Digital Humanities

Computational Platforms for Deciphering Ancient Greek Papyrus Fragments

Mshvidobadze Tinatin, PhD
Gori State University
Gori, Georgia

ORCID: 0000-0003-3721-9252

ABSTRACT

In 79 AD, the eruption of Mount Vesuvius instantly turned hundreds of papyrus scrolls from ancient Herculaneum into carbon-carbon fragments. The scrolls contain valuable writings by Greek philosophers, including works by the "Epicurean Philodemus". They were identified using X-ray phase contrast tomography. However, only a small part of the text hidden in the scrolls has been recovered. One of the main challenges in studying the Herculaneum papyri is their virtual unfolding, given their extremely complex structure and three-dimensional arrangement. In this paper, we discuss a computational platform for virtual unfolding and present the results of its application to two fragments of the Herculaneum papyri. The findings indicate potential for future applications and further interpretation of larger sections of text hidden in the carbonized Herculaneum papyrus. The paper also presents a computational pipeline for converting letter identifications into digital consensus transcriptions of papyrus fragments, using an interface to the Ancient Lives database. As a result, the paper explains the usefulness of the pipeline solution in the context of additional computational projects aimed at further accelerating the text fragment identification process.

Keywords: Greek texts, conservation, papyrus transcription, computer technology, crowdsourcing

DOI: https://doi.org/10.62343/cjss.2025.261

Received: March 30, 2025; Received in revised form: July 1, 2025; Accepted: July 3, 2025

INTRODUCTION

The ruins of Herculaneum offer a unique glimpse into the past with their remarkably well-preserved architecture and artifacts.

Ancient Greek carbonized papyrus scrolls survived the extraordinary conditions of the eruption of Mount Vesuvius in 79 AD and are now considered the last surviving ancient library. Although the scrolls have survived, they are unreadable due to the burning.

The paper aims to evaluate specific computational techniques for virtual unfolding and transcription.

Carbonization, due to the Vesuvius catastrophe and the natural effects of time on organic matter in the seventeenth century, has contributed to the degradation of the artifacts.

It should be noted that recently, thanks to creative technologies and innovative methods, the first word from the charred scrolls was readable - "Porphyras", or purple. Researcher Handmer described how he enhanced the ink's visibility. In doing so, he contributed to subsequent identification, including the term 'Pharythor' (Russell, 2023, pp. 145-156).

Modern conservation principles require preserving the integrity of historical documents when there is a risk of losing the information they contain. X-ray micro-computed tomography (micro-CT) provides digitization of the papyri and three-dimensional visualization of the scroll's interior. The first use of micro-CT on two intact Herculaneum papyri occurred in 2009 (Seales et al., 2011, pp. 223–235). Data analysis revealed a tortured shape, with no traces of writing. Subsequently, Mocella et al. and Bukreeva et al. used X-ray phase contrast tomography (XPCT) to study Herculaneum papyrus rolls and fragments (Mocella et al., 2015; Bukreeva et al., 2016, p. 277). The potential utility of this technique for detecting traces of hidden handwriting within the scrolls was confirmed.

More than a century ago, B.P. Grenfell and A.S. Hunt, of the University of Oxford, also discovered a vast hoard of papyri, numbering over 500,000 fragments, from the city of "Oxyrhynchus" (Bowman et al., 2007, p. 272). They began transcribing and editing the papyrus fragments, and to date, only a small portion of this vast treasure has been published. Transcribing the collection has been no easy task, as each fragment suffers from a unique level of deterioration due to different sections of the papyrus being lost or the manuscript text being illegible. The Ancient Lives project at the University of Oxford has developed a web-based interface for identifying letters on digital images of papyrus. Users can visit the Ancient Lives website and view transcriptions of ancient papyrus fragments at https://ancientlives.org. The identification of each letter and its associated features (x,y coordinates) is stored in a user identification database. To date, over 7 million letter identifications have been recorded internationally through the Ancient Lives interface.

