Unveiling hidden text in Frisian historic manuscripts by fused reflectance and transmittance imaging spectroscopy

Di Benedetto, Alessia; de Almeida Nieto, L.M.; Marsh, Daniel; Zwetsloot, Janneke; Janssen, Riemer; Popkema, Anne Tjerk; de Vries, Herre; Comelli, Daniela; Alfeld, Matthias

DOI
10.1038/s40494-025-02001-5

Publication date
2025

Document Version
Final published version

Published in
npj Heritage Science

npj | heritage science

Article

https://doi.org/10.1038/s40494-025-02001-5

Unveiling hidden text in Frisian historic manuscripts by fused reflectance and transmittance imaging spectroscopy

Check for updates

Alessia Di Benedetto[1,2], Luìs Manuel de Almieda Nieto[2], Daniel Marsh[2], Janneke Zwetsloot[2], Riemer Janssen[3], Anne Tjerk Popkema[3], Herre de Vries[3], Daniela Comelli[1] & Matthias Alfeld[2] ✉

In several manuscripts, text is obscured due to glued together leaves, making it difficult or impossible to read. In this context, this study explores the use of reflectance and transmittance imaging spectroscopy (RIS and TIS) in the visible and near-infrared range (400–1000 nm) to recover hidden texts. The method is applied to two cases of medieval Frisian legal codes from the Richthofen Collection, where we employed Non-Negative Matrix Factorization (NMF) for the analysis of single and spectrally fused datasets integrating both RIS and TIS. We further integrated spatial stitching of adjacent areas to enhance spatial resolution of the images. Our results demonstrate that factorization algorithms perform well on fused datasets, with spectral fusion proving essential in complex cases where individual analyses fail to clearly reveal hidden text.

Manuscripts can offer insight into the past's cultural, intellectual, social, and artistic practices. But a manuscript is more than its literal content, as its material aspects give insight into its creation process and treatment over the centuries. For example, before the 19th century bookbinding was a process independent from writing or printing, often resulting in texts being compiled into a single volume long after they were originally written or printed. Consequently, the study of manuscripts is a field of Heritage Science (HS) constantly growing and bridges the disciplines of history, art, and science[1].

Scientific investigations of manuscripts address different issues. The conservation of manuscripts for future generations requires knowledge of their composition and condition to develop proper storage and conservation plans. Furthermore, ink can either mechanically or chemically deteriorate. Sometimes this is done deliberately to re-use the support of the manuscript for a new text, a so-called palimpsest. In addition, in a few cases text was edited and deleted[2].

Old documents are delicate, and their fragility and value pose significant challenges when doing these investigations. In general, the stress these manuscripts are exposed to needs to be minimized by avoiding their transport and the exposure to intense light and heat sources. Many analytical techniques are employed to study how to make manuscripts more readable[2–4]. Among these, non-invasive imaging techniques stand out for their capacity to uncover hidden aspects of manuscripts and enhance their readability in a non-invasive way, without any physical contact. The most used imaging techniques are macroscopic X-ray Fluorescence mapping (MA-XRF)[5–8], effective as an elemental mapping technique, and reflectance imaging spectroscopy (RIS)[4,6,9–13], which can permit to distinguish between different chemical compounds.

Specifically, RIS in the visible and near IR spectral range (VNIR; 400–1000 nm)[11,12] or in the short-wave infrared spectral range (SWIR; 1000–2500 nm)[13] has gained prominence, due to its speed and ease of use. Indeed, a RIS setup consists of a continuous light source, such as halogen lamps, and a hyperspectral camera. The acquisition time necessary to scan an area is in the range of some seconds, which is short compared to many other imaging spectroscopy (IS) methods, including MA-XRF. In the recent past, RIS has been effectively applied to recover text in manuscripts damaged by accidents like carbonization, ink bleeding, corrosion, aging, and concealment in bookbinding[9,11,13], with notable examples including the restoration of illegible fragments of the Dead Sea Scrolls[14,15], Petra Church scrolls[15,16], and the carbonized Herculaneum papyri[17,18].

When analysing semi-transparent objects, IS in the VNIR range is commonly implemented in one of the two detection modes: reflectance (RIS) and transmittance (TIS)[19]. In particular, RIS is used to investigate the surface of an object, while TIS allows to put more weight on the internal/sub-surface features of a semi-transparent material like paper, that are not visible on the surface. The combined use of RIS and TIS is widely documented in literature in cases where it is necessary to detect differences in deeper regions of the sample. While this method is widely used in the agricultural sector[20–22], it remains largely unexplored in Heritage Science. In particular, TIS can be well-suited for the study of documents and manuscripts, as their (partial) transparency allows for the collection of transmitted light. Thus, by combining TIS and RIS it is

[1]Politecnico di Milano, Physics Department, Milano, Italy. [2]Delft University of Technology, Department of Material Science and Engineering, CD Delft, the Netherlands. [3]Pastei Research Group for Book History, Zuidhorn, the Netherlands. ✉e-mail: m.alfeld@tudelft.nl

possible to highlight differences between the visible surface of a page and hidden elements that are otherwise undetectable to the naked eye.

The treatment of IS data for manuscripts is based on statistical analysis and dimension reduction techniques, often followed by contrast enhancement achieved through commercial graphics editors. Principal component analysis (PCA) is the most often used unsupervised dimension reduction technique[23], but also other factorization techniques have been explored, such as independent component analysis (ICA)[24,25] and non-negative matrix factorization (NMF)[26]. The latter two are also labelled as blind source separation techniques, as in many cases they allow to separate mixed signals "blindly", i.e., without external information.

