

Automatic Classification System of Raman Spectra Applied to Pigments Analysis

Juan José González-Vidal^{(a,b)*}, Rosanna Pérez-Pueyo^(a), María José Soneira^(a)

^(a) Signal Theory and Communications Department, Universitat Politècnica de Catalunya, c/ Sor Eulàlia d'Anzizu s/n, D5, Campus Nord, 08034, Barcelona, Spain, e-mail: {rperez, soneira, juanjo.glezvidal}@tsc.upc.edu

^(b) DAPCOM Data Services, European Space Agency Business Incubation Centre Barcelona, Parc UPC-RDIT, c/ Esteve Terrades 1, 08860, Castelldefels, Barcelona, Spain, e-mail: juanjose.gonzalez@dapcom.es

* **Corresponding author:** Juan José González-Vidal

Address: Signal Theory and Communications Department, Universitat Politècnica de Catalunya, c/ Sor Eulàlia d'Anzizu s/n, D5, Campus Nord, 08034, Barcelona, Spain

Telephone: +34 934 016 442

Fax: +34 934 016 447

E-mail: juanjo.glezvidal@tsc.upc.edu (Juan José González-Vidal)

Abstract

Raman spectroscopy is one of the few non-destructive techniques capable of identifying pigments in art works. Raman spectra contain powerful information that can be used to identify unknown compounds and their chemical structures. However, the analysis of spectral data comes with some difficulties and therefore the spectral interpretation is not straightforward. Sometimes, there are very little differences in the spectral data concerning to specific identification objectives, for instance, in polymorphic discrimination or in the discrimination of natural and synthetic forms of certain pigments. Moreover, this discrimination is often performed manually so that the process can be repetitive, subjective, and particularly time-consuming. The result is an increasing motivation to automate the identification process involved in the classification of pigments in paint. In this paper, we propose a system to automatically classify the spectral data into specific and well-known classes, i.e. reference classes. The proposal is based on a combination of chemometric techniques, which provides a powerful way to achieve spectral separability so that it is possible to discriminate between very similar spectra in an automatic way. In this regard, a decision-making algorithm was specifically developed to select the corresponding reference class with no user input, which was successfully validated using simulated spectra. The implemented methodology was used to classify Raman spectra of pigments commonly present in artist's paints in experimental cases, providing reliable and consistent results. Therefore, the presented system can play a good auxiliary role in the analysts' endpoint classification.

Keywords: Pigments, automatic classification, Raman spectroscopy, principal component analysis, multiple discriminant analysis.

Introduction

Raman spectroscopy has proved to be a powerful technique for the non-destructive characterization of constitutive pigmentation in art works¹⁻⁵. In this sense, pigment identification is one of the main tasks for cataloguing, conserving and restoring works of art. It is well known that Raman spectroscopy is able to distinguish different molecular species based on the acquired Raman spectra. The discrimination between the pigments found in natural and synthetic forms⁶⁻⁸ or in different crystalline structures⁹⁻¹⁵ is an important topic in conservation science because the pigments may differ not only in their chemical and physical characteristics (such as stability, solubility and hue) but

also appeared at different times on the paint market and thus they may be used as chronological markers.

In art analysis, daily tasks concerning pigment identification are not automatized unlike in other scientific fields. The pigment discrimination is generally performed through visual comparison between Raman spectra generated on works of art with an appropriate set of reference Raman spectra. In general, these reference spectra are measured from pigment powders under certain measurement conditions. However, pigments in works of art are usually mixed with other materials (such as binding agents, varnishes, other pigments) so that a Raman spectrum measured directly on an analyzed art work can lose specific information due to different reasons. Indeed, these external agents may hamper the acquisition of the Raman spectra and may hinder their analysis. Moreover, measurement conditions may produce changes in the spectrum of a specific pigment which may lead to an erroneous interpretation. Furthermore, certain pigments can be found in different crystalline structures as the copper-phthalocyanine blue pigment for instance, and the differences in their spectral data may go unnoticed. Indeed, these little differences in the spectral data may occasionally lead to a subjective interpretation or to the need of aggregating data from different analytical methods, making the identification a costly and time-consuming process. Therefore, there is an increasing motivation to automate the human processes involved in the pigment classification. Thus, the development of classification tools that can help the analyst in making decisions has become a trending topic¹⁶⁻²¹.

Most papers that handle the classification issue are based on chemometrics where the identification features are manually retrieved from the spectra. As a result, a certain degree of subjectivity is still incorporated to infer the classification. Our premise, however, is that no user input should be required. This means that the process of assigning the class an unknown spectrum belongs should be fully automated. Hence, multivariate techniques were explored in this work in order to design an analytical method to automatically classify artistic pigments from their Raman spectra in a transparent way regarding the classification topic: the material's provenance, the crystalline structure, or any other classification matter.

One important difficulty which limits the classification of Raman spectra is that multiple users performing their measurements with different instruments can contribute to the same database. This may imply differences in bands intensity, in spectrometer resolution, and in signal pre-processing among others. Nevertheless, a good classification system should be able to deal with these differences and still be able to provide a reliable result. Of course, the successful outcome of these analyses is linked to the proper characterization of the reference classes according to the classification objective.

