

Applying Hyperspectral Reflectance Imaging to Investigate the Palettes and the Techniques of Painters

Michele Caccia¹, Simone Caglio¹, Anna Galli^{1,2}, Matteo Interlenghi², Isabella Castiglioni^{2,3}, Marco Martini¹

¹ Dipartimento di Scienza dei Materiali, Università degli Studi di Milano-Bicocca ² CNR-IBFM ³ Dipartimento di Fisica, Università degli Studi di Milano-Bicocca

Corresponding Author

Michele Caccia

michele.caccia@unimib.it

Citation

Caccia, M., Caglio, S., Galli, A., Interlenghi, M., Castiglioni, I., Martini, M. Applying Hyperspectral Reflectance Imaging to Investigate the Palettes and the Techniques of Painters. *J. Vis. Exp.* (172), e62202, doi:10.3791/62202 (2021).

Date Published

June 18, 2021

DOI

10.3791/62202

URL

jove.com/video/62202

Abstract

Reflectance Spectroscopy (RS) and Fiber Optics Reflectance Spectroscopy (FORS) are well-established techniques for the investigation of works of art with particular attention to paintings. Most modern museums put at the disposal of their research groups portable equipment that, together with the intrinsic non-invasiveness of RS and FORS, makes possible the *in situ* collection of reflectance spectra from the surface of artefacts. The comparison, performed by experts in pigments and painting materials, of the experimental data with databases of reference spectra drives the characterization of the palettes and of the techniques used by the artists. However, this approach requires specific skills and it is time consuming especially if the number of the spectra to be investigated becomes large as is the case of Hyperspectral Reflectance Imaging (HRI) datasets. The HRI experimental setups are multi-dimensional cameras that associate the spectral information, given by the reflectance spectra, with the spatial localization of the spectra over the painted surface. The resulting datasets are 3D-cubes (called hypercubes or data-cubes) where the first two dimensions locate the spectrum over the painting and the third is the spectrum itself (i.e., the reflectance of that point of the painted surface versus the wavelength in the operative range of the detector). The capability of the detector to simultaneously collect a great number of spectra (typically much more than 10,000 for each hypercube) makes the HRI datasets large reservoirs of information and justifies the need for the development of robust and, possibly, automated protocols to analyze the data. After the description of the procedure designed for the data acquisition, we present an analysis method that systematically exploits the potential of the hypercubes. Based on Spectral Angle Mapper (SAM) and on the manipulation of the collected spectra, the algorithm handles and analyzes thousands of spectra while at the same time it supports the user to unveil the features of the samples under investigation. The power of the approach is

illustrated by applying it to *Quarto Stato*, the iconic masterpiece by Giuseppe Pellizza da Volpedo, held in the Museo del Novecento in Milan (Italy).

Introduction

Reflectance Spectroscopy (RS) and Fiber Optics Reflectance Spectroscopy (FORS) are based on the detection of the light reflected by surfaces once illuminated by a light source, typically a tungsten-halogen lamp. The output of the acquisition system is constituted by spectra where the reflectance is monitored as a function of the wavelength in a range that depends on the characteristics of the employed experimental setup^{1,2,3}. Introduced during the last four decades^{4,5}, RS and FORS are typically used in combination with X-ray fluorescence and other spectroscopies to describe the materials and the techniques used by artists to realize their masterpieces^{6,7,8,9}. The study of the reflectance spectra is usually performed by comparing the data from the sample with a group of reference spectra selected by the user in personal or public databases. Once the reference spectra that comply with the realization period of the sample and with the *modus operandi* of the artist have been identified, the user recognizes the main features of the reflectance spectra (i.e., transition, absorption, and reflection bands^{1,2,10,11}) and then, with the help of other techniques^{6,7,8} they distinguish the pigments that have been used in the paintings. Finally they discuss the slight differences that there exist between the references and the experimental spectra^{7,9}.

In most cases, the experimental datasets are composed of a few spectra, collected from areas chosen by art experts and assumed to be significant for the characterization of the painting^{6, 12,13}. Despite the skills and the experience of the user, a few spectra cannot fully exhaust the characteristics of the whole painted surface. Moreover, the result of the

analysis will always be strongly dependent on the expertise of the performer. In this scenario, Hyperspectral Reflectance Imaging (HRI^{3,14,15}) could be a useful resource. Instead of a few isolated spectra, the experimental setups return the reflectance properties of extended portions or even of the whole artefact under investigation¹⁶. The two main advantages with respect to the acquisition of the isolated spectra are evident. On one hand, the availability of the spatial distribution of the reflectance properties allows the identification of areas that hide interesting features, even though they may not seem peculiar¹⁷. On the other hand, the hypercubes guarantee a number of spectra high enough to enable the statistical analysis of the data. These facts support the comprehension of the distribution of pigments within the painted surface^{18,19}.

With HRI, the comparison of the experimental data with the references could be hard to handle¹⁵. A typical detector returns hypercubes of at least 256 x 256 spectra. This would require the user to evaluate more than 65,000 reflectance spectra against each reference, a task almost impossible to be carried out manually in a reasonable time. Therefore, the request for robust and, possibly, automated protocols to manage and analyze HRI datasets is more than justified^{15,17}. The proposed method answers this need by handling the whole analytic procedure with the minimum involvement and the maximum flexibility.

An algorithm comprising a set of home-made codes (**Table of Materials**) reads, manages, and organizes the files returned by the experimental setup. It allows the fine selection of the

portions of the Fields of View (FOVs, one field of view is the area of the painting monitored by a single hypercube) to be studied and performs the analysis of the data based on the Spectral Angle Mapper (SAM) method^{20,21} and on the manipulation of the original spectra. SAM returns false color gray-scale images called similarity maps. The values of the pixels of these maps correspond to the spectral angles that are the angles between the spectra stored in the hypercubes and the so-called End Members (EMs, a group of reference spectra that should describe the features of the surface monitored by the hypercubes)²². In the case of RS applied to paintings, the EMs are the reflectance spectra of pigments that should match the palette of the Master. They are chosen based on the available information about the artist, the realization period of the painting, and the expertise of the user. Therefore, the output of the SAM is a set of maps that describes the spatial distributions of these pigments over the painting surface and that supports the user to infer the materials used by the artist and their organization in the artefact. The algorithm offers the possibility of employing all kind of references independently from their origin. The references can be specific spectra selected within the hypercubes, come from databases, be acquired by a different instrument on a different surface (such as samples of pigments or the palette of the artist, for instance), or be obtained employing any kind of reflectance spectroscopy, FORS included.

SAM has been preferred among the available classification methods because it has been demonstrated to be effective for characterizing pigments (refer to the book by Richard²³ to have an overview of the main available classification methods). Instead, the idea of developing a home-made protocol rather than adopting one of the many tools freely available on the net^{24,25} relies on a practical consideration.

