Communications to SIMAI Congress,
ISSN 1827-9015, Vol. 3 (2009) 211 (10pp)
DOI: 10.1685/CSC09211

Automatic Recognition of Color Pigments from Raman Spectrum Analysis

G.M. Farinella[1], G. Gallo[1], A.M. Gueli[2], F. Stanco[1]

[1] *Dipartimento di Matematica e Informatica, University of Catania, Italy*
gfarinella@dmi.unict.it, gallo@dmi.unict.it, fstanco@dmi.unict.it

[2] *LDL&BBCC, Dipartimento di Fisica e Astronomia*
University of Catania, Italy
anna.gueli@ct.infn.it

Abstract

In this paper, an algorithm to automatically recognize pigments from Raman spectrum analysis is presented. The Raman spectroscopy is a technique to confirm the presence of specific pigments in a painting. Pattern Recognition techniques are applied over an existing database of Raman spectra in order to find a set of characteristic features. These information are used to automatically recognize unknown pigments by using ad-hoc similarity measure.

Keywords: `Cultural Heritage, Raman Spectra, Pigments, Conservation and Restoration.`

1. Introduction

Raman spectrometry is a diagnostic technique used for non-invasive tests on paintings and polychrome surfaces. Experts analyze the peaks of the Raman spectra to obtain different information about the art work, such as the identification of the painting techniques that have been used (e.g. to be linked to specific artist or its school), the ongoing degradation of the art work as well as the origin of the materials that have been used to create the art work (to recognize ancient pigments that were produced only in a few specific places) [1,2]. Moreover, the Raman spectrometry can be useful to recognize the component pigments in a mixture of pigments that have been used by artists to realize specific hues and chromatics effects. In this context a fundamental role is accomplished by computer based systems that are able to perform automatic analysis of the Raman spectra obtained during the experiments on art works (Fig. 1). The Raman spectrometry hardware is usually coupled with commercial software that help experts on

Received DD MM 2009, in final form DD MM 2009
Published DD MM 2009

performing analysis of the data obtained by their experiments. Standard datasets of materials are usually used by these commercial software.

In this paper we propose a new algorithm for automatic recognition of color pigments from Raman spectra. Experimental results on a dataset containing samples of Raman spectra of different pigments, acquired using different excitation wavelengths and analysed dry and applied with four painting technique (egg tempera, casein tempera, oil and fresco), show that the proposed approach achieves good performances.

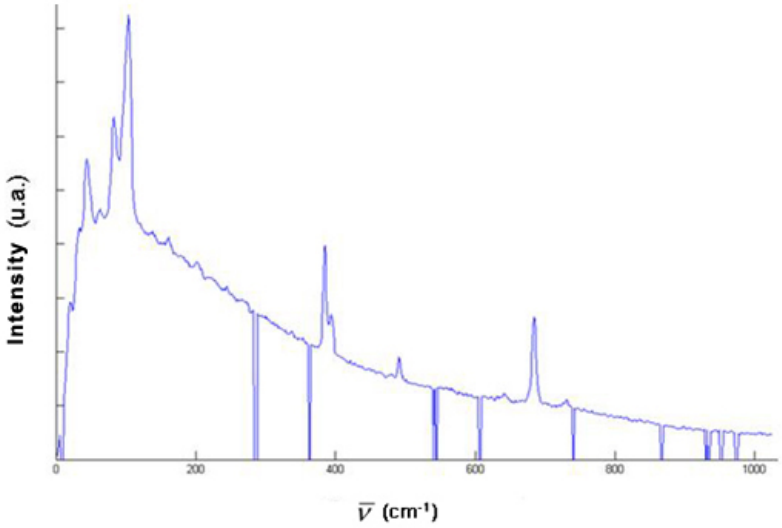


Fig. 1. Example of Raman spectra obtained during the experiments.

