Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 329 (2025) 125591

Contents lists available at ScienceDirect

Spectrochimica Acta Part A:
Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

Unravelling hidden text and figures in paper-based archival documents with micro-spatially offset Raman spectroscopy imaging

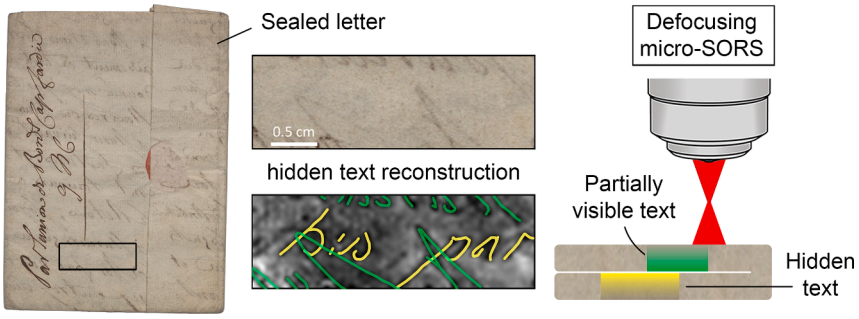
Marc Vermeulen [a,*], Claudia Conti [b], Pavel Matousek [c], Lora V. Angelova [a], Alessandra Botteon [b,*]

[a] The National Archives, Collection Care Department, Kew, Richmond TW9 4DU, UK
[b] Institute of Heritage Science, National Research Council (CNR ISPC), Via Cozzi 53, 20125 Milano, Italy
[c] Central Laser Facility, Research Complex at Harwell, STFC Rutherford Appleton Laboratory, Harwell, Oxford OX11 0QX, UK

HIGHLIGHTS

- Micro-SORS reveals non-invasively hidden text/images in historical documents.
- Raman signal, fluorescence and absorption aid in deciphering the obscured features.
- Chemometric analysis enhances visualization of hidden text despite planar challenges.
- Successfully reconstructed text in sealed 18th-century letters without opening them.
- Micro-SORS complements other imaging tools for cultural heritage conservation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Spatially offset Raman spectroscopy
micro-SORS
Subsurface imaging
Archives
Paper-based materials
Hidden text

ABSTRACT

The preservation of paper-based archival documents is crucial for safeguarding historical and cultural heritage. Some records possess visually inaccessible text or images because of previous conservation measures, their method of construction, or historic damage. Micro-spatially Offset Raman Spectroscopy (micro-SORS) has emerged as a promising method for probing below or through opaque material substrates non-invasively. This study explores the potential of micro-SORS to image hidden text and figures in paper-based archival documents, utilizing Raman signals, fluorescence emissions, and overall spectral intensity reflecting also sample absorption. We present case studies involving sealed letters and playing cards from historical collections, demonstrating the efficacy of micro-SORS in identifying pigments and deciphering hidden ink writings. Results show the successful mapping of vermilion pigment in playing cards and reconstruction of hidden iron gall ink text in sealed letters. Chemometric analysis further enhances the visualization of hidden text. Despite challenges such as the absence of Raman signal of the target materials, micro-SORS proves to be a valuable tool for accessing hidden information in paper-based artifacts, aiding in preservation efforts and historical research.

* Corresponding authors.
E-mail addresses: marc.vermeulen@nationalarchives.gov.uk (M. Vermeulen), alessandra.botteon@cnr.it (A. Botteon).

https://doi.org/10.1016/j.saa.2024.125591
Received 9 August 2024; Received in revised form 19 November 2024; Accepted 10 December 2024
Available online 11 December 2024

1. Introduction

The preservation of paper-based archival documents is an imperative endeavour, ensuring the safeguarding of our historical and cultural heritage for generations to come. Historical documents and books often pose analytical challenges as they possess non-visible text or images, which are inaccessible due to the nature of their construction, previous conservation measures or historical damages; this may include sealed documents or stuck pages due to deterioration of the materials. Over the years, numerous non-invasive techniques have been developed to assess the condition and content of these artifacts, for example through the use of multispectral imaging and ultraviolet radiation to enhance the poor readability of documents [1,2]. X-Ray Fluorescence (XRF) and Macro-XRF (MA-XRF) imaging, with its important depth penetration (a few micrometres down to several millimetres) have become techniques of choice to probe and image hidden layers in paintings, books and works on paper [3–7]. Reflectance imaging spectroscopy (RIS) has also been used, mostly using the short wavelength infrared range (1000–2500 nm) to highlight hidden features in paintings [8].

To date, access to hidden information in sealed, rolled or covered documents and books has often made use of MA-XRF,[6,9] X-ray-based micro-computed tomography (μ-CT) [10–18] and terahertz time-gated spectral imaging [19]. X-ray microtomography has also been applied to the study of letterlocking but limited interest was directed to reading the content of the letters [20].

However, most of these techniques lack the molecular specificity of Raman spectroscopy, which allows for the characterization of pigments [21,22] and synthetic organic pigments,[23–25] among others. However, while Raman spectroscopy has emerged as a powerful non-invasive tool for the analysis of various materials, including paper, it is limited by its inability to probe below the surface of the analysed material.

To merge the non-invasiveness and molecular specificity of Raman spectroscopy with sub-layer analysis of opaque surfaces, spatially offset Raman spectroscopy (SORS) was developed in the early 2000s [26,27]. While initially applied at the macroscale and often used for medical [28], pharmaceutical [29,30], or security [31–33] purposes, its variant, micro spatially offset Raman spectroscopy (micro-SORS) has garnered increasing attention as a breakthrough approach for non-invasive, high-resolution analysis of cultural heritage materials [34]. By adapting the principles of SORS to the micro-scale, micro-SORS offers the unique ability to interrogate subsurface layers of cultural heritage artifacts, providing valuable insights into their composition and degradation mechanisms. Despite the advantages offered by micro-SORS for the study of hidden layers in artistic production, its application has often, thus far, been limited to panel, canvas and mural paintings, [34–37] painted statues [38] and other decorated objects such as ceramics [34,39]. To date, except for the mock-up samples presented in [40,41], micro-SORS has not been applied to works on paper despite the auspicious results presented.

Micro-SORS, similarly to many other non-invasive analytical tools in the field of cultural heritage, i.e., XRF, reflectance spectroscopy and conventional Raman spectroscopy, saw its development also into an imaging technique for accessing hidden layers [39,40,42]. However, despite Raman spectroscopy being a well-suited technique for the non-invasive study of materials in works on paper [22,43–48], no application was found in using micro-SORS to recover hidden text in case studies. Part of the problem in applying micro-SORS to works on paper is related to the thinness of the text or coloured layers, which poses increased challenges regarding the acquisition of Raman photons at depth. However, recent research on mock-up samples by Botteon *et al.* showed that micro-SORS is well suited to image hidden texts and figures using fluorescence emission and overall spectral intensity, in addition to the Raman signal [41]. Contrast obtained using a difference in the overall spectral intensity and fluorescence profile rather than only monitoring any specific Raman bands were used to track the images within the hidden layer reflecting also underlying sample scattering and absorption features. This novel micro-SORS approach opens new prospects for the use of the technique for cultural heritage, with applications on paper-based objects such as books and documents, prone to high fluorescence emission of their substrates, thin colour or ink layers potentially lacking characteristic Raman signatures but exhibiting strong absorption patterns.

In response to the study by Botteon *et al.* [41], this paper further explores the burgeoning field of micro-SORS and its application in the realm of archival document analysis through the use of combined Raman signal, fluorescence emission and overall spectral intensity to image hidden text in historical documents. This paper delves into the capabilities, advantages, and challenges associated with micro-SORS probing and micro-SORS imaging as a cutting-edge analytical tool for archival collections through a series of case studies demonstrating its efficacy in identifying pigments through coloured paper pages and deciphering hidden inks in historical sealed letters. Micro-SORS imaging results were compared to the outcomes of conventional analytical techniques such as multiband imaging and MA-XRF mapping demonstrating technique's complementarity through the provision of chemical information.