When collecting imaging spectroscopy data with different modalities — such as diffuse reflected light and diffuse transmitted light or diffuse reflected light and fluorescence light — each spectral imaging dataset is typically processed separately, with dimensionality reduction applied before combining their representations for interpretation. However, recent studies[9,27–29] have explored approaches that directly integrate data from multiple modalities to generate a unified representation, demonstrating that spectral data fusion can leverage the complementary information from different modalities, enhance interpretability, and simplify the resulting representation.

In this context, this article explores the fusion of RIS and TIS data for the visualization and enhancement of hidden texts in manuscripts, making use of state-of-the-art multivariate analysis. The investigated objects are two cases of glued together leaves in handwritten Frisian legal codes from the medieval age, which are part of the Richthofen Collection in Tresoar (Leeuwarden, the Netherlands. https://richthofen.nl). In the first case, the *Fivelgo* manuscript, the text is hidden between a written and a blank leaf. Here, our study demonstrated that the use of RIS data is sufficient to reveal it. However, it requires the spatial fusion of multiple data sets to enhance the final spatial resolution. In the second case, the *Codex Phaesma*, where both sides of the glued leaf are written on, we proved that the spectral fusion of RIS and TIS data is needed to improve the detection and readability of obscured writings, demonstrating how spectral fusion leverages the complementary information contained in RIS and TIS data.

Methods

Manuscripts

This study focuses on illegible text from two books from the Richthofen Collection. The collection contains 10 bound manuscripts containing Frisian legal texts, dating from 1350 to 1600, predominantly in Old Frisian. The collection is located in the Tresoar Museum, archive and library Friesland (Tresoar Museum, argyf en biblioteek Fryslân), in Leeuwarden (the Netherlands). A codicological and paleographical investigation of the collection was recently published[30].

The *Fivelgo manuscript* (manuscript R 4, paginated, dimensions 188 x 140 x 22 mm, pages 52–67 of ref. 30) was likely written in the second quarter of the 15th century and rebound during the 4th quarter of the 18th century. A peculiar feature of this manuscript is that at an unknown moment in its history a blank bifolio was added after page 45 of which the first leaf is glued entirely over the manuscript's page 46. This covers an already written text that can be vaguely discerned through the blank leaf. The upper part of Fig. 1 shows the glued leaf under investigation, with writing on the first visible recto while the closing verso is blank. The second layer of text, hidden between the glued leaves, can be seen by looking at the closing verso in transmission mode.

The second case comes from manuscript R 8, recently dubbed *Codex Phaesma* (foliated, dimensions 247 x 201 x 27 mm, pages 127–142 of ref.[30]). Using transmission photography (Fig. 1F), it was possible to determine that folio 6 consists of two leaves glued together: indeed, extra lines of text are clearly visible. This was most likely done by adding an unfoliated leaf to folio 6 verso. Unlike the previous case, both folio 6 recto and the verso of the

Fig. 1 | Overview of the cases of study. Recto **A** and Verso **B** of glued leaf from R 4 Fivelgo manuscript, **C** Image of verso in transmission geometry. Recto **D** and Verso **E** of glued leaf from R 8 Codex Phaesma, **F** Image of verso in transmission geometry.

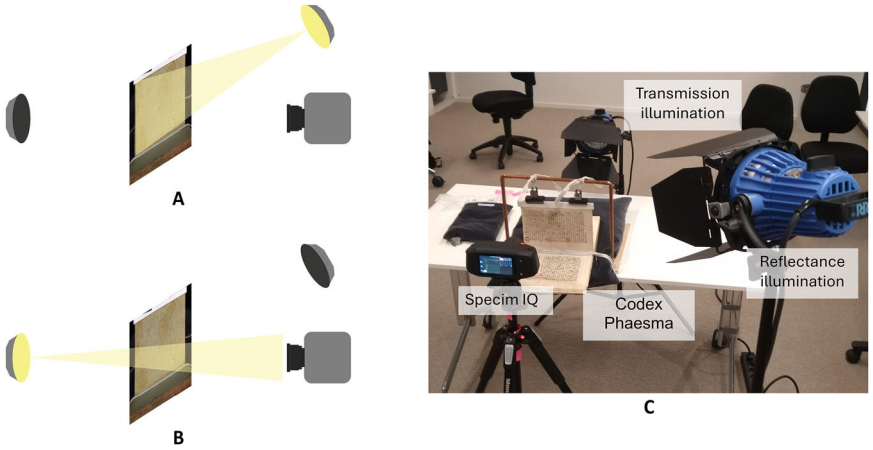

Fig. 2 | Experimental configuration. A Reflectance mode, **B** Transmission mode. **C** Picture of the experimental configuration.

added leaf have fully written sides. The foliation was dated roughly to the 17th–18th century.

In both cases the presence of the hidden text can be seen in the transmission photographs, but as it cannot be separated from the visible text, the concealed text remains illegible.

Experimental setup

For the acquisition of RIS and TIS datasets a Specim IQ[31] portable pushbroom hyperspectral camera (Specim ltd., Oulu, Finland) was used. The Specim IQ is sensitive to the VNIR range from 400 nm to 1000 nm, with a mean spectral resolution of 7 nm. The camera acquires IS data cubes of 512 × 512 pixels with 204 spectral channels. The pixel size in the data discussed in this article is approximately 0.25 mm. Two ARRILITE 70 Plus 240 V/800 W halogen lamps (ARRI AG., Munich, Germany) were employed for the illumination of the scene. The entire experimental set-up is comparably low cost and can be transported by two scientists in a private car or public transport. To correct the spectra in the reflectance mode, a white reference (ColorChecker Passport Photo 2019 edition) was placed in the field of view. TIS data were not corrected, as the reference measure - i.e., the light collected by the camera without the sample, was not acquired. We did not perform any flat-field correction, i.e., we did not correct for spatial non-uniformities in image intensity.

