TU WIEN Informatics

Restoration of Multispectral Images of Ancient Documents

DISSERTATION

zur Erlangung des akademischen Grades

Doktor der Technischen Wissenschaften

eingereicht von

Fabian Hollaus
Matrikelnummer 0326844

an der Fakultät für Informatik
der Technischen Universität Wien
Betreuung: a.o.Univ.-Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig

Diese Dissertation haben begutachtet:

Andreas Maier	Ioannis Pratikakis

Wien, 2. August 2021

Fabian Hollaus

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

TU WIEN Informatics

Restoration of Multispectral Images of Ancient Documents

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktor der Technischen Wissenschaften

by

Fabian Hollaus
Registration Number 0326844

to the Faculty of Informatics

at the TU Wien

Advisor: a.o.Univ.-Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig

The dissertation has been reviewed by:

Andreas Maier	Ioannis Pratikakis

Vienna, 2nd August, 2021

Fabian Hollaus

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Erklärung zur Verfassung der Arbeit

Fabian Hollaus

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 2. August 2021

Fabian Hollaus

Danksagung

An dieser Stelle möchte ich mich bei jenen Menschen bedanken, die mich direkt oder indirekt bei dieser Arbeit unterstützt haben.

Besonderer Dank gilt meinem Betreuer Robert Sablatnig, der mich stets fachkundig unterstütze und durch sein stetiges Nachfragen dazu motivierte diese Arbeit fertig zu stellen: *Jetzt ist's fertig.* Zusätzlich ist Robert hauptverantwortlich für das gute Arbeitsklima am CVL, das ich besonders schätze.

Des Weiteren möchte ich mich bei meinen ArbeitskollegInnen bedanken, die mich einiges lehrten und mit denen ich einiges leerte. Ganz besonders möchte ich mich bei Flo und Markus bedanken, die mich bei der Erstellung dieser Arbeit sehr unterstützt haben. Vielen Dank an Flo, dafür dass er sich die gesamte Arbeit angeschaut hat! Großer Dank gebührt auch Heinz Miklas und Melanie für die gute Zusammenarbeit und die gewährten Einblicke in die Welt der Philologie.

Ich bedanke mich ganz herzlich bei meinen Gutachtern Andreas K. Maier und Ioannis Pratikakis für die Begutachtung dieser Arbeit!

Zuletzt möchte ich mich noch meiner Familie danken:
Meinem Bruder Benedikt und meiner Mutter für ihre Unterstützung - auch, aber nicht nur während dieser Arbeit.
Julia, für die Unterstützung, Geduld und Gespräche - ohne diese hätte ich die Arbeit nicht fertiggestellt.

Für Hannah und Sophia.

Kurzfassung

Diese Arbeit befasst sich mit der Verarbeitung von Aufnahmen von historischen Dokumenten. Die historischen Schriften enthalten teilweise verblasste Schriftzeichen oder ihre Qualität ist durch Hintergrundvariationen vermindert. Die Multispektralaufnahme hat sich als wertvolles Werkzeug für die nicht-invasive Untersuchung von solchen alten Manuskripten erwiesen, da mit ihr Informationen erfasst werden können, die für das menschliche Auge unsichtbar sind. Die Dokumentenbilder, die in dieser Arbeit untersucht werden, wurden mit einem tragbaren multispektralen Aufnahmesystem aufgenommen. Die Aufnahme in schmalbandigen Spektralbereichen führte zu einer erheblichen Steigerung der Lesbarkeit. Die aufgenommenen Bilder bilden die Grundlage für zwei Arten von Verarbeitungstechniken, die in dieser Arbeit vorgestellt werden:

Zum einen wird eine Bildverbesserungsmethode vorgestellt, die die multispektralen Messungen anhand einer LDA basierten Transformation in einen niedriger dimensionalen Raum projiziert. Dadurch wird nicht nur die Dimensionalität der multispektralen Bilder reduziert, sondern auch die Lesbarkeit der degradierten Schriften erhöht. Eine qualitative Analyse, die von Philologen durchgeführt wurde, zeigt, dass die Methode teilweise bessere Ergebnisse erzielt als zwei andere Methoden zur Dimensionsreduktion.

Das zweite Ziel dieser Arbeit ist die Trennung der antiken Schriften vom restlichen Hintergrund. Dafür wurden mehrere Binarisierungsmethoden entwickelt: Zwei Methoden nutzen einen Target-Detection Algorithmus, mit dem festgestellt wird, ob Tinte in den multispektralen Messungen vorhanden ist. Eine weitere Binarisierungsmethode wird vorgestellt, die ein Gaussian Mixture Model (GMM) basiertes Clustering verwendet. Die vorgestellten Methoden nutzen räumliche und spektrale Informationen. Außerdem wird ein Fully Convolutional Network (FCN) für die Binarisierung verwendet. Die Methoden werden auf zwei Datenbanken evaluiert: Zunächst werden die Methoden auf dem MS-TEx Datensatz angewandt, wobei vielversprechende Ergebnisse erzielt werden. Die besten Resultate werden von den Methoden erzielt, die auf dem Target-Detection Algorithmus basieren. Diese Methoden nahmen am MS-TEx 2015 Wettbewerb teil, wo sie den ersten und zweiten Platz belegten. Zusätzlich werden die Methoden auf dem MSBin Datensatz evaluiert. Dieser Datensatz ist größer und ermöglicht ein erfolgreiches Training des FCN, welcher die übrigen Binarisierungsmethoden übertrifft. Dennoch sind die Ergebnisse aller Methoden den Resultaten überlegen, welche mit einem traditionellen Binarisierungsansatz erzielt werden.

Abstract

This thesis is concerned with the restoration of images of historical documents. The ancient writings imaged contain partially faded-out characters or are degraded by background variations. MultiSpectral Imaging (MSI) has proven to be a valuable tool for the non-invasive investigation of such ancient manuscripts, since it can be used to acquire information that is invisible to the human eye. The document images which are examined in this work, have been acquired with a portable MSI system. The imaging in narrow-band spectral ranges led to a considerable legibility increase. The images taken form the basis for two kinds of restoration techniques that are introduced in this work:

First, an enhancement method is proposed that projects the multispectral samples on a lower dimensional space by applying an Linear Discriminant Analysis (LDA) based transformation. Thus, not only the dimensionality of the multispectral images is lowered, but also the legibility of the degraded writings is increased. A qualitative analysis conducted by philologists shows that the method partially outperforms unsupervised dimension reduction methods, which are used in previous works. The method is also evaluated in a quantitative analysis, where it is shown that the performance of an Optical Character Recognition (OCR) system can be improved by applying the enhancement technique in a preprocessing step.

The second aim of this work is the separation of the ancient writings from the remaining background. Such binarization methods are used as a preprocessing step for other document image analysis methods, including OCR or writer identification. Multiple binarization methods have been developed for the multispectral document images considered: Two methods make use of a target detection algorithm, which is used to determine if ink is present within the multispectral samples. A further binarization method is introduced, which makes use of Gaussian Mixture Model (GMM) based clustering. The methods introduced make use of spatial and spectral information. Furthermore, a Fully Convolutional Network (FCN) is used for the binarization task. The methods are evaluated on two databases: First, the methods are applied on the MultiSpectral Text Extraction (MS-TEx) dataset, where the methods achieve promising results. The best performances are gained by the target detection-based methods. These methods participated in the MS-TEx 2015 contest, where they were ranked first and second. Second, the methods are evaluated on the MultiSpectral Document Binarization (MSBin) dataset. This dataset is larger and allows for a successful training of the FCN, which outperforms the remaining

binarization methods. Nevertheless, the results gained by all methods proposed are superior to the results which are gained by a traditional binarization approach that is designed for grayscale images.

Contents

CHAPTER 1

Introduction

This thesis deals with the analysis of multispectral images of ancient manuscripts. MSI has proven to be a valuable tool for the examination of historical manuscripts [RB05], [EN10]. This non-invasive investigation technique allows to capture information, which is invisible to the human eye [FK06]. Thus, the legibility of faded-out or erased text can be increased. The imaging in selected narrow-band ranges, ranging from UltraViolet (UV) to Near-Infrared (NIR), is used to raise the contrast of such ancient writings [LKSM08]. In addition, post-processing methods are applied on multispectral document images in order to further enhance their legibility [STB07]. Thus, the work of scholars is facilitated by MSI [EN10].

MSI is also used to improve the performance of automated document image analysis techniques: These techniques include document restoration, such as bleed-through removal [RDH18], [TBS09] or the separation of handwritings written with different ink types [GH15]. Another application is the differentiation between handwritings written with different ink types [GH15]. A further application that benefits from the additional spectral information is document image binarization [HNM+15].

This thesis is concerned with two aims: The first aim is the enhancement of ancient texts that are barely visible. Such texts have been imaged with a portable MSI system in order to increase their legibility. The images are further enhanced by an enhancement method that is proposed in this thesis. The second aim of this work is the binarization of multispectral document images.

1.1 Motivation

MSI, in the context of document analysis, has been successfully used to enhance the visibility of ancient writings that are barely visible [EN10]. It has also been used for the analysis of ancient writings contained in palimpsests [RB05]. Palimpsests are manuscripts

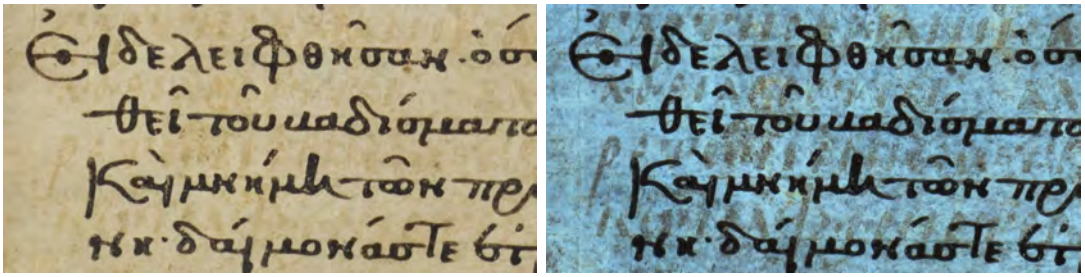


Figure 1.1: Suppl. Gr. 189. (a) Illuminated with white light. (b) UV fluorescence image.

in which an original writing is replaced by a younger writing. The old manuscripts were reused, because parchment was a costly material in former times [RB05]. The original writings were scrapped off, before the parchment was rewritten. Figure 1.1 shows an example for a palimpsest. The overwritten text is visible within the visible range, whereas the underwriting is only partially visible. The older text is best visible within the UV fluorescence image.

A prominent manuscript [FK06] that has been investigated with MSI is the Archimedes palimpsest. The Archimedes palimpsest is a copy of treatises created by the Greek scientist Archimedes of Syracuse. The original writing was erased and is partially not visible under ordinary white light illumination. Easton et al. [JCK11] have shown that the imaging in selected narrowband spectral ranges - especially the UV range - can be used to raise the contrast of the ancient writing.

While palimpsests texts have been erased, historical writings can also be degraded by further deteriorations including faded-out scripts, mold infestations, bleeding artifacts etc. [SLT13]. Exemplar images provided in Section 3, show that the imaging in selected spectral ranges can be applied in order to increase the visibility of such degraded writings. MSI data is highly correlated [LKSM08] and dimension reduction methods - such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA) - are used to lower the third dimension of the multispectral scan. Several researchers [JCK11], [STB07] show that a further legibility increase is gained by applying such dimension reduction methods. In this work a supervised dimension reduction technique, namely LDA, is used and it is shown that this method produces partially superior results, compared to a unsupervised dimension reduction by applying ICA or PCA. Since LDA is a supervised technique, it requires labeled training data. For this purpose a labeling strategy is proposed that automatically selects and labels a subset of the MSI data. The labeling strategy makes use of document image analysis methods, namely text line detection and binarization, and is introduced in Section 3.2. The results are grayscale images, which exhibit an increased foreground to background contrast.

Contrary, document image binarization methods are concerned with the separation of foreground and background. The resulting images are binary images and they are used as a preprocessing step for other document image analysis methods, including document

Figure 1.2: Exemplar images, used in DIBCO competitions.

layout analysis [BM20], character recognition [SLT10] or writer identification [HWS15], [FHGS13]. Additionally, binarization is used as a standalone application for noise removal in order to increase document readability [TM20].

Document image binarization methods are evaluated by means of pixel-based metrics or by means of OCR [NGP13]. The majority of the document image binarization techniques described in Section 2.2 are evaluated on datasets, which have been published in the course of the Document Image Binarization Contest (DIBCO) events. The first DIBCO competition was organized in 2009 [GNP09] and the published dataset allowed for the first time for an objective evaluation of binarization methods. Since then, ten DIBCO and Handwritten-Document Image Binarization Contest (H-DIBCO) competitions took place, which accelerated the research and development in the field of document image binarization [TM20]. The images in the DIBCO datasets are selected so that they *"should contain representative degradations which appear frequently"*[GNP09]. These degradations involve: background variations, shadows, smear, low contrast, bleed- and show-through [GNP09]. Figure 1.2 shows various examples for the degradations contained in the DIBCO datasets.

According to Tensmeyer and Martinez [TM20], a *"tremendous amount of progress has been made in the field of historical document binarization"* within the last decade. The majority of the binarization methods are developed for grayscale images [TM20]. Contrary, the binarization of multispectral images is rather a niche application: A limited number of works, for example [LS10], [MC15], are especially designed for MSI data. Hedjam et al. [HNM+15] organized the first - and only - binarization contest for multispectral document images in the course of the International Conference on Document Analysis and Recognition (ICDAR) 2015. The competition is entitled MS-TEx and the provided dataset allows for an objective evaluation of multispectral document binarization methods.

Since then, several methods [AAC19], [MC15], [HDS18] have been evaluated on the

MS-TEx dataset. These methods achieve better binarization results by incorporating spectral information, compared to traditional binarization methods that are designed for grayscale or RGB images.

The binarization of multispectral document images is used as a preprocessing step for other document image analysis techniques [BM20], [FHGS13], similar to the binarization of grayscale or RGB images. Furthermore, binarization of MSI data is used as a preprocessing step for enhancement methods that are applied on multispectral document images: For instance in [HNM+15] binarization is applied on a NIR channel in order to segment degraded regions that are best visible within this spectral range. The damaged regions are afterwards restored in a visible channel that shows the degraded text. In [KDB11] a semi-automated segmentation of the foreground is performed in order to identify regions that are afterwards enhanced by fusing NIR channels with channels from the visible range.

Closely related to the field of binarization is the identification and segmentation of different writings that are acquired by MSI or HyperSpectral Imaging (HSI) systems: In [GH15] hyperspectral images are used to segment two different ink types contained in a palimpsest. In [SPH+14] and [KKS19] hyperspectral images are used for forensic examination of modern documents that are written with different inks. This thesis is mainly concerned with the separation of a single handwriting, but the separation of two ancient writings is also briefly examined.

1.2 Aim of the Work

The methods proposed are developed to support the work of philologists and to be used as a preprocessing step for document image analysis methods. This work is concerned with two major aims: Enhancement and binarization. The method proposed for the former goal is designed to support the work of philologists and the methods suggested for the latter aim are used as a preprocessing step for document image analysis methods:

Enhancement The overall aim of the enhancement method proposed is to increase the legibility of degraded writings. The writings considered are in a poor condition, including faded-out script or mold infestations on the parchment. Hence, the visibility of the writings is limited and the method is used to increase the visibility and thus to facilitate the work of philologists and scholars.

Binarization Contrary to the enhancement method, the binarization techniques are not intended to increase the legibility of the writings. Instead, the binarization methods proposed are supposed to be used as a preprocessing tool for further document image analysis techniques, including for instance writer identification [FHGS13] or document layout analysis [BM20]. Additionally, a semantic segmentation approach is used for an automated separation of different ink types.

Figure 1.3: Exemplar input images used for enhancement (top row) and binarization (bottom row). White light images are shown on the left and UV images on the right.

The binarization methods and the enhancement method are designed for manuscripts which are in different conditions: The binarization methods are developed for manuscripts that are at least partially visible within the visible spectrum. The multispectral information is hereby used in order to increase the binarization performance. Contrary, the enhancement method is especially designed for faded-out writings that are barely visible within the multispectral channels. A binarization - and additionally a character recognition step - of these damaged writings is error-prone, because the contrast within the enhancement results is still partially low. Therefore, these images are not intended to be further processed by document image analysis methods, but instead they form the basis for a manual investigation by scholars. Note, that the enhancement method has also been used as a preprocessing step for an OCR system (see Section 3.3.2). However, it was found that the method can only be successfully applied on damaged image regions and gains nearly no readability increase on undamaged regions.

Figure 1.3 shows two examples for the multispectral document images, which are investigated in this thesis: The top row shows a portion of the manuscript for which the enhancement technique is mainly designed. The ink has discolored from black to white because of a chemical reaction, which can be attributed to water exposure [MGK+08]. The discolored ink is best visible within the image taken under UV light, shown on the right. The bottom row in Figure 1.3 shows an image that is belonging to the binarization dataset used. The writing is degraded, but it is in a better condition than the writing shown in the upper row. Again, the text is best visible within the image taken under UV illumination.

1.3 Methodologies Proposed and Contributions

The findings of this work are relevant for scholars and for document image analysis systems. Contributions include designing new enhancement and binarization methods as well as evaluating state-of-the-art binarization methods on a new dataset. The fundamentals of the methods proposed are described in the following.

1.3.1 Enhancement

In Section 3.2 an enhancement method is detailed that is published in [HGS13]. While previous works [JCK11], [STB07] apply unsupervised dimension reduction methods, the enhancement method proposed makes use of the supervised LDA approach. The LDA classifier requires labeled training data. Therefore, an approach is proposed, which automatically selects foreground and background samples. The enhancement method is especially designed for faded-out writings, which are barely visible within the multispectral channels. The writings are enhanced by applying the PCA transformation. Hence, the labeling is conducted on PCA images. Due to bad condition of the investigated manuscripts, the PCA images are corrupted by background variation and the faded-out text is only partially visible. Hence, a labeling by applying a binarization method is error-prone. Instead, the labeling step is based on the detection of text lines, since it was found experimentally that the text lines are better recognizable than single characters. The method is evaluated in a qualitative analysis that is conducted by philologists.

It should be noted that the LDA based dimension reduction could be replaced by other supervised techniques, including for example Support Vector Machines (SVM)'s or Random Forests. A non-linear combination of the multispectral channels might also increase the legibility and hence Neural Networks could also be applied on the multispectral data. These techniques might gain a further legibility increase, compared to a LDA based linear combination. However, an evaluation of other supervised techniques is out of the scope of this work, since this requires an additional human effort of philologists. Instead the contribution of this work is to show that an automated labeling of multispectral samples can be used for the successful application of supervised dimension reduction methods or classifiers.

1.3.2 Binarization

Different binarization methods for multispectral document images are introduced in this work: They are based on the following three underlying concepts:

Target Detection First, two binarization methods are introduced in Section 4.1, which are published in [HDS15] and [DHS16]. The methods make use of a target detection method stemming from the field of remote sensing [MTP+14]. Target detection methods are concerned with the identification of specific spectral signatures, named targets, within the acquired data. The binarization methods proposed in [HDS15] and [DHS16] combine a specific target detection method with a traditional binarization method [SLT10]. Two

different combination techniques have been developed and they are introduced in Section 4.1.1 and Section 4.1.2 respectively.

GMM A further binarization method is detailed in Section 4.2: The method is proposed in [HDS18] and it makes use of the traditional binarization technique [SLT10] and combines its output with the result of an GMM based clustering step. While, GMM's are used for the binarization of grayscale images [MP15], this is the first time that they are used for MSI data. It is shown in Section 4.2.2, that a GMM based clustering is sensitive to varying background intensities, stemming from uneven illumination or background noise. Therefore, a background compensation step is applied in order to resolve this drawback. The numerical evaluation shows that this preprocessing step has a major benefit on the binarization performance.

FCN Given the recent success of deep learning, a FCN based method is used for the binarization of multispectral document images in [HBS19]. While the architecture of the Convolutional Neural Network (CNN) used is a state-of-the-art approach, it is the first time that a deep neural network is used for the binarization of multispectral images of historical documents - to the best of my knowledge. This can be attributed to the circumstance, that the size of the MS-TEx training set is relatively small, i.e. 21 multispectral images, whereas deep learning-based methods typically make use of large training sets [ZPIE17]. In order to overcome this limitation an own dataset has been published in [HBS19], which consists of a larger training set, i.e. 80 multispectral images. It is shown experimentally in Section 4.3 that the size of the training set introduced is sufficient for a successful training of an FCN. Thus, the dataset allows for the application and development of further deep learning-based binarization methods. Additionally, the dataset allows for the application of ink separation methods, because it is comprised of two different ink types.

1.4 Thesis Structure

This thesis is structured as follows: First, the related work is detailed in Chapter 2: MSI systems that are used for the acquisition of historical and modern documents are introduced in Section 2.1, as well as ensuing post-processing techniques. Afterwards, binarization methods for grayscale images are introduced in Section 2.1. Binarization techniques that are especially designed for multispectral document images are detailed in Section 2.2.

In Chapter 3 the MSI system used is described and the enhancement method proposed is introduced: First, the fundamentals of multispectral imaging and the system used are depicted in Section 3.1. The enhancement method is explained in Section 3.2 and its usability as a preprocessing step within an OCR system is shown in Section 3.3.

In Chapter 4 the binarization methods proposed are first described in Sections 4.1, 4.2 and 4.3. Afterwards, a numerical evaluation of the techniques is given in Section 4.4. Finally, the thesis is summarized in Chapter 5.

CHAPTER 2

Related Work

This chapter provides an overview on MSI systems used for acquisition of document images and on binarization methods designed for grayscale and multispectral images. First, MSI and HSI systems for document images are described in Section 2.1. The imaging of historical document images is detailed in Section 2.1.1. The acquisition of modern documents is especially useful in the context of forgery detection and ink differentiation and is described in Section 2.1.2. The systems are designed for particular applications and multiple post-processing techniques have been developed for these applications. These methods are also summarized together with the corresponding imaging systems.

The remaining sections of this chapter are concerned with document image binarization. The binarization approaches presented in the following can be categorized into two groups: Methods designed for grayscale images and techniques that are especially developed for multispectral images. The methods belonging to the latter category are influenced by techniques for grayscale images. Hence, the methods designed for grayscale images are summarized first in Section 2.2. Classical approaches are discussed, as well as the winning methods of the DIBCO and H-DIBCO events. Following the current trend in binarization [PZK+19a], deep learning-based methods are described in an own section, namely Section 2.2.3.

Finally, binarization methods designed for MSI data are described in Section 2.3. This thesis is mainly concerned with the binarization of multispectral document images. Therefore, related approaches are explained in more detail, compared to the methods that are developed for grayscale images.

2.1 MultiSpectral Imaging

Spectral imaging of documents can be categorized based on the document types: Historical manuscripts and modern documents. This section follows this categorization. First,

imaging systems designed for historical and partially degraded writings are summarized in Section 2.1.1. Additionally, post-processing methods that are especially designed for ancient and degraded manuscripts are explained. Afterwards, imaging and processing of modern document images are detailed in Section 2.1.2.

2.1.1 Historical Manuscripts

The systematic photography of ancient documents started in the 19th century and the first descriptions of dedicated imaging systems can be found in [PG01] and [Kö14]. Kögel [Kö20] proposes to use UV illumination sources in order to enhance the contrast of faded-out palimpsest writings. Additionally, Kögel proposes to combine white light images with UV fluorescence images in order to remove the overwritten text in the resulting image. Figure 2.1 shows the imaging system used and exemplary results gained by the enhancement method. The imaging technique is non-invasive and replaced the invasive practice of applying destructive chemicals on the historical manuscripts [Pri79].

Originating from the field of remote sensing, modern spectral imaging of historical objects has been developed around 20 years ago [FK06]. The multiple spectral channels are either acquired by using a narrow-band LED illumination, or by filtering the reflected light with optical filters [Ber19]. According to Berns [Ber19], multispectral systems have a high spatial resolution and a low spectral resolution (i.e. a low number of channels), whereas hyperspectral systems have a lower spatial and a higher spectral resolution. Unfortunately, there is no official definition for MSI and HSI existing and both terms are used almost interchangeably [HK13].

While the majority of the historical objects imaged are paintings - for example [ABD+13] - MSI and HSI are also successfully used for the investigation of historical documents [FK06]. One of the most prominent historical documents imaged is the Archimedes palimpsest [FK06]. This document contains treatises of Archimedes and has been reused as a prayer book. Easton et al. [JCK11] have imaged the book with a MSI system that makes use of narrow-band LED illumination. By using this narrow-band lighting system, the thermal stress put on the manuscripts is reduced, compared to broadband tungsten illumination. The illumination system used provides 12 different spectral ranges from UV to NIR and the acquired multispectral images have a spatial resolution of 88 MegaPixel (MP).

The authors propose to apply PCA on an UV fluorescence image with three channels (RGB) in order to increase the legibility of the iron gall-based writing. Easton et al. note that the underwritten text is most visible in the second or third principal component. Since the underwritten text is enhanced by multiple principal components, Easton et al. suggest the merging of the three components into pseudo-color images. Furthermore, the authors propose to manually rotate the hue angle in order to increase the local contrast in the pseudo-color images. Figure 2.2 shows a manuscript portion of the Archimedes palimpsest. It is notable that the original underwritten text is best visible in the PCA pseudo color image.

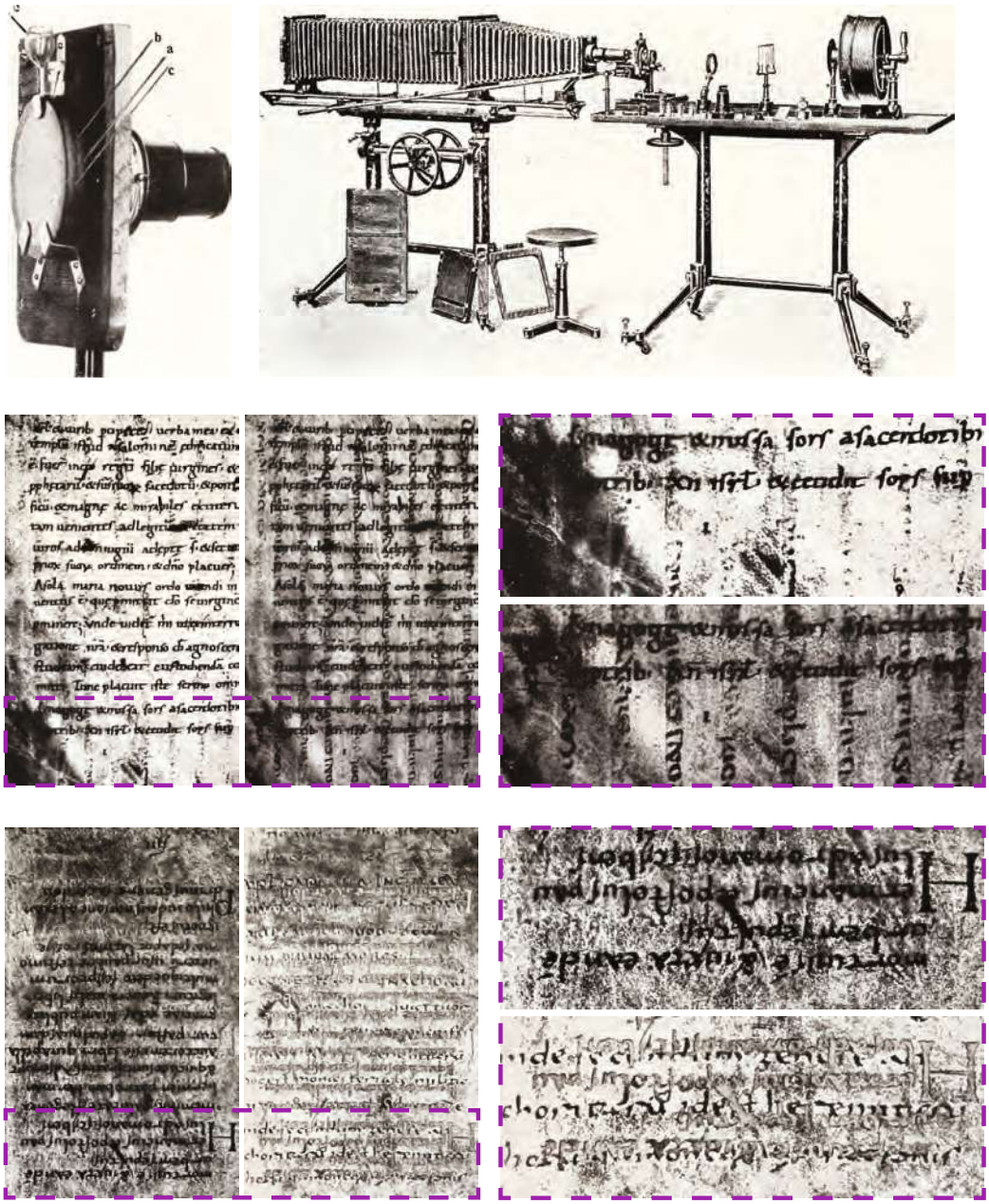


Figure 2.1: Historical palimpsest photography [Kö20]. (Top row) Camera with optical filter and illustration of the imaging system. (Middle row) White light and UV fluorescence images. (Bottom row) UV fluorescence image and enhancement result gained by combining a white light image and a UV fluorescence image. Note that the overwriting is removed in the enhancement result.

