ELSEVIER

Chemometrics and Intelligent Laboratory Systems 39 (1997) 103–114

Chemometrics and intelligent laboratory systems

Principal component analysis of visible and near-infrared multispectral images of works of art

S. Baronti, A. Casini, F. Lotti, Simone Porcinai *

Istituto di Ricerca sulle Onde Elettromagnetiche "Nello Carrara"-C.N.R., Via Panciatichi, 64, 50127 Firenze, Italy

Received 15 October 1996; accepted 20 February 1997

Abstract

Principal component analysis (PCA) was applied to a very simple case of a tempera panel painted with four known pigments (cinnabar, malachite, yellow ochre and chromium oxide). The four pigments were spread pure as well as dilute with carbon black (5% w/w, 10% w/w) thus creating 12 homogeneous areas of the same size. The panel was imaged by a Vidicon camera in the visible and near-infrared regions (420–1550 nm) resulting in a set of 29 images. PCA was applied by taking various subsets of the input data. From the analysis of this simple and predictable case study some guidelines are synthesized and proposed for the application to actual work of art. Results are presented for the painted panel. Preliminary results are also reported for the Luca Signorelli's "Predella della Trinità". The multivariate image analysis results in the visible and near-infrared regions show that it is possible to use the multispectral image data in order to get a segmentation and a classification of painted zones by pigments with different chemical composition or physical properties. © 1997 Elsevier Science B.V.

Keywords: Principal component analysis; Imaging spectroscopy

1. Introduction

After ten years of experience at our institute, we are in a position to affirm that fiber-optic reflectance spectroscopy (FORS) in the visible and short-wavelength infrared is a powerful technique for attaining good knowledge of the constituent materials and of the degradation products of paintings, without causing any damage to them [1,2].

Among non-invasive spectroscopic techniques, it is also the most functional, as it can be performed with transportable equipment. Its fiber-optic probes can be easily handled to collect data from as many points as needed, without limits to the size of the object. However a collection of local 1D spectrograms, though rich in information to a chemist, cannot be readily correlated to the visual appearance of the work of art.

A technique which promises to be even more efficient than punctual reflectance spectroscopy is its 2D extension: imaging spectroscopy or multispectral imaging, which supplies very readable data by retaining the visual aspect of the objects. Although suitable imaging spectroscopes are not yet available, valid test data can be collected using a set of optical filters which select the wavelength of the radiation detected by a camera.

* Corresponding author. Tel.: +39-55-4235273; fax: +39-55-410893; e-mail: porcinai@iroe.iroe.fi.cnr.it.

0169-7439/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.
PII S0169-7439(97)00047-6

Table 1
The four pigments used to paint the panel

Pigment	Formula	λ_D (nm)
Cinnabar	HgS	619
Malachite	$Cu_2CO_3(OH)_2$	510
Yellow ochre	FeO(OH)	583
Chromium oxide	Cr_2O_3	560

Table 2
The cumulative variance for the first five principal components from 29 VIS-NIR images of the panel

PCs	Var. Cum. (%)
1	86.83
2	93.49
3	97.72
4	98.92
5	99.62

For many years now, Principal component analysis (PCA) has been a standard method that is used for the compression and the enhancement of remotely sensed multispectral images [3–8]. In this paper we applied PCA to the multispectral images of a panel painted with four known pigments, in order to perform a methodological investigation of the sequence of images and of the homogeneous zones of the object. Finally, a preliminary study of an actual work of art on exhibit at the Uffizi Gallery, "Predella della Trinità" by Luca Signorelli, is presented.

2. Experimental

The imaging setup used for collection of the multiwavelength images was assembled around a PbO–