As seen in Fig. 2, the leaves under examination were raised, then both the recto and the verso of the glued leaves were analysed in reflectance and transmittance modes. Specifically, the established acquisition protocol in the case of tailored vertical data fusion involves fixing the camera, while the two lights were alternatively turned on, with one operating in reflectance and the other one in transmittance mode. This ensured the co-registration of the two datasets. Each side of the leaf was imaged through six different areas to combine them into a single data set with superior lateral resolution compared to a single image of the entire leaf. To allow for spatial stitching, the six zones were taken with some overlapping parts (at least 20%) by moving the camera to the next position, while manuscript and illumination staid fixed.

Data processing - Multivariate analysis

An IS dataset consists of a stack of hundreds of images each one acquired at a different wavelength, resulting in a hypercube where x and y directions represent the imaged area, and d direction is the wavelength collected. A manual inspection of such a number of images is highly inefficient and thus robust multivariate analysis methods are required to reduce redundancies and highlight correlations and latent features. Indeed, IS data can be expressed as a matrix multiplication of a smaller number of images and the resulting spectra. As the dimensionality of the data is reduced this process is also called data reduction.

When applying data reduction methods, the IS datacube is rearranged in data matrix D^{nxd}, composed of n ($=x * y$) pixels and d channels representing the spectral dimension. Matrix factorization decomposes the original high-dimensional data, consisting of d spectral dimensions, into a set of k fundamental components that preserve the informative content. The number of components k is selected to be significantly smaller than the original spectral dimension d. Thus, matrix factorization represents the data matrix D as the product of two lower dimensional matrices: the basis matrix (or endmembers matrix) W, which consist of k spectral endmembers, and the coefficient matrix (or abundances matrix) H, which contains the contribution, or abundance, of each endmember to each pixel and consists consequently of images. Matrix factorization is expressed in the following equation:

$$D^{n\times d} = H^{n\times k} \times W^{k\times d} + E^{n\times d}$$

Where the remainder E contains the difference between the original data and the representation.

Through matrix factorization it is possible to describe the IS datacube in a distinct representation where the components reflect the various features and patterns within the data[2]. This transformation models the IS dataset as a linear combination of different endmembers and thus facilitates the separation of overlapping features and the subsequent retrieval of images that allows to visualize a single and isolated layer. The methods used to calculate H and W, and consequently the approximated data matrix $H\,x\,W$, vary throughout matrix factorization techniques. Principal Component Analysis (PCA) is the most used approach. Nevertheless, other methods like Independent Component Analysis (ICA)[24,25] and Non-negative Matrix Factorization (NMF)[26] have recently gained popularity as data reduction methods.

ICA assumes the mutual independence of the endmembers, i.e., it aims to find basis vectors that are uncorrelated and independent. In the case of hidden text, the endmembers extracted highlight the various text layers, enabling their differentiation. In particular, the FastICA algorithm[32] has been shown to be a useful tool for differentiating between superimposed texts[33]. NMF is a factorization method that decomposes the non-negative data matrix (D) into two non-negative matrices. Specifically, NMF is easier to interpret than PCA since it preserves the non-negative nature of the data. Thus, NMF models data as a linear combination of endmembers and abundances that are more easily interpreted, as it corresponds to the underlying physics. To our knowledge, NMF has not yet been applied to separate written layers in manuscripts. However, it is widely employed for material separation, effectively separating overlapping spectral contributions and mapping them spatially[34,35]. Its ability to extract distinct components suggests strong potential for separating overlapping textual layers. Therefore, both ICA and NMF are powerful tools for manuscript analysis, as they

offer significant improvements in the separation and interpretation of complex datasets, enhancing the clarity of individual layers.

Ultimately, we employed NMF as matrix factorization technique to enhance the legibility of the hidden texts. It is noteworthy to mention that the goal of this study is not to retrieve the spectral signatures of the layers and that we do not expect that individual components can be associated with individual layers of the manuscripts but rather to reconstruct the images in a way that improves the readability of text. For this reason, in the presentation of results in the following discussion, we focused exclusively on the coefficient matrix *H*, which represents the spatial distribution of each extracted endmember, allowing us to isolate the different texts and background components.

Data processing – Data fusion protocol

A limiting factor for all methods of matrix factorization discussed above is that *H* and *W* are calculated simultaneously and are solutions optimized for a specific hypercube *D*. When dealing with multiple datasets *D*, each will have its own representation, leading to variations in *H* and *W* that make them not directly comparable. A method to ensure a uniform basis matrix W and an abundances matrix H is to fuse the data sets before subjecting them to the factorization method.

Specifically, both spectral and spatial data fusion can be considered. Spectral (or vertical) fusion combines datasets of different imaging techniques obtained on the same sample/area into a single hypercube, allowing us to exploit synergies between different methods and simplifying the interpretation of the data as only a single representation needs to be studied. Recent studies in the field of IS have demonstrated the effectiveness of this methodology, where data from different imaging modalities are combined in the early stage of the process, rather than during their interpretation[9,27,28]. When spectrally fusing multiple datasets, they are typically stacked along the spectral dimension (d). Thus, it is crucial to carefully consider the statistical weight each dataset contributes to the final fused cube. This involves taking the number of spectral channels each dataset contains into account, the spectral range and intensity values they cover. Thus, to achieve an effective fusion, normalization is critical to ensure balanced contributions. In our case, since both datasets have the same number of channels, the spectra were just scaled using pixel min-max normalization, where each pixel's spectral values are rescaled between 0 and 1, to better compare the spectral curve of each point. This approach ensures that neither dataset dominates the other during multivariate analysis, allowing for a balanced and unbiased integration of the spectral information. While normalization removes the physical interpretability of reflectance values, this is acceptable in applications focused on pattern recognition (e.g., revealing the hidden text), rather than quantitative material analysis. Vertical data fusion requires pixel-perfect alignment between the different modalities so that the spectral features coming from the same pixel are matched. In our experiments this process was simplified by acquiring RIS and TIS data without moving the camera.

