

From closed testaments to books: Virtual X-ray Reading as an alternate digitization technology for fragile documents

F. Albertin^{1*}, M. Romito¹, E. Peccenini^{2,3,4}, M. Bettuzzi^{2,3,4}, R. Brancaccio^{2,3,4}, M. P. Morigi^{2,3,4}, M. del Rio⁵, D. Raines⁶, G. Margaritondo⁷, D. Psaltis¹.

^{1*} School of Engineering, Ecole Polytechnique Fédérale de Lausanne - EPFL, Lausanne, Switzerland;

² Enrico Fermi Center, Rome, Italy;

³ Department of Physics and Astronomy, University of Bologna, Bologna, Italy;

⁴ Italian National Institute for Nuclear Physics - INFN, Section of Bologna, Bologna, Italy;

⁵ Venetian State Archive, Venice, Italy;

⁶ Department of Humanities, University of Ca' Foscari, Venice, Italy;

⁷ School of Basic Science, Ecole Polytechnique Fédérale de Lausanne - EPFL, Lausanne, Switzerland.

Abstract

In recent years, research and technology made considerable progress in increasing the speed and the safety of the entire digitization process of ancient collections. Despite this, imaging ancient, fragile or un-opened documents remains a formidable challenge. We employ an alternate digitization technique for handwritten documents, exploiting x-ray tomography: Virtual X-ray Reading. Thanks to the high penetration of x-rays, we can acquire 3-dimensional (3D) tomographic images of a multi-page document without opening it. The x-ray contrast necessary for the readability is produced by the chemical composition and the consequent strong x-ray absorption of the iron gall inks - largely used for European handwritten documents. We present the development of this technology, from the chemical investigations of the inks to the tomography of an unopened Venetian testament and of an 18th century, 200-page, handwritten book.

Introduction

The research on this innovative technique is part of the Venice Time Machine (VTM) project, a collaboration of several European institutions to make the Archivio di Stato (Venice) collection widely available to researchers and to the general public in an interactive, open-access, structured digitalized database [1].

The Archivio is an excellent example of the problems affecting many historical collections worldwide, and provides a chance to find solutions to optimize the overall digitization process by exploiting novel technologies. This collection includes more than 80 km of administrative handwritten documents on all aspects of the Republic of Venice, more than a millennium of history. But this wealth of information is largely unexplored because of its sheer size: the challenge of the VTM project is to digitize, transcribe and open all this information. Unfortunately, just for the first step - the digitization - the current estimated time exceeds 10 years.

Our work focuses on the development of new techniques based on x-ray tomography to achieve digitization in a totally non-invasive and potentially faster way. With x-ray tomography, we minimize the handling of the documents and acquire complete 3D images without opening them. Note that the technique is completely safe for the documents: we observed no radiation effects in our tests. Moreover, in the case of extremely fragile manuscripts

like scrolls or sealed documents, this technique constitutes the only alternate solution to read the documents avoiding the risky procedure of unwrapping. In summary, the page-by-page virtual x-ray reading opens up a new way to unveil the secrets of otherwise inaccessible ancient documents throughout the world.

Originally designed for medical imaging, x-ray tomography detects shapes and structures not only from the surface of the object but also from the inside, obtaining a complete 3D volume imaging of the analyzed object. Going beyond the medical field, in recent years it became a routine investigation tool for Cultural Heritage objects like sculptures, paintings and others [2–5].

Even if more challenging, the feasibility of this technique for analyzing weakly x-ray absorbing specimens like manuscripts was recently demonstrated by several studies [6–10]. The x-ray contrast required to distinguish between supports and writings is achieved thanks to the differences in x-ray absorption. The supports are indeed weakly absorbing materials like papers and parchments, whereas the iron gall inks are strong absorbers.

Pioneering research [6–8] provided the basis for two of the most impressive recent results: the x-ray tomographic reconstructions of a rolled and burned Herculaneum papyri by V. Mocella et al. [9] and the complete virtual reconstruction of a parchment scroll from En_Ghedi by W.B. Seales et al. [10].

