



The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: Investigation of a Time-Effective Approach to the Discrimination and Interpretation of Inkjet Printer Inks Using Micro Raman Spectroscopy to Produce Investigative Leads

Author(s): Patrick Buzzini, James Curran

Document Number: 255628

Date Received: October 2020

Award Number: 2016-DN-BX-0164

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Final Summary Overview

Project title: *Investigation of a time-effective approach to the discrimination and interpretation of inkjet printer inks using micro Raman spectroscopy to produce investigative leads*

Award Number: 2016-DN-BX-0164

Authors: Patrick Buzzini, James Curran

Purpose of the project

The main purpose of this project was to evaluate if Raman data obtained from the three colored microscopic dots of inkjet printed documents constitute a chemical signature sufficiently discriminating to provide reliable investigative leads in a time-effective and non-destructive manner. The following objectives were set:

1. Develop an objective chemometric approach that can reliably differentiate between samples from different sources based on minor spectral differences;
2. Evaluate if different ink color dots from the same source can be treated as independent pieces of information. That is, if there is any relationship (i.e., dependence) between the spectra for the different color components from the same source.
3. Compare with the traditional method of thin layer chromatography (TLC) to inform questioned document examiners about the complementarities between the two approaches.

Project design and methods

The project design included: sample collection, data acquisition, and data analysis oriented toward method comparisons at two levels: 1) Comparisons between different chemometric classifiers for Raman data and 2) comparisons between Raman and TLC methods.

Samples for this study consisted of 70 Original Equipment Manufacturer (OEM) inkjet printer ink samples in the form of printed test documents provided by the Counterfeit Forensic Section of the US Secret Service (USSS). Samples were selected according to their representativeness in casework. A collection of 171 reference colorants (19 pigment-based and 152 dye-based) was also obtained. This sample set was assembled based on an open-source survey on colorants used for inkjet printer technology.

Raman Analyses were conducted using an XploRA Raman instrument (Horiba Scientific, Edison, NJ) coupled to a confocal Olympus BX51 research light microscope (Olympus Inc., Center Valley, PA). A 50x objective lens, a near-infrared laser at 785 nm, and a 1800-250 cm^{-1} spectral range were used. 7 spectra for each colored cyan, magenta, and yellow (CMY) components were collected for each sample. Spectra were converted into data frames for both non-baseline and baselined corrected spectra (using a multi-point level baseline correction method). Spectra were also collected from the 171 reference colorants.

Thin layer chromatography (TLC) was applied to the inkjet samples and to the 152 dye-based reference colorants. A dilution factor for the dyes was first determined and resulted in 2 μL extractions of solvent (ethanol and water in a 1:1 ratio) per 0.1mg of sample. The used elution solvent was the mixture of ethyl acetate, ethanol, and water in 70:35:30 proportions. The eluted spots were image-recorded with white light and UV 312nm illuminations using the video spectral comparator (VSC) model 6000 from Foster

& Freeman (Evesham, UK).

Data analysis

A subset of 11 samples was selected for the initial phase of data analysis. These samples were selected to allow for comparisons between a) different printer manufacturers, b) same manufacturers and different models and c) same manufacturers and same models. Spectra of the CMY components were first compared visually and groupings were noted. This part allowed to a) understand the spectral features useful to discriminate different samples and b) set the expectations of the tested chemometric classifiers to detect these features and discriminate different samples in agreement with visual comparisons.

Spectral pre-treatment was conducted by means of normalization of the Raman spectra and four normalization methods were evaluated: normalization to the same frequency, normalization to area unity, normalization to unit sum, and standard normal variate (SNV). Two types of classification methods were evaluated: Linear Discriminant Analysis (LDA) and Ensemble Methods. Three LDA variants were used: 1) Principal component analysis (PCA) followed by LDA, 2) Partial least square discriminant analysis (PLSDA) and 3) "Sparse" LDA. Two ensemble methods were used: Random Forests and Gradient Boosting Machines (GBM).