Spatial (or horizontal) data fusion can be used to achieve higher spatial resolution by combining several hypercubes of details into one larger data set. Horizontal data fusion treats multiple data sets of the same modality (i.e., RIS or TIS) and fuses them into one block. To align the data from the different areas, one can exploit algorithms originating from computer vision, such as the scale-invariant feature transform (SIFT) that detect spatial features (key points) in images[36,37]. It is integrated into the Open Computer Vision library[38] and the open-source data evaluation package DataHandlerP[39]. This approach was used in a past study[27] and also explored in this project. However, we decided to stitch the data sets together next to one another without overlapping and aligned them afterwards, using the images from the resulting abundance matrix *H*, as detailed below.

Data processing – Image legibility improvement

Once *H* is calculated it can be exported in the form of images and be further processed in graphic editors, such as GIMP or Photoshop to increase the legibility of the images. Contrast enhancement was applied and, if necessary, we selectively adjusted the image's dynamic range, i.e., clipping the intensity within a specific range. More information about image legibility improvement can be found in the supplementary information. Care was taken to verify, by visual inspection, that features identified in enhanced images were also found in the NMF raw images, in order to not overinterpret artefacts introduced by image processing.

Results

Case 1: leaf from Fivelgo manuscript (manuscript R 4)

RIS and TIS data were taken on both recto and verso of the glued leaf in the *Fivelgo* manuscript. The potential of revealing the hidden text was explored by processing individual data sets with different matrix factorization methods. It was found that the RIS data of the leaf's (blank) verso was sufficient to reveal the text, while TIS data failed at separating the two texts. PCA, ICA and NMF were tried for this. PCA failed to reconstruct the hidden text, while NMF and ICA did so, but the former with a clearer contrast. The results of PCA, ICA and NMF performed on an area of the leaf are shown in Fig. S1 of the supplementary information.

As a single set of RIS data covering the entire leaf with 512 × 512 pixels was not expected to have sufficient lateral resolution to reveal minute details required to read the covered text, six scans of the leaf were taken to be horizontally fused after normalization to the external reference, as described above. As the blank leaf featured only a few points for aligning the data set, the RIS cubes were simply placed into one hypercube, next to one another without alignment, as described above.

The spectra were smoothed with a Savitzky-Golay filter (range 23 channels and order 3). The spectral range from 426 nm to 948 nm was selected to exclude channels in that the reduced sensitivity of the Specim IQ results in enhanced noise levels. The smoothed and normalized data was then factorized via NMF with 6 components. The number of components was empirically determined until a single component that allows for the visualization of the hidden text was found. The component was selected through visual inspection and it was exported into a graphics file (see Fig. 3).

The obtained image was processed in GIMP and subjected to a High Pass filter, a Symmetric Nearest Neighbor filter and a filter to mask the unsharp parts. This is illustrated in Fig. S2 of the supplementary information. Afterwards the six scans were separated and stitched in Photoshop to the result shown in Fig. 3. Photoshop was chosen for the latter operation as it provided a smooth stitching of overlapping areas and yielded a readable text. The first three lines of the text were transcribed as follows: "*[S]O we by sit arue oft anders//onreplycker gueden offt dessen [is?]//in gueder geboert ghekomen Ende*" with [S] indicating a capitalized letter that had not been executed when the leaves were glued together. This transcription allowed to identify the text as part of the medieval Frisian legal text *Oldambtster landrecht*, in a Middle Low German translation. A publication on the complete transcription and a proper interpretation of it are in preparation.

Case II: leaf from *Codex Phaesma* (manuscript R 8)

As shown in Fig. 1, recto and verso of the glued leaf under discussion of the *Codex Phaesma* feature text, and the transmitted image clearly shows additional hidden text that does not belong to either side of the leaf. We performed both RIS and TIS on each side of the leaf, imaging both recto and verso through six different areas. Fig. S3 presents a detailed comparison of the analysis performed on the individual reflectance and transmittance datasets. It is evident that the reflectance data from both front and back sides (Fig. S3A, C) do not provide any meaningful information about the hidden text layer, likely because reflected light is predominantly influenced by surface features. In contrast, the transmittance data from both sides partially reveal the hidden text, although the separation achieved is not sufficient for clear decipherment. Nevertheless, the transmittance analysis played a crucial role to assess where the concealed writing is located. By examining the orientation of the visible diagonal strikethroughs and characters (Fig. S3B, D), it became clear that only one of the two glued sides featured text. Specifically, the presence and position of capital letters in transmittance data

from verso of the glued leaf (Fig. S3D, highlighted in blue), indicates that the verso of its first leaf remained blank and the writing is actually on the glued recto of this second leaf. Despite these insights, the individual investigation of RIS or TIS datasets was not sufficient to fully separate the hidden text from the visible one and improve its legibility. For this reason, the RIS and TIS data from the six investigated areas from the verso side of the glued leaf were vertically fused and then analysed through NMF. We selected the spectral range from 426 nm to 948 nm to avoid noisy channels. Before the spectral merging, the spectra of the RIS and TIS dataset were smoothed with a Savitzky-Golay filter (range 17 channels and order 3) and each pixel's values were normalized between 0 and 1 for each dataset, separately. The smoothed and normalized data was then vertically fused and one of the six fused datacubes was factorized via NMF with 6 components. The use of a smaller dataset set allowed for a faster computation of the matrices W and H and optimization of factorization settings. Indeed, once an optimal set of H and W was found, each of the remaining five spectrally merged datasets was expressed as a linear combination of the spectral profiles W, yielding a homogeneous set of representations H.

