SPECIAL ISSUE PAPER

William J. Egan · Stephen L. Morgan
Edward G. Bartick · Rena A. Merrill · Hollis J. Taylor III

Forensic discrimination of photocopy and printer toners.
II. Discriminant analysis applied to infrared reflection-absorption spectroscopy

Abstract Copy toner samples were analyzed using reflection-absorption infrared microscopy (R-A IR). The grouping of copy toners into distinguishable classes achieved by visual comparison and computer-assisted spectral matching was compared to that achieved by multivariate discriminant analysis. For a data set containing spectra of 430 copy toners, 90% (388/430) of the spectra were initially correctly grouped into the classifications previously established by spectral matching. Three groups of samples that did not classify well contained too few samples to allow reliable classification. Samples from two other pairs of groups were similar and often misclassified. Closer examination of spectra from these groups revealed discriminating features that could be used in separate discriminant analyses to improve classification. For one pair of groups, the classification accuracy improved to 91% (81/89) and 97% (28/29), for the two groups, respectively. The other pair of groups were completely distinguishable from one another. With these additional tests, multivariate discriminant analysis correctly classified 96% of the 430 R-A IR toner spectra into the toner groups found previously by spectral matching.

W. J. Egan · S. L. Morgan
Department of Chemistry and Biochemistry,
The University of South Carolina, Columbia, SC 29208, USA

E. G. Bartick (✉) · R. A. Merrill
Counterterrorism and Forensic Science Research Unit,
FBI Laboratory, FBI Academy, Quantico, VA 22135, USA
e-mail: ebartick@fbiacademy.edu

H. J. Taylor III
Questioned Documents Unit, FBI Laboratory,
Quantico, VA 22135, USA

Present address:
W. J. Egan
Vertex Pharmaceuticals, 130 Waverly Street,
Cambridge, MA 02139-4242, USA

Introduction

The use of office and personal photocopying machines has increased dramatically over the last 20 years. As a result, photocopying documents has become simple, fast, and inexpensive. A major disadvantage of photocopy machines is that they are now more accessible for illegal activities such as counterfeiting, fraud, false documents, anonymous letters, confidential materials, and acts of terrorism [1, 2, 3]. Identification of the source of photocopied documents is not an easy task for forensic examiners since chemical and physical characteristics are very similar and numerous manufacturers of photocopy instruments and toner cartridges exist. The ability to match the chemical fingerprints of questioned toner samples to standards could be a valuable tool in questioned document investigations.

Toner analysis methods that are useful in forensic investigations must be rapidly performed and possess a known degree of accuracy. Totty [4] reviewed analytical techniques that have been used to characterize toners: visual examination, optical microscopy, scanning electron microscopy (SEM), magnetic viewers, infrared spectroscopy (IR), pyrolysis gas chromatography and/or mass spectrometry (Py-GC, Py-GC/MS, Py-MS), and differential scanning calorimetry (DSC). Early work by Kemp and Totty [5] found that 79 toners from various models of photocopier machines could be separated into 10 groups based on their IR spectra. Williams [6] identified numerous resins and the pigment Prussian blue based on characteristic IR absorptions. The possibilities of toner analysis and classification by IR spectroscopy and by diffuse reflection infrared Fourier transform spectroscopy (DRIFTS) have been described by other researchers [7, 8, 9, 10, 11, 12, 13, 14]. Merrill et al. [15] conducted a comparative study of three microscope-based IR techniques and DRIFTS for the analysis of toner samples. Reflection ab-

sorption infrared spectroscopy (R-A IR) performed best overall, in terms of low cost, rapid speed of analysis, nondestructiveness, and quality of spectra. Bartick and Merrill [16] also began development of an R-A IR spectral database for copy toners. More recently, toner samples extracted from photocopies with carbon tetrachloride have been analyzed by FT-IR to identify chemical constituents. [17] DRIFTS, SEM, and Py-GC have also been compared for differentiating photocopier toners [18].

Hundreds of copiers and printers are available from different manufacturers, varying in model, engine, and toner used. Comparison of a large number of spectra is tedious even when computer-assisted spectral matching is employed, and the accuracy of the classification results may not be quantifiable. Multivariate statistical methods offer a potential solution to these problems. As noted in a 1996 review on chemometrics, "[t]here were only a few papers that focused on the application of pattern recognition techniques to forensics, which is surprising in view of the potential impact that multivariate methods can have on this field" [19]. Paper samples have been differentiated on the basis of multivariate statistical analysis of their elemental compositions [20, 21]. Andrasko [13] has also applied two simple measures, the Euclidean distances between spectra and a similarity index, to differentiate black ink and color toner samples by reflection FTIR.

