

RESEARCH ARTICLE

Open Access



Quantitative analysis of mixed pigments for Chinese paintings using the improved method of ratio spectra derivative spectrophotometry based on mode

Shuqiang Lyu^{1,2}, Yiyi Liu^{1,2}, Miaole Hou^{1,2*}, Qinli Yin³, Wangting Wu⁴ and Xueyun Yang^{1,2}

Abstract

Background: Mixed pigment analysis is an important and complex subject in preserving and restoring Chinese paintings since the colors observed by our naked eyes or instruments such as hyperspectral cameras are usually a mixture of several kinds of pigments. The purpose of this study was to explore a more effective method to confirm the type of every pure pigment and their proportion in pigment mixtures on the surface of paintings.

Methods: Two endmember extraction algorithms were adopted to identify the types of pigments and an improved method of ratio spectra derivative spectrophotometry was used to determine their proportion.

Main works: (1) Extracting the pure pigment components from mixed spectrum by adopting two blind source separation algorithms: Independent Component Analysis and Non-negative Matrix Factorization; (2) matching the separated pure spectrum with the pigment spectral library built in our laboratory to determine the pigment type; and (3) calculating the proportions of mixed pigments using the Ratio Spectra Derivative Spectrophotometry based on Mode, which is improved based on the original algorithm. Finally, a comparison was made with two abundance inversion algorithms: Least Squares Algorithm and Minimum Volume Simplex Analysis. The correlation coefficient and root mean square error were used to provide evidence for accuracy evaluation.

Conclusions: (1) Non-negative matrix factorization is more suitable for endmember extraction; and (2) Ratio spectra derivative spectrophotometry based on mode is more suitable for abundance inversion.

Keywords: Hyperspectral imaging, Pigments unmixing, Endmember extraction, Abundance inversion, Chinese painting restoration

Introduction

Throughout history, human has created a large number of precious cultural heritages, such as paintings, ceramics, murals, utensils, fabrics, etc., which used a large number of mineral pigments, often accompanied by a mixture of various pigments during use, blending into rich colors.

The Chinese painting is one of the important classes of cultural heritages. However, its surviving time is limited, not mentioning the damages caused by mold, insects, smudges, etc. due to environmental, human or other factors. In order to better retain and repair them, it is often necessary to restore and duplicate colors, where judging the type and proportion of pigment is often the first step of cultural heritage restoration.

However, the colors observed by our naked eyes or instruments such as hyperspectral cameras are usually a mixture of several kinds of pigments due to pigment usage and painting skills. Liu et al. [1] illustrated that

*Correspondence: hougiao@bucea.edu.cn

¹ School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, No. 15 Yongyuan Road, Daxing District, Beijing 102616, China

Full list of author information is available at the end of the article



such pigments mixture were well-distributed and tight by performing three-dimension microscope analysis. Therefore, it is difficult to determine the type and proportion of the pigments on the surface of Chinese paintings before restoration.

In recent times, various techniques, such as X-ray diffraction analysis, fluorescent X-ray analysis, Raman spectroscopy, electron microscopic analysis, near-infrared spectroscopy, etc. [2–4], have been applied in pigment analysis. For example, Nam et al. [5] paid extreme care to use Raman spectroscopy to analyze pigment samples from Korean traditional paintings to ensure that the incident laser power was low enough to reduce the effects on cultural relics. However, these methods share some common problems. Firstly, most of these techniques require collection of physical samples from paintings to be analyzed by instruments, which can cause secondary damage to paintings. Secondly, sampling positions are mainly based on human judgment through visual inspection and lack scientific support. Finally, these methods often are time consuming and can only test some small local regions of paintings.

Recently, hyperspectral imaging technology, with the advantages of no-touching and no-sampling, has been widely applied in the field of preservation and restoration of paintings. Hyperspectral imaging overcomes the limitations of panchromatic and multi-spectral imaging, offering more detailed remotely sensed information with higher spectral resolution. Except for the spatial information included in a common image, it can also provide a continuous spectral response curve for each pixel in the image to subdivide and identify features. This makes the technology obviously advantageous in terms of object recognition, information extraction and sub-pixel target detection [6].

Similar to hyperspectral images in remote sensing, it is more common to find pixel mixing in the images of Chinese paintings due to the intentional use of mixed pigments during artists' painting process. It is therefore imperative for studying mixed pixel separation and type identification in different application areas. On the one hand, endmember extraction refers to a process of extracting pure pixels of target from hyperspectral image, or obtaining pure spectra from ground-measured data or spectral library. If we regard the mixed spectra as a result mixed by multi-source signals, the problem of extracting endmembers from mixed pixels can be understood as the problem of statistical signals separation, which can be solved by using the methods in the field of signal processing. The typical signal separation algorithms based on statistics are independent component analysis (ICA) and non-negative matrix factorization (NMF). Before extracting endmembers,

it is sometimes necessary to perform component number estimation. The currently used methods include orthogonal projection, spectral clustering, geometrical and statistical approaches [7–14], which are selected to estimate the “true” or “pure” spectral components contained on the image, and some prior knowledge is usually required in some steps to eliminate possibly unimportant endmembers [8, 13, 14]. On the other hand, each pigment has its specific reflective spectrum. Hence, the hyperspectral imaging technique characterizes the nature of pigment according to the spectral information. This provides the theoretical basis for pigment identification and can be used to determine pigment type by matching its spectrum with the standard spectra in a spectral library.

Abundance inversion is a process to calculate the proportion of each component in mixture. Ratio spectra derivative spectrophotometry (RSDS) is extended from derivative analysis and spectral mixing analysis, which can reduce the effect of background factors and efficiently enhance the differences between spectra to improve the accuracy of abundance inversion of target objects [15].

This paper follows a lab experimental approach to investigate solutions of determining the type and proportion of the mixed pigments on the surface of Chinese paintings, aiming at providing a new non-destructive, more effective method for mixed pigment analysis to obtain more scientific evidence to support restoration of Chinese paintings. The paper starts from an overall introduction of the context of the study, followed by a brief review of related works in the next section. “Materials and methods” section describes how the samples were made and the data collection process, and introduces the methods chosen for the study in detail. Further, the experiment and the results obtained are described, and the results are statistically analyzed in “Results and analysis” section. In the last section, the consideration about the applicability of the methods used in the study is discussed, the conclusion is drawn and the future work is outlined.

Related works

Mixed pigment analyses normally involve: (1) selection of a suitable spectral mixing model, (2) selection of an appropriate method for endmember extraction to obtain pure pigments according to the selected spectral mixing model, and (3) performing abundance inversion for each pure pigment according to the type of the obtained endmember to calculate its proportion in the mixed pigment. This section reviews the studies related to these

three aspects and intends to identify the problems and gaps existed or lessons learned.

Spectral mixing models

According to the degree of mixing, mixed pigments can be roughly divided into three types: compact, polymer and integrated. The latter two can be regarded as linear mixing because they ignore boundary effects. The compact type is often regarded as nonlinear mixing due to the complex interaction between objects. Both linear mixing model (LMM) and non-linear mixing model (NLMM) are commonly used for mixed pigment analysis. The difference between these two spectral mixing models lies in whether multiple refractions or reflections occur between incident photons and minerals. LMM has simple modeling and clear physical meaning, and is simpler for computation [16] in comparison with NLMM. Thus, many studies have been done based on the LMM. NLMM, on the other hand, takes into account the influence of spectral transmission, the interaction between spectra and some other factors, and can provide a good explanation of actual spectral mixing process. However, it requires more prior knowledge and is more complex and inconvenient to process. Based on the Radiosity Theory [17], proposed that the explicit modelling of the NLMM could be achieved by adding extra endmembers to the LMM. Some scholars transformed NLMM into LMM by using Hapke Radiative Transfer Models [18]. According to a study on laboratory spectral data, [19] showed that LMM is often used when actually processing mixed mineral spectra. Therefore, LMM can be considered when analyzing spectral data of mineral pigments.

Spectral unmixing method

In remote sensing, spectral unmixing method usually consists of two steps: endmember extraction and abundance inversion. Endmember extraction is to extract the pure pixel spectrum representing a certain feature from the image, while abundance inversion is to calculate the proportion of each endmember in mixed pixels.

Endmember extraction

Endmember extraction can be accomplished using the methods based on projection convex geometry analysis, sparse matrix regression analysis, or statistical learning analysis and so on [20, 21]. There are cases, paintings for example, where the images may contain less or even no pure pixel but mixed pixels. In such situation, the method based on statistical analysis for endmember extraction has drawn much attention. This kind of method considers mixed spectra as multi-source mixed signals, so the problem of extracting endmembers from

mixed pixels can be transformed to seek solutions to separate statistical signals, which can be solved by using the methods found in the field of signal processing. Furthermore, it has strong practical significance and application value to realize blind separation for hyperspectral mixed data without any prior knowledge, and consequently extract endmember spectrum. Blind source separation, which assumes several source signals received by one or more sensors after being mixed, can obtain source signals only by separating from these observed signals. ICA and NMF are two commonly used blind source separation algorithms.

The ICA algorithm, proposed in 1987, was used to address the problem in signal separation [22]. The algorithm works on observed mixed signals to obtain unknown source signals. Many researches showed that, when the source signals are independent from each other and meet the non-Gaussian requirements, ICA could separate the mixed signal into statistically independent signals. [23] proposed a fast ICA algorithm based on a fixed-point iterative algorithm, called FastICA algorithm, which has faster convergence speed and no need to manually set the parameters of step length adjustment. It overcomes the shortcomings of the traditional ICA and is suitable for hyperspectral data analysis. The NMF algorithm proposed by Lee DD [24] differs from other algorithms in that it requires that all elements in the matrix be non-negative. Without considering errors, the matrix is expressed as the result of the complete multiplication of two non-negative matrices. Since hyperspectral data is a kind of records of the object's reflectance data and meets the non-negative requirement, NMF can be used to extract endmembers from hyperspectral data.

Both ICA and NMF have been applied in endmember extraction of hyperspectral images for many years. However, few of these previous works made comparison between NMF and ICA in endmember extraction of pigments' mixtures.

Abundance inversion

Endmember's abundance value can be calculated after the process of spectral identification, that is, the determination of the proportion of mixed pigments. Depending on whether prior knowledge of endmembers is needed before unmixing, abundance inversion methods can be divided into supervised decomposition algorithm and unsupervised decomposition algorithms [25]. Since the mixed spectrum of mineral pigments does not completely meet the requirement of the LMM, the supervised decomposition algorithms, such as Least Square Algorithm (LSA), Minimum Volume Simplex Analysis (MVSA) and Ratio Spectra Derivative Spectrophotometry (RSDS), can be selected for the study.

Since the early 1990s, RSDS has been put forward as a kind of simple and effective spectral analysis method for determining the components concentration of objects [26]. Many scholars [27–29] utilized the position of zero intersection and spectral peak to measure the component abundance of mixtures. However, few scholars have directly applied it to determine the abundance of mineral components in mixtures. The traditional RSDS may result in measurement errors when the wavelength shifts or the ratio derivative of the test component value at the zero intersection wavelength is small. Further, it may result in low accuracy when dealing with hyperspectral data because hyperspectral data have a large number of bands and no unique peaks or intersections. Gypsum and Ephedra powder were taken as the experimental samples to calculate Pearson correlation coefficient and root mean square error (RMSE) between the obtained abundance and the actual abundance, which were considered as the precondition, to find out the strong linear band that is more consistent with the actual abundance [30]. However, they only tested with a few types of pigments. With different types of mineral mixing, whether the strong linear bands had the same characteristics remains to be further studied. Moreover, hyperspectral bands are numerous, and the specific band cannot be accurately chosen for abundance inversion without the basis of correlation coefficient and other preconditions.