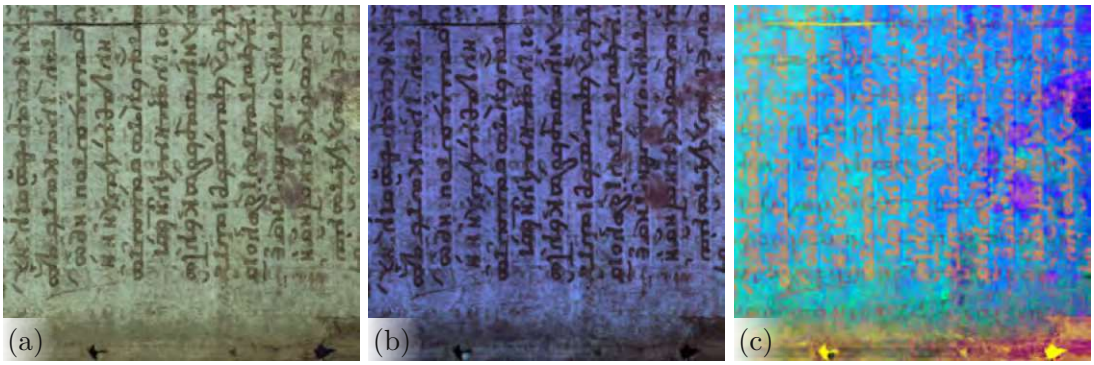


Figure 2.2: Portion of the folio 120 verso belonging to the Archimedes palimpsest. (a) Image acquired with tungsten illumination. (b) UV fluorescence image. [Arc] (c) PCA pseudo color image [Eas].

The Archimedes palimpsest is also investigated in [STB07]: While Easton et. [JCK11] apply the PCA transformation on an UV fluorescence image, Salerno et al. [STB07] apply the transformation on an entire multispectral image. Additionally, the authors suggest to apply the ICA transformation [HO00] on the multispectral images. The authors note that the first five eigenvalues are dominant and they propose to reduce the data with the PCA transformation to five components. This dimension reduction is used as a preprocessing technique for the ensuing ICA transformation.

Hedjam and Cheriet [HC13b] propose another enhancement method for multispectral document images. The MSI system used consists of a 6 MP camera, tunable lamps and a filter wheel providing 8 different spectral ranges between UV and NIR. The authors note that the iron-gall based ink is visible in the visible channels. The ink vanishes in the NIR spectral range, while corruptions like stamps, or bleed-through are still visible in this spectral range. Therefore, the enhancement method determines, which NIR exhibits no text but contains corruptions. The corrupted regions are found by applying a document image binarization method [SP00]. The regions found are then restored in the visible channels by applying a Total Variation (TV) based inpainting algorithm. Thus, character strokes that are partially occluded by corruptions (like stamps) can be restored by propagating the stroke regions into the damaged regions.

Kim et al. [KDB11] propose a different enhancement technique for documents that are captured with a HSI system. The system used allows for an imaging in 70 spectral ranges with a bandwidth of 10 nm, from UV to NIR. The images have a resolution of 4 MP for an area of 125 mm $\times$ 125 mm. The enhancement method proposed is designed for writings that are visible to the human eye, but are corrupted by ink-bleed, ink-corrosion and foxing. These degradations are less prevalent in the NIR range and Kim et al. suggest to fuse NIR images with RGB images. The method proposed detects degraded image regions and then composites the enhanced gradient map from a NIR image. Finally,

the resulting image is constructed from the composited gradients. In Figure 2.3 two enhancement results are shown, along with the original images.

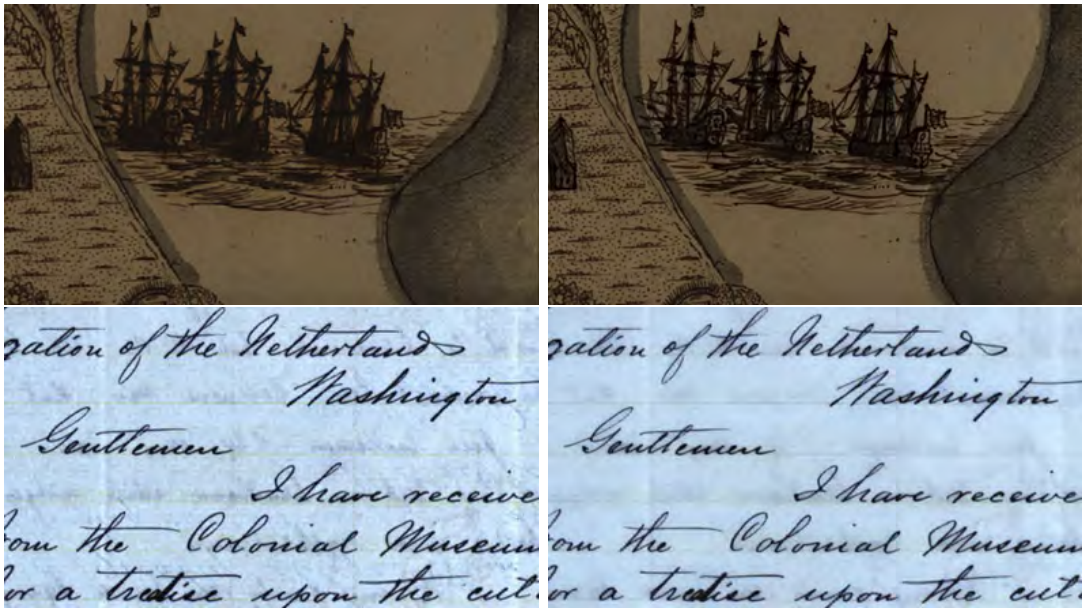


Figure 2.3: Enhancement result of Kim et al. [KDB11]. Original images (left) and corresponding fusion results (right).

George and Hardeberg [GH15] use a hyperspectral line scanning spectrometer for the acquisition of historical documents. The system acquires 160 spectral bands ranging from 400 nm to 1000 nm, and the scanned lines have a spatial resolution of 1600 pixels. The authors make use of spectral data in order to separate two different inks contained in a historical document originating from the 19th century. Figure 2.4 shows an illustration of the hyperspectral data cube and a plot of the spectral signatures of the inks. George and Hardeberg apply Spectral Angle Mapper (SAM) [KLB+93] as well as Spectral Information Divergence [Cha] in order to measure the distance between both signals. The authors show that these techniques can be used for the successful separation of both inks. An illustration of the hyperspectral images and a plot of the spectral signatures of the inks are given in Figure 2.4.

Hedjam et al. [HCK14] propose a different enhancement technique for handwritten text that is barely visible under regular white light illumination. The method is designed to emphasize text regions, while suppressing the remaining document background regions. Therefore, a linear filter is applied on the MSI data in order to emphasize the degraded writing. The filter technique is stemming from the field of remote sensing and is named Constrained Energy Minimization (CEM). In order to compute the filter, it is necessary to define a spectral signature, which is enhanced by the filter. Hedjam et al. propose to estimate this spectral signature with a self-referencing strategy: The strategy makes use of the spectral signature of handwriting in the test images as well as of spectral signatures

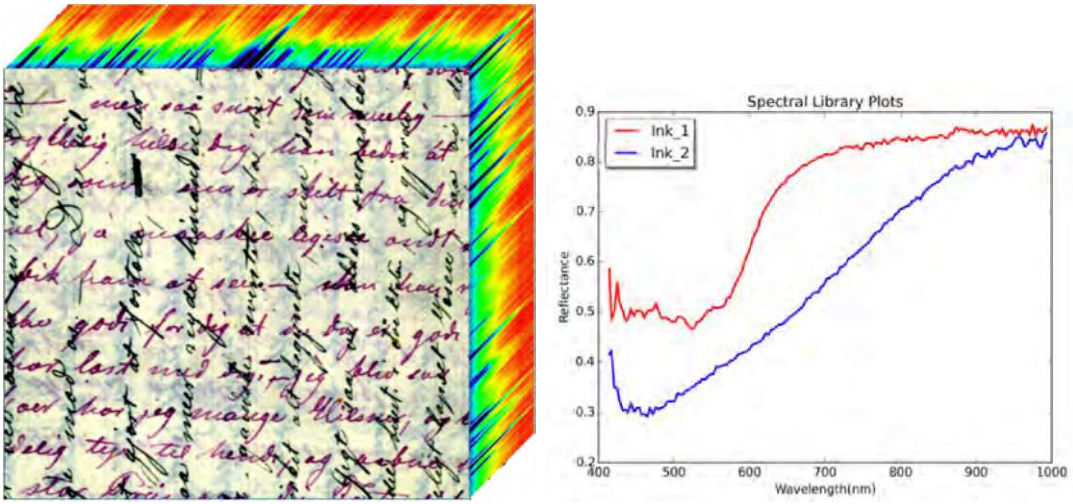


Figure 2.4: Hyperspectral data cube illustration (left) and spectral signatures of the two different inks (right) [GH15].

contained in a training database. Resulting images of the method are compared to images obtained by applying the ICA algorithm in [HO00] and gamma contrast-correction. The resulting images show that the method achieves a superior visibility enhancement.

Lettner et al. [LKSM08] propose another enhancement method for multispectral images of barley visible handwritings. The MSI system used consists of a 12 MP multispectral camera, a SLR camera, broadband and UV illumination and a filter wheel that contains 8 different filters. In a first step, a binary mask is computed that encodes text line regions. For the text line detection, the document image is binarized and connected components are grouped into a text line. Additionally, a-priori knowledge about the ruling scheme is used to add missing lines or to expand estimated lines. The binary mask found is then used as a weight mask that is required by the Multivariate Spatial Correlation (MSC) approach: This method is introduced by Wartenberg [War85] and is used to emphasize regions in multi- or hyperspectral images. Lettner et al. [LKSM08] show resulting images of their method that clearly outperform resulting images obtained with PCA. One example output is given in Figure 2.5 along with the corresponding multispectral images.

In [RDH18] a bleed-through removal algorithm for hyperspectral document images is proposed. The hyperspectral imaging system used is similar to the one in [GH15]. The authors combine Canny [Can86] edge detection and k-means to determine the regions that are affected by bleed through. The regions found are afterwards restored by applying an image inpainting algorithm [Tel04].

Marengo et al. [MMZ+11] propose to apply MSI in order to monitor the conservation of documents written on parchment. The authors state that PCA can be used in order to detect the degradation process of historical documents: First PCA is applied on the

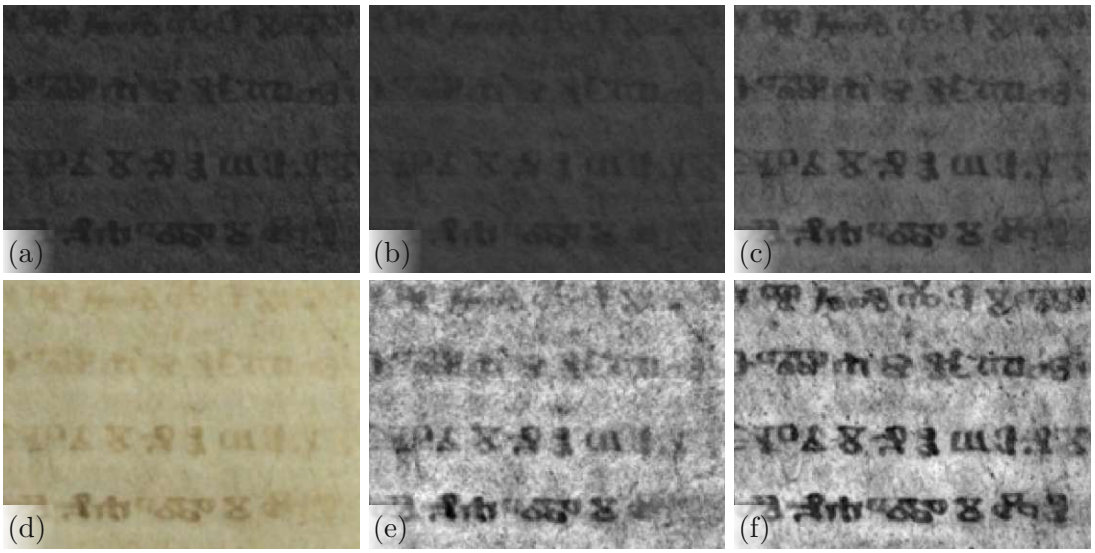


Figure 2.5: Enhancement results of Lettner et al. [LKSM08]. (a) UV reflectography. (b) UV fluorescence. (c) 450 nm. (d) RGB image. (e) PCA result. (f) MSC result.

original data and the relevant principal components are then used to reconstruct the original data. The difference between the original and the reconstructed data are the residuals and Marengo et al. [MMZ+11] state that these residuals can be analyzed in order to detect the degradation process of historical documents. The effectiveness of the method is demonstrated on a parchment document belonging to the Dead Sea scrolls.

MacDonald et al. [MGC+13] monitor the degradation of parchment documents: Therefore, parchment belonging to a historical document was degraded by multiple physical and chemical treatments. The samples were imaged before and after treatment with a multispectral imaging system and the collected database is made publicly available.

2.1.2 Modern Documents

While the works mentioned above are concerned with the investigation of historical manuscripts, spectral imaging is also used for the analysis of modern documents. Silva et al. [SPH+14] make use of HSI for the forensic analysis of document forgery. Hyperspectral images are acquired in the NIR range from 928 to 2524 nm. Artificial samples have been generated for three different types of forgeries: Obliteration by overwriting, adding text and intersecting lines. The authors apply PCA among other statistical tools in order to reveal the frauds. In case of overwritten text, 39 out of 90 overwritten texts are reconstructed. For the remaining two forgery types, over 80% of the forgeries are correctly classified, which indicates the potential of hyperspectral imaging for forgery detection. Figure 2.6 shows two example images, which contain forged text that was

added to an original writing: The forgery is not noticeable within the visible range, but the first and second PCA image indicate that writings are written with two different inks.

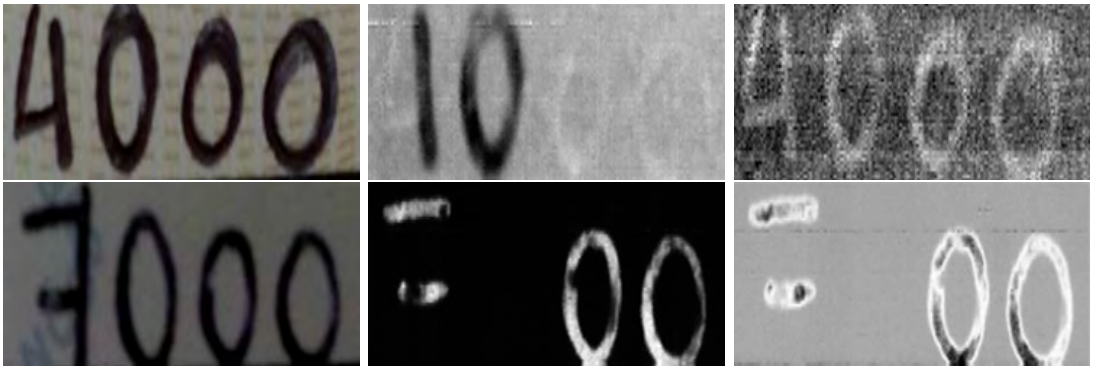


Figure 2.6: Forged numbers. From left to right: Image taken in the visible range, first principal component, second principal component. Based on a courtesy of Silva et al. [SPH+14].

Khan et al. [KSM13] introduce a hyperspectral image database of handwritten text written with various blue and black inks. The database contains 70 hyperspectral images of samples that contain a single sentence. The sentence is written either with multiple blue inks or with multiple black inks, which appear visually similar. The images have been captured with a 4 MP sensor and 33 spectral bands are acquired by using a tunable filter, which allows for an imaging between 400 and 720 nm. Khan et al. propose a simple approach for the identification of the different inks: First, the foreground is found by applying the binarization method of Sauvola and Pietikäinen [SP00]. Afterwards the k-means clustering algorithms is applied on the hyperspectral foreground samples, whereby the number of clusters k, corresponds to the number of contained inks. The authors show that this approach is capable of differentiating between two different inks contained in a questioned document. If the k-means algorithm is instead solely applied on corresponding RGB images it fails, because the inks are difficult to distinguish in the visible range.

In a follow-up paper, Khan et al. [KSM15] apply joint sparse PCA for dimension reduction and a band selection technique before k-means is applied on the reduced dataset. By applying these dimension reduction steps, the accuracy is improved up to 15%.

Luo et al. [LSM15] stress out that the approach of Khan et al. [KSM15] assumes that the same amount of text is written by the different inks. The authors state that in practical applications - such as document fraud - only minor modifications are performed. A further drawback of the approach of Khan et al. [KSM13] is that the number of contained inks, which is required for the k-means initialization, is manually set. Luo et al. argue that these two disadvantages limit the applicability of the method in [KSM13]. Instead, Luo et al. propose to combine anomaly detection algorithms [JTHW06] with unsupervised clustering techniques. The effectiveness of the method is demonstrated on

the dataset introduced in [KSM15].

A deep learning-based approach is proposed by Khan et al. in [KYAK18]. The authors propose to reshape the hyperspectral pixels with 33 dimensions to a 6x6 image patches, which are classified by a CNN. It is notable that by using this procedure, only spectral information is used and spatial information is neglected.

This drawback is overcome in a recent work of Khan et al. [KKS19], where spatial information is also used by a CNN: Therefore, the spectral signature of a pixel is considered together with the signatures of eight pixels in the local neighborhood. The authors show that by using the spectral information of the neighboring pixels, the overall performance is increased.

Malik et al. [MAS+15] apply HSI for the purpose of signature extraction in modern documents. The documents have been imaged within a spectral range between 400 and 900 nm and the approach is based on the observation that the handwritten signature vanishes in the NIR range whereas the machine printed text is visible in all spectral ranges. The authors apply the Speeded Up Robust Features (SURF) [BTG06] keypoint detector on all hyperspectral images. Afterwards the two images having the lowest and the highest number of detected keypoints are selected. The authors claim that the image with the lowest number of keypoints exhibits solely the machine printed text, whereas the image with the highest number of keypoints is the image where the signature is best visible. The two images are afterwards subtracted in order to extract the signature.

Butt et al. [BAS+16] propose an optimized version of the method by Malik et al. [MAS+15]. In a first step, Connected Components (CC)'s are found and for each connected component SURF [BTG06] features are computed. The features are afterwards classified as machine-printed text or handwritten text. If a CC contains both feature types, the component is assumed to belong to a region where the signature is overlapping the machine printed text. The signature in such overlapping regions is then segmented using the approach of Malik et al. [MAS+15]. The authors report a significantly increased recall compared to the method of Malik et al. [MAS+15].

Devassy et al. [DGH19] recently published a work in which five similarity measures are used for the separation of inks in modern handwritings. According to [DGH19] these similarity measures are used for the hyperspectral classification of remote sensing data, but only seldomly for ink classification - as in [GH15]. The evaluation is performed on samples containing 10 different ink types, whereby the number of samples is unfortunately not given. The HSI system used consists of a line scanner with a resolution of 1800 pixels and allows for imaging in 186 different spectral ranges from 400 to 1000 nm. The authors report that the highest accuracy is gained by the SAM technique.

2.1.3 Evaluation

The ink identification in modern documents (in [KSM13], [KSM15], [KYAK18], [KKS19], [LSM15], [DGH19]) is evaluated by comparing the reference ground truth with resulting

images. This is a pixel-based evaluation, since each pixel in the segmentation result is compared with the corresponding ground truth pixel. The performance is measured in terms of accuracy.

Contrary, there is no direct metric existing, which measures the enhancement of multispectral document images. The works mentioned in Section 2.1.1 provide solely exemplary resulting images and no numerical results, except for Hedjam and Cheriet [HC13b]: The authors evaluate their restoration method indirectly by means of binarization: Therefore, unprocessed multispectral images are binarized and the enhancement results. The performance gain is measured in terms of F-Measure (FM).

Arsene et al. [APA+16] applied 15 different dimensionality reduction methods on one multispectral image of an ancient palimpsest. The resulting images were assessed by 7 scholars, whereby each philologist assigned a numerical value to each resulting image. The dimension reduction method that gained the highest - and best - score was the Canonical Variates Analysis (CVA) technique, which is a supervised method for dimension reduction. To the best of our knowledge, the work in [APA+16] compares the highest number of dimension reduction methods, but it is not assured that its finding can be generalized, because only one document page has been assessed by scholars.

2.2 Binarization of Grayscale Images

This section is concerned with the binarization of grayscale images. First, classical approaches are introduced. More recent methods, including deep learning-based methods are summarized afterwards. Additionally, an overview on competitions for document image binarization is given.

2.2.1 Thresholding Methods

The general aim of document image binarization is to assign a label to each pixel, which indicates whether the pixel is belonging to the document foreground or background. In this section classical binarization methods are introduced that make use of global or local thresholds. Such thresholding techniques can be formally described by:

$$B(x,y) = \begin{cases} 1 & I(x,y) < T(x,y) \\ 0 & I(x,y) \geq T(x,y) \end{cases} \tag{2.1}$$

Global Threshold

In case of global thresholds the threshold remains constant over the entire image region ($T(x,y) = T$). Otsu [Ots79] introduce a global threshold technique that still remains popular [TM20]: The algorithm finds the threshold T by minimizing the intra-class variance and maximizing the inter-class variance. The Otsu threshold is not especially

designed for document images, but instead for a general foreground segmentation of natural images.

Cheriet et al. [CSS98] note that the method of Otsu [Ots79] is only suitable in case of homogenous foreground and background classes. In order to overcome this drawback, the authors suggest an extension, which allows for the segmentation of multiple classes and is especially designed for document images. The algorithm iteratively searches for histogram peaks and applies Otsu's [Ots79] method. This procedure is recursively applied until no new histogram peaks are found or the area of the image regions found is smaller than a predefined threshold. The method is successfully applied on document images that are corrupted by background variation.

Entropy based methods are another kind of global thresholding methods: The method in [Pun81] chooses the threshold value T as the value, which maximizes the entropy of the one dimensional histogram. Abutaleb [Abu89] notes that thus spatial information is neglected and extends the entropy-based thresholding to the two dimensional histogram.

The methods mentioned use a constant threshold and hence they are only suitable for document images that have a uniform background. However, they are usually not suitable for degraded documents, since these documents often do not possess a clear bimodal pattern [SLT13].

Local Threshold

Niblack [Nib85] suggest to overcome this drawback by making use of a local threshold that is defined as:

$$T(x, y) = m(x, y) + k * s(x, y), \tag{2.2}$$

where k is a constant factor (-0.2), m is the local mean value and s is the variance within a local window. It is stated by Gatos et al. [GPP06] that the algorithm is not sensitive on the choice of the size of the window, as long as it covers 1-2 characters. Gatos et al. [GPP06] also note that the method is not capable of efficiently removing background noise.

Similarly, Sauvola and Pietikäinen [SP00] note that the algorithm of Niblack [Nib85] fails in case of light background regions, where the threshold T is exceeded. Therefore, Sauvola and Pietikäinen [SP00] suggest an improvement over the original algorithm in [Nib85] to overcome this drawback:

$$T(x, y) = m(x, y) \cdot \left[1 + k \left(\frac{s(x, y)}{R} - 1\right)\right], \tag{2.3}$$

where R is the dynamic range of the standard deviation and $k = 0.5$. It is stated in [GPP06] and [Ten19] that the algorithm suggested by Sauvola and Pietikäinen [SP00]

is more robust to background noise compared to the original algorithm proposed by Niblack.

Figure 2.7 shows exemplar outputs gained by the methods of Otsu [Ots79], Niblack [Nib85] and Sauvola and Pietikäinen [SP00].

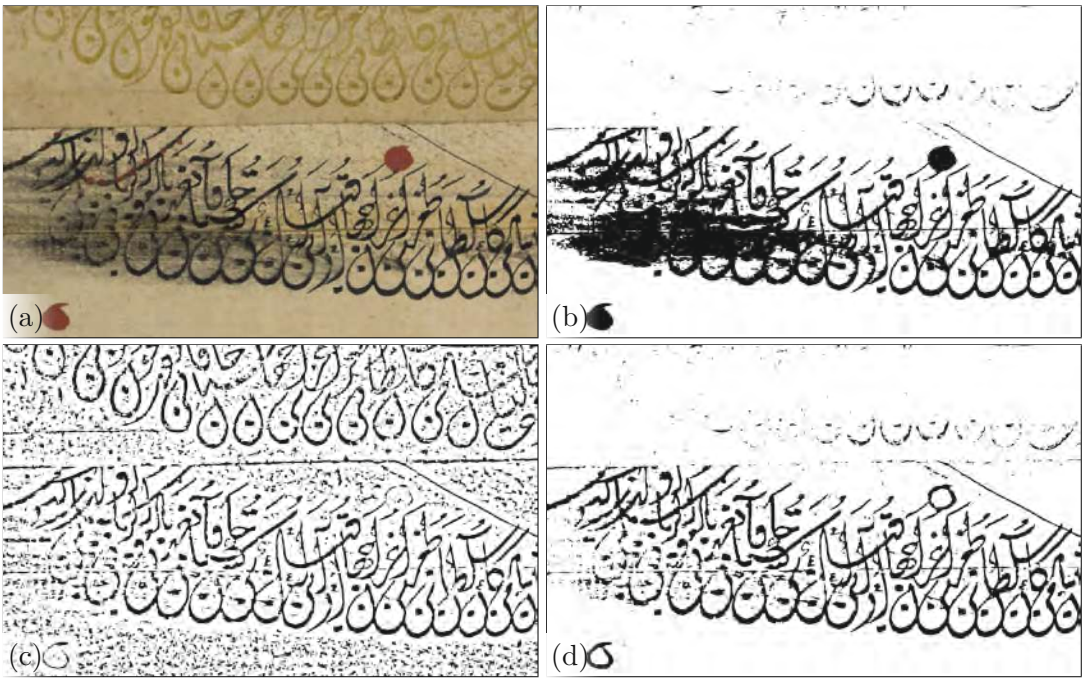


Figure 2.7: Binarization using global and local threshold techniques. (a) Original RGB image. (b) Result of Otsu [Ots79]. (c) Result of Niblack [Nib85]. (d) Result of Sauvola and Pietikäinen [SP00].

Shafait et al. [SKB08] propose an efficient implementation of the algorithm in [SP00]: The authors suggest to compute the mean and variance used in Equation 2.3 on integral images, which have been popularized in the field of computer vision [SKB08] by the work of Viola and Jones [VJ04]. Thus, the computational complexity is reduced from $O(W^2N^2)$ to $O(N^2)$.