In parallel, in the past four years our group successfully performed chemical investigations and x-ray tomography on a large number of Italian manuscripts from the 14th to the 19th century, from single-page documents to books and even to sealed testaments [11–13]. We present here two of our recent results: the tomography of a 1351 Venetian unopened testament and that of an 18th century scientific book.

Ink analysis

In the first stage of our research, we investigated the chemistry of ancient inks, obtaining an extensive picture of these historical materials for handwritten Italian documents over several centuries. To detect the presence of heavy elements in the inks, we used a non-invasive technique: X-Ray Fluorescence (XRF) [13]. Even though this technique could not provide a full quantitative analysis, a semi-quantitative comparison of the results was crucial for the evaluation of the mineral content of the inks and of the

x-ray contrast that could be expected in x-ray tomography. Luckily, the composition of the iron gall inks produces sufficient x-ray contrast for our technique, whereas carbon-based inks are indistinguishable from supports like paper, papyrus and parchment.

Our results demonstrate the ubiquitous presence of iron compounds in the analyzed documents, confirming that iron gall inks were used not only for high-level "official" or artistic documents but also for everyday handwritings, and opening the way to their tomographic analysis. Besides iron, we found other heavy elements in the inks, like potassium, calcium, manganese, copper and zinc. The presence of calcium and potassium are most likely linked to the use of specific fillers, while the low concentration of the other metals leads us to assume that they are impurities of the iron compound [14].

The XRF ink analysis was also extended to the iron gall inks of the Venetian testament and of the 18th century book. Since the testament was sealed, the XRF spectra were taken on a letter of the external-page text. The spectra of Fig.1 clearly show the presence of iron in the ink. Likewise, in Fig.2 the results of paper and ink measures for the 18th century book reveal the use of an iron-based ink.

Tomographic investigations

A 3D tomographic volume is obtained by processing with *ad hoc* algorithms a series of two-dimensional (2D) x-ray images of the object taken at different angles (projections). The volume corresponds, pixel by pixel, to the x-ray absorbing power of the related points of the real object.

Our investigations used several different x-ray tomography instruments. The machines for medical imaging are designed to leave the patient stationary while the system (the x-ray source and the detector) rotates to acquire the projections. Usually, for Cultural Heritage investigations the objects need not to remain motionless and one can simplify the design by rotating the work of art while the instrument is fixed.

Using this latter approach, we initially conducted successful tomographic acquisitions of manuscripts at large x-ray facilities (synchrotrons) [11]. These sources provide extremely high brilliance and resolution, but they prevent the application of the method for *in situ* analysis - an essential requirement for document safety. Therefore, we shifted to the use of laboratory-based instruments with the ultimate objective of designing an *ad hoc* transportable instrument for *in situ* analysis. This effort was crowned by success: with small sources, we performed x-ray tomography of several documents, from folded single foil specimens to a 200 pages book, spanning several centuries [12, 13], including the two specimens discussed in this manuscript.

Venetian testament

In Fig.3 (top), we show the 1351 Venetian testament, a paper document (about 5x10 cm²), sealed and folded six times. This is one of the thousands of testaments stored in the Archivio that are still unopened: x-ray tomography constitutes the only chance to read them without breaking their official seals.

We performed the tomographic imaging by acquiring in <1 hour $\approx 10^3$ x-ray projection images over 360° with a relatively low photon energy (40 keV). After the image processing, we obtained 3D volumetric reconstructions like the one shown in Fig.3 (bottom), achieving a spatial resolution of $\approx 60 \mu\text{m}^3$.

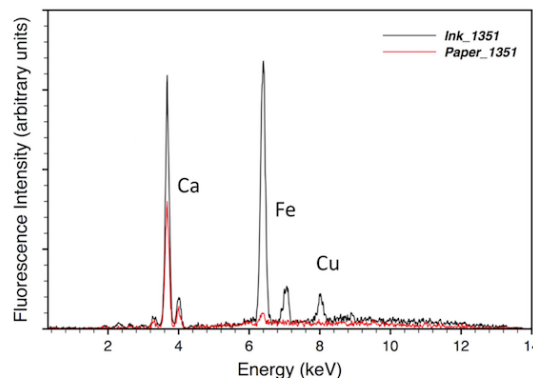
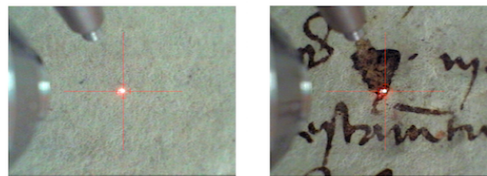


Figure 1: XRF analysis of the 1351 Venetian testament. The red spectrum was taken on the substrate point shown in the top part of the figure. The black one was taken on the ink point and clearly shows the presence of iron.