Merrill and Bartick [22] have previously compared prominent features and used computer-assisted spectral matching to divide 807 R-A IR spectra of toners into 98 subgroups based on the presence, absence, and ratios of peaks in 40 different spectral regions. Group assignments were summarized using a flowchart with nodes representing yes/no decisions with regard to presence, absence, or ratios of spectral peaks at specified wavenumber locations. The groups at the end point of branches represented clusters of similar copy toner spectra for which further discrimination was judged not possible. Linear discriminant analysis (LDA, also known as canonical variates analysis or CVA) is a multivariate statistical method that facilitates the objective evaluation of the classification of objects (in this case, spectra) into groups. We have previously applied LDA to R-A IR spectra from a subgroup of 60 toners having a poly(styrene-*co*-acrylate) base component [23]. The objective of the present paper is to evaluate whether groupings of R-A IR spectra of photocopy and printer toners for a larger subset of 430 toner samples can be reliably discriminated by statistical pattern recognition.

Experimental

Samples of dry photocopier and printer toners on paper were collected by the FBI Laboratory from verified sources that include original manufacturers. The R-A IR data set used in this study (430 spectra classified into 27 groups) is a subset of the complete library of 807 toner samples categorized into 98 groups previously described [22]. The entire set of 807 samples was not available at the time of this study. Toners were transferred from documents to reflective media (heavy duty aluminum foil affixed to standard glass slides with double-sided tape) using a temperature-regulated soldering iron set at 288 °C [15, 22, 24]. The soldering iron was

equipped with a screwdriver tip that had been ground off, leaving a flattened round head with a 4.8-mm diameter. Although other materials provide a suitable reflective surface for the reflection-absorption technique, aluminum foil is readily available, inexpensive, and permits the sample to be stored for further studies. The sample preparation is simple, fast, and essentially nondestructive. The document is still legible after transferring the toner sample and only minimal destruction is visible microscopically.

Samples were analyzed by R-A IR using a Spectra-Tech IR-Plan microscope with a medium-band MCT detector (Shelton, CT). The instrument collected 256 scans at 4-cm^{-1} resolution over the 650–4,000 cm^{-1} range for a total of 1,039 data points per spectrum.

Data analysis

Data was preprocessed and analyzed using OMNIC v. 3.0 (Nicolet Analytical Instruments, Madison, WI), Microsoft Excel (Microsoft Corporation, Redmond, WA), and programs written in MatLab (The MathWorks, Inc., Natick, MA).

The baselines of all R-A IR spectra were manually adjusted using the OMNIC software to remove background dispersion effects caused by carbon black in the samples [22]. Baseline adjusted spectra were then arranged in a matrix, X, whose n rows represent different spectra and m columns represent spectral frequencies (wavenumbers):

$$X = \begin{bmatrix} x_{11} & .. & .. & x_{1j} & .. & .. & x_{1m} \\ : & & & : & & & : \\ x_{i1} & .. & .. & x_{ij} & .. & .. & x_{im} \\ : & & & : & & & : \\ x_{n1} & .. & .. & x_{nj} & .. & .. & x_{nm} \end{bmatrix} \qquad (1)$$

The element x_{ij} of this matrix is the absorbance intensity at wavenumber j of spectrum i. Each spectrum (row) was then normalized to unit length by dividing each spectral intensity by the square root of the sum of the squares of all spectral intensities in that spectrum. Normalization removes systematic variation associated with size or amount effects in the spectra. After normalization, a typical preprocessing step is to autoscale the absorbances at each wavenumber by subtracting the mean spectral intensity for wavenumber j, and dividing by the standard deviation of spectral intensity for wavenumber j. Autoscaling produces a column autoscaled matrix with elements,

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \qquad (2)$$

having a mean of zero and a standard deviation of one for each feature (wavenumber). In Eq. (2), x_{ij} represents the intensity for wavenumber j in the ith spectrum, $\bar{x}_j$ is the mean intensity for wavenumber j, and s_j is the standard deviation of intensity for wavenumber j. This common data transformation removes inadvertent weighting caused by variations in the magnitude of intensity at different spectral frequencies [25]. In autoscaling the present data, the median was used instead of the mean in the numerator of Eq. (2); the median was also employed in place of the mean in calculations of the standard deviation. Error introduced by outliers is caused by a distortion of the location of the center of the data; the use of the median in these calculations may provide a more robust estimate of the true center of the data [26, 27].