Lazzara and Géraud [LG13] state that the method in [SP00] is sensitive to the size of the windows and propose a multiscale extension. Thus, the approach can be successfully applied on document images with varying font size, like for instance newspaper articles.

Another multi-scale approach for binarization is introduced in [MC10] and [MC12] by Moghaddam and Cheriet: The authors propose to use an adaptive and local version of Otsu's [Ots79] method in combination with a background estimation step. Thus, the performance is significantly increased, compared to the original algorithm of Otsu [Ots79].

Wolf et al. [WJC02] suggest an extension to the algorithm in [SP00]:

$$T(x,y) = m(x,y) - k\left(1 - \frac{s(x,y)}{R}\right)(m(x,y) - M). \tag{2.4}$$

By using the above formula, the contrast and mean intensity values are normalized. Thus, the algorithm is better suited for images with limited contrast [Ten19] compared to the original algorithm.

2.2.2 Miscellaneous Methods

The methods mentioned in the previous section calculate the threshold $T(x,y)$ directly on the unprocessed image intensities or on the corresponding first (mean) and second statistical moments (variance). Contrary, in the current section, methods are summarized that combine the aforementioned thresholding techniques or make use of other image processing techniques.

For example, Gatos et al. [GPP06] propose an adaptive binarization method that makes use of Sauvola's [SP00] algorithm: In a first step the input image is preprocessed by applying a low pass Wiener filter [Jai89] in order to enhance the contrast and to reduce noise. The resulting image is used to estimate the foreground with Sauvola's algorithm [SP00]. Afterwards, the background is estimated, whereby the unknown regions belonging to the estimated foreground are interpolated based on the known neighboring pixels. The preprocessed image is then subtracted from the background estimation and the resulting image is used in the next binarization step: If a gray value intensity of the resulting image exceeds a certain threshold the pixel is labeled as belonging to the foreground. The threshold is adaptive and is decreased for darker background regions.

Background estimation is also conducted by the algorithm in [NGP14]: First the input image is binarized using the approach suggest by Niblack [Nib85]. Afterwards, the foreground found is used as an inpainting mask and the background is propagated into the unknown regions by applying a simple inpainting technique introduced by the authors. The background estimation image and the input image are then combined and normalized in order to compensate background variations. The normalized image is then used to assist a global thresholding technique [Ots79] and a local [Nib85] thresholding method. The resulting images are afterwards combined in the final binarization step.

The methods above decide if a pixel is belonging to the foreground or background based on a threshold. Contrary, Hedjam et al. [HC11a] suggest a framework, which uses a soft decision based on a probabilistic model. The work assumes a Gaussian distribution of both classes and uses a Maximum Likelihood (ML) approach to assign a label to a pixel. The algorithm starts with the computation of an under-segmented binarization output by applying the method of Sauvola and Pietikäinen [SP00]. This output is then used to calculate the local mean and variance of both classes, which are needed for the final ML decision: This decision is dependent on the local mean and variance of the background as

well as on the local mean and the global and local variance of the foreground. Numerical results are presented, which show that the method outperforms the methods that have participated in the DIBCO 2009 contest. The authors note that they were available of the DIBCO 2009 dataset, whereas the participants were not.

The methods mentioned are directly applied on gray value intensities or on the mean or variance of these intensity values. Contrary, Sehad et al. [SCC14] propose to process the input image with a Gabor filter bank. The authors show that the performance of the binarization methods in [Nib85] and [SP00] is increased by applying the binarization methods on Gabor filtered images. However, the performance of Wolf et al. [WJC02] is decreased by applying the method on Gabor filtered images.

Multiple binarization methods have been proposed that make use of local image contrast or image gradients. An early binarization method falling in this category is suggested by Bernsen [Ber86]. The author defines the local contrast formally as:

$$C(x, y) = I_{max}(x, y) - I_{min}(x, y), \tag{2.5}$$

whereby $I_{max}(x, y)$ and $I_{min}(x, y)$ denote the maximum and minimum intensity value within a local neighborhood. The local contrast $C(x, y)$ is then used in a rule based binarization framework.

Su et al. [SLT13] note that the method is simple but not effective on degraded documents. Instead, they suggest another binarization method in [SLT10] that makes use of the local contrast as defined by:

$$C(x, y) = \frac{I_{max}(x, y) - I_{min}(x, y)}{I_{max}(x, y) + I_{min}(x, y) + \epsilon}, \tag{2.6}$$

whereby the infinitely small number ϵ avoids a division by zero. It is notable that Su et al. [SLT10] introduce a normalization term in the denominator that is used to compensate background variation. The algorithm proceeds with the detection of so-called high contrast pixels, which are typically located at stroke boundaries. These high-contrast pixels are found by applying a global Otsu [Ots79] threshold on the local contrast image. Afterwards, a pixel is classified as a foreground pixel, if there is a sufficient number of high-contrast pixels in its local neighborhood and if its gray value intensity is below the average intensity within the local neighborhood. It should be noted that the size of the local neighborhood is dependent on the dominant stroke width, which is estimated based on the local contrast image.

Kleber et al. [KDS11] stress out that the method of Su et al. [SLT10] relies mainly on the estimation of the stroke width parameter. The authors suggest a parameter independent approach that makes use of a scale space. The input image is binarized at every stage with the method of Su et al. and the information is propagated through the scales. Thus, the performance of the original algorithm is significantly increased.

The authors of [SLT10] have also proposed another way of finding pixels located at stroke boundaries in [LST10]. Instead of using Equation 2.6 the authors propose to find the high contrast pixels based on the L1-norm image gradient. Various document degradation (such as background variation) are compensated in a background estimation step. The overall foreground classification is then fulfilled by applying a rule based system. The method in [LST10] ranked first in the DIBCO 2009 contest.

Another binarization method of Su et al. is published in [SLT13]. The authors propose to combine the image encoding the high-contrast pixels based on Equation 2.6 with a Canny [Can86] edge image. Both binary images are multiplied in order to improve the detection of pixels located at stroke boundaries. Afterwards, the rule based heuristic described above is applied together with a post-processing step. The binarization technique ranked first in the H-DIBCO 2010 contest and a similar approach of Su et al. ranked first in the DIBCO 2013 competition.

Jia et al. [JSH+18] propose a binarization method that is also based on the detection of stroke boundaries. First, a background removal step is performed to compensate local variations. Afterwards the gradient magnitude is calculated and thresholded in order to determine the Structural Symmetric Pixels (SSP) candidates. The SSP pixels are pixels located at stroke boundaries and hence the corresponding gradient magnitude values assume local minima. Furthermore, the SSP's located on the left and right of a character stroke have gradients that point in opposite directions. The authors use this property and define decision rules to detect the SSP's. Similar to the methods proposed by Su et al. [SLT10], [SLT13] the stroke width is estimated and used in the heuristic determination of the SSP's. The SSP's are then used for the calculation of local thresholds that are required in the final binarization step.

Mitianoudis and Papamarkos [MP15] propose a framework consisting of three stages. First, background removal is performed. Afterwards, feature vectors are constructed for the pixels based on a local neighborhood representation that makes use of local contrast amongst other properties. The feature vectors are afterwards clustered using GMM based clustering, whereby the two Gaussians are used to describe the foreground and background. The output of the clustering step is then postprocessed in a final step in order to remove small CC's that are considered as binarization noise. The method achieved the fourth rank in the H-DIBCO 2014 contest.

The winning method of the DIBCO 2011 contest is proposed by Lelore and Bouchara in [LB13]: First pixels located at stroke boundaries are detected with the Canny [Can86] edge detection algorithm. Pixels in the near of the edges are then labeled as foreground or background based on the output of a clustering step. The remaining pixels are labeled as unknown. In the final binarization step, the CC's of the unknown regions are found. Each CC is then binarized based on the foreground to background ratio of its boundary pixels. The algorithm is especially designed for fast execution and the authors report an average runtime that is only eight times slower than the optimized Sauvola [SP00] algorithm introduced in [SKB08].

Howe [How11] suggests a Markov Random Field (MRF) model in which a graph cut model [BK04] is used for binarization: The energy function used consists of a data-fidelity term that relies on the Laplacian of the input image and a smoothness term that is based on the Canny [Can86] edge image. The labeling is then fulfilled by minimizing the energy function. In [How12] the algorithm is extended in such a manner that the parameters used are automatically fine-tuned based on a stability criterion heuristic. This approach won the H-DIBCO 2012 competition and gained the second rank in the DIBCO 2013 competition.

Mesquita et al. [MSMM15] propose a combination of the method proposed by Howe [How11] and an own thresholding method [MMA14] that is based on visual perception theory [HW12]. The authors make use of a racing algorithm [BYBS10] in order to select a parameter configuration. The parameters are tuned on the DIBCO 2011 dataset and the method won the H-DIBCO 2014 competition.

The winner of the H-DIBCO 2016 contest are Kligler and Tal [KKT18]. The authors propose to transform each 2D image point $I(x, y)$ into a 3D point with the coordinates $(x, y, I(x, y))$. This set of 3D points is then linearly transformed on spherical surface, whereby peaks correspond to bright image regions and local minima correspond to dark image regions. The regions latter mentioned are assumed to belong to text regions and they are found by applying an algorithm for visibility detection [KTB07]. The output of this step is a two-dimensional visibility score map, which can be used as an input for other binarization methods. The authors utilize the method of Howe [How12] and show that the performance is significantly increased by transforming the input image with the method proposed.

It is notable that in the H-DIBCO 2016 contest one method was submitted by Tensmeyer that uses an FCN [LSD15]. This was the first time that a deep learning-based method participated in a competition on document binarization. In the DIBCO 2017 contest, 6 out of 17 participating methods were based on deep learning and the winning method [BIN19] makes use of a FCN. The next section covers binarization methods that belong to this category.

Approximately a third of the participating methods of the DIBCO 2017 competition was based on deep learning. Contrary, no deep learning-based method was submitted to the H-DIBCO 2018 contest. Instead, the winning method submitted by Xiong et al. [XJX+18] applies an adapted version of the binarization method of Howe [How12] on a background compensated image. The background compensation is achieved by applying morphological bottom-hat transform, whereby the radius of the disk-shaped mask is calculated with the Stroke Width Transform [EOW10].

Color Conversion The majority of the binarization methods are designed for grayscale images and RGB images are usually converted to grayscale images by using simple conversion methods - such as the luminosity method [TM20], [HNKC15], which simply weights the three channels according to their wavelengths. Hedjam et al. [HNKC15]

stress out that these simple methods are designed for natural images and that they aim for preserving the contrast between different classes. Contrary, the method of Hedjam et al. [HNKC15] reduces the contrast within the text class. For this purpose, a linear filter is learned on a training set and the method is used to convert approximately 50 RGB images. Multiple binarization methods (including [How12], [LST10], [GPP06]) are applied on the converted images and on images, which are converted using conversion techniques for natural images. It is shown that each binarization method gains the highest performance on images which are converted using the method of Hedjam et al.

Mitianoudis and Papamarkos [MP14a] convert the RGB image to grayscale by applying the PCA transformation and using the first principal component as a grayscale image. Since the majority of the pixels within a document are background pixels, the first principal component is close to the background and the contrast is increased within the corresponding image [TM20]. Mitianoudis and Papamarkos do not evaluate the effect of the PCA transformation, because the conversion is only a preprocessing step within their binarization framework.

Recently, Bouillon et al. [BIL19] proposed another conversion method that is designed for handwritten documents. The conversion is performed in the YPQ color space [GD07] and the pixels in the color space are clustered. Afterwards, the pixel intensities are assigned by using the Mahalanobis distance between the pixel and the center of the background cluster. The authors apply four binarization methods (including [Ots79], [Nib85], [SP00]) on images that are converted using the method proposed and using the luminosity method. It is shown on the H-DIBCO 2016 that three out of four methods perform best on the images that are converted using the algorithm proposed.

2.2.3 Deep Learning

To the best of my knowledge, the first binarization approach, which makes use of neural networks was proposed by Hamza et al. in [HBS05]. The authors use a Self Organizing Map (SOM) to cluster pixels based on their (RGB) intensity values and the labeled output is then used to train a Multi Layer Perceptron (MLP). The MLP is then used to finally classify pixels as belonging to the foreground or background. Three output images are shown, which exhibit that the method is superior compared to the thresholding techniques proposed by Otsu [Ots79], Niblack [Nib85] and Sauvola and Pietikäinen [SP00].

Sari et al. propose another MLP based approach in [SKB12]. For each pixel, a one dimensional feature vector with nine elements is constructed that contains the gray value intensities within a local 3×3 neighborhood. These feature vectors are then used for the training of an MLP.

The two approaches mentioned are shallow neural networks. The first deep learningbased approaches are proposed by Afzal et al. [APS+15] and Pastor-Pellicer et al. [PBZ+15]. Afzal et al. [APS+15] suggest the use of an Long Short-Term Memory (LSTM) network, which belongs to the category of Recurrent Neural Network (RNN). Since the standard LSTM network architecture is designed for one dimensional feature vectors, Afzal et

al. propose to learn the spatial information with a 2D Bidirectional Long Short-Term Memory (BLSTM). The architecture used has two recurrent connections with two forget gates. The numerical results show that the method outperforms the method in [SP00] in terms of FM.

Similar to Afzal et al., Westphal et al. [WLG18] recently proposed to use another RNN architecture, namely the Grid LSTM architecture that is proposed by Kalchbrenner et al. [KDG16]. Contrary to the standard LSTM networks, Grid LSTM networks allow for a multidimensional processing. In [WLG18] a Grid LSTM is used for the first time for document image binarization. The loss function used is based on the Pseudo F-Measure (p-FM) metric. The authors stress out that it is necessary to account for the class imbalance between the foreground and background by weighting both classes. However, instead of simply adding a static weight to the loss function, the weights are calculated based on the measurements that are used for computation of the p-FM. Westphal et al. stress out that the loss focuses on the readability of the binarized document, similar to the p-FM measurement. The method is applied on the H-DIBCO 2016 dataset, where it achieves encouraging results.

Pastor-Pellicer et al. [PBZ+15] propose the use of a CNN for document image binarization. The architecture of the CNN is illustrated in Figure 2.8. It can be seen that the CNN consists of two convolutional layers and a MLP with two hidden layers and an output neuron. The method is applied on the DIBCO 2013 dataset where it achieves competitive results. However, the FM measure gained is 5% smaller than the performance gained by the DIBCO 2013 winning method of Su et al. [SLT13].

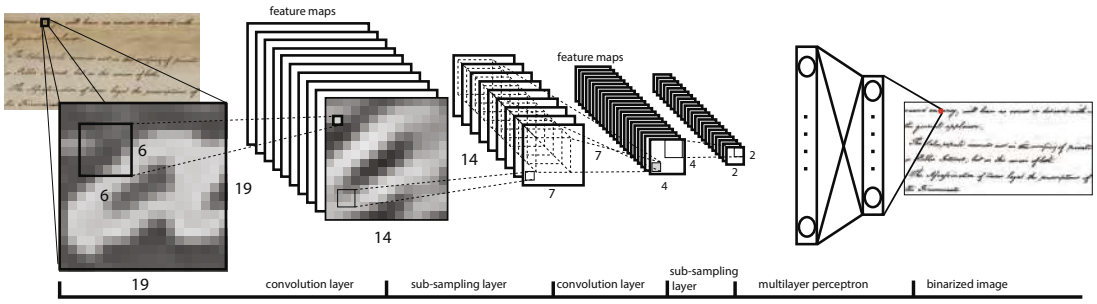


Figure 2.8: CNN architecture proposed Pastor-Pellicer et al. [PBZ+15]. Image taken from [PBZ+15].

While the method mentioned uses a CNN, Tensmeyer and Martinez [TM17] suggest a FCN for image binarization. The architecture has an U-shape and is illustrated in Figure 2.9. The cost function used is an adaption of the p-FM [NGP13] to compensate for the imbalance between foreground and background classes. The input fed into the FCN are unprocessed input images as well as locally computed Darkness features [WNRA16]. By using this features the performance on the H-DIBCO dataset is improved by approximately 0.5%, leading to an overall performance of 97.15% in terms of p-FM. A

similar method by Tensmeyer was submitted to DIBCO 2017, where it ranked in fourth place.

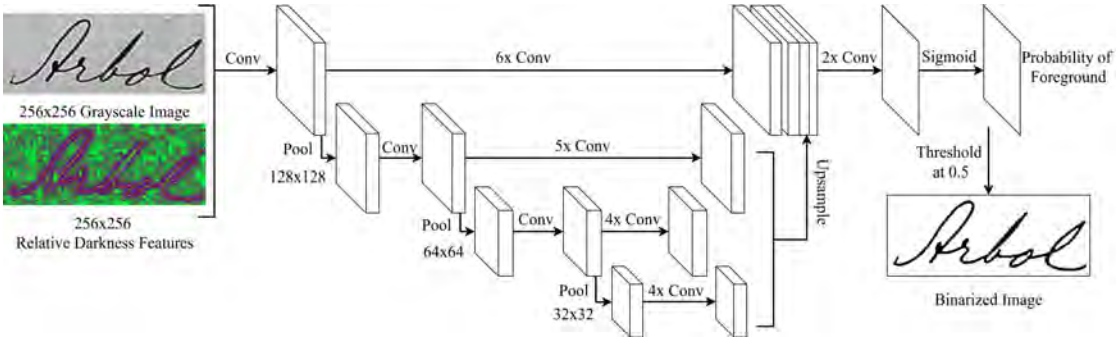


Figure 2.9: CNN architecture proposed by Tensmeyer and Martinez [TM17]. Image taken from [TM17].

The winning method of the DIBCO 2017 contest was submitted by Bezmaternykh et al. [BIN19] for Smart Engines Ltd.. The method makes use of the U-Net architecture that is proposed by Ronneberger et al. [RFB15] for biomedical image segmentation. The training data is augmented in order to increase its size. This strategy is also used in the original U-Net paper by Ronneberger et al. [RFB15], where the training images are transformed by rotation, shifting and elastic transformations.

Fifteen groups participated in the DIBCO 2019 contest and four of the submitted methods made use of the U-Net architecture. Interestingly, none of these approaches won the competition, but instead a clustering based approach that is proposed by Bera et al. [BGB+21]. According to the competition organizers, this can be attributed to fact that a subset of the dataset has no similarities with previous DIBCO datasets: For example, the dataset contains papyri fragments for the first time. The supervised approaches have been trained on the older datasets and were not capable of achieving the expected high performance [PZK+19b].

Peng et al. propose a convolutional encoder-decoder network for document image binarization in [PCN17]. The authors note that very deep networks suffer from the problem of vanishing/exploding gradients [GB10], [BSF94], which aggravates convergence [HZRS16]. In order to overcome this drawback, Peng et al. use the residual network architecture introduced in [HZRS16]. In the encoding path, max pooling is performed for the image down sampling and in the decoding part max unpooling layers are applied for upsampling. The output of the FCN is a confidence value for each pixel, which is obtained by a final softmax layer. This confidence map is then post-processed with a fully connected Conditional Random Field (CRF) in order to remove noise and to refine the text boundary. The method is applied on the H-DIBCO 2016 dataset, where it achieves competitive results. The results are superior compared to the FCN based network used by Tensmeyer and Martinez [TM17]. According to the Peng et al., this reveals that their deep convolutional encoder-decoder architecture is superior to other CNN based

binarization methods. However, it should be noted that the H-DIBCO 2016 dataset was available to Peng et al., whereas it was not available to Tensmeyer and Martinez.

Similar to the method of Peng et al. [PCN17], Calvo-Zaragoza and Gallego [CG19] propose an auto-encoder method for document image binarization. The authors note that traditional auto-encoders are used to learn the identity function and to reconstruct the input images. Contrary, the authors suggest to learn a selectional map, which encodes the class probability. The model is named Selectional Auto-Encoder and the loss function used aims at maximizing the FM and has been proposed by Pastor-Pellicer et al. [PZBB13]. The method is evaluated on multiple datasets compiled by the authors, where it achieves a performance that is superior to the results gained by Su et al. [SLT13], Howe [How12] and Kligler et al. [KKT18].

While the deep learning-based methods mentioned are directly applied on the input image, He and Schomaker [HS19] propose to learn degradations contained in document images and to produce uniform images, which do not contain degradations. Therefore, a U-Net is trained in order to enhance a degraded document image and thus to reconstruct an undamaged version of the input image. The ground truth images that are used in the training phase are synthetically generated, whereby the foreground and background pixel intensity values are replaced by the average foreground and background intensities. Thus, uniform images of the input images are constructed and the neural network learns how to reconstruct an undamaged version of the input image. The reconstructed image is then binarized by applying a global Otsu [Ots79] threshold. The method participated in the DIBCO 2019 competition, where it achieved the ninth rank.

Peng et al. propose another U-Net based binarization approach in [PWC19]. The authors suggest the use of convolutional attention layers and to down-sample the input image to three different sizes. The resampled images are fed into three U-shaped networks. The resulting three output feature maps are concatenated and classified by final softmax layer. The output is further refined by applying a convolutional CRF [TC19]. The loss function used is based on the Distance Reciprocal Distortion (DRD) metric in order to obtain a better visual perception quality.

Akbari et al. [ABAmO19] note that multispectral remote images have been successfully used in semantic segmentation based approaches like in [KSK18]. Motivated by this observation the authors propose to add a further channel to the grayscale input image: The additional channel used is the binarization output of the SSP based method proposed by Jia et al. [JSH+18]. Thus, the resulting image consists of two channels and is used as input for a CNN. Akbari et al. evaluate the performances of two different CNN architectures, namely the U-Net [RFB15] and the SegNet [BKC17] architecture, which is an encoder-decoder network proposed by Badrinarayanan et al. [BKC17]. The numerical results show that the binarization performance of both CNN architectures is improved by using the additional channel and that SegNet approach is slightly outperformed by the U-Net approach. In the case of the H-DIBCO 2016 dataset, the highest performance increase is measured, which is approximately 4% for both architectures in terms of FM.

Tensmeyer et al. [TBSM19] stress out that the application of deep learning-based methods for binarization is limited by the small amount of labeled data. Therefore, the authors propose to generate synthetic data with a Generative Adversarial Network (GAN) and extend the CycleGAN method proposed in [ZPIE17]. The synthetically generated data is used for pretraining a FCN based model as described in [TM17] by Tensmeyer and Martinez. The pretrained model is then finetuned on the DIBCO datasets. The method is applied on the DIBCO datasets and outperforms in 6 out of 9 cases the winning methods - in terms of FM.

2.2.4 Evaluation

This section contains an overview on evaluation metrics for document image binarization. Binarization metrics can be grouped into three different categories according to Ntirogiannis et al. [NGP13]. The first category is an OCR based evaluation, which is an end-to-end metric, since binarization is used as a preprocessing step for text recognition [Ten19]. This evaluation category is briefly explained in the following. Afterwards, the second category is detailed, namely pixel-based evaluation. This evaluation category is used by the majority of the binarization methods and it is based on the comparison between binarization outputs and the corresponding reference ground truth images. Pixel-based metrics are only briefly introduced in the following. A more thorough description of pixel-based metrics is given in evaluation part in Section 4.4, since these metrics are used for the evaluation of the binarization methods proposed.

The third category is evaluation by visual inspection [NGP13]. For instance, in [TJ95] binarized digits are manually categorized as correctly binarized, missed, broken etc. The evaluation based on visual inspection has been mainly used before the introduction of public datasets containing reference ground truth images, such as the DIBCO datasets. Such a visual inspection is subject to a subjective decision and hence cannot be fulfilled with a sufficient precision [NGP13]. Due to this shortcoming, this evaluation category is not further discussed in this thesis.

OCR-Based Evaluation

OCR based-evaluation is an indirect evaluation measure: A binarization result is used as an input for an OCR system and the OCR outcome is evaluated by means of text recognition metrics, such as character accuracy or word accuracy [NGP13]. In end-to-end OCR systems, binarization is used as a pre-processing step [SLT10] and OCR based evaluation can be used to measure the influence of binarization on end-to-end OCR systems.

The work of Trier and Jain [TJ95] is the first work, which provides a formal evaluation framework that is based on OCR. In this goal directed evaluation, eleven local threshold methods are applied on digit images. The resulting images are used as input for a commercial OCR system and the classification results are used to obtain performance metrics, including recognition rate and error rate. Gatos et al. [GPP06] also apply a

commercial OCR system (namely ABBY FineReader) on the binarization results of five different methods. The authors measure the OCR performance with the Levenshtein distance [Lev66], which measures how many single-character edits are necessary to transform one string into another. Similarly, Wolf et al. [WJC02] apply ABBY FineReader on binarized images and provide *Recall* and *Precision* values on a character level. The same OCR system is used by Gupta et al. [GJG07], where multiple binarization methods are compared and their performance is measured by a recognition rate that is an improved metric of the measure used by O'Gorman [O'G94].

It is noteworthy that the works mentioned make use of different metrics and datasets and that there is also no public dataset for OCR based evaluation available. This circumstance impedes a reproducible and comparable OCR based evaluation. Additionally, the methods mentioned are applied on modern and printed text and not on more challenging images of historical handwritings. Another aggravation is the fact that OCR based results are biased by the OCR engine used [Ten19]. Due to these shortcomings, OCR based evaluation was mainly used before the introduction of the first DIBCO dataset and more recent binarization methods are evaluated with pixel-based metrics.

Pixel-Based Evaluation

Several metrics have been proposed for the performance measure on a pixel level. These methods make use of a binary ground truth image, in which the foreground and background pixels are encoded using different intensity values. Such ground truth images are created manually, semi-automatically [NGP08] or automatically [TBSM19].

The frequently used datasets are the DIBCO and H-DIBCO datasets. The first DIBCO competition took place in 2009 and was organized by Gatos et al. [GNP09]. Since then, at each ICDAR a DIBCO competition took place, except for 2015. The DIBCO datasets contain handwritten and machine printed texts. Contrary, the H-DIBCO datasets contain solely images of handwritten text. The first H-DIBCO competition was organized by Pratikakis et al. [PGN10] at the International Conference on Frontiers in Handwriting Recognition (ICFHR) 2010. Figure 2.10 shows four exemplar input and ground truth image pairs taken from DIBCO and H-DIBCO datasets.

Table 2.1 shows which metrics have been used in the DIBCO and H-DIBCO competitions. Additionally, the numbers of test images and the winning methods are provided.

It can be seen that in recent competitions, starting with the DIBCO 2013 competition, the following metrics are used: FM, p-FM, Peak Signal-to-Noise Ratio (PSNR) and DRD. These metrics are used for the evaluation in this thesis and they are explained in Section 4.4. The remaining metrics have not been used in recent competitions. These metrics are: Negative Rate Metric (NRM), Misclassification Penalty Metric (MPM) and Pseudo F-Measure* (p-FM*)[1]. The interested reader is referred to [NGP08], [GNP09] and [GNP11] for a description of these out-dated metrics.

