Microchemical Journal 144 (2019) 411–418

Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc

Discriminating blue ballpoint pens inks in questioned documents by Raman imaging and mean-field approach independent component analysis (MF-ICA)

Carlos A. Teixeira, Ronei J. Poppi*

Institute of Chemistry, State University of Campinas, POB 6154, 13084-971 Campinas, SP, Brazil

ARTICLE INFO

Keywords:
Ink
Questioned document
Raman imaging
Independent component analysis

ABSTRACT

In the present work, Raman spectroscopy and chemometrics tools were explored as an analytical method to discriminate blue pen strokes on common office paper and bank check paper. The sample set was constituted of blue ballpoint pen inks available on local market and promotional pens as well. The samples were prepared to simulate an adulterated document using different pens. The proposal methodology was based on Raman imaging spectroscopy and Independent Components Analysis algorithm, which were able to recover the original spectrum of each pen, as well as images containing special information about each of the pens used. All recovered spectra could be match with references spectra at least with 0.8572 of correlation coefficient, while recovered images left no doubt about whether a forgery occurred. At the end of analysis, all the examined cases could be solved based on the discrimination between the pens inks, without any prior information.

1. Introduction

Forensic Chemistry is a branch of Chemistry that deals with applications of this subject in a legal setting. This branch of Chemistry basically attends the study areas of the Criminal and Legal Medicine. Forensic analysis is focused on the study of waste firearms and explosives [1,2], discrimination between authentic and counterfeit banknotes [3], fingerprinting identification [4,5], adulteration in fuels [6] and food [7] and the forensic analysis of documents [8]. This specific field is the part of the criminalistics that studies questioned documents to verify their authenticity or authorship.

A document is a paper or set of papers with written or printed information, that has a meaning or carry forward a message to someone [9,10]. Because of this broad definition, documents can be presented in various forms such as contracts, wills, checks, as well as petitions. Due to the importance of some documents, these are often pivots of investigations involving goods and values. Therefore, it is necessary to develop of physical and chemical methods of authenticity analysis.

Documents those suspected of being fraudulent or controversial in some way are classified as disputed document or questioned document [10,11]. Not all questioned documents are fraudulent indeed, but when it is, it is characterized as a change on any structural modification of part or all its contents. Regardless of the type of fraud, the study of questioned documents begins with a visual examination of the expert, followed by physical and chemical analyses. Depending on the type or importance of the document there is a range of techniques available for the study, such as Thin Layer Chromatography (TLC), Fourier-Transform Infrared Spectroscopy (FTIR), Gas Chromatography-Mass Spectrometry (GC–MS), Capillary Electrophoresis, X-ray Spectroscopy, as well as Raman Spectroscopy [12].

Due to documents of high value and unique, the use of analytical routines that preserve the integrity of forensic samples is required, and on that account the appeal for techniques such as Near-Infrared Diffuse Reflectance, Attenuated Total Reflectance Fourier Transform Infrared and Raman is increased.

Raman spectroscopy meets certain characteristics that are propitious to the forensic area. The technique allows obtaining information that helps in the identification of chemical substances, and it can be applied to solutions, solids, oils, body and fluids, normally without any specific pre-treatments. To obtain the spectra, a monochromatic light source (or a laser) interacts with the sample surface, and the *radiation scattered inelastically* [13] provides information about the chemical composition of the analysed sample.

The use of Raman spectroscopy has already proved to be useful oftentimes in questioned documents analysis field [14,15]. The major advantage of Raman over others spectroscopic techniques, specifically for questioned document analysis, is the rich spectral information, presented as a high number of vibrational bands. However, the

* Corresponding author.
E-mail address: rjpoppi@g.unicamp.br (R.J. Poppi).

https://doi.org/10.1016/j.microc.2018.10.002
Received 11 July 2018; Received in revised form 27 September 2018; Accepted 1 October 2018
Available online 03 October 2018
0026-265X/

presence of fluorescent compounds on the paper surface can mask the Raman signal of interest [16], decreasing its intensity. To overcome this problem, it is possible to use lasers on the infrared region, which decrease the undesired effect of fluorescence [17], or yet use Surface Enhanced Raman Spectroscopy which in addition to eliminating such types of interference, makes it possible to increase the Raman signal of selected bands, thus increasing its sensitivity [18,19].

Advances in optics, software and component development electronics, added to the use of chemometrics tools, made possible the emergence of new ways of using Raman spectroscopy as an analytical tool, as example Raman imaging spectroscopy. The technique enables to choose a specific area, where a spectrum is obtained at each point (pixel) on sample surface. The data obtained is a three-dimensional arrangement with two dimensional spatial coordinates and a dimension related to the wavenumbers of each spectrum.

Normally the differentiation between chemical compositions of pens inks, can be performed by visual analysis of the spectrum, in search of specific bands. However, when mapping samples that have crossed traces on a paper, for example, the scattered radiation will always be a mixed information from pens, paper and everything else on the surface of the sample. In addition, pigments are complex mixtures and sometimes their formulation is not disclosed, therefore the allocation of all bands of all spectra, seeking a univariate analysis can be a difficult task. In this sense, the use of multivariate analysis, such as PCA [20] and MCR-ALS [14,15], are of great use for the analysis of spectral data of inks.

According to the literature [21], considering only publications in the questioned documents area, most of these publications had as object of study the pigments of pens. Other sample types like jets, printer toners and the paper itself were less prominent. In this aspect, some papers discuss the use of images (whether near infrared, Raman or visible light) as non-destructive methods, associated with chemometric methods, to identify document forgery.