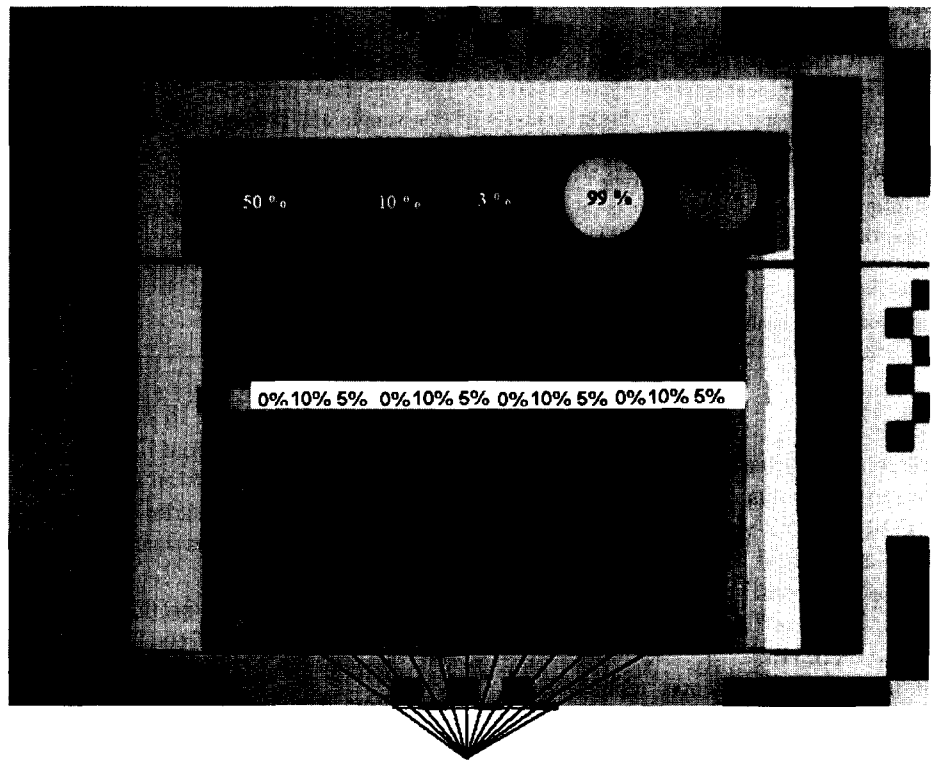

Fig. 1. The panel used for acquiring spectroscopic data which were used in PCA. The percentages in the middle of the panel indicate the content of carbon black. The shorter stripes placed above and the longer ones placed below the percentage labels are painted in the same way. The percentages on the SPECTRALON standards report their reflectances.

PbS vidicon camera (HAMAMATSU C2400-03). Spectral selection was realized according to a band sequential approach, by means of 29 narrow-band optical filters (range 420 nm–1550 nm) with full width at half maximum of ~ 10 nm. Their distribution in the 420–700 nm range is every 20 nm; in the 750–1200 nm range the central wavelength varies with steps of 50 nm. The other filters have central wavelengths at 1300 nm, 1400 nm, 1445 nm and 1550 nm.

Illumination was supplied by two projectors placed symmetrically at the sides of the camera with beam incidence angles of ~ 45°.

For noise reduction, each image was the result of 100 image averages, performed at video rate by a MATROX-LC image processing board.

In order to normalize the response of the collected images, five SPECTRALON reflectance standards (1.25″ D) are fixed above the panel by means of a cardboard frame: two of the standards (99% and 3% reflectance) are used for the normalization according to a linear model (without intercept) of the camera response. The other three reflectance standards (75%, 50%, 10%) are used to control the quality of the measurement.

Image spectroscopy tests were performed on a panel (196 mm × 147 mm) prepared with a gypsum layer and painted with four pigments using a tempera medium (Fig. 1). Several carbon black lines were drawn on the gypsum layer in order to simulate the presence of underdrawings, the search for which is one of the main objectives of our investigations. The pigments were chosen so that their dominant wavelengths (λ_D) were evenly distributed throughout the entire visible spectral range. The four pigments are reported in Table 1.

Three rectangular zones for each pigment were painted using the pure pigment and two mixtures, 5% w/w and 10% w/w of carbon black. We thus obtained 12 zones which cover almost all the panel's