The NMF analysis for each of the six sections of the glued leaf's verso can be seen in Fig. 4A. After the maps were found, the images were combined through SIFT and enhanced in Photoshop through contrast enhancement and by refining the dynamic range, particularly excluding high intensity values range. Indeed, looking at the NMF representation, the hidden text has low intensity values in the selected image, the paper background is constituted by the medium intensity pixels and the surface text has a high intensity values. By selecting a

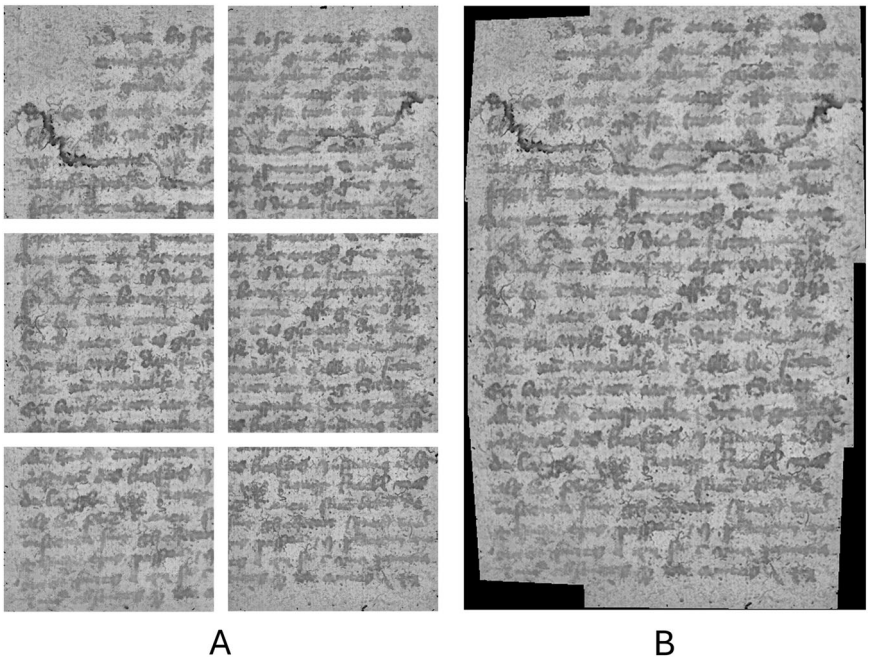

Fig. 3 | Analysis of reflectance measurements from the verso side of Case I. A Results of NMF analysis, with individual data blocks separated, enhanced by GIMP. **B** Horizontal fusion of the maps, achieved by Photoshop.

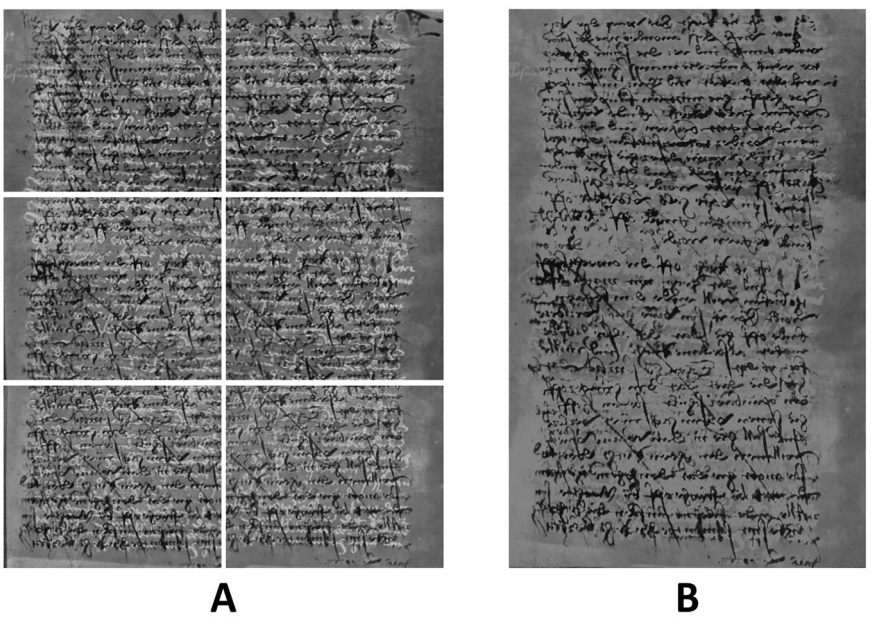

Fig. 4 | Analysis of fused reflectance and transmittance measurements from the verso side of Case II. A NMF analysis results, on each separated area. **B** Horizontal fusion of the maps, enhanced through Photoshop.