2. Proposed method

It is well known that experts look at the peaks of the Raman spectra to recognize the pigment that have been generated the signal. This section proposes an algorithm for the automatic identification of the peaks contained in a Raman spectra. The identified peaks are then represented in a multi-resolution way and used in pair with a recognition algorithm to automatically infer the corresponding pigment. The overall pigment recognition system is composed by the following two modules: 1) peaks detection and representation; 2) pigment recognition.

DOI: 10.1685/CSC09211

2.1. *Peaks Detection and Representation*

This module is used to identify the most important peaks of the considered Raman spectra and to represent them taking into account the position and the intensity of the signal. Since peaks of a Raman spectra obtained on a specific pigment can be shifted in their position, a multi-resolution approach is proposed to make more robust the final representation. The algorithm used to detect and represent the samples of a Raman spectra of a pigment is the follows:

Step 1. The Raman spectra is smoothed to deal with discontinuity due the absence of signal in the positions in which the spectra have intensity equal to zero (Fig. 2). The smoothed signal is obtained changing the values in the position having zero intensity with the value of the nearest position with intensity value different than zero.

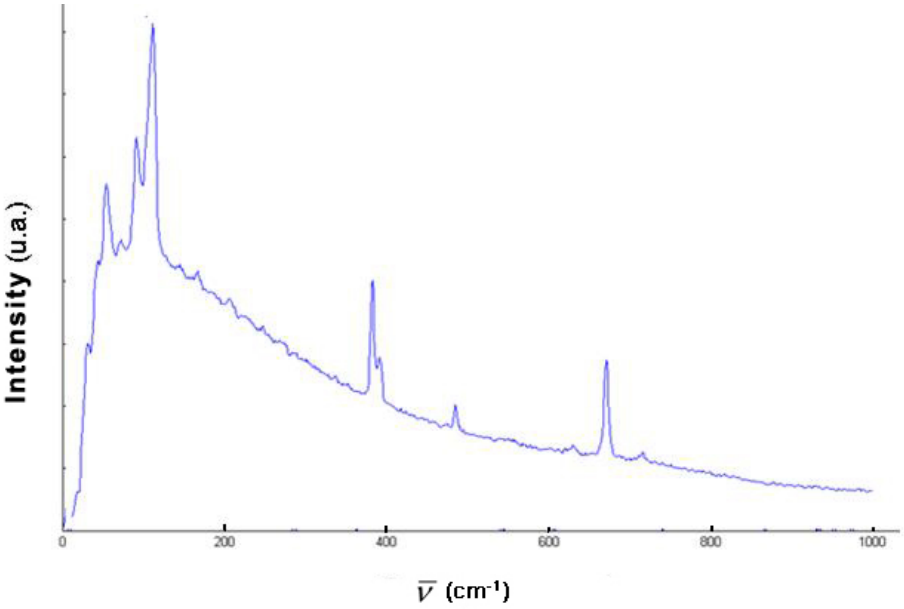


Fig. 2. Step 1: smoothing the positions with intensity equal to zero.

Step 2. To remove the noise of the Raman spectra a Local Regression Smoothing (LRS) approach is used ([3]). LRS performs a local regression of the intensities belonging to a predefined span windows of positions (see the example in Fig. 3). After this step, the most important peaks of the obtained signal (Fig. 4) are preserved and the noise is removed.

Step 3. The first and second order derivative of the Raman spectra is computed for peaks detection purpose (Fig. 5). Specifically, the positions

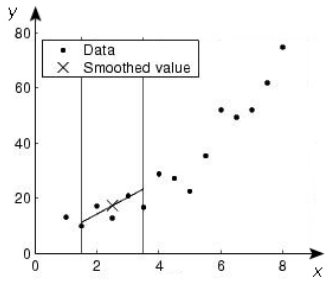


Fig. 3. An example of the algorithm Local Regression Smoothing (LRS) on a subspace of R^2.