2. Materials and methods

2.1. Letters

Three sealed letters from The National Archives Prize Papers Collection (HCA 32/156/5/UU4, HCA 32/156/5/UU9, and HCA 32/156/5/UU21, Fig. 1) were analysed in an attempt to access part of the written content without opening them. The letters, dated 1745, were seized from the ship *L'Union de Bordeaux* bound from Bordeaux to Le Cap, Saint Domingue (modern day Cap-Haïtien in Haiti) and Martinique. The thickness of each paper page used to create the letters (the letter is usually written on a single page that is folded and sealed with red wax) was measured using a calliper by accessing a single page where the paper would be slightly lifted on each side of the wax seal. Paper thicknesses were measured to be ~ 90 μm for letter UU4 and ~ 100 μm for both letter UU9 and letter UU21.

2.2. Playing cards

The reverse (design side) of a playing card from the Board of Trade (BT) Design Registers (BT43/420/48354, Fig. 2) was also analysed. According to the BT ledgers, the cards were deposited in the design register volume on December 20th, 1847, by James Boyd (Islington, London, UK). The playing cards, for which the back design was registered, had their numeral or Court faces glued onto the page of the board of trade volume, hence obstructing any observation of possible designs. The board of trade page on which the cards are glued were measured to have a thickness of ~ 180 μm. The paper used to produce the playing cards was ~ 460 μm in thickness. The measurements were done in the unglued corners.

2.3. X ray fluorescence (XRF)

2.3.1. Point analysis

Qualitative XRF point analysis was carried out using a Niton XL3t Ultra GOLDD + handheld XRF spectrometer. The built-in mining mode (Cu/Zn) was used, changing filters for main (Ag, 40 kV, 50 μA, 60 s acquisition time) and light elements (no filter, 15 kV, 100 μA, 60 s acquisition time), with a 3 mm spot size. Niton Data Transfer Version NDT_REL_8.2 Software was used for data processing.

2.3.2. MA-XRF

MA-XRF was carried out with a Bruker Elio energy dispersive X-ray fluorescence analyser, with a high-resolution, large area Silicon Drift Detector with 130 eV at manganese (Mn) Kα with 10 kcps input photon

Fig. 1. Visible light images of the front and back of the 3 sealed letters investigated: (a) HCA 32/156/5/UU4, (b) HCA 32/156/5/UU9, and (c) HCA 32/156/5/UU21. Photo credit: The National Archives.

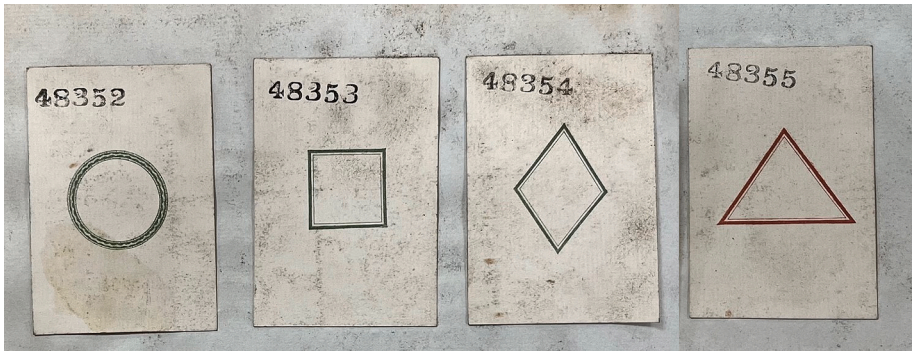


Fig. 2. Visible light image of playing cards glued onto the blue page of the Board of Trade volume showing the four basic geometric back design patterns being trademarked. From left to right: BT43/420/48352, BT43/420/48353, BT43/420/48354, BT43/420/48355. Photo credit: The National Archives.

rate (high resolution mode), 170 eV at Mn Kα with 200 kcps input photon rate (fast mode). The system is equipped with changeable filters, and a rhodium (Rh) transmission target with 50 kV maximum voltage and 4 W maximum power. The size of the analysed spot is ca. 1 mm in diameter. Elemental 2D mapping of the surface was achieved through automatic XY raster scanning with 1 mm spot size and 2.5 mm step size (exploratory scan), with a dwell time of 1 s/pixel, acquiring a scanned area of 65 × 85 mm (24 row and 36 columns) for the playing card and with 1 mm spot size and 0.5 mm step size (oversampling), with a dwell time of 1 s/pixel, acquiring a scanned area of 20 × 10 mm (40 row and 20 columns) for the sealed letter. The tube was operated at 50 kV and 80 μA. The maps were acquired using the Elio software and elaborated using PyMCA software suite [49].

2.4. Micro Raman spectrometer

Spectra were acquired with a HORIBA XploRA Plus Confocal Raman spectrometer, equipped with an Olympus 20x objective and a charge-coupled device (CCD) detector. The Raman spectrometer is equipped with an XY sample stage and a software-enabled Z focus control, permitting both automated mapping and micro-SORS measurements in defocusing mode [50]. The spectra were recorded with a near-infrared 785 nm laser in the spectral range 100–1300 cm^{-1} (playing card, area 1), 100–2300 cm^{-1} (playing card, area 2) and 500–1600 cm^{-1} (sealed letters) at 10 % of the maximum laser power (5 mW) when working in focus and 100 % of the laser power (50 mW) when working in defocusing mode. A 1200 gr/mm grating, a 50 μm slit and 500 μm pinhole were used for most accumulations, except for area 2 of the playing card, for which a 600 gr/mm grating was used to capture the extended range in a single accumulation.

In order to understand the optimum defocusing distance prior to imaging, the relative band intensity change of the pigments/paper was studied by acquiring sequences of Raman spectra at the image level (focus) and every 100 μm until a 1 mm defocusing was reached. The intensity ratio of Raman bands between target and top layers was then plotted against the defocusing distance to visualize the optimum defocusing distance, i.e., the sample distance at which the maximum ratio value is obtained. In the case of the letters, due to the overall lack of Raman signal for the ink, the optimized parameter was obtained by mapping small areas at different defocusing distances and checking the quality of reconstructed letters. In all the cases, the optimum distance was measured to be ~ 900 μm, which was then used for mapping on the playing cards and sealed letters.

The areas imaged, step size and acquisition times for each of the analysed areas were selected as follows, providing sufficiently clear images.

I. Playing card, area 1: 15.4 x 12.8 mm, 800 μm, and 2x8 sec, total acquisition time: 1h35min.
II. Playing card, area 2: 7.7 x 7.3 mm, 800 μm, and 2x8 sec, total acquisition time: 31 min.
III. Sealed letter UU4: 27.4 x 8.7 mm, 400 μm, and 1x12 sec, total acquisition time: 5h35min.
IV. Sealed letter UU9, area 1: 15.2 x 9.3 mm, 400 μm, and 1x7 sec, total acquisition time: 1h56min.
V. Sealed letter UU9, area 2: 34.8 x 7.2 mm, 300 μm, and 1x7 sec, total acquisition time: 5h14min.
VI. Sealed letter UU21: 8.6 x 4 mm, 250 μm, and 2x7 sec, total acquisition time: 3h21min.