Materials and methods

Figure 1 illustrates the overall process of our study, which has three main steps, including endmember extraction, spectral matching and abundance inversion.

1. Endmember extraction refers to the process of separating mixed spectra into pure spectra. The mixed spectra collected by hyperspectral instruments were considered as a matrix to be decomposed, and two kinds of blind source separation algorithms including ICA and NMF, were used to extract pure pigment components from mixed spectra. The separation results were then compared with those by applying the two algorithms to laboratory mixed pigment sample data.
2. After endmember extraction, the mixed spectrum was decomposed into several pure spectra. However, the pigment type remains unknown yet. Spectral identification is to determine pure pigment type by matching the components extracted from mixed pigments with the pigment spectral library built in our laboratory.
3. Abundance inversion calculates the proportion of each pure pigment in mixtures after its type is known. An improved algorithm named ratio spectra

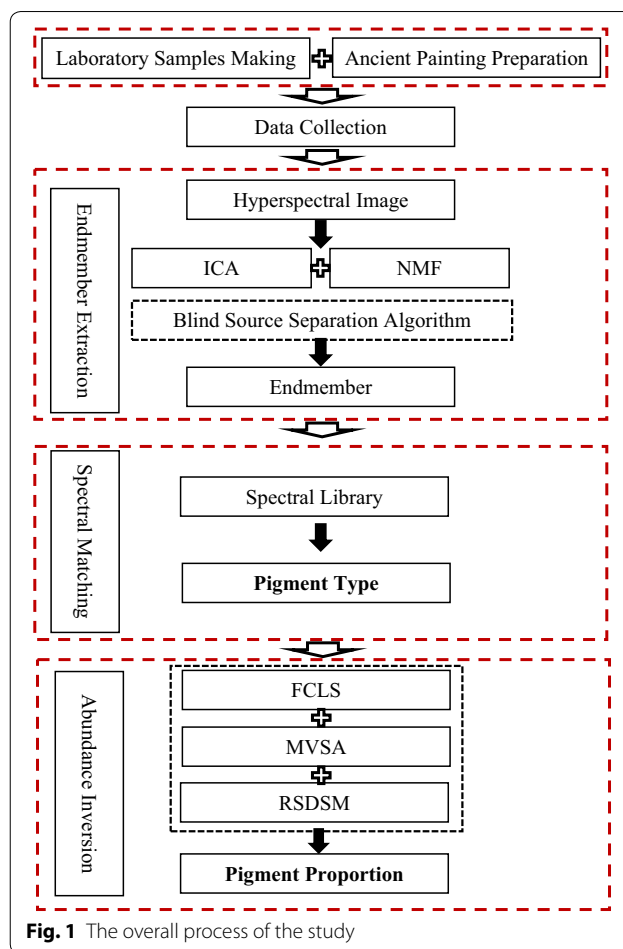


Fig. 1 The overall process of the study

derivative spectrophotometry based on the Mode (RSDSM) was developed to estimate the abundance fraction. In order to evaluate the accuracy of this algorithm, the Full-constrain least squares algorithm (FCLS) and the minimum volume simplex analysis (MVSA) were performed on the same data.

Samples preparation

Except for several unstable pigments such as lead white (which will turn black in the presence of sulfur) and red lead (which is easy to change color under certain humidity), a large number of studies have shown that most of the mineral pigments are relatively stable [31]. Their spectral characteristics also show a trend that is consistent with modern pigments. After the investigation of Chinese painting pigments, more than 30 kinds of pigments that were commonly used in ancient times and are still in use today were selected. Among them, the majority were mineral pigments, and a few were botanical pigments with high frequency of use.

Since the mineral pigments cannot be directly drawn on the paper, it was necessary to configure a certain concentration of glue before making experimental samples. The glue was then thoroughly mixed with the mineral pigments, followed by painting pigments evenly on 4×4 cm square rice papers and marking the names of pigments. The following is a brief introduction to some specific content of experimental samples. The picture of the samples is given in Additional file 1: Appendix S1.

Pigment selection

Choosing the suitable mineral pigment is crucial to make sample. Many modern Chinese pigments, although with the same traditional names, may add some modern chemical additives, which make them quite different from the pigments used in ancient times. After careful investigation and comparison, we selected pigment samples from a famous brand called Jiang Sixutang established 300 years ago. In order to make the basically same influence of the glue on the spectrum, most of the pigments we chosen were powdery, the glue preparation of which were consistent. And only two were solid pigments that had glue contained, which were used frequently in ancient time.

Glue selection

The material of the glue may be animal protein, natural plant or fat, etc. [32], and the type of glue produced is protein, bovine gum, bone glue, gelatin, peach glue, etc. In this experiment, gelatin was selected as the binder. When configuring the glue, we put a proper amount of gelatin in cold water for a while until the gelatin expanded and softened after absorbing water, and then placed it in hot water and stirred until it was dissolved.

Substrate selection

In order to be consistent with the drawing environment of Chinese paintings, using the rice paper as the substrate, the pigment mixture was drawn on the rice paper. Compared with the Pimade Xuan paper, raw Xuan paper has not been painted by alum, which conforms to the use of papers in Chinese paintings. Therefore, the raw Xuan type of paper from Rongbaozhai, an old stationery, calligraphy and painting shop in Beijing, China, was used in this study.

Data collection

The instrument used to acquire the spectral data in this study was ASD FieldSpec4 portable spectroradiometer, whose main specifications were listed in Table 1.

The data collection was carried out indoor to avoid the influence of other environmental factors. Two halogen lamps were selected as the artificial light source for

Table 1 Specifications of ASD FieldSpec4 spectroradiometer

| Name | Specifications |
|--------------------------------------|---|
| Spectral range (continuous coverage) | 350–2500 nm |
| Spectral bands | 2151 |
| Spectral width | 2.5 nm |
| Spectral resolution | 350–1000 nm @3 nm 1001–2500 nm @8 nm |
| Size | 12.7 × 35.6 × 29.2 cm |
| Weight | 5.44 kg |

spectral acquisition. In particular, when using the ASD spectroradiometer to collect spectral data, a well-sealed probe was performed 10 times on each sample to measure spectral reflectance. The data were then averaged as a standard spectral curve in order to prevent external light source and reduce operational error.

Model and methods

This sub-section first introduces the spectral mixing model of pigments used in our study for the analysis of mixed pigments, followed by the endmember extraction algorithm and spectral identification method. Then, the improved abundance inversion algorithm of RSDSM is discussed in detail.

Spectral mixing model

As mentioned above, most of the pigment mixtures showed a linear spectral mixing characteristic. Therefore, we selected LMM for spectral unmixing in this study.

In the LMM, the mixed spectrum r_{mix} can be described as a linear combination of its constituent components (i.e., endmembers) with pure spectral characteristics [33]. It can be expressed as follows:

$$r_{mix} = \sum_{i=1}^n r_i f_i + e = R \times f + e, \quad (1)$$

where, r_{mix} : the mixed spectrum for a given pixel, a column vector with $m \times 1$, m is the number of bands; n : the number of endmembers; r_i : the i th endmember of mixed spectrum r_{mix} ; f_i : the proportion (abundance fraction) of the endmember r_i ; R : the endmembers matrix ($m \times n$), where each column is a vector of endmember; f : the coefficient column vector, $f = (f_1, f_2, \dots, f_n)^T$; and e : the errors.

According to the actual physical meaning corresponding to the model, the abundance fraction of endmembers f_i is greater than zero, and their sum should be 1.

Endmember extraction algorithms

Two blind source separation algorithms for endmember extraction, including ICA and NMF, were adopted

to compare their performance in pigment composition extraction.

1. Independent component analysis algorithm.

The ICA algorithm assumes that the observed pigment spectrum is linearly mixed by several endmembers' spectra. The formula can be expressed as:

$$X = A \times S + e, \quad (2)$$

where, X : the observed signal; A : the matrix of endmember spectral abundance; S : the matrix of independent endmember spectrum; and e : the additive noise.

In practical applications, only the mixed signal X can be observed, which is used to estimate the value of A and S without any prior knowledge.

2. Non-negative matrix factorization algorithm.

The NMF algorithm can be expressed as:

$$X_{n \times m} = A_{n \times r} \times S_{r \times m} + E_{n \times m}, \quad (3)$$

where, X : the mixed spectral matrix to be decomposed, $X \in \mathbb{R}_+^{n \times m}$; $A \setminus S$: two endmember matrices after separation, $A \in \mathbb{R}_+^{n \times r}$, $S \in \mathbb{R}_+^{r \times m}$; and E : the error matrix, $\|E\|$ is as small as possible. (" $x \in \mathbb{R}_+$ " means " $x \geq 0$ ").

The range of r is much smaller than the values of n and m . In the sub-pixel unmixing, the value of r can be considered as the number of endmembers, which is usually obtained by prior knowledge or calculated by using endmember number estimation algorithm.

Spectral identification method

According to the property of the blind source separation algorithm, the types of the pure endmembers are still unknown after the endmember extraction. Therefore, it is necessary to match them with standard spectra in pigment spectral library to determine what kind of pigment type each endmember is. The pigment spectral library built by our team was based on collecting spectral data of mineral pigments in laboratory, which contains over thirty kinds of commonly used pigments such as Cinnabar, Ocher, Azurite, and Orpiment.

Specific pigment has its specific reflective spectrum, which is the theoretical basis for pigments identification. As a result, we can achieve the purpose of pigment recognition by comparing the characteristics of each spectral reflectance curve and calculating the correlation coefficient between the endmember curve extracted and the standard curves in pigment spectral library. The pigment curve with the highest similarity can be considered as the final matching type. However, some pigments have high similarities as they have the same ionic or functional

group compositions, such as, (1) Cinnabar and Vermilion; (2) Calcite and Lead white; (3) Calcite and Chalk; and (4) Clam meal and Chalk. Therefore, when these highly correlated spectra appear in the identification results, in order to reduce the risk in mistakenly identifying one pigment as another, it requires human intervention in the identification process and to take special care of.

Abundance inversion algorithms

After the processing of endmember extraction and spectral identification, only the abundance fractions of endmembers f_i remain unknown in Eq. (4). In this part, an improved algorithm is proposed to calculate the value of f_i .

RSDSM

After the Ratio Spectra Derivative Spectrophotometry was first proposed in 1990, [(15)] applied the method to mixture of lichen and rock for unmixing analysis based on hyperspectral data. Taking the binary mixture, which contains two kinds of endmembers, as an example, the method can be described as follows:

$$f_1 = \frac{d^2 \left(\frac{r_1(\lambda_i)}{r_2(\lambda_i)} \right)}{d\lambda^2} \bigg/ \frac{d^2 \left(\frac{r_{mix}(\lambda_i)}{r_2(\lambda_i)} \right)}{d\lambda^2}, \quad (4)$$

where, f_1 : the abundance of endmember 1; r_{mix} , r_1 , r_2 : the reflectance of mixture, endmember 1 and endmember 2 in corresponding wavelength position; λ : the wavelength; and i : the number of wavelength.

Equation (4) indicates that the abundance of endmember 1 is independent of the abundance of other endmembers in the mixture. Different from the original formula using the first-derivative spectra, it chooses the second-derivative spectra to estimate the abundance in Eq. (4), since some studies have demonstrated the advantages of high-order derivative spectra for characterizing the absorption features [34, 35]. At the same time, the effect on high-frequency noise and low-frequency background noise will be compounded as the order of the derivative increases.