In this work, a combination of multivariate techniques was used to characterize the reference classes. Specifically, principal component analysis (PCA)²² was used to provide a set of features from the reference Raman spectra, and multiple discriminant analysis (MDA)²³ was used to find the best combination of features to separate the reference classes. In this way, the combination of PCA and MDA used by the methodology proposed here, provided a suitable classification space for Raman spectra. Consequently, an unknown spectrum projected onto this classification space may be classified through a decision-making procedure specifically developed for achieving a

proper classification. In other words, one may efficiently extract selective information in a fully automatic way from the reference spectra which best differentiates the classes and allows to establish a criterion to classify unknown spectra.

The paper proceeds as follows: Firstly, we describe the classification methodology. In particular, we illustrate the clustering property of combining PCA and MDA applied to Raman spectra and describe how to solve the classification of unknown spectra. Here, the main interest of the current paper is to find an automated classifier of Raman spectra whose reference classes have tiny differences. Next, we show how the presented method was validated through a simulation environment. Then, we address some experimental cases of pigment classification by means of the proposed system. Finally, we discuss some conclusions.

Methodology

The main objective of the presented methodology is to automatically classify unknown spectra according to predefined classes in a consistent way. To do so, it is necessary to rely on mathematical tools and on a specific classification strategy that allow the objective comparison between unclassified spectra and reference classes. In this regard, it is crucial to make some spectral pre-treatment to properly address this comparison.

Noise is inherent to the acquisition of a Raman spectrum. Even assuming that it was collected under optimal conditions, a denoising should be performed to enhance the Raman information as much as possible. Furthermore, a uniform data format smoothes the progress of comparisons between unknowns and references. Consequently, a three-step pre-processing sequence must be followed in order to ensure the success of the automated classification. Baseline correction is the first pre-processing step. In this work, the methodology for baseline removal proposed in²⁴ was used. The second pre-processing step is interpolation, so that all spectra are stored in a compatible way. The interpolation ensures that all spectra have a common set of Raman shifts, which is crucial when spectra collected with different measurement systems are used. Finally, the last pre-processing step is normalization of intensities, which reduces the impact of measurement conditions so that the outcome of the classification is independent of the acquisition instrument. The normalization here used was the min-max normalization, where the minimum intensity is scaled to 0 and the maximum to 1, meaning that a normalized spectrum maintains the relative ratio between its Raman bands. These pre-processing steps ensure that all the spectra are baseline-corrected and fulfill a set of homogeneity conditions with respect to data format. As a result, the automatic classification of unknown spectra by means of the proposed methodology may be properly carried out.

The classification of artistic pigments fits the standard scheme of statistical classification²⁵, which is a supervised learning technique in the field of machine learning and statistics. It deals with the process of identifying to which of a set of classes an unclassified item belongs to, based on a training dataset containing references whose class membership is known beforehand. The standard classification scheme is built from two different stages: data acquisition and data processing. In the case of pigments analysis through Raman spectroscopy, the data acquisition stage is based on the Raman spectrometer. On the other hand, the data processing stage is composed of three different modules: *feature extraction*, *classifier* and *decision-maker*. First of all, the *feature extraction* is the process of defining a set of features, which most effectively

represent the important information for classification. We selected principal component analysis (PCA) for this purpose, as it is the technique that best fits the data dimensionality²⁶ requirements for the current work. Then, the *classifier* is the multivariate technique aimed at maximizing the inter-class distances whilst minimizing the intra-class differences from an appropriate set of class features. We selected multiple discriminant analysis (MDA) for finding a combination of features that separates the user-defined classes, i.e. training dataset. From the set of extracted features by PCA, MDA provides a new space, the so-called classification space. Finally, the *decision-maker* is the procedure in which an unknown or unclassified element is projected onto the classification space and is assigned to one of the classes according to some metrics that will be discussed hereafter.

Next, we describe the characterization of the training dataset in the classification space and the procedure of class assignment for unknown spectra.

Characterization of the classification space

In the case of pigments analysis through Raman spectroscopy, the training dataset is composed of sets of reference Raman spectra, i.e. reference classes. These reference classes are decided by the user according to the classification purpose. The training dataset is represented by a matrix, S , which is divided in sub-matrices. Each sub-matrix S_i identifies a known class where each row is a spectrum of the i -th reference class. The classification space is obtained by applying first PCA (*feature extractor*) to the training dataset and later MDA (*classifier*) over the PCA result. In this space, the training dataset for the i -th class is now represented by a matrix C_i where each row is a spectrum in the classification space. Each class is delineated by a region and is characterized by a centroid (the arithmetic center, μ_i) and a dispersion matrix (the auto-covariance matrix, Σ_i).

