UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR

INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN

EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES

Automatic Identification of Raman Spectra

Autor:

Isaac Hermosilla Rodríguez

Tutor:

Diego R. Llanos Ferraris, Fernando Rull Pérez

Valladolid, 10 de Septiembre de 2012

TTULO:	**Automatic Identification of Raman Spectra**
AUTOR:	**Isaac Hermosilla Rodríguez**
TUTOR:	**Diego R. Llanos Ferraris, Fernando Rull Pérez**
DEPARTAMENTO:	

Tribunal

PRESIDENTE:	**Dr. D.**
VOCAL:	**Dr. D.**
SECRETARIO:	**Dr. D.**
FECHA:	**10 de Septiembre de 2012**
CALIFICACIÓN:	

Resumen del TFM

Este trabajo consiste en un estado del arte sobre la identificación de compuestos a través de espectroscopía Raman. La espectroscopía Raman estudia la luz reflejada por un sistema cuando se incide sobre ella una haz de luz monocromático, de esta manera se conocen características químicas y estructurales del sistema. También contiene una propuesta de un algoritmo para identificar minerales a través de su espectro de Raman con una base de datos y una comparación con otro algoritmo recientemente publicado.

Palabras clave

Espectroscopía Raman, identificación de minerales.

Abstract

This work consist on a state of the art about compounds identification through raman spectroscopy. Raman spectroscopy studies the light reflected by a system when it is irradiated by a beam of monochromatic light, thus chemical and physical characteristics of the system are obtained. Also it has a proposal of an algorithm to identify minerals through its raman spectrum with a database and a comparison with another algorithm recently published.

Keywords

Raman spectroscopy, mineral identification

Agradecimientos

Quiero agradecer la ayuda y el tiempo prestado a mis tutores, Diego y Fernando, y al grupo de la Unidad Asociada UVA - CSIC a través del Centro de Astrobiología.

Contents

List of Figures

Chapter 1

Introduction

1.1 Resumen

La espectroscopia Raman es una técnica fotónica que permite conocer información qumica y estructural de casi cualquier material orgánico o inorgánico. El objetivo es analizar la luz dispersada por un material cuando es iluminado con un haz de luz monocromático. La mayor parte de la luz que incide sobre el material se dispersa en la misma frecuencia que la luz irradiada, y no da ninguna información acerca del material, pero el resto se dispersa en una frecuencia diferente de la luz irradiada y proporciona información sobre la composición molecular de la muestra. El principal objetivo es identificar automáticamente minerales a través de su espectro Raman. Este objetivo es aun más complicado cuando se trata de reconocer mezclas de distintos materiales en diferentes proporciones. En este trabajo se hará una descripción del estado del arte sobre la identificacin de distintos compuestos y se hará una propuesta de algoritmo para identificar minerales.

1.2 Importance of problem

The main problem is identify samples of minerals through its raman spectra. It is not an easy task due to the variability between two spectra of the same sample taken in different conditions, with different equipment... The problem is even harder when the spectrum corresponds to a mixture of more than one mineral.

The search of coincidences between spectra may be a tedious task especially when it is necessary make the comparison with a large database. This is a common task in the work of a specialist. So, the possibility of developing a reliable tool that makes easier this task is an important challenge. Besides, a tool of this kind, not only is useful for specialists, also it can be used by non experts and can be included inside a portable spectrometer.

1.3 Brief description of Raman Spectroscopy

Raman spectroscopy is a spectroscopic technique used to study vibrational, rotational, and other low-frequency modes in a system. It is widely used to provide information on chemical structures and physical forms, to identify substances from the characteristic

spectral patterns (fingerprinting), and to determine quantitatively or semi-quantitatively the amount of a substance in a sample. So, thinking in raman spectra like a fingerprint it is possible identify materials through the study of their raman spectra.

1.4 Aims

The main aims of this work are:

- Make a state of the art about recognition of different compounds through raman spectroscopy.
- Make an initial purpose to recognition of minerals.
- Develop an initial tool that gives clues and makes easier the decision-making process.
- Make initial tests with a large corpus.
- Make a comparison between our purpose and a method for identification that has been published.
- Lay the basis for a more accurate algorithm and for an algorithm for portable devices.

Chapter 2

Raman Spectroscopy

2.1 Resumen

La fascinación por el azul del mar Mediterráneo llevó a C.V. Raman a investigar la dispersión de la luz por los líquidos y así descubrir experimentalmente la dispersión de luz con cambio de frecuencia.

La espectroscopía Raman es una técnica fotónica que permite conocer información química y estructural de casi cualquier material orgánico o inorgánico.

El objetivo es analizar la luz dispersada por un material cuando es iluminado con un haz de luz monocromático. La mayor parte de la luz que incide sobre el material se dispersa en la misma frecuencia que la luz irradiada, y no da ninguna información acerca del material, pero el resto se dispersa en una frecuencia diferente de la luz irradiada y proporciona información sobre la composición molecular de la muestra.

La espectroscopía Raman tiene como ventaja la mínima manipulación y preparación de la muestra que se requiere.

2.2 Historical Background of Raman Spectroscopy

Curiosity about the explanation of the blue colour of the sky led Lord Rayleigh to formulate a classical theory of light scattering without change of frequency (Rayleigh, 1871) [15]. Fascination with the marvellous blue of the Mediterranean sea caused C. V. Raman to investigate the scattering of light by liquids and so to discover experimentally the scattering of light with change of frequency (Raman and Krishnan, 1928). An independent prediction of this phenomenon had been made a few years earlier (Smekal, 1923) using classical quantum theory.

Shortly after Raman and Krishnans publication a report of light scattering with change of frequency in quartz was reported by two Russian scientists, Landsberg and Mandelstam (1928), and in France, Raman and Krishnans observations were soon confirmed by Cabannes (1928) and Rocard (1928). The potential of the Raman effect in chemistry and physics was realized very rapidly. By the end of 1928 some 70 papers on the Raman effect had been published.

In 1928, when Sir Chandrasekhra Venkata Raman discovered the phenomenon that bears his name, only crude instrumentation was available. Sir Raman used sunlight as

the source and a telescope as the collector, the detector was his eyes. That such a feeble phenomenon as the Raman scattering was detected was indeed remarkable. [11]

Gradually, improvements in the various components of Raman instrumentation took place. Early research was concentrated on the development of better excitation sources. Various lamps of elements were developed as helium, bismuth, lead, zinc, but these proved to be unsatisfactory because of low light intensities.

In 1962 laser sources were developed for use with Raman spectroscopy . Eventually, the Argon and the Krypton lasers became available, and more recently the Nd- YAG laser has been used for Raman spectroscopy.

Progress occurred in the detection systems for Raman measurements. Whereas original measurements were made using photographic plates with the cumbersome development of photographic plates, photoelectric Raman instrumentation was developed after World War II.

These developments in Raman instrumentation brought commercial Raman instruments to the present state of the art of Raman measurements. Now, Raman spectra can also be obtained by Fourier transform (FT) spectroscopy. FT-Raman instruments are being sold by all Fourier transform infrared (FT-IR) instrument makers, either as interfaced units to the FT-IR spectrometer or as dedicated FT-Raman instruments.

2.3 Theory of Raman Scattering

The main spectroscopies employed to detect vibrations in molecules are based on the processes of infrared absorption and Raman scattering. They are widely used to provide information on chemical structures and physical forms, to identify substances from the characteristic spectral patterns (fingerprinting), and to determine quantitatively or semi-quantitatively the amount of a substance in a sample. Samples can be examined in a whole range of physical states; for example, as solids, liquids or vapours, in hot or cold states, in bulk, as microscopic particles, or as surface layers. The techniques are very wide ranging and provide solutions to a host of interesting and challenging analytical problems. Raman scattering is less widely used than infrared absorption, largely due to problems with sample degradation and fluorescence. However, recent advances in instrument technology have simplified the equipment and reduced the problems substantially. These advances, together with the ability of Raman spectroscopy to examine aqueous solutions, samples inside glass containers and samples without any preparation, have led to a rapid growth in the application of the technique. [25]

Raman spectroscopy is a spectroscopic technique used to study vibrational, rotational, and other low-frequency modes in a system [8]. It relies on inelastic scattering, or Raman scattering, of monochromatic light, usually from a laser in the visible, near infrared, or near ultraviolet range. The laser light interacts with molecular vibrations, phonons or other excitations in the system, resulting in the energy of the laser photons being shifted up or down. The shift in energy gives information about the vibrational modes in the system.

When light interacts with matter, the photons which make up the light may be absorbed or scattered, or may not interact with the material and may pass straight through it. If the energy of an incident photon corresponds to the energy gap between the ground

state of a molecule and an excited state, the photon may be absorbed and the molecule promoted to the higher energy excited state. It is this change which is measured in absorption spectroscopy by the detection of the loss of that energy of radiation from the light. However, it is also possible for the photon to interact with the molecule and scatter from it. In this case there is no need for the photon to have an energy which matches the difference between two energy levels of the molecule. The scattered photons can be observed by collecting light at an angle to the incident light beam, and provided there is no absorption from any electronic transitions which have similar energies to that of the incident light, the efficiency increases as the fourth power of the frequency of the incident light.