However, without the basis of correlation coefficient and other preconditions, the specific bands cannot be accurately chosen for abundance inversion because of over hundred bands for hyperspectral data. Therefore, based on the RSDS, an improved method was proposed to select characteristic band by determining interval in which the maximum mode is located. The mode, which is a statistical term, is a numerical value with a tendency to focus on the statistical distribution, i.e., the most frequently occurring value in a group of data. Taking the binary mixture as an example, the improved method called RSDSM is given as follows:

The improved algorithm considers the abundance of each wavelength to be different during the solution process. According to Eq. (4), the abundance fraction of endmember 1 can be calculated as f_1 . It must be noticed that the f_1 is not exactly the same at different wavelength due to the light nonlinear mixing phenomenon at some local bands of reflectance curve. The improved algorithm considers the abundance of each wavelength to be different during the solution process, which can solve the local nonlinear problem of the spectrum. So, the f_1 can be written as:

$$f_1 = (f_{11}, f_{12}, \dots, f_{1m}), \tag{5}$$

where, f_j : the vector abundance fraction of endmember 1; m : the total number of bands.

As well known, the abundance fraction should be between [0, 1]. We introduce an interval of Δf to divide the value range [0, 1] to $[0, \Delta f, 2\Delta f, \dots, 1]$. Then, an initial abundance fraction sub-range is confirmed by the following equation:

$$S_{range} = [j \times \Delta f, (j + 1) \times \Delta f], \quad (j = 0, 1, 2, \dots, k, 1/\Delta f), \tag{6}$$

The parameter of j can be calculated under the condition that the $Count_j$ in S_{range} is the maximum value among the range of j , and $Count_j$ can be expressed as follows:

$$Count_j = \sum_{i=1}^m \varepsilon((f_{1i} - j \times \Delta f)((j + 1) \times \Delta f - f_{1i})), \tag{7}$$

where the f_{1i} is the abundance fraction of endmember 1 at i th band, and $\varepsilon(t)$ is the unit step function as follows:

$$\varepsilon(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{8}$$

After that, we can get the initial abundance fraction range S_{range} in which the exact abundance fraction of endmember 1 must be. In order to confirm the right abundance fraction of endmember 1, a fine step length of Δs , which is depended on the human eye's discrimination for the color difference, was introduced to estimate the right position of abundance fraction of endmember 1 in S_{range} .

Thus, if we can calculate the step k , the final abundance fraction of endmember 1 might be written as,

$$f_{1k} = j \times \Delta f + k \times \Delta s, \quad k = 0, 1, 2, \dots, \Delta f / \Delta e, \tag{9}$$

With the constraint that the sum of abundance fraction of endmembers must be 1, the abundance fraction of endmember 2 should be,

$$f_{2k} = 1 - f_{1k}, \tag{10}$$

Then, we can get a simulated mixing reflectance r_k , which is,

$$r_k = f_{1k} \times r_1 + f_{2k} \times r_2, \tag{11}$$

where r_1 and r_2 are the reflectance of endmember 1 and 2 respectively. They are carried out by processing of endmember extraction and spectral identification mentioned above.

Obviously, the simulated reflectance r_k should look most likely the measured r_{mix} . Therefore, we choose their correlation coefficient c_k (shown in Eq. (12)) to evaluate the similarity and take the right result of k when the correlation coefficient reaches the maximum value among range of k .

$$c_k = \frac{Cov(r_{mix}, r_k)}{\sqrt{Var[r_{mix}]Var[r_k]}}, \tag{12}$$

Compared to the original method, RSDSM has been improved a number of ways. First the input spectrum is subtracted from the spectrum of the paper as a carrier to avoid its effects. Second, after obtaining the ratio derivative calculation result, the abundance range of the component to be tested is focused on a certain interval according to the statistical result of the mode. Third, the abundance values are estimated by calculating the correlation coefficient between the mixed spectrum and the simulated spectrum of the certain interval.

Parameter determining

At the above derivation of the improved algorithm, the values of parameter Δf and Δs are still not concerned. The interval parameter Δf is set to 0.1 in order to determine a coarse range for the abundance fraction. Another parameter Δs , however, is introduced to find the right position in the range. It is determined by an experiment considering the color discrimination of human being's naked eyes.

To determine the parameter Δs , blue pigment, which has the highest resolution of human eye, was selected for samples making. Figure 2 illustrates 21 pigment samples, which were mixed by Azurite and Clam white with different gradients. The total weight of the mixed pigment was 1 g. The weight of the Azurite gradually increased in units of 0.05 g as well as the weight of the Clam white decreased accordingly. Apparently, an increase of 0.05 g per weight of Azurite corresponded to an increase of 5% in its abundance.

The color difference values dE of pigments samples was measured by the AvaSpec-ULS3648 series spectroradiometer produced by Avantes BV in the Netherlands, whose spectral range was from 360 to 780 nm. Difference operation was performed on the color difference value of adjacent pigments samples that can make it more intuitive to observe that how the color difference value

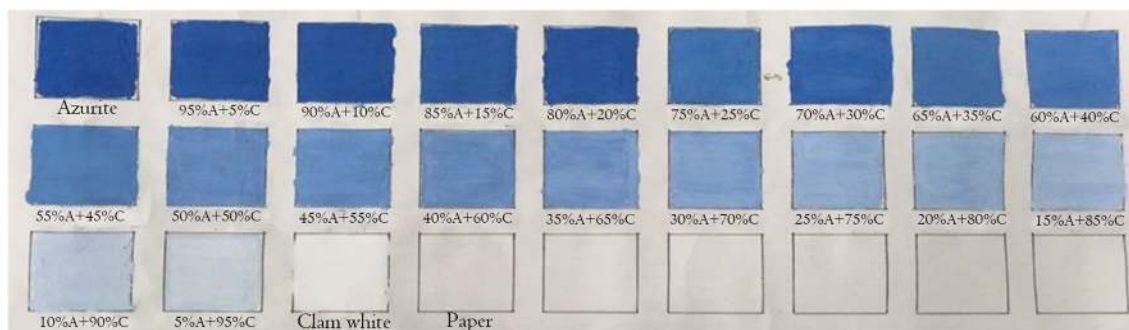


Fig. 2 Samples with different gradients of Azurite (A) and Clam white (C)

Table 2 The color difference values of different samples

| Pigment sample | <i>dE</i> | $ \Delta E $ |
|------------------------------|-----------|--------------|
| 100% Azurite | 1.3930 | |
| 95% Azurite + 5% Clam white | 2.5022 | 1.1092 |
| 90% Azurite + 10% Clam white | 3.0926 | 0.5904 |
| 85% Azurite + 15% Clam white | 7.4090 | 4.3164 |
| 80% Azurite + 20% Clam white | 6.7337 | 0.6753 |
| 75% Azurite + 25% Clam white | 9.6276 | 2.8939 |
| 70% Azurite + 30% Clam white | 14.3638 | 4.7362 |
| 65% Azurite + 35% Clam white | 17.0600 | 2.6962 |
| 60% Azurite + 40% Clam white | 18.3329 | 1.2729 |
| 55% Azurite + 45% Clam white | 19.7251 | 1.3922 |
| 50% Azurite + 50% Clam white | 22.2750 | 2.5499 |
| 45% Azurite + 55% Clam white | 24.6349 | 2.3599 |
| 40% Azurite + 60% Clam white | 26.6438 | 2.0089 |
| 35% Azurite + 65% Clam white | 28.1206 | 1.4768 |
| 30% Azurite + 70% Clam white | 30.9318 | 2.8112 |
| 25% Azurite + 75% Clam white | 39.0648 | 8.1330 |
| 20% Azurite + 80% Clam white | 39.3548 | 0.2900 |
| 15% Azurite + 85% Clam white | 41.4790 | 2.1242 |
| 10% Azurite + 90% Clam white | 51.2923 | 9.8133 |
| 5% Azurite + 95% Clam white | 56.2340 | 4.9417 |
| 100% Clam white | 67.3164 | 11.0824 |

changed with the change of pigment abundance. They are listed in Table 2.

Sun et al. [36] studied the color difference on the surface color and suggested that the color difference tolerance of the color sample chromaticity coordinates can be set as $\Delta E < 3.0$. As can be seen from Table 2, the color difference values of the majority of 19 sample points are within the range of $\Delta E < 3.0$. Therefore, it can be considered that the difference value 5% of the pigment abundance was the color discrimination thresholds of small color difference. It means that it is hard to recognize by

human being’s naked eyes when the mixing abundance difference of pigments is less than 5%. For the sake of precision, we selected 1% (0.01) of abundance difference as the value of the parameter Δs , because it is totally not discriminable for naked eyes if Δs is less than 0.01.

Accuracy evaluation index

We used the correlation coefficient and RMSE to evaluate the separation effect for mixed pixel.

The endmember spectra can be either the spectra collected in laboratory or the pure spectra extracted from an image. The larger the correlation coefficient is, the higher the similarity between the two spectra is. The smaller the RMSE is, the closer the distance between the two spectra is, and the smaller the deviation is.

Results and analysis

In this part, some mixed pigments samples made in laboratory and a Chinese painting created in the Qing Dynasty were used to test the performance of FastICA and NMF for endmember extraction. Further, the endmember separated from mixtures was matched with the standard spectra in spectral library to identify type. Finally, the RSDSM was used for abundance inversion on the laboratory mixed pigments samples data.

Endmember extraction

Experiment and result for pigments samples mixed in laboratory

In order to study the mixed spectral separation effect of actual mineral pigments by using FastICA and NMF, and to verify the validity and practicability of the two algorithms, three groups of mixed pigment samples, which were Cyanine and Ocher, Azurite and Orpiment, Zinnober and Gamboge, were made. Each pigment in each group was mixed with a certain proportion, and they respectively obtained five mixed spectra with different abundances. The endmember spectra of three groups and

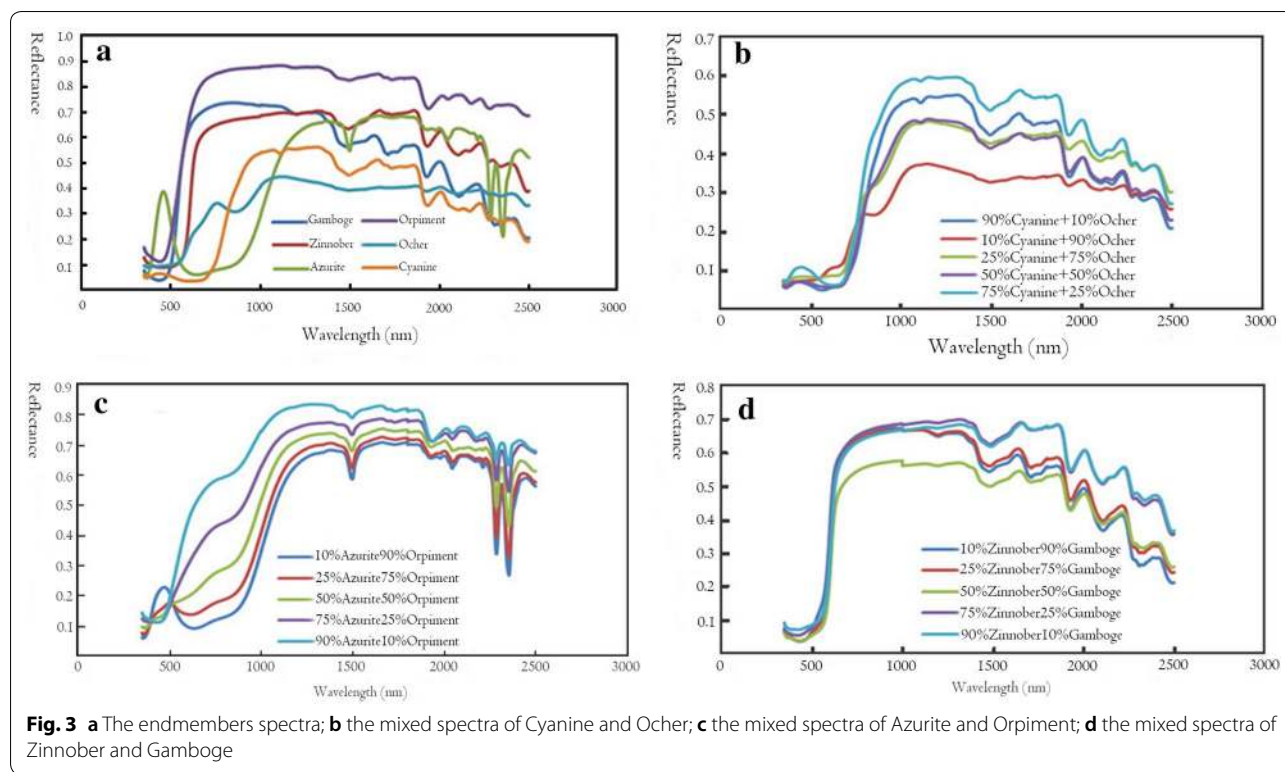


Fig. 3 a The endmembers spectra; b the mixed spectra of Cyanine and Ocher; c the mixed spectra of Azurite and Orpiment; d the mixed spectra of Zinnober and Gamboge

Table 3 The abundances of mixed pigments

| Mixture | A/% | B/% |
|---------|-----|-----|
| 1 | 90 | 10 |
| 2 | 75 | 25 |
| 3 | 50 | 50 |
| 4 | 25 | 75 |
| 5 | 10 | 90 |

their mixed pigments spectra with different abundances are shown in Fig. 3.