A computational pipeline has been used to interpret large amounts of letter-identification data, automating letter identification in digital consensus transcriptions of papyrus fragments using existing sources. The paper provides an interpretation of the results found in each assessment. Several fragments from the outer part of the scroll are analyzed, and the results of virtual unfolding of the papyrus fragments obtained with the developed computational platform are discussed. The letter identification components and all additional visualizations are implemented in the Python programming environment.

METHODS

This article uses descriptive, analytical, and explanatory methods to highlight the study's important issues. To formulate the concept, we examined the views of various researchers, which we used to illustrate the reasoning and conclusions.

All changes made to integrate the algorithms seamlessly are described in detail, and the relevant data were obtained from experiments.

The open-source software SYRMEP (Brun, 2017, pp. 125-134) and Tomo Project (STP) tools were used to reconstruct the 3D sample using the XPCT filtered back projection (FBP) method (Slaney, 2012, pp. 45-66).

Due to the complex structure and 3D arrangement of the papyrus sheets, a semi-manual segmentation procedure was used.

Image segmentation and surface extraction were performed with developed in-house macros and codes using the image processing software ImageJ and Python (Rueden, 2022, p. 529).

RESULTS

In Experiment 1, the surface parameterization process was used to generate a 3D mesh from 2D coordinates. In general, this operation produces geometric deformations that are responsible for image artifacts. The final result is obtained using appropriate methods. The results allow us to verify the reliability of the parameterization method with respect to a given sample.

Experiment 2. Let us consider the operation of the computational pipeline. As an evaluation criterion, 54 published fragments from the "Oxyrhynchus" collection were selected to measure the accuracy of consensus letter identification.

After processing, the pipeline outputs two files for each papyrus fragment. The first file contains the relative fragment consensus letter identifications with x and y coordinates. The second file, which is the final output of the pipeline, contains the relative fragment consensus line sequence, which closely resembles the original papyrus fragment (see Figure 2). The consensus letter identification components, the line sequence generation component, and all additional visualizations are implemented in Python using OpenCV image processing version 3.13.

1.py - C:/Users/ASUS/OneDrive/Рабочий стол/1.py (3.13.2)

File Edit Format Run Options Window Help

```python
import cv2
import numpy as np

img = cv2.imread('eGaIy.jpg', 0)
img = cv2.threshold(img, 127, 255, cv2.THRESH_BINARY)[1]  # ensure binary
num_labels, labels_im = cv2.connectedComponents(img)

def imshow_components(labels):
    # Map component labels to hue val
    label_hue = np.uint8(179*labels/np.max(labels))
    blank_ch = 255*np.ones_like(label_hue)
    labeled_img = cv2.merge([label_hue, blank_ch, blank_ch])

    # cvt to BGR for display
    labeled_img = cv2.cvtColor(labeled_img, cv2.COLOR_HSV2BGR)

    # set bg label to black
    labeled_img[label_hue==0] = 0

    cv2.imshow('labeled.png', labeled_img)
    cv2.waitKey()

imshow_components(labels_im)
```

Figure 1. *Visualization in Python.*

Pre-processing stage

The Ancient Lives interface is directly connected to a MySQL relational database that contains all transcription information.