Scattering is a commonly used technique. For example, it is widely used for measuring particle size and size distribution down to sizes less than 1 μ m. One everyday illustration of this is that the sky is blue because the higher energy blue light is scattered from molecules and particles in the atmosphere more efficiently than the lower energy red light. However, the main scattering technique used for molecular identification is Raman scattering.

Raman spectroscopy uses a single frequency of radiation to irradiate the sample and it is the radiation scattered from the molecule, one vibrational unit of energy different from the incident beam, which is detected. Thus, Raman scattering does not require matching of the incident radiation to the energy difference between the ground and excited states. In Raman scattering, the light interacts with the molecule and distorts (polarizes) the cloud of electrons round the nuclei to form a short-lived state called a virtual state., This state is not stable and the photon is quickly re-radiated.

The energy changes we detect in vibrational spectroscopy are those required to cause nuclear motion. If only electron cloud distortion is involved in scattering, the photons will be scattered with very small frequency changes, as the electrons are comparatively light. This scattering process is regarded as elastic scattering and is the dominant process. For molecules it is called Rayleigh scattering. However, if nuclear motion is induced during the scattering process, energy will be transferred either from the incident photon to the molecule or from the molecule to the scattered photon. In these cases the process is inelastic and the energy of the scattered photon is different from that of the incident photon by one vibrational unit. This is Raman scattering. It is inherently a weak process in that only one in every 10^6 - 10^8 photons which scatter is Raman scattered. In itself this does not make the process insensitive since with modern lasers and microscopes, very high power densities can be delivered to very small samples but it is does follow that other processes such as sample degradation and fluorescence can readily occur.

Since the virtual states are not real states of the molecule but are created when the laser interacts with the electrons and causes polarization, the energy of these states is determined by the frequency of the light source used. The Rayleigh process will be the most intense process since most photons scatter this way. It does not involve any energy change and consequently the light returns to the same energy state. The Raman scattering process from the ground vibrational state m leads to absorption of energy by the molecule and its promotion to a higher energy excited vibrational state (n). This is called Stokes scattering. Scattering from these states to the ground state m is called anti-Stokes scattering and involves transfer of energy to the scattered photon. The relative intensities of the

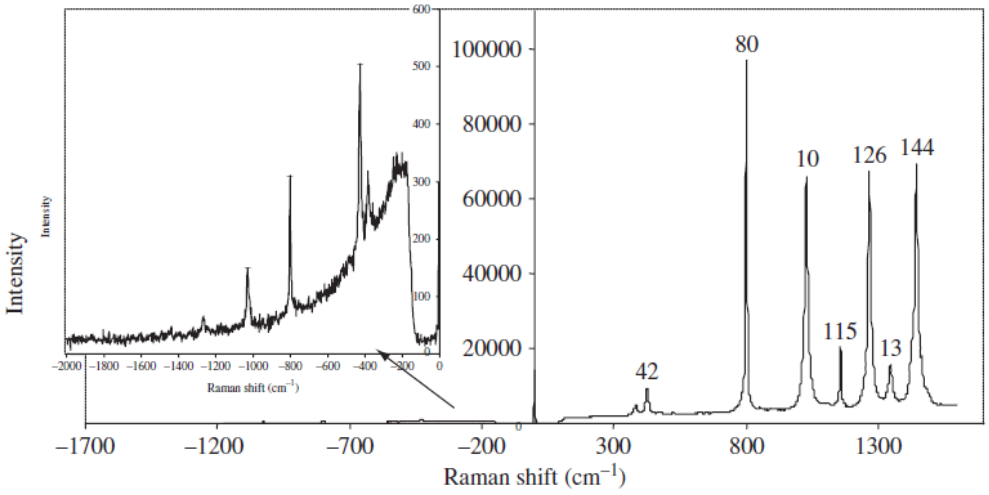


Figure 2.1: Stokes and anti-Stokes scattering for cyclohexane. To show the weak anti-Stokes spectrum, the y-axis has been extended in the inset.

two processes depend on the population of the various states of the molecule. Thus, compared to Stokes scattering, anti-Stokes scattering will be weak and will become weaker as the frequency of the vibration increases, due to decreased population of the excited vibrational states. Further, anti-Stokes scattering will increase relative to Stokes scattering as the temperature rises. Figure 2.1 shows a typical spectrum of Stokes and anti-Stokes scattering from cyclohexane separated by the intense Rayleigh scattering which should be offscale close to the point where there is no energy shift. However there is practically no signal close to the frequency of the exciting line along the x-axis. This is because filters in front of the spectrometer remove almost all light within about 200 cm1 of the exciting line. Some breakthrough of the laser light can be seen where there is no energy shift at all.

The scattering is measured as light detected by the spectrometer and the maximum amount of light detected is the highest point on the trace.

Strictly speaking, Raman scattering should be expressed as a shift in energy from that of the exciting radiation and should be referred to as $\Delta \mathrm{cm}^{-}1$ but it is often expressed simply as $\mathrm{cm}^{-}1$.

In the spectrum of the scattered radiation, the new frequencies are termed Raman lines, or bands, and collectively are said to constitute a Raman spectrum. Raman bands at frequencies less than the incident frequency are referred to as Stokes bands, and those at frequencies greater than the incident frequency as anti-Stokes bands.

2.3.1 Raman sample preparation and handling

Raman spectroscopy, as a scattering technique, is well known for the minimum of sample handling and preparation that is required. Typical Raman accessories are powder sample holders, cuvette holders, small liquid sample holders and clamps for irregularly shaped objects.

Many organic, and inorganic, materials are suitable for Raman spectroscopic analysis.

These can be solids, liquids, polymers or vapours. The majority of bulk, industrial laboratory samples are powders or liquids and can be examined directly by Raman spectroscopy at room temperature. Accessories for examination of materials by Raman spectroscopy are available across a wide range of temperature and physical forms. Sample presentation is rarely an issue in Raman spectroscopy of bulk samples.

In practice, modern Raman spectroscopy is simple. Variable instrument parameters are few, spectral manipulation is minimal and a simple interpretation of the data may be sufficient.

2.4 Identification through Raman Spectroscopy

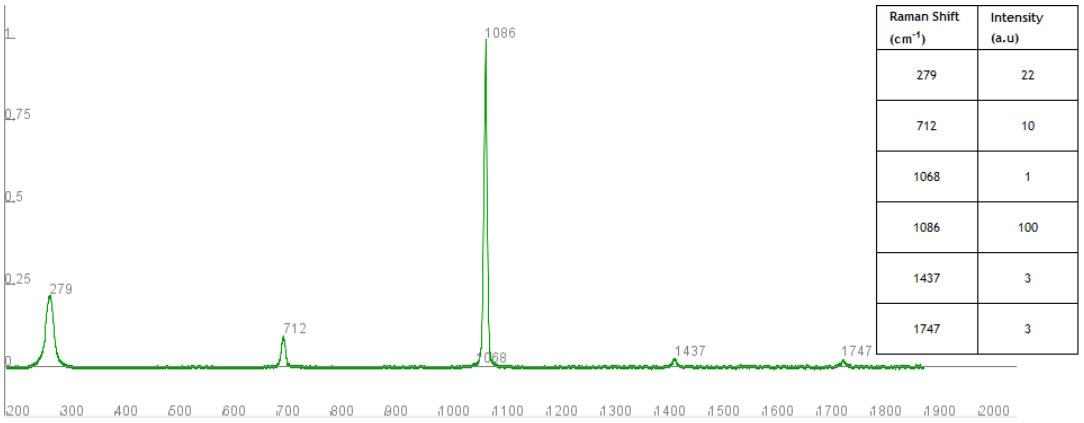
Raman spectroscopy is a photonic technique that allows to know chemical and structural information of almost any organic or inorganic material.

The aim is analyze the scattered light by a material when it is illuminated with a beam of monocromatic light. The major part of the light that impacts on the material is scattered in the same frecuency as the irradiated light, and doesn't give any information about the material but the rest is scattered in a different frecuency than the irradiated light and gives information about the molecular composition of the sample. This kind of scattered light is produced when the molecule absorbs energy and the final state is more energetic than the initial state, then the emitted photon of a higher wavelength generates a Stokes line. In the other hand, when the molecule loses energy generates an anti-Stokes line but this information isn't taken into account to analyse the spectrum. The raman spectra collects this phenomenon and represents the scattered intesity in a wavelenght which it is produced.

Each material has one or more points of high intensity in a determinated wavelength. The spectrum intensity also may be represented in Raman shift making a wave length transformation. So, each material can be represented by a spectrum which has at least one high intensity peak in a determinated wavelength or Raman shift. These high intensity peaks are also known as Raman Bands. In the figure 2.2 is shown the calcite raman spectrum and there is a table that contains the information about the intensity and the raman shift of each representative peak. Through the raman characteristics peaks we can identify materials using a database that contains information about the peaks of each material. This information can be defined as the raman signature of the material.

Some of the great advantages of Raman spectroscopy are that is not destructive with the material irradiated, doesn't need contact with the studied material and doesn't need that the sample has been preparated, is a rapid method because a spectrum can be taken in a few seconds. But also has an important disadvantage, with biological samples may appear fluorescence that makes harder the analysis.