Despite the effectiveness and scientific foundation of the existing GUIs and software, a single tool hardly satisfies all the needs of the user. There could be an Input/Output (I/O) issue because a tool does not manage the file containing the raw data. There could be an issue regarding the analysis of the data because another tool does not provide the desired approach. There could be a limitation in the handling of the data because the simultaneous analysis of multiple datasets is not supported. In any case, a perfect tool does not exist. Each method must be adjusted to the data or vice versa. Therefore, the development of a home-made protocol has been preferred.

The presented approach offers neither a complete set of analytical methods (see, for comparison, the tool proposed by Mobaraki and Amigo²⁴) nor an easy-to-manage user-interface (see, for comparison, the software employed by Zhu and co-workers²⁵), but, in exchange, it focuses on a still underrated aspect of hyperspectral data analysis: the opportunity to manipulate the detected spectra. The power of the approach is illustrated by applying it to the painting *Quarto Stato* by Giuseppe Pellizza da Volpedo (**Figure 1**), an iconic oil on canvas held in the Museo del Novecento in Milan, Italy. Note that, since the approach requires running home-made codes, the developer arbitrarily chose the names of the codes and both the input and output variables used in the description of the protocol. The names of the variables can be changed by the user but they must be provided as follow: the input and out variables must be written respectively within brackets and eventually separated by comma and within square brackets and eventually separated by a white space. On the contrary the names of the codes cannot be altered.

Protocol

1. Set the spatial resolution of the hypercubes

1. Perform a preliminary inspection of the painted surface (**Figure 1**) supported by art experts to identify the main features of the painting.

1. Recognize the pictorial techniques employed by the artist to create the painting.
2. Identify the different brush strokes of paint on the canvas.
3. Estimate, qualitatively, the characteristics of the brush strokes with particular attention to their size.

2. Mimic the pictorial technique used by the artist by creating *ad-hoc* test samples where the brush strokes show characteristics similar to those applied by the artist.

NOTE: Pellizza da Volpedo was a Divisionist painter. A restorer was asked to prepare some test samples that qualitatively reproduce the brush strokes of the canvas of interest (**Figure 2, column A**).

3. Set the distance between the surface under investigation and the acquisition equipment.

NOTE: The distance determines the spatial resolution of the hypercubes²⁶ and therefore the possibility to distinguish the brush strokes on the images and SAM maps of the painted surface.

1. Evaluate the distance between the surface of the sample and the acquisition equipment taking into account the characteristics of the hyperspectral camera²⁶ (**Table of Materials**) and the size of the brush strokes drawn in the test samples.

2. Put the acquisition stage and the hyperspectral camera at the distance evaluated in the previous step. Arrange the test samples on the stage and ensure uniform illumination of the surface of the samples.

3. Perform a white calibration using the white standard reference supplied with the hyperspectral camera. Acquire the hypercubes.

NOTE: For each FOV, the hyperspectral camera returns both raw and calibrated images. The latter have been used for the analysis.

4. Download the files returned by the instrument and save them in a dedicated folder.

4. Check whether the spatial resolution of the hypercubes can distinguish the different brush strokes on the images and SAM maps of the painted surface.

1. Inspect the RGB pictures returned by the hyperspectral camera to ensure that the brush strokes used to realize the test samples can be recognized (**Figure 2, column A**). If so, move to the next steps; otherwise go back to step 1.3.1 and restart.

2. List the files containing the hyperspectral data and the RGB images of the FOVs by running the data reading code, **HS_FileList**. Type the following command line (semicolon included) in the terminal window of the language used to develop the codes (**Table of Materials**) and press **Enter** to run the code:

[HS_DataList HS_ImageList] = HS_FileList;

1. No input is required and there are two outputs: the list of the files containing the hypercubes, **HS_DataList**, and the list of the

images returned by the hyperspectral camera, **HS_ImageList**.

NOTE: The size of each hypercube is 512 x 512 x 204 voxels where 204 is the number of channels used to monitor the reflectance signal. The channels span the wavelength range between 400 and 1,000 nm with a spectral resolution of 7 nm at FWHM²⁶.

3. Define the 3D portion of the hypercubes that must be analyzed by running the cropping code, **HS_Crop_png**. Define the desired portion of each data-cube by selecting an area over an interactive window that shows the 2D, RGB image of the FOV monitored by each hypercube. Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:

```
[HS_ImageList] = HS_Crop_png(HS_ImageList);
```

1. There is one input (the list of the images returned by the hyperspectral camera, **HS_ImageList**) and one output (the input list added with the spatial coordinates to eventually crop the hypercubes).
4. Apply the D65 illuminant and 1931 observer from CIE (International Commission on Illumination) standards to the hypercubes to retrieve the RGB images of the FOV(s) from the reflectance spectra by running the re-building code, **HS_RGB_rebuild**. Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:

```
[HS_ImageList] =  
HS_RGB_rebuild(HS_ImageList, HS_DataList);
```

1. There are two inputs (the list containing the images returned by the hyperspectral camera, **HS_ImageList**, and the list of the files containing the hypercubes, **HS_DataList**) and one output (the input list containing the images returned by the hyperspectral camera added with the RGB images of the surfaces of the hypercubes retrieved from the reflectance spectra).

NOTE: **HS_RGB_Rebuild** exploits the functions developed by Jeff Mather²⁷ to apply the D65 illuminant and 1931 observer from CIE to the data.

5. Manually select some reference spectra on the surfaces of the test samples (**White Circles** and **Numbers** in **Figure 2, column A**) by running the isolated measuring points selection code, **PointSel**. Select the measurement points by clicking an interactive window that shows, one by one, the 2D, RGB images of the FOV(s). Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:

```
[References] = PointSel(HS_DataList,  
HS_ImageList);
```

1. There are two inputs (the list containing the images returned by the hyperspectral camera, **HS_ImageList**, and the list of the files containing the hypercubes, **HS_DataList**) and one output (a variable, **References**, containing the spectra selected as references within the FOV(s)).
6. If desired, store the position of the references over the surface of the samples into a set of pictures by

running the dedicated code, **SavelmPoint**. Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:

SavelmPoint(References, HS_ImageList);

1. There are two inputs (the variable containing the reference spectra, **References**, and the list containing the images returned by the hyperspectral camera, **HS_ImageList**) and no outputs (the code saves .png images in the current work folder).
7. Organize the references into a matrix by running the conversion code, **RefListToMatrix**. Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:

```
[References_Matrix] =
RefListToMatrix(References,
HS_ImageList(1).WaveL);
```

 1. There are two inputs (the variable containing the reference spectra, **References**, and the list of the wavelengths at which the detector counts the photons during the data acquisition of the spectra, **HS_ImageList(1).WaveL**) and one output (the same reference spectra organized into a matrix, **References_Matrix**).

