Pattern Recognition Based on Principal Component Analysis in Raman Spectroscopy Applied to Artistic Pigments

J. J. González-Vidal, R. Pérez-Pueyo, M. J. Soneira and S. Ruiz-Moreno
Signal Theory and Communications Dept., Universitat Politècnica de Catalunya, Barcelona; Spain

Abstract

It is proposed an algorithm, based on the general scheme of pattern recognition and on the technique of principal components analysis, developed to automatically select the reference Raman spectrum corresponding to an unknown Raman spectrum, which is a 'fingerprint' of the analyzed material. It is applied to the identification of artistic pigments in art works. The performance of the implemented algorithm was studied in a simulation stage and verified with experimental cases of Raman spectra measured on paintings. The algorithm is fully automated and provides a factor of reliability on the identification, being a helpful tool for the analyst in the decision-making process.

Introduction

Scientific research on art works is done for different reasons; for instance, the artistic and historical interest to know the materials used in an art work, since these materials vary depending on their geographical or chronological location. Pictorial materials or pigments have been used from prehistoric times and throughout the History have been fundamental in graphic arts. Therefore, a detailed study of the pigments that constitute a work of art is important to reveal information regarding the restoration, conservation, dating and authentication of the art work.

In this sense, the analysis of the pigmentation of a work of art fits the scheme of pattern recognition, which is the process of categorization of observed samples by reference data (also called patterns). The recognition system assigns each observation to its category. In order to recognize the patterns, a set of processes is followed:

- Data acquisition: based on a sensor that collects the elements of the universe to be identified.
- Feature extraction: a mechanism that extracts the useful information, removing redundancies.
- Classification: the *a priori* unknown observation is assigned to the appropriate category.

In recent years, Raman spectroscopy has exponentially increased its application in the art world. This photonic technique, based on the Raman Effect, provides univocal information on the molecular composition of the analysed sample [1, 2]. The representation of this information is known as Raman spectrum and is composed of bands, whose positions recognize unequivocally the material. In addition, Raman spectroscopy uses low power lasers becoming a suitable technique for the non-destructive identification of pigments.

Traditionally, the artistic pigments identification has been carried out by visual comparison between the Raman spectra measured on art works with an appropriate set of reference Raman spectra. Nevertheless, since identifying Raman spectra of some kind of modern pigments may turn out to be a complex task due to their large number of bands located close together, the automation of the identification process has become a hot topic of research nowadays [3].

Materials & Methods

With the aim of speeding up the automatic identification process and saving computing resources it is used the technique of principal components analysis (PCA) as a data reduction tool. In this sense, PCA is applied to the set given by the Raman spectra of the reference pigments, with one high-quality Raman spectrum for each pigment, generating the reference spectral library. Nonetheless, it is needed to rely on mathematical tools that allow the spectral comparison between the unknown Raman spectrum and those of the reference spectral library. In consequence, in order to quantify the similarity between spectra of the Euclidean distance is chosen, and through the squared cosine the quality of the representation of a Raman spectrum once projected onto the reduced space is quantified.

The relations between the unknown Raman spectrum and the reference spectral library determine the candidate Raman spectra that may identify the unknown Raman spectrum. These relations are computed from the Euclidean distances and the squared cosines, bearing in mind that the lower the distance the more similar the Raman spectra, and the higher the squared cosine the better the representation in the reduced space. Thus, the identification criteria defined in this work allow recognizing the unknown Raman spectrum. In a few words: if the distance between the unknown Raman spectrum and a given reference Raman spectrum is lower than a certain boundary, then this given reference Raman spectrum matches the unknown Raman spectrum and the identification is achieved. Additionally, it is provided a reliability factor on the identification computed from the Euclidean distances as well. Consequently, the developed identification algorithm becomes a really useful tool to help the analyst to make a decision.

The whole method of pigment identification based on Raman spectroscopy and on the algorithm presented in this work may be seen as complete pattern recognition system (see Fig. 1).