[1]The asterisk is used here to indicate, that the p-FM* metric [NGP08] is different to the p-FM metric [NGP13], although both metrics are named similar

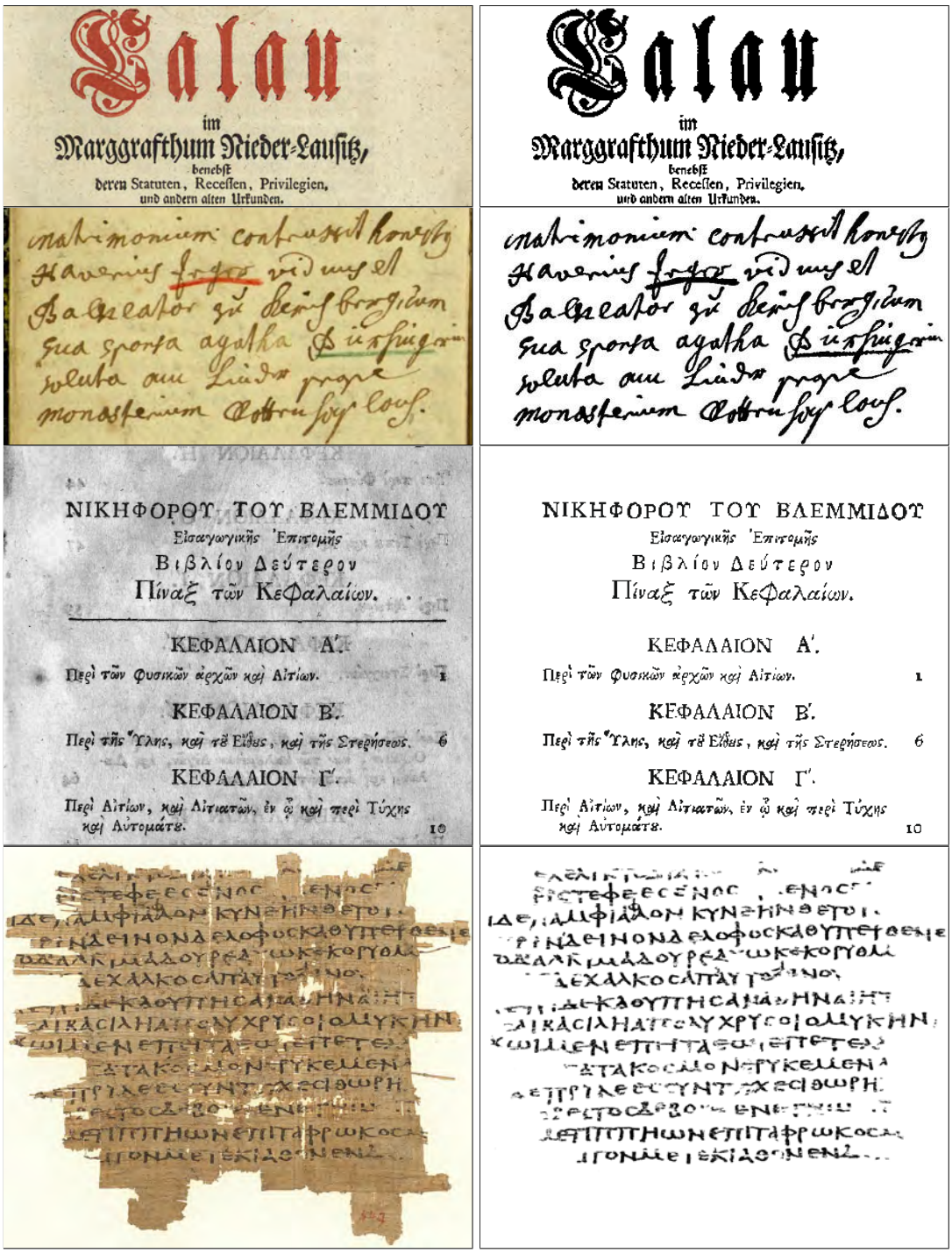


Figure 2.10: Input images and corresponding ground truth images. From top to bottom row: DIBCO 2009, H-DIBCO 2018, DIBCO 2019, DIBCO 2019 papyri dataset.

Competition	Metrics	Images	Winner
DIBCO 2009	FM, PSNR, NRM, MPM	10	Lu and Tan [LST10]
H-DIBCO 2010	FM, p-FM*, PSNR, NRM	10	Su et al. [SLT13]
DIBCO 2011	FM, PSNR, DRD, MPM	16	Lelore and Bouchara [LB13]
H-DIBCO 2012	FM, p-FM*, PSNR, DRD	16	Howe [How12]
DIBCO 2013	FM, p-FM, PSNR, DRD	16	Su et al.[2]
H-DIBCO 2014	FM, p-FM, PSNR, DRD	10	Mesquita et al. [MMA14]
H-DIBCO 2016	FM, p-FM, PSNR, DRD	10	Kligler and Tal [KKT18]
DIBCO 2017	FM, p-FM, PSNR, DRD	20	Bezmaternykh et al. [BIN19]
H-DIBCO 2018	FM, p-FM, PSNR, DRD	10	Xiong et al. [XJX+18]
DIBCO 2019	FM, p-FM, PSNR, DRD	10/10	Bera et al. [BGB+21]

Table 2.1: Overview on binarization competitions.

2.3 Binarization of MultiSpectral Images

The number of binarization algorithms designed for grayscale images exceeds by far the number of binarization algorithms designed for MSI data. This can be mainly attributed to the fact that MSI or HSI systems are more seldomly used for the acquisition of document images, compared to an acquisition with scanners or regular RGB sensors. Because binarization of MSI data is a niche application, it is possible to give an overview on all related works that have been found in the literature. To the best of my knowledge all related binarization approaches for MSI data are summarized in the following.

2.3.1 MRF

The first approach for multispectral document binarization is proposed by Lettner and Sablatnig [LS09]. The approach makes use of spectral and spatial information by applying a MRF, which allows for a combination of both features. In the following an introduction to MRF modelling is given. MRF modelling is based on Bayesian inference, which can be defined by:

The learning and inference algorithms are based on the Bayes inference, which can defined by:

$$p(x|y) \propto p(y|x)p(x). \tag{2.7}$$

whereby the posterior $p(x|y)$ depends on the likelihood $p(y|x)$ and on the prior $p(x)$.

In image restoration tasks - such as document image binarization - the aim is to find an optimal solution $x*$, which is maximizing the posterior probability $p(x|y)$. This approach is named Maximum A Posteriori (MAP) estimation and is given by:

$$x* = arg\ max\ p(x|y) \tag{2.8}$$

MAP estimation for MRF models is introduced by Geman and Geman [GG84] and is the most often used inference technique in MRF modeling [Li09]. For image segmentation - as defined in [KP06] - the prior describes the labeling process x and is modeled as a discrete random variable. The image process y is modeled as another random variable, which is a function of the labeling process. The overall aim of image segmentation is to find the optimal segmentation output $x*$ with MAP estimation.

The prior is based on the fact that the segmentation output should be locally homogeneous [KP06] and is formally defined by:

$$P(x) = \frac{1}{Z} \prod_{c \in C} f_c(x_{(c)})), \tag{2.9}$$

where Z is a normalizing constant and f_c is the potential function for a clique c. A clique $c \in C$ is a subset of neighboring nodes and the order of a MRF is defined by its maximum clique size.

Lettner and Sablatnig propose to use character stroke properties for the modeling of spatial dependencies of characters. The average stroke width in their dataset is five pixels and hence the authors argue that the prior should describe a neighborhood set of at least 4th order. The potential function used is based on the circumstance that similar classes are favored within cliques.

The estimation of the likelihood term is based on the assumption that $p(y|x)$ follows a normal distribution, which is modeled by a GMM. Thus, a GMM is used for the estimation of the average spectral signature and the corresponding covariance matrix for the foreground and background class. These vectors are found by applying the Expectation-Maximization (EM) algorithm.

Lettner and Sablatnig use the Iterated Conditional Modes (ICM) [Bes86] approach to find the MAP-MRF solution. More details on MRF modeling can be found in [KP06]. The binarization approach is evaluated on multispectral images of three different folios belonging to an ancient document. Each set of multispectral images contains 9 different spectral channels. The binarization performance is measured in terms of precision and recall. The evaluation shows that the results obtained are better compared to results that are obtained by the Sauvola and Pietikäinen [SP00] algorithm and a k-means based binarization approach [LBE04] which are applied on a single channel.

In a follow-up paper [LS10] Lettner and Sablatnig make use of Belief Propagation (BP) instead of ICM for the MAP estimation. This leads to a speedup compared to their

previous work. The algorithm is applied on four folios that are similar to the folios contained in the preceding work [LS09] of Lettner and Sablatnig. Again, the method is evaluated against the algorithm of Sauvola and Pietikäinen [SP00] and additionally against a Graph Cut (GC) based method [KKT09]. The algorithm of Sauvola and Pietikäinen [SP00] is clearly outperformed and the GC based method is partially outperformed. The ground truth data in both papers was manually created with the support of philologists, but was not made publicly available.

2.3.2 Maximum Likelihood

Hedjam and Cheriet [HC11b] propose to binarize multispectral document images by applying a maximum likelihood classifier. The authors stress out that classical binarization approaches are directly applied on grayscale values. Similar, Hedjam and Cheriet use a feature vector which contains the multispectral pixel values. The feature vector is additionally extended with two additional entries in order to improve the binarization performance.

The first feature which is introduced by Hedjam and Cheriet is named *pattern persistence* and encodes the variability of the spectral signature. The *pattern persistence* is denoted by $\mathcal{P}$ and is formally defined by:

$$\mathcal{P} = exp(-\nabla_T) \tag{2.10}$$

$$\nabla_T = \frac{1}{B-1} \sum_{i=1}^{B-1} \left(\gamma_{i+1} - \gamma_i + \frac{\gamma_{i+1} - \gamma_1}{i} \right), \tag{2.11}$$

where B is the number of wavelengths and γ_i denotes the reflectance value of the i th wavelength. The first term within the braces $\gamma_{i+1} - \gamma_i$ is named reflection variation and is the reflectance difference between adjacent wavelengths. The second term $\frac{\gamma_{i+1}-\gamma_1}{i}$ is named far reflectance variation and defines the difference between γ_{i+1} and the first reflectance value γ_1.

The second feature is named *pattern energy* and is related to the local structure of an image pattern which is encoded by a non-linear structure tensor [BW02]. The tensor is based on spatial image derivatives and is used for the extraction of local image features, such as edges or corners. Hedjam and Cheriet propose to calculate the tensors for each channel and use its eigenvalues λ_1 and λ_2 in order to encode the energy ϵ:

$$\epsilon^x = \sqrt{\lambda_1^x + \lambda_2^x}, \tag{2.12}$$

where x is a single channel of the multispectral image. The *pattern energy* ε is the average energy over all wavebands and is defined by:

$$\varepsilon = \frac{1}{B} \sum_{i=1}^{B} \epsilon_i^x \tag{2.13}$$

The two additional features are concatenated with the vector that contains the reflectance values. The multispectral images in the work of [HC11b] contain seven different channels. Hence the overall resulting feature vector contains nine elements. The feature vectors are afterwards classified as foreground or background by applying a ML classifier.

The performance is measured by means of FM. The algorithm is compared against the method in [MC10], which is a multi-scale version of the Otsu algorithm. Unfortunately, the paper contains no details on which channel the method in [MC10] is applied. The numerical results exhibit that the classical grayscale binarization method is clearly outperformed by the multispectral binarization technique. Additionally, the performance improvement of the two features introduced is evaluated: If the ML classifier is solely applied on reflectance values the performance is significantly lower, compared to the performance that is obtained by incorporating the two proposed features. Figure 2.11 shows an exemplar output that is obtained by using solely reflectance values (Figure 2.11 (e)) and by using the extended feature vector (Figure 2.11 (f)), It has to be mentioned that the dataset used is relatively small because it contains just three multispectral images.

2.3.3 Fusion Based

Mitianoudis and Papamarkos propose a fusion based binarization approach in [MP14b]. The method is designed for the multispectral document image set that is introduced by Hedjam and Cheriet in [HC13b]. The framework consists of three different stages:

In the first step, two synthetic images are created, namely an image showing the text and another image which contains solely the document background. The images are generated by applying an image fusion approach, which is proposed in [MS07]. The image fusion operates on a transform domain that is defined by self-trained ICA bases. The image containing the handwriting is generated by applying the fusion algorithm on the first six images of the multispectral images. These six images are acquired at wavelengths between 340 nm and 900 nm and the iron-gall based ink is visible within all these spectra. The background image is created by fusing the last two channels of the multispectral images, which are acquired at 1000 nm and 1100 nm. The iron-gall based ink is reflected in these NIR ranges and hence only the background is visible.

The second stage consists of a background subtraction. Mitianoudis and Papamarkos stress out the separation of the foreground and background can be posed as a Blind Source Separation (BSS) problem. The authors propose to separate the two components by applying an ICA algorithm, namely the FastICA [HO00] algorithm.

In the third stage the final binarization step is performed on the FastICA result image that shows the handwriting. Therefore, a spatial kernel K-harmonic Means clustering

Figure 2.11: Binarization output by Hedjam and Cheriet [HC11b]. (a) Multispectral channel in the visible range. (b) Multispectral channel in the NIR range. (c) *Pattern persistence* image. (d) *Pattern energy* image. (e) ML result using reflectance values. (f) Result.

algorithm [LMS07] is applied on the image. The clustering algorithm is an extended k-means algorithm which exploits spatial information by making use of a MRF model.

The authors evaluate their method on the dataset provided by Hedjam and Cheriet [HC13b]. They do not directly compare their results to the one obtained by Hedjam and Cheriet, because the latter method is concerned with image enhancement and is used as a preprocessing step for binarization. Instead, Mitianoudis and Papamarkos compare their results to results obtained by applying the algorithms in [GPP06] and [MP14a]. These binarization methods are designed for grayscale images and the authors use channels from the visible range as input images for the algorithms. The authors report that the FM score gained by their method is approximately 15% and 10% higher than the scores gained by the methods in [GPP06] and [MP14a]. The increased performance can be attributed to the additional spectral information that is used by the proposed method. In particular the NIR channels provide additional information which can be used to reconstruct solely the document background. Figure 2.12 shows intermediate results and a final binarization result of the method in [MP14a].

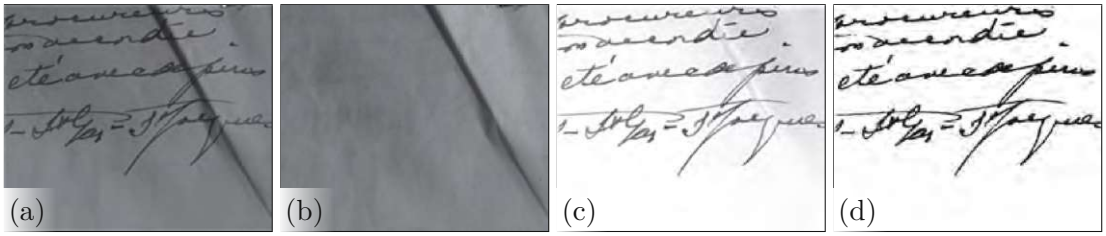


Figure 2.12: Multispectral document binarization proposed by Mitianoudis and Papamarkos [MP14b]. (a) Multispectral channel showing the handwriting. (b) Reconstructed background image. (c) FastICA output. (d) Final Result.

2.3.4 Multiple-Expert Binarization Framework

Moghaddam and Cheriet propose a multiple-expert binarization approach in [MC15]. The framework is trained and evaluated on the MS-TEx train and test set. The authors reduce the dimensionality of the multispectral images by means of three band subspace selection. The selected subsets are then binarized with a state-of-the-art binarization method. The binarized images are afterwards combined into a single image, which is the overall outcome. Moghaddam and Cheriet evaluate the performance of the method suggested by Howe [How12] and of the phase-based method in [NMC14], whereby the latter one performs better.

The method is evaluated on the MS-TEx test set, where it gains the lowest FM score of all methods that are designed for multispectral images. The authors claim that the weak performance can be attributed to the small number of training samples (i.e. 21 multispectral images) and that their method makes no use of a-priori information. Contrary the winning method [DHS16] uses the NIR channel for background subtraction, similar as the method in [MP14b].

2.3.5 Blind Source Separation

Recently, Abderrahmane et al. [AAC19] propose a binarization method that makes use of a BSS technique, namely Non-negative Matrix Factorization (NMF). NMF assumes the Linear Mixing Model (LMM), which is a simplistic model that is used by hyperspectral unmixing methods [MBC+14]. Abderrahmane et al. [AAC19] propose to use this model for the separation of the different sources that are contained in multispectral document images. The BSS technique used is NMF, which aims at decomposing the observed data into two non-negative low rank matrices. The two matrices serve as estimates of the endmembers and abundances. While NMF makes only use of the spectral information, Abderrahmane et al. [AAC19] propose to make additionally use of spatial information: Therefore, a graph regularized NMF [CHHH11] is applied which includes spatial information by building a nearest neighbor graph. The approach is applied on the MS-TEx dataset where two sources are labeled, namely the foreground and background

sources. Hence the authors propose to set the rank of the decomposed matrices to two. The resulting image of the BSS step that exhibits the document background is afterwards binarized with the binarization method that is proposed by Sauvola and Pietikäinen [SP00]. The approach is evaluated on the MS-TEx dataset and it is shown that the binarization performance is increased by incorporating spatial information, compared to the classical NMF approach. Abderrahmane et al. evaluate the method on the MS-TEx test set and show that it outperforms the method of the MS-TEx winner.

In a follow up paper, Salehani et al. [SAR+20] propose another NMF based framework. The framework consists of three stages: First an NMF based feature extraction is applied. Afterwards, the most appropriate row of the coefficient matrix is selected either manually or automatically. This selection corresponds to finding the source, which describes the spectral signature of the foreground. The automatic selection is fulfilled by comparing the corresponding binarization results with binarization results obtained by traditional binarization methods [SP00] [How12]. Finally, the source found is again binarized by applying the method of Howe [How12].

The authors show that the manual source selection leads to the best performance in terms of FM. This performance is superior compared to the results gained by the winning method of the MS-TEx competition, but it is gained in a semi-automated fashion. Contrary, the automatic source selection results in an F-Score that is 1% lower than the FM scored by the winning method.

2.3.6 Ground Truth Estimation

Hedjam and Cheriet [HC13a] propose a semi-automated approach for ground truth generation of multispectral document images. The method is applied on manually created ground truth images in order to refine the ground truth images and to correct mislabeled pixels. The authors propose to automatically select multispectral samples that have a high probability of being correctly labeled. The selected samples are used for the training of multiple classifiers, which are applied in the final binarization step.

Similar to [Smi10], Hedjam and Cheriet state that mislabeling occurs at stroke boundaries and that these regions can be classified differently by different human annotators. Contrary, regions that are located at the middle of character strokes are better distinguishable from the document background and the corresponding labeled regions are less error-prone, compared to stroke boundary regions. Based on this observation, Hedjam and Cheriet suggest to apply morphological operations on the ground truth image in order to select well-labeled samples. The multispectral samples are then used for the training of multiple classifiers, including k-nn. The outputs of the classifiers are then used in a majority voting step in order to assign the final labels. Several examples indicate that the multispectral information can be used to successfully readjust the stroke boundary in ground truth images.

2.3.7 MS-TEx Participants

In 2015 the ICDAR 2015 MS-TEx contest was organized by Hedjam et al. [HNM+15]. Five methods were submitted by four different teams. Two of these approaches are introduced in this work (see Section 4.1). The remaining approaches are unpublished, but they are briefly described in the following, because they are referenced in Section 4.4, where a numerical evaluation on the MS-TEx dataset is given. The summarization of the participating works is based on the description provided in the MS-TEx paper [HNM+15].

Raza The method first performs image fusion by applying a wavelet transform. Afterwards, noise removal filters are applied on the fused image. The resulting image is then binarized using the thresholding technique proposed by Niblack [Nib85] and noise filtering is again performed on the output image.

Zhang and Liu The algorithm applies the Canny [Can86] edge detector on an image acquired at 500 nm. Afterwards, the dark pixels in the local neighborhood are classified as foreground and bright pixels are classified as background. The remaining unlabeled pixels are inferred by Gaussian Random Fields. Noise is afterwards removed by using information that is contained in the NIR channels.

Wu et al. Wu et al. train three different classifiers on the MS-TEx training dataset: The first classifier is used to obtain statistical features, which indicate the likelihood that pixels belong to the foreground. More details on this classifier can be found in [WNRA16]. This classifier is applied on each image channel and the resulting multispectral features are classified by a second classifier. Afterwards, a third classifier is applied in a refinement step in which connected components are rejected.

2.3.8 Evaluation

The methods described above make all use pixel-based evaluation metrics: Lettner and Sablatnig [LS09], [LS10] use Recall, Precision and FM as performance metrics. The dataset was created manually, but is not made public available. Hedjam et al. [HC13b] evaluate their image restoration method indirectly by applying different binarization methods on restoration results. The authors show that the application of their method leads to an increased FM. The dataset in [HC13b] is named HISTODOC1 and is made publicly available[3]. The binarization approach of Mitianoudis and Papamarkos [MP14b] is evaluated on the HISTODOC1 dataset in terms of Recall, Precision and FM. It should be noted that the HISTODOC1 dataset was also published by the publishers of the MS-TEx dataset, whereby the latter replaces the former dataset, because the MS-TEx dataset contains a higher number of multispectral images. Moghaddam and Cheriet [MC15] evaluate their binarization approach in terms of FM on the MS-TEx dataset.

[3] http://www.synchromedia.ca/databases/HISTODOC1

The same database is used in [HNM+15], [AAC19] and [SAR+20]. These works measure the binarization performance in terms of FM, DRD and NRM. Additionally, in [HNM+15] the kappa coefficient is used, which was introduced in [Coh60].

2.4 Summary and Discussion

This chapter summarized MSI and HSI systems and processing techniques that are applied on document images. It was shown, that the acquisition systems are application driven and that their spatial and spectral resolutions are highly varying: The majority of the HSI systems described are used for the forensic analysis of inks. MSI systems offer a higher spatial resolution [Ber19] and are more often used for the enhancement of ancient writings. Recent MSI systems - such as [JCK11], [HBS19] - make use of imaging sensors with a spatial resolution of 50 MP or more, because scholars prefer high-resolution images for an adequate scholar reading [JCK11]. Additionally, it was shown that multiple enhancement methods have been developed for spectral images of historical documents. These methods are partially used for increasing the legibility of ancient texts. It was also explained, there is no direct performance metric existing, which measures the effect of legibility enhancement.

The second part of this section was concerned with the binarization of grayscale document images. The majority of the binarization methods described are evaluated on DIBCO and H-DIBCO datasets. It was shown that deep learning-based methods achieve state-of-the-art performance on these datasets - except for the DIBCO 2019 dataset.

The final part of this section was dedicated to the binarization of multispectral document images. Only a limited number of methods have been developed for this special purpose. The MS-TEx contest took place in 2015 and the published dataset enabled for the first time an objective performance evaluation of multispectral document binarization methods. Two recent works ([AAC19], [SAR+20] propose the use of NMF based source separation. The authors in [AAC19] claim that their method outperforms previous binarization methods. The performance of [AAC19] is listed in the numerical evaluation in Section 4.4, but my personal opinion is that the results are debatable because of the following reason: The presented F-Scores are gained on six different images belonging to the MS-TEx test set. The average of these 6 F-Scores is compared to the average FM of the MS-TEx winning method. However, the latter F-Score is gained on ten test images. Thus, the superior performance claimed by Abderrahmane et al. is doubtable, since it is probably not achieved on the entire test set[4]. It is notable that none of the methods designed for MSI data makes use of deep learning techniques. Eventually, this can be attributed to the circumstance that the MS-TEx dataset is relatively small and deep learning-based methods typically make use of large training datasets [ZPIE17].

[4] Unfortunately, the performance on the entire test set cannot be reproduced, since the source code of the method is not published.

CHAPTER 3

MultiSpectral Document Image Enhancement

The MSI system used in this work is especially designed for the acquisition of historical documents. These documents have been digitized within two projects: The first project is entitled *The Enigma of the Sinaitic Glagolitic Tradition* and was funded by the Austrian Science Fund (FWF) under grant P23133. The project was especially devoted to the investigation of documents written in Glagolitic, which is the oldest Slavonic script [Mik03]. Hence, the majority of the imaged manuscripts are written in Glagolitic. The second project is entitled *CIMA – Centre of Image and Material Analysis in Cultural Heritage* and was funded by the Austrian Federal Ministry of Science, Research and Economy. The origins, languages and scripts of the manuscripts are more varying compared to the first project. The languages of the imaged objects are amongst others: Greek, Slavonic, Latin, German and Aramaic. The manuscripts imaged in both projects originate between the 10th and the 19th century and their conditions are highly varying: While some manuscripts are easily legible in the visible range, others are affected by various degradations involving faded-out or erased characters, background clutter etc. This chapter contains several exemplar images that show the benefit of MSI for the investigation of historical and degraded writings.

Several methods [JCK11] [STB07] have shown that dimension reduction methods can be used to lower the dimension of the multispectral images and additionally to further increase the visibility of vanished writings. These methods make use of unsupervised dimension reduction techniques. Instead, in this chapter a supervised dimension reduction technique is applied, namely LDA. The LDA technique is used in an enhancement method, which is especially designed for strongly degraded writings. The main contribution of this work is to show that document image analysis methods can be used to automatically select and label a training set that is needed for the supervised dimension reduction. The enhancement method is also used in further applications, namely palimpsest enhancement

and as a preprocessing step for an OCR system. For the latter mentioned application, the method was adopted in order to make it applicable on writings, which contain degraded and non-degraded regions.

This chapter is structured as follows. First, the principles of MSI and the imaging modalities used are introduced in Section 3.1. Afterwards, the imaging system used is described in detail in Section 3.1.2. The enhancement method is detailed in Section 3.2. The enhancement of palimpsest texts is described in Section 3.3.1. In Section 3.3.2 the method is used as preprocessing step for an OCR system.

3.1 Image Acquisition

This section provides an explanation of the general principles of MSI. Several imaging modalities are especially useful for the acquisition of historical documents and they are briefly described in the following. The MSI acquisition system that was developed in the course of this work allows to make use of these imaging modalities. The acquisition setup evolved over time and the system as well as its evolvements are explained in this section. Afterwards, exemplary multispectral images are provided in order to demonstrate the capabilities of MSI for the investigation of ancient documents.

3.1.1 Principles

MSI systems acquire images in the visible spectrum as well as in the UV and NIR range. Unfortunately, different definitions fo these spectral ranges exist, particularly for the infrared range [FK06]. In this work the nomenclature of Fischer and Kakoulli [FK06] is used, where the visible spectrum lies between 400 and 700 nm and the NIR range is between 700 and 1000 nm. Below the visible spectrum is the UV range.

MSI systems acquire image data in three dimensions, which results a data cube with X, Y and λ dimensions. The narrow-band spectral ranges can either be acquired by filtering the incident light with narrow-band illumination or by filtering the reflected light with optical filters. According to Berns [Ber19], filters are either *"absorption filters, interference filters, or liquid-crystal tunable filters"*. Narrow-band illumination and optical filters can also be combined to further increase the number of channels. The cameras have monochrome sensors [Ber19], but color sensors have also been used for the imaging of historical documents [JCK11].

An illustration of MSI acquisition is shown in Figure 3.1, whereby the spectral radiance of the lighting source is denoted by $l_R(\lambda)$ and the spectral reflectance of the imaged object is $r(\lambda)$. The reflected light passes the optical system $o(\lambda)$ mounted in front of the camera and eventually a color filter $\varphi_k(\lambda)$ and is then acquired by the image sensor with a certain spectral sensitivity denoted by $a(\lambda)$.