NOTE: This step is mandatory because the code that evaluates the SAM maps requires the reference spectra to be organized into a matrix. The syntax of the second input, **HS_ImageList(1).WaveL**, is required to recall the variable **WaveL** from the list **HS_ImageList**. The number 1 within brackets refers to the first element of the list named as **HS_ImageList**; however, since all the hypercubes have the same wavelength range, it can be substituted by

each number minor or equal to the total number of listed images.

8. Extract the SAM maps using the whole spectra by running the standard SAM maps evaluation code, **SAM_Standard**. Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:
SAM_Standard(HS_ImageList, HS_DataList, References_Matrix);
 1. There are three inputs (the list containing the images returned by the hyperspectral camera, **HS_ImageList**; the list of the files containing the hypercubes, **HS_DataList**; and the matrix of the reference spectra, **References_Matrix**) and no output: the code saves the SAM maps as .png images in the current work folder.
9. Check whether the obtained similarity maps (**Figure 2, columns B - E**) display the details of the brush strokes used to realize the test samples. If this is the case, move to the next step of the protocol; otherwise go back to step 1.3.1 and restart.

2. Adjust the experimental parameters to the painting

1. Identify the Region(s) of Interest, ROI(s), of the painting to be studied (red rectangles in **Figure 3A**).
NOTE: It is common that more than one FOV is necessary to cover a single ROI.
2. Arrange the acquisition setup and the painting at the distance defined in the previous steps and perform the white calibration employing the white standard reference supplied with the hyperspectral camera.
NOTE: If the users must do an *in situ* acquisition (i.e., they must study a painting exposed in a museum or at an

exhibition), they can only manage the camera. This is the case of *Quarto Stato*, which is permanently exposed in a dedicated space at the Museo del Novecento in Milan, Italy.

3. Acquire the hyperspectral data from at least one FOV within the edge of each ROI(s) (unshaded areas within the red rectangles in **Figure 3A**).
4. Download the files returned by the instrument and save them into a dedicated folder.
5. Check whether the illumination of the surface of the painting has been uniformly set by looking at the RGB images returned by the hyperspectral camera. If this is the case, move to the next steps, otherwise go back to step 2.2 and restart.
- NOTE:** **Figure 4** illustrates the importance of this check (see the **Discussion** section for the details).
6. Repeat the sub-steps of step 1.4.
7. Check whether the data have a spatial resolution high enough to distinguish the brush strokes by observing the RGB pictures of the FOVs (**Figure 3B**) and the SAM maps (**Figure 3C**) related to the reference spectra selected within the FOVs (**green circles in Figure 3B**).
8. If the illumination and the spatial resolution have been properly set, complete the collection of the data acquiring the other FOVs necessary to cover the ROI(s); otherwise go back to step 2.2 and restart.

NOTE: When a ROI requires more than one FOV to be covered, ensure a certain degree of superposition between adjacent FOVs to easily stitch the resulting maps^{3,15}. The extent of the overlapping depends on the distance between the hyperspectral camera and the sample, on the translation, and the horizontal angle of view of the detector²⁸. In the case of the experimental

campaign conducted on *Quarto Stato*, the overlap has been set to be at least the 40% of the FOVs.

3. Hypercubes and the reference spectra management

1. Perform the I/O of the raw data: organize, read, and manage the hypercubes.
 1. Run the **HS_FileLister** code to store the list of the files containing the hypercubes and the related information into two variables at the disposal of the algorithm (see step 1.4.2 for the practical details).
NOTE: The hyperspectral camera returns hdr (high dynamic range) files that the code manages exploiting a revisited version of the script developed by Jarek Tuszynski²⁹.
 2. Run the **HS_Crop_png** code to select the portion of each FOV to be used in the analysis of the data (see step 1.4.3 for the practical details).
 3. Run the **HS_RGB_Rebuild** code to retrieve the RGB images of the FOVs from the reflectance spectra (see step 1.4.4 for the practical details).
2. Organize, read (if required), and manage the reference spectra.

NOTE: The reference spectra will play the role of the end members within the SAM method^{20,21}. This part of the algorithm is not univocally determined but depends on the selection mode and on the origin of the reference spectra.

 1. Run the **PointSel** code and click within the displayed interactive window to identify the reference spectra as isolated measuring points over the surface of the monitored area(s) (**Figure 5A**) (see step 1.4.5 for the practical details).

2. Automatically select the reference spectra as a regular reticulum of measuring points superimposed to the surface of the monitored area(s) by running the reticular selection code, **ReticularSel** (**Figure 5B**). Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:

```
[References] = ReticularSel(HS_DataList,
HS_ImageList, n_pixel);
```

1. There are three inputs (the list containing the images returned by the hyperspectral camera, **HS_ImageList**; the list of the files containing the hypercubes, **HS_DataList**; and the spacing of the reticulum, **n_pixel**, expressed in number of pixels) and one output: a variable containing the spectra selected as references within the FOVs, **References**.
3. Run the external references importer code, **Spectra_Importer**, to create a variable containing references from datasets and databases independent from the hypercubes acquired on *Quarto Stato*. Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:

```
[ExtReferences] =
Spectra_Importer(file_extension);
```

1. There is one input (the extension of the file containing the independent reference spectra, **file_extension**, written between apices) and one output (a variable containing the external references, **ExtReferences**).

NOTE: The external reference importer code has been optimized for importing tmr files but, if

necessary, it can be easily modified to deal with any kind of text file.

4. Run the **RefListToMatrix** code to put the references into a matrix, **References_Matrix** or **ExtReferences_Matrix**, as required by the code that evaluates the SAM maps (see step 1.4.7 for the practical details).
5. Wait for the **RefListToMatrix** code to equalize both the wavelength range and the spectral resolution (i.e., the number of components) of the hypercubes and the references.

NOTE: The code identifies the wavelength ranges of both the hypercubes and the references. The code compares the wavelength ranges and cuts off the wavelength interval(s) that are not monitored by both the hypercubes and the references. The code identifies the group of hyper-vectors (the hypercubes or the references) constituted by the lower number of components (i.e., characterized by the lower spectral resolution) in the equalized wavelength range. The code reduces the number of components of the longer hyper-vectors (the references or the hypercubes) to that of the shorter ones (the hypercubes or the references). This is done by keeping, for each wavelength of the shorter hyper-vectors, only the values of the longer hyper-vectors that correspond to the nearest wavelength to that of the shorter hyper-vectors.

1. The code automatically performs the equalization. If the references have been selected within the hypercubes, the wavelength range and the spectral resolution do not need to be equalized and they remain unchanged.