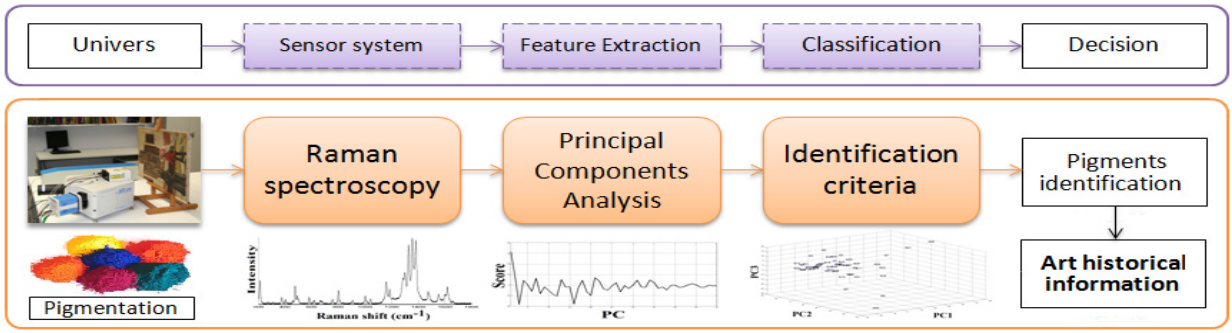

Fig. 1; General scheme of pattern recognition (purple) and its adaptation to pigments identification through Raman spectroscopy, principal components analysis and the identification algorithm (orange)

Results

A pigment identification example is shown in Fig. 2. After applying the implemented methodology, the system identifies the unknown spectrum measured on a painting as Indanthrene Blue, patented in 1909 by Rene Bohn as the first anthraquinone vat dye.

The proposed system has been proved in several experimental cases, verifying that allows the recognition of Raman spectra of artistic pigments.

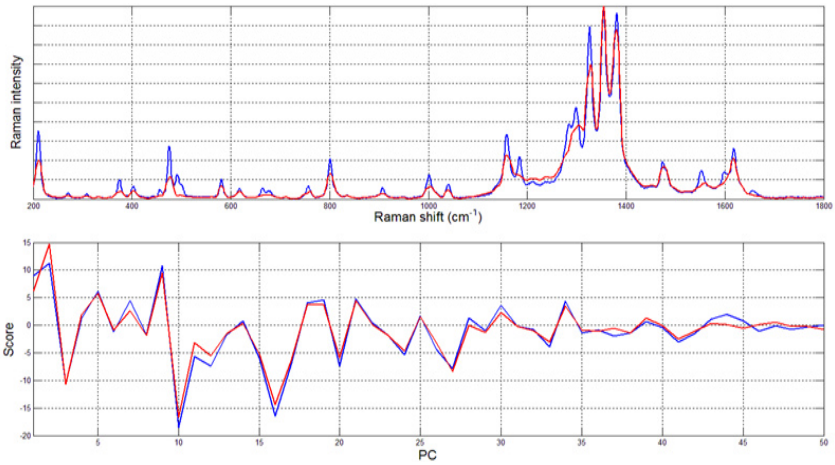

Fig. 2: Indanthrene Blue pigment (in blue) and unknown (in red) in spectral space at the top and in the reduced space at the bottom

Conclusions

The aim of this work was to explain a strategy to identify pigments in an automatic and systematic way. This strategy is based on the general scheme of pattern recognition and on a novel algorithm based on PCA and mathematical tools. The developed identification system has shown to be a really helpful tool in Raman spectroscopy applied to the analysis of artistic pigments, speeding up the pigments identification process.

References

1) J.M. Madariaga, Raman spectroscopy in art and archaeology, *J. Raman Spectrosc.*, 41(11), (2010) 1389-1393
2) H.G.M. Edwards, Analytical Raman spectroscopic discrimination between yellow pigments of the Renaissance, *Spectroc. Acta Pt. A-Molec. Biomolec. Spectr.*, 80(1), (2011) 14-20
3) P. Vandenabeele, Evaluation of a spectral searching algorithm for the comparison of Raman band positions, *Spectroc. Acta Pt. A-Molec. Biomolec. Spectr.*, 80(1), (2011) 27-31