Each classifier was evaluated with respect to the single-color components, their combination (i.e., merging the data sets of the individual components), spectral pre-treatment (baseline correction or not) and variable selection (full or reduced spectra with targeted spectral range). For each combination, a 3-fold cross-validation was carried out where 2/3 of the data were randomly selected for training the model and 1/3 was used for

testing the model. A single iteration of cross-validation consisted of randomly partitioning the data set into three parts (say A, B and C, where train AB, test on C; train AC, test on B; train BC test on A). Due to the random selection of the data, this process was repeated 10 times to mitigate the effect of the random partitioning of the data. For each method, the random number seeds were reset to the same starting point so that the same random partitions could be obtained, and each method could be tested according to the same training and testing combinations. The performance of the tested classifiers was evaluated by means of: a) The overall model accuracy; b) the number of false predictions or misclassifications; and c) the comparison with visual inspection groupings. The overall accuracy was recorded for each test set of the 3-fold cross-validation. The average across all 30 sets of testing results was obtained to calculate the mean accuracy. Confusion matrices were produced to check the rate of misclassifications also in the light of consideration of the groups observed during visual comparisons. The three best classifiers were then applied to the full datasets.

Project findings

Raman spectroscopy

Spectral patterns are differentiated based on the presence of Raman bands at a given wavenumber value (x-axis), their width (bands vs. sharp peaks) and their relative intensities (i.e., the height of a band relative to the height of another one). During visual comparisons, spectral differences were recorded in obvious cases where Raman spectra from different samples differ completely. Differentiations were also observed in cases where Raman patterns from different samples yielded similar Raman bands along with a variable number of differing bands. Finally, spectral differences were noted in cases

where the general Raman patterns of different samples were indistinguishable except for a low number (1 to 3) of minor peaks often of minor intensity. Differentiations were first noted for each single CMY color and then they were recorded considering the three components simultaneously. Differentiations given spot color resulted in 4 groups for the cyan components, 8 groups for the magenta components and 9 groups for the yellow components. In some cases, Raman bands attributed to the paper may have contributed to achieve discriminations. It appears that some ink dots be more opaque than those of other samples thus hiding the paper support. Simultaneous consideration of Raman data obtained from the three components was critical to improve the discriminating capabilities of the Raman method: all 11 samples could be differentiated (Buzzini *et al.*, 2018).

The use of the LDA and ensemble methods on the first 11 samples confirmed that the joint consideration of the spectral data of the three CMY components increased the rate of the discriminations. Overall, sparse LDA and GBM yielded the highest accuracy values. GBM offered a better performance without the necessity to pre-treat the spectra, which would result in increased time efficiency. Following analyses of the confusion matrices, the few observed misclassifications were mostly justified in the sense that they were in agreement with the groupings observed during visual comparisons.

Data analysis on the full sample set confirmed that GBM produced the highest rates of correct classifications when the spectral data from all the three colored inkjet components are considered simultaneously. However, consideration of the single colors showed that sparse LDA offered the highest accuracy for cyan and yellow components. The most correct classification for magenta components was obtained with both GBM and Random Forests.

Thin layer chromatography (TLC)

The population of the inkjet printer ink samples could be divided in 28 groups based on their elution patterns. Unsuccessful extraction was observed for 9 samples indicating their pigment-based nature. A major group of 11 samples was observed followed by 1 group of 6 samples, 1 group of 4 samples, 4 groups of 3 samples, 3 pairs and 17 individual groups. It is important to note the contribution of UV excitation (i.e., 312nm) for discriminating purposes. The TLC data of the ink samples were compared with those of the 152 dye-based reference colorants. Dye identifications were attempted, however, the current collection does not cover all the TLC profiles that were observed within the ink samples. Moreover, not all the provided samples had their 3 colored components separated and therefore elutions were carried out from the extracts of printed documents carrying dots of the three colored dots altogether.

Complementarities between Raman spectroscopy and thin layer chromatography

The complementarities between the Raman method and thin layer chromatography were studied at two levels: first, in terms of discriminating potential and second, in terms of colorant identifications. Considering the overall dataset, TLC could not be performed on all samples because of failure to extract. As mentioned, failure to extract was characteristic of pigment-based ink samples. As a result, Raman spectroscopy in conjunction with a chemometric approach not only offered a higher discriminating potential but offers the potential to identify and differentiate pigment-based samples.