Before starting with the sample recognition through its raman spectrum probably it would be necesary a previous processing. The common practices are the background and fluorescence removal.



Raman Shift (cm^{-1})	Intensity (a.u)
279	22
712	10
1068	1
1086	100
1437	3
1747	3

Figure 2.2: Calcite raman spectrum.

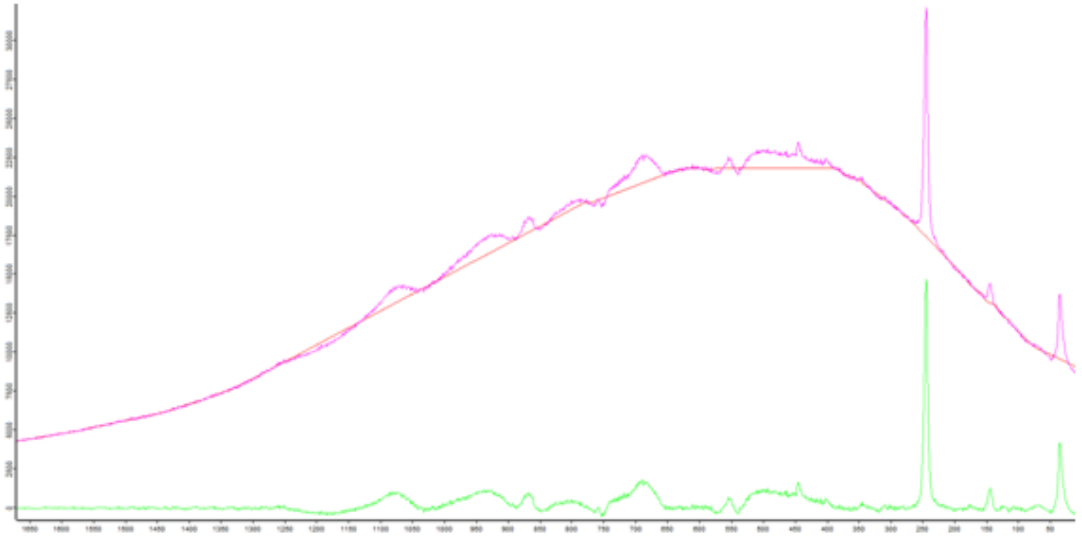


Figure 2.3: Baseline removal of magnetite spectrum.

2.5 Baseline removal

Before further spectrum analysis once acquired it is necessary the background removal because it can hinder the presentation, visualization and processing of relevant data (Fig. 2.3). This removal can be done manually or automatic, in the case of automatic removal it will be harder when appears random and systematic variations.

One important ratio to know how good a spectrum is can be the Signal-to-noise ratio (SNR 2.1), as higher this ratio will be easier remove the background from the spectrum.

$$SNR = \frac{MaximunIntensity}{\sigma} \tag{2.1}$$

To acquire spectra with better SNR it can be useful acquire it with more integration time or increase accumulation number. Integration time is the time during the sample is irradiated by the laser. Accumulation number is the number of spectra that have been made consecutively.

2.6 Fluorescence removal

As the background, the fluorescence must be removed. The presence of fluorescence makes that the spectrum has a characteristic curvature. When it appears the spectral analysis of the sample can be severely compromised.

The fluorescence may be because of the presence of organic materials but also can be observed in inorganic materials. In this cases can be attributed to the presence of impurities or external pollutants, and even environmental factors like relative humidity, temperature and sunlight.

In any case, the fluorescence must be removed during the preprocessing step before start with searching of similar spectra in our database.

2.7 Some issues about material identification

Once we have a spectrum with the baseline and fluorescence correction, we are able to start with the recognition of the sample.

All the materials have a unique spectrum which have one or more peaks with a known raman shift and intensity. So we can say that the spectrum signature is a set of pairs composed by raman shift and intensity for each characteristic peak.

The next step will be identify all the peaks in the spectrum. So, when we have a clean spectrum and the peaks identificated it may seem easy to find a similiar spectrum in a database, because we only have to search for spectrum with similar peaks. But isn't as easy as it may seem at the first sight.

Here there are some issues to take into account that makes hard the automatic recognition:

- There are many aspects that may affect to the raman shift and intensity in a spectrum like the integration time, the accumulation number, laser wavelength, irradiance, temperature, preasure, status of the material...

- With a mixture of materials, exists the possibility that more than one raman signals coincide, making hard the identification. Because a raman band can hide another through a combination of both.

- Develop a generic method to identify spectrums is complicated because there is a dependency in the characteristics of the equipment used to take the spectrum. Two spectra taken with different spectrometers may have differences.

- Peaks automatic detection depends on the previous steps, baseline and fluorescence removal. When more clean the spectrum be the more reliable the recognition be. So, it is necessary to have a reliable spectrum processing made by hand or automatic.

- Depending on the characteristics of the material, probably it would be necesary have several signatures under different conditions for the same material.

2.7.1 An illustrative example of processing Raman spectra

In [26] is described the whole process of raman spectra analyzing under the Exomars mission that ESA will launch in 2018 [10]. Most of all recently planned astrobiology missions to Mars are focused on the exploration of the surface and near subsurface in the search for biomarkers as indicators of exobiology. The search for signs of past and present life on Mars is one of the main goals of the ESA ExoMars Rover Mission. The Raman laser spectrometer (RLS), which will travel on the rover, has been ranked as *Fundamental* nondestructive analytical instrumentation for the mission.

In this article is presented a Raman signal processing package to perform in-depth processing of the spectra that aims to eliminate, insofar as possible, the noise and remove the baseline from the spectra. The process described needs filtering of raw data, baseline removal, peak finding and finally database search for sample identification.

Chapter 3

Raman Spectra Classification

3.1 Resumen

Hay muchos trabajos relacionados con la identificación de compuestos. Es posible clasificarlos teniendo en cuenta los diferentes métodos utilizados para reconocer un espectro problema. Los principales métodos están relacionados con inteligencia artificial, análisis de componentes principales y algoritmos basados en diversos coeficientes de correlación.

Hay un grupo de artículos publicados por un grupo de investigadores en el que han trabajado en la identificación de pigmentos por espectroscopía Raman. Todos los artículos tienen en común el uso de lógica difusa para identificar un espectro problema en una colección de espectros de referencia de pigmentos

3.2 Introduction

There are many works related with the identification of different compounds and materials. It is possible classify them taking into account the different methods used to recognize a problem spectrum. So, the main methods are related with artificial intelligence, principal component analysis, and various algorithms based on correlation coefficients.

In the next sections, some articles are commented grouped by the main technique used, so the next subsections are the followings: Fuzzy Logic 3.3.1, Artificial Intelligence 3.3.2, Principal component Analysis 3.4, Other Algorithms 3.5 and a brief introducction 3.2.

The article [32] talks about many scientific and industrial disciplines where Raman spectroscopy, and therefore recognition, find application. Some of these disciplines are:

- Chemistry: providing a chemical fingerprint for identification of a molecule, since vibrational information is specific to the chemical bonds and symmetry of molecules.

- Forensic sciences and Criminology: is used for identification of trace amounts of substances in evidential materials, etc. In-situ measurements can be realized, meaning no contamination of evidences during taking samples. Also can be used in identification of unknown or hazardous substances, by instance detection of explosives or drugs.

- Medicine: prognosis and diagnosis of carcinomas and other diseases.
- Geology and Mineralogy: serves for identification of the principal mineral phases or classification.
- Art and paints: examination of artworks and artefacts reveal worthy information for conservators or those of general historical interest.
- Robotic exploration of Mars: as a part of the scientific laboratory, a raman spectrometer will be included in the Exomars mission of ESA. [17]

In [31] are presented some issues resolved at the present time using Raman recognition. These examples are detection of small hexavalent chromium concentrations, monitoring of the curing process of epoxy resins, clasification of inks within the scope of revealing ink document falsifications...

As stated in section 2.3, each material can be represented by a spectrum and therefore each raman spectrum is like a fingerprint through a material can be identified. So, to facilitate the material recognition it is neccesary have a spectral database or a model to make possible the comparision between spectra.

In the past, databases were unreliable [20] owing to the variability in spectra obtained (ratio signal - noise, background...) by the earlier low throughput multi-grating systems. The first reliable Raman spectral databases have been produced for the new generation of spectrometers.

An example of Raman database is in [6], where Raman spectra of 43 excipients commonly used in pharmaceutical formulations are presented and their Raman bands are identified.

There is an important project, called RRUFF [5], that is creating a large database of Raman spectra from well characterized minerals. This is an open project and shares all the spectra in the web. Because is an open project, anybody can contribute with their spectra. One of their aims is allow users of Raman instruments compare their Raman patterns to those from RRUFF with confidence. But through the online database is not possible to make an spectra identification. On the other hand there is a free standalone program (Crystal Sleuth) developed by RRUFF that has a functionality that permits compare a problem spectrum with a large group of spectra obtained from the online database. In the section 3.5 the identification method of Cristal Sleuth will be discussed. An example of the use of RRUFF is in [16], where various raman spectra were acquired from 96 semi precious gemstones with the aim of classify them with reference spectra from the RRUFF database.