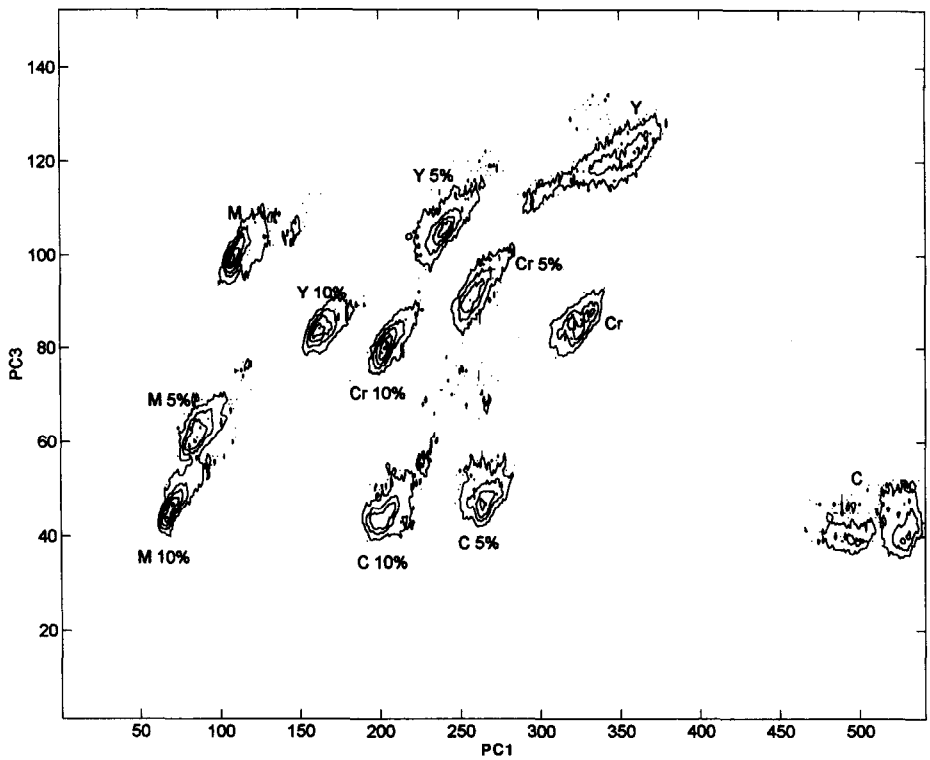

Fig. 2. The score plot PC3/PC1 from PCA on 29 VIS-NIR images of the panel. The level lines are related to the following pixel density values: 3, 20, 50, 100, 150, 190. (C) cinnabar; (Y) yellow ochre; (M) malachite; (Cr) chromium oxide; The composition of the mixtures pigments-carbon black is shown by the percentages (w/w) reported in the diagram.

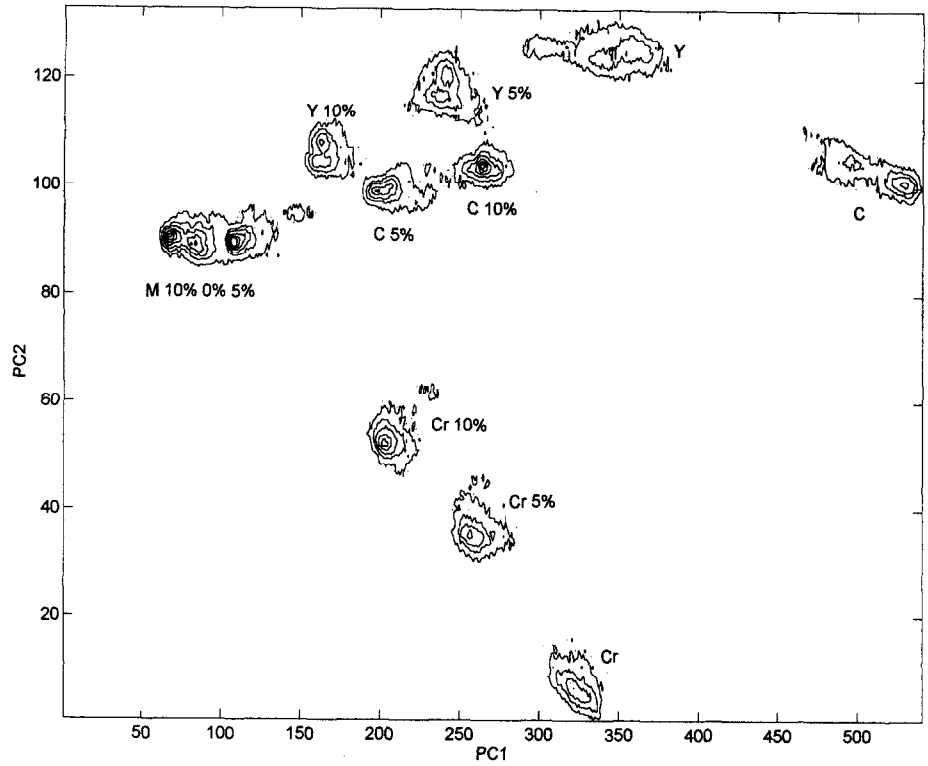

Fig. 3. The score plot PC2/PC1 from PCA on 29 VIS-NIR images of the panel. The level lines are related to the following pixel density values: 3, 20, 50, 100, 150, 190. (C) cinnabar; (Y) yellow ochre; (M) malachite; (Cr) chromium oxide; The composition of the mixtures pigments-carbon black is shown by the percentages (w/w) reported in the diagram.

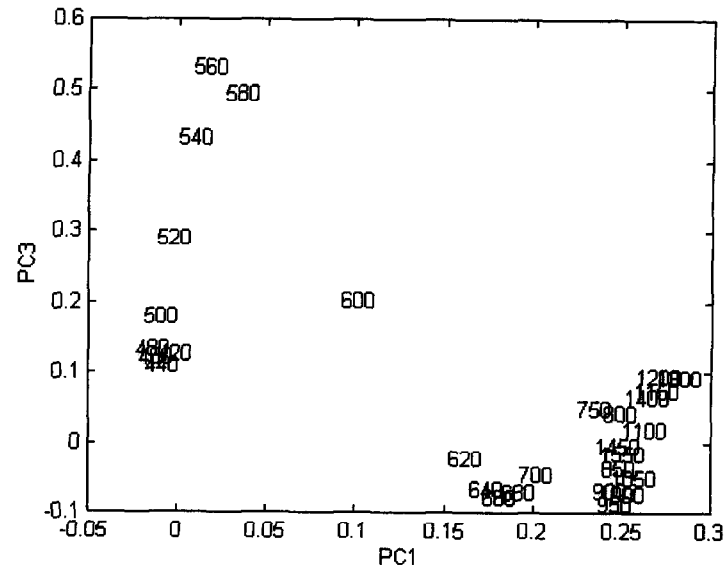

Fig. 4. The weight plot PC3/PC1 from PCA on 29 VIS-NIR images of the panel.