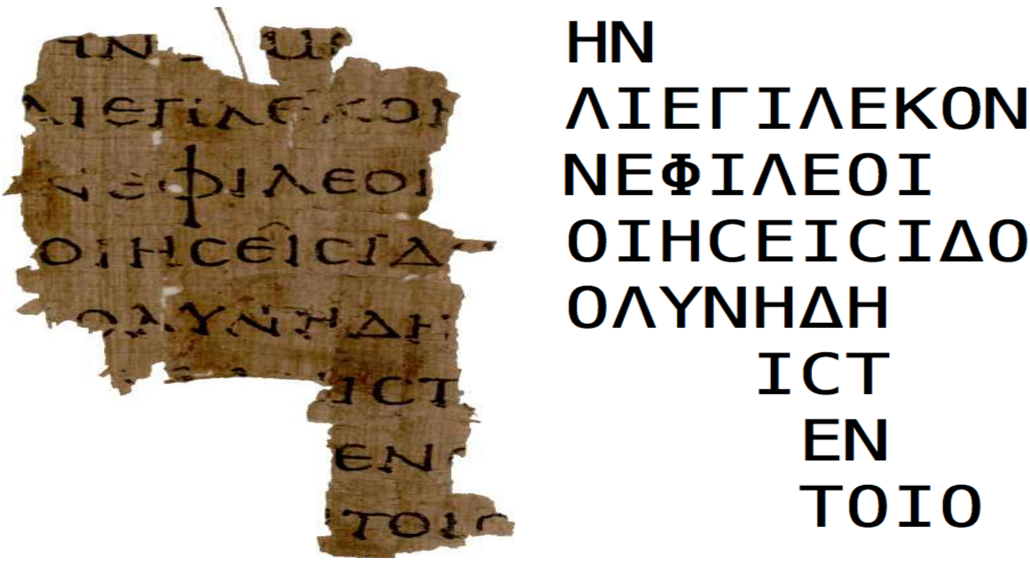

Figure 2. *An Ancient Lives fragment image (left). A digital consensus transcription for the same Ancient Lives fragment (right). Source: https://www.ancientlives.org/*

The MySQL database also reads the rotation information for each fragment.

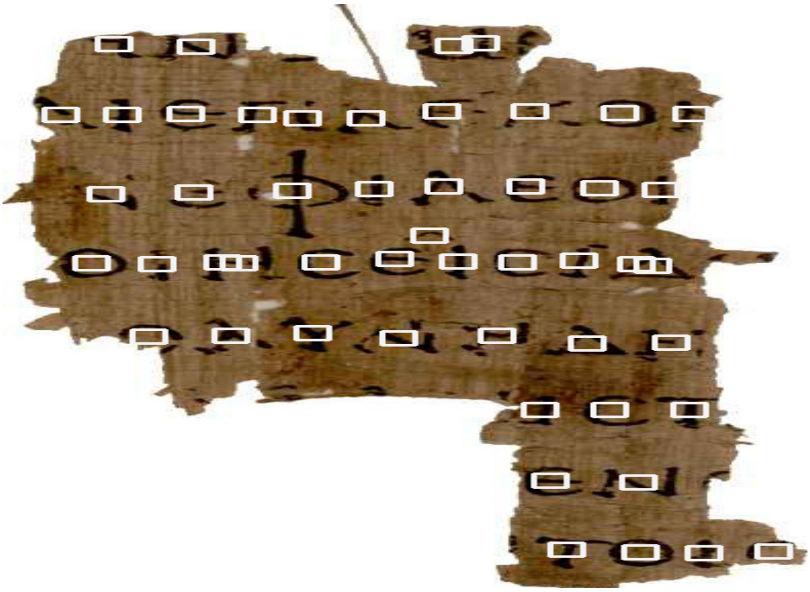

Figure 3. *A visualization of the processing performed by the pipeline using the same Ancient Lives fragment from Figure 2. Source:*

Each fragment is categorized into groups based on handwriting style and readability.

The established metrics, precision, recall, and F1 score, are used to determine the classification performance of each approach. (van Rijsbergen, 2004, pp. 365–373).

Recall is calculated by dividing the number of correct letter identifications by the total number of letter identifications relative to the fragment in the P. Oxy. transcription. Precision and recall are combined into a composite metric called the F1 score. The equation for calculating the F1 score for an individual fragment is:

$$\varphi_1 = 2\times \frac{C\times R}{C+R} \qquad (2)$$

Where C da R are the precision and recall, respectively, of the fragment click data. An F1 score of 0.0 can be interpreted as an approach to correctly classifying almost no letters. In contrast, an F1 score of 1.0 can be interpreted as an approach to correctly classifying most or all letters.

As a result, the described approach could produce duplicate lines when multiple line regions are identified from a single curvilinear line, due to incorrect measurements of the vertical spacing between line regions. To filter duplicate line regions, a final post-processing stage will remove a line region if it shares 70% or more identity with its neighboring line region.

The results of the evaluation for Stage 1 suggest that the stepwise aggregation approach achieves higher accuracy in correctly identifying consensus letters than the kernel-based approach, especially for fragments with cursive handwriting.