The pigments we used, which were not only the powder of mineral pigments Azurite and Orpiment without glue added but also the botanical pigments Cyanine and Gamboge with glue added, were grinded and sieved by a bowl and were consequently weighed by an analytical electronic balance with the 0.001-g precision. Then the pigments were added with the same amount of glue and painted on the rice paper of 4 × 4 cm to simulate the actual painting environment. The data collection environment was consistent with the environment where the spectral library was established. The specific abundances of mixed pigments samples are shown in Table 3.

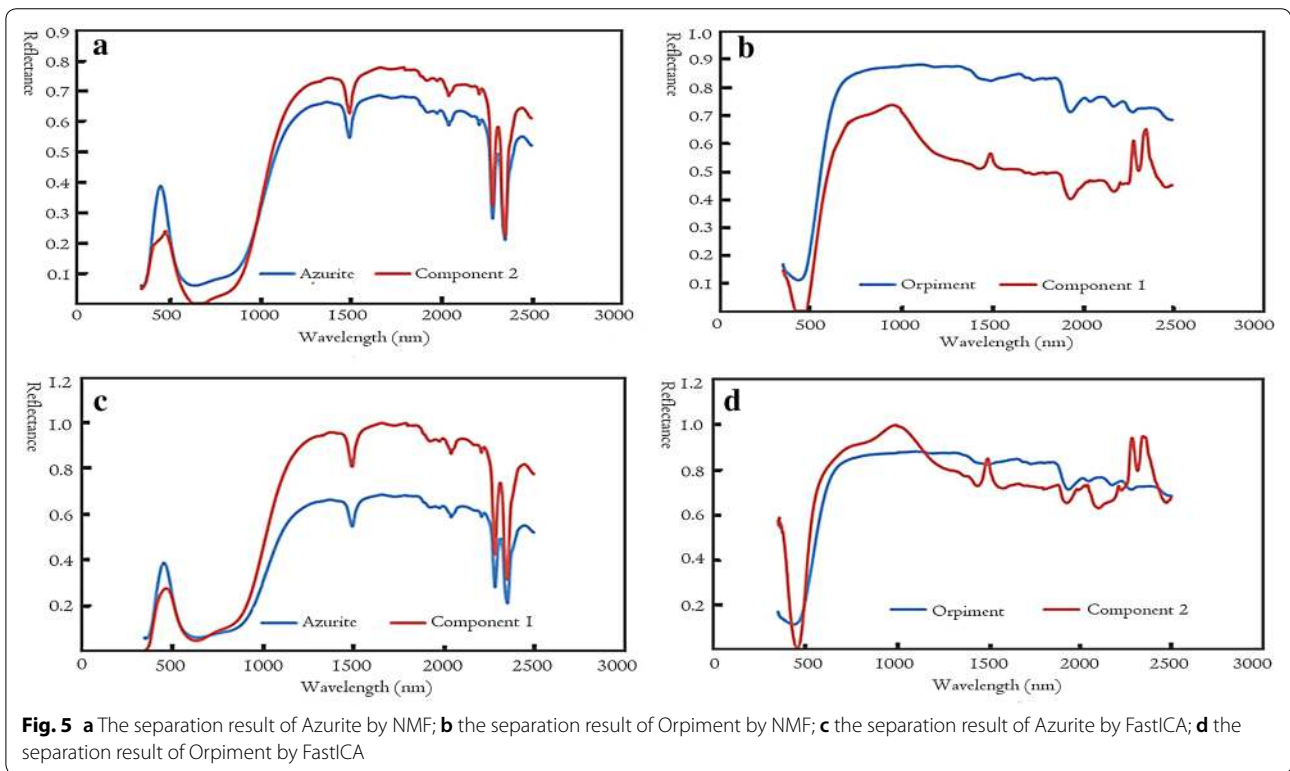
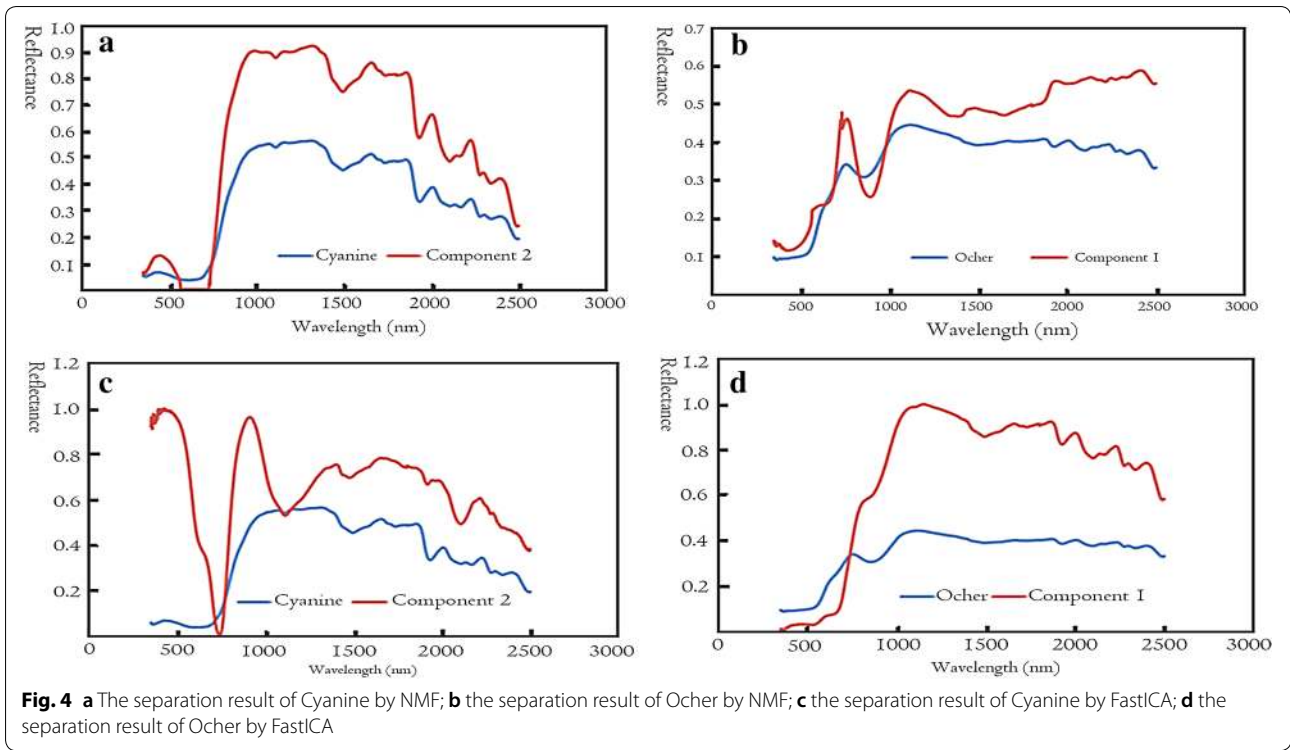
In this paper, we took the spectral data of the mixed pigments with different abundances as the original input spectral data. The results of the mixed pigments spectra

separated by NMF and FastICA in three groups (Cyanine and Ocher, Azurite and Orpiment, Zinnober and Gamboge) are shown in Figs. 4, 5 and 6, respectively. The blue curves are reflectance measured from the pure pigment while the red ones are from the endmembers extracted from the pigment mixtures by the algorithm.

In order to evaluate the separation effects of the two algorithms including NMF and FastICA more objectively, the correlation coefficient and RMSE are calculated and shown in Tables 4 and 5 and Additional file 1: Appendix S2. It should be noted that the correlation coefficient results in Table 4 are presented to illustrate the performance of the two algorithms for spectral separation. It is used to select a better algorithm for spectral separation. Therefore, in order to focus on comparing the separation effects of the two algorithms, only the components separated by the two algorithms and the known endmembers are used to calculate the correlation coefficients in the laboratory sample experiments, where two kinds of pure pigments contained are already known in the mixture.

The above experimental results can reveal the following observations:

1. By comparing the separation results of standard endmembers and components spectra (Figs. 4, 5, 6), it can be found that NMF is more suitable in spectral separation of pigment mixture than FastICA. The



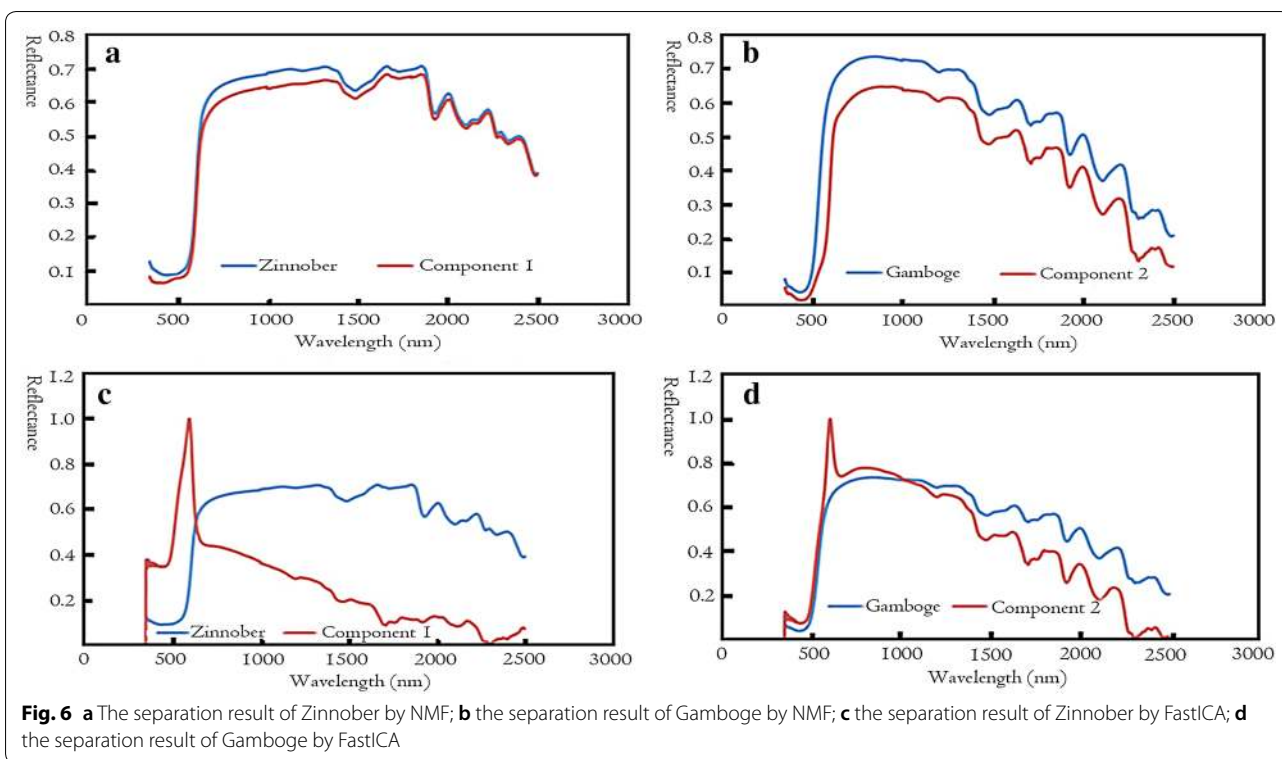


Table 4 The result of correlation coefficient (CC)