All the spectra were acquired using the HORIBA LabSpec 6 Spectroscopy Suite Software and the built-in MVA *Plus* software add-on was used for unsupervised multivariate processing of the acquired data. Principal component analysis (PCA), hierarchical clustering analysis (HCA), multivariate curve resolution (MCR) and K-mean clustering (KMC), were selected to process the acquired data. Details about each data processing approach are given below:

- PCA: decomposition method based on the calculation of eigenspectra of the mapped area, 3 components, data normalisation per area, no baseline correction, 32 iterations.
- HCA: clustering method based on calculation of similarities between each spectrum and the average spectrum of each class, 3 components, data normalisation per area, no baseline correction, 1 iteration.
- MCR: decomposition method with non-negativity constraint resulting in reference spectra that can be directly used for identification, 3 components, data normalisation per area, no baseline correction, 32 iterations.
- KMC: clustering method based on calculation of similarities between each spectrum and the average spectrum of each class, 3 components, data normalisation per area, no baseline correction, 32 iterations.

2.5. Multiband imaging

A Microbox X150 multiband imaging system was used to image the documents: a full-frame 150-megapixel colour camera (CMOS), two broadband LED illumination units (UV, VIS, IR) and software filter for multiband analysis (instead of mechanical lens filters). The pixel values of the sensor are evaluated on an analogue basis as native voltage values. Only the values of the desired spectral range were analysed and combined into one image.

Visible, transmitted light, as well as infrared reflected (IRR) and infrared false colour (IRFC) were acquired and compared. Fiji open-source software was used for image processing.

3. Results and discussion

Prior to analysis with micro-SORS, the material components of the sealed letter and playing cards were assessed. These include the paper, the ink, and potential pigments/colourants used in the paper.

3.1. Paper, ink, and pigments/colorants

XRF of the ink from the various letters yielded a slightly increased signal for iron (Fe) compared to the signal observed for the paper support (Fig. S1a–c). Based on the time period (mid-18th century) and the dark brown colour of the writing, this is likely to indicate the use of iron gall ink. Other elements including potassium (K), sulphur (S) calcium (Ca), silicon (Si) and phosphorous (P), were also observed in both paper and inks. Based on the increased signals for K and S in the ink of letters UU4 and UU9 (Fig. S1a and Fig. S1b), these elements are likely associated with the iron gall ink [51] while the limited increase in Fe and S in UU21 (Fig. S1c) may indicate a different iron gall ink composition. Multiband imaging in the infrared range (Fig. S2) shows variations in the absorption of the ink. However, the lack of additional spectral features between paper substrate and pure iron gall ink references in the infrared range (Fig. S3) tends to indicate that the variations observed in the case of these letters may come from a slight difference in ink compositions, such as the addition of carbon black, a material known for absorbing in the infrared. This is illustrated by the contemporary letter numbering carried out graphite, which appears dark in all three instances (Fig. S2). If carbon black would have been mixed with iron gall ink, one would expect to see the ink absorbing more in the infrared than if used alone. This may have been the case and may explain the variations observed, with HCA 32/156/5/UU9 possibly having a larger carbon content than HCA 32/156/5/UU4 or HCA 32/156/5/UU21, though the carbon remains undetected with XRF analysis.

The Board of Trade page on which the cards are glued contains cobalt (Co), arsenic (As), bismuth (Bi), potassium (K), and silicon (Si), indicative of the use of smalt (Fig. S1d). Smalt was a common pigment used in the production of blue paper in the 18th and 19th century but was later replaced by Prussian blue and synthetic ultramarine in more modern blue paper [52]. Other elements such as calcium (Ca), iron (Fe) and copper (Cu) are likely associated with the paper production process rather than the blue colourant. The point XRF of the design from the back of the cards (Fig. S4) highlighted the presence of a mercury (Hg)-based red, likely to be vermilion in the red design (Fig. S4a) and lead (Pb)-, chromium (Cr)-, and iron (Fe)-based green (Fig. S4b), likely to be a mixture of chrome yellow and Prussian blue, also often referred to as chrome green.

As common practice in cultural heritage studies, the letters and the playing cards were partially measured using XRF imaging. The elemental maps obtained are presented in Fig. 3 and Fig. S5 for the playing cards (BT43/420/48354) and a section of the sealed letter HCA 32/156/5/UU4, respectively.

For HCA 32/156/5/UU4, the elemental map for Fe suggests the use of an iron gall ink. Presence of Cu may indicate an ink production using copper sulphates [53]. However, the actual writing cannot be deciphered despite oversampling with a 500 μm step size. The presence of Fe throughout the scan also suggests the use of Fe in the paper itself, which may also prevent adequate visualisation of the iron-containing text. Instrumentation offering better lateral resolution may have yielded better results but was not available at the time of analysis. Cu was not observed in the point analysis because it may also not be evenly distributed across the text.

The imaging of the playing card BT43/420/48354 yielded satisfactory elemental maps. The elements highlighted include Fe, Pb, Cr, Hg and Ca. Contrary to the MA-XRF scanning of the sealed letters, the elemental maps obtained on the playing card allow for the visualization of the hidden design. Here, they reveal a standard single-ended English pattern court card of the queen of diamonds. All highlighted elements can be associated with a variety of pigments which remain unknown due to the lack of molecular specificity of the technique and their presence in many inorganic pigments. Only Hg in the red areas of the design can univocally be associated with the use of vermilion [54].

3.2. Micro-SORS imaging

3.2.1. Letters

The conventional Raman spectra collected on the inks used in the top layer of the three letters showed an intense fluorescence, obscuring all the Raman bands that could be used to track the ink distribution in the sublayers (Fig. S6). In this case, the reconstruction of the spatial distribution of hidden text, likely created using iron gall ink, was still achievable using difference in the overall spectral intensity and fluorescence profiles, as demonstrated earlier by Botteon *et al.* on mock-up samples [41]. Applying this approach to a section of letter HCA 32/

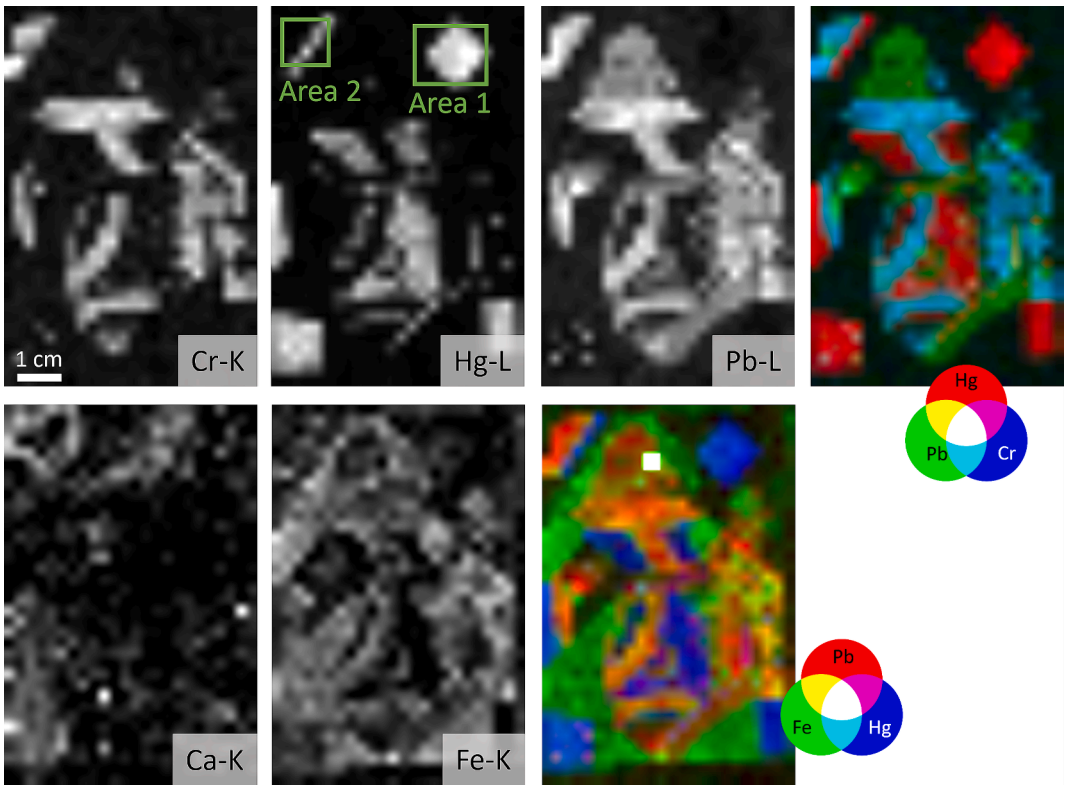


Fig. 3. Micro-XRF imaging elemental maps obtained for the playing card BT43/420/48354. The green squares in the Hg map indicate the two areas analysed with micro-SORS.

