Light scattering spectroscopy combined with Principal Component Analysis for animal species identification in historical parchments

Fernandez Alvarez, Angel Martin; Mayer, Alexandre; Deparis, Olivier

Published in:
Novel Optical Materials and Applications, NOMA 2019

DOI:
10.1364/noma.2019.not2b.4

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Light Scattering Spectroscopy Combined with Principal Component Analysis for Animal Species Identification in Historical Parchments

Ángel Martín Fernández Alvarez[1], **Alexandre Mayer**[2], **Olivier Deparis**[1]

[1]*Department of Physics & Heritages, Transmissions, Inheritances (PaTHs) Institute, University of Namur, Belgium.*

[2]*Department of Physics & Namur Institute for Complex Systems (naXys), University of Namur, Belgium.*

olivier.deparis@unamur.be

Abstract: Light scattering spectroscopic data collected from historical parchments were processed according to different principal component analysis schemes, leading to reliable parchment optical fingerprints. This method enabled animal species identification without resorting to molecular level analysis.

OCIS codes: 100.0100, 150.0150, 200.0200, 290.0290, 300.0300.

1. Introduction

Parchments are amongst the most emblematic and fascinating objects of world cultural heritage. Made from processed animal dermis and used as writing support during centuries, their outstanding longevity yet requires appropriate conservation conditions. The animal species from which a parchment is made (typically calf, sheep, goat) is a kind of information that is not only useful to conservators but also to scholars [1]. Though visual inspection is used to get such information, it often fails to identify the species correctly, especially in old parchments. For a better reliability, analytical methods based on genetics or proteomics have been developed, with increasing progresses towards non-invasive sampling protocols [2]. It is noteworthy that these methods require time-consuming, non-invasive "sampling" of parchments and have additional costs related to protein (collagen) or DNA extraction and analysis. In this context, we have recently introduced a non-invasive (optical) method, easy and straightforward to implement, for species identification in parchments [3]. Remarkably, the method enables species identification without resorting to any molecular level analysis. Rather, it achieves reliable and robust identification solely from an "optical fingerprint" which is obtained by processing light scattering spectroscopic data with principal component analysis (PCA). The method was validated by performing proteomic analyses from parchment collagen and 100% matching of the results was obtained for 20 historical parchments [3]. The key role of PCA is to eliminate data redundancy and to operate dimensionality reduction. This mapping to a lower-dimensional (PC: principal components) space is done by a linear transformation which is applied to the original data matrix. Different PCA representations can be obtained depending on the way original data are organized in the matrix. The purpose of this talk to investigate the role of data organization schemes in the method.

2. Method

2.1. Light scattering spectroscopic measurements

The method relies on recording an "optical fingerprint" of the parchment, which originates from the sample-specific non-invasive interaction of light with the parchment's anisotropic, hierarchical collagen structure. The recorded optical spectra (reflection, transmission, absorption) carry complex yet redundant information on the parchment, which depends not only on the animal species, but also on the fabrication process and the aging history [3]. A double beam spectrophotometer equipped with a 150 mm diameter integrating sphere (Fig. 1) is used for recording these spectra on both sides of parchment. For modern parchments (known species), flesh and grain sides can be distinguished easily at the naked eye. For historical parchments (unknown species), on the other hand, the sole distinction is between recto and verso. Samples under test consisted of 21 modern parchments and 20 historical parchments (12th c.- 16th c.). Five types of measurements are considered: absorption (A), total reflection (R), total transmission (T), diffuse (specular excluded) reflection (R_d) and diffuse transmission (T_d). Spectral range extends from UV (200 nm) to NIR (2350 nm).

2.2. PCA data processing

The raw data are the $M = n \times p$ optical measurements (p measured quantities, n samples) taken at N wavelengths. Data can be organized in a matrix X according to different schemes. A first scheme (the one used in our previous

