment

The statistical treatment was carried out using the SPSS® 12.0 software by Mathsoft Incorporated (Needham, USA). Distributions of area, perimeter, diameter, boxratio, and circularity were first visualized by boxplots. These plots enable a quick comparison of distributions of values between pages and to identify useful variables that may discriminate further. A linear discriminant analysis was performed on the measured variables. For each character ("a," "e," and "."), the result of the multivariate analysis was represented on a 2-dimensional scatter plot defined by the 1st and 2nd discriminant axes. The correlation of each variable with each discriminant axis was calculated. Furthermore, the Hotelling's T Squared test was used to test differences in means between pairs of pages. Finally, a cross-validation was applied to estimate the correct classification rate of the discriminant functions. This is a leaving-one-out method: discriminant functions were derived on the basis of N-1 of the characters, and then used to classify the character omitted. The process was then repeated for each additional character (Marquis and others 2005).

Results

Test A

The use of the optoelectronic device demonstrated that the toner on all 5 pages of the contract was single component. The examination using the stereomicroscope allowed pages 1, 2, and 5 to be distinguished from pages 3 and 4, as the latter presented a connection between letters "R" and "E" (Figure 1).

Another difference highlighted by the stereomicroscope was the size and the shape of the dots. Page 3 was characterized by the smallest dots, whereas page 4 presented the largest ones among the 5 pages (Figure 2). Size and shape of the internal loop of letter "e" also differed among the same groups of pages—this loop was very small on page 3, slightly larger on page 4, and much larger on pages 1, 2, and 5 (Figure 3). Similar differences in size and shape were found in the internal loop of letter "a." Subjectively, pages 1, 2, and 5 showed similar character morphology appearing different from that of pages 3 and 4, which in turn were different between them.

FTIR analysis allowed the 5 pages to be differentiated into 3 groups, confirming the results of the visual observations: Group I (pages 1, 2, and 5), Group II (page 3), and Group III (page 4).

Among all the variables obtained by the personal IAS™, the best discrimination was obtained when considering area, boxratio, and circularity (Figures 4, 5, 6, Table 3). Discriminant analysis was conducted by measuring the area, boxratio, and circularity of the internal loop of the "e" and allowed for separation among the 3 groups (I, II, and III). The results of Hotelling's T Squared test indicated that differences in the multivariate means between pairs of pages belonging to a different group were significant ($p<0.001$). Through cross-validation, all observations made on pages

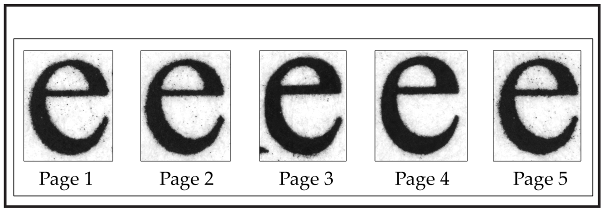

Figure 3. Illustration of size and shape of internal loop of letter "e" on the 5 pages of test A: this loop is small on page 3, a little bit larger on page 4, and much larger on pages 1, 2, and 5.

3 and 4 were allocated to their respective page; and none of the 30 measurements of pages 1, 2, and 5 were attributed to pages 3 or 4; i.e., this group could not be discriminated. This can be illustrated by the overlapping of their clouds of points (Figure 7). The 1st discriminant function accounted for 97.3% of the total variance, while the 2nd one explained the remaining 2.7%. The 1st and 2nd discriminant functions were mainly correlated to area and boxratio respectively (Table 4).

In blind Test A, the objective results of image analysis confirmed the results obtained by the classical routine methods. Both FDEs concluded that 3 computer-generated processes were used to print the 5-page contract. These results corresponded to the facts: pages 1, 2, and 5 were printed with a Canon ImageRunner 5000 PCL5e, page 3 with a Hewlett Packard LaserJet model 4200 PCL5e, and page 4 with a Hewlett Packard LaserJet model 4000N PCL6.

