Improving Palimpsest Readability via Image Preprocessing: An Investigation into Adjustment Techniques

Katia Rasheva-Yordanova[1[0000-0001-5859-9114]], Georgi P. Dimitrov[1, 2[0000-0002-4785-0702]], Paulina T. Tsvetkova[1, 2[0000-0001-5637-1925]], Milena Bankovska[1[0009-0001-9003-1024]], Pavel S. Petrov[3[0000-0002-1284-2606]]

[1] Department of Information Technologies, University of Library Studies and Information, 119, Tsarigradsko Shose, Sofia, Bulgaria
[2] Institute of Robotics, Bulgarian Academy of Sciences, Acad. Georgi Bonchev Str., Bl. 2, Sofia, Bulgaria
[3] Faculty of Information Technologies, University of Economics – Varna, 77 Kniaz Boris I Blvd., Varna, Bulgaria
k.rasheva@unibit.bg, geo.p.dimitrov@gmail.com, p.tsvetkova@unibit.bg, m.bankovska@unibit.bg, petrov@ue-varna.bg

Abstract. The palimpsests represent unique historical sources that hold the potential for new insights in the field of human history. These manuscripts, rewritten and reused over time, pose challenges in the research related to their readability and interpretation. The present study aims to investigate the readability of palimpsests through the use of image preprocessing techniques. The article focuses on methods for the preprocessing of palimpsests that could lead to a significant improvement in the readability of the 'hidden' text. The challenges encountered during the processing of palimpsests are explored and various techniques applicable to improving the readability of these manuscripts are analyzed.

The primary goal of preprocessing in this context is to separate the 'hidden' text from the visible one, neutralizing material defects and aging. The article presents specific methods such as extracting specific color range from a palimsest image. The experimental techniques are highlighted with sample codes illustrating the application of the respective technology. The current research attempts to advance the development of methods for processing palimpsests and opens up new perspectives for extracting information from those historically valuable manuscripts.

Keywords: AI, Palimpest, Manuscript, Image processing.

1 Introduction

Information and Communication Technology (ICT) utilized in tourism helps surmount cultural and language obstacles, ensuring accessibility to cultural heritage (Kaposi et al., 2013). Different papers describe how ICT functionalities could present Bulgarian cultural and science resources via digital libraries and mobile applications (Luchev et al., 2016; Márkus et al., 2015; Márkus et al., 2016; Pavlov et al., 2010; Pavlova-Draganova et al., 2007). Some authors even use VR technologies to help tourist in choosing their destinations (Nagy et al., 2016).

The etymology of the word "manuscript" comes from the Latin "manu scriptum," which literally translates to "written by hand." With the development of the first scripts and alphabets in various parts of the world, diverse materials and surfaces have been used for writing. In the past, these materials were so scarce that they were often reused multiple times. This phenomenon gives rise to a new, layered artifact known as a palimpsest, where the process of erasing the original manuscript is not entirely effective, allowing the original to often be discerned beneath the surface inscriptions. Collections of valuable palimpsests are widespread in historical archives worldwide and attract the interest of many researchers from various scientific fields. This is because ancient manuscripts can contain a significant amount of additional information beyond the visible texts. This information is not easily discernible to the naked eye, posing certain challenges for analysts, primarily depending on the set research objectives. Overcoming these challenges requires a specific approach tailored to the respective needs.

The main goal of studying palimpsests is associated with the restoration of the original text in the lower layer. This task poses a challenge on one hand due to the deterioration of the pigments and ink colors resulting from the aging of manuscripts, and on the other hand, due to the overlay of two or more text layers on top of each other. Reconstructing the text from the lower layer takes a significant amount of time and is costly, as it often requires the intervention of highly qualified specialists. This makes deciphering palimpsests an even more challenging task.