| Standard endmember | Methods | | | | CC of endmember |
|--------------------|-------------|-------------|-------------|-------------|-----------------|
| | NMF | | FastICA | | |
| | Component 1 | Component 2 | Component 1 | Component 2 | |
| Group 1 | | | | | |
| Cyanine | 0.566 | 0.995 | 0.929 | 0.241 | 0.831 |
| Ocher | 0.886 | 0.797 | 0.951 | -0.283 | |
| Group 2 | | | | | |
| Zinnober | 0.997 | 0.839 | -0.334 | 0.935 | 0.791 |
| Gamboge | 0.753 | 0.970 | 0.306 | 0.905 | |
| Group 3 | | | | | |
| Azurite | -0.091 | 0.993 | 0.990 | -0.122 | 0.321 |
| Orpiment | 0.901 | 0.357 | 0.085 | 0.838 | |

component spectra separated by NMF are highly similar to the standard spectra, and the shapes of spectra are very similar. At the same time, the correlation coefficient resulted from NMF is also higher than that from FastICA.

- By comparing the correlation coefficients between standard endmembers and the component spectra (Table 4), it can be found that the components separated by NMF are distinguishable, while one of components separated by FastICA may be similar to

two standard endmembers, which is difficult to distinguish. For example, the correlation coefficients in Group 1 of the component 1 separated by FastICA and the two standard endmembers are 0.929 and 0.951, respectively, which are both high, therefore, the type of component 1 cannot be judged by the correlation coefficient result alone. In addition, the component 2 separated by FastICA in Group 2 also has the similar indistinguishable problem. However, it is easy to make a distinction of the components

Table 5 The result of RMSE

| Standard endmember | Method | | | |
|--------------------|-------------|-------------|-------------|-------------|
| | NMF | | FastICA | |
| | Component 1 | Component 2 | Component 1 | Component 2 |
| Cyanine | 0.0303 | 0.0690 | 0.1402 | 0.1406 |
| Ocher | 0.0132 | 0.1110 | 0.1680 | 0.1509 |
| Zinnober | 0.0009 | 0.0340 | 0.1906 | 0.0719 |
| Gamboge | 0.0189 | 0.0125 | 0.1241 | 0.0194 |
| Azurite | 0.0890 | 0.0062 | 0.0549 | 0.1840 |
| Orpiment | 0.0674 | 0.1429 | 0.1172 | 0.0119 |

separated by FastICA in Group 3, probably because that the correlation coefficient between the standard endmember Azurite and Orpiment is small (0.321). In contrast, the correlation coefficients of standard endmembers in Group1 and Group 2 are higher (0.831, 0.791), which makes the separation result of FastICA difficult to distinguish. The separability of the two algorithms can be described by charts shown in Additional file 1: Appendix S2. It can be seen that the NMF algorithm has better separation effect on mixed spectra. By comparing the RMSE results between standard endmembers and component spectra (Table 5), the observations obtained above can also be demonstrated.

- From the principle of algorithms, FastICA uses different non-quadratic functions, which also has an impact on the separation results. It depends on whether the Gaussianity of the endmember spectrum is closely related to the variation law of different non-quadratic functions. In addition, the initial value of the separation matrix used in this paper was a 0 to 1 matrix generated by that random function. The initial value of each iteration was not the same, so the separation result was easy to fall into the local optimum, and the calculation result may also be not converging. If the initial values of the objective function and the separation matrix are reasonably selected, a better separation result can be obtained. NMF has only the non-negative requirement for the input matrix. However, the objective function used in this algorithm is non-convex and has multiple extremum points. The calculation result may not be the global optimal solution. Choosing a suitable initial value matrix will make it easier to obtain a better separation result.

According to the above analysis, although NMF is also easy to fall into the local optimal solution and exists other issues, it is more suitable than FastICA after concerning several factors including the accuracy,

reproducibility and stability in the processing of pigments mixtures separation.

Experiment and result for real Chinese painting

A landscape painting created in late Qing Dynasty was selected for testing FastICA and NMF. Considering the fact that neighbouring pixels own not only the endmembers with correlations but also approximated fractional abundances, which was illustrated in [37], six regions of interest (ROIs) with the same colour tone but in different positions on the surface of the painting were chosen and tagged in Fig. 7. Their spectra were acquired using the ASD spectroradiometer with its own light source in a dark room. And the data were averaged by multiple measurements to reduce noise.

Figure 8a shows the original spectra of ROIs and Fig. 8b shows the spectra of ROIs after continuum removal, which can enhance the spectral features.

From the above six ROIs, two regions named “1-Top of the pavilion” and “2-Rock in front of the pavilion” were selected as the spectral matrix to be separated. First of all, the number of separated endmember is need to be determined before performing NMF algorithm. Different from the aforementioned cases of estimating the pigment kinds used in the whole image, our research focuses on the specific regions, the component number on which mainly depends on how many kinds of pigments are usually mixed together on paintings. Traditional Chinese paintings rarely mix a variety of pigments together, usually only two or three, that's because mixing too many will result in the colors on the surface unclear and untidy [38, 39]. Generally, in order to avoid chemical reactions between pigments, and dirty or unstable on painting's surface, oil paintings won't mix more than three kinds of pigment too at the same time [40]. After observing and testing the cross sections of two real oil paintings, Cavaleri et al. [41] found that there are two superimposed green pigment layers under the whitewash layer on the surface.



Fig. 7 Six selected ROIs in the Chinese painting

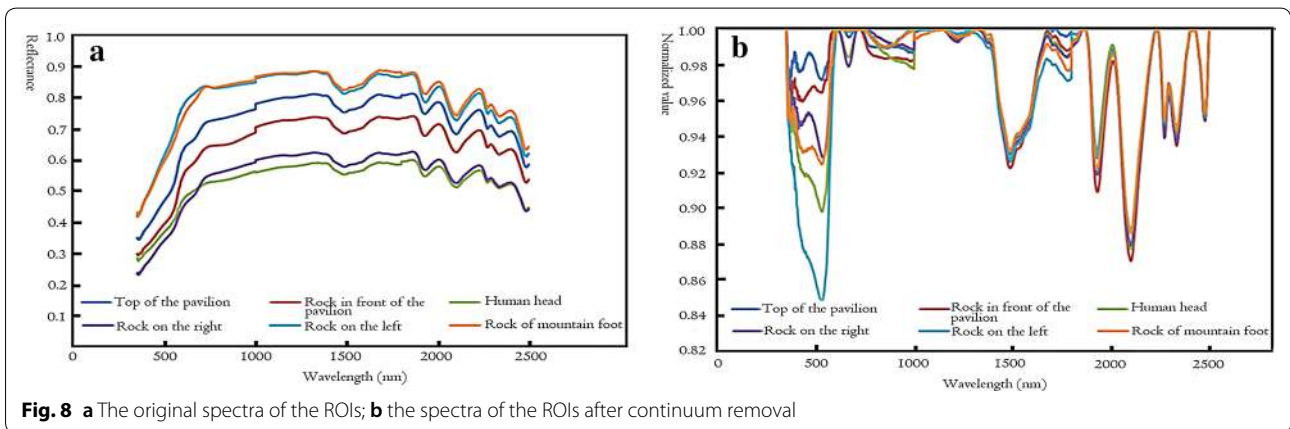


Fig. 8 **a** The original spectra of the ROIs; **b** the spectra of the ROIs after continuum removal

Inspired by the two artworks, they created a series of reference mixtures made of two or three oil pigments to simulate some areas of paints really detected in works

of art. Anglos et al. [42] prepared mixture samples consisting of two or three of the cadmium pigments for research.

Table 6 The match results of NMF

| Components | Result 1 | Result 2 | Result 3 |
|------------|--|---|--|
| NMF 1 | Terracotta (Fe ₂ O ₃) | Ocher (Fe ₂ O ₃) | Orpiment (As ₂ S ₃) |
| NMF 2 | Lead powder (Pb(OH) ₂ ·2PbCO ₃) | Yunbai | Calcite (CaCO ₃) |

Table 7 Component evaluation results

| | NMF 1 | NMF 2 | ICA 1 | ICA 2 |
|-------------------------|--------|--------|---------|--------|
| Correlation coefficient | | | | |
| Ocher | 0.9590 | 0.5098 | 0.51066 | 0.8224 |
| Lead powder | 0.7102 | 0.9642 | -0.3971 | 0.9015 |
| RMSE | | | | |
| Ocher | 0.0112 | 0.3967 | 0.2251 | 0.0925 |
| Lead powder | 0.0365 | 0.0234 | 0.0691 | 0.0506 |

It can be concluded that the color on surface of paintings are usually a mixture of two or three pigments. Therefore, the number of separated endmember in our study of the Chinese painting was set to 2. And then, we used the NMF algorithms to separate these two original spectra (rather than continuum-removed spectra) to get endmembers.

Spectral identification

The separated components of the Chinese painting were matched with the spectra in the standard spectral library using the classic spectral angle-matching algorithm.

Since it can be seen from the above experiments that the NMF algorithm was more effective than the FastICA algorithm, the component match result of the NMF was used here. The top three match results are given in Table 6.

The results show that the best match results of the two components NMF 1 and NMF 2 were Terracotta (Fe₂O₃) and Lead powder (Pb(OH)₂·2PbCO₃). However, the Ocher with the same main composition of Fe₂O₃ is more used as the pigments. The correlation coefficient and RMSE between the components by NMF and the endmembers from spectra library were calculated respectively. The results are shown in Table 7. For the correlation coefficient, the components NMF 1 and Ocher, NMF 2 and Lead powder have high similarity, and the spectral curves are consistent. In addition, the RMSE results show the same. The correlation coefficient and RMSE results of the components separated by FastICA show that the ICA 2 component has high similarity with Ocher and Lead powder, which is similar to the previous results of two experiments and shows that the separation effect of ICA is no better than that of NMF.

In order to verify whether the match results were correct, Raman Spectroscopy was performed on the same area of the head, the rock, and the roof of the selected Chinese painting. The laser light source at 532 nm was selected for measurement. The result on the head is shown in Fig. 9. The similar Raman results of rock and roof are shown in Additional file 1: Appendix S3.

Raman result in Fig. 9 shows obvious peaks around 223, 407 and 608 cm⁻¹, which are the characteristic peaks of the flexural vibration caused by Fe–O of Hematite [43].