But for the material identification is not enough searching similarities between the problem spectrum and a set of reference spectra. In many cases the problem spectrum may belong to a mixture of more than one material and in this cases identify the materials that have a contribution in the raman spectra is a harder task. In a first view, seems reasonable that the contribution of each material is a lineal contribution taking into account the proportion of each material in the mixture. But is not true in all the cases. The article [1] talks about Principle of superposition in the Raman effect with mixtures. This concept can be interpreted into two ways, first way qualitatively refers to the fact that the bands of each individual material appear in the spectrum of the mixture, or second way, quantitatively

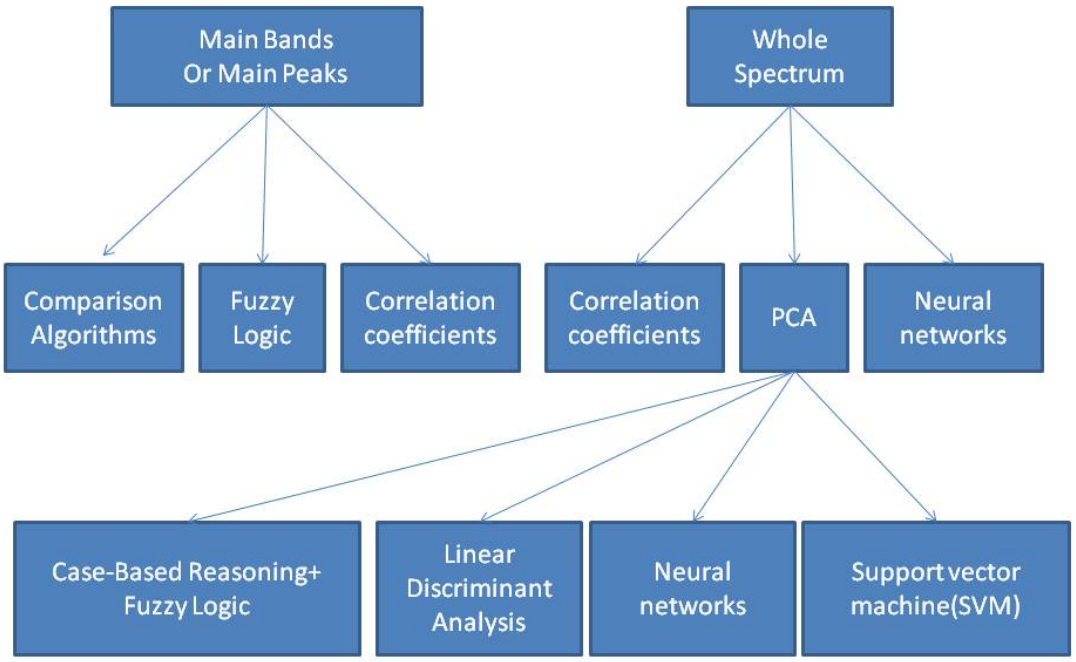


Figure 3.1: Summary of the main methods.

that is, the mixture spectrum is the pondered sum of the individual spectra according to the proportions in which they appear in the mixture. But, in some cases, the Principle of superposition is not quantitatively verified and qualitatively. Specific properties of the materials, like reflectance, absorption, transmission and grain size and others are responsible for the non-linear effect that does not agree with the principle of superposition.

For the identification of spectra, it is possible divide the methods into two main groups. The first group is searching for similarities between whole spectrum and the reference spectra and second group searching similarities between the main bands or peaks of the problem spectrum and main bands of the reference spectra. In the next sections some methods will be discussed always bearing in mind the difference between this kind of methods.

In figure 3.1 it is shown a summary of the main methods found in the articles.

3.3 Artificial Intelligence

3.3.1 Fuzzy Logic

There are a group of articles published by a group of researchers in which they have worked in the identification of pigments by raman spectroscopy. All the articles have in common the use of fuzzy logic to identify a problem spectrum in a reference library spec-

tra of pigments. They have experimented with the fuzzy logic in main methods to identify raman spectra, taking into account the whole spectrum and only the main spectrum bands. Besides they have supplemented the fuzzy logic with other interesting methods like Principal Component Analysis (PCA).

The article [19] talks about a fuzzy logic system for band detection of Raman spectra. The reasoning used looks for the peaks in the spectrum in a conventional visual way. This is an important step when is intended to compare spectra by raman bands or peaks. They make a formulation of input membership functions that transform the system inputs into fuzzy variables. The input fuzzy variable is then mapped into a fuzzy output variable by means of rules, which describe the relation between these variables. Both membership functions and rules are problem dependent. In this article the location of Raman bands is based on the quadratic coefficients of the parabolic approximation of the spectrum in consecutive segments. The quadratic coefficient is negative when the shape of the spectrum in the segment is convex. Bearing in mind that the shape of a Raman band is always convex, in a segment with a negative quadratic coefficient a Raman band can exist. In particular, the general statement for detecting bands is: *IF three consecutive quadratic coefficients are negative and similar THEN there is a Raman band*. The inference engine provides the way in which rules are combined and evaluated in order to obtain the output fuzzy sets and finally, the crisp output value is obtained using a centroid technique from output fuzzy sets. The method is practically independent of the spectrum baseline and therefore is not necessary to eliminate it previously.

The article [3] talks about identification of artistic pigments using fuzzy Logic and principal component analysis (PCA) . They reduce each dimension spectra by means of a data reduction tool called the principal component analysis (PCA). The most important use of this chemometric technique is to represent the N-dimensional data in a smaller number of dimensions, usually two or three, without loss of information. A correlation coefficient is used to estimate the degree of similarity between problem spectra and reference spectra. The fuzzifier assigns to every numerical input value a degree of membership to each input Fuzzy sets through the membership functions associated to them. The inference engine provides the way in which rules are combined and by means of different logical operators modifies the defined output fuzzy sets. Finally, the defuzzifier maps output fuzzy sets into crisp numbers applying mathematicalmechanisms in order to obtain the final result. If there is more than one candidate, a new library is made by mixing them. This system can only identify binary mixtures that are the most common mixtures made by artists in the pigments field.

The article [18] again uses fuzzy logic to identify spectra but now comparing only raman bands. First, by means of a fuzzy system based on If Then rules, the system selects as candidates all the pigments in a considered database that have bands in the same positions as the unknown spectrum. Second step is calculate the similarity degree, it numerically determines the coincidence in the wavenumber position of the bands in the compared spectra. That is, the greater the coincidence between bands, the closer the similarity degree value is to 1. And finally, the output variable called the matched bands number provides the ratio of the number of matched bands between the unknown spectrum and the candidate to the total number of candidate bands, expressed as a percentage. All of these variables are input variables in the following stage (identification fuzzy system),

which makes the decisions to the identity of the analyzed pigment by selecting it from among the candidates and provides the identification reliability. Thus, the output of the whole system is the pigment (or pigments) identified as the unknown pigment with a certain confidence, which is expressed as a number from 0 to 10.

In the article [22] an algorithm was tested with a database of 32 pigments. The algorithm was tested both pure pigments and mixtures. The system is designed as a software complement in a portable micro-Raman instrument for studying pigments. The Raman bands that will be used in the fuzzy system to perform the automatic identification. Before the fuzzy procedure is applied, a previous denoising process is necessary along with a baseline correction to remove background signals usually caused by fluorescence. Peak picking is performed manually. The user selects the Raman bands that are considered to be characteristic of the compound. The algorithm first calculates the intersection between the fuzzy set of the unknown spectrum with each of the crisp sets stored in the reference library. As a result, a score is obtained for each band in the computed intersection. These scores are then weighted using a factor that represents the probability of occurrence of the Raman band. Finally, final result is weighted with a factor that is the number of reference bands encountered in the unknown pigment spectra divided by the total number of bands in the reference pigment spectrum. The value final is between [0, 1] and is classified using three categories, *present*, *possibly present*, or *not present*.

The article [27] makes a comparison between two strategies based on fuzzy logic to identify unknown Raman spectra. First one is compare the whole spectrum of the analyzed sample with the spectra of standard materials or, once the wavenumber positions of the raman bands of the unknown spectrum are localized, to compare them with those of the reference spectra. The identification system comparing the whole spectrum is made by estimating the degree of similarity between two spectra, that is the correlation coefficient. To make the correlation independent of the Raman intensities each spectrum is normalized between 0 and 1. The crisp inputs of the system are the correlation coefficients between the unknown spectrum and each of those of the chosen library. The system has four rules to make a decision. Then, for an obtained correlation coefficient, each of these four rules is interpreted by implication, using the product operator. For each implication, a fuzzy set is obtained and all of them are aggregated, by means of the sum, into a single output fuzzy set. Finally, the output crisp value, the degree of similarity between problem spectrum and reference spectrum, is calculated using the centroid method. In the other hand, identification system locating the Raman bands needs only the wavenumber position of the Raman bands. So the system only looks for the coincidence in the position of all the bands of the two compared spectra. The input variables of the system are the wavenumbers, on which the Raman bands of the spectrum under analysis are centred. The fuzzy system for bands detection used in this article has been described in this section [19] . The output variables are the pigments among those catalogued in the library with which the unknown spectrum shares some bands. The higher the number of common bands, the more similar are the compared spectra. The membership functions assign the greatest degree of membership (value 1) to the exact position of the band, a smaller degree of membership to the values of adjacent wavenumber positions in a spectral range and a degree of membership zero for the wavenumber positions out of that range. The output variables are, on one hand, the names of each pigment with which the system finds

coincidences, and on the other hand, for each of those pigments, a *degree of security* in their identification. One important result is that in cases of the mixture, the system locating the bands identifies the presence of two or more pigments, whereas the other one only recognizes one of them , the one in which the similarity degree is higher.