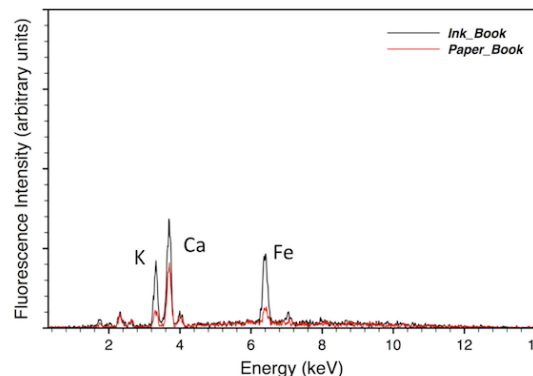
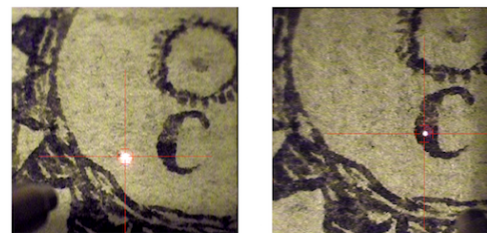


Figure 2: XRF analysis of the 18th century book for the substrate and ink points shown at the top. Once again, the ink spectrum clearly shows the presence of iron.

The tomography image in Fig.3 (bottom) clearly shows characters on the external page and the strongly absorbing "verso" of the seal. The next step of our analysis posed a formidable challenge: the segmentation of the 3D volume to separate the warped and closely-packed individual pages. This challenge was met and the 3D reconstruction was separated in six individual pages that could then be analyzed one-by-one. Figure 4 shows the individual segmented pages on the left and, on the right, parts of the texts and a complete internal page, extracted from the tomographic data. The results on the internal pages are, of course, particularly important since they provide the first glance of a text that has been inaccessible for more than six centuries.

The first lines of the reconstructed text (Fig.4, top-right) correspond to the beginning of the document with a standard formula including the date, the place and the name of the recording officer. This external portion of the text gave us the opportunity to validate the results of Virtual X-ray Reading, which reveals equally well all pages, internal and external.

Note that the segmentation yields text immune from letter superposition from other pages. The background "noise" is caused by the characteristic inhomogeneity of the handmade paper. The observable contrast fluctuation from letter to letter and between different parts of individual letters is caused by two different factors: the inhomogeneity of the ink (a mixture of a fluid binder with a non-entirely-soluble, partially precipitated iron pigment) and the mechanics of the handwriting process. These fluctuations, combined with the noisy background, augment the difficulties of the virtual reading process.

The real beginning of the testament is on the first internal page and the first words are perfectly readable in x-ray tomography: "A nome de Dio" - "In the name of God" (Fig.4 - middle-right). In the bottom-right part of Fig.4, we show one of the six reconstructed entire pages: lines of text are clearly visible and, here again, many letters and several words are directly readable, others inferable.

200-pages book

To demonstrate the real capabilities of the technique in the investigation of large objects, we investigated an 18th century handwritten book, a 200-pages physics manual (25x15x5 cm³).

We performed the tomographic analysis by acquiring $\approx 10^3$ x-ray projection images over 360° with a higher photon energy (70 keV). To improve the signal-to-noise ratio, we decided to acquire 16 frames for each projection and to use the average of them. After the image processing we obtained the 3D volumetric reconstruction with a spatial resolution of $\approx 60 \mu\text{m}^3$.

The results of Fig.5 clearly demonstrate the feasibility of using the technique for large books. The 3D visible-light visualization reproduces the external aspects of the object (Fig.5 - left), but the 3D tomographic analysis achieves much more: we can navigate through all the 200 pages, where all the writings are clearly visible and word reading is feasible (Fig.5 - right).