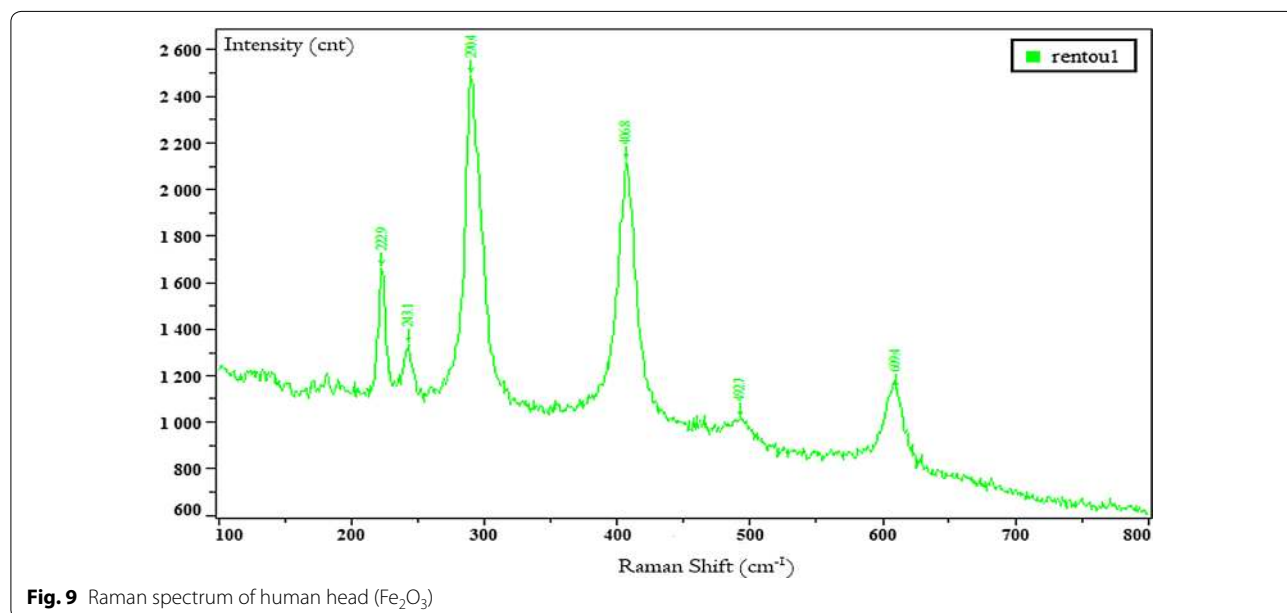


Fig. 9 Raman spectrum of human head (Fe₂O₃)

In addition, the main component of Ocher is hematite, which is Fe_2O_3 .

There are three possible reasons why Lead powder was not detected: (1) there are relatively low content and the signal of Lead powder is weak in the band of $100-700\text{ cm}^{-1}$; (2) the wavelength range of the Raman Spectroscopy used for this identification is in $100-700\text{ cm}^{-1}$, but some literature shown that the strong signal peak of Lead powder appears in 1054 cm^{-1} and other ranges [44]; and (3) to add white pigments into some heavy-toned colors is widely used in the actual painting, which often play the role of harmonize and dilute. Combining with the whole picture, the color of the selected area is lighter, so it cannot be ruled out the possibility that Lead powder was added for dilution purpose.

It can be seen from the above discussion that although the range of detection equipment is limited and the type of pigment contained in the area cannot be verified completely, combined with the pigment matching result and the painting knowledge, it can be deduced that the types

of pigments are Ocher and Lead powder, which is consistent with the result of NMF. As a result, NMF algorithm can also have a good performance on the mixed pigments separation of the actual painting.

Abundance inversion

In order to verify the validity and practicability of the RSDSM algorithm, seven groups of mineral pigment samples including Azurite and Orpiment were made, two of which were pure pigment samples (regarded as endmembers) and the other five were mixed pigments samples. The mixed pigment samples are shown in Fig. 10, and the spectra of endmember and mixed pigment are shown in Fig. 11. Mineral pigments used for this experiment were glue-free Azurite and Orpiment powder, which were accurately weighed by the same analytical electronic balance mentioned above. The pigment samples mixed with same amount of glue were painted on the area of $4 \times 4\text{ cm}$ of the rice paper to simulate the actual painting. The specific proportions of mixed pigments samples are shown in Table 8.

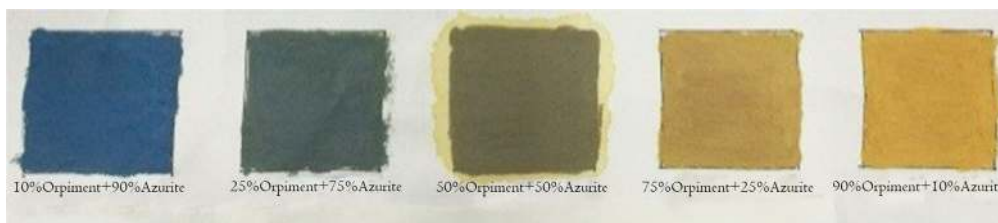


Fig. 10 The pigment samples

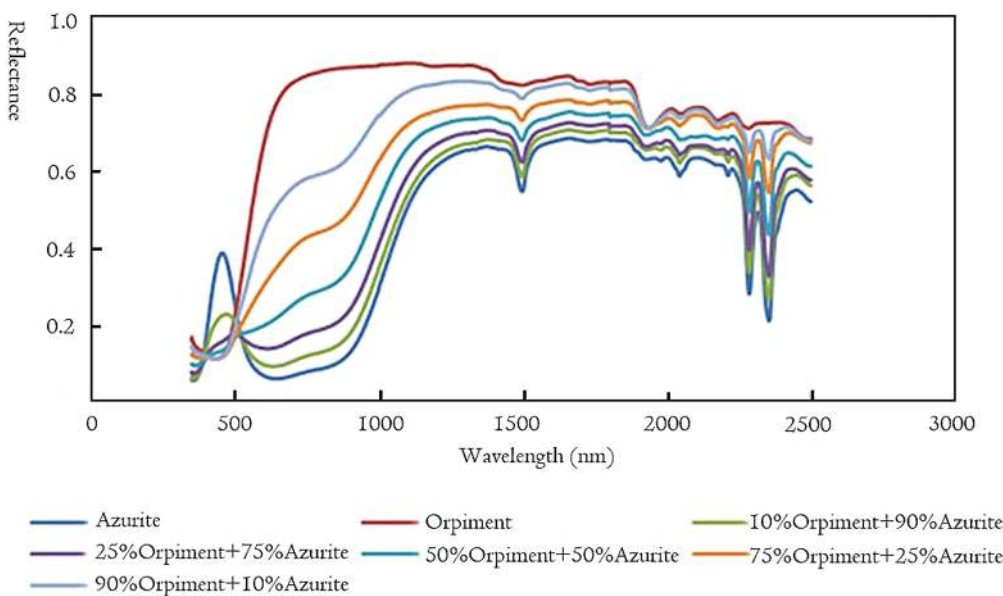


Fig. 11 The spectra of endmember and mixture

Table 8 The abundance inversion results and their RMSE of the RSDSM

| Sample ratios | Orpiment | 10% | 25% | 50% | 75% | 90% |
|-----------------|----------|------|------|------|------|------|
| | Azurite | 90% | 75% | 50% | 25% | 10% |
| Ratios by RSDSM | Orpiment | 10% | 20% | 30% | 70% | 94% |
| | Azurite | 90% | 80% | 70% | 30% | 6% |
| Error | | 0.00 | 0.05 | 0.20 | 0.05 | 0.04 |
| RMSE | 0.0685 | | | | | |

The final abundance inversion results and their RMSE by the RSDSM are shown in Table 8.

Discussion

Discussion of the component numbers setting

It is concluded in "Introduction" section that the number of separated endmember of the Chinese painting was set to 2 for the color on surface of paintings, which is usually a mixture of two or three pigments. In order to verify the impact of the number of components chosen on the result, we designed two experiments of mixed pigments. In the first experiment, two pigments including Azurite 2 and Malachite 2 were mixed. In addition, in the second experiment, three pigments, including Azurite 2, Malachite 2 and white pigment, were mixed. The number marked after the pigment name is to distinguish the different particle sizes of the pigment. As the number gets larger, the particles get finer. Then we deliberately set the number of endmembers, which we have already known, to be correct and incorrect, respectively, to study what would happen if a wrong number of components had been set.

The experimental results are shown in Figs. 12 and 13 below. In the first experiment, when the number of endmember is set to 2, the identification results are Malachite 2 and Azurite 2. When the number of endmember

is set to 3, two of the three extracted components have a similar trend, and their identification results both are Malachite 3 (see the results in Fig. 12).

In the second experiment, the right number of components is 3, and we set the number to 2, 3, and 4, respectively. The results are list as follows:

1. When the number of endmember is set to 2, the identification results are Malachite 2 and Malachite 3.
2. When the number of endmember is set to 3, the identification results are Malachite 2, Azurite 3, and white pigment, respectively.
3. When the number of endmember is set to 4, the four components are identified as Malachite 2, Malachite 3, Azurite and white pigment, respectively (see the results in Fig. 13).

It can be concluded that when the number of endmembers is set less than the true number of components, as shown in Fig. 13a, the shapes of the two components reflectance would be similar and the identification results are the same (or just different in particle size).

When the number of endmember is set larger than the true number of components, as shown in Figs. 12b and 13c, the extra endmembers would be similar with the