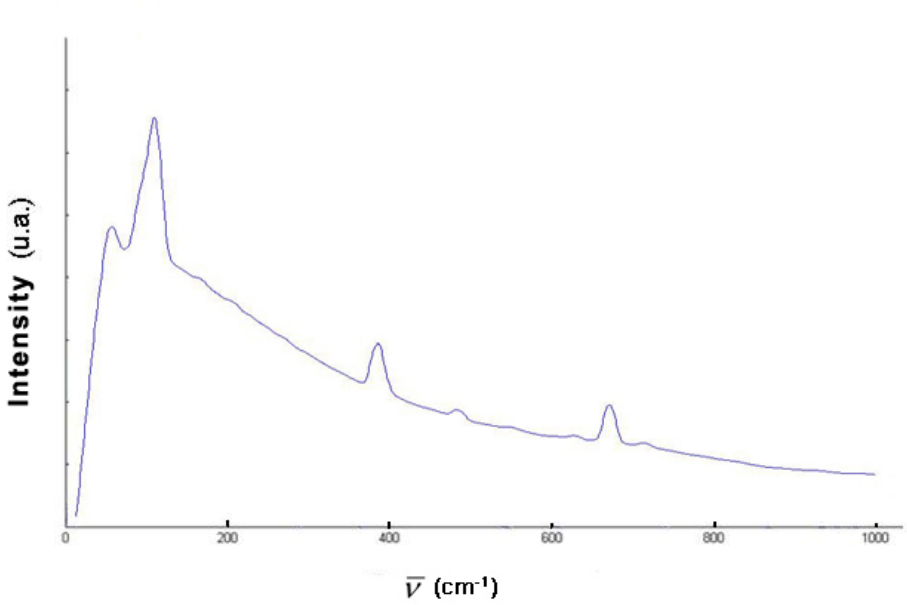
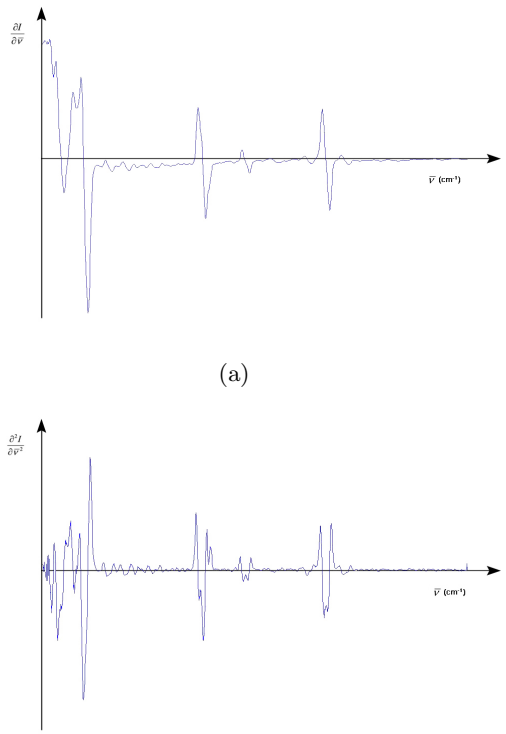


Fig. 4. Step 2: The Raman spectrum in Fig. 2 is processed with a LRS to deal with noise.

in which the second order derivative is negative are selected. The absolute value of the second order derivative of the selected peaks are used to establish how much the peaks are relevant (Fig. 6) by using a Gaussian probabilistic approach. The mean and variance of the Gaussian distribution are computed on the absolute value of the second order derivative of the selected peaks. The peaks having probability greater than a predefined threshold are selected as relevant peaks.

Specifically, a pick is considered significant if the corresponding prob-

DOI: 10.1685/CSC09211



(a)

(b)

Fig. 5. The first (a) and second (b) order derivative of the signal in Fig. 4.

ability is greater than the probability of the Gaussian distribution corresponding to the value $3\sigma^2$. A weight, proportional to the corresponding absolute value of the second order derivative, is assigned to each important peak. The sum of all weights is normalized to be equal to 1. Fig. 6 shows an example of the distribution obtained by using the described process on the signal of Fig. 5.