In order to perform an efficient classification, proper class separability in the classification space should be obtained. This class separability is checked by computing the Jeffries-Matusita distance, JMD ^{27,28} (see Supporting Information), which ranges from 0 to 2. The classes in the classification space are totally distinguishable when JMD is equal to 2 while lower values indicate a worse separability. We use the class separability as a parameter to generate an adequate classification space, selecting a proper number of features. As we said previously, the PCs scores obtained in the *feature extraction* module are the distinctive features for each class. Then, we tune the number of PCs scores, successively until there is no improvement in the class separability. We considered a JMD value greater than 1.75 for achieving good class separability. In this way, starting from a number of PCs scores equals to the minimum number of spectra in $C_i \forall i$, the JMD is calculated in the tentative classification space obtained by performing PCA followed by MDA. The number of PCs scores used is increased by one until the desired JMD value is achieved or until the number of PCs scores is greater to the maximum number of spectra in $C_i \forall i$. If the class separability is achieved through the obtained number of PCs scores, it means that the user-defined reference classes allow the classification of unknown spectra. Otherwise, no class separability is achieved with the defined classes and must be re-defined. The organization chart of this procedure is outlined in Fig. 1.

Additionally, taking into account that outliers in a class can deform the class characterization, a basic statistical rule for outlier rejection is applied to each reference

class in the classification space. This rule is defined as: if $x > \mu_i + 2\Sigma_i$ then x is rejected (being x a spectrum of the i -th class in the classification space). When a reference spectrum is rejected, the class parameters (centroid and dispersion matrix) are automatically recomputed for that class.

Class assignment

Once the classification space is characterized, we defined a classification rule to assign an unknown spectrum to a reference class. For this purpose, we developed an autonomous matching technique based on distance metrics. Specifically, the Euclidean and Mahalanobis distances were used (see Supporting Information). First, we compute the Euclidean distance, which provides a measure of the distance between an unknown spectrum in the classification space (x) and the class centroid (μ_i). Second, we calculate the Mahalanobis distance, which provides a measure of the distance between an unknown spectrum in the classification space (x) and the class (\mathbf{C}_i). These metrics express intuitive notions about the concept of distance. While the Euclidean distance expresses how far apart an unknown spectrum and the center of a class are, the Mahalanobis distance takes into account the class dispersion and expresses how far apart an unknown spectrum and a class region are. Bearing in mind these meanings we define a classification distance combining the results of these two different distances, specifically:

$$CD_i(x) = ED(x, \mu_i) \cdot MD(x, \mathbf{C}_i)$$

In order to classify an unknown spectrum we define a toolkit based on the above distance which allows to explore the matching of the unknown to a class. To do so, we firstly compute the classification distance (CD) between classes, the so-called *InterClassCD*, which provides a notion on how close the classes are. Note that there are as many values of *InterClassCD* as defined classes and the minimum value is due to the closest classes. Also, we calculate the so-called *IntraClassCD*, which provides an idea on how close a spectrum is to its own class. The farthest spectrum from a given class gives the maximum value of *IntraClassCD* for that class. Then, the classification of an unknown spectrum is performed by exploring the classification distance between the unknown and each class. The assignment of the unknown spectrum to a reference class is performed by a matching function defined as:

$$MF_i(x) = \begin{cases} 1 & \text{if } CD_i(x) \leq \max IntraClassCD_i \\ \left(1 - \frac{CD_i(x)}{\min InterClassCD}\right) & \text{if } CD_i(x) \leq \min InterClassCD \\ 0 & \text{otherwise} \end{cases}$$

where $CD_i(x)$ is the minimum value of CD between the unknown (x) and the reference classes, $\min InterClassCD$ is the minimum value of *InterClassCD* and $\max IntraClassCD_i$ the maximum value of *IntraClassCD* for the i -th class. The matching function expressed in % is intended to help the analyst in the decision-making process.

The methodological scheme of the classification system is shown summarized in Fig. 2. It illustrates the standard classification design together with the approach proposed in the current work.

Results and discussion

Verification and validation

In order to diagnose the performance of the proposed methodology, it was analyzed in a simulation stage. This study was performed using simulated Raman spectra, specifically, fluorescence-free simulated spectra. Briefly, a simulated spectrum was generated by combining a variable number of Lorentzian peaks with random locations, amplitudes, and bandwidths (full width half-maximum (FWHM)), although we constrained the variables such that the spectra appeared qualitatively similar to those found in real Raman spectra (baseline subtracted). In particular, three different classes were created. To do so, three different spectra were generated, which simulated spectra measured from three different pigments. The only difference between these spectra was the amplitude of two selected bands between 650 cm^{-1} and 800 cm^{-1} , as can be seen in Fig. S1 (Supporting Information). Then, ten different simulated spectra were generated for each class, simulating different realizations for each of the three pigments. These different realizations were generated through random variations in band locations, amplitudes and bandwidths. Specifically, normal distribution functions were used giving random variations of $\pm 5\text{ cm}^{-1}$ in band locations, ± 0.05 a.u. in normalized intensities, and $\pm 2\text{ cm}^{-1}$ in bandwidths (see inset figure in Fig. S1 (Supporting Information)).