In particular, even if the segmentation procedure poses more severe challenges than in the previous case, the extraction of parts of text and complete readable words is proven by Fig.6.



Figure 3: 1351 Venetian testament. Top: visible images of the recto and verso of the document. Bottom: 3D tomographic reconstruction with writings on the recto clearly visible.

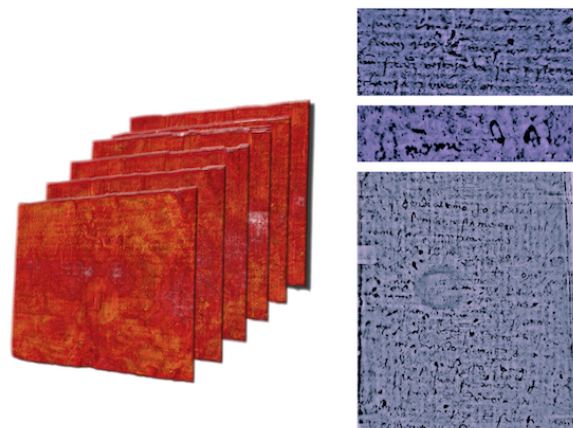


Figure 4: Left: Successful segmentation of the six pages of the Venetian testament. Right, from top to bottom: reconstructed text from the front page, a sentence from the inside and the complete reconstruction of an internal page.



Figure 5: The 18th century physics book. On the left: 3D reconstruction with visible light. On the right: a sectional view of the x-ray tomography reconstruction, where all the writings from the inside are visible.

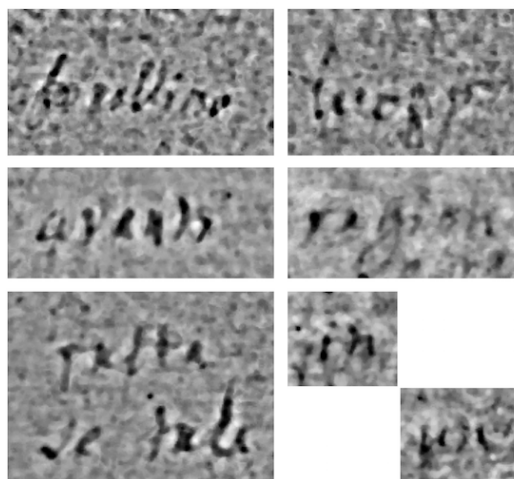


Figure 6: Clearly readable words extracted from the 3D reconstructed volume of the 200-page book.

Conclusion

The research conducted by our group in the past four years, from the extensive analysis of ancient inks to the tomography imaging of tens of handwritten documents, proved the feasibility of this technique for the digitization of ancient handwritten documents.

Notwithstanding the promising results obtained so far, a robust segmentation methodology is still needed and must be developed. Automatic separation of the pages is a fundamental requirement for the large-scale application of the technique - and a formidable challenge. Moreover, the XRF analysis and the x-ray images highlighted strong iron content fluctuations in the inks - not only from document to document but also between letters. Together with the inhomogeneity of the handmade supports, this feature clearly requires processing to reduce the noise of the images and improve the readability of the writings.

The impact of this research and the applications of the technology will extend well beyond the Archivio: the approach is applicable to the vast majority of European manuscript and, potentially, to collections worldwide.