Step 4. A multi-resolution representation of the most important peaks is obtained to deal with problems due small variation of peaks in terms of positions. The probabilities of couples of adjacent bins are iteratively grouped until a predefined resolution is reached (Fig 7). The final representation is obtained as concatenation of the distribution obtained during this process. The final representation of each Raman spectra is normalized in order to have sum equal to 1. The final representation of each Raman spectra is stored as template of the corresponding pigment.

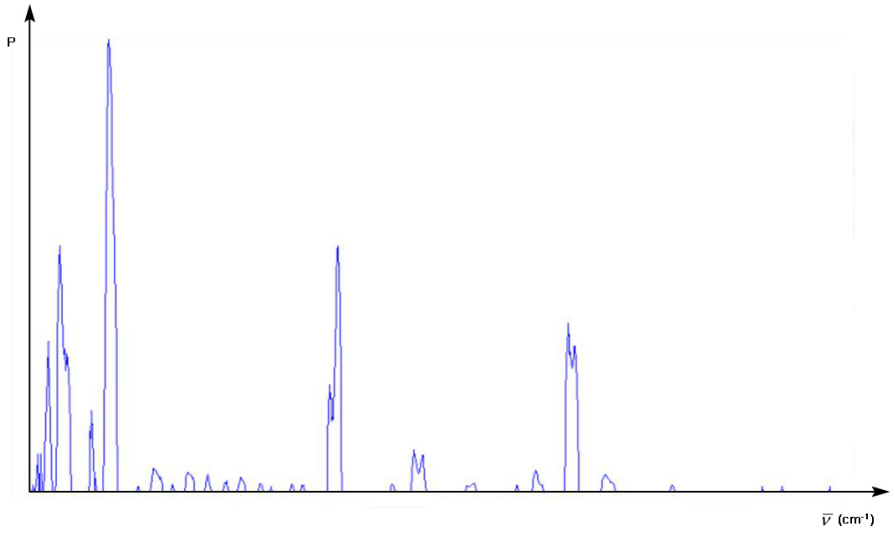


Fig. 6. In the y axis is reported the probability to observe relevant peaks given a position $\overline{\nu}$.

2.2. *Pigment Recognition*

We adopt the K-nearest neighbor approach ([4]) to recognize an unknown pigment represented ad described above. Specifically, the similarity between a Raman spectra of an unknown pigment and samples of Raman spectra of known pigments (stored in a database) is computed after representing them with the approach proposed in previous section. We propose to use a similarity measure based on the Bhattacharaya coefficient ([5]) to compute similarity between samples:

$$d(P,Q) = \sum_{l} \frac{1}{2^{l-1}} \sqrt{1 - \sum_{i} \sqrt{P_l^i * Q_l^i}} \tag{1}$$

where P and Q are the distribution representation of two Raman spectra and l indicates the level of resolution under consideration in the multiresolution representation. The idea beyond the proposed similarity measure is that the similarity between two distribution computed with the metric based on Bhattacharaya coefficient on greater resolution should be weighted more than the similarity computed on distribution at lower resolutions. This means that peaks of different Raman spectra having the same position in greater resolution are weighted more in terms of similarity with