After median autoscaling, the data was further preprocessed using principal component analysis (PCA) [28, 29] via singular value decomposition (SVD) [30] to project the spectra, initially consisting of intensities at 1,039 wavenumbers, into a space of reduced dimensionality. SVD decomposes the data matrix into the product of three matrices:

$$X_{n \times p} = U_{n \times k} S_{k \times k} V'_{k \times p} \qquad (3)$$

where the matrix X refers to the data matrix after any preprocessing. PCA creates linear combinations of the original spectral variables, called eigenvectors or principal components (PCs), which successively account for increasing amounts of variation in the

data. The matrix S contains the square roots of the eigenvalues of X (the singular values), ordered largest to smallest from top left to bottom right along the diagonal; the square of these values define the proportional variance explained by each PC. The columns of the matrix V contain the weights of the original variables (the loadings) necessary to form the principal component scores ($U \times S$). If the first few PCs are found to explain a substantial proportion of the variation in the data, the projection of points representing the samples in a two- or three-dimensional plot may be informative concerning their similarity. For each analysis, a number of PCs was retained for discriminant analysis to capture an adequately large fraction of variance in the data.

After preprocessing and data compression using PCA, linear discriminant analysis was used to construct axes which best separate the groups by maximizing the ratio of their between- to within-group variances [31, 32, 33]. Discriminant analysis is well suited for the analysis of grouped data and has a long history of use in analytical chemistry [20, 34, 35, 36, 37, 38]. The implementation of LDA is based upon three matrices,

$$T = (X - \bar{X})' \, (X - \bar{X}) \tag{4}$$

$$W = (X_j - \bar{X}_j)' \, (X_j - \bar{X}_j) \tag{5}$$

$$B = (\bar{X}_j - \bar{X})' \, (\bar{X}_j - \bar{X}) \tag{6}$$

representing the total, within-groups, and between-groups sums of squares and products matrices, respectively. The matrix $\bar{X}$ represents the mean of X, while j designates a group of samples. The canonical variates (CVs) are defined by the eigenvectors of the matrix $W^{-1}B$, with the proportion of variance accounted for each CV proportional to the eigenvalues. As with PCA, if a sufficiently large proportion of the variability associated with the first few CVs, a projection of the data points (the spectra) in the two- or three-dimensional space of the CVs permits the researcher to visualize clustering and similarity of the data. The clustering of similar samples can be assessed by comparison to the distances between samples (spectra) judged different from one another.

Group definitions followed the groups experimentally established from the R-A IR spectra by visual comparison and computer-assisted spectral matching [22]. Jackknife cross validation was used to estimate the predictive ability of the LDA model by removing each sample from the data set in turn and recomputing discriminant functions based on the remaining samples [39, 40]. Estimates of the classification error for each sample are obtained without using that sample to calculate the discriminant model. This "leave-one-out" method is useful when a separate test set of data is unavailable, because it provides a way to estimate the classification error rate from the available data. The similarity between the jackknifed sample and the mean vectors for each group, calculated by scaling each vector to unit length and multiplying them together, was used to assign group membership.

Elliptical confidence regions around the spectra of toner groups were calculated as follows. Each ellipse represents distances which are statistically equidistant from the group mean for a predetermined level of probability. Confidence regions were calculated by transforming each group to a principal component representation. A confidence circle based on Hotelling's T^2 statistic (the multivariate generalization of Student's t statistic) was calculated and then transformed back into the original variables, forming an ellipse [41]. Note that even if the original two variables are already PCs, the method calculates new PCs for each group separately. Hence, the confidence region may take the form of an ellipse, not a circle, when plotted on the original spectral variables.

Univariate Fisher ratios [31, 32, 37] were used as an indicator of which spectral features (wavenumbers) were individually most discriminating for separation of any specified groups. Fisher ratios, calculated as the ratio of between- to within-group variability for selected spectral wavelengths, range from zero (a nondiscriminating feature) to an unbounded upper value. Larger Fisher ratios indicate more discriminating features.

Results and discussion

Representative spectra of toners from this data set have been previously shown by Merrill and Bartick [22]. R-A IR microspectroscopy enables discrimination of toners by their organic and polymeric components. Although an experienced analyst becomes expert at recognizing distinguishing features in these complex IR spectra, the pattern recognition task is subjective and becomes quite difficult and time-consuming when numerous samples are compared. Pattern recognition techniques such as principal component analysis and multivariate discriminant analysis offer approaches to handling this complexity and to assess the statistical validity of differences observed between different samples.

