Kodikologie und Paläographie im digitalen Zeitalter 3

Codicology and Palaeography in the Digital Age 3

Schriften des
Instituts für Dokumentologie und Editorik

herausgegeben von:

Band 10

Schriften des Instituts für Dokumentologie und Editorik — Band 10

Kodikologie und Paläographie im digitalen Zeitalter 3

Codicology and Palaeography in the Digital Age 3

herausgegeben von | edited by

Oliver Duntze, Torsten Schaßan, Georg Vogeler

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Johanna Puhl, Patrick Sahle

2015
BoD, Norderstedt

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.d-nb.de/** abrufbar.

Herstellung und Verlag: Books on Demand GmbH, Norderstedt
ISBN: 978-3-7347-9899-3
Einbandgestaltung: Johanna Puhl, basierend auf dem Entwurf von Katharina Weber
Satz: LaTeX, Bernhard Assmann und Torsten Schaßan

Inhaltsverzeichnis

Appendices

Readability Enhancement and Palimpsest Decipherment of Historical Manuscripts

Fabian Hollaus, Melanie Gau, Robert Sablatnig
William A. Christens-Barry, Heinz Miklas

Abstract

This paper presents image acquisition and readability enhancement techniques for historical manuscripts developed in the interdisciplinary project "The Enigma of the Sinaitic Glagolitic Tradition" (Sinai II Project).[1] We are mainly dealing with parchment documents originating from the 10th to the 12th centuries from St. Catherine's Monastery on Mount Sinai. Their contents are being analyzed, fully or partly transcribed and edited in the course of the project. For comparison also other mss. are taken into consideration. The main challenge derives from the fact that some of the manuscripts are in a bad condition due to various damages, e.g. mold, washed out or faded text, etc. or contain palimpsest (=overwritten) parts. Therefore, the manuscripts investigated are imaged with a portable multispectral imaging system. This non-invasive conservation technique has proven extremely useful for the examination and reconstruction of vanished text areas and erased or washed off palimpsest texts. Compared to regular white light, the illumination with specific wavelengths highlights particular details of the documents, i.e. the writing and writing material, ruling, and underwritten text. In order to further enhance the contrast of the degraded writings, several Blind Source Separation techniques are applied onto the multispectral images, including Principal Component Analysis (PCA), Independent Component Analysis (ICA) and others. Furthermore, this paper reports on other latest developments in the Sinai II Project, i.e. Document Image Dewarping, Automatic Layout Analysis, the recent result of another project related to our work: the image processing tool *Paleo Toolbar*, and the launch of the series *Glagolitica Sinaitica*.

Zusammenfassung

In dieser Arbeit werden Bildaufnahmemethoden sowie Methoden zur Verbesserung der Lesbarkeit für historische Manuskripte vorgestellt, die im Rahmen des interdisziplinären Projekts "The Enigma of the Sinaitic Glagolitic Tradition" (Sinai II Projekt)

[1] <http://www.caa.tuwien.ac.at/cvl/research/sinai/>

Kodikologie und Paläographie im Digitalen Zeitalter 3 – Codicology and Palaeography in the Digital Age 3. Hrsg. Oliver Duntze, Torsten Schaßan, Georg Vogeler unter Mitarbeit von Bernhard Assmann, Johanna Puhl und Patrick Sahle. Schriften des Instituts für Dokumentologie und Editorik 10. Norderstedt: Books on Demand, 2015. 31–46.

angewandt und entwickelt werden. In dem Projekt werden hauptsächlich Pergamentdokumente analysiert, die zwischen dem zehnten und zwölften Jahrhundert im Katherinenkloster am Berg Sinai entstanden. Im Rahmen des Projektes werden die betreffenden Texte analysiert, vollständig oder teilweise transkribiert sowie ediert. Die größte Herausforderung stellt hierbei der schlechte Zustand der Handschriften dar, da diese teilweise von Schimmel befallen sind, verblichenen oder entfernten Text enthalten oder palimpsestiert sind. Deshalb werden die betreffenden Handschriften mithilfe eines portablen multispektralen Aufnahmesystems digitalisiert. Diese nicht invasive Konservierungsmethode hat sich als besonders nützlich für die Untersuchung und Rekonstruktion von solchen schlecht lesbaren Handschriften erwiesen. Verglichen mit normalem Weißlicht, bietet eine Untersuchung in ausgewählten spektralen Bereichen die Möglichkeit gewisse Details eines Dokuments, wie verblasste Schrift oder Palimpsesttext, besser sichtbar zu machen. Um den Kontrast zwischen den verblichenen Schriften und dem restlichen Dokument weiter zu verstärken, werden sogenannte Blind Source Separation Techniken - wie Principal Component Analysis (PCA) oder Independent Component Analysis (ICA) - angewandt. Des Weiteren werden in dieser Arbeit die jüngsten Entwicklungen im Sinai II Projekt beschrieben: Diese beinhalten Techniken für Document Image Dewarping und Automatic Layout Analysis, die Herausgabe der Edition *Glagolitica Sinaitica* sowie das Resultat aus einem anderen Projekt, das sich mit einer ähnlichen Zielsetzung beschäftigt: Das Bildverarbeitungsprogramm *Paleo Toolbar*.