surface, that were differentiated for their λ_D value (of the pigment) and for their brightness (of the carbon black percentage). A set of control measurements were performed on 5 points for each stripe using a Zeiss mod. MMS1 fiber optic spectrometer. The source is a 20 W tungsten–halogen lamp and the light beam is sent to and collected from the sample via two optical fiber bundles operating in the range of 400–2200 nm. The received radiation is diffracted by a holographic concave grating and then detected by a photodiode array which measures 512 values over a spectral interval from 400 nm to 1000 nm. The wand-type probe was modified by connecting the distal end of the optical fiber bundle to a reflection-measuring head. This device makes it possible to exclude the radiation from the environment, and to collect the scattered radiation with a viewing angle of 45° with respect to the incident direction.

The initial data matrix used for PCA is a $n \times v$ matrix where n is the number of objects (number of pixels considered in the PCA: about 86000 in the case of the panel; about 237000 in the case of the Luca Signorelli's predella) and v is the number of variables (wavelengths or spectral bands). No pretreatment was adopted before PCA.

3. Results and discussion

The PCA of the panel VIS-NIR (420 nm–1550 nm) images explains 99.62% of the data matrix variance after extraction of the first five components (Table 2).

Owing to the very high number of pixels considered in the PCA on the multispectral images of the panel, a very clear representation of the classes of

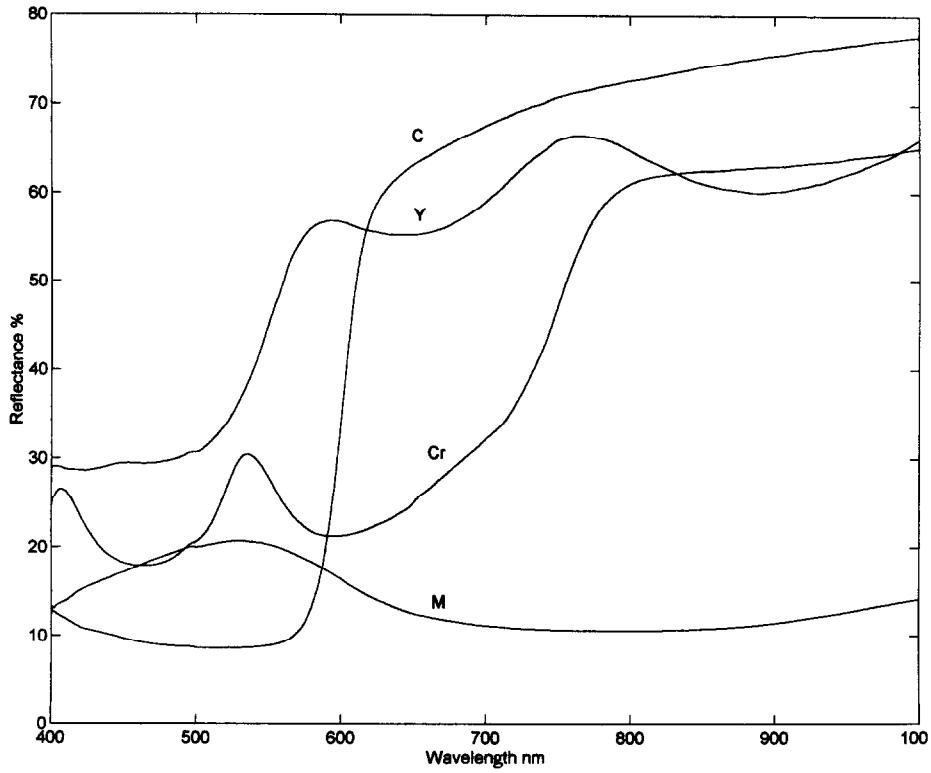

Fig. 5. The NIR spectra of the four pigments used for painting the panel. (C) cinnabar; (Y) yellow ochre; (Cr) chromium oxide; (M) malachite.

pixels on the score plots can be obtained by means of density pixel level lines (Fig. 2). Usually two (or more) classes of pixels, which appear superimposed one upon the other, are characterized by two distinct pixel density peaks. The score plot which offers the clearest separation into clusters of the 12 painted zones on the panel is the projection on the PC3/PC1 plane. In fact on the score plot PC1/PC2 (Fig. 3) the chromium oxide clusters (pure chromium oxide and two mixtures 5% 10% with carbon black) are well separated from the clusters of the other pigments by means of PC2 axis, whose eigenvector is very similar to the chromium oxide spectrum.