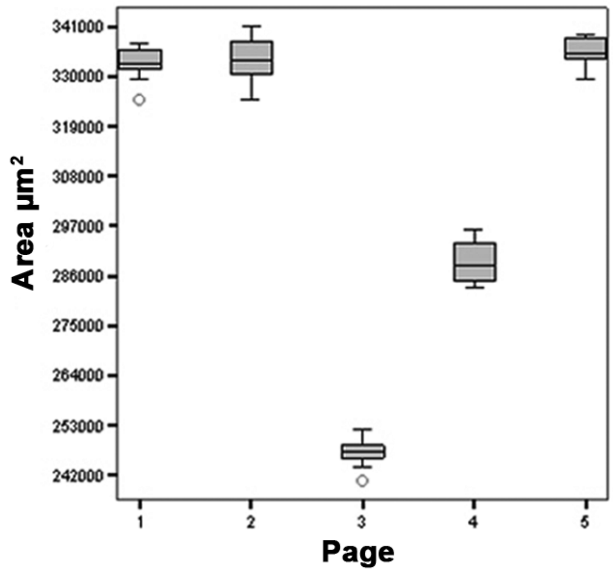

Figure 4. Boxplot of area values of internal loop of 10 "e" letters measured on each of the 5 contract pages of Test A.

Test B

The toner on all 6 pages of the contract was a dual-component toner. Examination through the stereomicroscope disclosed an identical acquired repetitive defect on pages 1 and 3. The possible cause of the defect was attributed to the organic photoconductor drum (OPC). It was concluded that pages 1 and 3 were printed by the same laser printer; therefore, no additional examination was conducted on these pages. The other 4 pages (2, 4, 5, and 6) could not be differentiated by optical exami-

	Area mm^2		Boxratio		Circularity	
	$\overline{X}$	S.D.	$\overline{X}$	S.D.	$\overline{X}$	S.D.
Page 1	0.332	0.004	0.591	0.010	2.056	0.112
Page 2	0.334	0.005	0.593	0.014	2.033	0.168
Page 3	0.247	0.003	0.519	0.011	1.910	0.122
Page 4	0.289	0.004	0.605	0.011	1.796	0.133
Page 5	0.335	0.003	0.594	0.015	2.030	0.190

Table 3. Summary statistics ($\overline{X}$ = mean and S.D. = standard deviation) of area, boxratio, and circularity of loop of 10 "e" letters measured on each of the 5 contract pages of Test A.

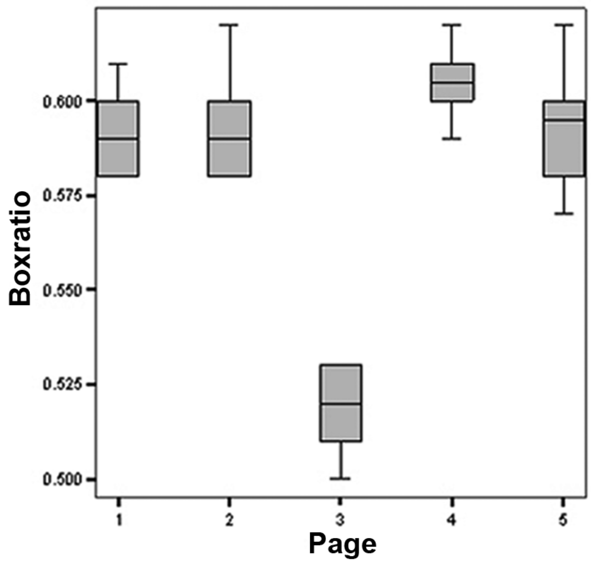

Figure 5. Boxplot of boxratio values of internal loop of 10 "e" letters measured on each of the 5 contract pages of Test A.

