G Model
CULHER-3328; No. of Pages 12
ARTICLE IN PRESS

Journal of Cultural Heritage xxx (2017) xxx–xxx

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Original article

Automatic pigment identification from hyperspectral data

Bartosz Grabowski [a], Wojciech Masarczyk [a,*], Przemysław Głomb [a], Agata Mendys [b]

[a] Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5, 44-100 Gliwice, Poland
[b] National Museum in Krakow, Laboratory of Analysis and Nondestructive Investigation of Heritage Objects (LANBOZ), Piłsudskiego 14, 31-109 Kraków, Poland

INFO ARTICLE

Historique de l'article :
Reçu le 25 septembre 2017
Accepté le 5 janvier 2018
Disponible sur Internet le xxx

Keywords:
Pigment identification
Hyperspectral imaging
Classification
Endmember estimation
Spectral unmixing

ABSTRACT

Art objects conservation or historical analysis necessitates a thorough knowledge of materials used by the artist and their subsequent changes. In the case of paintings this requires the ability to correctly identify the pigments that were used for creation or later restoration of the artwork. This is a challenging problem, as the applied method should be non-contact, robust for the wide variety of chemical substances used and straightforward in the interpretation. Recently, the hyperspectral imaging has emerged as a promising measuring methodology for this kind of the artwork analysis; the combination of acquiring spectral information and planar (photography-like) pixel arrangement provides a lot of potential for material characterization. While initial studies of hyperspectral imaging application to art objects analysis are encouraging, the difficulties of working with its multidimensional data are acknowledged; in many cases complex algorithms are required to fully utilize its potential. In this paper, we study the problem of algorithm design for pigment identification based on a hyperspectral image of a painting. We combine various processing steps to achieve a robust solution requiring minimal user intervention. Using a special set of paintings and a reference pigment database we demonstrate the viability of applying this method in the pigment recognition setting. Our results confirm the potential of using hyperspectral imaging in the art conservation setting, and based on them we discuss the potential construction and elements of such an algorithm.

1. Introduction

Paintings are one of the crucial elements of the tangible cultural heritage (CH). Due to their importance, uniqueness and physical characteristics they require special tools for their safeguarding, restoration or understanding of their historical context. There are many problems related to the conservation and restoration of paintings, e.g. the removal of dirt and old varnish without affecting the paint layer; paint stabilization by consolidating loose paint and removing consolidant residues from previous treatments [1]; selection of materials for retouching that minimizes metamerism and with ageing behaviour similar to the original paint layer [2]; selecting proper storage conditions, both in terms of humidity and temperature, as well as exposure to light [3,4]. All of the aforementioned problems require a thorough knowledge of the materials used by the artist and during the subsequent changes, their chemical composition and determination of their preservation state. Therefore, the conservation process nowadays is supported by other fields of science like chemistry, physics or information science. Since artworks are unique and usually very complex systems, the techniques used in their investigations are still modified to meet specific requirements and new techniques are studied. Due to the nature of cultural heritage objects taking samples from them is not advised and very often not allowed. Non-invasive, or even non-contact methods, that can be applied in-situ are therefore necessary. What is more, cultural heritage objects are very often heterogeneous in composition which increases interest in full-field imaging methods.

A very important problem in the process of this kind of art analysis is the proper pigment identification. From the conservators' point of view, this knowledge is important for choosing the proper materials for restoration and to assess the original material's sensitivity to environmental conditions. From the point of view of an art historian – to study the artist's craft and to help in artwork's authenticity verification. One of the most popular laboratory methods for pigment identification is X-ray fluorescence spectrometry (XRF). However, as was shown in round robin tests [5], interpretation of its results is not always straightforward and can be misleading; verification with other complementary methods is necessary.

* Corresponding author.
Adresses e-mail : bgrabowski@iitis.pl (B. Grabowski), wmasarczyk@iitis.pl (W. Masarczyk), przemg@iitis.pl (P. Głomb), amendys@muzeum.krakow.pl (A. Mendys).

https://doi.org/10.1016/j.culher.2018.01.003
1296-2074/

Pour citer cet article : B. Grabowski, et al., Automatic pigment identification from hyperspectral data, Journal of Cultural Heritage (2017), https://doi.org/10.1016/j.culher.2018.01.003

One of the alternative methods that that have been recently gaining popularity in the CH field is the hyperspectral imaging (HSI) in the visible and near infra-red regions [6,7]. Hyperspectral imaging captures image data in hundreds of narrow contiguous spectral bands, producing at each pixel a precise light reflectance information. This data can be used for detailed analysis of the imaged object's surface, including substance degradation estimation, material classification or anomaly detection. It is widely used as an aid in e.g. identification of pigments, dyes and binders [8,9], analysis of synthetic as well as natural polymers [10,11], detection of cracks on paint layers [12] or artwork authentication [13].