156/5/UU4 allowed to recreate words (Fig. 4) that were present not only on the reverse of the page analysed (partially visible in Fig. 4a), but also on an internal page. Here, the reconstruction of the hidden word was independent of the Raman shift used, as suggested by the same map observed for the three wavenumbers selected, 535, 895, and 1480 cm^{-1} (Fig. 4b,c and d, respectively). The darker appearance of the words compared to the brighter surrounding areas illustrates well the variations of spectral overall intensity, with the iron gall ink words (darker) being more absorbing when compared to the cream colour paper, phenomenon already discussed by Botteon *et al.* [41]. In-focus analysis, subsets of SORS-scanned areas as suggested by green dashed lines around the inserts in Fig. 4, shows that the visualization of the hidden words would not be possible in conventional Raman modalities, with and without changes in brightness and contrast sometimes helping with enhancing slight variations in Raman intensities. This illustrates the great value of micro-SORS for revealing the content of sealed letters without opening them, despite the lack of molecular information often associated with this analytical method.

To avoid any biased selection of the spectral area of interest, various chemometric approaches were used to reconstruct hidden words based on the Raman signal over the complete spectral range. Results are presented on Fig. 5. In this case, both multivariate curve resolution (MCR) and principal component analysis (PCA) allowed to visualise hidden text, identical to what was observed using the overall spectra intensity at selected wavenumbers (Fig. 4-b,c,d). For similar results, the straightforwardness of the overall intensity at selected wavenumber approach would appear as the easiest approach to reconstruct the hidden text. However, the multivariate analysis approach provides additional information through the associated loading score plots, showing that, in this case, spectral intensity and overall fluorescence profile is the best parameter to reconstruct the text, based in the overall lack of Raman bands and small variations across the entire measured area. Areas of text appear darker against a brighter background corresponding to the more intense signal of the paper background (Fig. 5-MCR-1, MCR-2 and PCA-1). Small but distinctive changes in the background signals observed in the 700–1400 cm^{-1} range (Fig. S6) between the areas of text (ink) and paper allow for the positive imaging of the areas of text with PCA (Fig. 5-PCA-2), as suggested by the positive loadings in this particular range. Areas for which the signal mostly correspond to the paper signal (Raman bands included between 1000–1200 and 1300–1400 cm^{-1}) does not provide any information on the hidden layer due to its presence in both the top and hidden layers (Fig. 5-PCA-3).

Both the overall spectral intensity and multivariate analysis methods have also been successfully applied to other sealed letters. They have highlighted a complex partial signature in HCA 32/156/5/UU9 (Fig. S7) and have indicated the presence of some text in HCA 32/156/5/UU21 (Fig. S8). However, for the latter, the use of the overall spectral intensity at selected wavenumber did not allow for the elucidation of the hidden letters while the multivariate approaches allowed for the identification of letters that seem to be "g" and "a". For all multivariate approaches, based on the associated loading plots, except for the 1100 cm^{-1} Raman band associated with the paper (Fig. 5-PCA-3, Fig. S7-PCA-2, Fig. S8-PCA-3), all the letter extractions appear to be based on small overall changes in intensity and background fluorescence signal.

Aside from the moderately long scanning times, one of the main limitations of the technique remains the necessity for planarity across the imaged area. When extending the area scanned for HCA 32/156/5/UU9 beyond the analysed area presented in Fig. S7, the technique was not as successful at reconstructing hidden text (Fig. S9) and was not even able to reconstruct the partial signature observed before. This is due to the area being slightly raised on the right hand-side, consequently changing the overall defocusing and the optimum distance to enhance the contrast between the upper and hidden layers. One possible approach to bypass this technical problem would be to work with an automatic focus at each measured point. This would, however, drastically increase the overall measuring time. A second option would be to employ the point-like micro-SORS approach as described by Botteon *et al.* [41]; however, this technique is not yet available within The National

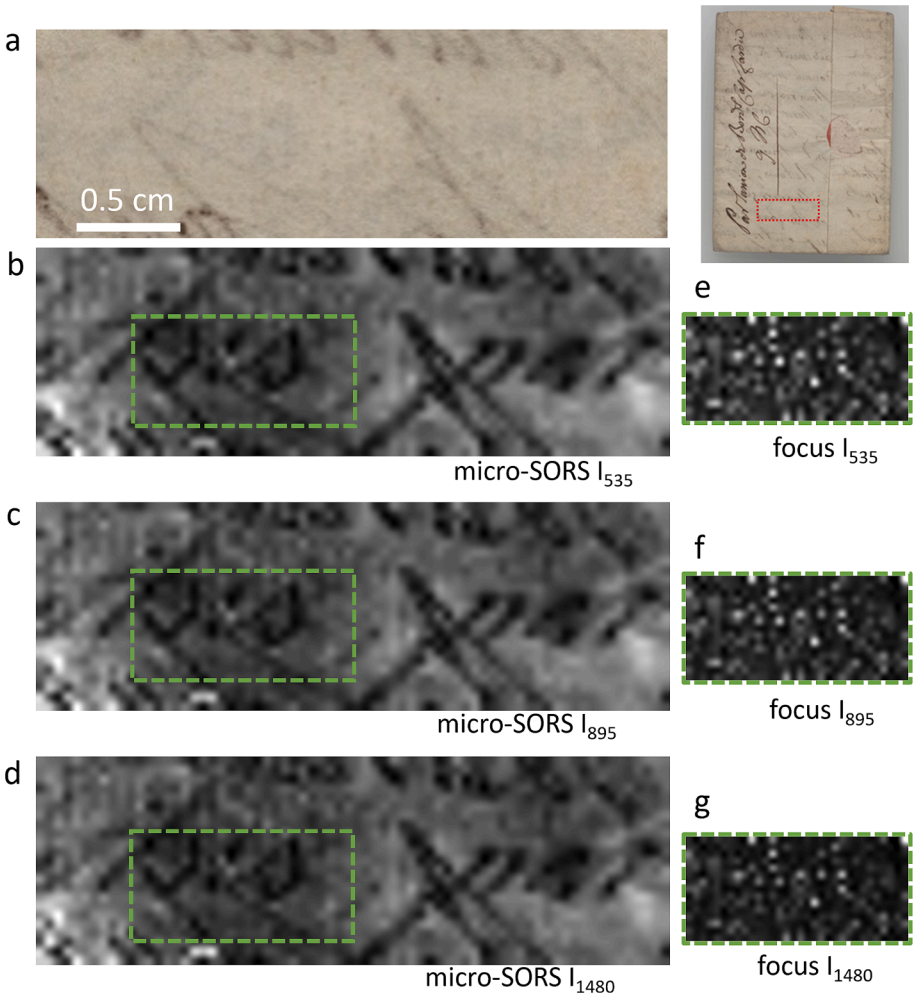


Fig. 4. (a) Visible light image of the measured section of sealed letter HCA 32/156/5/UU4 and (b, c and d) micro-SORS imaging using the overall spectra intensity at (b) 535 cm^{-1}, (c) 895 cm^{-1} and (d) 1480 cm^{-1}. The dotted areas in b, c and d correspond to the mapping of selected areas measured in focus mode respectively presented in (e, f, and g). The latest images were mapped with spectra intensity at 535 cm^{-1}, 895 cm^{-1} and 1480 cm^{-1}, respectively.