PCA was applied as a dimensionality reduction technique to the data set of 430 R-A IR toner samples spanning 27 assigned groups. The wavenumber regions 2,200–2,750 and 3,201-3,998 cm^{-1} were deleted from the spectral data prior to PCA, because no peaks were located in these regions. The first three PCs comprised 51.13% of the variation, PCs 1–28 comprised 95.18% of the variation, and PCs 1–139 comprised 99.95% of the variation. The first 139 PCs were used as inputs for LDA, reducing the number of variables from the original 1,039 wavenumbers while preserving the majority of the variation in the data. Fig. 1 shows projections into the plane of the first two PCs for the entire R-A IR data set. While the plot is crowded, some clustering by R-A IR groups is revealed. In the preceding paper, R-A IR spectra were categorized into groups using five flowcharts. In Fig. 1, the greatest separation is between spectra from groups appearing in charts 1–3 (left side of the lower plot) compared to spectra from groups appearing in charts 4–5 (right side).
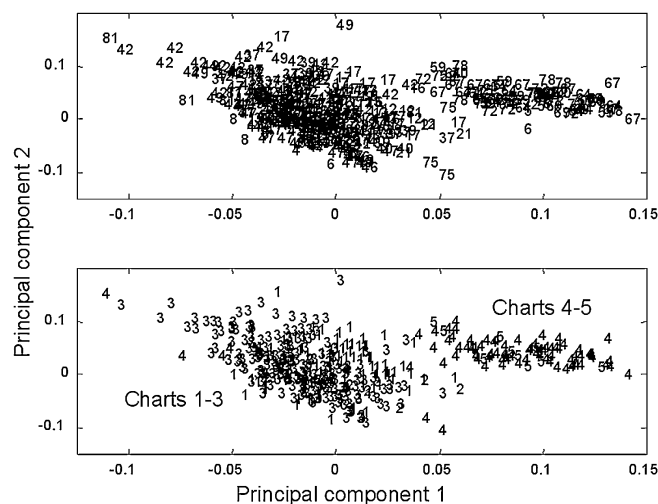

Fig. 1 The projection of all 430 R-A IR toner spectra into the space of the first two PCs: (*upper*) samples labeled by their group numbers; (*lower*) samples labeled by chart designation from the grouping assigned by visual comparison and computer-assisted spectral matching [22]
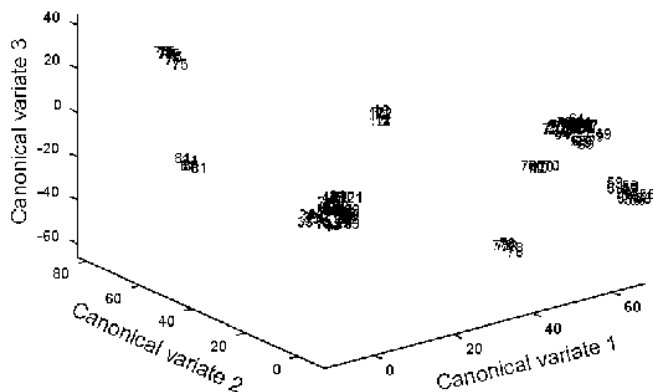
Fig. 2 The projection of all 430 R-A IR toner spectra into the space of the first three canonical variates, labeled by their group numbers. Three CVs explain 63.51% of the total dispersion. R-A IR toner groups 12, 56, 59, 70, 75, 78, and 81 are well separated on these axes

LDA of the entire data set produced 26 canonical variate axes. Such a large and complicated data set would be expected to require a large number of CVs to discriminate between groups. Projection of the spectra into the space of the first three CVs (Fig. 2) explains just 63.51% of the dispersion in the data, supporting this reasoning. However, a number of R-A IR groups are well separated by projections on only the first three CVs. Table 1 summarizes the classification accuracy for jackknifed cross validation of classification by LDA using all 26 CVs. The results were quite good: 90.23% (388/430) of the toners were correctly identified into the predetermined groups. Groups 6 and 8 had poor correct classification percentages, but contained only a few samples. It is necessary to increase the number of toners in these groups before better conclusions can be drawn concerning their classification.