6. If desired, store the position of the references over the surface of the samples into a set of pictures by running the dedicated code (see step 1.4.7 for the practical details).

NOTE: This option is available only if the references have been selected within the hypercubes (steps 3.2.1 and 3.2.2).

4. SAM analysis

1. Run the **SAM_Complete** code to evaluate the similarity maps. Type the following command line (semicolon included) in the terminal window and press **Enter** to run the code:

```
SAM_Complete(HS_ImageList, HS_DataList,
References_Matrix);
```

1. There are three inputs (the list containing the images returned by the hyperspectral camera, **HS_ImageList**; the list of the files containing the hypercubes, **HS_DataList**; and the reference matrix, **References_Matrix** or **ExtReferences_Matrix**) and no outputs (the code saves the SAM maps as .png files in the current work folder).

NOTE: Other than the three described input variables, the **SAM_complete** code must be fed with few additional parameters to tailor the analysis protocol according to the preferences of the user (see the next steps).

2. When required, feed the code with the pre-processing option by typing the number 0 or 1 in the dialog box depending on the desired pre-processing operation and press **Enter** to continue.

1. Pre-processing option set to 0: the area subtended by each reflectance spectrum is normalized to 1.
2. Pre-processing option set to 1: the area subtended by each reflectance spectrum is normalized to 1 and then the normalized spectrum is derived one time.

NOTE: Both the hypercubes and the references undergo the same pre-processing option.

3. Select the end members to be used for the SAM analysis among the reference matrix by feeding the code with the numbers of the columns that correspond to the desired spectra. When required, enter into the dialog box the sequence of numbers corresponding to the desired columns by typing the numbers separated by a white space. Press **Enter** to continue.

NOTE: The sequence [1 2 3] corresponds to the selection of the first three columns of the reference matrix; an empty vector corresponds to the selection of all the columns of the reference matrix.

4. Feed the code with a string containing the first part of the name that will identify the sets of maps to be saved (i.e., the common part of the name of the .png files returned by **SAM_Complete**). When required, insert the string in the dialog box. Press **Enter** to continue.

NOTE: If the user types test, then the name of all the output .png images will start with test.

5. When required, feed the code with the method selected to handle the data by typing the number 0, 1, or 2 in the dialog box depending on the desired handle method and press **Enter** to continue.

1. Set the method to 0 for no manipulation of the data.
2. Set the method to 1 to require manual selection of the wavelength range(s) of the spectra to be considered before starting the analysis (**Figure 6**).

3. Set the method to 2 to require the algorithm to order the data on the basis of a specific criterion before the evaluation of the SAM maps (**Figure 7**).
 6. Wait for the protocol to process the data and to save the SAM maps in the current work folder as .png files.
- NOTE:** If the handle method has been set to 0 or 2, the user must just wait. If it has been set to 1, the user must select the portion(s) of the spectra to be employed for evaluating the SAM maps by clicking on an interactive window (**Figure 6**).

Representative Results

The proposed protocol offers a set of interesting features for the management and the analysis of HRI data. The I/O (step 3.1) of the raw data is always the first problem that must be solved before applying any analysis method and it can become a critical issue when dealing with large amounts of data. In the present case, the only task regarding the raw data is to store the experimental results into a dedicated folder and select it by browsing the hard disk when running the reading code (step 3.1.1). Thereafter, the cropping and the RGB rebuilding codes allow the refinement of the selection of the data to be analyzed (step 3.1.2) and checks that the experimental conditions have been properly set at the moment of the acquisition of the hypercubes (step 3.1.3, see **Figure 4** and the **Discussion** section for further details).

Once verified that the data-cubes have been correctly acquired, the algorithm offers different possibilities to select the end members for the SAM analysis^{20,21} (step 3.2). The first two options (steps 3.2.1 and 3.2.2) retrieve the references among the hypercubes by manually selecting some isolated measuring points (**Figure 5A**) or by automatically sampling the surface of the painting providing a reticular selection of

measuring points within one or more FOVs (**Figure 5B**). The analysis based on isolated measuring points is faster than the reticular based one, but it implies a careful and, possibly, informed observation of the FOV(s) to identify the significant spectra; this means good experience dealing with pigments and painted surfaces. The reticular selection makes the algorithm time-consuming and forces the user to observe a lot of output images to retrieve a handful of useful similarity maps. However, the reticular selection provides a complete screening of the hypercubes and, mostly, it can be carried out without experience of the experimental context. In principle, once the sampling distance, **n_pixel**, is decided, the user can neglect the observation of the FOV(s) with a very low probability of losing details.

In addition to the selection of the reference spectra within the hypercubes, the algorithm offers the opportunity to compare the data from the sample under investigation with references belonging to other sources (step 3.2.4). The external reference spectra importer code manages the I/O of references that do not belong to the surface of the painting. The matrix converter code equalizes the wavelength ranges and the spectral resolution of both the hypercubes and the external references (step 3.2.4). This possibility extends the capabilities of the user regarding the characterization of the sample. Indeed, the user can exploit every kind of available resource in terms of reflectance data. The hypercubes can be compared with public databases, with the spectral archives of the user, with new data collected on *ad hoc* prepared samples or even on other objects (paintings, palettes, hues, or whatever) belonging to the author or to other artists. Moreover, the external references can be obtained exploiting any kind of reflectance techniques so much so that the references shown in **Figure 6** and **Figure 7** have been

acquired by a portable FORS miniature spectrometer (**Table of Materials**) and not by the camera used for the HRI data.

Beyond the data management, the algorithm offers an original approach to the data analysis too. It allows manipulation of the spectra before evaluating the SAM maps (steps 4.1–4.5). This possibility finds its rationale in the choice of the SAM method to investigate the distributions of the pigments. In fact, SAM considers the reflectance spectra as they would be vectors in a multi-dimensional space (i.e., hyper-vectors with a number of components equal to that of the acquisition channels). Therefore, if the principal aim of the analysis is to compare different but similar references to distinguish which one best matches the pigments used by the artist, then the almost identical components of the reference spectra (i.e., the wavelengths that correspond to almost identical values in the hyper-vectors) should not be particularly useful and the algorithm allows to exclude these components from the analysis.