According to [HSB+99] the camera response c at a pixel can be modeled by:

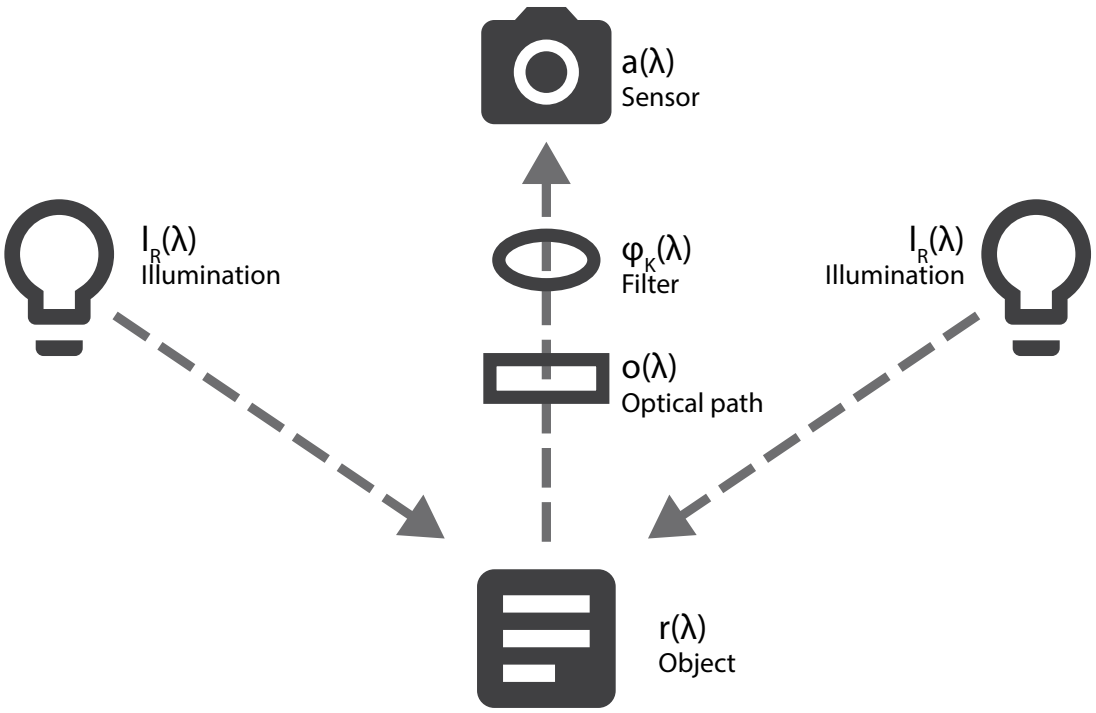


Figure 3.1: Based on an illustration by Hardeberg et al. [HSB+99]

$$c = \int_{\lambda_{min}}^{\lambda_{max}} l_R(\lambda) r(\lambda) o(\lambda) \varphi_k(\lambda) a(\lambda) d\lambda \tag{3.1}$$

The term reflectance describes *"the ratio of the energy reflected or scattered by the surface to the incident energy"* [FK06]. Spectral reflectance is defined as the reflectance per unit wavelength and a spectral reflectance curve is the spectral reflectance over a certain spectral range. Such spectral reflectance curves are used to analyze material properties, including color and composition [FK06]. It should be noted that by using narrow-band LED illumination the spectrum between λ_{min} and λ_{max} in Equation 3.1 is under-sampled. The transformation of these signals to evenly sampled data, i.e. spectral reflectance curve, is an inverse mapping problem [Ber19] and the interested reader is referred to [RS08] for an overview on approximations for this problem.

Imaging Modalities

- **Fluorescence Imaging**
  Luminescent materials are materials, which absorb light and emit it at longer wavelengths. If the emitted light ceases with the excitation, the effect is called

fluorescence[1] [Ber19]. UV florescence imaging is successfully applied on parchment objects, because the UV photons are absorbed by parchment, which emits the light in the visible spectrum [EN10]. While parchment is a luminescent material, iron-gall based ink is not fluorescent and hence the contrast between parchment and iron-gall based ink is increased with UV fluorescence imaging. In order to enable UV fluorescence imaging the light source must contain energy in the appropriate UV range to emit light in the blue region of the visible spectrum [Ber19]. For example in [APA+16] and [JCK11] the peak frequency of the UV lighting sources is 365 nm. Additionally, for UV fluorescence imaging it is necessary to absorb the reflected UV light with a long-pass filter. Thus, only the fluorescent light is acquired by the camera sensor.

- **Reflectography Imaging**
  Contrary, the imaging of the reflected UV light is termed UV reflectography. In order to absorb the fluorescent UV light it is necessary to remove the visible light by applying a short-pass or a band-pass filter. UV reflectography can be used to gather information about object surfaces, because *"surface roughness and stains can be enhanced by the technique" [Stu07].* According to Stuart [Stu07], UV reflectography is less often used for documents and paintings, compared to UV fluorescence.

  Contrary, NIR reflectography and short-wave infrared (1000 - 2500 nm) reflectography have proved to be useful for the investigation of paintings [Stu07], [FK06]. Infrared reflectography offers the advantage, that infrared rays are able to partially penetrate the surface of a painting. Thus, this imaging modality enables the visualization of layers of paintings below the visible surface [Mai00], including for examples underdrawings [FK06]. While NIR reflectography is a common investigation technique for paintings[FK06], it is less commonly used for the enhancement of ancient writings. One exemplar manuscript, which is successfully imaged within the NIR range, is belonging to the Dead Sea Scrolls: Caine and Magen [CM11] report that a particular manuscript is probably degraded by adhered parchment, which covers partially an ancient Aramaic writing. By imaging the object in the NIR range the overlapping parchment was penetrated and the underlying text partially restored.

It is notable that other imaging modalities are also used for the analysis of cultural heritage objects: Easton et al. [EN10] have successfully applied X-ray fluorescence imaging on selected folios of the Archimedes palimpsest which are obscured by forgeries. Underdrawings in paintings can be visualized by imaging within the short-wave infrared range [FK06]. In [SCM+18] 3-D X-ray micro-CT is used to image ancient Chinese bamboo scrolls. The method is used to enhance the legibility of the writings without any manual conservations steps (i.e. unwrapping or cleaning of the scrolls).

[1] The opposite effect is called phosphorescence, which occurs if there is still light emitted after the excitation ceases.

3.1.2 Acquisition Setup

The MSI acquisition system used at the Computer Vision Lab (CVL) evolved over time: Kleber et al. [KLD+08] introduce an MSI system, which makes use of broadband illumination in combination with optical filters. This thesis contains multiple multispectral images that have been acquired by the system developed by Kleber et al. [KLD+08]. Hence, the original system is explained in the following and afterwards the modifications of the CVL MSI system are described.

The multispectral camera used in [KLD+08] is a Hamamatsu C9300-124. This camera is a grayscale NIR camera with a spectral response between 300 nm and 1000 nm and a spatial resolution of 4000×2672 px. Two different lighting sources are used, namely an UV lamp and a broadband tungsten illumination. The spectral ranges are provided by optical filters that are obstructed in a filter wheel, which is mounted in front of the camera. The filters used are: Four band pass filters with peaks at 450 nm, 550 nm, 650 nm and 780 nm. Two long pass filters with a cut-off frequency of 400 nm and 800 nm. The former mentioned filter is used for UV fluorescence and the latter mentioned filter is used for NIR reflectography. UV reflectography images are provided by a short pass filter with a cut-off frequency of 400 nm.

In addition to the grayscale multispectral camera, a Single-Lens Reflex (SLR) camera is used for taking RGB images. The SLR camera is a Nikon D2Xs with a spatial resolution of 4288×2848 px. This camera is used for taking white light images and UV fluorescence images. The main purpose of the Nikon D2Xs camera is to acquire images that can be used for visualization and facsimile editions.

The SLR camera was replaced by another SLR camera, namely a Nikon D4, which provides a higher resolution of 4928×3280 px. The broadband illumination was also replaced by a LED illumination [HGS12]: Two Eureka Light^{TM}LED panels constructed by Equipoise Imaging LLC are used to obtain 11 different spectral ranges. The spectral ranges have peaks at 365 nm, 450 nm, 465 nm, 505 nm, 535 nm, 570 nm, 625 nm, 700 nm, 780 nm, 870 nm and 940 nm. Additionally, four supplemental white light LED panels are used for white light images. The spectra of the multispectral LED panels are given in Figure 3.2.

The usage of narrow-band LED illumination has two advantages compared to the previously used broadband illumination. First, the incident light is already filtered and hence it is not necessary to filter the reflected light with optical filters. The usage of optical filters leads to geometric distortions [BSA08] and hence resulting images have to be registered with an image registration algorithm. Such an image registration step is not necessary if optical filters are omitted. In our MSI introduced in [HGS12] two filters are used for UV reflectography and UV fluorescence. The optical filters used are a short pass filter and a long pass filter, both with a cut-off frequency of 400 nm (similar as in [KLD+08]). The second advantage of LED illumination is the circumstance that the thermal stress is reduced.

The entire setup is illustrated in Figure 3.3. It can be seen that two diffusers are mounted

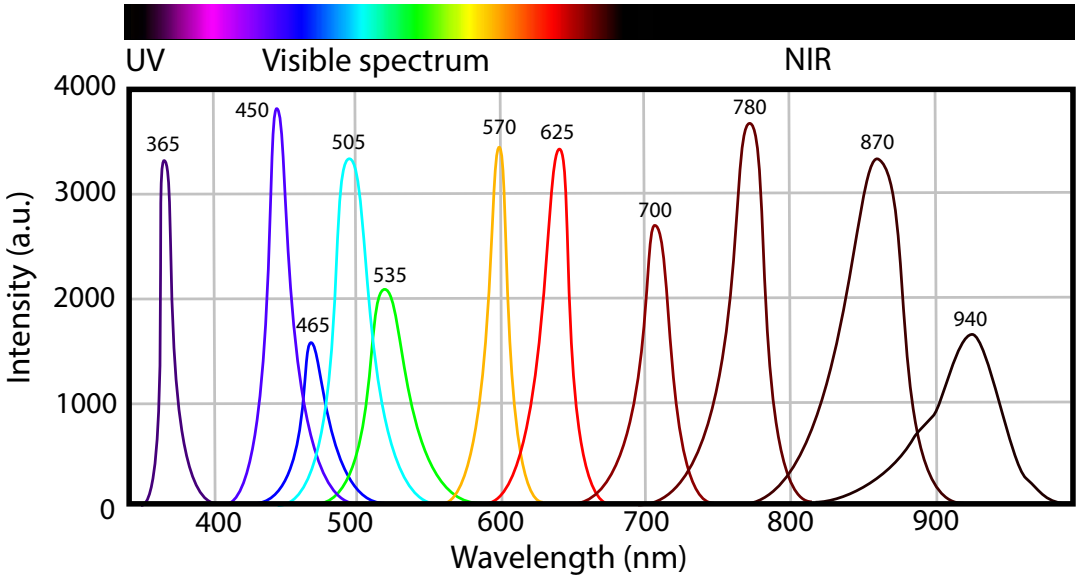


Figure 3.2: Spectra of the multispectral LED panels.

in front of the LED panels. These diffusers are used in order to soften the light and to reduce the effect of uneven illumination caused by the LED illumination. The historical documents are placed on a board that is mounted onto a linear unit. The linear unit allows for a movement between the two cameras. This is especially useful for historical books, because solely the book pages have to be flipped and a manual positioning after turning a page is avoided.

Image Registration Since two optical filters are used for the acquisition of the UV reflectography and fluorescence images, these images have to be registered on each other in order to resolve the geometric distortions [BSA08]. These images have been registered on the remaining images with image registration methods [DLS07], [HJB+12]. Additionally, the images captured with the SLR camera are also registered on the multispectral images by using the method introduced in [DLS07].

3.1.3 Exemplar Documents

This section contains exemplar images of historical documents in order to gain insights on the capabilities of MSI. First, a portion of a palimpsest manuscript is given in Figure 3.4. The overwriting is a Cyrillic text and the ancient underwritten text is written in Greek. The underwritten text is written in a horizontal direction and the overwriting is oriented vertically. The older text vanishes faster with increasing wavelengths, whereas the overwriting is partially visible within NIR range. The red characters are best visible in the RGB white light images and they are partially vanishing under red and NIR light,

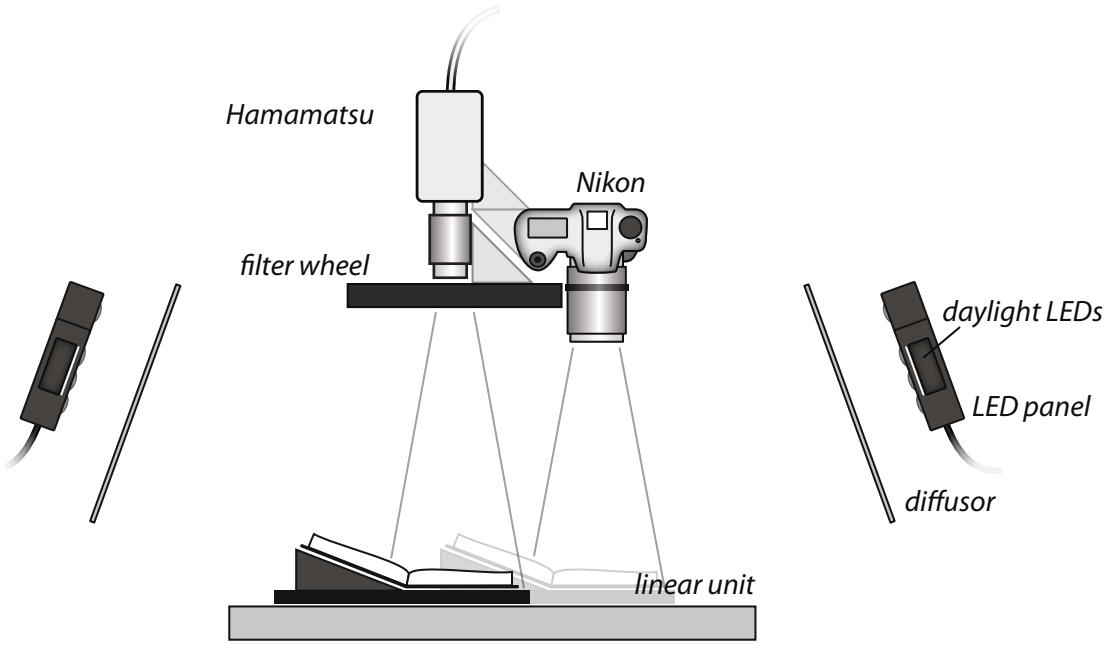


Figure 3.3: Illustration of the MSI system.

because this light is partially reflected by the red ink.

An example for Glagolitic writings is given in Figure 3.5. The faded-out writing is best visible in the UV fluorescence image. A Greek palimpsest is shown in Figure 3.6. The over- and underwriting are oriented in the same direction. The ancient writing is again best visible in the UV fluorescence image. The examples shown in Figure 3.4, Figure 3.5 and Figure 3.6 contain ancient writings that have been created between the 11th and 13th century. Contrary, the palimpsest text that is shown in Figure 3.7 is considerably younger: The palimpsest text is a forgery that was created by Constantine Simonides, who lived in the 19th century. Simonides was a paleographer and produced artificially generated palimpsests, by overwriting original handwritings with a brighter ink that look visually plausible as underwritings. This can be seen in Figure 3.7: The forged underwriting is partially visible under white light and is best visible under UV light - similar to the authentic writings shown above. While it is clear that the underwriting is a forgery, a multispectral analysis conducted by us in [HS17] was not capable of determining if the overwriting is also a forgery created by Simonides. The examples shown are all written on parchment and are best visible under UV light. Contrary, Figure 3.8 shows a leather fragment. The ancient writing is only partially visible under white light, but it is best visible within the NIR range.

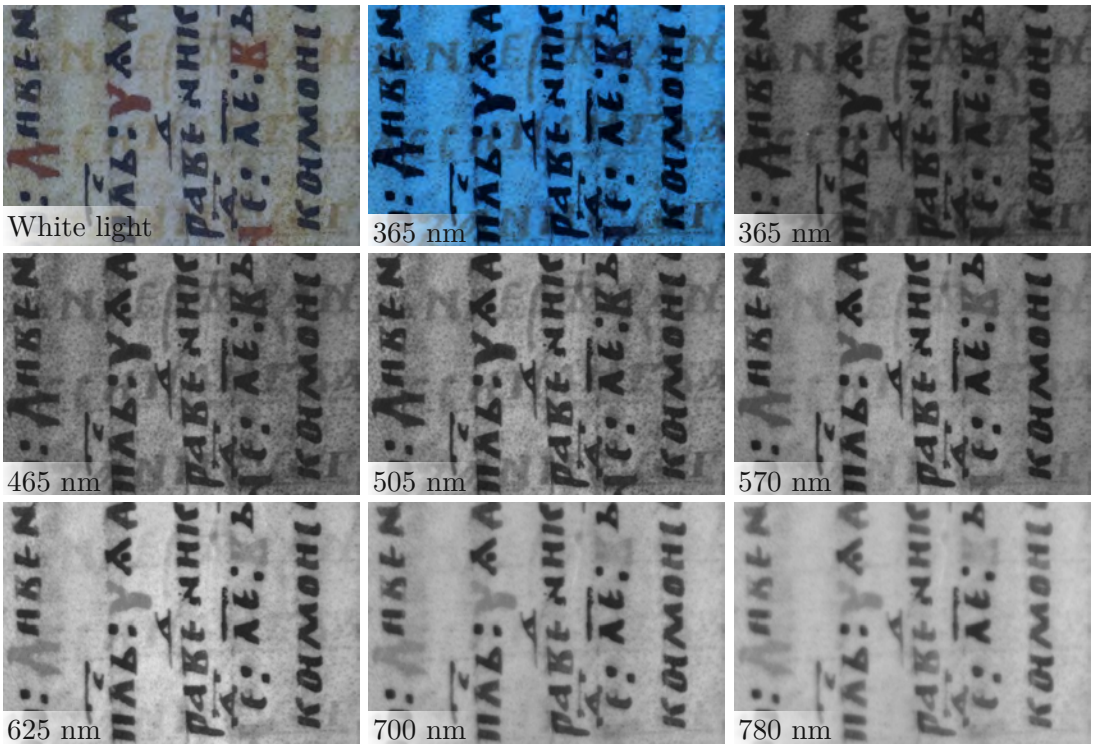


Figure 3.4: Manuscript with shelf mark NBKM880, illuminated with varying wavelengths. The first two images show images captured by the Nikon camera, whereas the remaining photographs are taken by the Hamamatsu camera. The images illuminated with UV light (365 nm) are UV fluorescence images.

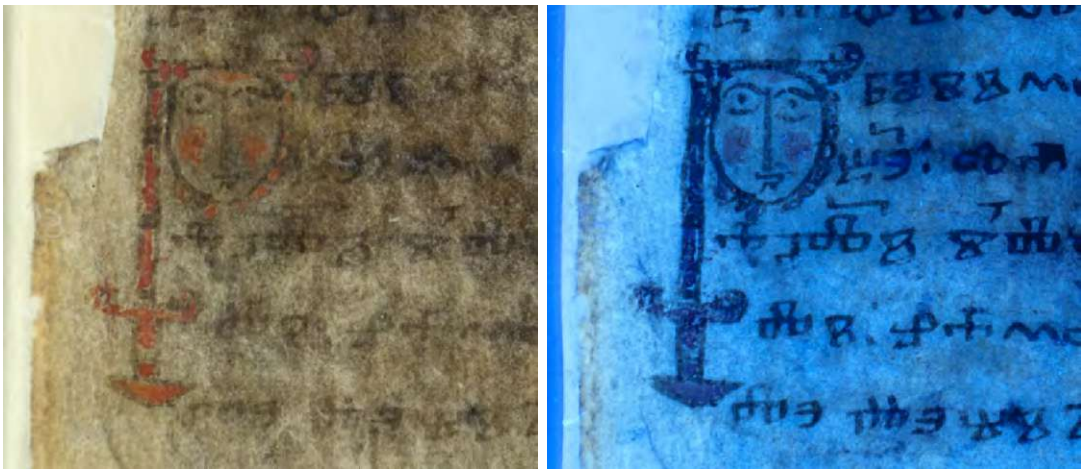


Figure 3.5: Cod. Slav. 136 - 1 recto. (Left) Image taken under white light. (Right) UV fluorescence image.

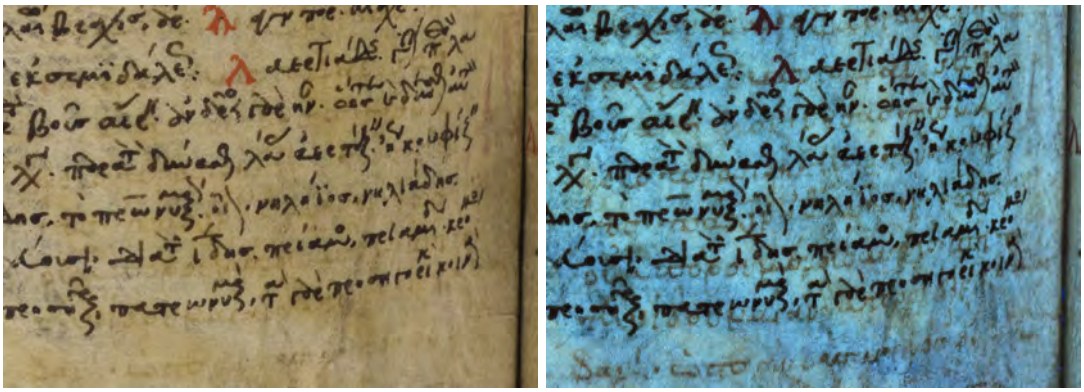


Figure 3.6: Phil. Gr. 158 - 141 verso. (Left) Image taken under white light. (Right) UV fluorescence image.

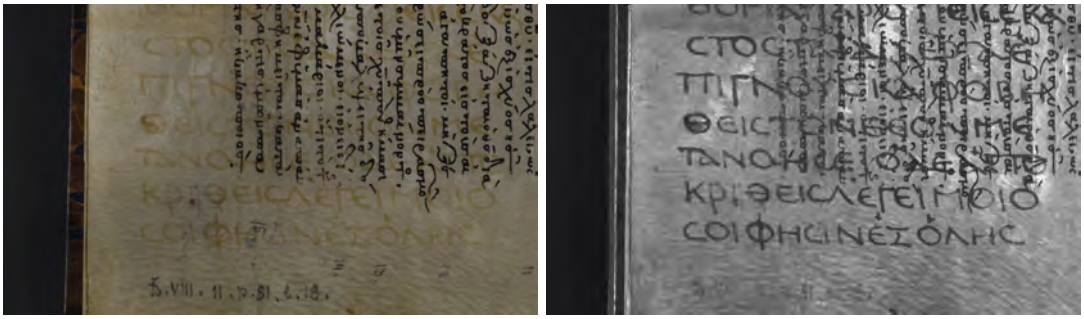


Figure 3.7: Suppl. Gr. 119 - 1 recto. (Left) Image taken under white light. (Right) UV fluorescence image

Figure 3.8: Fragment with the id SR22 A560, found in Egypt. (Left) Image taken under white light. (Right) NIR reflectography image (870 nm).

3.2 Enhancement

The exemplar images in the previous section reveal that MSI can be used for the enhancement of degraded writings. The images shown are unprocessed multispectral images[2]. Dimension reduction techniques - including PCA [JCK11] and ICA [STB07] - have been successfully applied on multispectral images in order to further enhance the visibility of degraded writings. While these two dimension reduction methods are unsupervised techniques, we have applied the supervised LDA technique in [HGS13]. Since LDA is supervised method it is necessary to label a subset of the multispectral samples as belonging to the foreground or background. For the labeling of the sample points a semi-automated method was proposed that makes use of document image analysis methods, namely text line detection and document image binarization. Thus, the enhancement method is not only based on spectral information, but also on spatial information.

The method is especially designed for degraded writings that are partially barely visible within the multispectral images. Therefore, the labeling is not directly applied on multispectral images, but instead on PCA images, since the text is partially better visible within these images. The PCA and LDA methods are explained in the following along with the enhancement method proposed.

3.2.1 Dimension Reduction

The overall aim of the dimension reduction methods applied is to find a transformation of the multispectral data that removes the correlation and enhances the degraded writing. This transformation is formally defined by:

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \tag{3.2}$$

whereby $\mathbf{y}$ denotes the transformed data, $\mathbf{x}$ is the transformation matrix and $\mathbf{x}$ is the unprocessed multispectral data.

Principal Component Analysis For the PCA transformation, the columns of $\mathbf{W}$ are filled with eigenvectors of the covariance matrix of the zero mean normalized data. The eigenvectors are sorted in a descending order based on the corresponding eigenvalues. The dimensionality of the multispectral data is reduced by using only a limited number of the first eigenvectors.

Linear Discriminant Analysis PCA aims at maximizing the scatter of the transformed data $\mathbf{y}$. Contrary, the LDA transformation is a projection $\mathbf{W}$ which maximizes the ratio between the between-class scatter and the within-class scatter. The within-class scatter is formally defined by:

[2]The only image processing techniques that were applied on the images are the conversion from raw image formats and image registration.

$$\mathbf{S}_W = \sum_{i=1}^{c} \sum_{\mathbf{x_k} \in \mathbf{X_i}} (\mathbf{x}_k - \mathbf{m}_i)(\mathbf{x}_k - \mathbf{m}_i)' \tag{3.3}$$

where c is the number of classes, $\mathbf{x}_k$ is a sample belonging to a class X_i and $\mathbf{m}_i$ is the mean of the class. The between-class scatter is defined by

$$\mathbf{S}_B = \sum_{i=1}^{c} N_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})' \tag{3.4}$$

where N_i depicts the number of samples in the class X_i. The projection $\mathbf{W}$ is found by maximizing the criterion function:

$$\mathbf{J}(\mathbf{W}) = \frac{\mathbf{W}'\mathbf{S_B}\mathbf{W}}{\mathbf{W}'\mathbf{S_W}\mathbf{W}}. \tag{3.5}$$

For the solution of the criterion function $\mathbf{J}(\cdot)$ the function can be converted to a generalized eigenvalue problem. The interested reader is referred to [DHS12] for solutions of the two class problem and the general c class problem.

Duda and Hart [DHS12] note that the PCA transformation finds components that are useful for the representation of the data Contrary, LDA is searching for components that are convenient for discrimination. Belhumeur et al. [BHK97] demonstrate the advantage of using a class specific projection with a two class problem: The problem is illustrated in Figure 3.9. The samples of both classes are randomly perturbed in a direction that is orthogonal to a linear subspace. LDA and PCA have been applied on the samples in order to project the two-dimensional samples on an one dimensional space, which are illustrated by dashed lines. The resulting projections exhibit that the LDA projection gains a higher between-class scatter $\mathbf{S}_B$, whereas PCA smears both classes together [BHK97]. Thus, projected LDA samples can be successfully classified, whereas the PCA projected points cannot be correctly classified by using a linear decision boundary.

3.2.2 Text Enhancement

In order to obtain the LDA projection, which discriminates between foreground and background, it is necessary to extract and label a subset of the multispectral data. Document image binarization algorithms are designed for such labeling problems, and are applied on a single channel of the multispectral data. However, the method is especially designed for degraded documents and applying a document binarization method is error-prone because of two reasons: (a) The contrast between foreground and background is partially considerably low, which impedes a correct binarization. (b) The writings are partially barely visible and it is not predefined in which spectral range the writing is best visible. While iron gall based ink is usually best visible within the UV fluorescence range [JCK11], this assumption is not entirely correct for the writings considered: The method

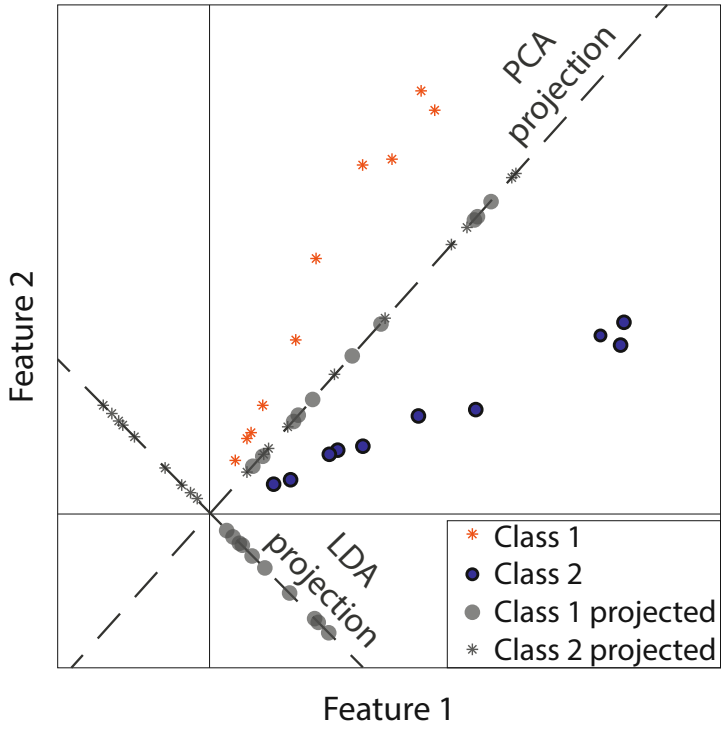


Figure 3.9: LDA and PCA applied on a classification problem. Based on an illustration created by [BHK97].

is mainly designed the Glagolitic writings contained in a manuscript that is named 'Missale Sinaiticum'. The book was created between the 10th and 11th century and its legibility is partially limited: The ink in this book has partially discolored from brown or black to white [MGK+08] and the characters are best visible in varying wavelengths.