1. Introduction

Apart from the description and classification, the main task of the work with historical manuscripts consists of the decipherment and edition of their contents. This may be difficult due to paleographic or linguistic peculiarities, but also because of the bad preservation state of the relevant source. Thus, it is essential to develop and evaluate methods of image and readability enhancement for this kind of documents.

In our interdisciplinary projects "Critical Edition of the New Sinaitic Glagolitic Euchology (Sacramentary) Fragments with the Aid of Modern Technologies" (P19608, 2007-2010) and "The Enigma of the Sinaitic Glagolitic Tradition" (P23133; 2011-2014), both funded by the Austrian Science Fund (FWF), we have been investigating various approaches of manuscript analyses with scientific and computational methods, among them Image Binarization (cf. Lettner 2009), Character Classification (cf. Vill 2008), Optical Character Recognition (OCR) (cf. Diem 2011), Layout Analysis (cf. Kleber 2009, Garz 2011), and Fragment Puzzling (Kleber 2008). In the scope of both Sinai projects also chemical analyses have been executed on the ancient documents by material chemists from the Vienna Academy of Fine Arts, namely X-Ray Fluorescence

Analyses (XRF) and, in the latter project also InfraRed Spectroscopy (FTIR). The results are being published in the new series *Glagolitica Sinaitica* (GlagSin) by Holzhausen publishing house and will comprise not only facsimile and critical editions, but also a separate technical volume on analysis methods and computational results.

The manuscripts investigated in this study were all written on parchment and originate from the 10^{th} to 14^{th} centuries. They are written in Glagolitic, the oldest Slavic script (cf. Miklas). The documents are in a poor condition and its texts partly barely or not at all discernible, due to partially faded or washed out ink, stains, mold, and other corruption by background clutter. The readability of multiple texts in palimpsests poses additional difficulties. One object in particular, the so called *Missale Sinaiticum* (Cod. Sin. Slav. 5/N), has long been almost impossible to read and transcribe, due to the mysterious discoloration phenomenon of brown or black ink that has turned to white (cf. Miklas). The second object of investigation is one of the Budapest Glagolitic Fragments (Duod. Slav. 2, Nr. 5), which contains characters, that are faded out but still darker than the background.

Both objects were digitized with MultiSpectral Imaging (MSI) technology that has proven a valuable basis for the investigation of historical manuscripts using computational methods. This approach is valued because it is non-invasive and more capable of enhancing the contrast of the degraded writings than methods that rely on normal white light illumination (Easton; Rapantzikos). Using MSI, the transcription of the contents is considerably easier and more comprehensive, as the results of the Archimedes Palimpsest project also have shown (cf. Easton). Our custom built and portable MSI system is described in detail in Section 2 (cf. Lettner; Hollaus).

Successful approaches to contrast enhancement of badly legible historical documents are dimension reduction techniques like Principal Component Analysis (PCA) and Independent Component Analysis (ICA) as demonstrated in Easton, Salerno, and Lettner.

Easton et al. use the PCA approach for the separation and enhancement of the diverse textual layers in the Archimedes palimpsest. They suggest merging those of a set of PCA images that typically emphasize the writing into one pseudocolor image. Salerno et al. when investigating the Archimedes Palimpsest applied the PCA approach and also analyzed the ICA method, which used the FastICA methodology (cf. Hyvärinen). Rapantzikos et al. compared the performance of the PCA technique to the performance of the Linear Spectral Mixture Analysis for the separation of the text layers contained in a palimpsest-ms. and state that the PCA approach is better suited for this task.

Lettner et al. proposed a technique for the enhancement of one or more text layers using not only spectral but also spatial information. In this procedure, a priori knowledge provided by philological experts is incorporated into a text line model used for the generation of a binary mask to encode the text regions. The enhancement

itself is performed by applying the Multivariate Spatial Correlation technique (cf. Wartenberg). The authors were able to show that their technique is superior to the PCA method.

Similar to Lettner we implement spatial correlation, since our approach is also based on text line detection. But in contrast to Lettner we do not require a priori knowledge, as the text lines are automatically found in the investigated regions. To the best of our knowledge this is the first time that the Fisher Linear Discriminate Analysis (LDA) approach is used for the enhancement of historical texts. So far it had been successfully used for face recognition tasks: Belhumeur et al. show that their LDA based dimensionality reduction outperforms the Eigenface method, which utilizes the PCA transformation.

While PCA and ICA are unsupervised dimension reduction tools, we apply the LDA approach, which is a supervised dimension reduction method and hence requires a labelled subset of the multispectral samples. In order to find a training set for the LDA classifier, we apply a semi-automated procedure, in which a subset of multispectral observations is labelled as belonging to text or background regions. The labelling procedure was especially designed for the *Missale Sinaiticum*, where a correct labelling of the fore- and background regions is difficult since in parts characters are barely visible within the multispectral images or best visible in different spectral ranges. Therefore, the labelling is performed on PCA images with their enhanced contrast. Due to the bad condition of the *Missale Sinaiticum* the PCA images are also corrupted by noise; hence labelling by application of a binarization algorithm would have been too error-prone. Instead, the labelling is performed in an iterative manner and is based on the detection of text lines, since the detection of text lines is more robust against noise — e.g. background clutter or lighting variations — compared to binarization methods.

In the last chapter, several further document analysis and image enhancement results will be presented, namely Document Image Dewarping, Automatic Layout Analysis, and two different toolboxes for paleographic document analysis.