Archives laboratory.

Nonetheless, even if spatially limited, the access to hidden information is an appropriate first step in a long decision-making process often associated with opening historical sealed letters. The identification of a signature or some key words may be instrumental as the required proof whether these invaluable objects should be opened for access to their complete content.

3.2.2. Playing card

Based on previous XRF analyses (point and imaging), most of the pigments used to create the hidden design of the playing card could be inferred, with vermilion used in the red areas, chrome yellow likely to be used in the yellow areas, lead white for the white areas, and Prussian blue likely used in the blue areas. Nevertheless, this playing card was used to better understand the potential and limitations of micro-SORS for the characterisation and spatial imaging of hidden pigmented layers in works on paper, as, until now, most micro-SORS imaging on works on paper has been carried on mock-up samples.

The use of the Raman specific bands at 253 and 343 cm^{-1} (as visible in extracted spectra presented in Fig. S10) allowed to confirm the use of vermilion and to visualize its distribution in both areas 1 and 2 (Fig. 6a-c). For area 2, only the band at 253 cm^{-1} was used to confirm and visualize the use of vermilion due to the lack of clear band at 343 cm^{-1} (Fig. S10).

When mapping the distribution of the Raman bands of interest, one can decide to either use a local baseline to subtract the background or not for the relevant signal. In the case of areas 1 and 2, both approaches allow for the visualization of the pigmented area (Fig. 6a-f). However, for the Raman bands with the strongest signals (bands at 253 cm^{-1}), the use of a local baseline offered a better contrast (Fig. 6a vs Fig. 6d) and appeared the only way to back-map the band in the case of high fluorescence, likely due to the paper (Fig. 6c vs Fig. 6f; Fig. S10). For area 1, the use of the local baseline (Fig. 6a-b) also highlights variation within the design itself, with a stronger signal on the right-hand side. This may be associated with variation in pigment thickness or in sample flatness, which is difficult to assess non-invasively. Similar features can be observed when using the various multivariate analysis approaches, especially K-Means clustering or multivariate curve resolution (Fig. S11). These methods all allowed for the mapping of the red diamond based on the vermilion spectrum, according to the associated loading plots (HCA-1&2, KMC-1, MCR-1&2 and PC-1 in Fig. S11). However, the techniques also allowed to recreate the diamond-like shape using the Raman bands of cellulose at 1095, and 1121 cm^{-1} as main component of the spectrum (PC-3 in Fig. S11), in the absence of any of the vermilion bands. In this case, areas of paper appear brighter and the diamond area darker. These observations parallel those made by Botteon *et al.* in their recent article [41].

While not successful in reconstructing the vermilion band in area 2 (Fig. 6f), the back mapping of the band intensity without a local baseline was successful for both the 253 and 343 cm^{-1} Raman bands in area 1

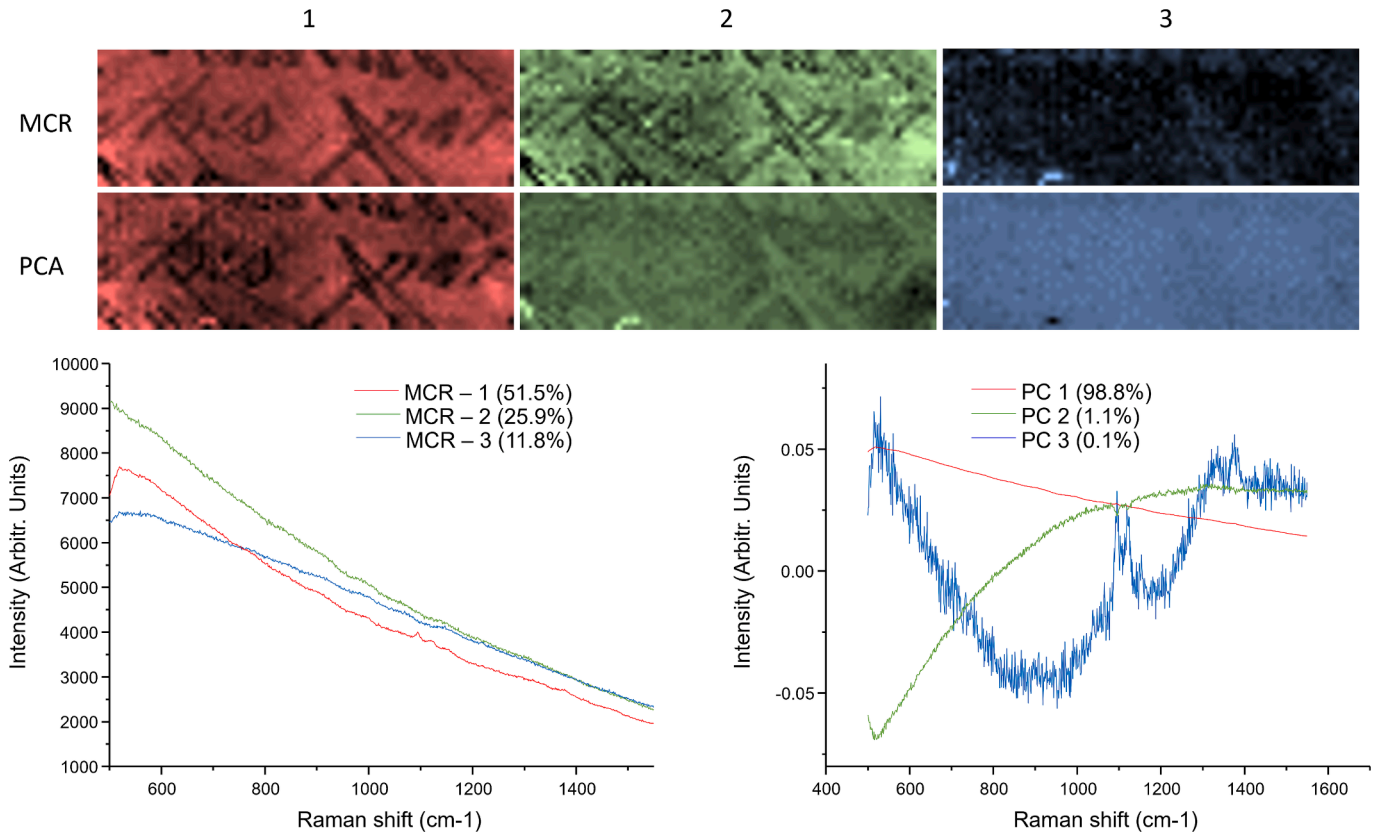


Fig. 5. Micro-SORS images (score plots) obtained through the application of multivariate analysis methods: multivariate curve resolution (MCR) and principal component analysis (PCA) and their associated loading plots (components 1, 2 and 3).

(Fig. 6d-e). While the contrast is not as good as that observed when using the local baseline (especially for the stronger band at 253 cm^{-1}), this approach shows that the overall fluorescence from the paper substrate can be an issue and mask the weaker Raman band intensities, hindering the identification and/or visualization of a pigment distribution. This issue can be bypassed using the whole spectrum with all characteristic bands (rather than a single band) as illustrated by the multivariate approach results, especially HCA-1&2, KMC-1, MCR-1&2 and PCA-1 (Fig. S11), all using the full spectrum of vermilion to visualise its distribution.