The 'Missale Sinaiticum' manuscript images have been acquired by Lettner et al. [LKSM08] with the MSI system that is described in Section 2.1.1. Two exemplar multispectral images belonging to the 'Missale Sinaiticum' manuscript are shown in Figure 3.10. The image portions contain writings, where the ink discolored to white. It is notable that the writings are best visible in the UV reflectography images, but they are only partially visible in these images.

The contrast of the degraded writings is enhanced by applying the PCA transformation. Figure 3.11 shows PCA images that have been generated by applying the PCA transformation on the MSI data shown in Figure 3.10. It can be seen that the text is not visible within the first PCA images. It was found that the text is only emphasized by the first principal component if the text is visible within multiple spectral ranges, which is not the case for the white ink contained in the 'Missale Sinaiticum'. In the remaining PCA images the text is partially better visible compared to the unprocessed

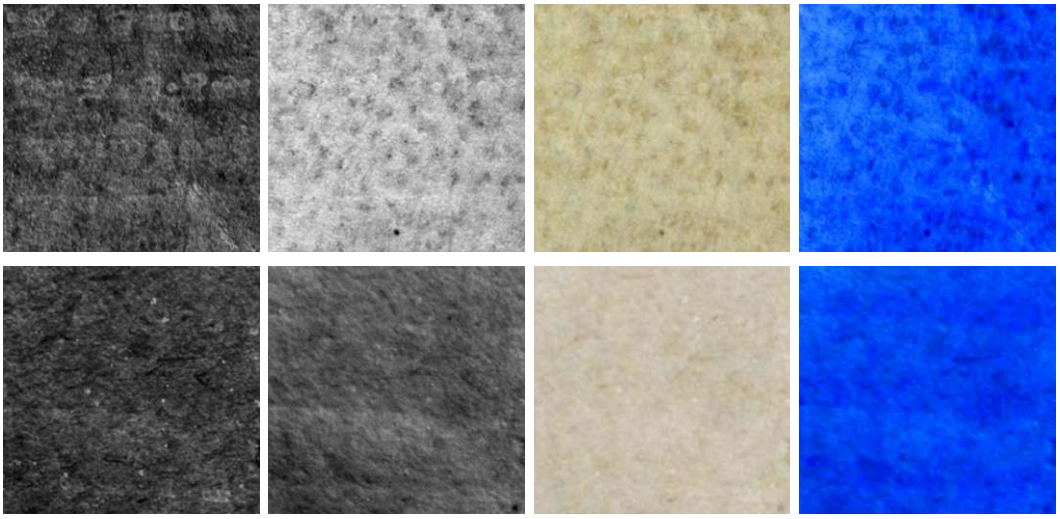


Figure 3.10: Two multispectral samples belonging to the 'Missale Sinaiticum'. From left to right: UV reflectography, UV fluorescence, white light RGB image and UV fluorescence RGB image.

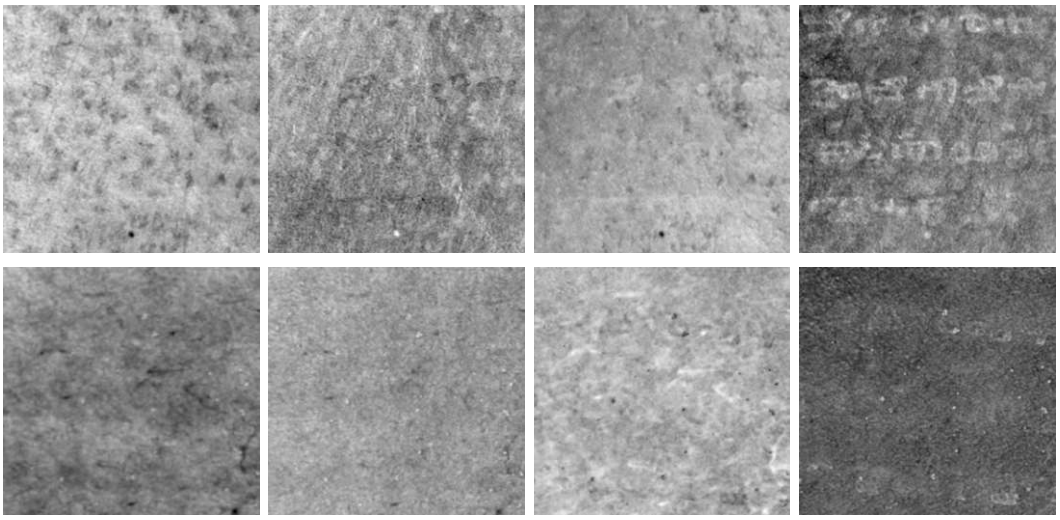


Figure 3.11: PCA images. Images obtained using the first four principal components.

multispectral images. Additionally, it is notable that the text is enhanced by multiple principal components.

Since the text is better recognizable in the PCA images, the labeling procedure is performed on PCA images. This labeling procedure is described in the following section.

Coarse Enhancement

A degraded writing is often emphasized by multiple principal components [LKSM08] [JCK11]. Therefore, the first k PCA images are considered in the coarse enhancement stage. In the case of the 'Missale Sinaiticum', k was set to five, since it was experimentally found that the degraded text is partially emphasized by the first five principal components. The method is also evaluated on a second set of manuscript images, which is named 'Glagolitic Fragments'. The spectral signatures of the writings in 'Glagolitic Fragments' are less varying and hence k was set to three for this manuscript. In Figure 3.11 and 3.12 only the first four PCA and LDA images are shown. The fifth PCA and LDA image are omitted because of layout constraints. However, please note that the enhancement results are obtained on five PCA and LDA images.

The text lines in the PCA images are better recognizable than the faded-out characters. Therefore, the multispectral samples are labeled as belonging to a text line or to an intermediate region between text lines. For the text line detection a method is applied that makes use of Local Projection Profiles (LPP), whereby the LPP calculation is similar to the one described in [YHKD09]:

First, the LPP is calculated for each pixel, whereby the calculation is performed in the following way: For the leftmost vertical stripe the LPP is calculated for the pixels, which are located at the horizontal center of the stripe: Therefore, all intensity values of the stripe are simply summed up along the x-axis and the resulting column vector forms the LPP of the considered pixels. In the next step, the LPP of the next vertical stripe is calculated by adding the column vector to the right of the sliding stripe and subtracting the vector to the left. The resulting image is called a smearing image, whereby the smearing is performed along the x-axis. Local minima and maxima in the smearing image are located at the vertical center of text lines and intermediate regions between text lines. In order to localize these minima and maxima, the smearing image is filtered with a Gaussian column kernel and zero crossings of its first derivative are found. Afterwards, false positives are removed by applying non-maximum and non-minimum suppression. The lines found are then dilated with disk structuring element, whereby its radius is smaller than the half of the average text lines. The resulting regions are then labeled as belonging to local minima and maxima. The parameters for the text line detection are manually set for each manuscript considered.

The corresponding multispectral samples are used for the training of an LDA classifier. Afterwards, all multispectral samples are projected on the hyperplane of the LDA classifier. The projected one dimensional samples are reshaped into an image, which is hereafter denoted as LDA image. It should be noted that the training set contains partially data that is labeled incorrect, because the text lines contain also background samples. However, the text is still enhanced compared to the PCA images. Therefore, this stage of the method is called coarse enhancement step.

The LDA images, which have been obtained on the PCA images shown in Figure 3.11, are shown in Figure 3.12. It can be seen that the writing is enhanced if the PCA image

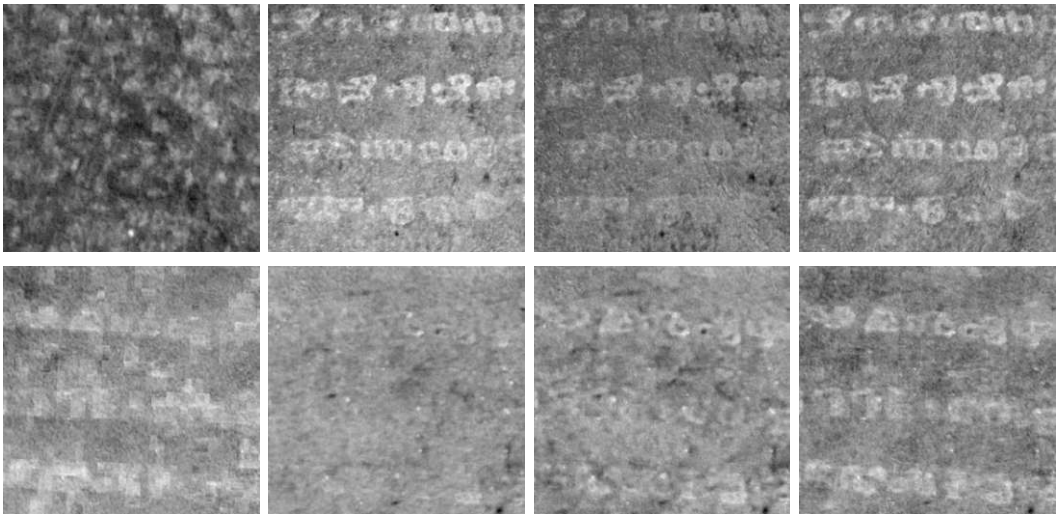


Figure 3.12: LDA images.

exhibits remnants of the degraded handwriting.

Since the writing is visible within multiple LDA images, it is desirable to generate one image out of the multiple images for further analysis. For this purpose, the PCA transformation is applied on the k LDA images. Since the writing is emphasized by the majority of the LDA images, it is enhanced by the first principal component. The resulting PCA image is the final output of the coarse enhancement stage.

Fine Enhancement

The training data in the previous stage contains partially samples with incorrect labels. The so-called fine enhancement step aims at resolving this drawback. In order to obtain accurate training samples and labels, the binarization method of Su et al. [SLT10] is applied on the resulting image of the coarse enhancement stage. This binarization algorithm assumes that the foreground is darker than background. This assumption is not valid for the enhancement results and also not for the multispectral images, since the 'Missale Sinaiticum' manuscript contains partially white ink regions. Therefore, a user is requested to define, which of the two classes is brighter than the other. Afterwards, the binarization method is applied on the result of the coarse enhancement step. The resulting image is not directly used for the labeling, since it contains false positives and negatives. Instead, the LPP method is again applied and the background samples are solely taken from regions between two text lines. Thus, the number of false positives located within text line regions is reduced.

Figure 3.13 shows the outputs of the coarse and fine enhancement stages in the third and fourth columns. It can be seen that the foreground to background is slightly enhanced in the latter mentioned images. Additionally, the multispectral channel and the PCA image

Figure 3.13: Enhancement results. From left to right: UV reflectography image. PCA image. Coarse enhancement result. Fine enhancement result.

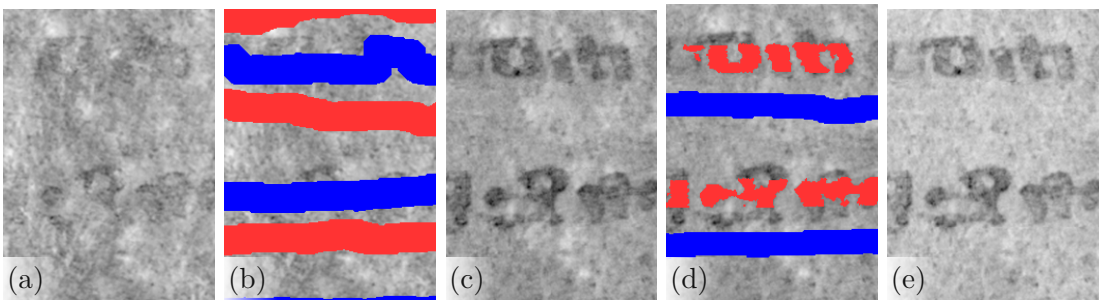


Figure 3.14: Example for the line detection and binarization steps. (From left to right) PCA input image. LPP based text line detection result obtained on the PCA image. Coarse enhancement result. LPP line detection result and binarization result obtained on the coarse enhancement result. Fine enhancement result.

in which the writing is best visible are also given in order to enable a vivid comparison to the enhancement result. These two images are provided in the first and second columns.

Figure 3.14 shows exemplar results of the LPP based text line detection and of the binarization step. The projection profiles in the second image are gained on the PCA image which is shown first in Figure 3.14. The projection profiles and the binarization result shown in the fourth image are gained on the coarse enhancement result, which is given in the third image. It is notable that the foreground to background contrast is slightly enhanced in the fine enhancement result, compared to the coarse enhancement result.

3.2.3 Results

It was already mentioned in Section 2.1.1 that there is no metric existing, which measures the legibility or enhancement increase. The enhancement strategy is especially designed for increasing the legibility of writings that are investigated by scholars. Hence, the performance is evaluated by using a quality assessment of the legibility that was conducted by philologists.

It would have been possible to fulfill the quality assessment on entire portions of manuscripts. However, due to varying spectral signatures contained on single manuscript pages, it cannot be assumed that the visual quality of all characters within a document is similar. Therefore, the evaluation is performed on a character basis instead of evaluating entire image patches or documents. For this purpose, 212 single characters were extracted from 7 panels that have been gathered from different pages belonging to two different Glagolitic manuscripts. All panels contain partially degraded characters. The evaluation was performed by two philologists, which are specialized in the investigation of Glagolitic writings. The philologists were asked to compare single characters in a pairwise manner and to assign a one to the character, which is better recognizable and a zero to its counterpart. For each character the sum of the assigned scores is calculated. The test was performed blindly, meaning the scholars did not know from which channel or enhancement result a character was taken and the characters were presented to them in a randomized order.

The enhancement method is compared against two unsupervised dimension reduction methods, namely PCA and ICA, that have been successfully used for task of legibility enhancement [JCK11] [STB07]. Similar to [STB07], the ICA implementation used is the FastICA algorithm introduced in [HO00]. It cannot be assumed that the PCA and ICA transformations that are based on local statistics gain better performances than transformations, which are based on global statistics. Therefore, both transformations have been calculated on image portions and on entire folios from which the portions were extracted. The LDA transformation is solely computed based on local statistics, since we noted that the performance is better on image portions compared to entire manuscript folios.

It is not known a-priori, which principal or independent component emphasizes the writing best. Hence, the corresponding images were selected manually before the characters were extracted. Additionally, the multispectral channel in which the text is best visible was also manually selected and added to the test set. All test images were normalized between zero and one before the characters were extracted and the resulting image patches were not further processed. Figure 3.15 shows an example of the test set along with the scores assigned by one philologist.

The first test set is comprised of panels belonging to different folios of the 'Missale Sinaiticum' manuscript. The average scores assigned by the two philologists are given in Table 3.1. The abbreviations *G.* and *L.* denote that the PCA or ICA results are obtained based on global or local statistics. The maximum average score that can be gained is

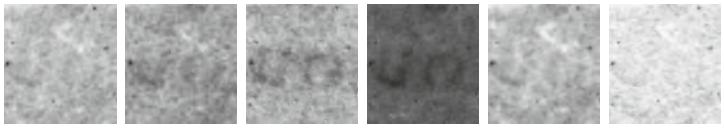


Figure 3.15: Example of the test set. The scores assigned are from left to right: 1, 3, 5, 4, 2 and 0.

#	G. PCA	L. PCA	G. ICA	L. ICA	MSI	LDA
1	2.3/1.0	2.3/1.4	1.5/1.1	2.6/1.9	0.5/0.2	**3.9**/**3.1**
2	**3.7**/**3.0**	1.3/1.2	1.5/0.6	2.5/2.6	2.8/2.6	2.0/1.3
3	1.3/1.9	1.4/1.6	2.6/2.9	2.3/1.7	0.8/1.1	**3.8**/**3.9**
4	2.5/2.0	2.6/2.2	2.8/1.8	0.5/0.3	2.6/4.3	**4.0**/**4.4**

Table 3.1: Average scores gained on 'Missale Sinaiticum' test panels

#	G. PCA	G. ICA	MSI	LDA
1	0.5/0.5	1.2/1.5	1.3/1.2	**2.7**/**2.5**
2	0.4/0.8	**2.1**/1.7	1.7/**1.9**	1.8/1.2
3	1.3/1.7	1.7/1.9	0.9/0.5	**2.1**/**2.0**

Table 3.2: Average scores gained on 'Glagolitic Fragments' test panels

five. The enhancement method yields on four out of five multispectral test images the highest average score. The method is partially outperformed on the second test panel. This can be attributed to fact that the PCA images contain a relatively large amount of background variations, which leads to wrong text line detection results. Hence, a large amount of the training data is falsely labeled and the writing is not sufficiently enhanced in the overall enhancement result. It is also not obvious in Table 3.1 if the unsupervised dimension reduction techniques gain a better performance based on global or local statistics.

The second test set contains images from a manuscript that is named 'Glagolitic Fragments'. These images have been acquired with the LED based system introduced in [HGS12] but the multispectral images have all been imaged with a Nikon D4 camera. Figure 3.16 shows an exemplar portion of the data set. The images in the top row of Figure 3.16 are multispectral channels and the bottom row shows PCA, ICA and LDA enhancement results. This test set contains only relatively small fragments and hence the unsupervised methods have been applied on the entire fragments. Thus, the maximum score that can be assigned to a character is three. The scores assigned by the scholars are given in Table 3.2. It can be seen that the enhancement method gains in two out of three cases the highest score.

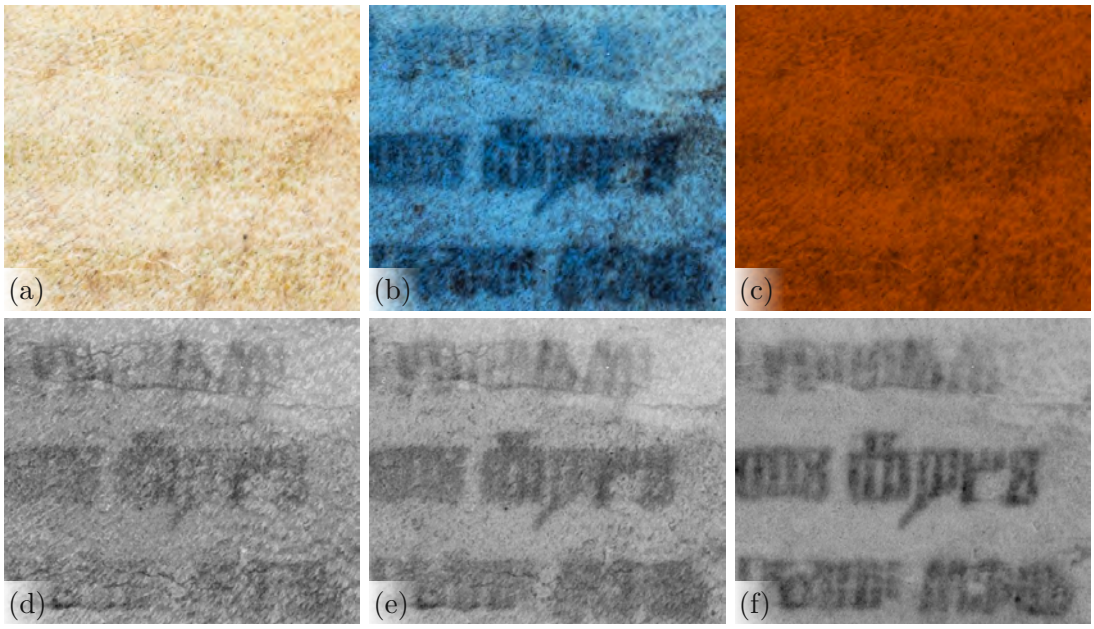


Figure 3.16: Fragment portion belonging to the 'Glagolitic Fragments'. (a) White light image. (b) UV fluorescence image. (c) 570 nm. (d) PCA image. (e) ICA image. (f) Enhancement result.

3.3 Further Applications

The enhancement method introduced in the previous section has been also used and extended for two other applications: First, the method is used in [HS14] for the enhancement of palimpsest underwritings. Second, the technique is used in [HDS14] as a preprocessing step in order to improve the performance of an OCR system. The original method proposed in [HGS13] had to be adapted in order to meet the application specific requirements. The two applications are introduced in the following, whereby the palimpsest enhancement is described in Section 3.3.1 and the OCR based application is detailed in Section 3.3.2.

3.3.1 Palimpsest Enhancement

In [HS14] a method is proposed that is concerned with the enhancement of palimpsest underwritings. The previous method in [HGS13] is designed for a single text and hence the LDA classifier is applied on a two class problem. For the separation of two palimpsest writings, i.e. an over- and underwriting, a three class problem has to be considered. One example containing such palimpsest texts is given in Figure 3.17 (a) - (c). The palimpsest is taken from a manuscript named 'Suppl. Gr. 200' and both writings are in Greek language. It can be seen that the overwriting is visible within the UV and visible range.

The underwriting is instead only partially visible under UV illumination.

In order to label the three classes, the following procedure is applied: In a first step the overwriting is detected by applying the binarization method in [SLT10] on a multispectral channel showing solely the overwriting. The younger overwriting is usually visible within the visible range and partially within the NIR range. Contrary, the underwriting is not visible within this range. Hence, the binarization approach can be successfully applied on a channel taken from the visible range. For the 'Suppl. Gr. 200' manuscript, the binarization method was applied on the red channel of the RGB image.

The labeling of the underwriting is similar to the labeling of a single writing that was introduced above. However, the PCA images show partially the overwritten text, which impairs the LPP based text line detection. Therefore, a preprocessing step is applied before the LPP based text line detection is conducted: The detected overwritten areas in the PCA images are simply filled with the median gray value of the remaining document. Afterwards, the enhancement procedure (introduced in Section 3.2.2) is applied. In the fine enhancement step, the binarization approach is not applied. Instead only the text line and intermediate regions are used in the final labeling step, because the condition of the 'Suppl. Gr. 200' manuscript prevents successful binarization. While this leads to a relatively large amount of wrongly labeled training data, the underwriting is still partially better visible within the LDA images compared to the outputs of PCA and ICA.

This can be seen in Figure 3.17 (d) - (f), where the results of the different dimension reduction methods are shown. A further enhancement result is given in Figure 3.18. The underwritten text is best visible in the UV fluorescence image shown in (a), compared to the remaining multispectral images. The writing is partially enhanced by the ICA and LDA image, whereby the underwriting in the latter mentioned has a considerably higher contrast to the remaining background. Nevertheless, it is not obvious which enhancement result is superior. The palimpsest enhancement results were not evaluated in a user study - as in Section 3.2 - because this would have caused a huge manual effort.

3.3.2 OCR Improvement

The LDA based method was also used in [HDS14], where it is evaluated if the method can be used as a preprocessing step for an OCR system, which is introduced in [DS10]. While the original enhancement method is especially designed for degraded writings, the OCR system is also applied on non-degraded writings. It was found that the method in [HGS13] is not sufficient for such non-degraded writings and hence it had to be adapted. This adaption is detailed in Section 3.3.2. The OCR system used is explained in Section 3.3.2, where it is evaluated if the enhancement method can be used to improve the accuracy of the OCR system.

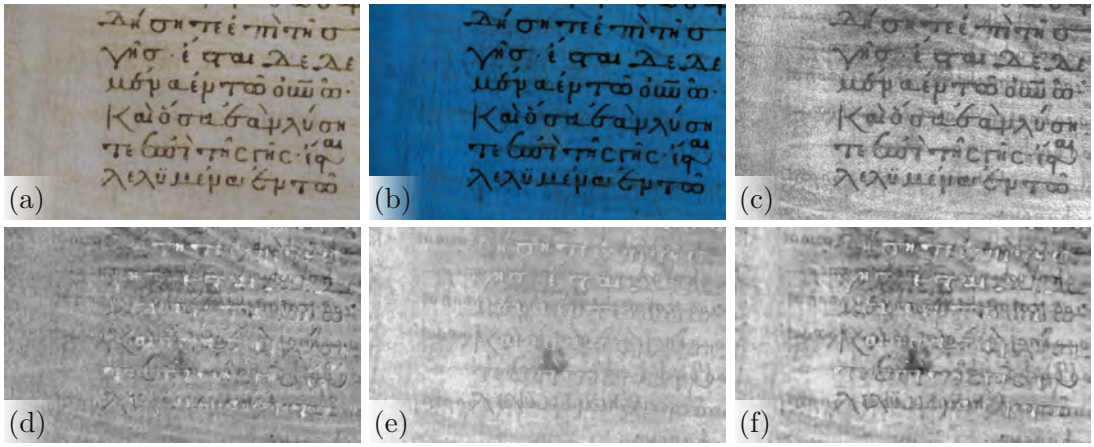


Figure 3.17: Palimpsest belonging to Suppl. Gr. 200. (a) White light image. (b) UV fluorescence image. (c) 570 nm. (d) PCA image. (e) ICA image. (f) Enhancement result.

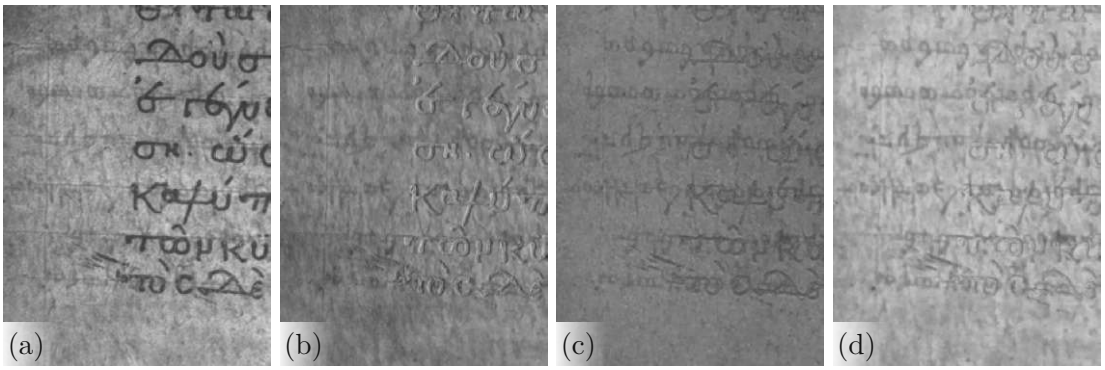


Figure 3.18: Palimpsest, belonging to Suppl. Gr. 200. (a) UV fluorescence image. (b) PCA image (c) ICA image. (d) LDA image.

Methodology

The OCR system is applied on images which contain image regions where the writing is faded-out and regions that are only partially affected by slight degradations - including background variation and bleeding artifacts. Similar to [HC13b] it was found that these two kinds of image regions require different restoration techniques. Therefore, the image is split into image patches and the patches are classified into non-degraded and degraded image regions. The segmentation into non-degraded and degraded regions is detailed in Section 3.3.2. Afterwards, the corresponding enhancement techniques are applied on the images. The retouching of both degradation types is explained in Section 3.3.2 and Section 3.3.2. Finally, both restoration results are fused together based on the segmentation map that is computed in the first step.

Segmentation Non-degraded regions can be characterized as regions containing writings that are visible within the UV and visible range. Contrary, degraded writings are only visible within selected spectral ranges, which is in our case the UV range. In the case of non-degraded regions, the variance between the foreground and background samples is considerably higher compared to degraded regions. Non-degraded writings are typically emphasized by the first principal component, whereas degraded writings are emphasized by multiple principal components [LKSM08]. Additionally, we noted that the eigenvalue of the first principal component is higher in the case of non-degraded writings.