The protocol supports two options for manipulating the data (step 4.5): the user can define the wavelength portion(s) of the reflectance data to be analyzed manually (**Figure 6**) or automatically (**Figure 7**). The manual selection is straightforward. The pre-processed reference spectra or their first derivatives, depending on the selected pre-processing option (step 4.2), appear on an interactive window, **Figure 6A**, and the user selects one or more wavelength interval(s), **Figure 6B**, by clicking on the graph surface. The automatic selection is based on the mathematical criterion of the maximum variance applied to the pre-processed reference spectra or their first derivatives, depending on the selected pre-processing option (step 4.2). The algorithm computes the variance (normalized and displayed as a **Dashed Line** in **Figure 7A**) within the selected references and order

all the spectra (both the references and the hypercubes) accordingly to this criterion (the **Dashed Line** in **Figure 7B** represents the normalized and ordered variance). In other words, if the maximum variance corresponds to the n th wavelength, the content of the n th component of each pre-processed spectrum (references and hypercubes) will be moved to the first position of a re-arranged hyper-vector and so on (the colored portions of the background in **Figure 7A** and **Figure 7B** graphically explain the re-arrangement of the data). Practically, the components of the pre-processed spectra are ordered similar to principal component analysis³⁰.

Once the spectra have been manipulated, the algorithm evaluates the SAM maps. Following the manual manipulation (**Figure 6**), the protocol returns three sets of maps: two corresponding to the groups of selected and rejected wavelengths and one obtained employing the whole spectra. Otherwise, following the automatic manipulation (**Figure 7**), the algorithm applies a floating threshold to the variance values and evaluates the SAM maps at the increasing of the threshold for both the re-arranged hyper-vector components corresponding to the over threshold (i.e., automatically selected) and to the under threshold (i.e., automatically rejected) values of the variance. These sets of maps, together with that obtained from the whole spectra (always returned by the algorithm), result in a total of $(2N + 1)$ sets of maps where N is the number of values assumed by the threshold. The sets of similarity maps obtained at the increasing of the threshold (**Figure 8**) illustrate that data manipulation does not alter the content but rather provides new insights into the details of the mapped area(s) and, consequently, can help to distinguish similarities and differences between the samples and the references.



Figure 1: *Quarto Stato*. A picture of the painting, 1899-1901, 293 x 545 cm, oil on canvas, Giuseppe Pellizza da Volpedo, Museo del Novecento, Milan, Italy. [Please click here to view a larger version of this figure.](#)

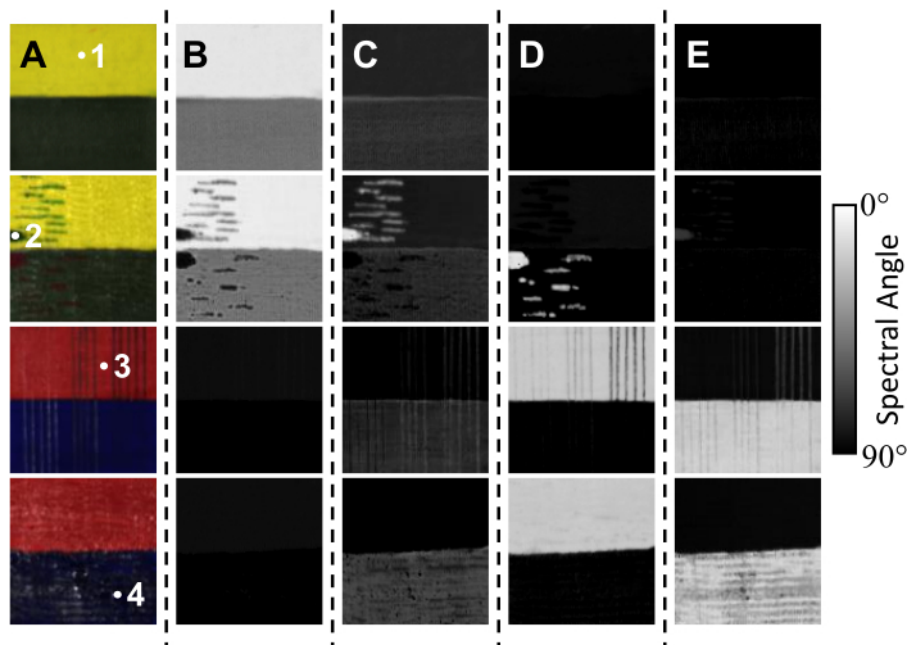


Figure 2: Definition of the experimental conditions. (A) The *ad hoc* prepared test samples; the **White Circles** and **Numbers** identify the measuring points corresponding to the spectra selected as references. (B) The SAM maps evaluated with respect to reference spectrum number 1, (C) number 2, (D) number 3, and (E) number 4. The **Gray Color Bar** indicates the range of values of the spectral angles. [Please click here to view a larger version of this figure.](#)

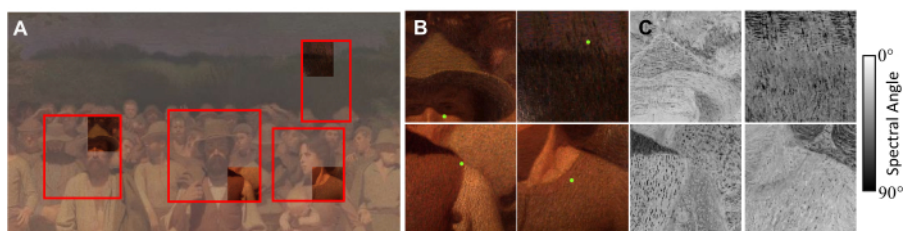


Figure 3: The application of the defined experimental conditions to Quarto Stato. (A) The ROIs selected for the experimental campaign (**Red Rectangles**); in each rectangle a FOV of those necessary to cover the ROI has been highlighted (**Unshaded Areas**). (B) The RGB pictures of the four **Unshaded Areas** of panel A. (C) The SAM maps evaluated with respect to a reference spectrum selected within each FOV (**Green Circles**). The **Gray Color Bar** indicates the range of values of the spectral angles. [Please click here to view a larger version of this figure.](#)