Two other similar pairs of R-A IR groups were difficult to classify, and, in fact, were most often misclassified as the other group: groups 42 and 49, and groups 64 and 67. These four groups, however, did have enough samples to make further investigation worthwhile. Visual inspection of the mean spectra for groups 42 and 49 shows that the two groups differ only in two small regions: group 42 possesses small peaks at 1,115 and 1,270 cm^{-1}, whereas the peaks present in group 49 do not. Univariate Fisher ratios based on these two groups show that the two peaks are, indeed, the most important spectral features separating these two groups. The groups' mean spectra and univariate Fisher ratios are shown in Fig. 3. Fig. 4 is an expanded view of Fig. 3. The presence of a peak at 1,268–1,272 cm^{-1} is a distinguishing characteristic between the groups. The univariate Fisher ratios indicate that the 1,095–1,115 cm^{-1} is also important for separating the two groups.

To assess the ability of the 1,115 and 1,270 cm^{-1} peaks to differentiate samples from groups 42 and 49, we selected wavenumber regions with univariate Fisher ratios exceeding 0.15 (1,097–1,115 and 1,261–1,277 cm^{-1}) for further analysis. When PCA was used to compress the data, a bi-plot (Fig. 5) of the projections of the spectra on the

Table 1 Linear discriminant analysis results from classifying the full R-A IR data set into groups previously enumerated [22]. Classification accuracies were calculated using jackknifed cross validation

| Group	Correct	Total Samples	Percentage	Incorrectly Chosen
4	7	7	100.00	–
6	4	6	66.67	4,47
8	4	5	80.00	6
12	9	9	100.00	–
13	12	12	100.00	–
16	4	4	100.00	–
17	46	49	93.88	13×2,16
21	8	8	100.00	–
33	4	4	100.00	–
37	29	32	90.63	42×2,46
39	20	20	100.00	–
40	5	5	100.00	–
41	7	7	100.00	–
42	79	89	88.76	33,41×3,49×4,78,81
46	5	5	100.00	–
47	43	43	100.00	–
49	21	29	72.41	37,41,42×4,46,75
56	4	4	100.00	–
59	7	7	100.00	–
64	9	16	56.25	67×7
67	12	20	60.00	64×8
69	7	7	100.00	–
70	6	6	100.00	–
72	16	16	100.00	–
75	8	8	100.00	–
78	6	6	100.00	–
81	6	6	100.00	–
Total	388	430	90.23	
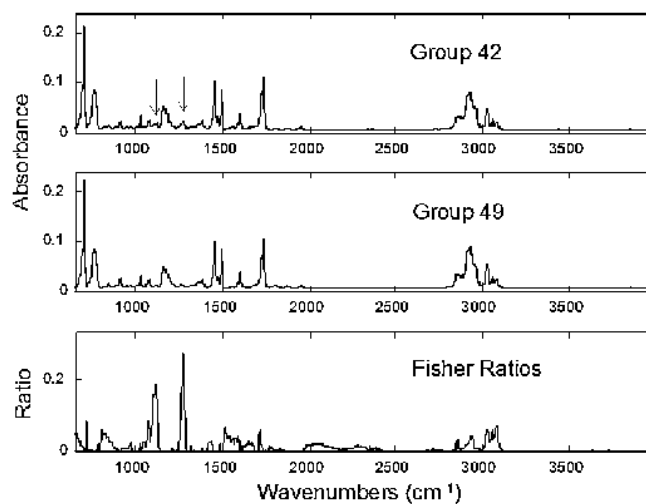
Fig. 3 Mean spectra and univariate Fisher ratios for groups 42 and 49. Note the small peaks centered at 1,115 and 1,270 cm^{-1} which have high univariate Fisher ratios, indicative of their importance for discriminating between the two groups
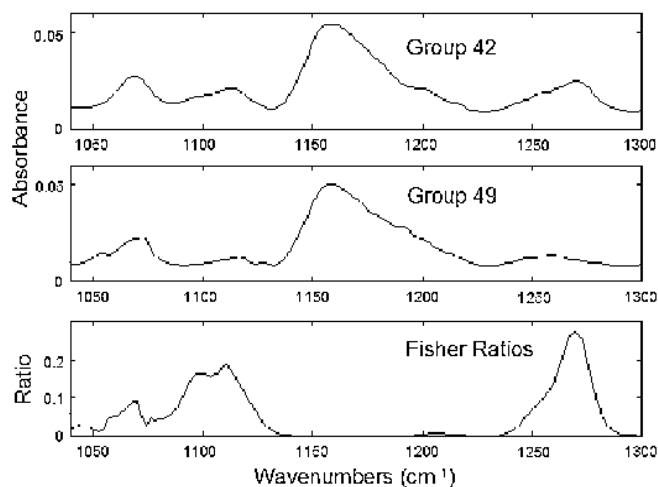
Fig. 4 Expanded view of the mean spectra and univariate Fisher ratios for groups 42 and 49
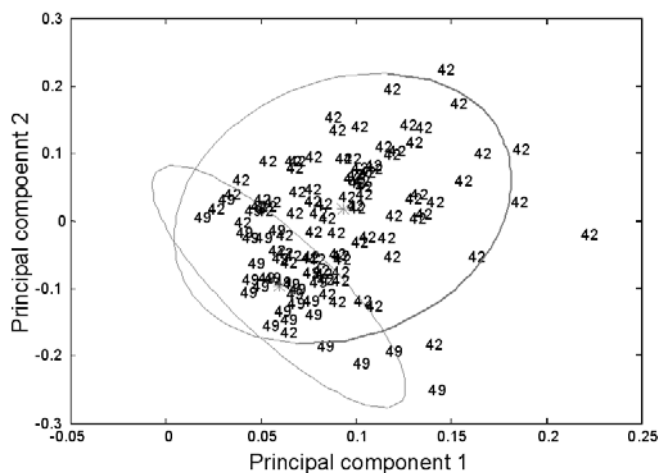