With the simulated classes, we applied the Lachenbruch procedure (also called *leave-one-out* cross-validation)²⁹, which is a standard model validation technique for assessing the predictive performance of a methodology. It involves using one spectrum as the test set (for which we certainly know the corresponding class) and the remaining spectra as the training dataset. Specifically, based on this cross-validation procedure, we applied the following five-step sequence:

- 1) Let the i -th spectrum form the test set (test spectrum)
- 2) Get the classification space using the remaining spectra (29 spectra)
- 3) Apply the classification criterion for class assignment on the test spectrum
- 4) Check the classification outcome with respect to the expected class
- 5) Repeat step 1 for $i=1, \dots, n$ with n being the total number of spectra ($n=30$)

We obtained a success rate of 100%, which shows a good predictive performance of the presented methodology in an under-controlled environment using simulated data.

Experimental cases

In order to show the performance of the implemented methodology in experimental environments we applied the developed classification system to unknown Raman spectra acquired from hand-made paints²⁰ and art works. The spectra for the reference classes were acquired from reference pigment powders. The experimental spectra used in this work that were measured by the authors were recorded using the portable Raman equipment iHR320 (Horiba Jobin Yvon). The optical source employed for spectral acquisition was a He-Ne laser (632.8 nm), which provided approximately 17 mW. The light from the laser was guided using an optical fiber to the optical head and directed to the sample. The scattered light was collected and filtered using the corresponding edge filter inside the optical head. It was then guided using an optical fiber to the monochromator and detected by a thermoelectrically cooled charge-coupled device (CCD). Acquisition times were around 30 s with five accumulations (150 s) on inorganic pigments and 300 s with five accumulations (1500 s) for the organic pigments to achieve the best trade-off between signal-to-noise ratio in spectra from the sample and measurement time.

No assumptions regarding the input data are made by the classification system, which processes the data blindly through the presented automatic approach in a fully transparent way. To highlight this point, two experimental cases are presented hereafter with completely different classification purposes.

In the first experimental case, we distinguish among ultramarine blue pigment in its natural form (as lapis lazuli) and in its synthetic form. Therefore, two reference classes were built. The natural form class was composed of six spectra acquired from Afghan, Siberian and Chilean lapis lazuli samples. The synthetic form class was composed of six spectra as well, which were acquired from several synthetic ultramarine blue pigment powders manufactured by Nubiola. The *Feature Extraction* module provided a 6-dimensional PCs space with an accumulative variance of 99.5%. The classification space is described by a straight line with two separated regions (one for each class) with a JMD equals to 2. The classification methodology was applied to twelve unknown spectra measured in our laboratory from different art works (see Fig. 3). Specifically, one of these unknown spectra was acquired from a Chilean art figure whilst the remaining unknown spectra were measured from different oil paintings. Fig. 4 shows the projection of the unknown spectra onto the classification space. The classification results are reported in Table SI (Supporting Information). The consistency of the results was assessed by inspection of the measured areas using a Leica MZ-12 stereomicroscope with a photomicrographic resolution of 600 magnifications. All the unknown spectra were successfully classified although with different value of the matching function, ranging from 47.32% to 100%. Specifically, the spectrum classified with the minimum matching value (painting 7 in Fig. 3) was deeply affected by undesired artifacts (unknown peaks) in the pre-processed spectrum.

In the second example, we distinguish between three different crystalline structures of copper-phthalocyanine blue pigment, specifically α -, β - and ϵ -modifications. Thus, three reference classes were built. Additionally to the reference spectra recorded by the authors from pigment powders supplied by different manufacturers (Kremer, Sennelier and Mongay), a significant set of reference spectra were supplied by three different researchers and therefore recorded using different acquisition systems under different measurement conditions.

It is well-known that there may be differences between Raman spectra recorded with different instruments, which may become a handicap for the purposes of pigment classification. Indeed, instrument resolution, excitation wavelength or even laser power of the excitation source can strongly influence the Raman bands as stated in³⁰. Nevertheless, the classification methodology presented here is not affected by these issues as illustrated by the results shown hereafter as long as the user-defined reference classes are properly defined and represented in the classification space: the implemented system automatically picked up the spectral markers for classification by means of the PCs scores regardless of the heterogenic input data and discriminated the CuPc classes in the classification space.

Specifically, the α -modification class consisted of 27 spectra: nine spectra recorded using a 532 nm excitation wavelength, ten spectra recorded using a 633 nm excitation wavelength, and eight spectra recorded using a 785 nm excitation wavelength. The β -modification class consisted of 38 spectra: eleven spectra were recorded using a 532 nm excitation wavelength, thirteen spectra recorded using a 633 nm excitation wavelength,

and fourteen spectra recorded using a 785nm excitation wavelength. Finally, the ϵ -modification class consisted of 14 spectra: ten spectra were recorded using a 532 nm excitation wavelength, one spectrum recorded using a 633 nm excitation wavelength, and three spectra recorded using a 785nm excitation wavelength. The *feature extraction* module provided a 23-dimensional PCs space with an accumulative variance of 99.19%. The classification space is described by three class regions (one for each reference class) with a minimum JMD of 1.99. Table SII (Supporting Information) presents the classification results for Raman spectra supplied by Marta Anghelone (Academy of Fine Arts Vienna) measured on hand-made samples. These hand-made samples were prepared with PB15:1, PB15:3 and PB15:6 pigment powders manufactured by Kremer Pigments, which were mixed with several binding agents in different proportions and subjected to a UV aging process as reported in²⁰. We applied the presented methodology to a total of 36 spectra (see Fig. 5) and we obtained a success rate of 100%, showing the consistency of the implemented classification system. Fig. 6 shows the projection of the unknown spectra onto the classification space. The spectra classified with lower matching values were affected by band shifting, band spreading and intensity inversions.