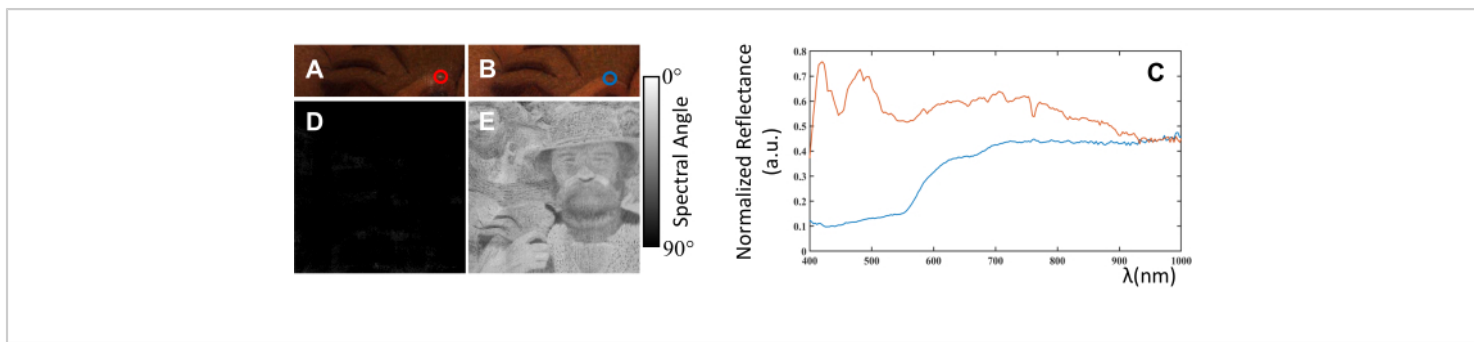


Figure 4: Proper versus improper illumination of the surface of the sample. (A) A portion of the FOV where a small fraction of the painted surface (**Red Circle**) is affected by altered reflectance properties due to improper illumination. (B) The same small fraction of the painting (**Blue Circle**) as it results when the FOV is properly illuminated. (C) The reflectance spectra of the measuring point at the center of the circles when the FOV is improperly and properly illuminated (**Red** and **Blue Line** respectively). (D) The SAM map of the FOV obtained using the spectrum of the improper illuminated measuring point as reference. (E) The SAM map of the FOV obtained using the spectrum of the proper illuminated measuring point as reference. The **Gray Color Bar** refers to (D) and (E) and indicates the range of values of the spectral angles obtained comparing the first derivatives of the spectra of the hypercube of the selected FOV and the first derivative of the spectrum of the measuring point at the center of the colored circles in (A) and (B). [Please click here to view a larger version of this figure.](#)

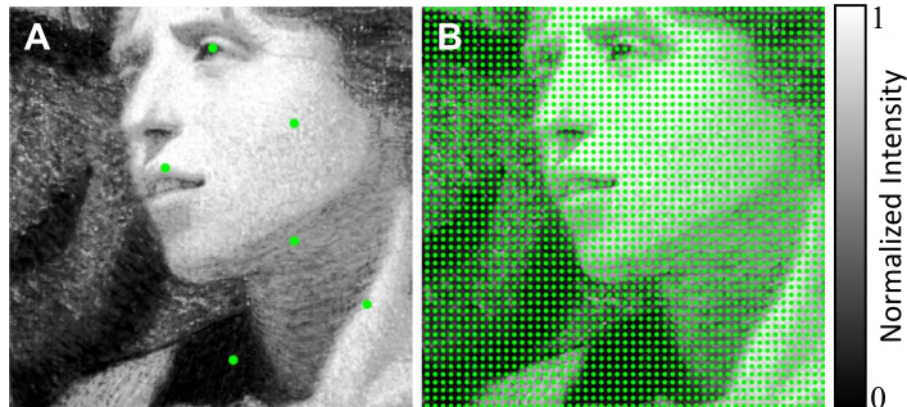


Figure 5: References selection within the hypercubes. (A) The isolated measuring points selection mode; the **Green Circles** indicate the location of the reference spectra manually selected on the FOV shown. (B) The reticular selection mode; the **Green Circles** indicate the location of the reference spectra selected by applying a reticulum with the sampling interval (n_{pixel}) set to five pixels to the FOV shown. The image reported in both (A) and (B) is the grayscale conversion of the RGB image of the FOV retrieved from the reflectance spectra applying the D65 illuminant and 1931 observer from CIE standards to the hypercube; the **Gray Color Bar** refers to the normalized intensity of this image. [Please click here to view a larger version of this figure.](#)

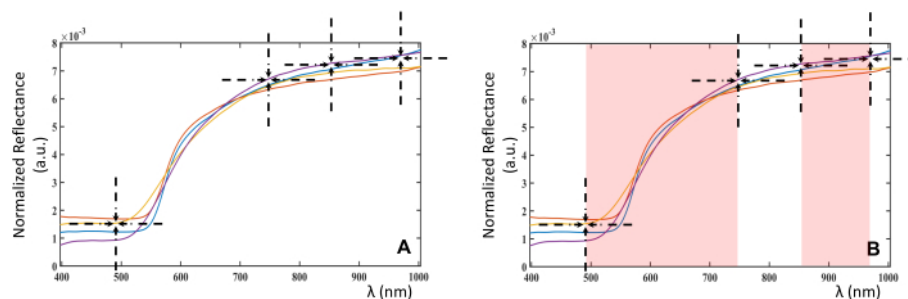


Figure 6: The manual data manipulation mode. (A) The aspect of the interactive window that allows the user to divide the reference spectra into the selected and rejected fractions of wavelengths. (B) The same references of (A) where the portions of data selected for evaluating the SAM maps have been highlighted by a pink background. (A) and (B) display the pre-processed spectra of the references. [Please click here to view a larger version of this figure.](#)

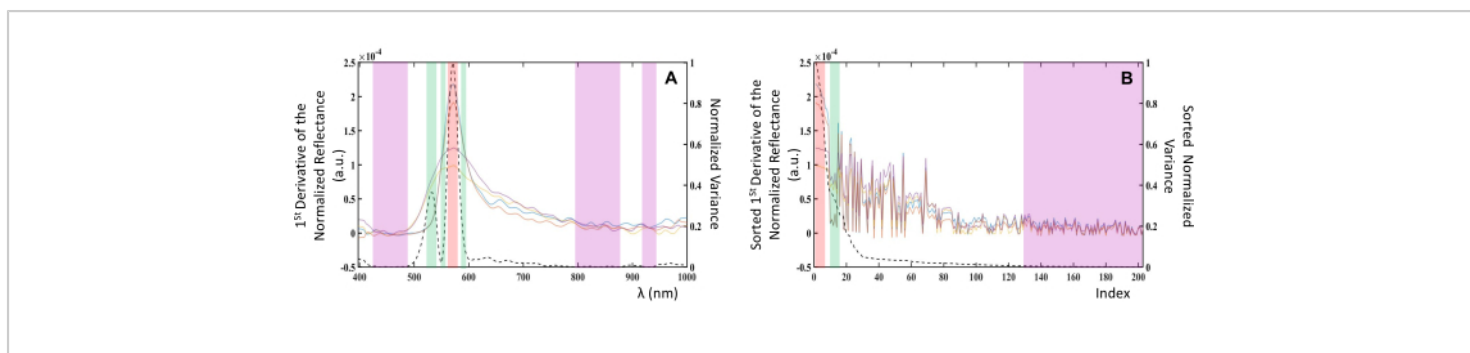


Figure 7: The automatic data manipulation mode. (A) The first derivatives of the four normalized references reported in **Figure 6 (Coloured Lines)** and their normalized maximum variance (**Black Dashed Line**). **(B)** The same derivatives of **(A)** sorted following the criterion of the maximum variance; the sorted values of the normalized maximum variance have been reported too (**Black Dashed Line**). Some portions of the background have been colored with different hues in the attempt to visually illustrate the re-arrangement of the hyper-vectors. [Please click here to view a larger version of this figure.](#)