In [2] is presented a three-phase methodology that automates the spectral comparison based on one of the most powerful paradigms in machine learning, the case-based reasoning (CBR). The spectra are processed in order to minimize the noise effects and eliminate the fluorescence baseline before being used by the system. The first step is normalization, interpolation and reduction of the number of data points of the spectra without loss information with PCA. Once the spectra have been processed, the system has to capture among all the spectral patterns those that are the most similar to the analyzed spectrum and it returns an identification degree that may help the user to identify the unknown pigment. For each unknown spectrum is proposed a solution taking into account previous knowledge. The methodology works as a case-based reasoning system finding out the best solution for in each case by following an inference strategy. Is composed of the following phases

- Retrieves the most similar cases from the case memory.
- Adapts them to propose a new solution.
- The user checks if this solution is valid.

In this article it is not the objective to work with large spectral libraries, actually, they use some characteristics that can be used to exclude a large number of pigments. This system make use of the complete spectra and not only their band positions, but to avoid handling a large number of points for each spectrum they make use of principal component analysis (PCA). The most similar pigments are retrieved using Pearson coefficients. Candidate by candidate, the inference process deals independently with four rules and leads, for each of them, to a fuzzy set output. Finally the user takes his final decision by evaluating the solutions proposed by the system.

3.3.2 Neural networks and others

The article [7] describes application of artificial neural networks and statistical methods to analyse a Raman spectra database and to use the results of the analysis in structure elucidation process. A statistical algorithm is used to search a database of spectra for reliable regularities required for determining the presence of spectrasubstructure correlations. The database contains 156 Raman spectra of organic compounds. The spectra were transformed into vectors encoded in the binary system and divided in intervals. Each interval is assigned a 0 value or a 1 value. If the spectrum of a compound has at least one significant peak in an interval, the interval code is 1. If no peak appeared in an interval, the 0 value is given as the code. A set of vectors representing Raman spectra and coded information about the presence or absence of a selected substructure was divided into a training set and a testing set. With these sets, a correlation vector is created by rejecting the components whose values are below a given threshold value. Then the generation of a rule knowledge base consistent with the regularities represented by the correlation vector is created. The fixed spectrumsubstructure correlations were automatically transformed

into rules if - then. The user inputs into the computer the spectral parameters (band locations and intensities) of the Raman spectrum obtained for the investigated compound. The inference engine utilizing the knowledgebase produces the list of substructures that probably form a part of the molecule of the examinated substance. In the other hand, in the neural network approach, designing and training of artificial neural networks was carried out with the use of a commercial program Stastica Neural Networks. As a result, the neural network method achieved more reliable results for almost all substructures.

In [23] an integrated software system for processing, analyzing, and classifying Raman spectra is presented. The system is open source and extensible, allowing the community of Raman researchers to make continual improvements to the software It has features to subtract background and fluorescence. Implements several analysis and classification techniques, such as principal component analysis and support vector machines (SVM). For each of these techniques, a model is created by training the algorithm with selected groups of spectra. The system currently supports three different algorithms for spectral classification, one based on linear discriminant analysis (LDA), another implemented an artificial neural network (ANN) and the last one employs support vector machines (SVMs).

The article [24] says that identification of an unknown species based on spectroscopic data is a common statistical problem. Most of these statistical methods can be separated in two main groups, unsupervised or exploratory and supervised methods. Unsupervised methods are used for studying experimental spectral data without a prior knowledge of the object. Hierarchical cluster analysis (HCA), Density Based Spatial Clustering of Applications with Noise or PCA based methods of the dimensionality reduction are more commonly used approaches. Supervised methods utilize a prior knowledge about the system by developing classification models based on known spectra . These methods include Linear DA (LDA), Direct LDA (DLDA), Kernel-based LDA , Multivariate Analysis of Variance ... Each of these algorithms is most efficient for a certain type of data. In cases when the data set characteristics are not known, selection of the algorithm is usually done using trial and error.

An example of the use of support vector machines is in [21]. This method utilizes the entire spectrum and determining the state of the cells based upon the similarities/differences of the examined spectra versus an established database. The SVM technique is used to classify. The classification intends to identify the cell mortality type, by distinguishing among apoptotic versus healthy cells and necrotic versus healthy cells.

In this experiment, the data is used in its raw form without any further processing because the baseline, fluorescence, and intensity range of the spectra are characteristics of the cellular state. Therefore removing them will reduce the convoluted spectral information.

3.4 PCA: Principal Component Analysis

Principal component analysis is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Normally is related with identification made through the whole spectrum. Below are some works related to

this technique.

In [30] is made an evaluation of a searching algorithm based on principal component analysis for identification of organic pigments. Before processing the principal components calculations a baseline-correction is performed to avoid background and fluorescence. The spectra are normalized. The training set of reference spectra is used for the principal components analysis. 28 PCs are extracted from each spectrum. For the problem spectrum is made a similar pretreatment. Comparison between the unknown and the references is made using the Euclidean distance between two data points as a measure of their similarity. They obtained bad results when the spectra have low signal to noise ratio.

In [4] uses PCA and cluster analysis to classify glasses and glazes. With cluster analysis creates a hierarchical tree diagram that assembles samples as a function of the shortest distance. Clusters presented are built using a Euclidean distance.

In [12] is made an experiment to identify *phylogenetically homogeneous Bacillus subtilis-group*. First normalize spectra and the first derivatives were calculated to minimize the influence of background signal caused by slight sample fluorescence. Then data reduction is performed using principal component analysis. During internal evaluation the complete dataset (219 experiments) was randomly splitted in a training set (3/4) and a test set (1/4). To search similarities is used the euclidean distance. The identification was regarded as reliable at the species level when at least three out of five matches pointed to the same species. They say that this method should not be considered as a stand-alone technique for taxonomic purposes.

In [13] spectra were reduced using principal component analysis and six linear discriminant analysis models (LDA) were calculated to construct an identification scheme. Taking conventional microbiological identification as the reference method, the accuracy of this identification scheme was 90%.

3.5 Other algorithms

In [29] is presented an evaluation of a spectral searching algorithm for the comparison of Raman band positions. The algorithm evaluates all reference spectra one after the other, searching for coincidences in bands positions. Only reference products with at least one corresponding band are included in the list of results. In order to be able to investigate mixtures of compounds, combinations of the products with at least one identified Raman band can be made and are treated like the reference spectra. Finally, the selected (combinations of) reference products are evaluated and sorted, by using an appropriate measure of similarity. Different measures of similarity or dissimilarity can be used, as the number of bands of the unknown that can be explained by the reference product, number of bands of the unknown spectrum that are not present in the corresponding reference spectrum, bands that are present in the reference spectrum, but that cannot be observed in the spectrum of the unknown, deviation between an identified Raman band position of the unknown and the reference product and the root-mean-squared (RMS) deviation between the Raman band position of the reference product and the band position of the unknown. If not all Raman bands of the unknown product can be assigned to a single reference product (i.e. the number of unidentified bands greater than 0), an iterative approach can be used. In that case, combinations of reference products with at least one identified band

are made (the lists of reference band positions are merged), and this combination is evaluated. The algorithm was thoroughly evaluated with different test samples, recorded on a different spectrometer than the reference spectra. In general, the algorithm is able to identify the pigments in the sample, even in a mixture. But when some reference products have a very high number of Raman bands, accidentally they can quite easily give rise to a certain number of identified bands, masking the presence of a compound with only a limited number of Raman bands. Also, in mixtures weak bands are not always detected and the most intense features of a certain reference product may in a mixture only be present as a band of minor intensity.

In [14] it is talked about the design of a portable spectrometer. For this, use a treatment of three main stages. Initially, the background has to be removed from the spectrum. The background of the spectrum is approximated by a first-degree polynomial and then the background is subtracted from the measured spectrum. The second step of pre-processing is spectrum smoothing and is performed to reduce its random component and accidental spikes. A Savitzky and Golay smoothing filter is chosen. Third, the pre-processed spectrum is parameterized to establish a set of parameters (positions of the spectra peaks, their relative amplitudes and their widths). To find positions of the spectral lines a derivative of the spectrum is examined. Chemicals are detected by comparing the registered spectrum with the reference spectra available in the database. The spectra are compared by summing up, according to different norms like absolute difference value search, first derivative absolute value search or least squares search. For a portable device the data processing time of each algorithm is a very important parameter so they conclude that the algorithms based on the first derivative of Raman spectra are most efficient for chemicals detection.

Chapter 4

Solution Definition

4.1 Resumen

El algoritmo de nuestra propuesta se basa en la comparación de bandas raman. Este algoritmo tiene dos fases principales, una búsqueda inicial y una búsqueda recursiva.