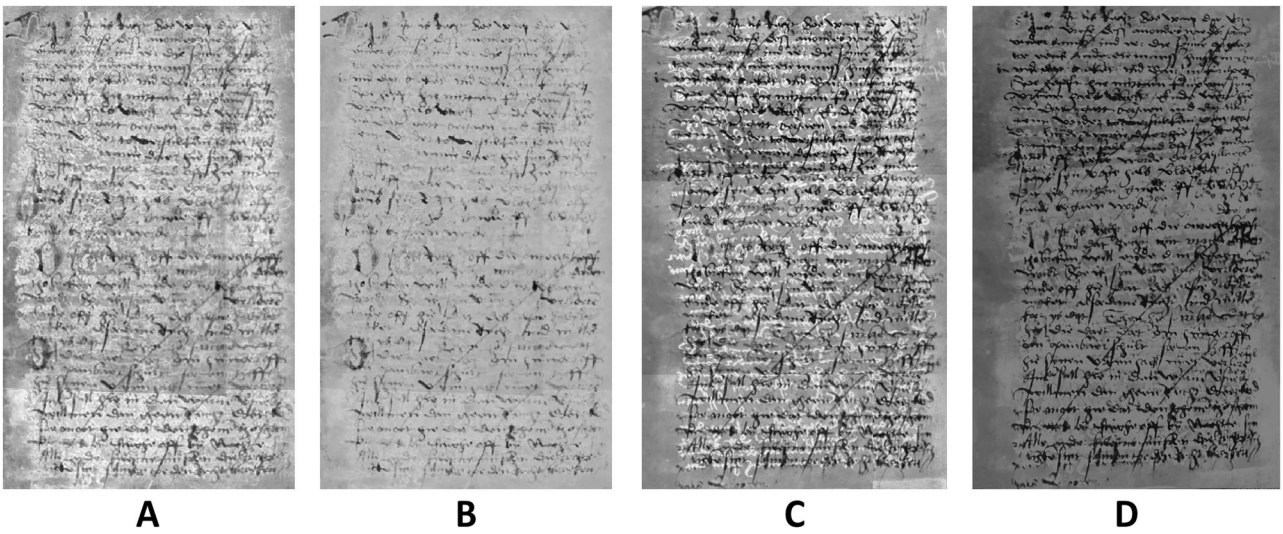

Fig. 5 | Comparison between spectrally non-fused and fused dataset. A NMF on TIS data from the verso, **B** Contrast optimized NMF on TIS data from the verso. **C** NMF on fused RIS and TIS data from the verso. **D** Contrast optimized NMF on fused RIS and TIS data from the verso. All the images are already mirrored in order to follow the text direction.

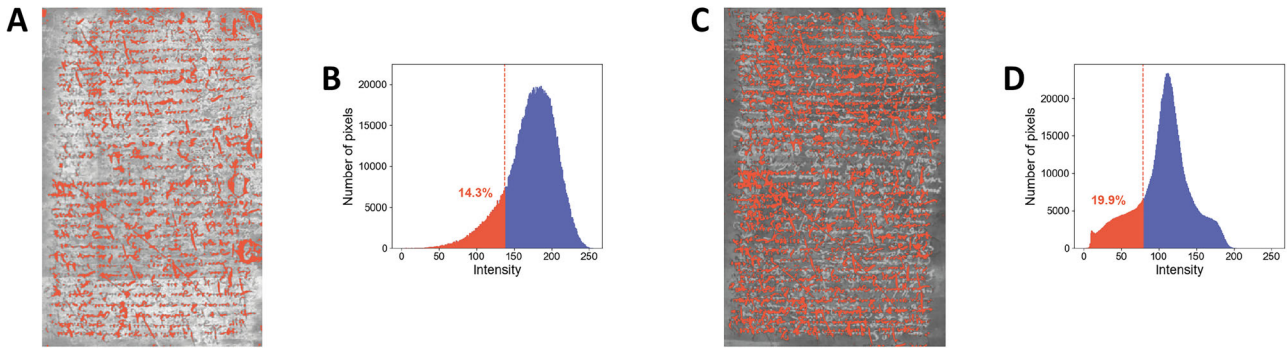

Fig. 6 | Quantitative evaluation of results from TIS and fused RIS-TIS data. A Result of spatially registered NMF analysis of TIS data with highlighted in red the low intensity pixels corresponding to the hidden text; **B** Histogram showing the intensity distribution of the pixels, with highlighted in red the low intensity pixels; **C** Result of spatially registered NMF analysis of fused RIS-TIS data with highlighted in red the low intensity pixels corresponding to the hidden text; **D** Histogram showing the intensity distribution of the pixels, with highlighted in red the low intensity pixels.

proper dynamic range of the image, the hidden text could be isolated. A detailed discussion on the manipulation of the dynamic range to enhance the hidden text is provided in the supplementary information. The first three lines were transcribed as: "*DJt is Recht Soe waer die vrye// vriese dess auondes nae der sonnen//onder ganck ind eer die sonne opgaet*." This text is, save matters of spelling, a near identical copy of the text on the visible folio 6 recto of the Codex Phaesma. Also here, a publication on the complete transcription and a historical interpretation is in preparation.

Discussion

This study of the two cases of glued leaves from the Richthofen Collection highlights the effectiveness of non-invasive RIS and TIS combined with a multivariate data processing of the collected IS data.

We show a portable user-friendly setup consisting of a low-cost commercial push-broom camera, paired with two halogen lamps, enabling flexible and rapid acquisition. This portable configuration provides high-quality imaging data without requiring complex instrumentation, making it accessible for on-site cultural heritage investigations. Further, we showcase the potential of spatially stitching datasets (horizontal data fusion) of adjacent areas. This approach allows for improved spatial resolution and the study of larger regions with greater detail.

We demonstrated our data acquisition and analysis protocol on two cases of text hidden between glued leaves in manuscripts. In the Case I, the hidden text was vaguely visible to the naked eye from the reverse side of the leaf. Thus, using spatially stitched RIS data from its verso side, it was possible to improve the legibility and to identify the hidden text.

The second case, however, was far more intricate. The hidden text is located on the glued recto of the second leaf, but as its verso is written, it is not possible to read the text well by analysing either RIS or TIS data alone. We made the decision to combine the two hypercubes, specifically the transmission and reflectance modes on the verso of the glued leaf, to improve the layer separation. In Fig. 5, the comparison between the mono-modal analysis and the one performed on fused datasets is shown. The fusion of RIS and TIS was able to extract and discriminate the text on the visible verso of the second leaf from the hidden content on its glued recto, enhancing the legibility of the hidden text.