The next section introduces a number of image enhancement techniques based on MultiSpectral Imaging (MSI).

2. Image Enhancement Techniques based on MultiSpectral Imaging

2.1. MSI Approach and System Setup

The MSI approach has proven its applicability for the investigation of ancient and problematic sources, since it is a non-invasive analysis tool capable of increasing the legibility of indecipherable texts. Our portable MSI system employs two sets of

LED panels, which provide 11 different narrowband spectral ranges. Compared to an MSI setup that applies optical filters, the LED illumination has two major advantages: First, the heat put on the manuscripts is reduced since broadband illumination is not required. Second, geometrical distortions stemming from optical filters are avoided and an image registration step is not necessary since the optical characteristics are not changed during acquisition. The images are acquired by two cameras, namely a Hamamatsu greyscale camera and a Nikon D4 SLR camera (cf. Figure 1). As a result the multispectral images already provide improved contrast and readability of the ancient writings in several spectral ranges — compared to ordinary white light illumination.

Nevertheless, due to their bad condition, some manuscript portions still remain unreadable. So we have implemented three different enhancement techniques, all of which are based on dimension reduction and applied statistics of the MSI results.

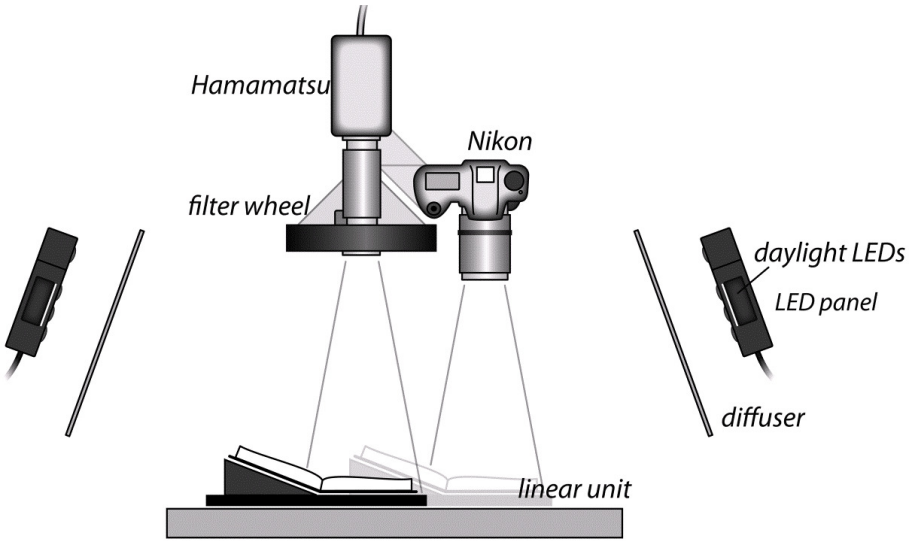

Figure 1. Portable MSI system setup of the Sinai Project.

2.2. Comparison of Image Enhancement Techniques

This section contains a performance comparison between three different dimension reduction techniques: PCA, ICA, and LDA.

These techniques are used to lower the third dimension of a multispectral scan in order to extract the relevant information. For folios containing only one single

textual layer, the MSI scans are reduced to just one image to emphasize the content in question, in our case the text. For palimpsests, on the other hand, the third dimension of the MS scan is reduced to two images, one emphasizing the underwritten, the other the overwritten text. Dimension reduction techniques can be grouped into unsupervised (PCA and ICA) and supervised (LDA) dimension reduction approaches.

Unsupervised Dimension Reduction Techniques

This technique category makes no use of class information. This category of enhancement techniques finds a transformation of the multispectral scan that removes any contained correlation. Such correlation stems mainly from material spectral correlation.

PCA

The PCA transformation finds an orthogonal transformation, on which the data is projected, and removes redundancies contained in the MS scan. The transformation of the multispectral data x is formally defined by:

$$y = Wx$$

where y is the transformed signal, and W is the transformation matrix. The columns of the transformation matrix are filled with the eigenvectors of the covariance matrix of the zero mean normalized data. The eigenvectors are sorted in descending order according to their corresponding eigenvalues.

In this step, the first k PCA images are considered, since one writing is usually emphasized by multiple principal components (cf. Lettner; Easton). For the *Missale Sinaiticum* it was experimentally found that five PCA images have to be considered. For the Budapest Glagolitic Fragment k was set to three, because the spectral signature of the writing in this manuscript is less varying than in the *Missale Sinaiticum*, and the writing is described by the first three eigenvectors. We noticed that in the PCA images, the ruling scheme is better recognizable than the degraded characters. Therefore, in the coarse enhancement stage the samples are labelled as belonging to text lines or belonging to intermediate regions instead of belonging to text or background regions.

We noticed that the PCA output is generally ordered, meaning the first PCA image shows the text layer - if the writing is in relative good, i.e. rich-in-contrast condition. If the document contains a great amount of background clutter or faded-out characters, however, the first image may show the enhanced background clutter instead, whereas the image showing the writing is emphasized by another principal component.

ICA

In contrast to the PCA transformation, the ICA approach finds a projection matrix W that is not necessarily orthogonal. PCA follows the assumption that sources within the data (e.g. the text(s)) are uncorrelated, whereas ICA assumes that the sources are

statistically independent. Those sources that carry any specific content (e.g. text or palimpsest text) are detected by maximizing their independence. Similar to Salerno we applied this method in order to separate the two layers of a palimpsest. We applied the FastICA algorithm (for more details on ICA cf. also Hyvärinen; Salerno; Hollaus).