Consequently, the discrimination among the 12 stripes is not as good as using the PC1/PC3 score plot. In particular, by using the score plot PC1/PC2, we did not obtain a clear separation among the three clusters of malachite pigment and its mixtures with carbon black.

By projecting each class onto the image space, it can be noted that a distinct dense class of pixel exists for each different zone on the panel. For each class on the PC3/PC1 score plot many low pixel density zones around a central high-density zone can be seen. The pixels outside the classes were identified as pixels near the boundary of the painted zones on the

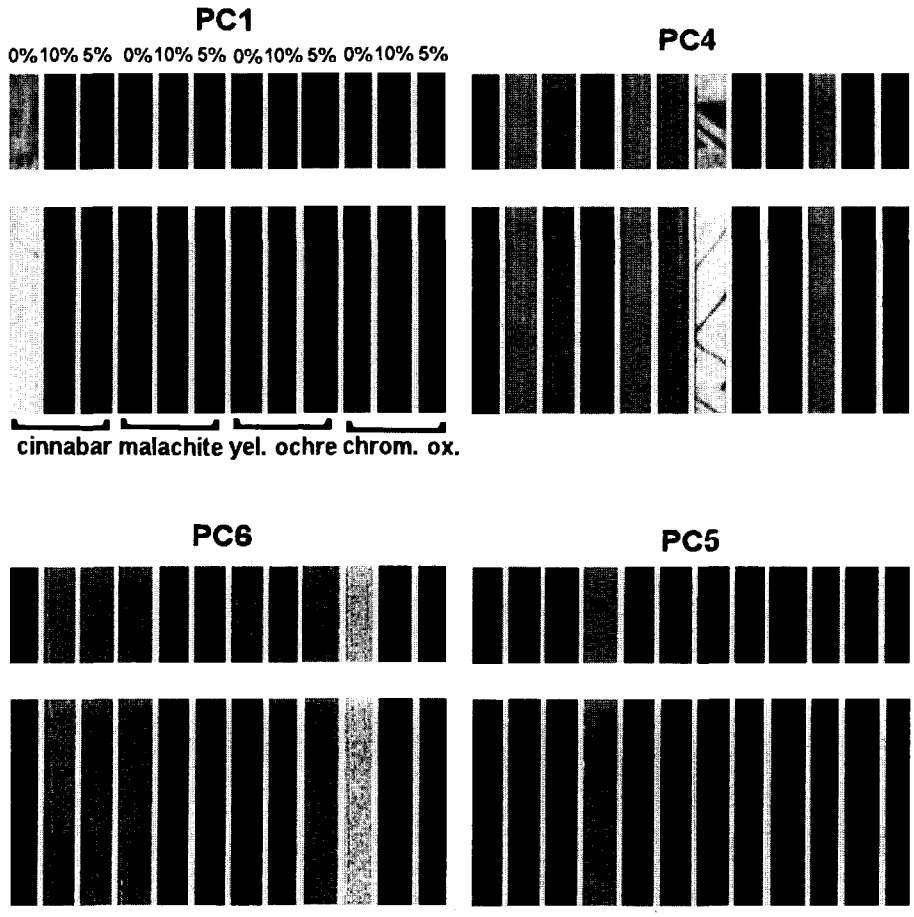

Fig. 6. The PC1, PC4, PC5 and PC6 images of the panel from PCA on 29 VIS-NIR images.