The mapping in focus mode did not yield satisfactory results (Fig. 6g) showing that micro-SORS imaging, with or without a local baseline, was the preferred method to non-invasively identify and locate the distribution of the pigment, despite a 180 µm thick smalt-containing paper layer covering the coloured area. While identifying vermilion using a Raman-based technique is not surprising due to its good Raman scattering properties, its identification with micro-SORS through an almost 200 µm thick layer is very encouraging and extends further the capabilities of the technique proving its feasibility for paper-based materials. One could also expect that results may be even better when using full micro-SORS, rather than defocussing micro-SORS, based on its higher surface/subsurface contrast as demonstrated by Botteon *et al.* [41].

According to the exploratory MA-XRF scan presented in Fig. 3, the crown area (area 2) should also contain other pigments, potentially chrome yellow ($PbCrO_4$) and Prussian blue ($Fe[Fe^{3+}Fe^{2+}(CN)_6]_3$), both good Raman scatters with well-defined and well-known Raman bands [21]. However, the mapping of the band intensity at 841 cm^{-1} (chrome yellow, Fig. 6h) and 2154 cm^{-1} (Prussian blue, Fig. 6i) did not yield any maps that would indicate the use of such pigments, independently of the use of local baseline, which provided good results in the case of vermilion. Additional micro-SORS analysis in the areas of the crown, where these pigments were expected, did not yield any Raman bands that could be associated with the pigments, despite both pigments having been identified through the use of micro-SORS in painted surfaces [38]. The reason behind this inability of the technique to identify (and subsequently map) both chrome yellow and Prussian blue in this case is difficult to explain. There could be several reasons behind this, including i) a substrate layer (paper) being too thick to allow the identification of the pigments, ii) a weak Raman signal being covered by the overwhelming fluorescence from the paper layer, or iii) a matrix absorption effect. While all of these are worthy of further investigation, they are beyond the scope of this article. Despite not being able to identify and visualize their distribution within the selected area of the playing card using defocusing micro-SORS, the use of point-like micro-SORS could address the limitation observed here as it provides a higher degree of surface/subsurface contrast [41].

4. Conclusions

Micro-Spatially Offset Raman Spectroscopy presents a groundbreaking approach for the non-invasive analysis of paper-based archival documents. Case studies involving 18th century sealed letters and 19th century playing cards revealed the potential of micro-SORS to provide valuable insight into the composition and hidden content of paper-based materials. The use of distinct Raman bands allowed for molecular identification of the materials in the playing cards, though it was not possible to identify the inks in the sealed letters. Nevertheless, despite the inability to molecularly characterize certain materials, micro-SORS imaging of the sealed letters enabled the reconstruction of text obscured by overlapping paper layers using alternative phenomena such as photons absorption and fluorescence emission. Chemometric analysis further enhanced the visualization of hidden text, despite challenges such as surface planarity. The findings of this study highlight the immense potential of micro-SORS as a non-destructive and informative

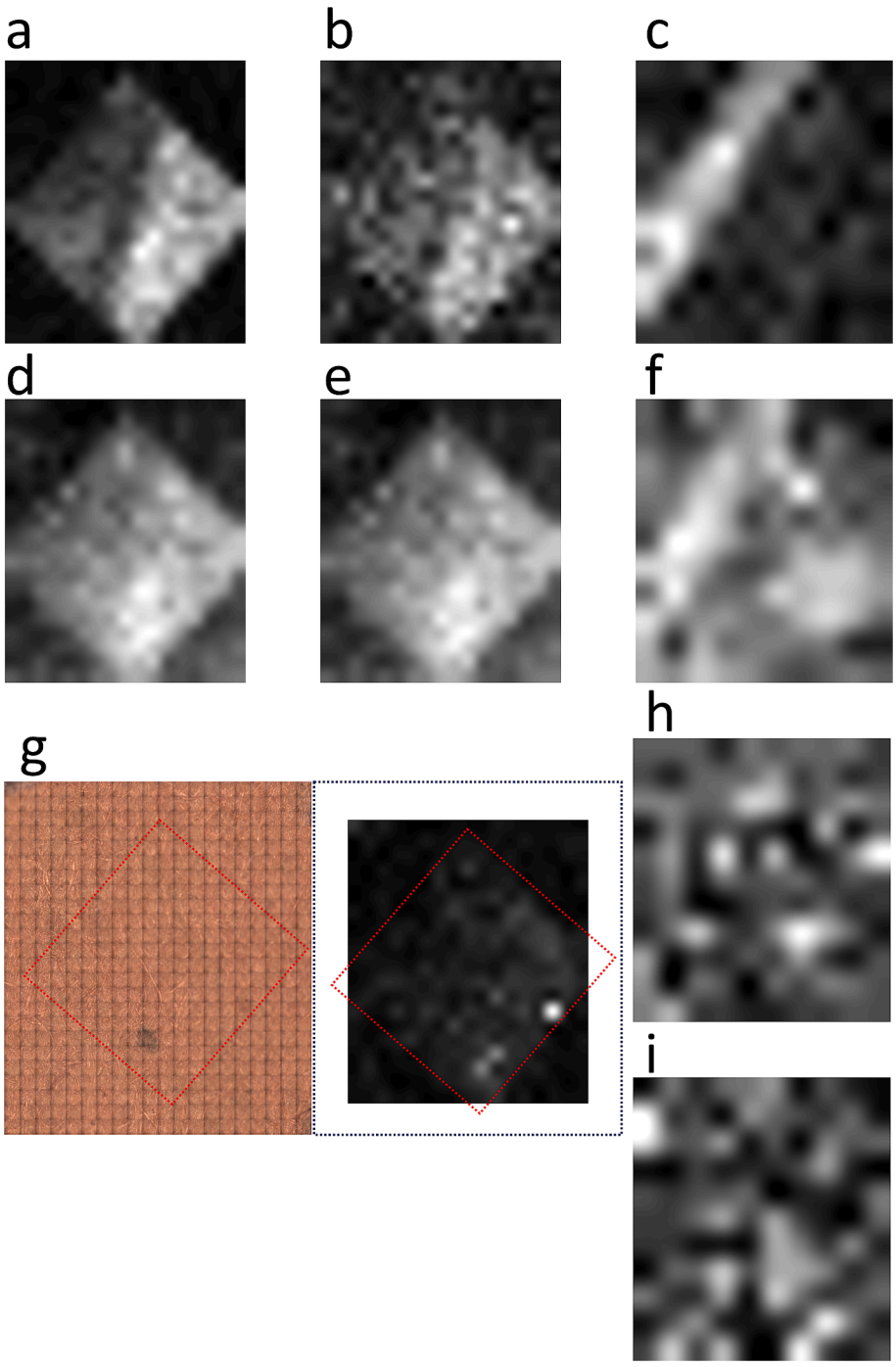


Fig. 6. Intensity maps obtained at characteristic Raman shifts for the two areas analysed: (a) area 1, intensity at 253 cm^{-1}, local baseline, (b) area 1, intensity at 343 cm^{-1}, local baseline, (c) area 2, intensity at 253 cm^{-1}, local baseline, (d) area 1, intensity at 253 cm^{-1}, no local baseline, (e) area 1, intensity at 343 cm^{-1}, no local baseline, (f) area 2, intensity at 253 cm^{-1}, no local baseline, (g) area 1, optical image with indication (red dotted line) of the red diamond location, and intensity at 253 cm^{-1} in focus mode, local baseline, (h) area 2, intensity at 840 cm^{-1}, local baseline, and (i) intensity at 2154 cm^{-1}, local baseline.