Conclusion

In the present contribution, a methodology to automatically distinguish between Raman spectra showing small differences was presented. According to predefined reference classes, the method is able to classify unknown spectra from a single spectral observation, with no user input or previous knowledge of the analyzed sample. The developed model is based on automated matching of unclassified spectra using PCA and MDA, and it is computationally efficient and conceptually simple. The results showed that the method is suitable for art works analysis as it successfully classified the analyzed Raman spectra in a consistent way. Moreover, the implemented method is an easy-to-use system and it is straightforward to update when new spectral data become available.

The implemented classification system has been applied to experimental Raman spectra, and the obtained results showed that it may play a good auxiliary role in the analysts' endpoint classification. Therefore, the system may become a useful tool to help in the decision-making process, in order to ease the management of pigment classification from Raman spectra whose reference classes are very similar.

Finally, it is worth noting that the methodology makes no assumptions with respect to the input data, applying a blind treatment of the Raman spectra and processing them in a transparent way regardless of the classification purposes. Consequently, it is perfectly capable of dealing with spectra from different sources, i.e. recorded with different acquisition systems and measurement conditions. This fact may represent a significant advantage of the presented automated system in the application of pigment classification in art analysis through Raman spectroscopy, as it is independent of the measurement system and the configuration used for the acquisition of Raman spectra.

Acknowledgements

The authors are very grateful to Wim Fremout and Steven Saverwyns from the Royal Institute for Cultural Heritage, Nadim C. Scherrer from the Bern University of Applied Sciences, and Marta Anghelone from the Academy of Fine Arts Vienna for kindly

providing a really valuable set of Raman spectra from copper-phthalocyanine blue pigments that were used in the current work.

References

- [1] H. G. M. Edwards, P. Vandenabeele, J. Jehlička, T. J. Benoy, *Spectrochim. Acta, Part A*. **2014**; 118, 598.
- [2] L. Medeghini, P. P. Lottici, C. De Vito, S. Mignardi, D. Bersani, *J. Raman Spectrosc.* **2014**; 45, 1244.
- [3] S. A. Centeno, DOI: 10.1002/jrs.4767.
- [4] M. C. Caggiani, C. Valotteau, P. Colomban, *J. Raman Spectrosc.* **2014**; 45, 456.
- [5] C. Conti, J. Striova, I. Aliatis, E. Possenti, G. Massonnet, C. Muehlethaler, T. Poli, M. Positano, *J. Raman Spectrosc.* **2014**; 45, 1186.
- [6] M. Bacci, C. Cucci, E. Del Federico, A. Ienco, A. Jerschow, J. M. Newman, M. Picollo, *Vib. Spectrosc.* **2009**; 49, 80.
- [7] E. M. A. Ali, H. G. M. Edwards, *Spectrochim. Acta Part A*. **2014**; 121, 415.
- [8] A. R. De Torres, S. Ruiz-Moreno, A. López-Gil, P. Ferrer, and M. C. Chillón, *J. Raman Spectrosc.* **2014**; 45, 1279.
- [9] S. Švarcová, Z. Čermáková, J. Hradilová, P. Bezdička, D. Hradil, *Spectrochim. Acta Part A*. **2014**; 132, 514.
- [10] G. Beaulieu-Houle, D. F. R. Gilson, I.S. Butler, *Spectrochim. Acta Part A*, **2014**; 117, 61.
- [11] S. Quillen Lomax, J. F. Lomax, A. De Luca-Westrate, *J. Raman Spectrosc.* **2014**; 45, 448.
- [12] L. B. Brostoff, S. A. Centeno, P. Ropret, P. Bythrow, F. Pottier, *Analytical Chemistry*, **2009**; 81(15), 6096.
- [13] C. Defeyt, P. Vandenabeele, B. Gilbert, J. Van Pevenage, R. Cloots, D. Strivay, *J. Raman Spectrosc.* **2012**; 43(11), 1772.
- [14] W. Fremout, S. Saverwyns, *J. Raman Spectrosc.* **2012**; 43(11), 1536.
- [15] N.C. Scherrer, Z. Stefan, D. Francoise, F. Annette, K. Renate, *Spectrochim. Acta Part A*, **2009**; 73, 505.
- [16] M. Hoehse, A. Paul, I. Gornushkin, U. Panne, *Anal. Bioanal. Chem.* **2012**; 402, 1443.
- [17] M. R. De Almeida, D. N. Correa, W. F. C. Rocha, F. J. O. Scaffi, R. J. Poppi, *Microchem. J.* **2013**; 109, 170.
- [18] G. Barone, D. Bersani, J. Jehlička, P. P. Lottici, P. Mazzoleni, S. Raneri, P. Vandenabeele, C. Di Giacomo, G. Larinà, *J. Raman Spectrosc.* **2015**; 46, 989–995.
- [19] C. Defeyt, J. Van Pevenage, L. Moens, D. Strivay, P. Vandenabeele, *Spectrochim. Acta Part A*. **2013**; 115, 636.
- [20] M. Anghelone, D. Jembrih-Simbürger^a, M. Schreiner, *Spectrochim. Acta Part A*, **2015**; 149, 419.
- [21] F. de Souza, L. Borba, R. S. Honorato, A. de Juan, *Forensic Sci. Int.* **2015**; 249, 73.
- [22] K. Pearson, *Philos. Mag.* **1901**; 2(11), 559.
- [23] R. A. Fisher, *Annals of Eugenics.* **1936**; 7(2), 179.
- [24] R. Perez-Pueyo, M. J. Soneira, S. Ruiz-Moreno, *App. Spectrosc.* **2010**; 64, 595.
- [25] R. O. Duda and E. H. Shortliffe, *Science.* **1983**; 220, 4594.