This is why an increasing number of contemporary researchers (Starynska, A., Messinger, D., & Kong, Y., 2021), (Easton, Roger & Knox, Keith & Christens-Barry, William & Boydston, Ken., 2018), , are seeking new and non-invasive methods and techniques suitable for studying palimpsests with the aim of acquiring new knowledge, information, and facts from the past , .

Over the past twenty years, various techniques and methods for analyzing palimpsests have emerged (Brenner & Miklas, 2019; Brenner et all, 2019; Easton et all, 2018) allowing us to obtain additional information about the history of ancient manuscripts, as well as insights into the living conditions of that time. Digital technologies play a key role in their study, providing new opportunities for understanding, preserving, and accessing this type of cultural heritage . However, a more valuable aspect is that through digitization, it becomes possible to study the hidden information in manuscripts. Often, this requires significant preprocessing.

The significance of image preprocessing for the legibility of palimpsests is crucial within the context of their uniqueness and complexity. Preprocessing plays a key role in separating the "hidden" text from the visible, providing clarity in the study of palimpsests. This process neutralizes defects and aging in the manuscripts, improving the quality and readability of the images. By enhancing contrast and clarity, preprocessing facilitates the comprehension of the text. The experimental methods and techniques involved in this process allow for testing and optimizing approaches to achieve the best results in the investigation of palimpsests, uncovering new perspectives for extracting information from these historically valuable manuscripts.

The aim of this article is to explore and present methods and techniques for preprocessing, with a focus on improving the readability of the "hidden" text in palimpsests, addressing challenges related to their nature and condition.

2 Exposition of the Investigation

2.1 Object of the Study

The research in our case was done at the National Library in Sofia as the particular palimsest is the Dragothian Mineum, a manuscript from the beginning of the 12th century, is preserved as a palimsest in the Dragothian Apostle (manuscript NBCM No. 880). It contains text in Cyrillic and Glagolitic. The Glagolitic in this minae is the second layer, which is in practice largely hidden by the newer Cyrillic inscription. A captured page from Dragothian Mineum is shown further down (Fig. 1.).

Fig. 1. Page from particular palimsest - Dragothian Mineum.

The goal we have set in our case was to make the individual elements of the manuscript accessible to researchers and or computer programs working at semantic level. This entails tasks such as recognizing handwritten characters or discerning whether the text was authored by a single individual or multiple persons, achieved through the processing of captured images using specialized cameras (Dyer, 2013). For this purpose, the captured images should be processed in a way that maximizes the contrast between the elements of interest, such as letters and background, as distinct as possible.

2.2 Why Preliminary Preparation of Captured Palimpsests Is Necessary

Due to the disappearance of colors in the palimpsests, sophisticated techniques are needed to restore the colors and contrast lost when the previous text was erased. Additionally, the original palimpsests are of animal skin, many faded colors, making processing even more difficult. Also, a new text has been written over the original text. One of the most common methods used to improve contrast is standard histogram equalization. This method enhances the image contrast globally by distributing the most common intensity values. Unfortunately, it often introduces additional noise to the captured palimpsest image. Additionally, a third layer is frequently detected in a significant portion of palimpsests. This layer is typically practically invisible to the naked eye.

To overcome the noise and enhance the contrast of the letters in the text, our study proposes the use of contrast-limited adaptive histogram equalization (CLAHE) for pre-clearing artifacts. We sequentially apply both methods, followed by additional image processing using Adobe Illustrator.

In this paper, we propose a novel method that integrates the classical technique of noise reduction using cv2.fastNlMeansDenoising which employs the Non-local Means Denoising algorithm (http://www.ipol.im/pub/algo/bcm_non_local_means_denoising/ with several computational optimizations.) with CLAHE (contrast-limited adaptive histogram equalization). These methods are applied sequentially to the image to achieve improved results. The primary objective is to initially reduce inherent natural noise, followed by enhancing contrast. Subsequently, denoising is applied to mitigate the significant noise introduced by CLAHE, facilitating the subsequent process of deciphering the ancient manuscript.