Supervised Dimension Reduction Techniques

LDA

PCA and ICA as unsupervised dimension reduction techniques find the relevant information within the multispectral scans by themselves. LDA as a supervised dimension reduction algorithm requires a prepared subset of the samples with labelled fore- and background pixels as a training set for the classifier. Such labelling could be achieved by applying a simple document binarization technique (cf. Su), but since historical documents can be corrupted with background clutter, simply applying a binarization algorithm is error-prone. That is why we suggest a more careful procedure instead: In a first step, the areas in an input image with the most visible content, e.g. text lines, are determined. Text line regions are found with a text line detection algorithm similar to the one in Yosef et.al. Depending on the document, both UV fluorescence images and the output of the PCA transformation can provide suitable input images. The pixels of this input image are labelled as belonging to text line or intermediate, i.e. background, regions. Then the LDA-based dimension reduction is applied in order to generate an enhanced image by bringing out the degraded text and enhancing the contrast.

The PCA transformation maximizes the so-called scatter of the transformed feature vector y. In contrast, the LDA method as a supervised dimension reduction tool seeks a projection W that maximizes the ratio of the between-class scatter to the within-class scatter, i.e. it exploits the class information in order to select discriminant features, e.g. for text and background (for more details on LDA cf. also Duda).

If the manuscript is a palimpsest, a three-class problem is considered, where overwriting, underwriting, and background classes are considered. Since the overwriting of the manuscript investigated is typically most visibly separable from other content under Near Infrared (NIR) illumination, whereas the underwriting is not, the labelling of the overwriting is performed on NIR images by applying a binarization algorithm (cf. Su). Afterwards, the underwriting is found by applying the procedure explained above.

2.3. Results

For image enhancement of degraded documents, dimension reduction techniques have yielded two advantages: On the one hand, an investigation of the entire scan

is avoided, while on the other hand, the contrast and visibility of the degraded characters is enhanced and the resulting images show a considerable increase in legibility compared to the images contained in the multispectral scan.

In this study we compared three different dimension reduction techniques applied on the entire multispectral scan on two different manuscripts: PCA, ICA, and LDA. The results show that all three approaches are capable of considerably enhancing the contrast compared to the unprocessed multispectral images. In the best case the output is one single image showing the relevant image content, e.g. the text.

We noticed that the PCA output is generally ordered, meaning the first PCA image shows the writing — if the writing is in relative good, i.e. rich-in-contrast condition. However, if the document contains a great amount of background clutter or faded-out characters it is not guaranteed that the first image shows the degraded writing, since the background clutter may be enhanced in the first image, whereas the image showing the writing is emphasized by another principal component.

The ICA approach does not order the resulting images; hence the ICA image showing the respective content most clearly must be selected manually.

In contrast, the LDA approach considers a two class problem and thus its output is a single image which enhances the writing, given a previously applied line detection algorithm separated the classes correctly. A qualitative analysis showed that the LDA approach achieves better performance in the case of background clutter and faded-out characters.

The results of the image enhancement techniques were evaluated by using a qualitative legibility assessment, conducted by a philologist experienced in reading the Glagolitic script. In order to avoid bias the scholar did not evaluate entire image patches, but instead single characters. We extracted 212 single characters from 7 test panels, each containing several hard to read characters. The test panels were normalized between the minimum and maximum intensity value. A grouping into categories of visual quality was not used, due to high variety of the enhancement results. Instead, the scholar was asked to pairwise compare two single characters, which were produced by different techniques. The philologist was not told which method was used for the enhancement of a particular character. For each compared letter pair, the scholar decided which letter was found more legible and assigned a 1 to the superior result image and a 0 to the inferior counterpart. Afterwards, the sum of the assigned scores was calculated.

The performance of the LDA based technique improves if it is solely applied to a region in which the characters have a similar contrast. Hence, we applied the algorithm exclusively to the test panels. For the PCA and ICA approaches it cannot be assumed that a transformation based on local statistics gains a better enhancement result than a transformation which has been calculated on the basis of the entire folio. Therefore, the PCA and ICA techniques have been applied in a local and global manner.

Since the text layer is emphasized by several principal components, we manually selected the best PCA outcome. The output of the (Fast)ICA algorithm depends on a random initialization and on several parameters we applied on the MSI data and selected the best results for the evaluation task. Additionally, the multispectral image with the best visible text was added to the test set. Two examples of Missale Sinaiticum can be seen in Figure 2 and Figure 3.