DOI: 10.1685/CSC09211

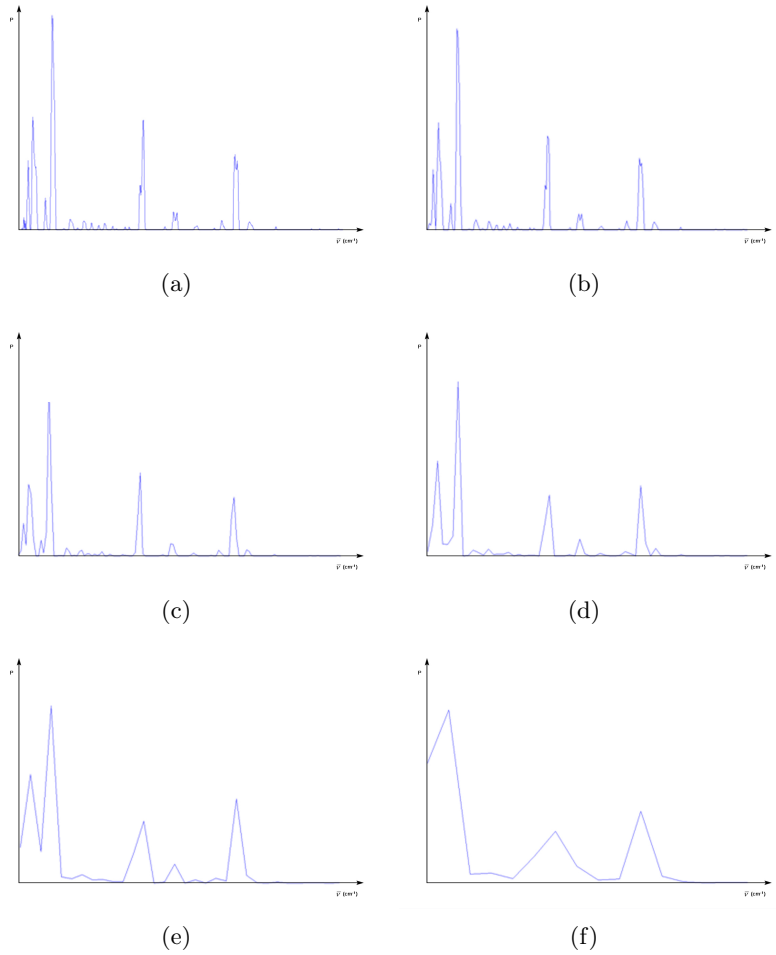


Fig. 7. Multi-resolution representation of the probability distribution in Fig. 6. In our experiment the following resolution are used: (a) 512 bins; (b) 256 bins; (c) 128 bins; (d) 64 bins; (e) 32 bins; (f) 16 bins.

respect to peaks of different Raman spectra having the same position in lover resolution. In the experiments we have used 7 level of resolution ($l = 1, 2, 3, 4, 5, 6, 7$ with corresponding $bins = 1024, 512, 256, 128, 64, 32, 16$). The distribution and weight at level $l + 1$ is obtained from the distribution and weight at level l.

3. Experimental results

To test the proposed approach we have used 196 different Raman spectra obtained on different pigments applied with different methods (60 different samples have been applied with casein, 49 with oil, 57 with egg, 30 with fresco). Recognition performances have been obtained applying the leave one out cross validation methodology. Fig. 8 shows examples of some pigments used in our experiments. The recognition performances of the proposed approach are reported in Table 1.

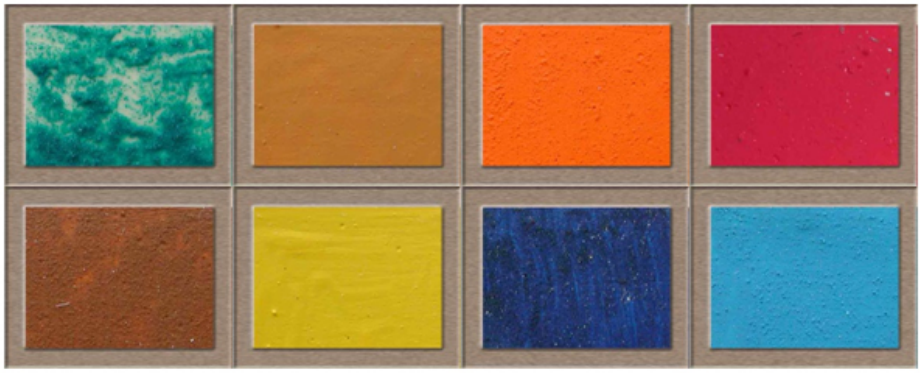


Fig. 8. Examples of pigments used in our experiments.

The accuracy obtained considering all the experiment is 71,43%. The best accuracy have been obtained when pigments are applied with oil (accuracy 81,63 %), whereas the worst accuracy have been obtained when pigments are applied with a fresco technique (46,67 %).