To provide a more quantitative evaluation, we measured the amount of low-intensity pixels in the NMF-registered images derived from TIS data alone and from the spectrally fused RIS-TIS datasets. As shown in Fig. 5, the text recovered from TIS-only analysis appears significantly less contrasted, with several text portions missing or blending into the grey background. In Fig. 6 we highlighted in red the low-intensity pixels, corresponding to hidden text, for both the TIS dataset (panel A) and the fused RIS-TIS dataset

(panel C). The histograms (panels B and D) reveal that the concealed text accounts for 14.3% of total pixels in the TIS-only image, increasing to 19.9% in the fused data. This clearly demonstrates that spectral fusion enhances the clarity and level of hidden text recovery by integrating complementary reflectance and transmittance information. The details of this calculation are in the *supplementary information*.

It is noteworthy to mention that the fusion protocol can be also exploited for the combination of other types of datasets collected with different techniques. Indeed, there are studies in which X-ray fluorescence and reflectance have been used in an earlier stage of the analysis and show how that improved the material identification[27]. Similarly, Li et al.[29] recently proved the effectiveness of spectral fusion in the detection of degradation phenomena in dyed textiles, combining RIS and photoluminescence imaging.

We recognized that the strength of the SIFT algorithm is its ability to automatically and reliably identification of key points to facilitate data fusion. Our measurements purposely included sufficient overlapping regions between adjacent areas: this overlap is essential for SIFT and similar approaches to detect and match similar characteristics and enable precise spatial alignment of the results. Moreover, the DataHandlerP software significantly simplifies the stitching workflow by integrating SIFT-based alignment and image warping into an intuitive, semi-automatic user interface.

In conclusion, this work highlights how advanced analytical techniques can effectively support the investigation of historical and cultural aspects of cultural heritage artifacts. We leverage consolidated multivariate analytical techniques and a novel protocol involving the spectral fusion of complementary modalities (RIS and TIS), to effectively unveil hidden text from manuscripts. We were able to contribute to the knowledge of the practices of documenting and storing legal texts of medieval Frisia. This approach not only offers valuable insights into Frisian history but also demonstrates the broader potential of these techniques, properly adapted on the case study, in revealing artistic and cultural elements within other historical artworks. Our findings suggest that the continued development and application of such methodologies could significantly advance the field, enabling deeper exploration and preservation of cultural heritage.

Data availability

The data used in this publication are available from the 4TU repository: https://doi.org/10.4121/3ee53917-a217-4b58-96c2-361ead8faf3b.

Received: 19 March 2025; Accepted: 14 August 2025;
Published online: 12 September 2025

References

1. Easton, R. L. & Noel, W. Infinite Possibilities: Ten Years of Study of the Archimedes Palimpsest. *Proc. Am. Philos. Soc.* **154**, 50–76 (2010).
2. Tonazzini, A. et al. Analytical and mathematical methods for revealing hidden details in ancient manuscripts and paintings: A review. *J. Adv. Res* **17**, 31–42 (2019).
3. Perino, M., Pronti, L., Moffa, C., Rosellini, M. & Felici, A. C. New Frontiers in the Digital Restoration of Hidden Texts in Manuscripts: A Review of the Technical Approaches. *Heritage* **7**, 683–696 (2024).
4. Knox, K. T., Easton, R. L. & Christens-Barry. W. A. *Image restoration of damaged or erased manuscripts 16th European Signal Processing Conference* 1–5 (2008).
5. Magkanas, G., Bagán, H., Sistach, M. C. & García, J. F. Illuminated manuscript analysis methodology using MA-XRF and NMF: Application on the Liber Feudorum Maior. *Microchem. J.* **165** (2021).
6. de Viguerie, L. et al. XRF and reflectance hyperspectral imaging on a 15th century illuminated manuscript: combining imaging and quantitative analysis to understand the artist's technique. *Herit. Sci.* **6**, 11 (2018).
7. Turner, N. K., Patterson, C. S., MacLennan, D. K. & Trentelman, K. Visualizing underdrawings in medieval manuscript illuminations with macro-X-ray fluorescence scanning. *X-Ray Spectrom.* **48**, 251–261 (2019).
8. Bergmann, U., Manning, P. L. & Wogelius, R. A. Chemical Mapping of Paleontological and Archeological Artifacts with Synchrotron X-Rays. *Annu. Rev. Anal. Chem.* **5**, 361–389 (2012).
9. Pouyet, E. et al. Revealing the biography of a hidden medieval manuscript using synchrotron and conventional imaging techniques. *Anal. Chim. Acta*. **982**, 20–30 (2017).
10. Starynska, A., Messinger, D. & Kong, Y. Revealing a history: palimpsest text separation with generative networks. *Int. J. Doc. Anal. Recognit. ((IJDAR))* **24**, 181–195 (2021).
11. Joo Kim, S., Deng, F. & Brown, M. S. Visual enhancement of old documents with hyperspectral imaging. *Pattern Recognit.* **44**, 1461–1469 (2011).
12. Wijsman, S., Neate, S., Kogou, S. & Liang, H. Uncovering the Oppenheimer Siddur: using scientific analysis to reveal the production process of a medieval illuminated Hebrew manuscript. *Herit. Sci.* **6**, 15 (2018).
13. Tournié, A. et al. Ancient Greek text concealed on the back of unrolled papyrus revealed through shortwave-infrared hyperspectral imaging. *Sci. Adv* **5**, 1–8 (2019).
14. Bearman, G. H. & Spiro, S. I. Archaeological Applications of Advanced Imaging Techniques. *Biblic Archaeol.* **59**, 56–66 (1996).
15. Chabries, D. M., Booras, S. W. & Bearman, G. H. Imaging the past: recent applications of multispectral imaging technology to deciphering manuscripts. *Antiquity* **77**, 359–372 (2003).
16. Chabries, D. M. & Booras, S. W. The Petra Scrolls *Image Processing, Image Quality, Image Capture Systems Conference* (2001).
17. Booras, S. W. & Seely, D. R. Multispectral imaging of the Herculaneum papyri *Cronache*. *Ercolanesi* **29**, 95–100 (1999).
18. Macfarlane, R., Del Mastro, G., Antoni, A. & Booras, S. Update Report on the Use of the Multi-spectral images of the Herculaneum papyri *XXIV International Congress of Papyrology, Helsinki, Finland* (Helsinki, Finland) 579–586 (2004).
19. Schaare, P. N. & Fraser, D. G. *Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (Actinidia chinensis)* 20, (2000).
20. Zhang, H. et al. Detection of seed purity of hybrid wheat using reflectance and transmittance hyperspectral imaging technology. *Front Plant Sci.* **13**, 1–13 (2022).
21. Leiva-Valenzuela, G. A., Lu, R. & Aguilera, J. M. Assessment of internal quality of blueberries using hyperspectral transmittance and reflectance images with whole spectra or selected wavelengths. *Innovative Food Sci. Emerg. Technol.* **24**, 2–13 (2014).
22. Lu, R. & Ariana, D. P. Detection of fruit fly infestation in pickling cucumbers using a hyperspectral reflectance/transmittance imaging system. *Postharvest Biol. Technol.* **81**, 44–50 (2013).
23. Pottier, F., Michelin, A. & Robinet, L. Recovering illegible writings in fire-damaged medieval manuscripts through data treatment of UV-fluorescence photography. *J. Cult. Herit.* **36**, 183–190 (2019).
24. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10**, 626–634 (1999).
25. Alfeld, M. et al. MA-XRF and hyperspectral reflectance imaging for visualizing traces of antique polychromy on the Frieze of the Siphnian Treasury. *Microchemical J.* **141**, 395–403 (2018).
26. Feng, X. R. et al. Hyperspectral Unmixing Based on Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE J. Sel. Top. Appl Earth Obs Remote Sens* **15**, 4414–4436 (2022).
27. Alfeld, M., Pedetti, S., Martinez, P. & Walter, P. Joint data treatment for Vis–NIR reflectance imaging spectroscopy and XRF imaging acquired in the Theban Necropolis in Egypt by data fusion and t-SNE. *C. R. Phys.* **19**, 625–635 (2018).
28. Catelli, E. et al. Towards the non-destructive analysis of multilayered samples: A novel XRF-VNIR-SWIR hyperspectral imaging system