This observation is used in the segmentation step: First, the image is split up into non-overlapping image patches. In our case the patch size used is $30 \times 30px$. This relatively small patch size is used to prevent that a single patch is overlapping more than one text line. Afterwards, the PCA transformation is applied on each image patch and the first eigenvalue is considered: If this value exceeds a certain threshold - in our case 0.1 - the patch is classified as non-degraded or otherwise as degraded. Figure 3.19 shows an example for the segmentation step: The image in Figure 3.19 (a) is taken under white light and the contrast between foreground and background is considerably larger in the inner image region compared to the outer region, where the text vanishes. This is also notable in the image in Figure 3.19 (b) which visualizes the first eigenvalues of the patches. It can be seen that the eigenvalues of the patches in the outer regions are considerably lower than in the inner regions. The final output of the thresholding step is shown in Figure 3.19 (c) and (d).

Enhancement of Non-Degraded Regions The writing in the non-degraded regions is detected by applying a document binarization method. Therefore, the binarization method of Su et al. [SLT10] is applied on an image that has been acquired with a band pass filter at 650 nm. The output of the binarization method is not directly used as input for the training of an LDA classifier, since faded-out characters might not be found by the method of Su et al. [SLT10]. In order to exclude such false negatives from the training set, the background samples are solely taken from intermediate regions between text lines. These intermediate regions are found by applying the LPP based text line

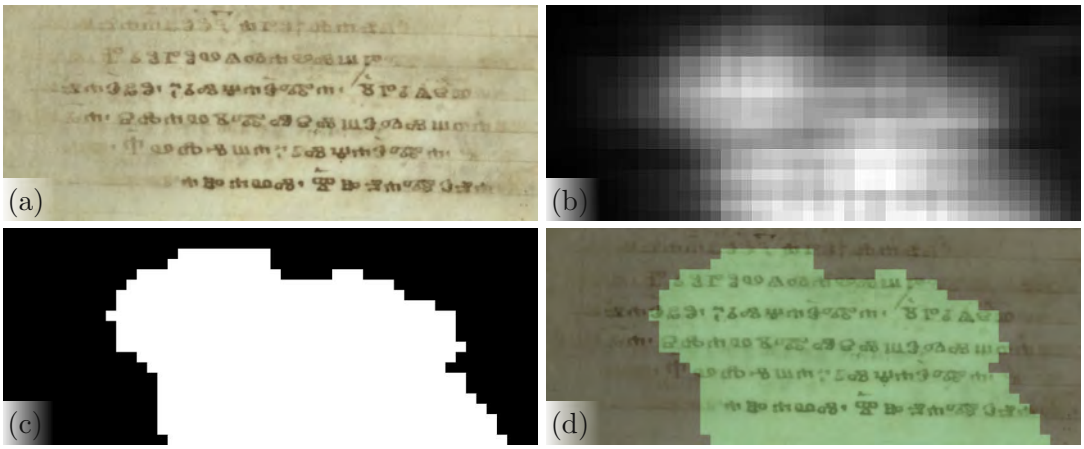


Figure 3.19: Segmentation result. (a) RGB image taken under white light. (b) Map of the first eigenvalues. (c) Segmentation output. (d) Input image with superimposed segmentation result.

detection step. The foreground image found by the binarization method is additionally eroded by a disk structuring element with a radius of two pixels. The erosion is performed in order to select only multispectral samples as foreground training samples that have a pure spectral signature. According to [HC13b] a pure pixel contains a single material, whereas mixed pixels *"may contain a mixture (i.e., a linear combination) of different materials or classes"*.

We assume that pure pixels are located in the middle of character strokes, whereas pixels located at stroke boundaries are mixed pixels. The mixed pixels are mixtures of the fore- and background and are excluded from the training set in order to obtain a resulting image with a considerably high contrast between foreground and background. Afterwards, the selected training samples are used for the training of an LDA classifier and all multispectral samples are projected on the one-dimensional hyperplane found. Figure 3.20 shows an example for the restoration of a non-degraded image region. It can be seen that the majority of the background clutter contained in the visible image shown in Figure 3.20 (left) is removed in the enhancement result shown in Figure 3.20 (right).

Enhancement of Degraded Regions For the enhancement of degraded regions, the labeling procedure introduced in Section 3.2.2 is applied on the considered image: Therefore, first the PCA images are calculated and afterwards the LPP based line detection is applied in order to generate the LDA images. Afterwards, the PCA transformation is applied on the LDA images and the first PCA image is considered. Since this image shows the writing, the overall text line detection result is obtained on this image by applying the LPP based method introduced in Section 3.2.2. In Section 3.2.2 the method is applied on entire image patches. Contrary, the dataset used in [HDS14] consists of larger image

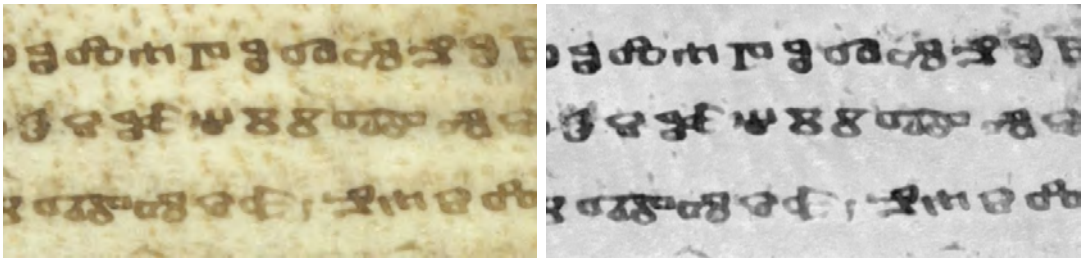


Figure 3.20: Restoration of a non-degraded image region. (Left) Image acquired at the visible range. (Right) Enhancement result.

portions and the spectral signature of the writings within single portions is more varying. In order to account for this higher spectral variability, the LDA projection is only applied on image patches. Therefore, the image is first divided into non-overlapping patches with a size of $100 \times 100px$. This patch size is used, because the average distance between two adjacent text lines in our dataset is approximately $90px$.

For each patch a distinct LDA transformation is calculated in order to solely use the spectral signatures of the pixels belonging to the considered patch. Therefore, it is first determined which pixels are belonging to text line or intermediate regions. The pixels found are then used for the training of an LDA classifier and the entire patch data is projected on the hyperplane found. The resulting image patch is used to further refine the training data: Therefore, the mean intensity values of the text line and intermediate regions are determined. Only pixels that have a smaller or higher intensity value than the average value are labeled as foreground and background pixels. The pixels found are again used in the training of an LDA classifier. This time, not only the pixels belonging to the patch are projected, because this would lead to an inferior result, where the patch-wise application is notable. One example for a patch-wise projection of the samples is given in Figure 3.21: The image in (a) is the coarse enhancement result. The image in (b) shows how the patch-wise application would look like.

Instead of applying LDA projections patch-wise, the following procedure is used: For each patch, the LDA projection is calculated and the samples belonging to the entire image are projected. The resulting images are afterwards fused into a single image, whereby the pixels are linearly weighted: For each image the Euclidean distances between all pixels and the center of the corresponding patch - from which the samples were taken - are calculated. Afterwards, the distances are subtracted from the maximum distance and the resulting weights are normalized between zero and one. Thus, the highest weights are assigned to the center of a patch, i.e. a weight of one. Lower weights are assigned to more distant pixels. The weights are multiplicated with the corresponding projected images and the normalized sum of the weighted images is the overall result. Figure 3.21 (c) shows the effect of the weighting procedure. It is notable that the contrast between writing and background is considerably higher in Figure 3.21 (c), compared to (a). However, it is also notable in Figure 3.21 (c) that contrast between foreground and background is lower

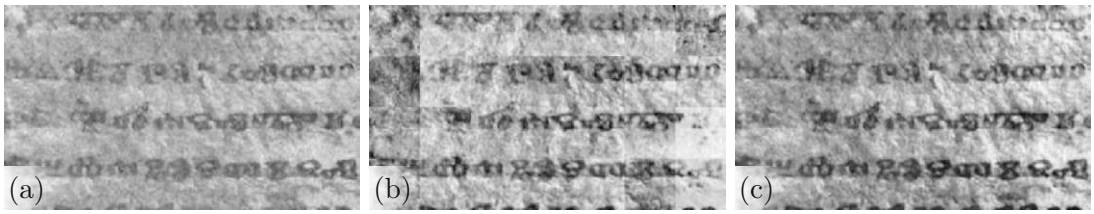


Figure 3.21: Restoration of degraded image. (a) Coarse enhancement result. (b) Result obtained by applying the projections patch-wise. (c) Fusion result.

in the left image region, compared to the remaining image. This can be attributed to the fact that the LPP based text line detection failed in the left image region. Hence, the writing is not enhanced in the corresponding patches, as it can be seen in Figure 3.21 (b).

Merging of Restoration Results In the final step of the method, the two different enhancement results are combined into a single image. This step is only necessary if two different degradation kinds are found by the segmentation step. Otherwise, the single enhancement result is used as the overall output. If two degradation kinds are present, the two results are merged based on the segmentation output: The segmentation map is first thresholded ($T = 0.5$) and the binarization result is filtered with a two-dimensional Gaussian kernel. A relatively large $\sigma = 100$ is used for the Gaussian kernel in order to enable seamless transitions between degraded and non-degraded image regions. The Gaussian filtered map is then multiplied with the output of the method for non-degraded regions and the inverted map is multiplied with the enhancement result of degraded image regions. Both multiplication results are summed up and divided by two.

Figure 3.22 shows outputs of the methods designed for non-degraded and degraded image regions. The overall resulting image is shown in 3.22 (d). The segmentation mask used was already shown in Figure 3.19.

Results

The OCR system used is especially designed for the Glagolitic writings contained in the 'Missale Sinaiticum' and was proposed by Diem and Sablatnig in [DS10]. Diem and Sablatnig [DS10] propose an OCR system that is based on local descriptors, namely Scale Invariant Feature Transform (SIFT) features proposed by [Low04]. The SIFT features are found on grayscale images and the features are classified by multiple SVM's, whereby each SVM has been trained on a single character in an on-against-all scheme. The SVM's have been trained on single characters extracted from photographs taken under white light tungsten illumination. The classification outputs of the SVM's are clustered and the final class labels are determined by applying a voting scheme. The interested reader is referred to [DS10] for more details on the OCR system.

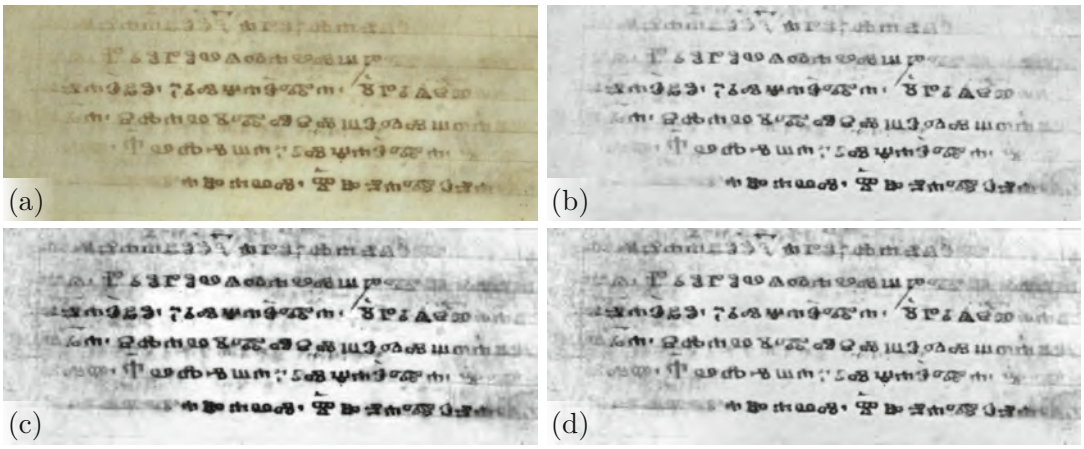


Figure 3.22: Enhancement of a degraded manuscript portion. (a) Resulting image obtained by solely projecting multispectral samples belonging to each patch considered. (b) Resulting image by projecting all samples and fusing the resulting images.

The OCR performance is evaluated by means of precision, recall and character accuracy. The method is applied on two sets of test panels. Each of the two test sets contains 11 test panels and the total number of characters is 1301. The first set contains non-degraded panels and the second set contains several degradations, including faded-out characters and background clutter. The background clutter would lead to a relatively large number False Negatives (FN) that are located on background clutter regions. Therefore, the OCR system is only applied on characters that are manually segmented. Feature points that are located on the outside of a segmented character are rejected and not classified. A character is only treated as a FN if a certain uncertainty criterion is met. Thus, the number of FN is relatively low, which results in relatively high recall values. The recall values are provided for the sake of completeness.

The performance of the method is compared to results that have been gained by applying the PCA and ICA transformation. The best output of both unsupervised dimension reduction methods was manually selected for the performance evaluation. Additionally, the performance gained on photographs taken under white light illumination is evaluated. Similar to [JCK11] it was observed that the writings are best visible within the blue channel of the UV fluorescence images. Hence, this channel is also used in the evaluation. The LDA based enhancement strategies for both degradation kinds are also evaluated separately: The enhancement result for non-degraded regions is denoted by LDA-nd and the result for degraded regions is denoted by LDA-d.

The performance gained on the non-degraded test set is given in Table 3.3. It can be seen that the best performance in terms of precision is gained on the ICA outputs. The performance gained on the VIS, UV and LDA is smaller but similar. The character

	VIS	UV	PCA	ICA	LDA	LDA-nd	LDA-d
Precision	0.87	0.89	0.83	**0.89**	0.88	0.88	0.83
Recall	**0.96**	0.96	0.94	0.94	0.94	0.94	0.96
Accuracy	0.83	**0.87**	0.78	0.80	0.83	0.85	0.80

Table 3.3: Performance gained on the non-degraded test panels

	VIS	UV	PCA	ICA	LDA	LDA-nd	LDA-d
Precision	0.61	0.57	0.57	0.55	**0.65**	0.62	0.58
Recall	0.76	0.87	0.88	0.80	**0.89**	0.88	0.87
Accuracy	0.51	0.52	0.54	0.43	**0.60**	0.57	0.52

Table 3.4: Performance gained on the degraded test panels

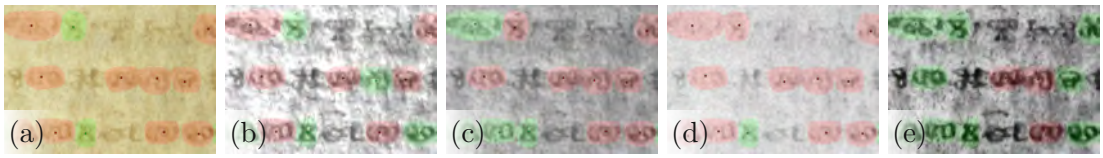


Figure 3.23: OCR results. (a) Visible image. (b) UV fluorescence image. (c) PCA result. (d) ICA result. (e) LDA result.

accuracy is more varying, whereby the best performance is achieved on the UV fluorescence images. The accuracy gained by the LDA based enhancement method is inferior. The contrast between foreground and background is higher in the UV images, compared to LDA images. The latter mentioned images exhibit more background variations, which explains the inferior performance gained on the enhancement results.

The OCR performance gained on the degraded test is shown in Table 3.4. It can be seen that the method proposed gains the highest performance values. It is also notable that the combination of both enhancement methods increases the performance. Figure 3.23 shows OCR results that are gained on an image belonging to the degraded test set. Green depicts correctly classified characters and red shows wrong OCR outputs. Uncolored characters are not annotated and have been excluded in the numerical evaluation. The annotation was conducted on RGB images that have been illuminated with white light. It is notable that the characters are best visible in the LDA image. This circumstance is also reflected by the superior output of the OCR system on the LDA image.

3.4 Summary

The advantage of using MSI as an investigation method for the analysis of ancient documents was shown in this chapter. MSI was used to increase the visibility of degraded

writings and a further contrast enhancement can be gained by reducing the dimensionality of the multispectral images. For this purpose, a new enhancement method was proposed that is based on supervised LDA dimension reduction, which requires labeled training samples. In order to automatically select and label these samples a method was introduced. The labeling is performed on PCA images and makes use of the circumstance that text lines are better recognizable in these images than characters. Therefore, an LPP based text line detection is applied on these images and the samples are labeled as belonging to text line or intermediate regions. In a final step, a finer labeling is performed by applying a binarization method.

The technique was evaluated by means of a user study, which was conducted by two scholars. The evaluation was performed on a character level in order to account for the varying spectral signatures of the writing contained in single folios. The method was compared to two unsupervised dimension reduction techniques, which were partly outperformed by the method. The results gained in the user study indicate that spatial information can be used in order to enable a supervised dimension reduction.

Additionally, the method is also evaluated by means of an OCR system. Since the method is originally designed for degraded writings, it was adopted in order to make it applicable on non-degraded images. The numerical evaluation showed that the method is only partially capable of increasing the performance of the OCR system: In the case of non-degraded writings the performance gained on enhanced images is inferior to the results gained on unprocessed multispectral images. Contrary, the results gained on degraded writings are promising.

The performance of the method strongly depends on the output of the LPP based text line detection. The method is additionally only suited for handwritings with a regular text line structure. In a future work, the enhancement method has to be improved by incorporating other state-of-the-art methods for text line detection or document layout analysis. Furthermore, the performances of other classifiers have to be analyzed - including for example SVM's or neural networks.

CHAPTER 4

MultiSpectral Document Image Binarization

This chapter is concerned with the binarization of multispectral document images. Several binarization methods have been developed during the course of this work: First, a method from the field of remote sensing, namely Adaptive Cosine Estimator (ACE) target detection, is combined with a binarization method that is designed for grayscale images. This combination is fulfilled in a rule-based system, which is described in Section 4.1.1. In Section 4.1.2 the method is extended: The rule-based system is replaced with a spatial segmentation step that makes use of the GrabCut [RKB04] algorithm. In Section 4.2 a method is introduced that makes use of GMM based clustering. An FCN approach for semantic segmentation is explained in Section 4.3. The network architecture used was introduced in [OSK18] and the network serves as a baseline for deep learning-based binarization. The binarization methods are evaluated in Section 4.4. First, the datasets and metrics used are described. Afterwards, the methods are evaluated separately. A comparison of the performances gained is given in the final section.

4.1 Target Detection

Target detection methods are used in remote sensing applications, where the aim is to determine if a rare object with a specific spectral signature is present in the acquired data [MTP+14]. According to Manolakis [MTP+14] the term rare describes the circumstance that the amount of target pixels is relatively small compared to the total number of pixels. Theoretically, target detection can be interpreted as a binary classification problem, where each pixel is classified as belonging to the target class or the non-target class, whereby the non-target class is the union of different specific background classes [MTP+14]. Thus, the aim of target detectors is similar to the aim of document image binarization, but

classic target detection algorithms (including matched filter and ACE) solely make use of spectral information and do not use spatial information.

This aim of target detectors can be formally defined by:

$$y = T(\mathbf{x}, \mathbf{s}), \tag{4.1}$$

where y is the detector output which describes the probability that the test pixel $\mathbf{x}$ is belonging to the target class. $\mathbf{s}$ is the spectral signature of the target class and T is the transformation of the target detector. If y exceeds a certain threshold, the pixel $\mathbf{x}$ is classified as belonging to the target class or otherwise as belonging to the non-target class.

Multiple target detectors have been developed, including Matched Filter (MF) and CEM. These methods make use of first and second order statistics and apply a linear transformation of the multispectral data. Contrary, the ACE target detector finds a non-linear transformation. Since ACE target detection is partially superior to MF and CEM based target detection [CS05], [TFF07], the ACE target detector is used in the method proposed.

ACE Target Detector

The ACE target detector is formally defined by:

$$y(x) = \frac{sign\left[(\mathbf{s}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]\left[(\mathbf{s}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^2}{\left[(\mathbf{s}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{s}-\boldsymbol{\mu})\right]\left[(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]}, \tag{4.2}$$

$$y_{ACE}(x) = \begin{cases} 1 & y(x) > 1 \\ 0 & y(x) < 0 \\ y(x) & otherwise \end{cases} \tag{4.3}$$

where $\mathbf{x}$ is the spectral signature of the pixel, $\boldsymbol{\mu}$ is the average spectral signature of the data and Σ^{-1} is the inverted covariance matrix of the background with $\Sigma = (\mathbf{x}-\boldsymbol{\mu})' * (\mathbf{x}-\boldsymbol{\mu})/(N-1)$. Singular Value Decomposition is used to compute the inverse of the covariance matrix.

Figure 4.1 shows an example for ACE based target detection. The number on the top, i.e. *1672*, is an annotation and is written with a different ink than the remaining handwriting. This circumstance is also visible in the target detection result, because the class probability assigned by the ACE detector is relatively low. However, it can also be seen that thin strokes belonging to the iron gall based ink have also a relatively low class probability. Hence, the output of the ACE detector cannot be directly used by applying a simple threshold.

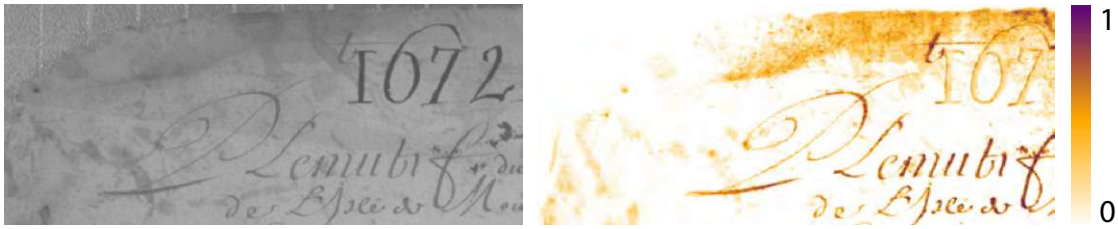


Figure 4.1: ACE based target detection. (Left) Manuscript imaged at 500 nm. (Right) ACE result. The result is color coded as indicated by the bar on the right.

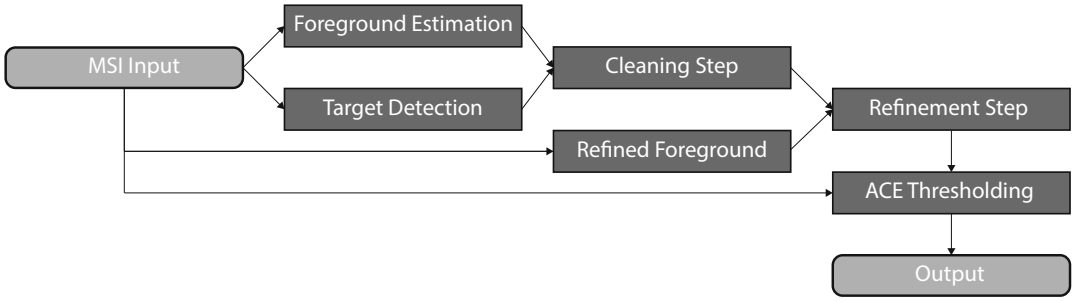


Figure 4.2: Outline of the ACE base method.

Instead, we propose two different methods that make use of spatial information in order to refine the output of the ACE target detector. The method in [HDS15] was published first and is hereafter denoted by base method. This method is first introduced in Section 4.1.1. The second method [DHS16] is an extension of the base method and is described in Section 4.1.2.

4.1.1 Base Method

The method consists of several consecutive steps, which are outlined in Figure 4.2. The technique makes use of the ACE target detector and hence it is necessary to define the target signature $\mathbf{s}$. Since the spectral target signatures are varying between different multispectral images, the target signatures are directly estimated on the multispectral images considered. This estimation is based on the outcome of a foreground estimation step. The output of the target detection is then binarized using Otsu [Ots79] threshold. Afterwards, a cleaning and a refinement step are applied in order to reduce the number of false positives and false negatives. The output of this stage is then finally combined with a threshold result that is gained on the target detection image.

Target Detection

In order to estimate the target signature $\mathbf{s}$ the binarization method suggested by Su et al. [SLT10] is first applied on a selected channel of the multispectral image. The binarization

output is hereafter denoted by S and serves as the basis for the remaining steps below as well as for the target detection. It should be noted that other state-of-the-art binarization methods can also be used instead of the method of Su et al. [SLT10]. The binarization method used should gain a high precision rather than a high recall, because false positives lead to an altered estimation of the target signature. The method of Su et al. incorporates the idea of stroke width and hence the algorithm does not over-segment regions, which are affected by background variations. This is especially advantageous in our case, because large and dark areas - such as ink stains - are not affecting the target signature estimation.

S is found by applying the binarization method on a channel, where the iron gall based ink is best visible. The channel used for binarization is hereafter denoted by f_{text}. The image that is used for the MSBin dataset is acquired at 365 nm, because the foreground to background contrast is maximized in this channel. For the MS-TEx dataset, the handwriting is best visible in the channel that is acquired at 500 nm.

However, other foreground objects are also visible within f_{text}, including annotations, stamps or degradations. Since these objects are also partially classified as foreground regions in S, the target signature is not directly estimated from the output of the binarization method. Instead, the following outlier removal step is used to remove false positives. Thus, the target signature is estimated from foreground regions, which are likely true positives, and s is not affected by spectral outliers that are belonging to other classes, such as stamps.

First, the median spectral signature $\tilde{\mathbf{x}}$ of the foreground pixels found is calculated. Afterwards, the absolute point-wise distances between the foreground samples and the median signature $\tilde{\mathbf{x}}$ are calculated.

$$D(x_i) = |x_i - \tilde{x_i}| \quad \forall i = 1, ..., n \tag{4.4}$$

whereby n is the number of channels, e.g. the number of channels of the multispectral image. Thus, for each foreground pixel an n-dimensional distance vector to the median value is calculated. A multispectral sample $\mathbf{x}$ is classified as an outlier if 50% or more of the elements of its corresponding distance vector $D(\mathbf{x})$ are exceeding the following threshold:

$$t_{\sigma_i} = \sigma(x_i)/3 \quad \forall i = 1, ..., n \tag{4.5}$$

where $\sigma(x_i)$ is the standard deviation of the foreground pixels of the i-th channel.

The outlier removal assumes that the majority of the foreground pixels found by the binarization algorithm is correctly labeled. This assumption is valid for the datasets considered, because the number of pixels belonging to annotations or stamps is always lower than the number of pixels belonging to the iron gall-based ink. The target signature is afterwards determined, by calculating the average spectral signature of the result of the outlier removal step.

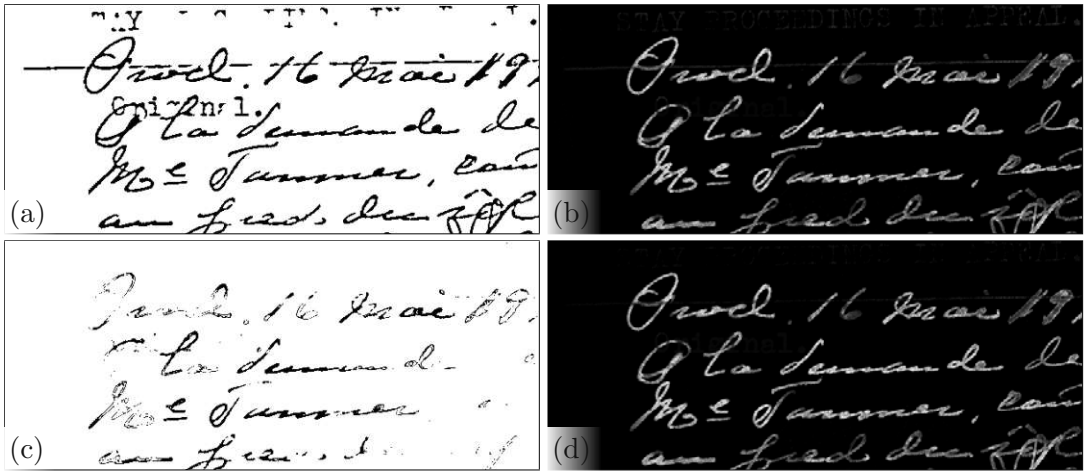


Figure 4.3: Influence of the outlier removal step on the target detection. (a) Initial binary image. (b) Corresponding ACE image. (c) Output of the outlier removal step. (d) Resulting target detection image.