With regards to colorant identification, despite it was not possible to find a corresponding colorant for all samples using both Raman and TLC data, in some

instances the Raman method helped differentiating samples with indistinguishable TLC spots. For example, the Raman method allowed differentiating blue phthalocyanine cyan dyes *C.I. Acid Blue 249* from dyes *C.I. Acid Blue 185*, *C.I. Direct Blue 86* and *C.I. Direct Blue 199*, these latter being indistinguishable.

Implications for criminal justice policy and practice in the United States

This study is expected to contribute considerably toward the adoption of a sensible chemometric approach to utilize a large quantity of spectroscopic data (i.e., Raman patterns) in the early stage of an investigation. Maximizing time-effectiveness of forensic analyses in a reliable fashion was a top priority of this study. Although the starting point of this research was the improvement of the current analysis of inkjet-printed counterfeit banknotes, the approach is definitely transferable to questioned document examiners working in forensic laboratories who are concerned with casework involving any type of document produced with inkjet printer technology. They will also gain insights about the complementarities between Raman spectroscopy and the routine TLC method. The proposed methodology fulfills the expectations of the counterfeit specialists, and questioned document examiners, because it is *objective, time-effective and non-destructive* and can be easily integrated in an already existing investigative framework based on sourcing, individualization and case linkage.

Questioned document examiners are not the only beneficiaries of this study. The proponents had the opportunity to observe the problems investigated in this study in the forensic discipline of trace evidence. While it is easy to mention the possibility to identify the main colorants from spectra collected from unknown specimens, it is subtler to value the presence of minor peaks that are observed in Raman patterns. These minor peaks

play a crucial role for discriminating purposes.

Scholarly products

Three articles are expected to be published in scientific peer-reviewed journals in addition to the following one:

Buzzini P, Polston C, Schackmuth C. On the criteria for the discrimination of inkjet printer inks using micro Raman Spectroscopy. *Journal of Raman Spectroscopy* 2018; 49(11): 1791-1801. DOI: 10.1002/jrs.5458.

Also, the following presentations have been delivered at scientific meetings:

Buzzini P, Curran J, Polston C. *The classification of Raman patterns of inkjet printer inks: comparing visual inspection and different variants of linear discriminant analysis methods.* Poster presented at the 77th Annual General Meeting of the American Society of Questioned Document Examiners (ASQDE), Cary, NC, August 4-8 2019.

Buzzini P, Curran J, Polston C. *The classification of Raman patterns of inkjet printer inks: comparing visual inspection and different variants of linear discriminant analysis methods.* Presented at the Inter/Micro 2019 conference. McCrone Research Institute, Chicago, IL, June 10-12 2019.

Buzzini P, Curran JM, Polston C. *The classification of Raman patterns from inkjet printer inks using visual comparisons of spectra and different statistical methodologies.* Presented at the 71st meeting of the American Academy of Forensic Sciences, Baltimore, MD, February 18-23 2019.

Buzzini P, Polston C, Schackmuth M. *The discrimination of inkjet printer inks using micro Raman spectroscopy*. Poster presented at the joint meeting of the American Society of Questioned Document Examiners and the Southwestern Association of Forensic Document Examiners. Park City, UT, August 18-23 2018.

Buzzini P, Polston C, Schackmuth M. *Evaluation of the criteria for the discrimination of inkjet printer inks using micro Raman spectroscopy*. Presented at the Inter/Micro 2018 conference. McCrone Research Institute, Chicago, IL, June 11-15 2018.

Buzzini P, Curran J, Polston C. *The application of t-stochastic node embedding and random forest statistical methods to classify Raman spectra of inkjet printer inks for purposes of identification and production of investigative leads*. Presented at the 70th meeting of the American Academy of Forensic Sciences, Seattle, WA, February 19-24 2018.

On the forensic contribution of micro-Raman spectroscopy to the chemical characterization of inkjet printer inks - an investigative approach. Webinar presented to the Center for Security Printing and Anti-Counterfeiting Technology (SPACT). Rapid City, SD, July 20 2017.

The interested reader may access to them at: <http://www.ifrti.org/research.html>.