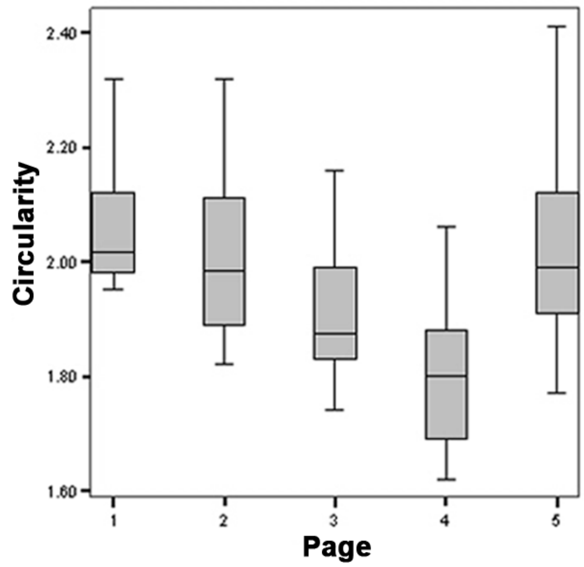

Figure 6. Boxplot of circularity values of the internal loop of 10 "e" letters measured on each of the 5 contract pages of Test A.

nation. FTIR analysis could not provide further discrimination of the remaining pages; therefore, they were classified into the same group.

The best discrimination using the personal IAS™ was obtained when area, boxratio, and circularity parameters were measured on the internal loops of the letter "a." Discriminant analysis using these variables provided a good separation among the 3 groups: page 2 (I), page 4 (II), and pages 5 and 6 (III). Results of the Hotelling's T Squared test showed significant differences in multivariate means between pairs of pages belonging to a different group ($p<0.001$). No significant difference was found between pages 5 and 6.

The FDEs concluded that at least 4 computer-generated processes were used to print the 6 contract pages. In fact, 4 different machines were used; pages 1 and 3 were printed with a Lexmark Optra T632 N° 11, page 2 with a Lexmark Optra T632 N° 10, page 4 was generated by a Lexmark Optra T632 N° 8, and pages 5 and 6 were printed with a Lexmark Optra T630 N° 6. In the 2nd blind test, the objective results provided by image analysis allowed additional discrimination of the results obtained by classical routine methods.

Test C

As indicated previously, the objective of Test C was to determine which laser printer out of the possible 13 was used to print the 3 pages of the contract. The authors were confronted with a closed set test. It was found that the toner on page 1 was single component, whereas the toner on pages 2 and 3 was a dual component. The latter could not be distinguished by optical examination. FTIR analysis was used to identify the organic binder of page 1. It was concluded that page 1 had been printed by the Canon IR400 laser printer. The FTIR spectra of toners from pages 2 and 3 were identical.

Discriminant analysis using area, boxratio, and circularity measurements of the internal loop of letter "a" could not be used to discriminate between page 2 and the pages printed with the Lexmark Optra T630 N° 7, as well as between page 3 and the pages printed with the Lexmark Optra T632 N° 10.

Discussion

The results for the blind Tests B and C lead to the conclusion that image analysis can be used as a discrimination technique. The results were complementary to those obtained by conventional techniques. A limitation of this technique is its dependence on software and hardware. As indicated by Flynn (2006), "The modern FDE must be aware of such things as the operating system, the word processing program, the version of the word processing program, the digital font and version of the font file, the printer driver, etc. All can have an impact on the appearance of the printed text, as can the many variables associated with the