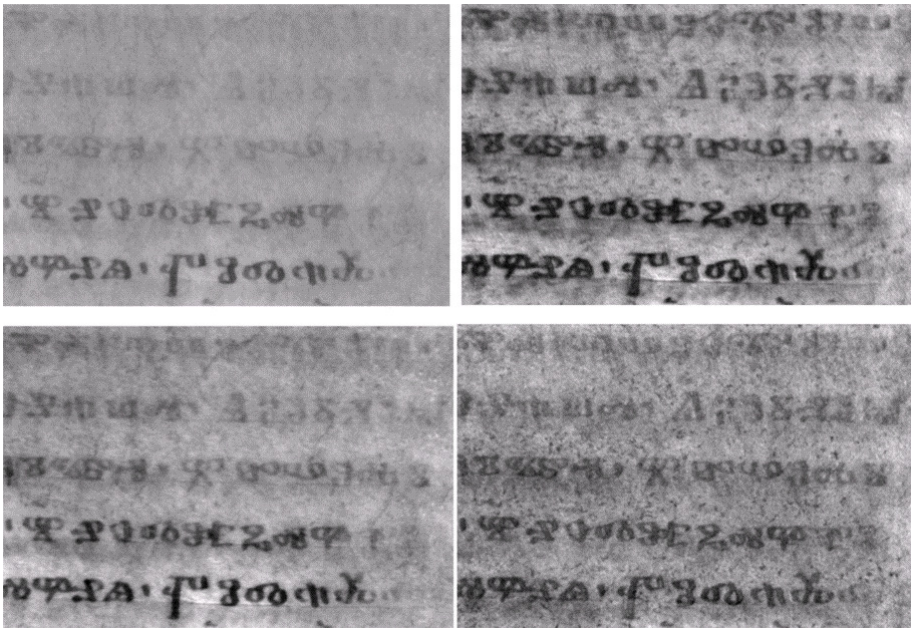

Figure 2. Missale Sinaiticum: top left: white light image. Top right: LDA, bottom left: PCA, bottom right: ICA output.

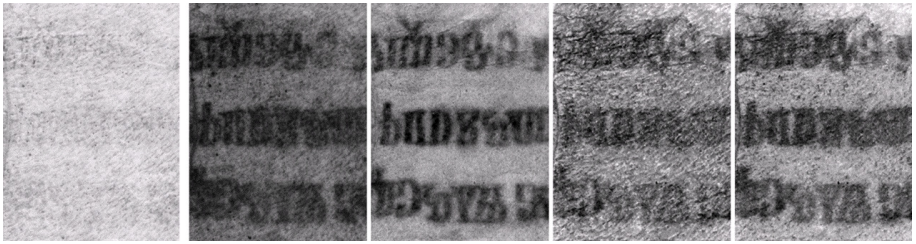

Figure 3. Budapest Glagolitic Fragment: from left to right: white light image. UV fluorescence image. LDA, PCA, ICA results.

In both cases the LDA output is superior to the other enhancement techniques, since it is capable of restoring the degraded text parts. In Figure 3 the PCA performs poorly, with lower readability than in the UV fluorescence image.

3. The Sinai Project II – Related Results

Since our last report on the Sinai Project several new developments have been made. They will be described shortly in the following.

3.1. Document Image Dewarping

The other readability enhancement method presented in this work, namely document image dewarping, is concerned with the rectification and smoothing of images of documents that are deformed due to age and bad storage conditions. The proposed algorithm has been adapted to artificially warped documents showing a conventional Latin computer font, e.g. *Times New Roman*, as well as images of Glagolitic script in two variants, a) artificially warped documents written in a standardized computer font, e.g. *Glagolica Bulgarian*, for experimental purposes and the binarized images of the ancient Sinai manuscripts (cf. Figure 4). The warping of all three types of documents could be straitened successfully and the readability thus improved.

The evaluation of the dewarping of the Glagolitic documents showed a particular improvement by reducing the Straightness Standard Error. An evaluation of the results on the artificially warped Latin documents by means of applying standard OCR on the original as well as on the dewarped text shows that the straightened text had a higher recognition rate by 51% on average (cf. Steinböck).

To further enhance the dewarping results on the Glagolitic document images, we plan to also take into account alignment and orientation of the characters, as the Glagolitic script can be not only a standing script, but also a hanging one.

3.2. Document Layout Analysis

The layout entities considered in this approach include body text, embellished initials, plain initials and headings. These textual elements are disassembled into segments, and a part-based detection is done which employs local gradient features known from the field of object recognition, the Scale Invariant Feature Transform (Sift).

These features describe the structures in a scale-, rotation- and illumination invariant manner. Hence, this approach does not rely on a binarization step, but is directly applied to the gray scale image, and it is robust to variations in shape, illumination, writing orientation, and (background) noise, too. Thus, it is suitable for ancient handwritten documents with varying layouts and degradation effects.