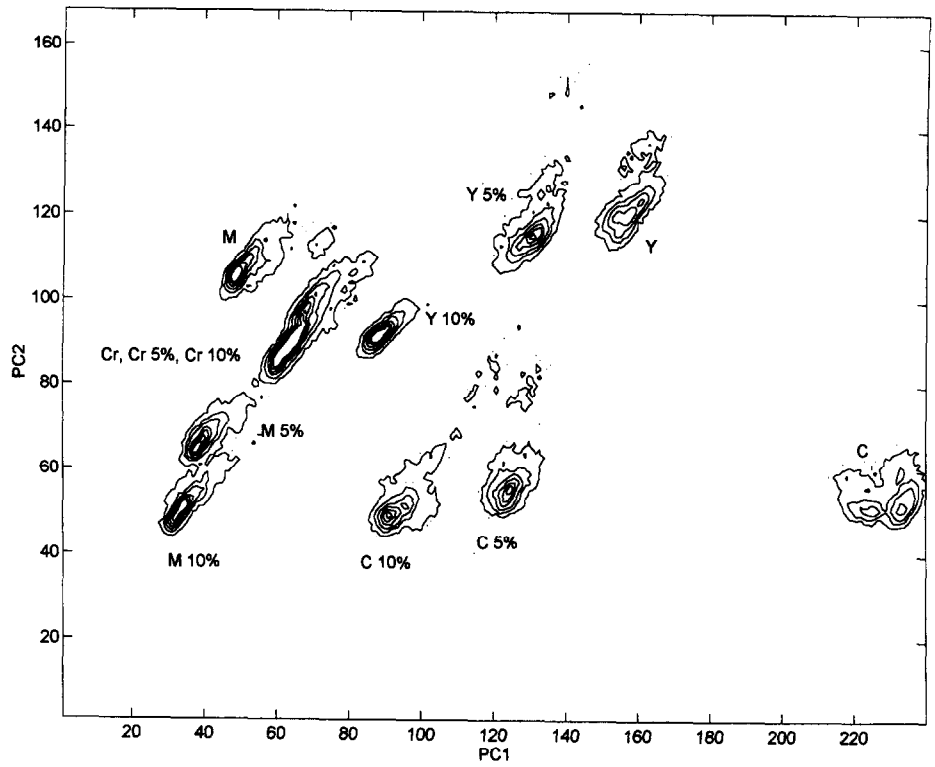

Fig. 7. The score plot PC2/PC1 from PCA on 15 VIS images of the panel. (C) cinnabar; (Y) yellow ochre; (M) malachite; (Cr) chromium oxide; The composition of the mixtures pigments-carbon black is shown by the percentages (w/w) reported in the diagram.

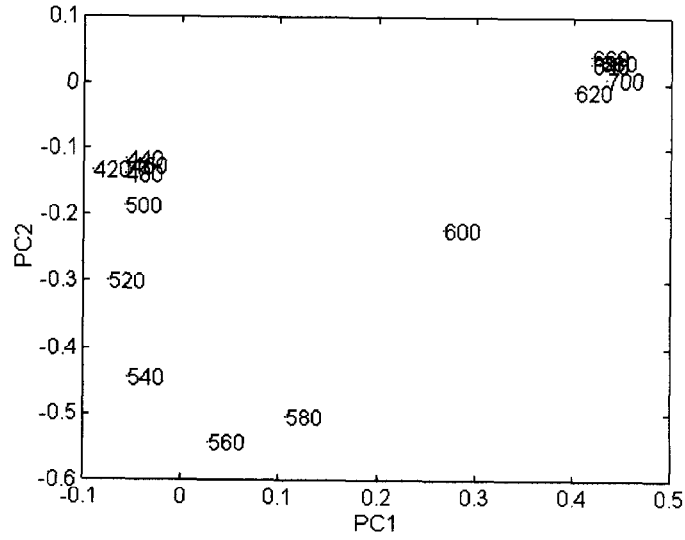

Fig. 8. The PC2/PC1 loadings plot from PCA on 15 VIS panel images.

tablet. Their outlier position on the score plot is due to the anomalous reflectance value at the boundary between two contiguous zones with different but homogeneous reflectivity. Furthermore, for the object investigated, it is possible to appreciate the differences, not only between different pigments, but also between zones painted by the same pigment mixed with a different percentage of carbon black.

The variables (wavelength bands in our case) loadings plot can be helpful in understanding the meaning (i.e., the common features) of the groups. Fig. 4 shows how the different variables (wavelength bands) influence the PC1 and PC3 eigenvectors. The PC1 eigenvector receives a greater contribution from the near-infrared bands (83.8%) than the visible ones (16.2%); all the near-infrared bands have a positive and almost equal weight on the PC1, which represents an 'average grade' in this spectral region. Accordingly, each pure pigment cluster is characterized by a larger value of PC1 coordinate compared to its mixtures with carbon black (Fig. 2). Indeed, the latter absorbs the electromagnetic radiation over the entire near-infrared region. The visible bands 420–600 nm have a strong influence (92.7%) on the PC3 (the sum of the squared corresponding elements is equal to the 92.7% of square norm of the PC3 eigenvector). In particular, the bands at 560 nm and 580 nm make the highest contribution in the visible region. In fact, the cinnabar and the yellow ochre pure pigments have the lowest and the highest reflectivity at 560–580 nm region respectively, among the four pigments used, (Fig. 5); the former is positioned at the bottom, while the latter is located at the top of the PC3/PC1 score plot.

It is interesting to note how the underdrawing is revealed on the PC1, PC4 and PC5 component images, which are heavily weighted by the near-infrared bands. The underdrawings are especially visible on the yellow ochre stripe. Moreover, by comparing the PC4 and PC5 images with the PC6 image, it is evident how the internal spatial correlation is still high in the PC4 and PC5 images, while the PC6 begins to be noisy (Fig. 6).