En la fase inicial se recuperan de las base de datos todos los espectros que tienen una intensidad relativa igual a 1 con el mismo desplazamiento raman que la banda de mayor intensidad del espectro problema. A continuación se buscará enlazar el resto de picos del espectro problema con los del espectro referencia. Una vez se han encontrado todas las coincidencias posibles entre los espectros, si aún quedan bandas por identificar se pasa a la fase recursiva.

La fase recursiva renormaliza el espectro en intensidad y busca enlazar con otros espectros referencia los que no se han podido identificar en las fases anteriores. Esto se repetirá hasta que no queden bandas del espectro problema por identificar o hasta que no haya candidatos.

4.2 Initial search

The algorithm of our proposal is based on bands comparisons. This algorithm has two main phases, an initial search and a recursive search. In the two next sections both phases will be explained.

The algorithm starts retrieving the spectra that have a band that matches with the band with the maximum intensity in the problem spectrum. If, besides of the band with the maximum intensity, all the bands of the problem spectrum have matched, the reference spectrum contains all the bands of the problem spectrum and therefore the problem spectrum is a pure material. In the other hand, if after matching all the rest of bands of the reference spectrum still there is one or more bands of the spectrum problem that have not been matched, the problem spectrum probably belongs to a mixture of two or more materials.

From the set of possible candidates, there will be discarded those whose characteristics bands don't match with the problem spectrum bands. We define characteristics bands as the eighty per cent of the number of bands of the reference spectrum that has a normalized

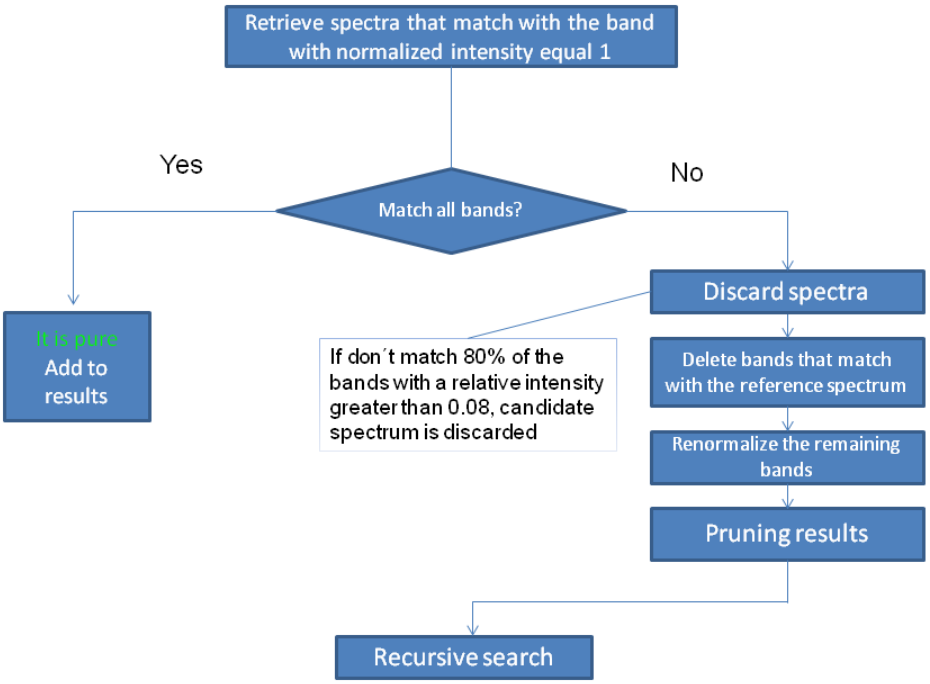


Figure 4.1: Initial search.

intensity greater than eight. This rule is not applicated to spectra of materials that have less than five raman bands.

The next step is delete from the search those bands that have matched with the reference spectra, so, after this step, there only will be bands of the problem spectrum that have not matched.

With this set of not matched bands, a new normalization is made without the intensities of the matched bands. The aim of this renormalization is make possible new matches. The reference bands spectra in the database are stored normalized and so, as we have deleted from the search the bands that belongs to other spectra, without them, the next most intense peak should have intesity equal to 1 to match it in the database.

Finally, the last step will be pruning of candidates. To save time with large databases, when there are a big number of candidates we pruning those that don't have a number of matches greater or equal than the half of matches of the candidate with the maximum of matches.

Once the initial phase has finalised, the rest of the bands that not matches are processed in the recursive phase. One flow chart with a summary of the initial phase is in figure 4.1.

4.3 Recursive search

The recursive search starts retrieving all the spectra that have at least one coincidence with the not matched bands from the previous phase or iteration. There is a summary flow

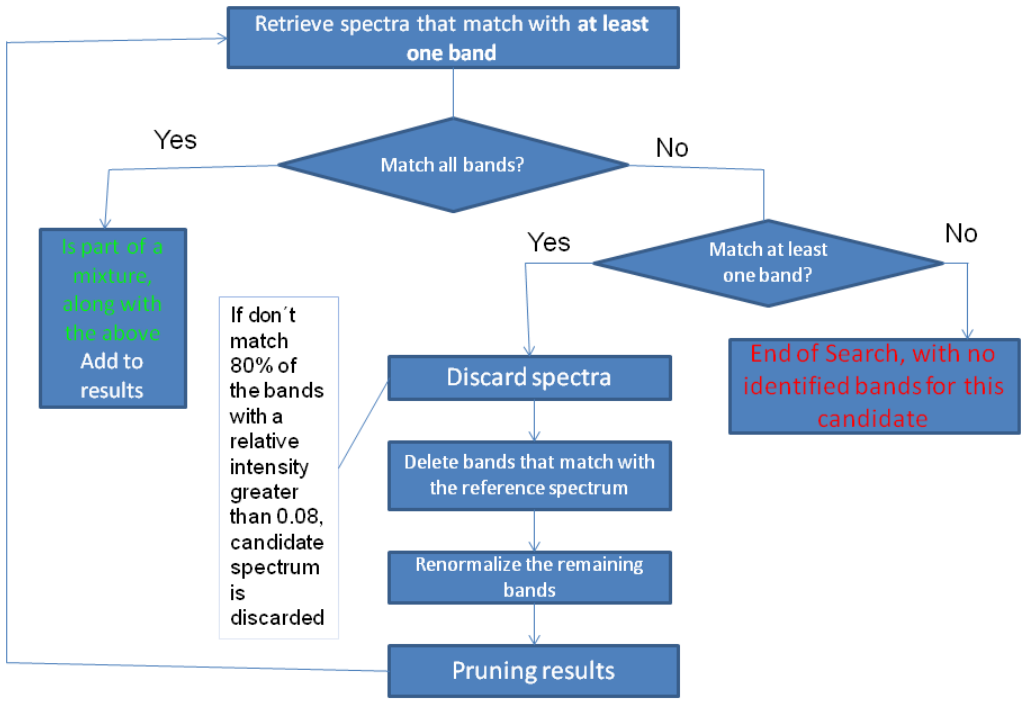


Figure 4.2: Recursive search.

chart in figure 4.2.

When the rest of the not matched bands are matched, the reference spectrum is included in the solution as a part of a mixture. But when still there are not matched bands, a new recursive search is made. Before recursive search, a process similar to the initial search is done with the remaining bands. First, the candidate spectra may be discarded if don't match with the characteristic bands (the eighty per cent of the bands that has intensity greater than 0.08). With this rule we avoid non representative matches.

After it, the matched bands are deleted from the next recursive search and like the previous search, the rest of the spectrum is now renormalized with the same aim. Again a pruning of results is made to avoid non representative mixtures composed by a high number of materials. The pruning made is the same than in the initial search.

The recursive search will finish in one of two cases. When all the problem spectrum bands are identified or when there is not more candidates for the not matched bands.

As a result of this recursive algorithm a tree is created with the reference spectra that cover all the bands of the problem spectrum. Each tree level may have a set of reference spectra that match with a diferents sets of bands.

With this algorithm a detailed information is given to the user. For each band of the problem spectrum, a list of reference spectra is shown to help the user to identify the material.

Chapter 5

Experimental Results

5.1 Resumen

Se ha desarrollado una infraestructura tecnológica basada en una base de datos MySQL y PHP para implementar la propuesta de algoritmo.

Se ha contado con un corpus de unos 2346 espectros raman para probar el algoritmo. Este corpus se ha obtenido del proyecto RRUFF.

En primer lugar se ha probado el algoritmo de identificación con los espectros del propio corpus. En todos los casos los espectros han sido identificados, sin embargo en el 21 por ciento de los casos, el número de resultados obtenido ha sido demasiado numeroso.

En segundo lugar se ha comparado el algoritmo propuesto con uno publicado recientemente que también se basa en comparación de bandas raman. Sin embargo este algoritmo no utiliza la intensidad de las bandas raman. Probando este algoritmo con el mismo corpus, se ha observado que el número de resultados muy numerosos se ha incrementado hasta el 45 por ciento.

5.2 Experimental Design

For testing the recognition algorithm, has been created an technological infrastructure based on a web interface. The spectrum database has been mainly developed with MySQL and the algorithm has been implemented with object-oriented PHP and Java.