2.3 Using UW Light in Image Capture

In the present investigation UV images were selected for analysis due to their inherent advantageous attributes. UV images often exhibit higher contrast compared to visible light images, which can facilitate better detection and segmentation of features (https://midopt.com/solutions/monochrome-imaging/ultraviolet-imaging/). UV imaging can minimize background noise from ambient light sources, leading to clearer and more precise image analysis (Sellami & Rhinane, 2023). Fig. 2. shows the page from Fig. 1, but captured with UV light. In the current study we observed that UV images of palimpsests exhibit better performance compared to infrared images.

2.4 Denoising

Denoising of an image refers to the process of reconstruction of a signal from noisy images. Denoising is done to remove unwanted noise from image to analyze it in better form. It refers to one of the major pre-processing steps. There are four functions in opencv which are used for denoising of different images. In our case we have used cv.fastNlMeansDenoising() – that technique works with a single grayscale image. The grayscale image helps in simplifying algorithms and eliminate the complexities associated with computational demands. That image improves visual clarity by distinguishing between shadow nuances and highlight within the images due to its primarily two-dimensional (2D) nature, as opposed to three-dimensional (3D). The original and the denoised images are shown in figure 3,4. The grayscale conversion proves particularly beneficial for captured images where detailed color matching is unnecessary (https://www.isahit.com/blog/why-to-use-grayscale-conversion-during-image-processing).

Fig. 2. Page from Fig. 1., but with UV light.

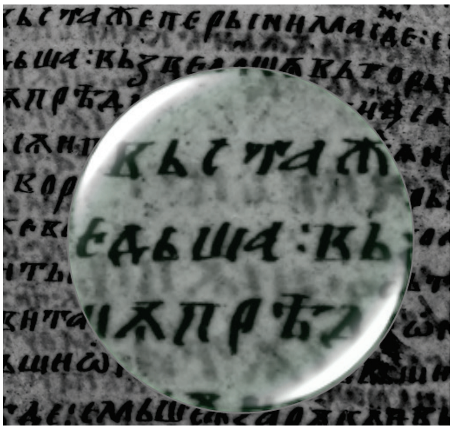

Fig. 3. Original image.

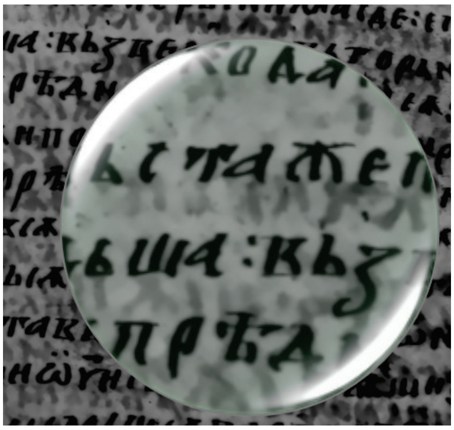

Fig. 4. Denoised image.

2.5 CLAHE (Contrast Limited Adaptive Histogram Equalization)

One image improvement technique based on the histogram equalization method is CLAHE. CLAHE can solve the problem of increasing contrast that is too over, namely by giving the boundary value on the histogram. This limit value is called the clip limit which states the maximum height of a histogram. How to calculate the clip limit of a histogram can be defined by the Equation (1) below:

$$\beta = \frac{\mathrm{M}}{\mathrm{N}}\left(1 + \frac{\alpha}{100}(S_{max} - 1)\right) \quad (1)$$

The variable M represents the region size, N represents the grayscale value (256), and α is a clip factor indicating the addition of a histogram, which ranges from 0 to 100. The process of CLAHE begins with initializing the region size and clip limit, followed by shaping the histogram for each region. Next, the histogram is clipped according to the clip limit, and any excess values (considered to be excess) are redistributed. The new histogram is then mapped to the image, and pixel interpolation is performed in neighboring regions. This process results in the generation of the CLAHE image.

Fig. 5. CLAHE.