combined with multiblock data processing. *Anal. Chim. Acta* **1239**, 340710 (2023).
29. Li, Z. et al. Multimodal HSI Combined with Multiblock Data Fusion: A New Tool for the Study of Time-Dependent Alteration Processes in Dyed Textiles. *Anal. Chem.* **96**, 19939–19946 (2024).
30. Pastei, Popkema, A. T., Janssen, R., & de Vries, H. De Richthofenkolleksje, Codicologische en inhoudelijke beschrijvingen van de tien handschriften, inclusief bezittersgeschiedenis (Version 2023) Zenodo. https://doi.org/10.5281/zenodo.7764244 (2023).
31. Behmann, J. et al. Specim IQ: Evaluation of a New, Miniaturized Handheld Hyperspectral Camera and Its Application for Plant Phenotyping and Disease Detection. *Sensors* **18**, 441 (2018).
32. FastICA algorithm https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.FastICA.html.
33. Tonazzini, A., Bedini, L. & Salerno, E. Independent component analysis for document restoration. *Int. J. Doc. Anal. Recognit.* **7**, 17–27 (2004).
34. Szymańska-Chargot, M., Pieczywek, P. M., Chylińska, M. & Zdunek, A. Hyperspectral image analysis of Raman maps of plant cell walls for blind spectra characterization by nonnegative matrix factorization algorithm. *Chemometrics Intell. Lab. Syst.* **151**, 136–145 (2016).
35. Di Benedetto, A., Pozzi, P., Valentini, G. & Comelli, D. Integrating multimodal Raman and photoluminescence microscopy with enhanced insights through multivariate analysis. *J. Phys.: Photonics* **6**, 035019 (2024).
36. Lowe D. G. 1999 *Object Recognition from Local Scale-Invariant Features*.
37. SIFT - OpenCV https://docs.opencv.org/4.x/d7/d60/classcv_1_1SIFT.html.
38. Open C. V. https://opencv.org/.
39. DataHandlerP https://sourceforge.net/projects/datahandlerp/.

Acknowledgements

A.D.B. Ph.D. research was financially supported by the European Union's NextGenerationEU Programme with the I-PHOQS Infrastructure [IR0000016, ID D2B8D520, CUP B53C22001750006] “Integrated infrastructure initiative in Photonic and Quantum Sciences”.

Author contributions

R.J., A.T.P., H.d.V. researched the context of the case, initiated the collaboration, transcribed the text from the final results and put it into the historical context. M.A., L.M.d.A.N., D.M., J.Z. performed the measurements of the books by RIS and TIS. A.D.B. and M.A. evaluated the data and wrote the main manuscript text. M.A. designed the data evaluation workflow. All authors revised the manuscript text together and reviewed and agreed on the submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s40494-025-02001-5.

Correspondence and requests for materials should be addressed to Matthias Alfeld.