In addition to achieving higher accuracy, the stepwise aggregation approach has faster execution time than the kernel-based approach.

DISCUSSION

The main objective of the work was to develop a computational platform to perform virtual segmentation and flattening of a single sheet of a carbonized papyrus fragment.

Experiment 1. Sara Stabile et al., performed XPCT and micro-CT experiments to reconstruct three-dimensional images of two historical papyrus fragments, PHerc.1103 and PHerc (Stabile et al., 2021, p.143-147).

PHerc.1103 was scanned using synchrotron-based XPCT (Synchrotron radiation-based X-ray phase-contrast tomography imaging is an innovative modality for the quantitative analysis of three-dimensional morphology), while PHerc.110 was scanned using micro-CT using a laboratory source. According to technical expertise, the fragments belonged to different scrolls from Herculaneum. Both fragments have a highly compact internal structure and consist of multiple papyrus layers glued together, which cannot be separated by hand.

The most popular angle-preserving parameterization methods are used: least-squares conformal mapping (LSCM) and angle-based flattening (ABF). (Lévy et al., 2002; Sheffer et al., 2011, p.326–337).

A surface parameterization process is used to represent the 3D mesh in 2D coordinate space.

Recently, various digital procedures have been proposed to provide a digital version of the hidden text inside the rolled papyri, beginning with the non-invasive acquisition of their images. Nevertheless, the structure of the carbonized Herculaneum Papyrus is quite different from that of a conventional roulette wheel, where the standard "virtual scrolling" algorithm is used.

Future developments in specialized software may further enhance the ability to read the text from the Herculaneum Papyrus without actually reading it. The success lies in adapting the basic principles of XPCT, such as "virtual scrolling," to the Herculaneum Papyrus and, consequently, to virtually open, read, and decipher parts of the text hidden within the unrolled papyrus.

X-ray phase contrast tomography is based on absorption and is a well-known tool for imaging the internal structure of thick objects with hard X-rays.

Standard absorption tomography is an inadequate technique for distinguishing details of similar density, for example, distinguishing carbon-fiber-based papyrus foil from carbon-based ink used for writing on papyrus. In this study, better contrast was achieved by imaging the phase modulation induced by the object in a coherent or partially coherent beam. There are several experimental approaches to detecting phase contrast with X-rays. A series of experiments was carried out at ID17 of the European Synchrotron Radiation Facility (ESRF) in Grenoble (F), using a free-space propagation setup.

The results of these experiments allow us to read and decipher letters and numbers previously written on a papyrus roll and to optimize the experimental conditions. Due to the flow of CO2, this complex internal structure was ideal for developing and testing new numerical algorithms for "virtual" unrolling and for adjusting and optimizing them to reveal the hidden writing.

XPCT experiments at PHerc. 375 and PHerc. 495, were performed using a monochromatized incident X-ray beam with an energy of 73 keV.

Image processing. 3D rendering and segmentation of the impurities were performed using Visage Amira 6 (an interactive system from FEI) and VolView (an interactive volume visualization system from Kitware Inc.). The 3D structure of the papyri was obtained through a segmentation procedure. Images were generated by associating colors with different grayscale ranges in the 3D model.

Virtual Unrolling. In recent years, the problem of virtual unrolling using volumetric scanning in the virtual restoration and preservation of ancient artifacts, such as parchments or papyri, has been a highly active area of research. The use of X-ray computed tomography (CT) and X-ray microcomputed tomography for data digitization has facilitated the development of restoration algorithms. Various digital procedures have been proposed to provide a digital version of the text hidden within the rolled scrolls, starting with non-invasive acquisition.

Digital restoration. Surface flattening and rolling can be interpreted as an isometric mapping (i.e., maintaining distances) from 3D to 2D images, which reduces text distortion in the parchment.

Various approaches have been proposed to solve these problems. One of the most promising is described by Seales and colleagues. The authors developed software that combines the flattening and rolling functions based on the simulation of the surface of the mass source. Algorithms proposed by O. Samko et al. allow solving the problem of contact points between adjacent sheet layers (Samko, 2014, pp. 248–259).