Fig. 6. Denoised after CLAHE.

The output of CLAHE applied on RGB color model (Fig. 5) and Denoised after CLAHE (Fig. 6)

3 Results

The initial phase of the present investigation involved the acquisition of an ancient palimpsest employing infrared (IR) and ultraviolet (UV) light (Fig. 7. and Fig. 8). We observed that UV light provided better contrast between different layers of text on the palimpsest. This could be due to the fact that UV light imaging can offer higher resolution compared to infrared, allowing for finer details to be captured. This is particularly important when dealing with faded text, like the one in our case, where every detail matters in deciphering the content accurately. In addition, UV light is absorbed differently by various materials, making it useful for selective imaging. Certain pigments used in ancient inks may fluoresce under UV light, while others remain relatively unaffected. This selective absorption can help in isolating specific layers of text or imagery. The image was then converted in a grayscale image for better visual clarity and easier analysis, as stated earlier (Fig. 9.). The palimpsest may contain noise or artifacts that obscure the underlying information, so we applied noise reduction algorithms. They are used to remove the unwanted elements while preserving the essential details of the text. Furthermore, the use of the grayscale image additionally helped reduce the impact of noise present in individual color channels. By combining the information from multiple channels into a single intensity channel, the resulting image exhibited reduced noise, leading to improved processing results. We observed that the clarity of the underlying writings was enhanced without damaging the integrity of the original image, as shown in (Fig. 10.).

Fig. 7. Original image – infrared light.

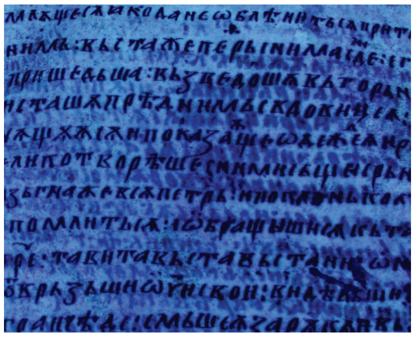

Fig. 8. Original image – UV light.

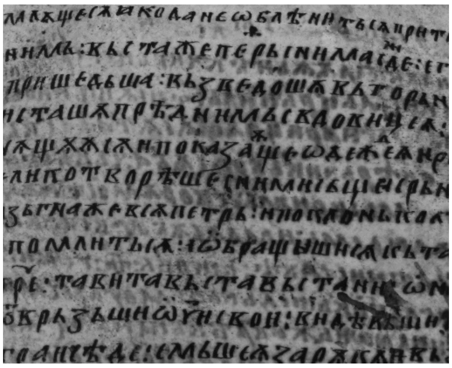

Fig. 9. Original image – Gray scale.

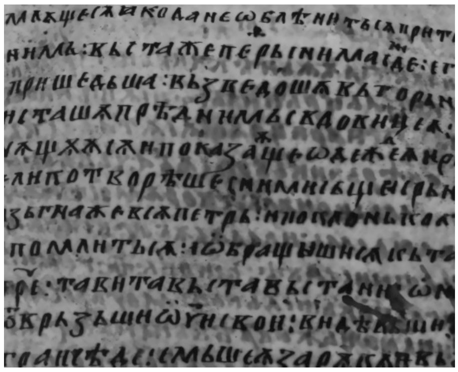

Fig. 10. Image after first denoise.

After that we used Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve the contrast of the image (Fig. 11.) CLAHE works by enhancing local contrasts and bringing out the details in regions that are darker or lighter than the surrounding areas, while preventing over-amplification of noise. CLAHE divides the image into smaller regions and applies histogram equalization locally, allowing for better control over contrast enhancement. We found out that CLAHE was beneficial for enhancing the visibility of details in the palimpsest with uneven lighting and low contrast. While CLAHE can improve contrast and visibility, it may also introduce some degree of noise or artifacts, especially in regions with low texture or detail. To address this, we applied additional noise reduction techniques on the image affected by CLAHE enhancement (Fig. 12.) These techniques minimized noise while preserving the benefits of contrast enhancement achieved through CLAHE.