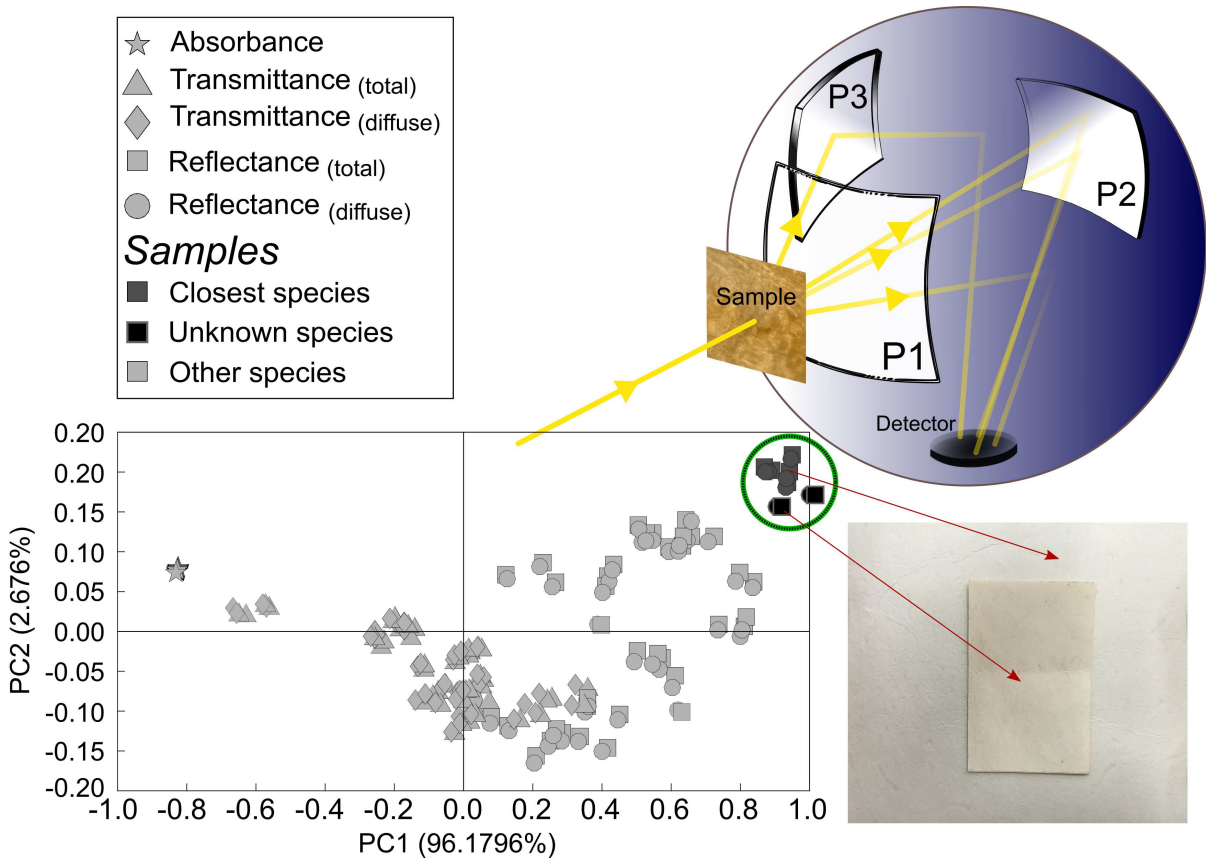

Fig. 1. Spectroscopic measurements on parchment using integrating sphere and PCA representation of the data.

work [3]) consists in gathering the measured spectra (A, R, R_d, T, T_d) along the lines of X, quantity after quantity, sample after sample, with the spectral values along the columns of X. The calculated correlation matrix $S_X = XX^t/(N-1)$ has therefore a straightforward physical meaning. In the related PCA representation, the so-called PCA scores gather into clusters associated to each type of measurements (Fig. 1). Proximity values are defined between the sample under test and samples of known species. The identification relies on the shortest Euclidian distance calculated in the complete PC space, for the most representative quantity along PC1 [3]. Another scheme, which will be investigated, consists in gathering measured spectra in a single vector that contains the complete information collected on a sample (recto or verso, grain or flesh sides). These super-vectors of size $D = p \times N$ obtained for the different samples are then gathered line by line in a matrix X' of size $n \times D$. At this point, PCA or various machine learning techniques can be applied in order to analyze the data and perform species identification.

3. Results

A typical PCA representation of data using the first scheme is shown in Fig. 1. Data cluster of both the unknown species and the closest species is highlighted by a circle. When calculated using PC1 and PC2, the proximity value is equal to 94.5%, and is associated with a modern sheep parchment sample (flesh side) [3]. More results will be presented at the conference and the effect data organization schemes will be discussed.

References

1. S. Fiddyment, *et al*, "Animal origin of 13th-century uterine vellum revealed using noninvasive peptide fingerprinting," PNAS **112**, 15066–15071 (2015).
2. M. Teasdale, *et al*, "The York Gospels: a 1000-year biological palimpsest," R. Soc. Open Science **4**, 170988 (2017).
3. A. M. Fernandez Alvarez, J. Bouhy, M. Dieu, C. Charles and O. Deparis, "Animal species identification in parchments by light," Sc. Rep. **9**, 1825 (2019).