- [26] L. J. P. van der Maaten, E.O. Postma, H.J. van den Herik, *J. Machine Learning Research*. **2009**; 10, 66.
- [27] I. L. Thomas, et al., *Int. J. Remote sensing*. **1987**; 8, 331.
- [28] N. A. Thacker, F. J. Aherne and P. I. Rockett, *Kybernetika*, **1997**; 34, 363.
- [29] Y. Zhang, Y. Yang, *J. Econometrics*, **2015**; 187, 95.
- [30] N. C. Scherrer, *Book of Abstracts: RAA2011*, (Eds: Bologna, ISBN: 978-88-97162-20-9), Parma (Italy). **2011**, pp. 203-204.

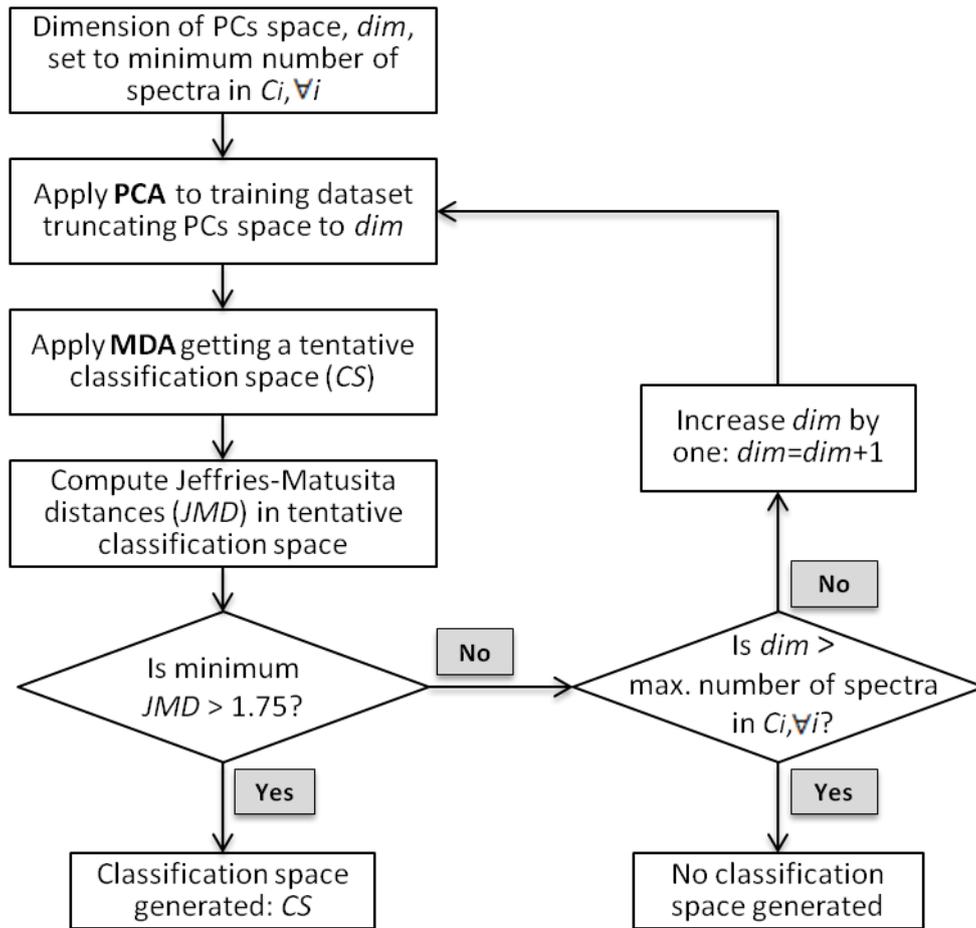


Figure 1: Classification space generation from training dataset