Fig. 5 Projection of spectra from groups 42 and 49 into the space of the first two PCs using the 1,097–1,115 and 1,261–1,277 cm^{-1} spectral ranges. The ellipses represent 95% confidence bounds on samples from each group

first two PCs (99.94% of the total variation) showed that the 95% confidence ellipses overlap. However, using the first 17 normalized PCs (99.99% of the total variation) as inputs for LDA, the jackknifed classification results for groups 42 and 49 improved considerably. The previous correct classification percentages were 88.76% (79/89) and 72.41% (21/29) for group 42 and group 49, respectively, based on the classification model created using all 430 samples. Use of only the two specific spectral regions improved the correct percentages to 91.01% (81/89) and 96.55% (28/29), for groups 42 and 49, respectively. The group 42 copy toners have a poly(styrene:acrylate) base component and were classified in a single group because previous visual analysis could not distinguish them [22]. However, linear discriminant analysis applied to R-A IR spectra was able to discriminate toners from several group 42 toners including AB Dick, Brother, Copystar, Okidata, Newgen, and Texas Instruments [23].
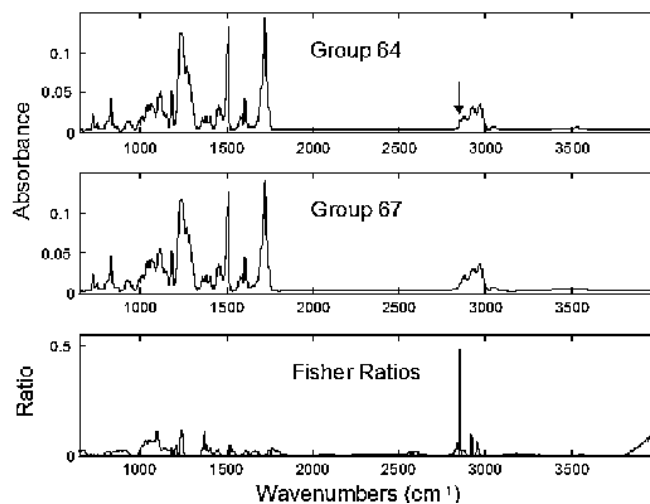

Fig. 6 Mean spectra and univariate Fisher ratios for groups 64 and 67. Note the small peak centered at 2,850 cm^{-1} which has a high univariate Fisher ratio, indicative of its importance for discriminating between the two groups
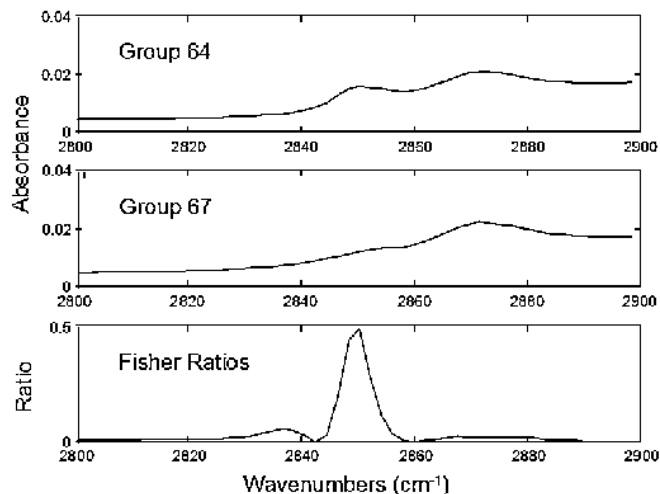