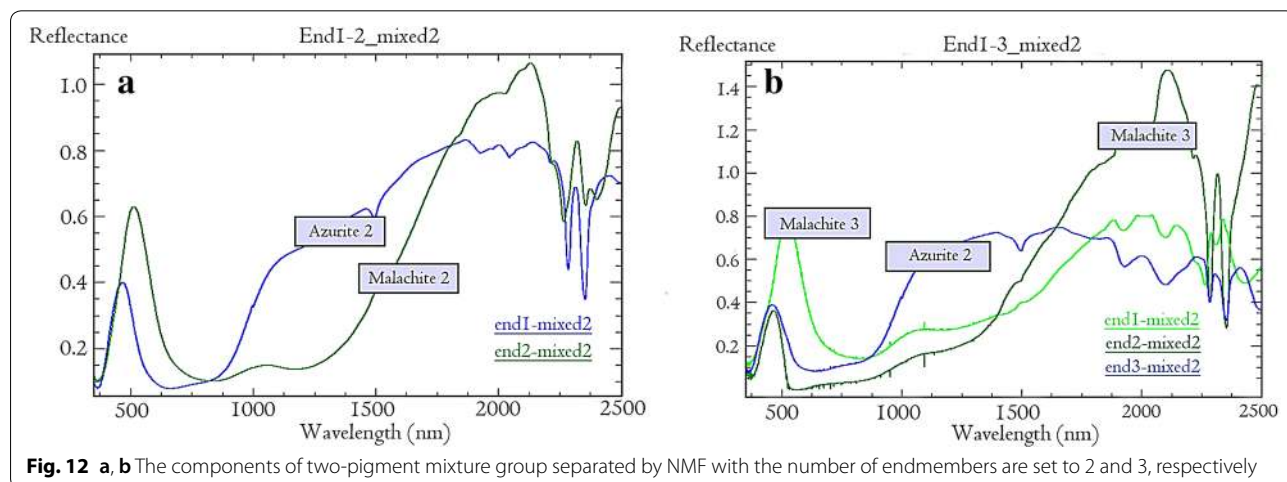
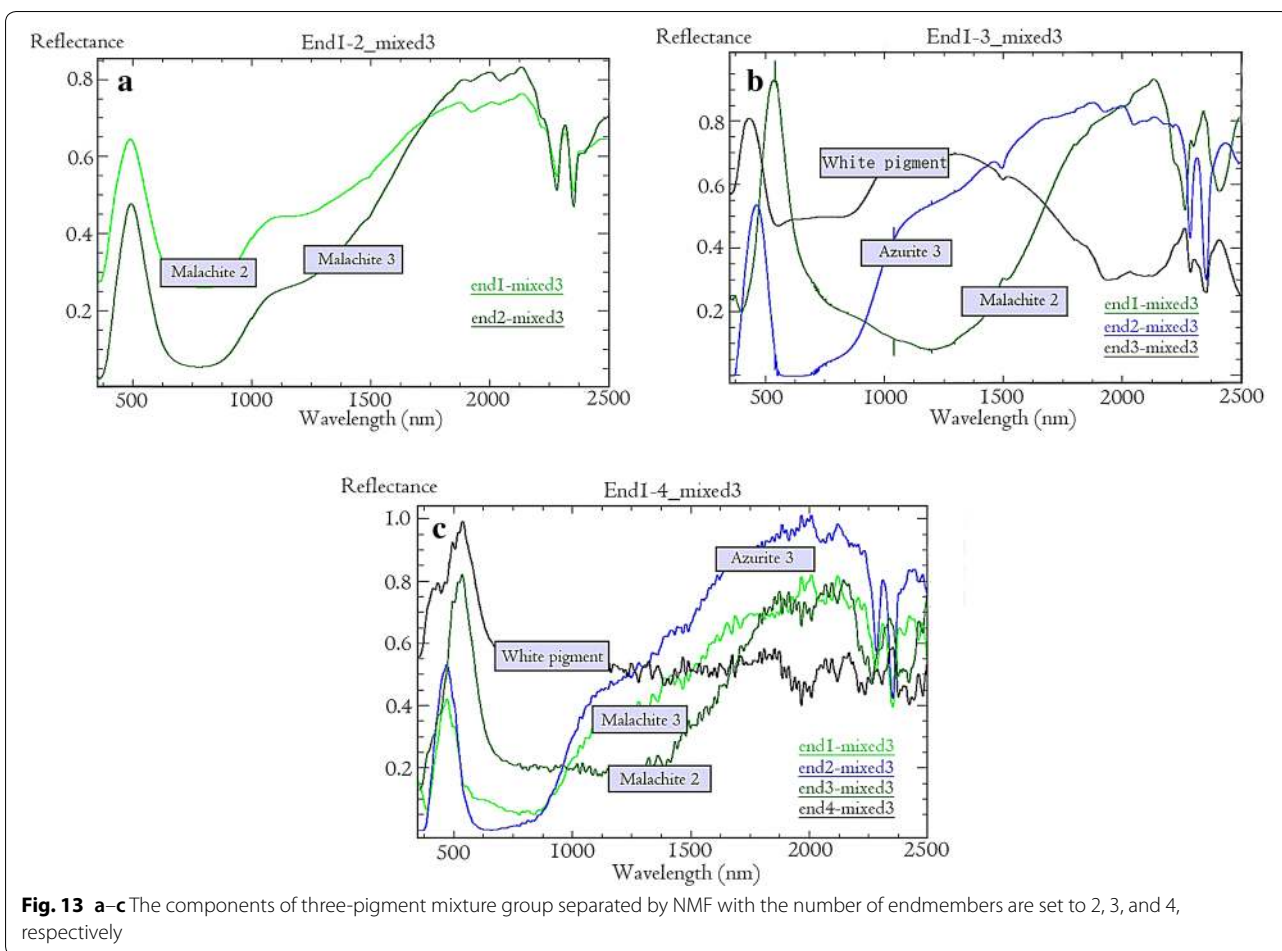


Fig. 12 a, b The components of two-pigment mixture group separated by NMF with the number of endmembers are set to 2 and 3, respectively



shape of one of the components (usually with the larger content component in the mixture), and their identification results are the same (or just different in particle size).

Therefore, if a wrong number of components had been set, there will be obvious errors, which can be easily observed and corrected.

Discussion of the applicability of NMF algorithm

The applicability of the NMF algorithm to the spectral matrix with single abundance needs further study. From the application point of view, the best way is to directly determine the endmember type and abundance of the ROIs on the painting by inputting spectral matrix constructed by multi-times measurements on the same ROIs. Unfortunately, due to the restrictions of the principle of the NMF algorithm, the decomposed solution of NMF is not unique when the input matrix to be decomposed is a matrix with exactly the same column vectors. Therefore, the objective function is more obviously non-convexities and easier to fall into the local optimum in the solving process. At present, a common way to solve this problem is to add some constraint conditions, such

as, smoothness and sparseness. Therefore, the applicability of NMF algorithm from above aspects should be improved in future.

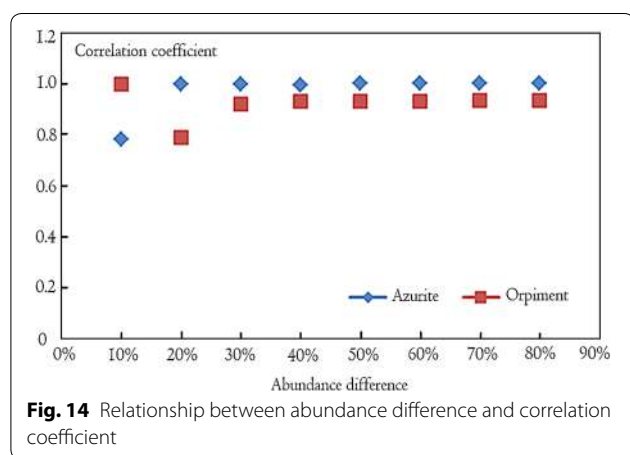
In this paper, the spectra regarded as the input spectral matrix to be decomposed were mixed spectra, which were collected on different ROIs distributed on different positions on the painting with the same tone but different saturation. The spectral matrix collected in this way has some abundance difference, which provides input matrix with unequal column vectors.

Discussion of spectral identification

In "Abundance inversion algorithms" section, we mentioned that one of the reasons way the Raman spectroscopy failed to detect Lead powder may be that the content of Lead powder is too low while the correlation coefficient between the component separated from the mixture and the endmember in spectral library are high. In order to prove this point, we studied how the abundance difference in the mixture affects the correlation coefficient between the extracted component and the endmember in spectral library. We made 8 groups of

Table 9 The correlation coefficient between the endmember and the corresponding component separated from different abundances by using NMF

| Group | Abundance difference (%) | Composition | Correlation coefficient of Azurite and corresponding component | Correlation coefficient of orpiment and corresponding component |
|-------|--------------------------|---|--|---|
| 1 | 10 | 10% Azurite, 90% Orpiment + 20% Azurite, 80% Orpiment | 0.7781 | 0.9965 |
| 2 | 20 | 10% Azurite, 90% Orpiment + 30% Azurite, 70% Orpiment | 0.9956 | 0.7885 |
| 3 | 30 | 10% Azurite, 90% Orpiment + 40% Azurite, 60% Orpiment | 0.9973 | 0.9176 |
| 4 | 40 | 10% Azurite, 90% Orpiment + 50% Azurite, 50% Orpiment | 0.9917 | 0.9274 |
| 5 | 50 | 10% Azurite, 90% Orpiment + 60% Azurite, 40% Orpiment | 0.9986 | 0.9278 |
| 6 | 60 | 10% Azurite, 90% Orpiment + 70% Azurite, 30% Orpiment | 0.9998 | 0.9290 |
| 7 | 70 | 10% Azurite, 90% Orpiment + 80% Azurite, 20% Orpiment | 0.9988 | 0.9309 |
| 8 | 80 | 10% Azurite, 90% Orpiment + 90% Azurite, 10% Orpiment | 0.9984 | 0.9330 |



mixed samples with different abundances using Azurite and Orpiment, and calculated the correlation coefficients of the endmember and corresponding components (as shown in Table 9 and Fig. 14).

It can be seen that, except for the first group, the correlation coefficient is not directly related to the content of endmember within mixture. Moreover, when the abundance difference in the mixture is greater than 30%, the correlation coefficients of the endmember and the corresponding component are both higher than 0.92, even if one has less content in the mixture.

Therefore, the result that the correlation coefficient of component 2 and lead powder in Table 7 is higher than that of Fe₂O₃ could not indicate that the content of Lead powder is higher than that of Fe₂O₃, so

that sufficient signal may not be generated and not be detected by Raman spectroscopy.

Discussion of abundance inversion

In order to further verify the performance of the RSDSM algorithm, other two algorithms include FCLS and MVSA, which are also commonly used for abundance inversion, were applied on same samples. The results are shown in Table 10, and the RMSE are calculated and shown in Table 11.

According to above experiments, we mainly analyzed the results from the following aspects:

1. For the RMSE results of the three algorithms, the best algorithm is the RSDSM, whose RMSE is 0.0685. RMSE of FCLS and MVSA are 0.0791 and 0.0941 respectively. For the error ranges of the three algorithms, the FCLS is better than MVSA with the smallest error of 1.79% and the largest error of 23.5%. MVSA can quickly calculate and get the abundance value, but the accuracy is not high. Its largest error value of abundance reaches to 25.3%. The RMSE of the RSDSM is the smallest. Its smallest error value of abundance is 0. Compared with the results of the other two algorithms, most of the abundance values used by the RSDSM are more accurate, and the overall accuracy is highest.
2. The range setting of the statistical interval Δf for abundance inversion is also an importance aspect which can play a vital role in the unmixing accuracy. Too large statistical interval can satisfy the require-

Table 10 The abundance results of different algorithms

| Method | Sample ratios | | | | | |
|--------|---------------|--------|--------|--------|--------|--------|
| | Orpiment | 10% | 25% | 50% | 75% | 90% |
| | Azurite | 90% | 75% | 50% | 25% | 10% |
| FCLS | Orpiment | 0.0821 | 0.1748 | 0.3234 | 0.5150 | 0.7169 |
| | Azurite | 0.9179 | 0.8252 | 0.6766 | 0.4850 | 0.2831 |
| MVSA | Orpiment | 0.1075 | 0.1943 | 0.3219 | 0.4970 | 0.6865 |
| | Azurite | 0.9498 | 0.8765 | 0.7689 | 0.5972 | 0.3940 |
| RSDSM | Orpiment | 0.1000 | 0.2000 | 0.3000 | 0.7000 | 0.9400 |
| | Azurite | 0.9000 | 0.8000 | 0.7000 | 0.3000 | 0.0600 |

Table 11 The RMSE result of different algorithms

| Ratios by methods | Sample ratios | | | | | | RMSE | Error range (%) |
|-------------------|---------------|--------|--------|--------|--------|--------|--------|-----------------|
| | Orpiment | 10% | 25% | 50% | 75% | 90% | | |
| | Azurite | 90% | 75% | 50% | 25% | 10% | | |
| FCLS | | 0.0179 | 0.0752 | 0.1766 | 0.2350 | 0.1831 | 0.0791 | 1.79–23.5 |
| MVSA | | 0.0075 | 0.0557 | 0.1781 | 0.2530 | 0.2135 | 0.0941 | 0.75–25.3 |
| RSDSM | | 0.0000 | 0.0500 | 0.2000 | 0.0500 | 0.0400 | 0.0685 | 0–20 |

ments of the accuracy but it will be time consuming and inefficiency. On the other hand, too small statistical interval Δf will lead to a larger error. Therefore, it is very important to determine a suitable range of statistical interval.

3. The operational error often occurs in the pigment sample-making period. For example, the powdery mineral pigment is easy to spill during the weighing and mixing period. In addition, there might be some partial uneven region during the drawing process. The error is still inevitable in despite of making samples under strict operation instructions and painting in several times.
4. The choice of pigments has a dramatic effect on the visual compensation of paintings. However, traditional inpainting is a step-by-step trial process often from shallow to deep, which is a gradual process of exploration. The improved method in this paper can provide a roughly abundance interval for the inpainting process of the paintings restoration, which will narrow the test color range. It is helpful to improve the efficiency of the restoration work.

Conclusions

Typical endmember extraction methods of ICA and NMF were both used for pigments mixtures separation to obtain their components. Then the pigment’s types were determined by calculating the similarity between

components and standard endmembers in the pigment spectral library. Moreover, we used laboratory mixed pigments data to compare the separation effect of the two algorithms. We then analyzed the mixed pigment of a painting created in the late Qing Dynasty and verified the separation results by Raman Spectroscopy. The experimental results show that, compared with ICA, NMF has a higher performance both in separation quality and separation stability. Therefore, NMF is more suitable for spectral separation of mixed pigments.