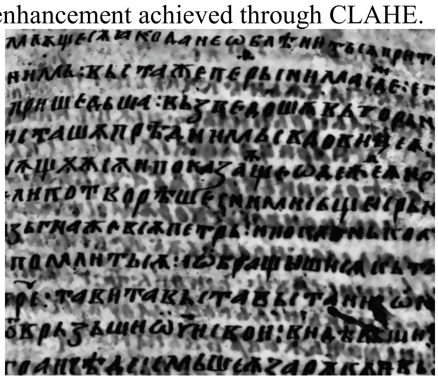

Fig. 11. Image after preprocesing with CLAHE algorithm.

Fig. 12. after preprocesing with CLAHE algorithm and second denoise.

Since denoising algorithms may inadvertently remove fine details in an image along with noise, a second CLAHE pass can help enhance these fine details, making the image appear sharper (Fig. 13.). CLAHE is adaptive in nature, meaning it adjusts its parameters locally based on the image content. By applying CLAHE after denoising, we ensured that the process of fine detail enhancement took into account the denoised characteristics of the image, resulting in more accurate and visually pleasing image. After preprocessing the image, we wanted to isolate specific color channels. For that purpose we used Photoshop (Fig. 14.) Software tools like Photoshop allow for the separation of color channels and that enabled us to enhance individual components such as red, green and blue. In addition, Photoshop helped for the isolation of specific wavelengths of light, which further enhanced the visibility of the palimpsest's text.

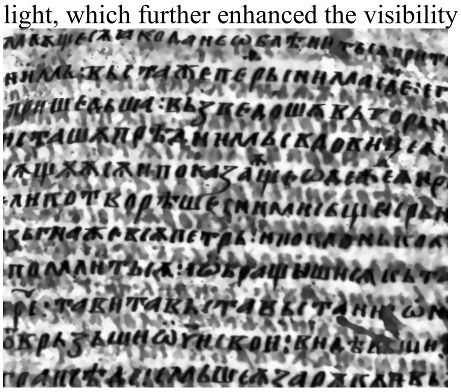

Fig. 13. Image after preprocesing with second CLAHE algorithm.

Fig. 14. Image after restore color.

4 Conclusion

In conclusion, the process of capturing and enhancing images of ancient palimpsests involved a series of sophisticated techniques aimed at revealing and preserving valuable historical information. By utilizing ultraviolet light, we managed to unveil hidden layers of text and that provided insights to us into historical documents that were not previously investigated. Subsequent denoising of the images helped us to remove unwanted artifacts while retaining essential details. Contrast enhancement through CLAHE improved visibility, although it introduced some noise, which was mitigated through additional noise reduction methods subsequently. Finally, software tools like Photoshop facilitated further analysis by allowing us to isolate and manipulate color channels, aiding in the interpretation and restoration of ancient artifacts. These techniques collectively enabled the exploration of ancient manuscripts, enriching our understanding of history and cultural heritage.

Acknowledgements.

This paper is supported by the Science and Research Fund at the Ministry of Education and Science of Bulgaria, Project КП-06-Н60-09/2021: “Interdisciplinary Methods and Means for Research of Manuscript Monuments” and the Erasmus+ project "The Future is in Applied Artificial Intelligence - FAAI - 2022-1-PL01-KA220-HED-000088359". We also acknowledge the valuable help of our colleagues from The Centre of Image and Material Analysis in Cultural Heritage – Austria (https://cima.or.at).

References

Starynska, A., Messinger, D., & Kong, Y. (2021). Revealing a history: palimpsest text separation with generative networks. *International Journal on Document Analysis and Recognition*, Volume 24, Issue 3,Sep 2021, pp 181–195. doi:https://doi.org/10.1007/s10032-021-00379-z

Brenner S., & Miklas H. (2019). Multispectral Imaging of Degraded Manuscripts in the Ivan Vazov National Library in Plovdiv. *odišnik na Narodna biblioteka „Ivan Vazov“ (2014-2018)*, 32-48.