technique for accessing hidden information in cultural heritage objects, contributing to preservation efforts and historical research. The consistent reconstruction of sections of hidden text in sealed letters proves that this non-invasive approach is a good alternative to the opening of letters to access their content. The time required for the analysis, which is partially a consequence of the necessity for high spatial resolution to recover hidden, hand-written text, as well as the necessity for plane extended areas, represent the main limitations of the technique for a large-scale use. However, the technique can be considered as an additional tool to non-invasively explore and assess the content of a letter prior to its potential opening or consideration of deploying other imaging tools such as X-Ray tomography, terahertz imaging or even invasive analyses. Further advancements in instrumentation and methodology are warranted to address the technical limitations and expand the applicability of micro-SORS in the field of heritage conservation and analysis, but this study represents a breakthrough in the application of micro-SORS to cultural heritage studies.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Dr Oliver Finnegan, Prize Papers Records Specialist and Olivia Gecseg, Visual Collections Records Specialist, both from the Collections Expertise and Engagement Department from at The National Archives for access and background information on the sealed letters from the Prize Papers Collection and from the playing cards in the Board of Trade Design Registers. The authors would also like to thank Michele Gironda, Dr Roald Tagle and Mareike Gerken, from Bruker Nano Analytics, for access to the Elio MA-XRF. This work was supported by the Arts and Humanities Research Council [grant number AH/V012010/1]. Finally, the authors would like to acknowledge the Grants and Funding Office from The National Archives for providing research funding through their Strategic Research Fund scheme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2024.125591.

Data availability

Data will be made available on request.

References

[1] C. Duffy, Multi-spectral Imaging at the British Library, in: 2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018), 2018, pp. 1-4.
[2] L. Pereira Pardo, P. Dryburgh, E. Biggs, M. Vermeulen, P. Crooks, A. Gibson, M. Fort, C. Vlachou-Mogire, M. Bertasa, J.R. Gilchrist, J. Danskin, Advanced imaging to recover illegible text in historic documents. The challenge of past chemical treatments for ink enhancement, J. Cult. Herit. 68 (2024) 342–353.
[3] K. Trentelman, K. Janssens, G. Van der Snickt, Y. Szafran, A.T. Woollett, J. Dik, Rembrandt's an old man in military costume: the underlying image re-examined, Appl. Phys. A, Mater. Sci. Processing 121 (2015) 801–811.
[4] G. Van der Snickt, S. Legrand, I. Slama, E. Van Zuien, G. Gruber, K. Van der Stighelen, L. Klaassen, E. Oberthaler, K. Janssens, In situ macro X-ray fluorescence (MA-XRF) scanning as a non-invasive tool to probe for subsurface modifications in paintings by P.P. Rubens, Microchem. J. 138 (2018) 238–245.
[5] G. Van der Snickt, A. Martins, J. Delaney, K. Janssens, J. Zeibel, M. Duffy, C. McGlinchey, B. Van Driel, J. Dik, Exploring a hidden painting below the surface of rené magritte's le portrait, Appl. Spectrosc. 70 (2016) 57–67.
[6] J.R. Duivenvoorden, A. Käyhkö, E. Kwakkel, J. Dik, Hidden library: visualizing fragments of medieval manuscripts in early-modern bookbindings with mobile macro-XRF scanner, Heritage Sci. 5 (2017) 6.
[7] S. Pessanha, M. Costa, J.M. Sampaio, M.L. Carvalho, Revealing the hidden preliminary version of Eça de Queiroz "The Illustrious House of Ramires" using X-ray micro-analysis, Nucl. Instrum. Methods Phys. Res., Sect. B 371 (2016) 396–400.
[8] H. Barrett, What was Vermeer hiding in The Milkmaid?, in: The Telegraph, 2023.
[9] E. Pouyet, S. Devine, T. Grafakos, R. Kieckhefer, J. Salvant, L. Smieska, A. Woll, A. Katsaggelos, O. Cossairt, M. Walton, Revealing the biography of a hidden medieval manuscript using synchrotron and conventional imaging techniques, Anal. Chim. Acta 982 (2017) 20–30.
[10] D. Stromer, V. Christlein, C. Martindale, P. Zippert, E. Haltenberger, T. Hausotte, A. Maier, Browsing through sealed historical manuscripts by using 3-D computed tomography with low-brilliance X-ray sources, Sci Rep 8 (2018) 15335.
[11] C.S. Parker, S. Parsons, J. Bandy, C. Chapman, F. Coppens, W.B. Seales, From invisibility to readability: Recovering the ink of Herculaneum, PLoS One 14 (2019) e0215775.
[12] J.E. Ensley, K.H. Tachau, S.A. Walsh, H. Zhang, G. Simon, L. Moser, J. Atha, P. Dilley, E.A. Hoffman, M. Sonka, Using computed tomography to recover hidden medieval fragments beneath early modern leather bindings, first results, Heritage Science 11 (2023) 82.
[13] I. Bukreeva, M. Alessandrelli, V. Moso, R. Graziano, A. Cedola, Investigating Herculaneum papyri: An innovative 3D approach for the virtual unfolding of the rolls, arXiv (Cornell University), 2017.
[14] I. Bukreeva, A. Mittone, A. Bravin, G. Festa, M. Alessandrelli, P. Coan, V. Formoso, R.G. Agostino, M. Giocondo, F. Ciuchi, M. Fratini, L. Massimi, A. Lamarra, C. Andreani, R. Bartolino, G. Gigli, G. Ranocchia, A. Cedola, Virtual unrolling and deciphering of Herculaneum papyri by X-ray phase-contrast tomography, Sci Rep 6 (2016) 27227.
[15] D. Allegra, E. Ciliberto, P. Ciliberto, F.L.M. Milotta, G. Petrillo, F. Stanco, C. Trombato, Virtual unrolling using X-ray computed tomography, in, EURASIP, pp. 2864-2868.
[16] O. Samko, Y.-K. Lai, D. Marshall, P.L. Rosin, Virtual unrolling and information recovery from scanned scrolled historical documents, Pattern Recogn. 47 (2014) 248–259.
[17] D. Baum, N. Lindow, H.-C. Hege, V. Lepper, T. Siopi, F. Kutz, K. Mahlow, H.-E. Mahnke, Revealing hidden text in rolled and folded papyri, Applied physics, A, Materials Science & Processing 123 (2017) 1–7.
[18] V. Mocella, E. Brun, C. Ferrero, D. Delattre, Revealing letters in rolled Herculaneum papyri by X-ray phase-contrast imaging, Nat. Commun. 6 (2015) 5895.
[19] A. Redo-Sanchez, B. Heshmat, A. Aghasi, S. Naqvi, M. Zhang, J. Romberg, R. Raskar, Terahertz time-gated spectral imaging for content extraction through layered structures, Nat. Commun. 7 (2016) 12665.
[20] J. Dambrogio, A. Ghassaei, D.S. Smith, H. Jackson, M.L. Demaine, G. Davis, D. Mills, R. Ahrendt, N. Akkerman, D. van der Linden, E.D. Demaine, Unlocking history through automated virtual unfolding of sealed documents imaged by X-ray microtomography, Nat. Commun. 12 (2021) 1184.
[21] L. Burgio, R.J.H. Clark, Library of FT-Raman spectra of pigments, minerals, pigment media and varnishes, and supplement to existing library of Raman spectra of pigments with visible excitation, Spectrochim. Acta A Mol. Biomol. Spectrosc. 57 (2001) 1491–1521.
[22] M. Vermeulen, M. Leona, Evidence of early amorphous arsenic sulfide production and use in Edo period Japanese woodblock prints by Hokusai and Kunisada, Heritage Science 7 (2019).
[23] F. Schulte, K.-W. Brzezinka, K. Lutzenberger, H. Stege, U. Panne, Raman spectroscopy of synthetic organic pigments used in 20th century works of art: Raman spectroscopy of synthetic organic pigments, J. Raman Spectrosc. 39 (2008) 1455–1463.
[24] S.A. Centeno, V.L. Buisan, P. Ropret, Raman study of synthetic organic pigments and dyes in early lithographic inks (1890–1920), J. Raman Spectrosc. 37 (2006) 1111–1118.
[25] R. Pause, I.D. van der Werf, K.J. van den Berg, Identification of pre-1950 synthetic organic pigments in artists' paints. a non-invasive approach using handheld raman spectroscopy, Heritage 4 (2021) 1348–1365.
[26] P. Matousek, Inverse spatially offset raman spectroscopy for deep noninvasive probing of turbid media, Appl. Spectrosc. 60 (2006) 1341–1347.
[27] P. Matousek, M.D. Morris, N. Everall, I.P. Clark, M. Towrie, E. Draper, A. Goodship, A.W. Parker, Numerical simulations of subsurface probing in diffusely scattering media using spatially offset raman spectroscopy, Appl. Spectrosc. 59 (2005) 1485–1492.
[28] K. Buckley, J.G. Kerns, P.D. Gikas, H.L. Birch, J. Vinton, R. Keen, A.W. Parker, P. Matousek, A.E. Goodship, Measurement of abnormal bone composition in vivo using noninvasive Raman spectroscopy, IBMS BoneKey 11 (2014) 602.
[29] P. Matousek, A.W. Parker, Bulk raman analysis of pharmaceutical tablets, Appl. Spectrosc. 60 (2006) 1353–1357.
[30] P. Matousek, F. Thorley, P. Chen, M. Hargreaves, C. Tombling, P. Loeffen, M. Bloomfield, D. Andrews, Emerging Raman techniques for rapid noninvasive characterization of pharmaceutical samples and containers, Spectroscopy 26 (2011) 44.
[31] L. Lee, A. Frisby, R. Mansson, R.J. Hopkins, Through-barrier detection of explosive components for security screening applications, in, SPIE, Bellingham Wash, pp. 81890V-81890V-81815.
[32] A. Frisby, L. Lee, C. Howle, A. Martin, R. Hopkins, Spatially offset Raman spectroscopy (SORS) for through-barrier proximal chemical and explosive detection, in, SPIE, Bellingham Wash, pp. 81890B-81890B-81814.
[33] P.W. Loeffen, G. Maskall, S. Bonthron, M. Bloomfield, C. Tombling, P. Matousek, Spatially offset Raman spectroscopy (SORS) for liquid screening, in, SPIE, Bellingham Wash, pp. 81890C-81890C-81810.
[34] C. Conti, A. Botteon, C. Colombo, D. Pinna, M. Realini, P. Matousek, Advances in Raman spectroscopy for the non-destructive subsurface analysis of artworks: Micro-SORS, J. Cult. Herit. 43 (2020) 319–328.
[35] A. Botteon, C. Colombo, M. Realini, C. Castiglioni, A. Piccirillo, P. Matousek, C. Conti, Non-invasive and in situ investigation of layers sequence in panel paintings by portable micro-spatially offset Raman spectroscopy, J. Raman Spectrosc. 51 (2020) 2016–2021.
[36] A. Botteon, C. Colombo, M. Realini, S. Bracci, D. Magrini, P. Matousek, C. Conti, Exploring street art paintings by microspatially offset Raman spectroscopy, J. Raman Spectrosc. 49 (2018) 1652–1659.
[37] D. Chiriu, G. Desogus, F.A. Pisu, D.R. Fiorino, S.M. Grillo, P.C. Ricci, C. M. Carbonaro, Beyond the surface: Raman micro-SORS for in depth non-destructive analysis of fresco layers, Microchem. J. 153 (2020) 9.
[38] C. Conti, C. Colombo, M. Realini, P. Matousek, Subsurface analysis of painted sculptures and plasters using micrometre-scale spatially offset Raman spectroscopy (micro-SORS), J. Raman Spectrosc. 46 (2015) 476–482.
[39] A. Rousaki, A. Botteon, C. Colombo, C. Conti, P. Matousek, L. Moens, P. Vandenabeele, Development of defocusing micro-SORS mapping: a study of a 19th century porcelain card, Anal. Methods 9 (2017) 6435–6442.