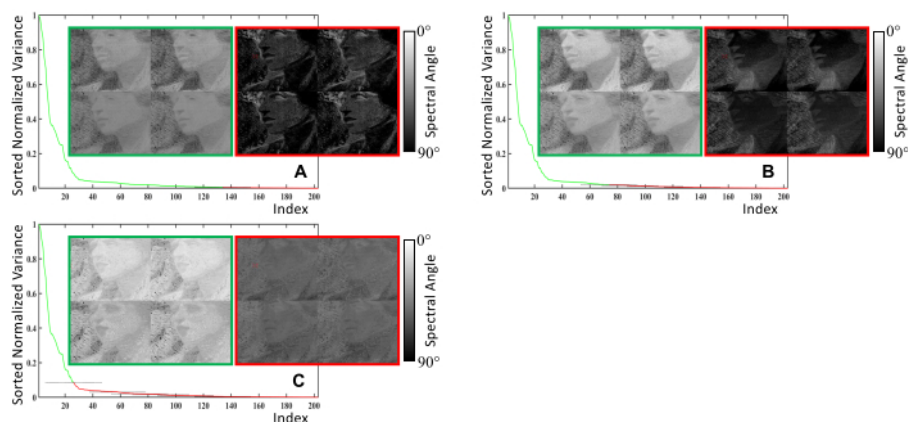


Figure 8: The SAM maps obtained by the automatic data manipulation mode. (A-C) The sorted values of the normalized maximum variance evaluated within the first derivatives of the reference spectra of **Figure 7**; the **Green and Red Sections** of the curve indicate, respectively, the selected (over threshold values) and rejected fraction of the data (under threshold values). The panels show what happens at the increasing of the threshold (**Black Dotted Segment**); each panel reports the SAM maps for both the groups of values obtained for the four derivatives of the spectra of **Figure 7**; the **Green Edged Maps** refer to the over threshold fractions while the **Red Edged Maps** refer to the under threshold ones. The **Gray Color Bars** indicate the range of values of the spectral angles. In this example, the step that determines the increase of the threshold is equal to 0.5% of the normalized maximum variance; the threshold values reported in **(C)** is 0.09 and it is the last considered threshold value because a further increase would reduce the number of selected components of the hyper-vectors below the arbitrarily fixed lower limit of 20 values, i.e., 10% of the total number of the acquisition channels of the hyperspectral camera.

[Please click here to view a larger version of this figure.](#)

Discussion

Hyperspectral reflectance imaging datasets are large reservoirs of information; therefore, the development of robust and, possibly, automated protocols to analyze the data is a key turn to exploit their potential^{15,17}. The proposed algorithm answers this need in the field of cultural heritage with particular attention to the characterization of the pigments of paintings. Based on SAM^{20,21}, the algorithm supports the user during the whole analysis process from the setting of the experimental conditions to the evaluation of the distribution of pigments. Though the algorithm still does

not have a complete graphical interface and that it does not provide a tool for viewing the results (for this purpose an open-source software has been used³¹ and it is recommended, see **Table of Materials**), the set of possibilities implemented to modulate the approach to the data analysis extensively balances these drawbacks.

The protocol sets the acquisition system according to the characteristics of both the sample and the detector. On one hand, the Divisionist technique employed by Pellizza Da Volpedo to create *Quarto Stato* requires that the hypercubes distinguish between small brush strokes of different pigments

placed side by side. On the other hand, the hyperspectral camera has a focus range between 150 mm and infinite with a manual adjustment system that at 1 m distance to the target detects an area of 0.55 x 0.55 m with a spatial resolution of 1.07 mm²⁶. The application of the algorithm to few hypercubes acquired on the test samples (**Figure 2**) helps establish a suitable working distance for the data acquisition. The observation of the measurements allows to set the working distance for the experimental campaign to 30 cm, corresponding to a resolution of 0.31 mm at the target. This working distance was also successfully adopted during the experimental sessions conducted on *Quarto Stato* (**Figure 3**). Once the working distance was defined, the illumination of the surface of the sample remains a critical issue^{3,15}. When a portion of a FOV shows uneven (**Red Circle in Figure 4A**) instead of uniform illumination (**Blue Circle in Figure 4B**), the reflectance properties change dramatically (**Figure 4C**) and the whole procedure is compromised (**Figure 4D** vs. **Figure 4E**). The protocol prevents uneven illumination (and more in general against artefacts in the monitored areas) during the acquisition of the data (by returning RGB, step 1.4.1, and SAM maps, step 1.4.9, that can be checked by the user) and *a posteriori* by excluding the compromised portions of the FOVs from the analysis by means of the cropping code (steps 1.4.2 and 3.1.2).

The protocol allows the user to select the references (i.e., the end members used for the evaluation of the SAM maps) with the maximum freedom. On one hand, the EMs can be chosen within the edges of the hypercubes by two manners: isolated measuring points selection (**Figure 5A** in steps 1.4.5 and 3.1.2) or reticular measuring points selection (**Figure 5B** in step 3.1.3). The first can be defined as informed selection because it requires some expertise in the user to manually identify the significant measuring points. The latter can be

defined as blind selection because the reticular sampling of the FOVs requires only the value of the sampling interval to be performed. On the other hand, the EMs can be retrieved from outside the painting under investigation (step 3.1.4). During the experimental campaign conducted on *Quarto Stato*, a portable miniature FORS spectrometer (**Table of Materials**) was used to collect spectra from draft samples belonging to the artist and currently kept in the Studio Museum located in Volpedo (Pellizza da Volpedo Studio Museo, Volpedo (AL), Italy). These reflectance data have been used for the evaluation of the SAM maps and some of them are reported in **Figure 6** and **Figure 7**. Since it limits the importance of the absolute intensity and of the baseline of the spectra, the pre-processing is mandatory for both the hypercubes and the EMs, especially if they have been obtained from slightly different setups or operative conditions³².

The last main feature of the protocol is the chance to manipulate the experimental data. For manipulation, it is intended that the identification of the most significant components of the EMs (i.e., of those portions of the spectra of the end members should help characterize the materials used by the artist). This task can be accomplished manually (**Figure 6**) or automatically (**Figure 7**). In the first case, the algorithm takes advantage of the expertise of the performer while, in the second case, it is a statistical criterion that determines the components that, time by time, will be used to evaluate the SAM maps. In both cases, the manipulation increases the number of the resulting similarity maps and consequently extends the capability to disclose the information carried by the hypercubes. In particular, the criterion-based selection generates a great number of insights of the painted surface (**Figure 8**).