Table 1. Automatic classification results.

Type	Number of Samples	Accuracy
Casein	60	80,00%
Oil	49	81,63%
Egg	57	66,76%
Fresco	30	46,67%
TOTAL	196	71,43%

Fig. 9 shows an example where the proposed method fails. As shown in Fig. 9 the main errors occur when the unknown Raman spectra of a specific class (the class Test number 25 in Fig. 9) is very similar to a Raman spectra belonging to a different class, making the recognition task difficult also for the expert in the field. In the example of Fig. 9 the pigment of Iron Oxide/Mars Orange (the red signal in Fig. 9), is erroneously associated by

DOI: 10.1685/CSC09211

the algorithm to the pigment Hostaperm Orange

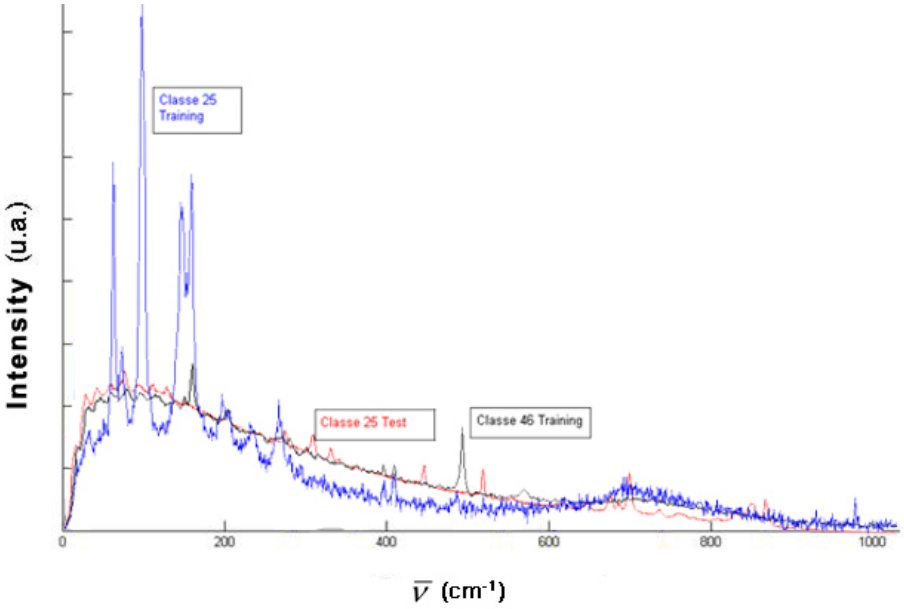


Fig. 9. Misclassification example. The class 25 (Iron Oxide / Mars Orange) have been associated to the class 46 (Hostaperm Orange).

4. Conclusions and future works.

In this paper we proposed a new approach for automatic recognition of color pigments from Raman spectra. The overall accuracy obtained by our experiments is 71,43%. Future work will be devoted on improving the representation extracted to better recognize the challenging Raman spectra (e.g. pigments applied with a fresco) and to identufy the component pigments from a mixture of them.

REFERENCES

1. F. Rull Perez. Application of IR and Raman Spectroscopy to the study of Medieval Pigments, Handbook of Raman Spectroscopy, vol. 28, pp. 835-862, 2001.
2. A. Bhattacharaya. On a measure of divergence between two statistical population defined by probability distributions. Bulletin of Calcutta Mathematical Society, vol. 35, pp. 99-110, 1943.
3. W.S. Cleveland. Robust locally weighted regression and smoothing scatterplots. In journal of the American Statistical Association, vol. 74 pp.

829-836, 1979.

4. C.M. Bishop, Pattern Recognition and Machine Learning. Springer.
5. D. Comaniciu, V. Ramesh, and P. Meer. Kernel-Based Object Tracking. In IEEE Trans. Pattern Analysis Machine Intell., Vol. 25, No. 5, pp. 564-575, 2003.