[40] A. Botteon, C. Conti, M. Realini, C. Colombo, P. Matousek, Discovering Hidden Painted Images: Subsurface Imaging Using Microscale Spatially Offset Raman Spectroscopy, Anal. Chem. 89 (2017) 792–798.
[41] A. Botteon, M. Vermeulen, L. Cristina, S. Bruni, P. Matousek, C. Miliani, M. Realini, L. Angelova, C. Conti, Advanced Microspatially Offset Raman Spectroscopy for Noninvasive Imaging of Concealed Texts and Figures Using Raman Signal, Fluorescence Emission, and Overall Spectral Intensity, Anal. Chem. 96 (2024) 4535–4543.
[42] C. Corden, P. Matousek, C. Conti, I. Notingher, Sub-Surface Molecular Analysis and Imaging in Turbid Media Using Time-Gated Raman Spectral Multiplexing, Appl. Spectrosc. 75 (2021) 156–167.
[43] R. Mulholland, D. Howell, A. Beeby, C.E. Nicholson, K. Domoney, Identifying eighteenth century pigments at the Bodleian library using in situ Raman spectroscopy, XRF and Hyperspectral Imaging, Heritage Sci. 5 (2017) 43.
[44] C. Yan, Z. Cheng, S. Luo, C. Huang, S. Han, X. Han, Y. Du, C. Ying, Analysis of handmade paper by Raman spectroscopy combined with machine learning, J. Raman Spectrosc. 53 (2022) 260–271.
[45] D. Chiriu, P.C. Ricci, G. Cappellini, Raman characterization of XIV–XVI centuries Sardinian documents: Inks, papers and parchments, Vib. Spectrosc 92 (2017) 70–81.
[46] N.S. Daly, M. Sullivan, L. Lee, K. Trentelman, Multivariate analysis of Raman spectra of carbonaceous black drawing media for the in situ identification of historic artist materials, J. Raman Spectrosc. 49 (2018) 1497–1506.
[47] M. Vermeulen, D. Tamburini, E.M.K. Müller, S.A. Centeno, E. Basso, M. Leona, Integrating liquid chromatography mass spectrometry into an analytical protocol for the identification of organic colorants in Japanese woodblock prints, Sci Rep 10 (2020) 20921.
[48] M. Vermeulen, L. Burgio, N. Vandeperre, E. Driscoll, M. Viljoen, J. Woo, M. Leona, Beyond the connoisseurship approach: creating a chronology in Hokusai prints using non-invasive techniques and multivariate data analysis, Heritage Science 8 (2020) 62.
[49] V.A. Solé, E. Papillon, M. Cotte, P. Walter, J. Susini, A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra, Spectrochim. Acta B at. Spectrosc. 62 (2007) 63–68.
[50] S. Mosca, C. Conti, N. Stone, P. Matousek, Spatially offset Raman spectroscopy, Nat. Rev. Methods Primers 1 (2021) 21.
[51] V. Corregidor, R. Viegas, L.M. Ferreira, L.C. Alves, Study of Iron Gall Inks, Ingredients and Paper Composition Using Non-Destructive Techniques, Heritage 2 (2019) 2691–2703.
[52] I. Brückle, Historical manufacture and use of blue paper, The Book and Paper Group Annual 12 (1993) 5–7.
[53] G.V. Fichera, M. Malagodi, P. Cofrancesco, M.L. Weththimuni, C. Guglieri, L. Olivi, S. Ruffolo, M. Licchelli, Study of the copper effect in iron-gall inks after artificial ageing, Chem. Pap. 72 (2018) 1905–1915.
[54] E. Gliozzo, Pigments — Mercury-based red (cinnabar-vermilion) and white (calomel) and their degradation products, Archaeol. Anthropol. Sci. 13 (2021) 210.