Taken individually, the enumerated features could appear as mere technical benefits, but together they imply at least two main key points. The algorithm can be successfully applied by any kind of user and it can significantly broaden the scenario of the analysis. In fact, the main steps of the protocol (i.e., the selection of the references and the manipulation of the data) can be performed automatically, disregarding the skills and the experience of the user. With the possibility to drive the analysis with spectra from outside the hypercubes, all the reflectance data in the disposal of the researchers can be exploited for the characterization of the sample under investigation.

In summary, the protocol can be an extremely flexible tool. With some improvements regarding the graphical interface and the number of supported analysis methods, it can be a step beyond the state of the art regarding the handling and the analysis of data obtained from painted surfaces by means of hyperspectral reflectance imaging.

Disclosures

The authors have nothing to disclose.

Acknowledgments

This research was funded by Regione Lombardia in the framework of the Project MOBARTeCH: una piattaforma mobile tecnologica, interattiva e partecipata per lo studio, la conservazione e la valorizzazione di beni storico-artistici - Call Accordi per la Ricerca e l'Innovazione.

The authors are grateful to the staff at Museo del Novecento for the support during the *in situ* experimental sessions and to the Associazione Pellizza da Volpedo for the access to Studio Museo.

References

1. Picollo, M. et al. Fiber Optics Reflectance Spectroscopy: a non-destructive technique for the analysis of works of art. in *Optical Sensors and Microsystems*. Springer. 259-265 (2002).
2. Bacci, M. et al. Non-destructive spectroscopic investigations on paintings using optical fibers. *MRS Online Proceedings Library Archive*. **267** (1992).
3. Liang, H. Advances in multispectral and hyperspectral imaging for archaeology and art conservation. *Applied Physics A*. **106**, 309-323 (2012).
4. Bullock, L. Reflectance spectrophotometry for measurement of colour change. *National Gallery Technical Bulletin*. **2**, 49-55 (1978).
5. Saunders, D. Colour change measurement by digital image processing. *National Gallery Technical Bulletin*. 66-77 (1988).
6. Appolonia, L. et al. Combined use of FORS, XRF and Raman spectroscopy in the study of mural paintings in the Aosta Valley (Italy). *Analytical and Bioanalytical Chemistry*. **395**, 2005-2013 (2009).
7. Pouyet, E. K. et al. New insights into Pablo Picasso's La Miséreuse accroupie using X-ray fluorescence imaging and reflectance spectroscopies combined with micro-analyses of samples. *SN Applied Sciences*. **2**, 1-6 (2020).
8. Garofano, I., Perez-Rodriguez, J. L., Robador, M. D., Duran A. An innovative combination of non-invasive UV-Visible-FORS, XRD and XRF techniques to study Roman wall paintings from Seville, Spain. *Journal of Cultural Heritage*. **22**, 1028-1039 (2016).

9. Dupuis, G., Elias, M., Simonot, L. Pigment identification by fiber-optics diffuse reflectance spectroscopy. *Applied Spectroscopy*. **56**, 1329-1336 (2002).
10. Bacci, M., Picollo, M. Non-destructive spectroscopic detection of cobalt (II) in paintings and glass. *Studies in Conservation*. **41**, 136-144 (1996).
11. Cosentino, A. FORS spectral database of historical pigments in different binders. *E-Conservation Journal*. 54-65 (2014).
12. Leona, M., Winter, J. Fiber optics reflectance spectroscopy: a unique tool for the investigation of Japanese paintings. *Studies in Conservation*. **46**, 153-162 (2001).
13. Cheilakou, E., Troullinos, M., Kouli, M. Identification of pigments on Byzantine wall paintings from Crete (14th century AD) using non-invasive Fiber Optics Diffuse Reflectance Spectroscopy (FORS). *Journal of Archaeological Science*. **41**, 541-555 (2014).
14. Kubik, M. Hyperspectral imaging: a new technique for the non-invasive study of artworks. *Physical Techniques in the Study of Art, Archaeology and Cultural Heritage*. **2**, 199-259 (2007).
15. Fischer, C., Kakoulli, I. Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications. *Studies in Conservation*. **51**, 3-16(2006).
16. Daniel, F. et al. Hyperspectral imaging applied to the analysis of Goya paintings in the Museum of Zaragoza (Spain). *Microchemical Journal*. **126**, 113-120 (2016).
17. Baronti, S., Casini, A., Lotti, F., Porcinai, S. Principal component analysis of visible and near-infrared multispectral images of works of art. *Chemometrics and Intelligent Laboratory Systems*. **39**, 103-114 (1997).
18. Mansfield, J. R. et al. Near infrared spectroscopic reflectance imaging: supervised vs. unsupervised analysis using an art conservation application. *Vibrational Spectroscopy*. **19**, 33-45 (1999).
19. Clodius, W. B. Multispectral and Hyperspectral Image Processing, Part 1: Initial Processing. *Encyclopedia of Optical Engineering: Las-Pho, pages 1025-2048*. **2**, 1390 (2003).
20. Kruse, F. A. et al. The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. in *AIP Conference Proceedings*. (1993).
21. Yang, C., Everitt, J. H., Bradford, J. M. Yield estimation from hyperspectral imagery using spectral angle mapper (SAM). *Transactions of the ASABE*. **51**, 729-737 (2008).
22. Delaney, J. K. D. et al. Integrated X-ray fluorescence and diffuse visible-to-near-infrared reflectance scanner for standoff elemental and molecular spectroscopic imaging of paints and works on paper. *Heritage Science*. **6**, 1-12 (2018).
23. Richards, J. A. Remote sensing digital image analysis. *Springer*. **3** (1999).
24. Mobaraki, N., Amigo, J. M. HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems*. **172**, 174-187 (2018).
25. Zhu, C. Y. et al. Optimization of a hyperspectral imaging system for rapid detection of microplastics down to 100 μm . *MethodsX*. **8**, 101175 (2021).

26. Behmann, J. et al. Specim IQ: evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection. *Sensors*. **18**, 441 (2018).
27. Mather, J. Spectral and XYZ Color Functions (<https://www.mathworks.com/matlabcentral/fileexchange/7021-spectral-and-xyz-color-functions>). *MATLAB Central File Exchange*. (2021).
28. Chen, C.-Y., Klette, R. Image stitching-Comparisons and new techniques. in *International Conference on Computer Analysis of Images and Patterns*. (1999).
29. Tuszynski, J. read_envihdr (https://www.mathworks.com/matlabcentral/fileexchange/38500-read_envihdr). *MATLAB Central File Exchange*. (2021).
30. Jolliffe, I. T., Cadima, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. **374**, 20150202 (2016).
31. Schneider, C. A., Rasband, W. S., Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. **9**, 671-675 (2012).
32. Rinnan, Å., Van Den Berg, F., Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*. **28**, 1201-1222 (2009).