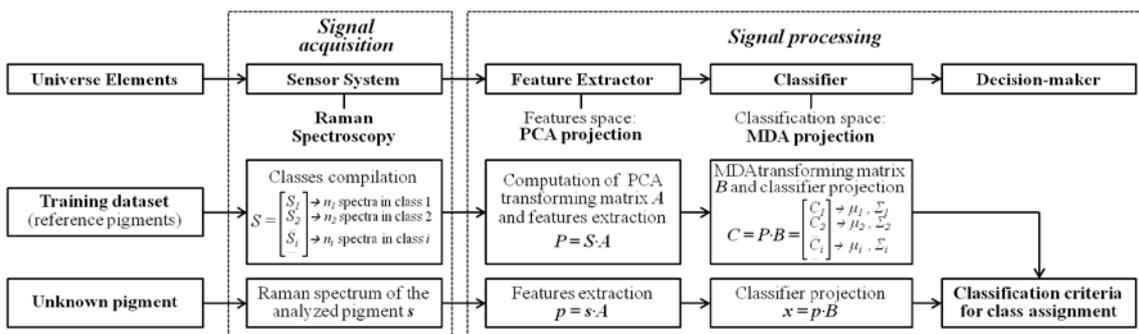


Figure 2: Overview of the classification scheme

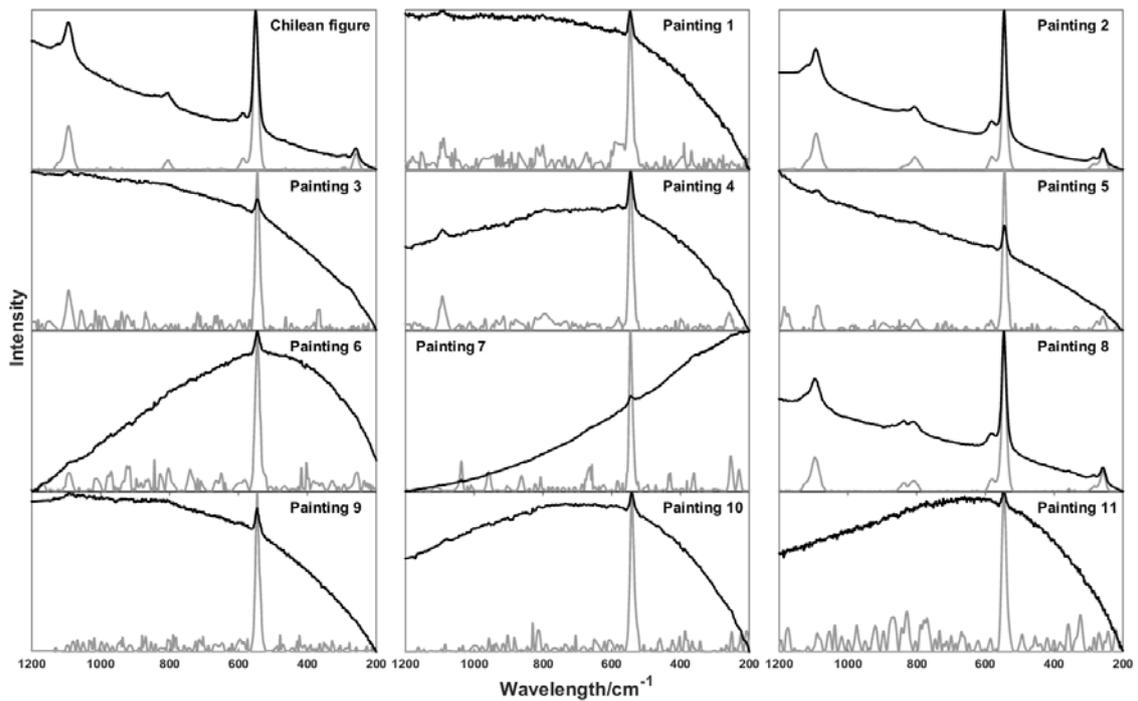


Figure 3: Experimental Raman spectra from ultramarine blue measured on a Chilean art figure and oil paintings: acquired spectra (black) and pre-processed spectra (gray)

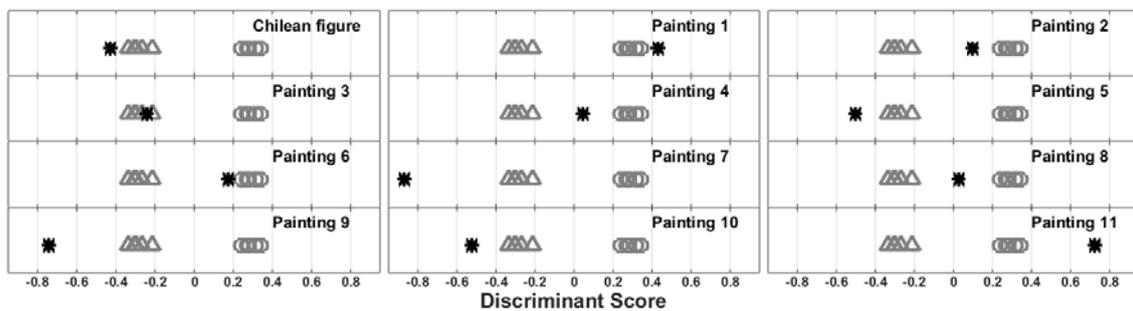


Figure 4: Projection of experimental Raman spectra from ultramarine blue onto the classification space: natural form class (triangles), synthetic form class (circles) and unknowns (asterisks)