With PHP has been implemented the algorithm and all the necessities of management and treatment of the spectra, like normalization, baseline removal, automatic detection of raman bands...

With Java has been developed the graphical interface to easily interact with the spectra. Some examples of this functionalities are visualize Raman bands, plot baselines, make zoom on the spectra, visualize results...

We have made tests in two different corpus. On the one hand we have a database created at *Unidad Asociada UVa-CSIC a traves del Centro de Astrobiologia* [28] that contains 224 raman spectra of 117 samples, on the other hand, we have used the set of spectra that the program *Crystal Sleuth*, developed by the RRUFF project [5], contains to make identifications.

But the set of spectra that *Crystal Sleuth* contains does not have information about raman bands, it only has raw spectra. So, to make possible the utilization of this corpus with our algorithm, it was necesary process this set of spectra. Therefore we have developed an automatic process that extracts the information about raman bands in the set of spectra of *Crystal Sleuth*.

As this set of spectra has been obtained from the RRUFF database, it has a heterogeneous spectra set, so there are spectra with a good ratio signal-noise where it is easy to automatize the automatic recognition of peaks but also there are a group of noisy spectra where our automatic process does not work as wanted.

Taking this into mind, from the original set of 2643 spectra have been filtered 297 spectra that have not a good signal-noise ratio. So our corpus has 2346 spectra from 1571 samples.

5.2.1 Testing algorithm with the corpus

For testing our proposed algorithm we have randomly selected 100 spectra from the set of Crystal Sleuth program which are part of the whole corpus.

For the automated identification process we have used a margin of error of 5 cm^{-1} for raman shift and a margin of error of 0.05 over 1 for intensity. With the automatic process we have evaluated the identifying this set of 100 spectra, getting the following results:

- All of them have been correctly identified. All results lists have the original spectrum as a suggestion that matches with the raman bands of the spectrum problem.
- In 55 of the 100 spectra there is only one suggestion to identify the problem spectrum . So, in these cases the problem spectrum is identified unequivocally.
- In 21 cases the number of results is too large taking values from 28 to 628 suggestions, so for this cases the result may be unacceptable.

In 5.1 it is shown an example recognition of jarosita in the system with 5 cm^{-1} of margin of error for raman shift, 5 percent of error for intensity and 3 times the varianze to search peaks in the problem spectrum.

In conclusion, this algorithm is able to identify the spectra from the corpus but it seems reasonable find a way to make more accurate the threshold that decides when a problem spectrum matches with a reference spectrum.

5.2.2 Comparison between propolsals

As discussed in section 3.5 in [29] is presented an algorithm for the comparison of Raman band positions. As the algorithm presented in this work, this solution is based in comparisons between raman bands , an iterative comparison and a measure to know the similarity between two spectrums. But between these two proposals there are two main differences:

- Algorithm in [29] evaluates all reference spectra one after the other while the algorithm presented here only evaluates those which have matched with the highest intensity band and with the derivates that still have matched with the remaining

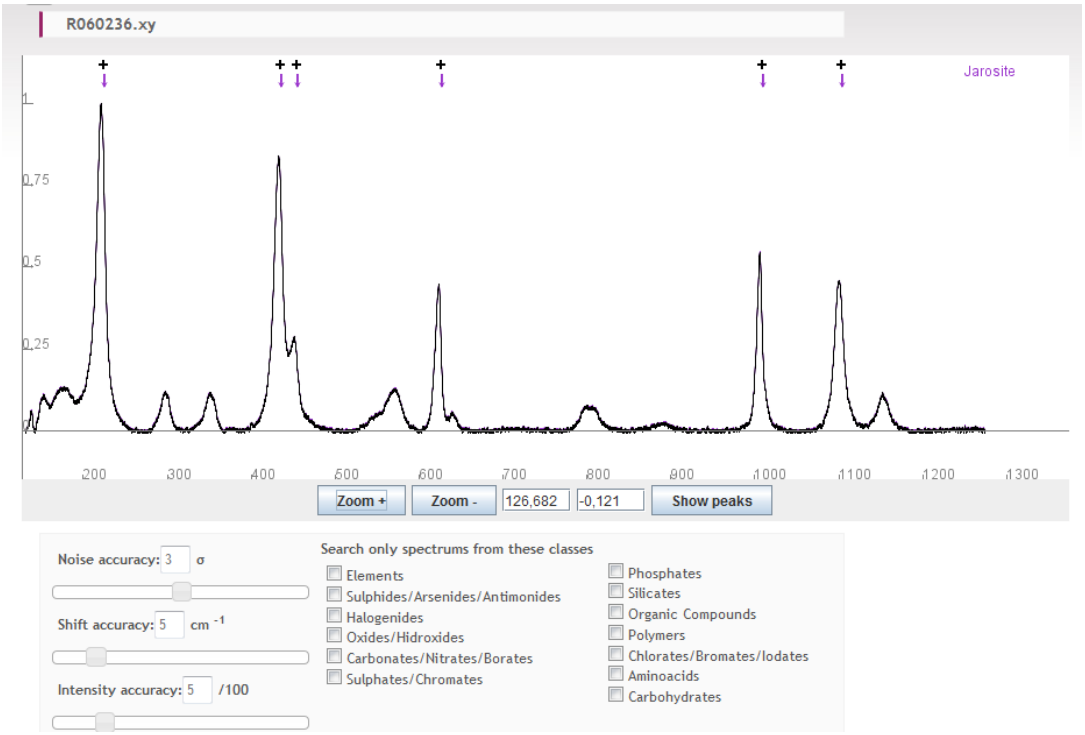


Figure 5.1: Example identification of a jarosite spectrum through the developed system.

bands that have not matched with the first reference spectrum. This way of evaluate the reference spectra searching for matches saves time and is an important issue when there are more than 2000 spectra, like in the used corpus.

- As the article itself mentions, the main disadvantage of this approach is that the intensity of the bands is not taken into account. The reason of not using intensities for looking for matches is that the intensity of a spectrum may change depending on the spectrometer used. But in the other hand, the intensity is a very powerful threshold to reduce the number of candidates for a spectrum problem, even more when there is a big database of reference spectra. To prove it , the same experiment (5.2.1) than in the previous section has been performed but for this time the raman bands matches have been searched only taking into account raman shift. The results are the followings, in figure 5.2 there is a summary with the results of both experiments:

  - Like in the previous experiment, all of them have been correctly identified. All results lists have the original spectrum as a suggestion that matches with the raman bands of the spectrum problem.
  - None of them has only one suggestion to identify the problem spectrum. The minimum of suggestions is three.
  - In 45 cases the number of results is too large taking values from 26 to 153 suggestions, so for this cases the result may be unacceptable.

	With Intensity	Without Intensity
Identified	100%	100%
Only one suggestion	55%	0%
Too large results	21 %	45 %

Figure 5.2: Comparison between searching raman bands with and without intensity

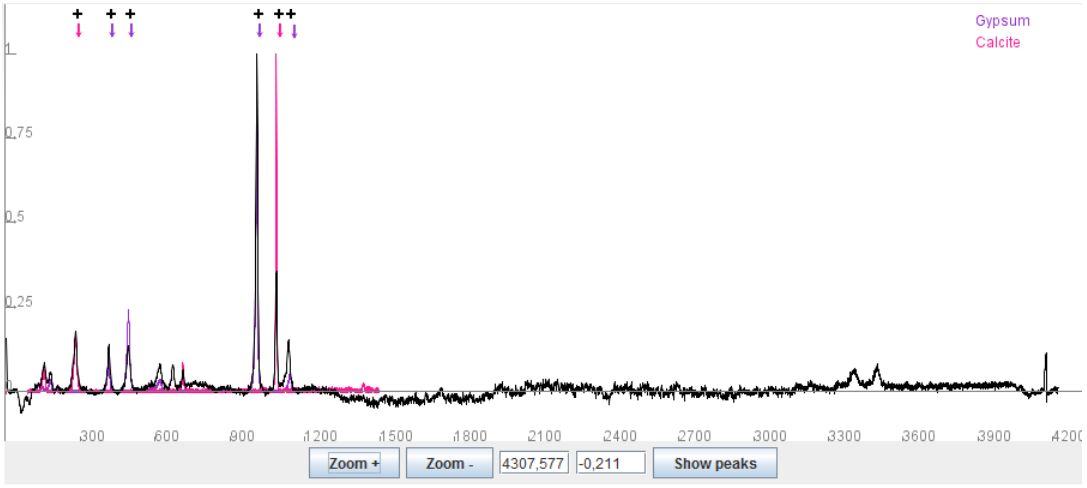


Figure 5.3: Mixture of calcite and gypsum.

With this results, make the searches of raman bands with coincidence in intensity seems necessary when there is a large amount of spectra to compare at least in the first iteration of the search. The problem of having spectra from different spectrometers and therefore probably with different intensities of the raman bands can be avoided in part through normalizing the spectra and with definition of tolerance margins.

5.2.3 Identification of mixtures of compounds

The algorithm proposed here has been developed bearing in mind the possibility of identify samples compounded of more than one mineral. These mixtures can be compounded of minerals in different proportion that can modify the absolute intesity of a raman band. So to achieve identify more than a compound in one spectrum the recursive search was developped as is described in section 4.3.