Following this research line, the aim of this work was to use Raman imaging spectroscopy and Mean Field approaches Independent Component Analysis (MF-ICA) as multivariate tool, to identify forgeries. The application of MF-ICA in Raman Imaging documentoscopy dataset brought a new perspective, mostly about how a pen leaves the trail that makes a handwritten character. The analyses were conducted by evaluating the addition of numeric characters, firstly on common office paper, and then in bank check sheets, simulating an application closer to real conditions. The proposed methodology could be able to discriminate, chemically and spatially the graphic instruments employed to prepare and adulterate the document in question, even when both spectra had high correlation coefficient.

2. Data analysis

2.1. Mean-field approach independent component analysis (MF-ICA)

Independent component analysis (ICA) belongs to the same class of algorithms as principal component analysis (PCA), but on the first we assume that the components are fully independent rather than just not correlated [22]. In this case independence means that the variation of one component has no influence on the variations of the other component [23], which is fair assumption for a sample consisting of two different pen inks over a piece of paper. That is probably the major advantage of ICA over others projection pursuit approaches. PCA uses first and second moments from acquired data set, trusting almost exclusively on the Gaussian distribution of the data. On the other hand, ICA explores intrinsically non-Gaussian characteristics and thus using higher moments [24]. Once the distribution of independent signals is not Gaussian, it turns in an advantage for the algorithm over others curve resolution solutions [23].

The main idea of the algorithm is to reduce data dimension, describing them in terms of independent components. The model assumes that a set of observed signals **M** is formed by contributions from linear unknown sources **S** and their proportions **A**, which in turn are mutually and statistically independent. Mathematically we have:

$$\mathbf{M} = \mathbf{A}\mathbf{S}^{\mathbf{T}} \tag{1}$$

In this case **M** is the observed signals, **A** is the mixing matrix (concentrations) and **S** is the pure signals (spectra). The purpose of ICA is to extract information (**A** and **S**) from a set of mixed signals of unknown proportions, using exclusively the data contained in **M**, as show in Fig. 1.

Since the information about coefficients and independent components are unknown, it is necessary to find an unmixing matrix **W** which performs a linear transformation in **M** extracting an estimative of independent components $\widehat{\boldsymbol{S}}$:

$$\widehat{\boldsymbol{S}}^{\mathbf{T}} = \mathbf{W}\mathbf{M} \tag{2}$$

The main objective of the PCA is to minimize the error projections of compressed data, while the purpose of the ICA is to minimize the statistical dependence between the base vectors. Mathematically ICA tries to find a linear transformation, through matrix **W**, which minimizes the statistic dependence between the lines of **M** [23].

Form the **S** matrix, the proportions in **A** can then be calculated by:

$$\mathbf{A} = \mathbf{M}\mathbf{S}^{\mathbf{T}}(\mathbf{S}\mathbf{S}^{\mathbf{T}})^{-1} \tag{3}$$

Different from PCA, the bases vectors of ICA are not orthogonal,

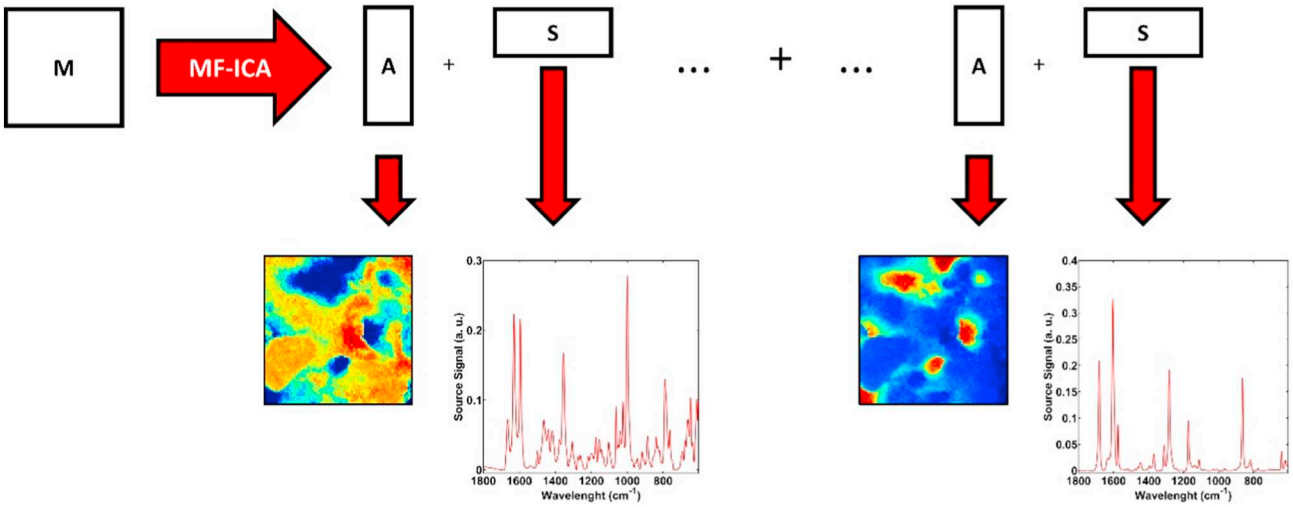

Fig. 1. Schematic representation of ICA as the sum of contained components in **A** and **S**, and their respective chemical meanings.