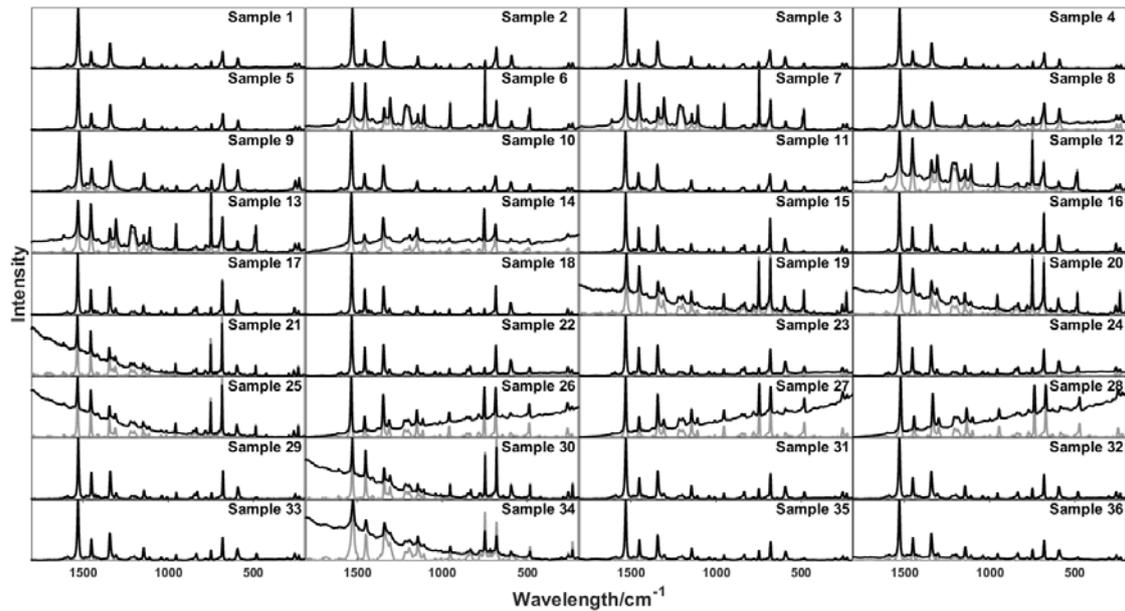


Figure 5: Experimental Raman spectra from copper-phthalocyanine blue measured on hand-made samples: acquired spectra (black) and pre-processed spectra (gray)

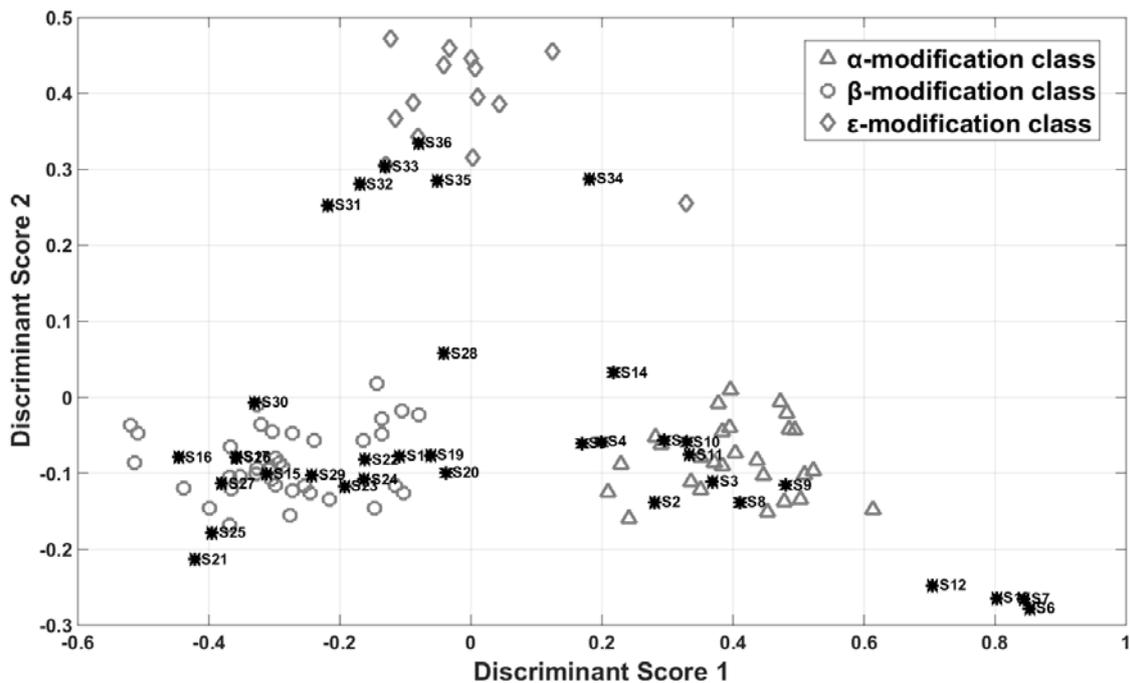


Figure 6: Projection of experimental Raman spectra from copper-phthalocyanine blue onto the classification space: α -modification class (triangles), β -modification class (circles), ϵ -modification class (diamonds) and unknowns (asterisks)