The internal structure of the Herculaneum papyrus is quite different from a conventional roll, where a simple "virtual rolling" was used. However, the complex internal arrangement of the chaotic bundles of layers still consisted of layers that were similarly bent, torn, and twisted. The central part of the scroll is better preserved, and the papyrus layers are separated and loosely wound. The outer part of the papyrus is tightly rolled and stacked. It is clear that the virtual detection of the text requires different approaches and has its own challenges.

According to the experiment, the letters can be deformed due to thermal and pressure effects and due to the irregular surface of the fibers. It is not possible to know a priori the aspect of the letters and even whether there is text in the analyzed part of the rolls.

The unsigned procedure for detecting hidden text was implemented using computational algorithms in MATLAB and various commercial software. Phase-contrast tomography provides a 3D volume of the books, revealing the Greek text hidden inside.

This method uses non-destructive X-ray imaging techniques, such as micro-CT and XPCT, to digitize fragile artifacts without damaging them, and employs a computer algorithm that virtually unfolds the Herculaneum papyrus fragments. Our analysis suggests that this technique is suitable for processing the scorza, the most damaged yet valuable part of the papyrus.

Searching for text within papyrus is the ultimate goal of virtual investigation. The widespread use of carbon-based black ink in the Herculaneum papyri is one of the most challenging issues in writing.

In terms of scientific accuracy, the analysis of the texts has revealed that XPCT has provided image contrast for lighter materials such as carbon. The technique can detect slight variations in density that distinguish areas of carbon ink from carbonized papyrus sheets. However, future XPCT experiments with higher spatial resolution are needed to differentiate text from papyrus structures reliably.

Choosing the proper flattening procedure is essential, as it helps distinguish between writing and papyrus and avoid misinterpretation of the data.

The segmentation procedures were carried out using specially developed ad hoc computational programs and free software.

For these purposes, we have developed and offer an accessible algorithm that can be reproduced, optimized, and improved.

Let us assume that the object consists of a single quasi-uniform material with a constant ratio between the real and imaginary parts of the refractive index.

(ϛ/γ). With this condition, we develop the transport intensity equation (TIE) to investigate the phase:

$$\boldsymbol{\theta}(\boldsymbol{x}) = \frac{1}{2} \ln \left(PT^{-1} \left\{ \frac{PT[I_N(x)/I_0(x)]}{\frac{\gamma}{c} + |f|^2 \left(\frac{\vartheta M}{4p}\right)} \right\} \right) \quad (1)$$

Where PT and PT^{-1} denote forward and inverse Fourier transformation, respectively, $f = (f_x,\ f_y)$ re the Fourier coordinates, $I_N(x)/I_0(x)$ is the normalized intensity detected at distance D.

In this experiment, which was carried out at ID17 of the European Synchrotron Radiation Facility in Grenoble, PHerc. 1103 was measured with a monochromatic incident X-ray energy of 80 keV, selected with a Si(111) monochromator.

A two-step image reconstruction was applied to the XPCT experimental data: phase retrieval and tomographic reconstruction.

CONCLUSION

In this paper, we present a novel computational pipeline for decoding transcriptions of ancient Greek papyri represented in digital images that closely resemble the original papyrus format. Palynologists can use the digital consensus transcriptions produced by the pipeline to explore, modify, and publish fragments with confidence more quickly.

The Ancient Lives project was implemented to help palynologists transcribe and evaluate fragments more quickly.

Several complementary tools have been developed to support and accelerate the transcription process, such as Greek-BLAST (Williams, 2020), a popular genetic sequence alignment tool specifically designed to identify literary papyrus fragments.

Consensus transcriptions of literary fragments generated by Ancient Lives can be directly fed into Greek-BLAST and rapidly aligned against databases of ancient Greek literary manuscripts (i.e., the Perseus Digital Library (Smith, 2021, pp. 15-25).