Fig. 7 Expanded view of the mean spectra and univariate Fisher ratios for groups 64 and 67

We performed a similar analysis to investigate discrimination between the other problematic groups: 64 and 67. The 2850–2852 cm^{-1} range has been used to distinguish between samples in these two toner groups [22]. The LDA model for the entire data set misclassified over half the samples in each of these two groups as belonging to the other group. Fig. 6 shows the mean spectra and univariate Fisher ratios for the two groups. An expanded view of the group 64 and 67 samples is shown in Fig. 7. When the 2,841–2,860 cm^{-1} spectral region was used as the sole input for PCA, the differences between the groups were highlighted and discrimination between the two groups was possible. A bi-plot (Fig. 8) of the spectra projected on the first two PCs (99.94% of the total variation) shows that the 95% confidence ellipses do not overlap; groups 64 and 67 are totally separable when just the 2,841–2,860 cm^{-1} range is used.
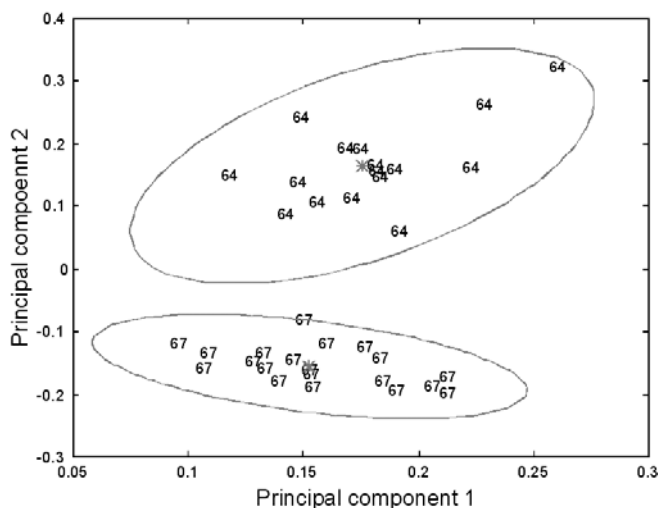
Fig. 8 Projection of spectra from groups 64 and 67 into the space of the first two PCs using the 2,841–2,860 cm^{-1} spectral range. The ellipses represent 95% confidence bounds on samples from each group

Despite the difficulties in separating the two very similar pairs of groups, the overall LDA classification results for the R-A IR data are excellent. When all groups were considered, a 90.23% (388/430) correct classification rate was achieved. When more specific analyses highlighting the importance of smaller spectral regions were performed on groups 64, 67, 42, and 49, the overall percentage of correctly classified toners rose to 95.81% (412/430).

Conclusions

We have demonstrated the successful application of multivariate statistical methods to the differentiation of photocopy and printer toners using reflection-absorption infrared spectroscopy. Discriminating among different types and manufacturer's brands of toner by visual examination of these relatively complex IR spectra can be time-consuming and subjective. Additionally, the forensic analyst may not be able to fully utilize the fine structure of the pattern due to its complexity. Multivariate pattern recognition methods can take into account the entire spectrum and thus potentially have more information to use for discrimination, but are also sensitive to minor but discriminating spectral features.

The focus of this work is development of statistical-based strategies for data handling offering improvements in method validation and ease of interpretation. Multivariate techniques provide a greater ability to discriminate between groups in large sample sets compared to visual analysis. Interpretation time can be reduced because the approach has the potential to be automated. The examples discussed illustrate the potential for computer-assisted data interpretation of forensic analytical data to provide decisive forensic identification of questioned samples. For each set of data, a visually interpretable map displaying

the quantitative similarity of the IR spectra of forensic samples can be created. In this work, linear discriminant analysis, combined with some further tests based on specific spectral regions, was able to correctly classify 95.81% of the 430 R-A IR toner spectra into the groups previously established by visual comparison and computer-assisted spectral matching. This work has demonstrated the statistical validity of the groups of toner spectra assigned by previous work at the FBI Laboratory [22]. Further work in our laboratories involves comparisons of the discrimination achieved by R-A IR to that achieved using copy toner elemental compositions determined by scanning electron with X-ray dispersive analysis (SEM-EDX) and organic polymer composition analyzed by Py-GC/MS [42].