In order to visualize the variability among the different zones on the panel, we tried to represent the panel image by using false colors, by associating the RGB (Red, Green and Blue) channels to the first three PCs. The color image obtained contains 97.72% of

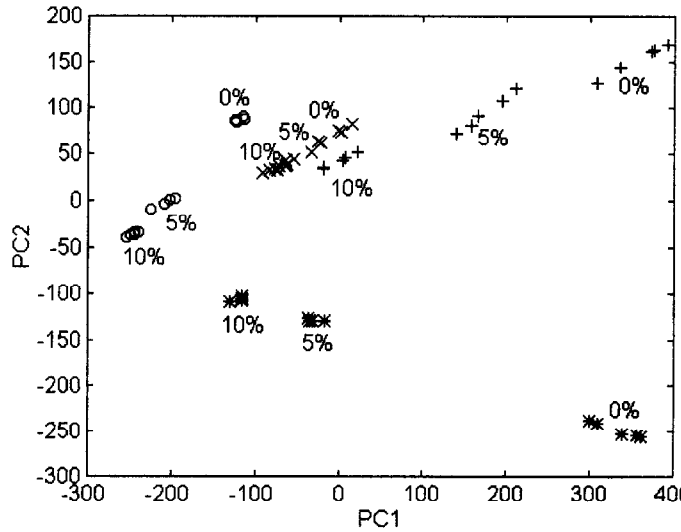

Fig. 9. The score plot PC2/PC1 from PCA on visible spectra recorded by FORS on the panel stripes. (*) cinnabar; (+) yellow ochre; (×) malachite; (○) chromium oxide. The composition of the mixtures pigments-carbon black is shown by the percentages (w/w) reported in the diagram.

the information in the original 29 spectral bands. The principal component technique is thus one of the most efficient ways for selecting data to be assigned to the three fundamental colors in order to produce color displays of multi-band images.

As regards the images recorded in the visible range only, Fig. 7 (PC2/PC1 score plot) shows that the samples are roughly separated into two groups along the PC1 coordinate: the yellow and red pigments (yellow ochre, cinnabar and their mixtures with carbon black) are characterized by a PC1 value higher than that of the green pigments (malachite, chromium oxide and their mixture with carbon black). Again, this separation is due to the different reflectivity values in the 600 nm–700 nm region, as can be seen by observing the PC1 weight (Fig. 8) and the visible spectra of the aforesaid two groups of samples (Fig. 5): the cinnabar and yellow ochre spectra have a higher reflectivity than malachite and chromium oxide spectra in this region. Only in the case of chromium oxide pigment, we did not obtain a complete separation of the pure pigment from its carbon-black mixtures.

To verify the results obtained from imaging spectroscopy, we performed the PCA on visible spectra collected with the spectrometer on five points for each stripe of the panel. The PCA results from both spectroscopic techniques are in good agreement. In Fig. 9 are reported the score plots PC2/PC1 and in Fig. 10 the eigenvectors' shape for the first three components from both spectroscopic data sets. There are many analogies in the relative position of the groups of the different pigments in the principal component space as well as in the shape of PC eigenvectors.

The methodology explained above can be profitably applied to actual works of art even if the interpretation of the results becomes more difficult. For the study of Signorelli's "Predella della Trinità", 29 multispectral images were recorded in the visible and near-infrared regions. The plot of the first eigenvector (PC1) after application of PCA to the 29 VIS-NIR images shows that the main contribution to the total variance of the data set is given by the NIR bands, which have similar weight, showing that the variance accounted from them is mainly due to the variations of the average reflectance (Fig. 11). Due to the high penetration of NIR radiation, it is interesting to evaluate with precision the spectral behavior of the pixels in this region, to evidence information from deeper layers. To this purpose, and to avoid that higher variations in the visible range could mask

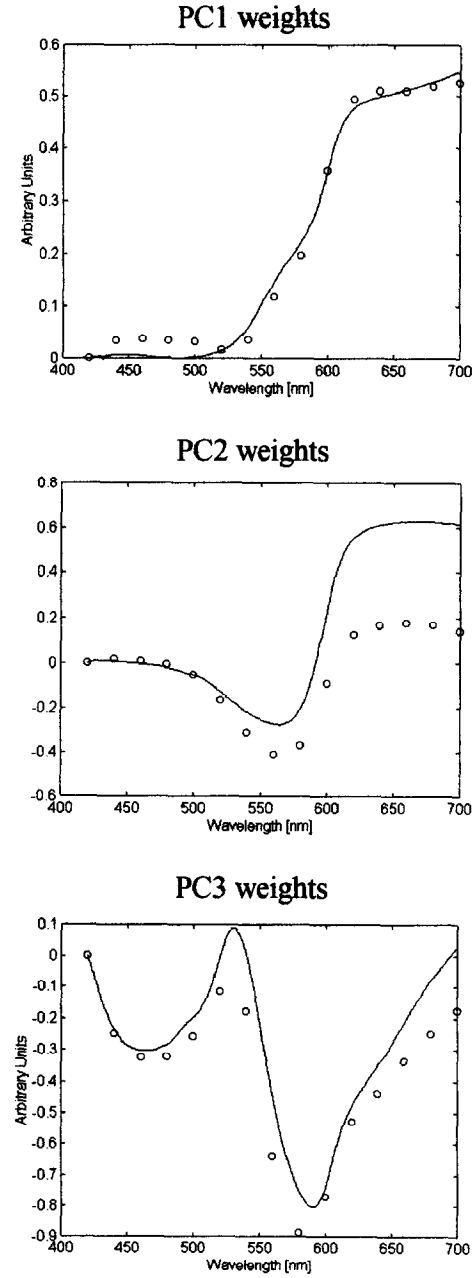

Fig. 10. The shape of the eigenvectors for the first three components from both spectroscopic techniques. (○) eigenvectors from PCA on multispectral images; (—) eigenvectors from PCA on fiber-optic reflectance visible spectra.