In addition, researchers at the University of Minnesota have developed a web-based tool to curate digital consensus transcriptions generated by the computational pipeline rapidly. Selected consensus transcriptions are stored in a database for later data mining.

Finally, the consensus transcriptions produced through Ancient Lives serve as the basis for many fragments that will be further studied, edited, and published in The Oxyrhynchus Papyri series and on Proteus. This new interactive, web-based platform uses advanced computational methods and techniques for both the study and analysis of ancient texts and the subsequent generation of digital texts.

A key component of future work is to improve the pipeline's line sequencing phase. The final redesign of the pipeline will be completed after additional classification information (e.g., methods for identifying linear information or missing papyri) is incorporated into the Ancient Lives framework.

Ethics Approval and Conflict of Interest

This study was conducted in accordance with relevant ethical standards. The authors declare that there are no financial, personal, professional, or institutional conflicts of interest that could have influenced the design, conduct, interpretation, or publication of this work.

Financing

The research was carried out without financial aid.

Declaration of competing interest

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

REFERENCES

Brun, F., et al. (2017). Syrmep tomo project: A graphical user interface for customizing reconstruction workflows. *Advanced Structural and Chemical Imaging*, *3*, 125–134.

Bowman, A., Coles, R., Gonis, N., Obbink, D., & Parsons, P. (2007). *Oxyrhynchus: A city and its texts* (Vol. 93). Egypt Exploration Society.

Bukreeva, I., et al. (2016). Virtual unrolling and deciphering of Herculaneum papyri by X-ray phase-contrast tomography. *Scientific Reports*, *6*, 277. https://doi.org/10.1038/srep27227

Kak, A., & Slaney, M. (2012). *Principles of computerized tomographic imaging* (pp. 45–66). Society for Industrial and Applied Mathematics.

Lévy, B., Petitjean, S., Ray, N., & Maillot, J. (2002). Least squares conformal maps for automatic texture atlas generation. *ACM Transactions on Graphics*, *21*, 362–371.

Mocella, V., Brun, E., Ferrero, C., & Delattre, D. (2015). Revealing letters in rolled Herculaneum papyri by X-ray phase-contrast imaging. *Nature Communications, 6*, 5895. https://doi.org/10.1038/ncomms6895

van Rijsbergen, C. J. (2004). Foundations of evaluation. *Journal of Documentation*, *30*(4), 365–373.

Rueden, C. T., et al. (2022). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, *23*, 529. https://doi.org/10.1186/s12859-017-1934-z

Russell, M. (2023). First word from “unreadable” scrolls that survived Vesuvius has been decoded. *IFLScience*. (Online article; page numbers omitted in APA for web publications.)

Samko, O., Lai, Y. K., Marshall, D., & Rosin, P. L. (2014). Virtual unrolling and information recovery from scanned, scrolled historical documents. *Pattern Recognition*, *47*, 248–259. https://doi.org/10.1016/j.patcog.2013.06.015

Seales, B. W., Griffioen, J., Baumann, R., & Field, M. (2011). Analysis of Herculaneum papyri with X-ray computed tomography. In *Proceedings of the International Conference on Nondestructive Investigations and Microanalysis for the Diagnostics and Conservation of Cultural and Environmental Heritage* (pp. 223–235).

Sheffer, A., & de Sturler, E. (2011). Parameterization of faceted surfaces for meshing using angle-based flattening. *Engineering with Computers*, *17*, 326–337. https://doi.org/10.1007/PL00013391

Smith, D., Rydberg-Cox, J., & Crane, G. (2021). The Perseus Project: A digital library for the humanities. *Literary and Linguistic Computing*, *15*(1), 15–25.

Stabile, S., Palermo, F., Formoso, V., Bartolino, & Cedola. (2021). A computational platform for the virtual unfolding of Herculaneum papyri. *Scientific Reports*, *11,* 143–147.

Williams, A., Carroll, H., Wallin, J., Brusuelas, J., Fortson, L., Lamblin, A., & Yu, H. (2020). Identification of ancient Greek papyrus fragments using genetic sequence alignment algorithms. In *Proceedings of the 1st Workshop on Digital Humanities and e-Science*.