1.1. Hyperspectral imaging for pigment identification

Hyperspectral imaging has been recently used as a base for the methods supporting pigment component identification in the art conservation setting. In [14] a case study of pigment identification on the famous 'The Scream' (1893) by Edvard Munch is presented. The authors compare the methods of: fully constrained unmixing with pre-set parameters and two standard spectral distance measures (Spectral Angle Mapper and Spectral Correlation Mapper) in matching areas of the painting to the pigment library. They note the necessity of using unmixing to properly process inhomogeneous regions. The authors of [15] create infra-red false-colour (IRFC) images for easier distinction between different pigments. Using the data from the pigment spectral library, they manually identify three spectral bands which maximize distinction between a selected group of pigments (e.g. red pigments of cinnabar, red ochre, red lake and minium). The proposed approach is easy to use and has good performance in comparison to traditional methods. In [16] a variety of paintings were processed for pigment identification, based on hyperspectral images in VNIR and SWIR ranges. A tailored endmember identification algorithm was used to get endmember spectra, which were subsequently evaluated and compared with an analysis done using other methods, e.g. XRF. They note the utility of hyperspectral data, especially in the SWIR range, for improved false colour visualization of changes from paintings having large and complex reworkings made by the original artist or by a subsequent conservator. Two different hyperspectral imaging systems were used in the analysis of Goya's paintings [17]; the constructed false-colour images were used to confirm XRF results and additionally to identify the non-restored zones of the paintings. A detailed analysis of pigment-binder mixtures is presented in [18]. The authors analyze six pigments (red and yellow ocher, chrome yellow, mineral blue, cobalt green and malachite) and four binders (Arabic gum, gouache, egg tempera and linseed oil). The presented results show good separability (qualitatively evaluated on the Principal Component Analysis plots) between binders knowing the pigment, and good supervised classification of binders (done using Partial Least Squares Discriminant Analysis, PLS-DA). Very recently [13] have successfully attempted to use a state-of-the-art pattern recognition algorithm (Support Vector Machine, SVM) for spectral classification of pixel data into pigment classes in the context of art authentication. It should be noted that the problem of pigment component identification is not restricted to paintings; for example, [19] use spectral signatures of chromogenic colour photographs to identify photographic paper manufacturers, while [20] use spectral processing to identify colour dyes in the collections of indigenous textiles.

1.2. Multivariate data analysis for hyperspectral imaging

Typically, for the analysis of large chemical data sets chemometric methods are used i.e. methods based on statistical data analysis. In the case of hyperspectral data, the analysis of spectra from many image pixels requires the use of multivariate statistical analysis techniques. These methods are widely used and well-established in the field of HSI, and their description and examples of applications can be found in many textbooks, e.g. [21]. Nevertheless, for each specific type of application, the development of specific algorithms or sequences of algorithms is necessary, in order to achieve an optimal performance.

The task of identifying pigments from hyperspectral data is complicated due to the fact that each observed spectrum is a combination of different sources. To extract the desired source, we apply spectral unmixing methods. There exist a number of methods tailored specifically to practical hyperspectral unmixing and source estimation problems [22], notably geometrical and statistical approaches. In geometrical approaches it is assumed that investigated spectra are a linear combination of material sources, and we can try to minimize the volume of a simplex enclosing them so that vertices of the final simplex approximate original sources. These methods can be divided into two categories: the ones with the pure pixel assumption – that data includes points that contain pure component spectra – and those that do not require it. Examples of the former are Vertex Component Analysis (VCA) [23], which iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined and NFINDR [24], which outputs minimum volume simplex spanning the pixels. One of the algorithms without pure pixel assumption is the Simplex Identification via Variable Splitting and Augmented Lagrangian (SISAL) [25], which minimizes the simplex enclosing most (but not necessarily all) data points, resulting in robustness to noise in the data. In contrast to SISAL, the Minimum Volume Enclosing Simplex (MVES) [26] finds the simplex which encloses all the points. The statistical approaches form another group of methods which in some cases (esp. highly mixed scenarios) are reported to obtain better results at the expense of a higher computational complexity. Among those methods, Bayesian approaches can be highlighted, e.g. Bayesian Linear Unmixing [27].

To ensure the optimum performance of unmixing methods, we need to provide a homogeneous region of the same pigment. To do so, we can use one of the dimensionality reduction algorithms followed by clustering. The widely applied dimensionality reduction algorithms are for example: t-Stochastic Distributed Neighbourhood Embedding (t-SNE) [28], which projects high-dimensional data into a two- or three-dimensional space in such a way that spectrally similar points are close to each other, Principal Component Analysis (PCA) [29], which transforms the data into orthogonal basis. Selection of homogeneous regions can be done e.g. with the Agglomerative Clustering [30], which groups similar samples into clusters. Other methods of clustering designed to differentiate between groups in data are, e.g. K-Means [30], an algorithm which for every point assigns the nearest centre of the formed clusters, or DBSCAN [31], which creates clusters based on the regions of high density. Using these algorithms, we obtain separate groups of similar spectra.

In many cases, the final requirement is not only estimation of the pigments present but labelling pixels that correspond to each pigment, so that image areas are highlighted. To classify the whole image, we can use classification algorithms. Trained on the labelled samples provided by the user, they can classify previously unseen data. One of the classification algorithms is Support Vector Machines (SVM) [32], trying to separate data using decision boundary with the largest margin. If one wants to obtain a non-linear boundary, the 'kernel trick' and corresponding special non-linear kernel functions may be applied. If transparency of the results is desired, random forests [33] or K-Nearest Neighbours (K-NN) [34] algorithms can be used. The former method works by separating data using the most differentiating features, while the latter assigns labels to data considering its closest neighbours. Another approach is to use Artificial Neural Networks (ANN) [35], which is a category

Pour citer cet article : B. Grabowski, et al., Automatic pigment identification from hyperspectral data, Journal of Cultural Heritage (2017), https://doi.org/10.1016/j.culher.2018.01.003