In the case of tests made with different real samples of calcite and gypsum with different proportions, the algorithm is able to identify both minerals but also gives as a solution around other sixty spectra. The result of recognition the mixture can be seen in figure 5.3.

Also, with mixtures can occur that with a low proportion of a compound not all raman bands appear in the spectrum and so the rule used in the algorithm of having at least 80 per cent of number of raman bands can leave out from suggestion the real compound.

Chapter 6

Conclusions and future work

6.1 Resumen

Este trabajo hace un estado del arte de los principales artículos sobre identificación de minerales a través de espectros raman.

Se ha hecho una propuesta de algoritmo de identificación capaz de identificar minerales a través de su espectro raman pero que necesita de alguna mejora para reducir el número de resultados para un espectro problema en un 21 por ciento de los casos. No obstante, lo desarrollado puede ser una potente herramienta para espectroscopistas expertos.

Este trabajo será presentado en el congreso European Planetary Science Congress 2012 en Septiembre de 2012.

Como trabajo futuro se buscará mejorar el algoritmo e incluir en el procesado de los espectros la eliminación automática de la linea de base de los espectros y la detección automática de bandas raman.

6.2 Conclusions

This work makes a compilation of some articles related with the identification of different materials or compounds through its raman spectrum. Also, taking into account the ideas found in the articles an algorithm has been developped to identify single minerals and mixtures. Beside the algorithm, an technological infrastructure has been created to test the algorithm and to help user to processing raman spectra. This algorithm always takes as a precondition that the baseline has been correctly removed from the problem spectrum and its raman bands has been correctly plotted.

As the results showed, this algorithm is able to identify minerals through its raman spectrum but seems necessary make a improvement that makes a more strict threshold that decreased the number of suggestions in some cases.

Although the algorithm with the technological infrastructure developed can be a powerful tool with a combination of it and some ideas from the articles presented here can make more reliable and a better usability.

This work will be presented at European Planetary Science Congress 2012 [9] as *Raman spectra processing algorithms and database for RLS-ExoMars*. The poster also will

include a study and an proposal of an algorithm to automatic baseline removal and automatic peak detection. It will be presented during the session *Planetary in situ measurements*

6.3 Future work

- Develop an algorithm to automatic background removal. To do it, the starting point will be [33].
- Improve algorithm automatic raman bands detection.
- Develop an algorithm and a database for a portable device.
- Develop an online database with mineral identification capabilities.

Bibliography

[1] Mónica Breitman, Sergio Ruiz-Moreno, and Alejandro López Gil. Experimental problems in raman spectroscopy applied to pigment identification in mixtures. *Spectrochimica Acta Part A 68*, 2007.

[2] M. Castanys, R. Perez-Pueyo, M. J. Soneira, E. Golobardes, and A. Fornells. Identification of raman spectra through a case-based reasoning system: application to artistic pigments. *Journal of Raman Spectroscopy*, 2011.

[3] M. Castanys, M. J. Soneira, and R. Perez-Pueyo. Automatic identification of artistic pigments by raman spectroscopy using fuzzy logic and principal component analysis. *Laser Chemistry*, 2006.

[4] Philippe Colomban, Aurelie Tournie, and Ludovic Bellot-Gurlet. Raman identification of glassy silicates used in ceramics, glass and jewellery: a tentative differentiation guide. *Journal of Raman Spectroscopy*, 2006.

[5] Database of Raman spectra, X-ray diffraction and chemistry data for minerals Last Access 27 February 2012. http://rruff.info/about/about_software.php.

[6] Marleen de Veij, Peter Vandenabeele, Thomas De Beer, Jean Paul Remonc, and Luc Moensa. Reference database of raman spectra of pharmaceutical excipients. *Journal of Raman Spectroscopy*, 2008.

[7] B. Debska and B. Guzowska-Swider. Searching for regularities in a raman spectral database. *Journal of Molecular structure*, 2005.

[8] Howell G. M. Edwards. *Handbook of Raman Spectroscopy*, volume 1. Marcel Dekker, Inc, 1th edition, 2001.

[9] European Planetary Science Congress 2012 Last Access 15 July 2012. http://www.epsc2012.eu/home.html.

[10] European Space Agency Web Last Access 12 January 2012. http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=45103&fbodylongid=2130.

[11] John R. Ferraro, Kazuo Nakamoto, and Chris W. Brown. *Introductory Raman Spectroscopy*, volume 1. Elsevier, 3th edition, 2003.

[12] Didier Hutsebaut, Joachim Vandroemme, Jeroen Heyrman, Peter Dawyndt, Peter Vandenabeele, Luc Moens, and Paul de vos. Raman microspectroscopy as an identification tool within the phylogenetically homogeneous bacillus subtilis-group. *Systematic and Applied Microbiology*, 2006.

[13] M. S. Ibelings, K. Maquelin, H. Ph. Endtz, H. A. Bruining, and G. J. Puppels. Rapid identification of candida spp. in peritonitis patients by raman spectroscopy. *European Society of Clinical Microbiology and Infectious Diseases*, 2005.

[14] Andrzej Kwiatkowski, Marcin Gnyba, Janusz Smulko, and Pawe Wierzba. Algorithms of chemicals detection using raman spectra. *METROLOGY AND MEASUREMENT SYSTEMS*, 2010.

[15] Derek A. Long. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*, volume 1. John Wiley Sons Ltd, 1th edition, 2002.

[16] Steve Lowry, Dick Wieboldt, Dave Dalrymple, Renata Jasinevicius, and Robert T. Downs. The use of a raman spectral database of minerals for the rapid verification of semiprecious gemstones. *Spectroscopy*, 2009.

[17] Fernando Rull Pérez and Jesus Martinez-Frias. Raman spectroscopy goes to mars. *Spectroscopy Europe*.

[18] R. Perez-Pueyo, M. J. Soneira, M. Castanys, and S. Ruiz-Moreno. Fuzzy approach for identifying artistic pigments with raman spectroscopy. *Applied Spectroscopy*, 2009.

[19] R. Perez-Pueyo, M. J. Soneira, and S. Ruiz-Moreno. A fuzzy logic system for band detection in raman spectroscopy. *Journal of Raman Spectroscopy*, 2004.

[20] G.D. Pitt, D.N. Batchelder, R. Bennett, R.W. Bormett, I.P. Hayward, B.J.E. Smith, K.P.J. Williams, Y.Y. Yang, K.J. Baldwin, and S. Webster. Engineering aspects and applications of the new raman instrumentation. *IEE Proceedings*, 2005.

[21] Georgios Pyrgiotakis, O. Erhun Kundakcioglu, PanosM. Pardalosc, and BrijM. Moudgil. Raman spectroscopy and support vector machines for quick toxicological evaluation of titania nanoparticles. *Journal of Raman Spectroscopy*, 2007.

[22] Pablo Manuel Ramos, Joan Ferré, Itziar Ruisnchez, and Konstantinos S. Andrikopoulos. Fuzzy logic for identifying pigments studied by raman spectroscopy. *Applied Spectroscopy*, 2004.

[23] Luke A. Reisner, Alex Cao, and Abhilash K. Pandya. An integrated software system for processing, analyzing, and classifying raman spectra. *Chemometrics and Intelligent Laboratory Systems*, 2010.

[24] Vitali Sikirzhytski, Kelly Virkler, and Igor K. Lednev. Discriminant analysis of raman spectra for body fluid identification for forensic purposes. *Sensors*, 2010.

[25] Ewen Smiths and Geoffrey Dent. *Modern Raman Spectroscopy A Practical Approach*, volume 1. John Wiley Sons Ltd, 1th edition, 2005.

[26] Pablon Sobrón, Francisco Sobón, Aurelio Sanz, and Fernando Rull. Raman signal processing software for automated identification of mineral phases and biosignatures on mars. *Applied Spectroscopy*, 2008.

[27] M. Castanys Tutzo, R. Perez-Pueyo, M. J. Soneira, and S. Ruiz Moreno. Fuzzy logic: a technique to raman spectra recognition. *Journal of Raman Spectroscopy*, 2005.

[28] Unidad Asociada UVa-CSIC a traves del Centro de Astrobiologia: Espectroscopia Avanzada en Ciencias de la Tierra y Planetarias Last Access 3 May 2012. http://tierra.rediris.es/erica.

[29] Peter Vandenabeele. Evaluation of a spectral searching algorithm for the comparison of raman band positions. *Spectrochimica Acta Part A 80*, 2011.

[30] Peter Vandenabeele, An Hardy, Howell G. M. Edwards, and Luc Moens. Evaluation of a principal components-based searching algorithm for raman spectroscopic identification of organic pigments in 20th century artwork. *Applied Spectroscopy*, 2001.

[31] Hana Vaskova. A powerful tool for material identification: Raman spectroscopy. *International journal of mathematical models and methods in applied sciences*, 2004.

[32] Hana Vaskova. Raman spectroscopy as an innovative method for material identification. *Recent Researches in Automatic Control*, 2004.

[33] Andrew T. Weakley, Peter R. Griffiths, and D. Eric Aston. Automatic baseline subtraction of vibrational spectra using minima identification and discrimination via adaptive, least- squares thresholding. *Applied Spectroscopy*, 2012.