References

- [1] F. Kaplan, The Venice Time Machine, Proceedings of the 2015 ACM Symposium on Document Engineering. (2015).
- [2] M. Bettuzzi, F. Casali, M. P. Morigi et al., Computed tomography of a medium size Roman bronze of Cupid, Appl. Phys. A, 118 (4) (2015).
- [3] A. Re, F. Albertin, C. Avataneo et al. X-ray tomography of large wooden artworks: the case study of "Doppio corpo" by Pietro Piffetti, Herit Sci 2 (2014).
- [4] M. P. Morigi, F. Casali, M. Bettuzzi et al., Application of X-ray Computed Tomography to Cultural Heritage diagnostics, Appl. Phys. A 100 (3) (2010).
- [5] M.P. Morigi, F. Casali, M. Bettuzzi et al., CT investigation of two paintings on wood tables by Gentile da Fabriano, Nucl. Instr. Meth. Phys. Res. A, 580(1) (2007).
- [6] R. Baumann, D.C. Porter, W.B. Seales, The use of micro-ct in the study of archaeological artifacts. Proceedings of 10th International Conference on NDT of Art (2008).
- [7] W.B. Seales, W. Griffioen, R. Baumann, M. Field, Analysis of Herculaneum papyri with x-ray computed tomography, Proceedings of 10th International Conference on NDT of Art (2011).
- [8] D. Mills, O. Samko, P. Rosin, K. Thomas, T. Wess, G. R. Davis, Apocalypso: revealing the unreadable, Developments in X-Ray Tomography VIII Proc. of SPIE, 8506 (2012).
- [9] A. Mocella, V. Emmanuel, B. Ferrero, C. Delattre, Revealing letters in rolled Herculaneum papyri by x-ray phase-contrast imaging, Nature communications 6(5895) (2015).
- [10] W. B. Seales, C. S. Parker, M. Segal, E. Tov, P. Shor, Y. Porath, From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi. Science Advances, 2(9) (2016).
- [11] F. Albertin, A. Astolfo, M. Stampanoni, E. Peccenini, Y. Hwu, F. Kaplan, G. Margaritondo, Ancient administrative handwritten documents: X-ray analysis and imaging. Journal of Synchrotron Radiation, 22(2) (2015).
- [12] F. Albertin, E. Peccenini, Y. Hwu, T. T. Lee, E. B. L. Ong, J. H. Je, G. Margaritondo, The Venice Archivio Di Stato: In-

novating Digitization With X-Ray Tomography, Proceedings of 2015 Digital Heritage Congress, 1 (2015).

- [13] F. Albertin, A. Patera, I. Jerjen, S. Hartmann, E. Peccenini, F. Kaplan, M. Stampanoni, R. Kaufmann, G. Margaritondo, Virtual reading of a large ancient handwritten science book, *Microchemical Journal*, 125 (2016).
- [14] D. Ruggiero, Gli inchiostri ferrogallici negli archivi e nelle biblioteche - Laboratorio di fisica dell' Istituto per il Restauro e la Conservazione del Patrimonio Archivistico e Librario (2002).

Author Biography

F. Albertin: She obtained her Ph.D. in Physics at the Ferrara University developing x-ray instruments for pigment investigations. At present, she is a member of the research staff Post-Doc at the EPFL, developing innovative technologies for digitizing ancient documents.

M. Romito: She is a Ph.D. student at the EPFL working on cochlear imaging and manuscript tomography.

E. Peccenini: She obtained her Ph.D. in Physics at the Ferrara University developing an infrared scanner for painting investigations. At present, she is a member of the research staff of the Bologna University x-ray imaging group.

M. Bettuzzi: He obtained his Ph.D in Physics at the Bologna University and at present is a member of the research staff of the Bologna University x-ray imaging group, developing x-ray tomography instruments for Cultural Heritage investigations.

R. Brancaccio: She obtained her Ph.D. in Physics at the Bologna University and at present is a member of the research staff of the Bologna University x-ray imaging group, developing algorithms for Cultural Heritage x-ray investigations.

P. Morigi: She is an Associate Professor of Physics and the director of the x-ray imaging group of the Bologna University.

M. del Rio: She obtained her degree in Classics - Byzantine Philology - at the Padova University and at present is an archivist at the Archivio di Stato in Venice.

D. Raines: She obtained her Ph.D. in History in the Ecole des Hautes Etudes en Sciences Sociales. She is an assistant professor at the Ca' Foscari University specializing in the history of libraries and archival science.

G. Margaritondo: He is an Emeritus Professor of Physics at the EPFL, where he held several managerial positions up to the vice-presidency. His fields of expertise include solid-state physics, synchrotron radiation and x-ray imaging.

D. Psaltis: He is Professor of Optics, Director of the Optics Lab. and Dean of the Engineering School at the EPFL. His current research interests include imaging, holography, biophotonics, and optofluidics.