J10 The Classification of Raman Patterns From Inkjet Printer Inks Using Visual Comparisons of Spectra and Different Statistical Methodologies

Patrick Buzzini, PhD, Sam Houston State University, Huntsville, TX 77340; James M. Curran, PhD, University of Auckland, Auckland 1142, NEW ZEALAND; Carrie Polston, BA, Sam Houston State University, Huntsville, TX 77341*

Learning Overview: After attending this presentation, attendees will understand the criteria to differentiate Raman patterns measured on inkjet-printed documents from different sources and the ability of different multivariate statistical methods to classify the collected Raman spectra.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by discussing the comparison between traditional visual inspection and different statistical methods to differentiate and classify spectroscopic patterns in an investigative context.

Inkjet printers are devices that are commonly encountered in society. Therefore, it is not surprising that they are likely to be involved in different types of illicit activities, such as threats or extortions by means of anonymous letters, fraud in the context of disputed contracts, alterations to or counterfeit of ID documents, or counterfeit of currency. The printing process used in inkjet technology involves the production of a constellation of ink dots of micrometric size that are projected without contact to the receiving surface (i.e., paper). When coupled with a microscopical approach, Raman spectroscopy has been demonstrated to be a suitable method for obtaining a chemical signature *in situ* from the three main colored components (cyan, magenta, and yellow) of inkjet printer inks. Although the Raman technique is already a relatively well-established method for the characterization of colorants (both dyes and pigments), this study highlights the contribution of minor peaks within Raman spectra to improve the discriminating capabilities of the technique.

In the present phase of this project, 231 Raman spectra were collected from the cyan, magenta, and yellow color components of 11 inkjet printer ink samples provided by the Counterfeit Forensic Section of the United States Secret Service, using a Near-Infrared (NIR) laser wavelength at 785nm. Spectra were first compared visually, and groupings were formed for each individual color and for the three colors jointly considered. Since the overall goal of this project is to evaluate if Raman data gathered from these three components constitute, together, a chemical signature of sufficient discriminating quality to provide reliable investigative leads, many spectral comparisons need to be carried out. However, visual comparisons are impractical and tedious for this purpose, and a sensible statistical classifier is then required for conducting spectral comparisons. To facilitate the implementation of a statistical approach, two main problems need to be considered. The first is that most differentiations between spectra of different inkjet printer ink samples are made by consideration of minor peaks since most of the Raman bands present in a spectrum are shared by multiple samples. This is due to the prevalent use of few colorants in the inkjet manufacturing industry and/or to the higher scattering property of these colorants compared to other present within the ink formulation. The second problem is that Raman spectra often exhibit large intra-source variation with regard to the absolute intensity of the signal, although their relative intensities and wavenumber positions present high repeatability. This implies that appropriate data pre-treatment and feature selection must be carefully investigated.

In this study, spectra have been submitted to baseline correction, and four normalization methods have been investigated: normalization to the same frequency, normalization to area unity, normalization to unit sum, and Standard Normal Variate (SNV). Data dimension reduction techniques of Principal Component Analysis (PCA) and t-Stochastic Node Embedding (t-SNE) have been utilized for visualizing the collected high-dimensional spectral data. The potential of the classification methods of Linear Discriminant Analysis (LDA), random forests, and Naïve Bayes classifiers were evaluated. At this stage, none of the selected combinations of statistical methods resulted in the classifications observed by visual inspection of spectra. For example, for random forests (applied after normalization to area unity and t-SNE), ten-fold repeated cross-validation with three repeats per fold were used to achieve an overall accuracy of the model of approximately 75% for the cyan color. A following step of this study will consist of selecting lower portions of the spectral range that are expected to yield less inter-sample redundancy.

This research is funded by an award from the National Institute of Justice.

Questioned Documents, Raman, Inkjet