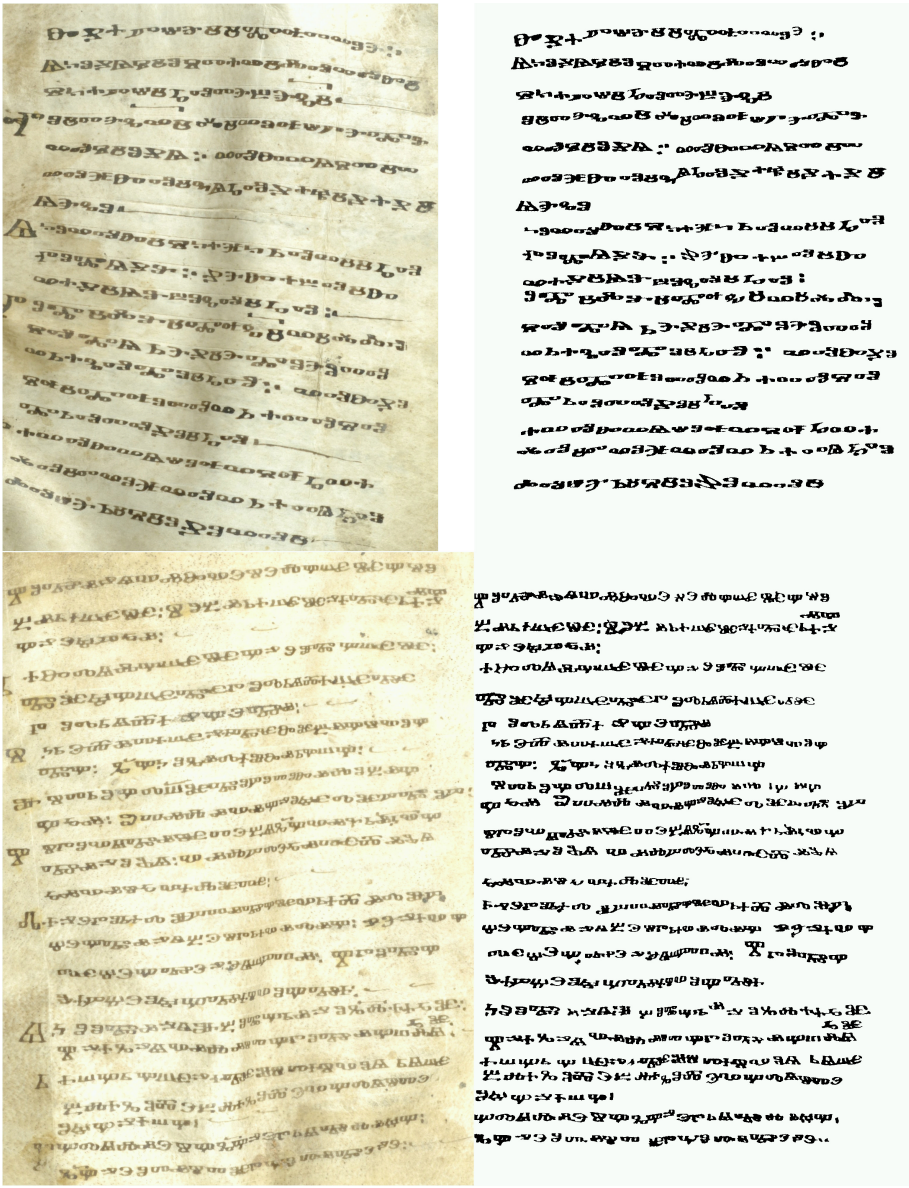

Figure 4. Results of the dewarping algorithm on 2 folia of the Psalter of Demetrius (Cod. Sin. Slav. 3/N): original images (left), rectified images (right).

As the whole entity cannot directly be inferred from the mere positions of the interest points, a localization algorithm is needed that expands the interest points according to their scales and the classification score to regions that encapsulate the whole entity. Therefore, a cascading algorithm is proposed that successively rejects weak candidates applying voting schemes (cf. Garz).

The evaluation shows that the method is able to locate main body text in ancient manuscripts. The detection rate of decorative entities is not as high as for main body text but already yields to promising results (cf. Figure 5).

3.3. *nomacs image lounge* and Toolbox for Manuscript Analysis

The image viewer *nomacs image lounge* (<http://www.nomacs.org>) developed by image processing experts at the Computer Vision Lab at Vienna UT already features several useful functions for multispectral image analysis, e.g. image panning (also between connected computers), and manuscript analysis, e.g. a false color tool for enhancing the visual contrast between e.g. text and palimpsest. The latest development is a toolbox for manuscript analysis. It was developed in collaboration with manuscript researchers in order to make certain processing and reading steps better and easier. The toolbox includes an automatic line detection tool based on the text line detection technique similar to Yosef et.al. The calculated lines, depending on the script either below or above the text, can be visualized directly on the shown manuscript image. It can also measure the factual distance between two points of a loaded page in relation to the image resolution. Another tool makes cutting out characters for the creation of character comparison tables very fast by simply clicking into the according letter and automatically extracting it in a bounding box. The toolbox will be included in the next release of *nomacs*.

3.4. External Project: Paleo Toolbox

Recently, one of the authors (Christens-Barry) has developed an imaging and processing software tool, *Paleo Toolbox*, to aid researchers having varying levels of computer skill. *Paleo Toolbox* implements workflows developed for spectral imaging projects, including the Archimedes Palimpsest Project and the St. Catherine's Palimpsests Project being conducted by EMEL[2], with particular focus on the identification and increased legibility of undertext. It uses a simple tool palette interface that builds upon the plugin architecture of ImageJ, a widely used open source tool for all major platforms[3].

---

[2] <http://rsbweb.nih.gov/ij/>.

[3] Available at <http://rsb.info.nih.gov/ij/>.