An improved method named RSDSM was proposed to finish abundance inversion of mixed spectra. Moreover, two commonly used methods of FCLS and MVSA were introduced to compare the performance with the method improved in this paper. The experiment shows that the method proposed in this paper has a better performance.

Therefore, the area with similar tone on the surface of a painting can be separated into several components by the NMF. After identified their kinds by compared the similarities with spectra in spectral library, the abundances of endmember pigments can be calculated by the RSDSM proposed. The pigments and their proportion will be figured out in this way.

At the same time, it should be noticed that the RSDSM also has some constraints in some aspects. The method proposed at present can only deal with the unmixing problem within the similar tone areas in an image one by one, and it should be improved in future to solve the problem for a whole image. Further, the

unmixing accuracy depends on the input samples spectra selected manually, the accuracy of the endmember extraction, and the statistical range of the initial abundance value. The automation and practicality of this method should be improved by combining with endmembers automatic extraction and other methods in future.

Furthermore, we will also exploit other spectral unmixing methods, which are more suitable for pigment mixing properties, to improve our results in subsequent study.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40494-020-00372-5>.

Additional file 1. Additional figures.

Abbreviations

ICA: Independent component analysis; NMF: Non-negative matrix factorization; RSDS: Ratio spectra derivative spectrophotometry; LMM: Linear mixing model; NLMM: Non-linear mixing model; LSA: Least square algorithm; MVSA: Minimum volume simplex analysis; RMSE: Root mean square error; RSDSM: ratio spectra derivative spectrophotometry based on the mode; FCLS: Full-constrain least squares algorithm; MVSA: Minimum volume simplex analysis; ASD: Analytical spectral devices; ROIs: Regions of interest.

Acknowledgements

The authors would like to thank Professor Songnian Li from Ryerson University, Canada for his constructive comments on the preparation and many rounds of revisions of the paper and guidance during the research.

Authors' contributions

Conceptualization: LSQ, LYY, HML, and YQL. Methodology: LSQ, LYY, and YQL. Validation: LSQ, LYY, YQL, WWT, and YXY. Formal analysis: LSQ, LYY, YQL, and YXY. Resources: WWT; Writing—original draft preparation and writing—review: all authors. The paper was approved by all authors. All authors read and approved the final manuscript.

Funding

This research was funded by the Research Fund of the National Key Research and Development Program (No. 2017YFB1402105), the Fundamental Research Funds for Beijing University of Civil Engineering and Architecture (No. X18024) and Beijing Advanced Innovation Center for Future Urban Design (No. UDC2016030200).

Availability of data and materials

Data sharing is not applicable to this article as some of the data was collected on the surface of real cultural relics.

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication

The consent for the publication of details and images in the manuscript are obtained from all participants.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, No. 15 Yongyuan Road, Daxing District, Beijing 102616, China. ² Beijing Key Laboratory for Architectural Heritage Fine Reconstruction & Health Monitoring, No. 15 Yongyuan Road, Daxing

District, Beijing 102616, China. ³ Construction Comprehensive Survey Research and Design Institute (CIGIS) Co., Ltd., No. 177 Dongzhimen nei Street, Dongcheng District, Beijing 100007, China. ⁴ Capital Museum, No. 16 Fuxingmenwai Street, Xicheng District, Beijing 100045, China.

Received: 29 January 2020 Accepted: 16 March 2020

Published online: 30 March 2020

References

- Liu ZF, Zhang H, Zhou WH, et al. Pigment identification on an undated Chinese painting by non-destructive analysis. *Vib Spectrosc*. 2019;101:28–33.
- Wang JY, Wei L, Liu ZJ. Raman spectra of mineral pigments used in ancient Chinese artworks. *J Light Scatt*. 2012;24:1 (in Chinese).
- Xia Y, Wu SC, Cui SK, et al. Study on the pigments in polychrome potteries of West Han Dynasty tomb in Weishan, Shandong province. *Sci Conserv Archaeol*. 2008;20(2):13–9 (in Chinese).
- Aliatis I, Bersani D, Campani E, et al. Green pigments of the Pompeian artists' palette. *Spectr Acta Part A Mol Biomol Spectrosc*. 2009;73(3):532–8.
- Nam JY, Han K, Ji JE, et al. Raman spectroscopy of the pigments on Korean traditional paintings. *Vib Spectrosc*. 2016;85:181–4.
- Zhang LP, Du B, Zhang LF. *Hyperspectral remote sensing image processing*. Beijing: Science Press; 2014.
- Delaney JK, Ricciardi P, Glinsman LD, et al. Use of imaging spectroscopy, fiber optic reflectance spectroscopy, and X-ray fluorescence to map and identify pigments in illuminated manuscripts. *Stud Conserv*. 2014;59(2):91–101.
- Pottier F, Michelin A, Tournié A, et al. Macroscopic reflectance spectral imaging to reveal multiple and complementary types of information for the non-invasive study of an entire polychromatic manuscript. *J Cult Herit*. 2019;35:1–15.
- Romano C, Lam T, Newsome GA, et al. Characterization of zinc carboxylates in an oil paint test panel. *Stud Conserv*. 2019. <https://doi.org/10.1080/00393630.2019.1666467>.
- Grabowski B, Masarczyk W, Glomb P, et al. Automatic pigment identification from hyperspectral data. *J Cult Herit*. 2018;31:1–12.
- Deborah H, George S, Hardeberg JY. Spectral-divergence based pigment discrimination and mapping: a case study on The Scream (1893) by Edvard Munch. *J Am Inst Conserv*. 2019;58(1–2):90–107.
- Ritter GX, Urcid G. Lattice algebra approach to endmember determination in hyperspectral imagery. *Adv Imag Electr Phys*. 2010. [https://doi.org/10.1016/S1076-5670\(10\)60004-3](https://doi.org/10.1016/S1076-5670(10)60004-3).
- Bai D, Messinger DW, Howell D. Hyperspectral analysis of cultural heritage artifacts: pigment material diversity in the gough map of Britain. *Opt Eng*. 2017;56(8):081805. <https://doi.org/10.1117/1.OE.56.8.081805>.
- Bai D, Messinger DW, Howell D. A hyperspectral imaging spectral unmixing and classification approach to pigment mapping in the Gough & Selden Maps. *J Am Inst Conserv*. 2019;58(1–2):69–89.
- Zhang J, Rivard B, Sanchez-Azofeifa A. Derivative spectral unmixing of hyperspectral data applied to mixtures of lichen and rock. *IEEE Trans Geosci Remote Sens*. 2004;42(9):1934–40.
- Adams JB, Gillespie AR. *Remote sensing of landscapes with spectral images: a physical modeling approach*. Cambridge: Cambridge University Press; 2006.
- Borel CC, Gerstl SA. Nonlinear spectral mixing models for vegetative and soil surfaces. *Remote Sens Environ*. 1994;47(3):403–16.
- Lin HL, Zhang X, Sun YL. Hyperspectral sparse unmixing of minerals with single scattering albedo. *J Remote Sens*. 2016;20(1):53–61 (in Chinese).
- Wang RS, Gan FP, Yan BK, et al. Hyperspectral mineral mapping and its application. *Remote Sens Land Resour*. 2010;22(1):1–13 (in Chinese).
- Craig MD. Minimum-volume transforms for remotely sensed data. *IEEE Trans Geosci Remote Sens*. 1994;32(3):542–52.
- Khoshsokhan S, Rajabi R, Zayyani H. Sparsity-constrained distributed unmixing of hyperspectral data. *IEEE J Select Topics Appl Earth Observ Remote Sens*. 2019;12(4):1279–88.
- Herault J, Jutten C. Space or time adaptive signal processing by neural network models. *AIP Conf Proc*. 1986. <https://doi.org/10.1063/1.36258>.
- Hyvriinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Netw*. 2000;13(4–5):411–30.

24. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999;401(6755):788–91.
25. Wu FF, Guo PY, Wei C. Hyperspectral image analysis and application. Tokyo: Science Press; 2013.
26. Salinas F, Nevado JB, Mansilla AE. A new spectrophotometric method for quantitative multicomponent analysis resolution of mixtures of salicylic and salicylic acids. *Talanta*. 1990;37(3):347–51.
27. Nevado JB, Cabanillas CG, Salcedo AC. Simultaneous spectrophotometric determination of three food dyes by using the first derivative of ratio spectra. *Talanta*. 1995;42(12):2043–51.
28. Nikam A, Pawar SS, Gandhi S. Estimation of paracetamol and aceclofenac in tablet formulation by ratio spectra derivative spectroscopy. *Indian J Pharm Sci*. 2008;70(5):635–7.
29. Galera MM, Vidal JM, Frenich AG. First derivative of the ratio spectra method for resolving iodide and thiocyanate in binary mixtures. *Talanta*. 1994;41(9):1545–51.
30. Zhang LF, Huang CP. Research on the model of spectral unmixing for minerals based on derivative of ratio spectroscopy. *Spectr Spectr Anal*. 2013;33(1):172–6 (in Chinese).
31. Lei AW, Tong H, Wan XX. Study of the material foundation of Dunhuang murals and aging test of red lead and the study of color representation method of cultural heritage. *Sci Technol Innov Herald*. 2016;17:178–9 (in Chinese).
32. An JJ. Analysis and characterization of the binding media of ancient polychrome artworks by MALDI-TOF-MS. Kirkland: Northwest University; 2012 (in Chinese).
33. Adams JB, Smith MO, Johnson PE. Spectral mixture modeling: a new analysis of rock and soil types at the Viking Lander 1 site. *J Geophys Res Solid Earth*. 1986;91(B8):8098–112.
34. Tsai F, Philpot W. Derivative analysis of hyperspectral data. *Remote Sens Environ*. 1998;66(1):41–51.
35. Huguenin RL, Jones JL. Intelligent information extraction from reflectance spectra: absorption band positions. *J Geophys Res Solid Earth*. 1986;91(B9):9585–98.
36. Sun XR, Lin ZD, Zhang JY, et al. Discrimination of color difference of surface. *Acta Psychol Sin*. 1996;28(1):9–15 (in Chinese).
37. Li JJ, Li YS, Song R, et al. Local spectral similarity preserving regularized robust sparse hyperspectral unmixing. *IEEE Trans Geosci Remote Sens*. 2019;57(10):7756–69.
38. Liao MY, Zeng QR. The use of color in Chinese painting and its application techniques. *Art Educ Res*. 2011;6:42–3 (in Chinese).
39. Bi JX. Relationship between color and ink color of Chinese painting—Status of Chinese painting color theory. *Trad Chin Paint*. 1998;6:4–9 (in Chinese).
40. Performance and taboos of commonly used pigments in oil painting. http://www.360doc.com/content/15/0211/12/19787817_447891606.shtml. Accessed 6 Nov 2019.
41. Cavaleri T, Giovagnoli A, Nervo M. Pigments and mixtures identification by Visible Reflectance Spectroscopy. *Proc Chem*. 2013;8:45–54.
42. Anglos D, Solomidou M, Zergioti I, Zafropoulos V, Papazoglou TG, Fotakis C. Laser-induced fluorescence in artwork diagnostics: an application in pigment analysis. *Appl Spectrosc*. 1996;50(10):1331–4.
43. Chang JJ. Raman studies of the pigments in ancient wall paintings and dyes. *J Jilin Univ*. 2010;8:201 (in Chinese).
44. Hu WY, Wang Y. Research and analysis of Raman spectroscopy in the murals of the Tombs of the Northern Qi Dynasty in Shuiquanliang. *Silicon Valley*. 2012;17:153–4 (in Chinese).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