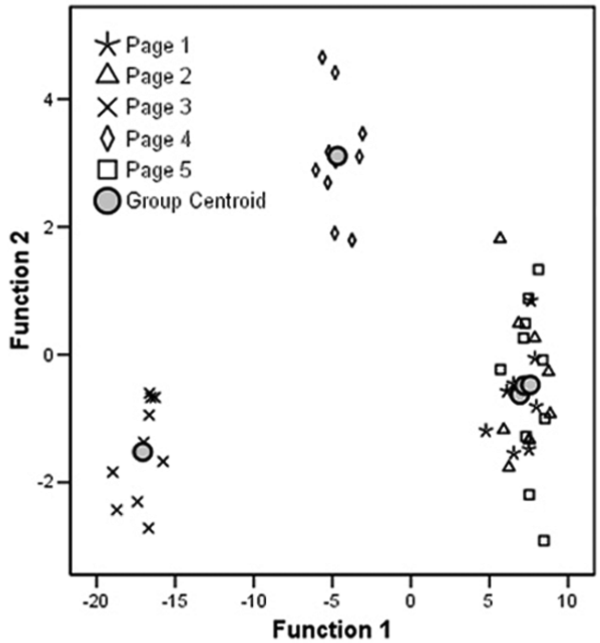

Figure 7. Results of the discriminant analysis performed on area, boxratio, and circularity of internal loop of 10 "e" letters measured on each of the 5 contract pages of Test A.

printer itself." This preliminary research paper focused only on the printer variable; i.e., on the different output documents that can be generated by different laser printers. Given that documents of Tests B and C were produced with the same paper, personal computer, word processor, and font, the differences highlighted were undoubtedly caused by the printer.

The advantage of image analysis is that documents printed by different laser printers of the same make and model can be objectively distinguished. Differences that can be highlighted are likely to be due to the printer cartridge and/or toner type. Two brand new identical OPC drums may have different photosensitivity properties that cause stroke width differences. Similarly, a different batch of toner can change the output from the same laser printer. From this study, image analysis appears to provide positive results in cases where 1 or more pages of a contract were prepared differently or a part of a text was altered. In cases where the aim is to identify the laser printer, several parameters related to hardware and software must be known to properly interpret the results of an image analysis process, as the printer is not the only factor influencing the features of interest in image analysis.

	Function 1	Function 2
Area μm^2	0.904	-0.148
Boxratio	0.206	0.949
Circularity	0.048	-0.296

Table 4. Discriminant analysis of area, boxratio, and circularity of internal loop of 10 "e" letters on each of the 5 contract pages of Test A showing the correlation of variable with the 1st and the 2nd discriminant functions.

Conclusions

This paper highlights the complementarities of forensic image analysis to classical methods employed for the analysis of laser-printed documents. Forensic image analysis is a realistic application in closed-set cases; i.e., where the number of potential laser printers can be clearly defined or in cases of page substitution or text addition.

References

Flynn, W. (2006). The Examination of Computer-Generated Documents. *Scientific Examination of Questioned Documents*, 2nd Edition, Taylor & Francis, p. 194.

Marquis, R.; Schmittbuhl, M.; Mazzella, W.D.; Taroni, F. (2005). Quantification of the Shape of Handwritten Characters: A step to objective discrimination between writers based on the study of the capital character O. *Forensic Science International*, Vol.150, pp. 23-32.

Merrill, R.A.; Bartick, E.G.; and Mazzella, W.D. (1996). Studies of Techniques for Analyzing Photocopy Toners by IR. *Journal of Forensic Sciences*, Vol. 41, pp. 264-271.

Neumann, C.; Mazzella, W.D. (2004). *Questioned Documents Encyclopedia of Analytical Science*, Academic Press, pp. 456-473.

Quality Engineering Associates, Inc. (2004). *Personal IAS® User's Guide*, 2004.

Trzcinska, B. (2006). Classification of Black Powder Toners on the Basis of Integrated Analytical Information Provided by Fourier Transform Infrared Spectrometry and X-Ray Fluorescence Spectrometry. *Journal of Forensic Sciences*, Vol. 51, pp. 919-924.

Tse, M.K.; Forrest, D.J.; She, K.Y. (1995). Use of an Automated Print Quality Evaluation System as a Failure Analysis Tool in Electrophotography. Paper presented at the 11th IS&T Congress, October 29-November 3, Hilton Head, South Carolina.