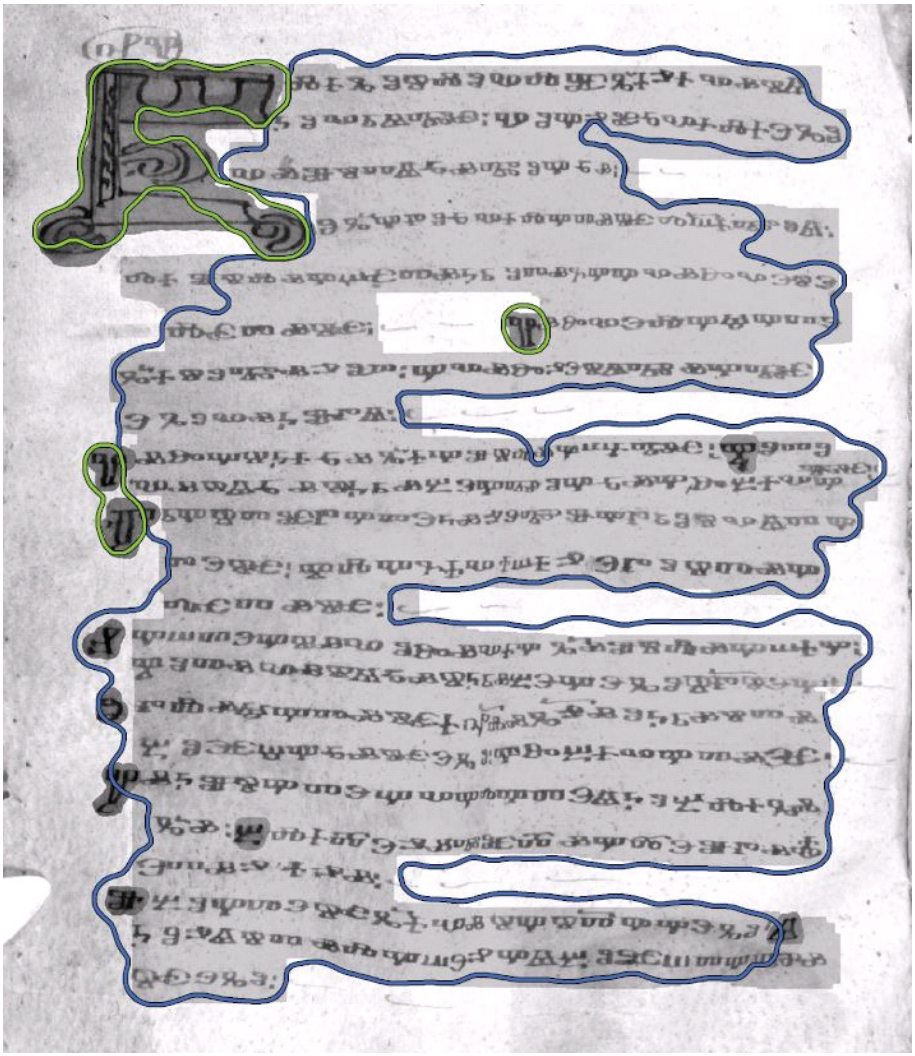

Figure 5. Automatic Layout Detection on Psalterium Demetrii Sinaitici fol. 43v.

Paleo Toolbox operations offer the large number of image viewing and image processing operations that are available in the standard ImageJ libraries and plugins, without requiring great expertise of the user. These rely largely on quantitative manipulation, measurement, and visualization of features of interest. It includes a set of image analysis (statistical approaches such as PCA), visualization (pseudocolor display and color channel manipulation) and markup tools (textual annotation; region-of-interest recording and replay). It more generally allows users to perform operations that assist in codicological and paleolographic investigations. *Paleo Toolbox* can be adapted to the needs of individual scholars: ImageJ's plugin architecture facilitates the development or adaptation of processing routines to match the goals of particular studies or the content of particular texts.

Its design is driven by the recognition that a linear model of image analysis, in which images are initially processed solely by computer scientists and are then subsequently viewed solely by humanities scholars, is both static and inadequate. *Paleo Toolbox* is designed to allow scholars with little expertise in image manipulation to dynamically explore images, to adjust display features (contrast, zoom, pan, color scheme etc.), but also to create annotations and markup. These can be used to guide computer scientists in adaptively choosing particular processing procedures based on the content.

Paleo Toolbox is based on an iterative model that envisions feedback from scholars that informs further processing. This model is more similar to the interplay between the human eye and the human brain than conventional approaches that conceive vision as a unidirectional flow of information from image processors (technologists) to visual interpreters (scholars). A further, central goal at the heart of *Paleo Toolbox* is to assist the collaboration between technologists and scholars, while adopting standard image and metadata formats that facilitate exchange, storage, and retrieval of images, annotations, and working materials.

Based on the insights of one of the authors (Miklas) at a recent working conference on the St. Catherine's Palimpsests Project, *Paleo Toolbox* is presently being configured to allow users to choose between a basic viewing environment, suited to the needs of non-experts to view and annotate images in a simplified setting, and a more sophisticated processing environment encompassing the full range of processing operations. This is intended to ease the joint use of a common tool by technologists and scholars, at their various levels of comfort in using software tools, during both individual and collaborative projects. (Please contact Bill Christens-Barry to obtain a copy).

3.5. Series Glagolitica Sinaitica

Finally, we would like to announce that the first edition of our new series *Glagolitica Sinaitica* (GlagSin) was issued by Holzhausen publishing house in August 2012

containing the facsimile edition of the Psalterium Demetrii Sinaitici and the inserted Medical Folia (Sin. Slav. 3/N); the critical edition of both manuscripts is well under way. The following editions will comprise not only other facsimile and critical editions of Sinai manuscripts, but also a separate volume on analysis methods, material chemistry, and computational results.

Acknowledgments

The authors would like to thank the Austrian Science Fund (FWF) for funding the project under grant P23133.