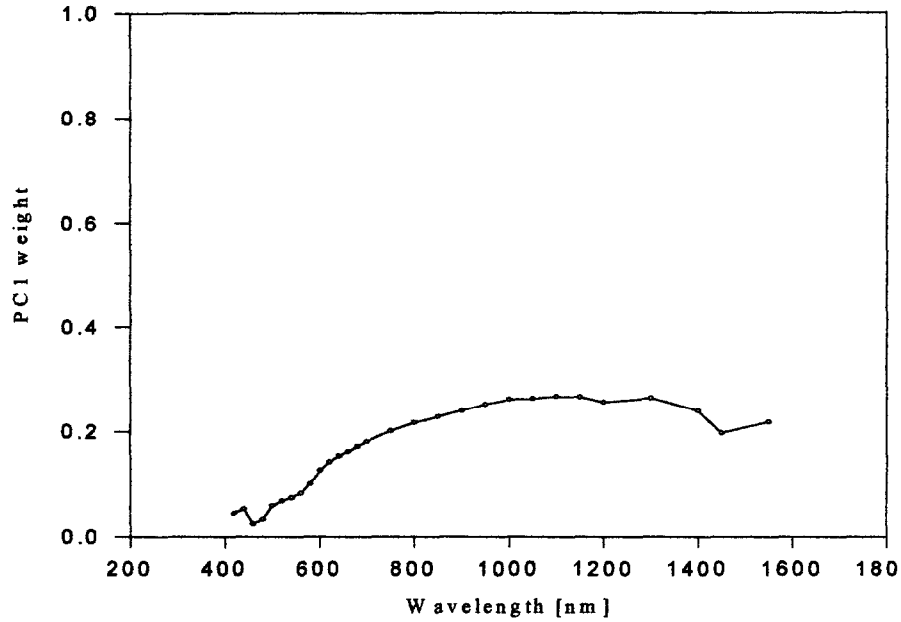

Fig. 11. PC1 Eigenvector from PCA on 29 VIS-NIR images of the Signorelli's predella.

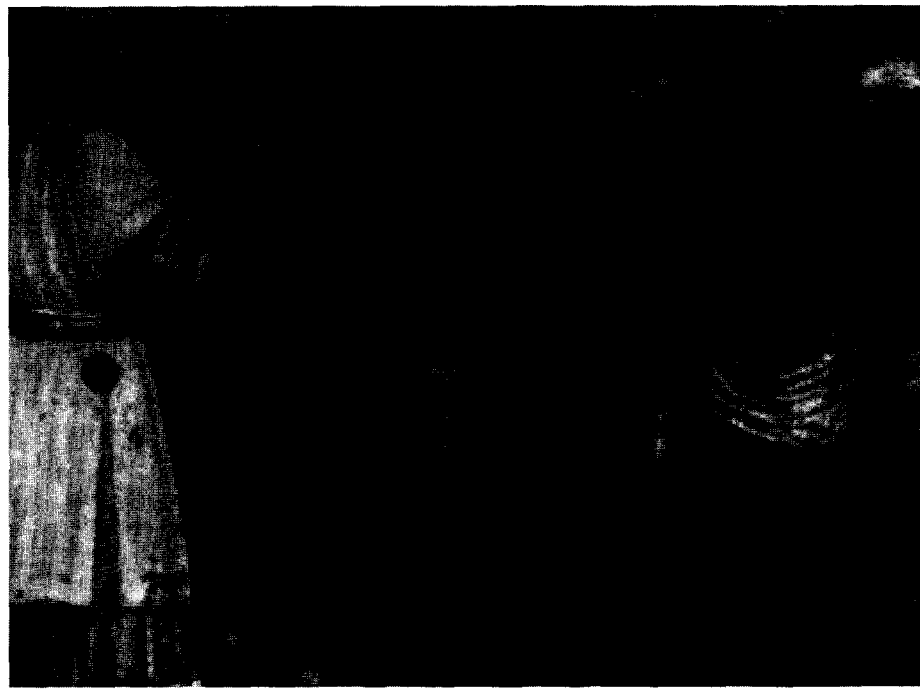

Fig. 12. The PC1 image of the "Predella della Trinità" from PCA on 14 NIR images.