Brenner S., Sablatnig R., Cappa F., Vetter W., Frühmann B., Schreiner M., & Miklas H. (2020. Virtual Conservation and Restoration via Multispectral Imaging and Spectroscopy. In M.V. Korogodina (Ed.) *Issledovanie i restavraciâ rukopisej: materialy konferencii - 2019 [Sankt-Peterburg, 17–18 sentâbrâ, 2019 g.]* (pp. 15-26) [Manuscript research and restoration: conference materials - 2019 [St. Petersburg, September 17–18, 2019] (pp. 15-26)]. SPb.: BAN. https://spbiiran.ru/issledovanie-i-restavracziya-rukopisej-sbornik-materialov-nauchnoj-konferenczii-2019-goda/

Dyer, J. V. (2013). *Multispectral Imaging in Reflectance and Photo-induced Luminescence modes: a User Manual.* https://api.semanticscholar.org/CorpusID:134752316

Easton, R. L. Jr., Knox, K. T., Christens-Barry, W. A., & Boydston, K. (2018). Spectral Imaging Methods Applied to the Syriac Galen Palimpsest. *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies 3*(1), 69-82. https://doi.org/10.1353/mns.2018.0003

Kaposi, G., Szkaliczki, T., Márkus, Z. L., Luchev, D., Goynov, M., & Paneva-Marinova, D. (2013). Mobile Exploring of the Bulgarian Iconography through QR Codes in the GUIDE@HAND Tourist Guide Application. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, *3*, 44–52. https://doi.org/10.55630/dipp.2013.3.4

Luchev, D., Paneva-Marinova, D., Pavlov, R., Kaposi, G., Márkus, Z., Szántó, G., Szkaliczki, T., & Veres, M. (2016). Game-based Learning of Bulgarian Iconographical Art on Smart Phone Application. In *Proceeding of the International Conference on e-Learning'16, September, 2016, Bratislava, Slovakia:* (pp. 195-200). University of Ruse.

Márkus, Z. L., Kaposi, G., Szántó, G., Szkaliczki, T., Veres, M., Weisz, Z., Luchev, D., Paneva-Marinova, D., Pavlov, R., & Pavlova, L. (2016). Mobile Exploring of Bulgarian Cultural and Scientific Assets. *Serdica Journal of Computing, 10*(2), 93-124. https://doi.org/10.55630/sjc.2016.10.93-124

Márkus, Z. L., Kaposi, G., Szkaliczki, T., Luchev, D., & Pavlov, R. (2015). BOOK@HAND BIDL: Mobile Exploring of the Bulgarian Iconography by Using Panorama Pictures. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, *5*, 109–120. https://doi.org/10.55630/dipp.2015.5.10

Nagy, J., Márkus, Z. L., Kaposi, G., Szántó, G., Szkaliczki, T., & Vass, N. (2016). New Tourist Service Based on Virtual Reality Glasses in the Town of Miskolc, Hungary. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, *6*, 81–92. https://doi.org/10.55630/dipp.2016.6.8

Pavlov, R., Paneva-Marinnova, D., Goynov, M., & Pavlova-Draganova, L. (2010). Services for Content Creation and Presentation in an Iconographical Digital Library. *Serdica Journal of Computing, 4*(2), 279-292. https://doi.org/10.55630/sjc.2010.4.279-292

Sellami, E. M., & Rhinane, H. (2023). A New Approach for Mapping Land Use / Land Cover Using Google Earth Engine: A Comparison of Composition Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W6-2022, 343–349. https://doi.org/10.5194/isprs-archives-xlviii-4-w6-2022-343-2023

Received: March 15, 2024
Reviewed: April 07, 2024
Finally Accepted: May 10, 2024