Bibliography

Belhumeur, P. N., J. P. Hespanha, and D. J. Kriegman. "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection." In: *ECCV* 1 (1996). 45–58.

Diem, M., R. Sablatnig, M. Gau, and H. Miklas. "Recognizing Degraded Handwritten Characters." *Codicology and Palaeography in the Digital Age 2.* Ed. F. Fischer, C. Fritze, G. Vogeler. Norderstedt: BoD, 2011. 295–306. <http://kups.ub.uni-koeln.de/4337/>.

Duda, R.; Hart, P.; Stork, D. (ed.) (2001): *Pattern Classification*^{2}. New York: Wiley-Interscience, 2001.

Easton, R., W. Christens-Barry, and K. Knox. "Spectral Image Processing and Analysis of the Archimedes Palimpsest." *Proceedings of the 19th European Signal Processing Conference, EUSIPCO. 2011.*

Garz, A., M. Diem, and R. Sablatnig. "Layout Analysis of Ancient Manuscripts Using Local Features." Holappa, M. (Ed.): *Eikonopoiia. Digital Imaging of Ancient Textual Heritage 2010: Technological Challenges and Solutions.* (Commentationes Humanarum Litterarum. Vol. 129). Helsinki: Societas Scientiarum Fennica, 2011. 150–162.

Glagolitica Sinaitica. Wien: Holzhausen Verlag, 2012ff. <https://shop.verlagholzhausen.at/hhshop/Psalterium-Demetrii-Sinaitici.-Monasterii-s.-Catharinae-cod.-Slav.-3/n-adiectis-foliis-medicinalibus.htm?x3ce42=f72b414588eef8293a09877c464a828a>.

Hollaus, F., M. Gau, and R. Sablatnig. "Multispectral Image Acquisition of Ancient Manuscripts." *EuroMed* 2012. 30–39.

Hyvärinen, A.; Oja, E. "Independent Component Analysis: Algorithms and Applications." *Neural Networks.* Vol. 13/4-5 (2000). 411–430.

ImageJ. Image Procession and Analysis with Java. <http://rsb.info.nih.gov/ij/>.

Kleber, F., and R. Sablatnig. "Scientific Puzzle Solving: Current Techniques and Applications." *Computer Applications and Quantitative Methods in Archaeology (CAA). Making History Interactive.* Williamsburg (VA), 2009. (CD publication).

Kleber, F., R. Sablatnig, M. Gau, and H. Miklas. "Ruling Estimation for Degraded Ancient Documents Based on Text Line Extraction." In: *2nd Int. Conf. on Electronic Visualisation and the Arts (EVA). Digital Cultural Heritage – Essential for Tourism.* Vienna. 79–86.

Lettner, M.; Diem, M.; Sablatnig, R.; Miklas, H. (2008): "Registration and Enhancing of Multispectral Manuscript Images." *16th European Signal Processing Conf. (EUSIPCO)*. Lausanne. (CD publication).

Lettner, M.; Sablatnig, R. (2009): "Document Image Binarization in Multispectral Images using Markov Random Fields." *33rd Workshop of the Austrian Association for Pattern Recognition (AAPR): Challenges Visual Learning*. Ed. Roth, P.M., T. Mauthner, and T. Pock. (Schriftenreihe der OCG. Vol. 254). Stainz, 2008. 85–96.

Miklas, H., M. Gau, F. Kleber, M. Diem, M. Lettner, M.C. Vill, R. Sablatnig, M. Schreiner, M. Melcher, and E.-G. Hammerschmid. "St. Catherine's Monastery on Mount Sinai and the Balkan-Slavic Manuscript-Tradition." *Slovo. Towards a Digital Library of South Slavic Manuscripts*. Ed. H. Miklas and A. Miltenova. Sofia: Bulgarian Academy of Science, Institute of Literature, 2008. 13–36, 286 (Res.).

Rapantzikos, K., and C. Balas. "Hyperspectral Imaging: Potential in Non-Destructive Analysis of Palimpsests." *Int. Conf. on Image Processing (ICIP)* 2 (2005). 618–621.

Salerno, E., A. Tonazzini, and L. Bedini. "Digital Image Analysis to Enhance Underwritten Text in the Archimedes Palimpsest." *Int. Journal on Document Analysis and Recognition*. Vol. 92/4 (2007). 79–87.

Steinböck, D.: *Document Image Dewarping*. Bachelor Thesis. Vienna Univ. of Technology. 2012.

Su, B.; S. Lu, and C.L. Tan. "Binarization of Historical Document Images Using the Local Maximum and Minimum." *Document Analysis Systems*. (2010) 159–166.

Vill, M.C., and R. Sablatnig. "Static Stroke Decomposition of Glagolitic Characters." *2nd Int. Conf. on Electronic Visualisation and the Arts (EVA). Digital Cultural Heritage – Essential for Tourism*. Ed. R. Sablatnig, J. Hemsley, P. Kammerer, E. Zolda, and J. Stockinger. Vienna, 2008. 95–102.

Wartenberg, D. "Multivariate Spatial Correlation: A Method for Exploratory Geographical Analysis." In: *Geographical Analysis*. Vol. 17/4 (1985), 263–283.

Yosef, I. B., N. Hagbi, K. Kedem, and I. Dinstein. "Line Segmentation for Degraded Handwritten Historical Documents." In: *ICDAR* (2009). 1161–1165.