References

1. Cantu AA (1991) Anal Chem 63:847A-854A
2. Totty RN (1990) Forensic Sci Int 46:121–126
3. Gilmour CL (1994) Can Soc Forens Sci J 27:245–259
4. Totty RN (1990) Forensic Sci Rev 2:1–23
5. Kemp GS, Totty RN (1983) Forensic Sci Int 22:75–83
6. Williams RL (1983) Forensic Sci Int 22:85–95
7. Zimmerman J, Mooney D, Kimmett MJ (1986) J Forensic Sci 31:489–493
8. Lennard CJ, Mazzella WD (1991) J Forensic Sci Soc 31:365–371
9. Shiver FC, Nelson MS, Nelson LK (1991) J Forensic Sci 36:145–152
10. Mazzella WD, Lennard CJ, Margot PA (1991) J Forensic Sci 36:449–465
11. Mazzella WD, Lennard CJ, Margot PA (1991) J Forensic Sci 36:820–837
12. Andrasko J (1994) J Forensic Sci 39:226–230
13. Andrasko J (1996) J Forensic Sci 41:812–823
14. Trzcinska BM, BrozekMucha Z (1997) Mikrochim Acta Suppl 14:235–237
15. Merrill RA, Bartick EG, Mazzella WD (1996) J Forensic Sci 41:264–271
16. Bartick EG, Merrill RA (1995) In: Jacob B, Bonte W (eds) Advances in forensic sciences, vol 3: Forensic criminalistics 1. Verlag, Berlin, pp 310–313
17. Tandon G, Jasuja OP, Sehgal VN (1997) J Forensic Document Examiners 3:119–126
18. Brandi J, James B, Gutowski SJ (1997) Int J Forensic Document Examiners 3:324–344
19. Brown SD, Sum ST, Despagne F, Lavine BK (1996) Anal Chem 68:21R–61R
20. Duewer DL, Kowalski BR (1975) Anal Chem 47:526–530
21. Simon PJ, Giessen BC, Copeland TR (1977) Anal Chem 49:2285–2288
22. Merrill RA, Bartick EG, Taylor HJ III (2003) Anal Bioanal Chem Doi 10.1007/s00216-003-2074-z (in press)

23. Bartick EG, Merrill RA, Egan WJ, Kochanowski BK, Morgan SL (1998) In: de Haseth JA (ed) Fourier transform spectroscopy: 11th international conference, American institute of physics Woodbury, pp 257–259 (13 August 1997); AIP Conference Proceedings 430:257–259
24. Bartick EG, Tungol MW, Reffner JA (1994) Anal Chim Acta 288:35–42
25. Sharaf MA, Illman DL, Kowalski BR (1986) Chemometrics. Wiley, New York
26. Rousseeuw PJ (1991) J Chemom 5:1–20
27. Egan WJ, Morgan SL (1998) Anal Chem 70:2372–2379
28. Jolliffe IT (1986) Principal component analysis. Springer, Berlin Heidelberg New York
29. Wold S, Esbensen K, Geladi P (1987) Chemom Intell Lab Syst 2:37–52
30. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical recipes: the art of scientific computing. Cambridge University Press, New York, pp 52–64
31. Fisher RA (1936) Ann Eugenics 7:179–188

32. Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis. Academic Press, San Diego
33. Campbell NA, Atchley WR (1981) Syst Zool 30:268–280
34. Windig W, Haverkamp J, Kistemaker PG (1983) Anal Chem 55:81–88
35. Vallis LV, MacFie HJ, Gutteridge CS (1985) Anal Chem 57:704–709
36. Devaux MF, Bertrand D, Robert P, Qannari M (1988) Appl Spec 42:1015–1019
37. Sahota RS, Morgan SL (1992) Anal Chem 64:2383–2392
38. Ma CY, Bayne CK (1993) Anal Chem 65:772–777
39. Tukey JW (1958) Ann Math Stat 29:614
40. Mosteller F, Tukey JW (1986) In: Jones LV (ed) The collected works of John W. Tukey. Wadsworth & Brooks Monterey, vol IV, chap 15
41. Jackson JE (1980) J Quality Technology 12:200–213
42. Egan WJ, Galipo RC, Kochanowski BK, Morgan SL, Bartick EG, Ward DC, Mothershead RF II, Miller ML (2003) Anal Bioanal Chem (in press)