Figure 4.3 shows an example for the influence of false positives on the target detection result. The upper image half contains a stamp region, which is partially labeled as belonging to the foreground. The image in Figure 4.3 (a) shows the output of the binarization algorithm of Su et al. [SLT10]. Figure 4.3 (b) shows the corresponding ACE output, which is gained by using the mean spectral signature of the foreground pixels as target signature. The stamp is partially visible within the image. The result of the outlier removal step is given in Figure 4.3 (c). The corresponding ACE result is shown in 4.3 (d). It is notable that the stamp is less visible than in the ACE output that is gained by using the output of the binarization algorithm.

Cleaning Step

The target detection resulting image $y_{ACE}(x, y)$ is used to remove false positives that are found by the binarization method suggested by Su et al. [SLT10]. This step of the method is hereafter denoted as 'Cleaning Step'. In the first step the target detection image y_{ACE} is binarized by applying a global Otsu [Ots79] threshold:

$$P(x, y) = \begin{cases} 1 & y_{ACE}(x, y) \geq T \\ 0 & otherwise \end{cases} \tag{4.6}$$

where $P(x, y)$ is the result of applying the global Otsu threshold T. The foreground pixels found are similar to the target signature and are likely true positives. These pixels are hereafter denoted as 'true positive candidates'. These pixels are then used to eliminate

false positives contained in the binarization output S. The elimination of false positives assumes that actual text pixels are located in the near of 'true positive candidates'.

The text pixels are found by applying the following procedure: First, the number of foreground pixels in $P(x, y)$ within a local neighborhood is calculated. The resulting image is denoted by $N_P(x, y)$. Afterwards, each foreground pixel in the binarization output S is analyzed: A pixel $S(x, y)$ is only classified as foreground pixel if the number of true positive candidates $N_P(x, y)$ exceeds a certain threshold. The heuristic is formally defined by:

$$PN(x,y) = \begin{cases} 1 & N_P(x,y) \geq N_{min} \wedge S(x,y) = 1 \\ 0 & otherwise \end{cases} \tag{4.7}$$

where N_{min} is the minimum number of true positive candidates. A local neighborhood size of 21×21 is used and N_{min} is set to 1. These parameters were empirically chosen.

The heuristic used is inspired by a similar heuristic used in the binarization method suggested by [SLT10]: Therein, a pixel is considered as foreground candidate, if the number of high contrast pixels within a local neighborhood exceeds a certain threshold.

The overall output of the 'Cleaning Step' is the union of $P(x, y)$ and $PN(x, y)$:

$$C(x,y) = P(x,y) \vee PN(x,y), \tag{4.8}$$

where $C(x, y)$ is the final result of the 'Cleaning Step'. Figure 4.4 (b) shows an example of the 'Cleaning Step' that is gained on the output of the method of Su et al. [SLT10]. The binarization results are color coded: Red depicts false positives, black visualizes false negatives and green depicts true positives. Note that this color coding is also used for further visualizations of binarization results within this work.

Refinement Step

The resulting image of the preceding steps exhibits mainly foreground pixels, which are located at relatively thick and dark strokes. This can be attributed to the fact that the binarization method of Su et al. [SLT10] favors these pixels, instead of pixels that are located at thin and elongated strokes. In order to resolve this drawback, it is necessary to identify such false negatives. This identification is performed in the so-called 'Refinement' step:

First, the binarization algorithm in [SLT10] is again applied on the same input image, which was used in the overall first step. Contrary, the contrast image - defined in Equation 2.6 - is not used. Instead, the detection of the high contrast pixels is based on the ACE output: Therefore, the gradient magnitude image of $y_{ACE}(x, y)$ is calculated and Otsu [Ots79] thresholded. The thresholded image is afterwards dilated with a disk structuring element with a radius of 2px. The foreground pixels found in this image are then used as

the high contrast pixels. The high contrast pixels are then used for computation of the binarization output, as it is suggested by Su et al. [SLT10]: A pixel in the input image is classified as belonging to the foreground, if a sufficient number of high-contrast pixels are located in its local neighborhood and if its intensity value is smaller than the average intensity within the local neighborhood.

Afterwards, the binarization algorithm proceeds as described in Section 2.2.2 and its output is combined with the output of the 'Cleaning Step', $C(x, y)$: If a CC in $C(x, y)$ is connected with a CC in $C(x, y)$, the latter is added to the result of the 'False Positive Elimination'. Thus, only existing strokes (or CC's) are extended in the 'Refinement Step'. The overall output of the 'Refinement Step' is denoted by $R(x, y)$. Figure 4.4 shows an example output of the 'Refinement Step'.

ACE Thresholding

In the final step of the method, the ACE image $y_{ACE}(x, y)$ is combined with the 'Refinement Step' resulting image $R(x, y)$. If the target detection assigned a relatively high or low likelihood that a pixel contains the target signature it is assumed to be labeled correctly. Otherwise, the corresponding output in R is used. Thus, the last step of the method is formally defined by:

$$B(x,y) = \begin{cases} 1 & y_{ACE}(x,y) \geq t_{high} \\ 0 & y_{ACE}(x,y) \leq t_{low} \\ R(x,y) & otherwise \end{cases}, \tag{4.9}$$

where $B(x, y)$ is the final output of the ACE based binarization method. t_{high} and t_{low} are thresholds that are empirically chosen based on the evaluations, which are conducted on the training datasets. Figure 4.4 (d) shows the final example output. It can be seen that the initial binary image contains false positives, which are arising from a machine-written text. The corresponding ink is different from the iron-gall based ink that was used for the handwritten text. The binarization method of Su et al. [SLT10] is not capable of differentiating between the two ink types and hence the machine-written text is partially classified as belonging to the foreground. Contrary, the method proposed correctly classifies the machine-written as belonging to the background.

4.1.2 GrabCut Based Method

The method introduced above makes use of heuristics in order to combine an ACE target detection result with the output of a state-of-the-art binarization method. Contrary, in [DHS16] a method is introduced that combines ACE based target detection image with a segmentation algorithm that is based on energy minimization. The method is outlined in Figure 4.5. Similar to the base method, the target detection is based on the output of the foreground estimation, which is gained on a single channel. Afterwards, the spatial segmentation is performed that is guided by the information which is gained

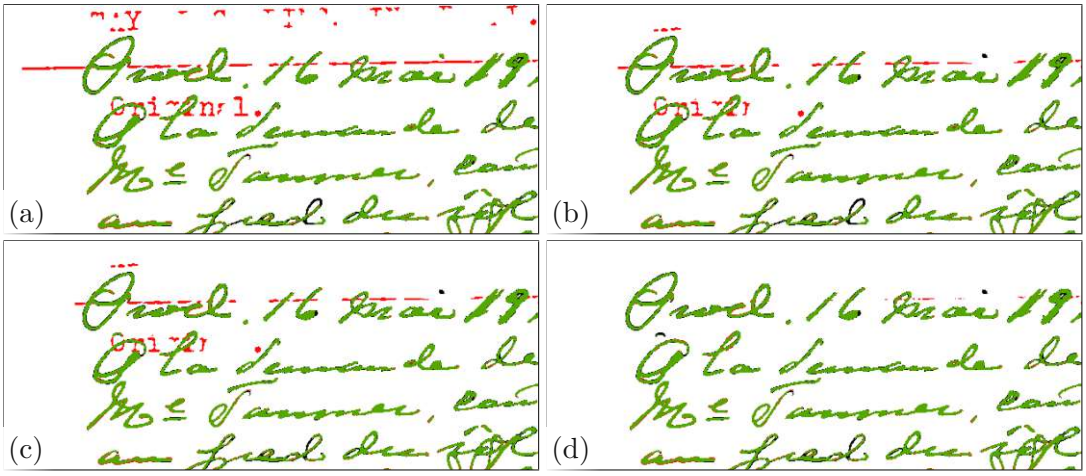


Figure 4.4: Binarization steps. (a) Initial binary image. (b) Result of the 'Cleaning Step'. (c) 'Refinement' result. (d) Final output.

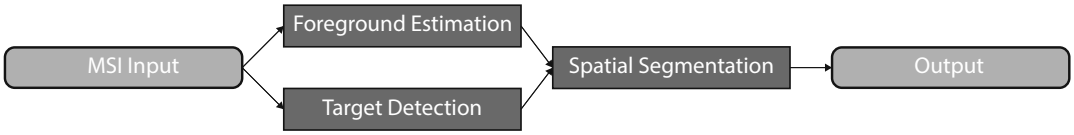


Figure 4.5: Outline of the GrabCut based method.

in the target detection and foreground estimation steps. The overall method is detailed in the following.

Target Detection

Similar to the previous ACE based binarization method, the binarization method of Su et al. [SLT10] is applied on the channel f_{text} where the text is most visible. The binarized image S is again analyzed in order to remove spectral outliers. Spectral inliers are foreground pixels, which fulfill the following condition:

$$f_i(x) = q_{25} - w \cdot iql < f(x) < q_{75} + w \cdot iql \tag{4.10}$$

where $f(x)$ is the multispectral observation at a location x, q_{25} and q_{75} being the first and the third quartiles respectively and iql being the interquartile range $q_{75} - q_{25}$. w is a weighting factor, which is set to 1.5. By using Equation 4.10 up to 50% of the foreground pixels in $B(x, y)$ can be identified as being spectral outliers. The spectral inliers are used for the ACE target detection, which was already defined in Equation 4.2.

Spatial Segmentation

The segmentation algorithm used is named GrabCut and is proposed by Rother et al. [RKB04]. The GrabCut algorithm is an extension of a Graph Cuts based segmentation algorithm introduced by Boykov and Jolly [BJ01]. The Graph Cut algorithm in [BJ01] poses the segmentation problem as an energy minimization problem. Note that the document image binarization algorithm of Howe [How11] is also based on the Graph Cut algorithm [BJ01] - as it was described in Section 2.2.2. The original Graph Cut algorithm is designed for monochrome images, whereas its extension by Rother et al. [RKB04] is designed for color images. The original data term in [BJ01] is based on histograms and it is replaced by a color GMM in [RKB04]. Additionally, Rother et al. [RKB04] propose an iterative estimation, which reduces the user interaction needed - compared to the algorithm in [BJ01]. The GrabCut is still semi-automated, because user guidance is required within the segmentation process.

GrabCut segmentation is designed for RGB color images, whereas the multispectral data considered has a greater number of channels. In order to make the GrabCut applicable, the MSI data is first converted to a pseudo-color image p using:

$$p_b(x,y) = f_{text}(x,y) - f_{nir}(x,y) \tag{4.11}$$
$$p_r(x,y) = \mu(x,y) \tag{4.12}$$
$$p_g(x,y) = \sigma(x,y) \tag{4.13}$$

where $\mu(x,y)$ and $\sigma(x,y)$ denote spectral mean and standard deviation. The blue channel of the pseudo-color image is gained by subtracting the NIR channel f_{nir} from the visible channel f_{text}. The NIR channels used are acquired at 1100 nm for the MS-TEx dataset and at 870 nm for the MSBin dataset. The subtraction is based on the observation, that the iron-gall based ink is not visible in the NIR channel, whereas background variation and other degradations are partially visible in the latter mentioned image. Note that this subtraction is similar to background removal or subtraction, which is conducted by multiple binarization methods (as described in Section 2.2).

The target detection result y_{ACE} is not used for the generation of the pseudo-color image, because this would bias the GrabCut result towards the target detection result and the advantage of the spatial segmentation would be lost. Instead y_{ACE} is used to generate a segmentation mask:

In the original GrabCut algorithm the user is requested to define a rough initial segmentation mask. Contrary, the binarization method does not require user guidance, since the segmentation mask is generated by combining the target detection image (y_{ACE}) with the initial binarization output (S):

$$m(x,y) = \begin{cases} \text{F} & \text{if } y_{ACE}(x,y) > t_f \wedge S(x,y) = 1, \\ \text{B} & \text{if } y_{ACE}(x,y) < t_b \wedge S(x,y) = 0, \\ \text{PF} & \text{if } y_{ACE}(x,y) > t_{pf} \vee S(x,y) = 1, \\ \text{PB} & \text{else.} \end{cases} \tag{4.14}$$

where $m(x,y)$ is the segmentation mask and the mask labels are abbreviated with: Foreground (F), Background (B), Potential Foreground (PF) and Potential Background (PB).

The initial mask is then used in the iterative energy minimization of the GrabCut algorithm. After each step of the energy minimization the resulting mask is altered in the following manner: The union of the foreground and potential foreground pixels is eroded until it overlaps less than 5% of the binarization output S. If the remaining image contains foreground pixels, these pixels are set as belonging to the definitive background (B). The energy minimization is stopped if no remaining pixels are found. This strategy makes use of the stroke width idea that was introduced in [SLT10]. Thus, the GrabCut does not segment large regions as foreground. After the energy minimization converges, the final binarization foreground is the union of foreground and potential foreground pixels.

Figure 4.6 illustrates the spatial segmentation steps. The pseudo-color image that is used as input image is given in Figure 4.6 (a). Figure 4.6 (b) - (d) show GrabCut segmentation masks. The color encoding used is: White depicts B, yellow PB, pink PF and violet F. In the initial segmentation mask in Figure 4.6 (b) the stamp in left image half is labeled as potential foreground. After, the first GrabCut iteration (shown in Figure 4.6 (c)) the majority of the pixels belonging to the stamp is labeled as potential background. In the final GrabCut output (given in Figure 4.6 (d)) nearly all pixels belonging to the stamp are labeled as potential background. The binarization output of the GrabCut algorithm is shown in Figure 4.6 (e).

The different stages of the method are illustrated in Figure 4.7. It is notable that the algorithm of Su et al. [SLT10] (shown in Figure 4.7 (a)) is not capable of differentiating between the ink of the stamp and the iron-gall based ink. Contrary, the method proposed classifies the stamp pixels correctly as background pixels - as it can be seen in the overall output, which is shown in Figure 4.7 (e).

4.2 Gaussian Mixture Model

Another binarization method is proposed in [HDS18]: The method makes use of GMM based clustering and groups spectral signatures by fitting two GMM's with EM. It was found that the GMM's are sensitive to background variations and tend to model these variations with multiple Gaussians. This leads to over-segmented backgrounds

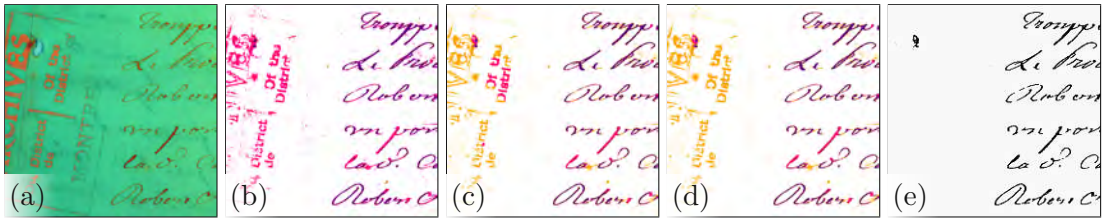


Figure 4.6: GrabCut inputs and (intermediate) results. (a) Pseudo-color image used for the segmentation. (b) Initial segmentation mask. (c) First GrabCut iteration. (d) Final GrabCut output. (e) Binarization result.

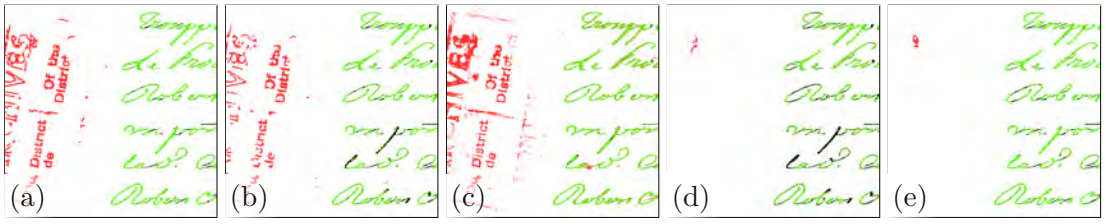


Figure 4.7: Binarization steps. (a) Initial binary image. (b) Spectral inliers. (c) ACE result thresholded with t_{pf}. (d) ACE result thresholded with t_f. (e) Final result obtained by applying the GrabCut algorithm.

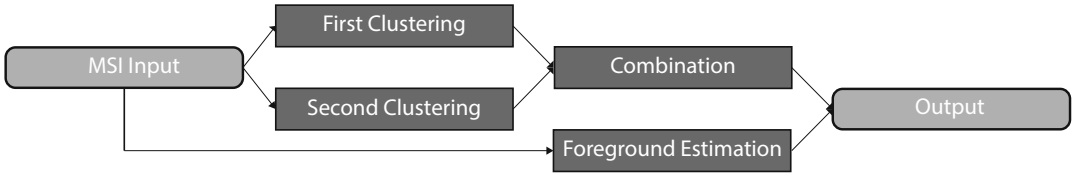


Figure 4.8: Outline of the GMM based method.

and impedes a separation between foreground and background. Therefore, a simple preprocessing step is applied before the GMM's are fitted on the multispectral data.

The method is outlined in Figure 4.8. First two different GMM's are fitted (here denoted with 'First Clustering' and 'Second Clustering'). Afterwards, the outputs are combined and the resulting image is combined with foreground estimation that is gained on a single channel.

This section is structured as follows. First, the underlying theory of GMM and EM are detailed. Afterwards, the preprocessing step is described. The clustering steps are introduced in Section 4.2.3 and the combination of both steps is detailed in Section 4.2.4.

4.2.1 GMM Theory

A GMM is a density model, which combines a number of N Gaussian distributions:

$$p(\mathbf{x}) = \sum_{i=1}^{1} \alpha_i \mathcal{N}(\mathbf{x}|\mu_i, \mathbf{\Sigma}_i), \tag{4.15}$$

where $\mathcal{N}$ is a Gaussian normal distribution with the covariance $\mathbf{\Sigma}_i$ and a mean μ_i. D is the number of dimensions, which is in our case the number of multispectral channels. α_i is a weighting factor, with $\sum_{i=1}^{N} \alpha_i = 1$.

The parameter set $\theta = \mathbf{\Sigma}_i, \mu_i, \alpha_i : i = 1, ..., N$ is learnt for each of the N Gaussian The parameters are obtained by maximizing the log-likelihood of the training data $\mathcal{X} = x_1, ..., x_N$:

$$p(\mathcal{X}|\theta) = \prod_{n=1}^{N} p(x_n|\theta) \tag{4.16}$$

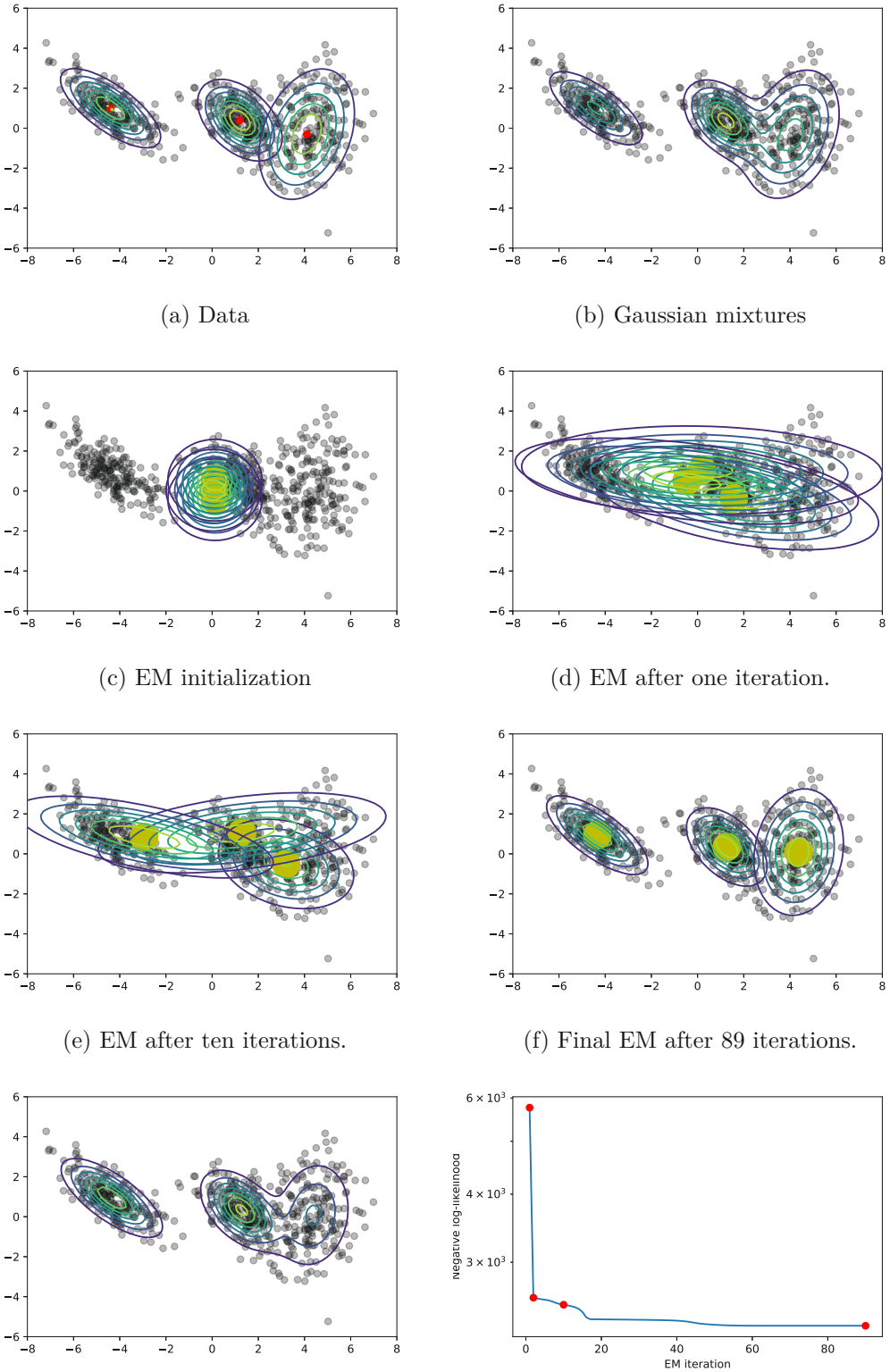
There is no closed-form solution for Equation 4.16 existing [DFO20]. Instead, the parameters are estimated by applying the EM algorithm: The EM algorithm is proposed by Dempster et al. [DLR77] and *"is a general iterative scheme for learning parameters (maximum likelihood or MAP) in mixture models"* [DFO20]. Basically, the EM algorithm consists of the following two steps:

E-Step Compute the posterior probability of a data point belonging to mixture component k [DFO20].

M-Step Use the updated posterior probabilities to estimate $\mathbf{\Sigma}_k, \mu_k, \alpha_k$.

These two steps are iterated until the increase of the log-likelihood is smaller than a predefined threshold. Dempster et al. [DLR77] have shown that the log-likelihood is increased after each EM iteration or is at least left unchanged [NH98]. More details on the EM algorithm can be found in [DLR77] and [DFO20].

Figure 4.9 shows an example for a GMM that is estimated by applying the EM algorithm. The data is two-dimensional and has been generated by randomly sampling from three different Gaussian distributions. The components of the GMM are shown in 4.9 (a) and the GMM are given in Figure 4.9 (b). The EM algorithm is initialized with three components, whose distributions are illustrated in Figure 4.9 (c). Particular EM iterations are given in Figure 4.9 (d) - (f), whereby the algorithm converged after 89 iterations. The overall estimated GMM is shown in Figure 4.9 (g). The negative log-likelihood is given in Figure 4.9 (h), whereby the red dots indicate the iterations that are shown in (d) - (f). It is notable that the estimated GMM is similar to the real GMM although inappropriate values were used in the EM initialization.



(a) Data

(b) Gaussian mixtures

(c) EM initialization

(d) EM after one iteration.

(e) EM after ten iterations.

(f) Final EM after 89 iterations.

(g) Final estimated GMM.

(h) Negative log-likelihood.

Figure 4.9: GMM estimation by applying the EM algorithm. The plots are created on the basis of a script that is provided as a supplementary material in [DFO20].

Figure 4.10: Clustering of multispectral image. (a) Image acquired at 500 nm. (b) k-means output. (c) GMM output.

In the example above, the components were initialized with inappropriate values in order to show that the EM algorithm is still capable of estimating the Gaussian distributions. However, a proper initialization of the EM algorithm is desirable: According to Bishop [Bis07] *the K-means algorithm itself is often used to initialize the parameters in a Gaussian mixture model before applying the EM algorithm.* In the binarization method proposed, the components are also initialized with the k-means clustering results. The k-means algorithm itself is initialized with the k-means++ [AV07], which is an improved seeding method for the k-means algorithm.

The k-means algorithm is a clustering method, which assigns each data point to a certain class. Contrary, the GMM makes a soft assignment to each sample [Bis07]. In this work, GMM is used as a clustering method: Therefore, each sample is labeled as belonging to the component with the maximum posterior probability. One example for GMM based clustering is given in Figure 4.10. The clustering result is color-coded[1]. The multispectral image is clustered using k-means and using GMM, whereby for each clustering method the number of components was set to eight. It can be seen that both clustering methods assign the foreground pixels as belonging to multiple components. This can be attributed to the fact that outer and thin stroke regions contain a *mixed* spectral signature, whereas inner stroke regions have a *pure* spectral signature. The k-means assigns the foreground pixels and the stamp pixels partially to the same classes. This output is inaccurate, because the handwriting and the stamp have been created with different inks. Contrary, the GMM output is superior, since the handwriting and stamp pixels are assigned to different classes. Additionally, the remaining classes, i.e. the background and bleed through, are better separated in the GMM output. It should be noted that both outputs have been gained on a multispectral image, which was preprocessed. This preprocessing is detailed in the following.

[1]The colors have been manually selected in order to achieve a sufficient contrast between the assigned components.

4.2.2 Preprocessing

The GMM based clustering of the multispectral images is error-prone, because the GMM tends to create components, which model background variations. Thus, the background is over-segmented and it was found that *mixed* foreground pixels are also partially assigned to components, which are modeling the background variations. In order to lower the influence of background variations, a simple background suppression method is introduced: First, the background is estimated for each channel of the multispectral image, by filtering each channel with a two-dimensional median filter.

$$bg_i(f_i) = med(f_i) \tag{4.17}$$

whereby $bg_i(f_i)$ is the background estimation result and med depicts the two-dimensional median filter. The filter sizes are experimentally determined. The sizes of the filters were found empirically, because it twas found that the filtered images are solely dependent on background pixels and not influenced by foreground pixels. The background compensated image fc_i is afterwards constructed by subtracting the background estimation image from the input image:

$$fc_i = f_i - bg_i(f_i) \tag{4.18}$$

Figure 4.11 shows the effect of the background compensation step. The top row in Figure 4.11 contains unprocessed channels belonging to a multispectral image of the MS-TEx database. The second row in Figure 4.11 contains the same channels, whereby the contrast was manually increased in order to visualize the background variation. The background variation is partially caused by uneven lighting, which is barely visible in the top row of Figure 4.11. The background compensated images are given in the third row of Figure 4.11, whereby their contrast was also manually increased. It can be seen that the background variation is considerably lower compared to the images in the second row of Figure 4.11. Figure 4.11 (k) and (l) show the GMM results gained on unprocessed and processed images, respectively. It is obvious that the latter image is not affected by the background variations. It can also be seen that the foreground is not well separated in the result obtained on unprocessed images: The foreground pixels are assigned to components, which are also modelling background regions (colored dark red and green in Figure 4.11 (k)).

4.2.3 Clustering

The preprocessed images are used in the GMM based clustering, whereby the EM algorithm is initialized with parameters that are found by applying the k-means algorithm. It was found experimentally that by using the same covariance matrix $\mathbf{\Sigma}$ for all components, the influence of background variations can be further lowered. One example for this circumstance is given in Figure 4.12, where the image on the left is gained by using the same covariance matrix for all components. Contrary, the output gained by using different covariance matrices is shown in Figure 4.12 (right).

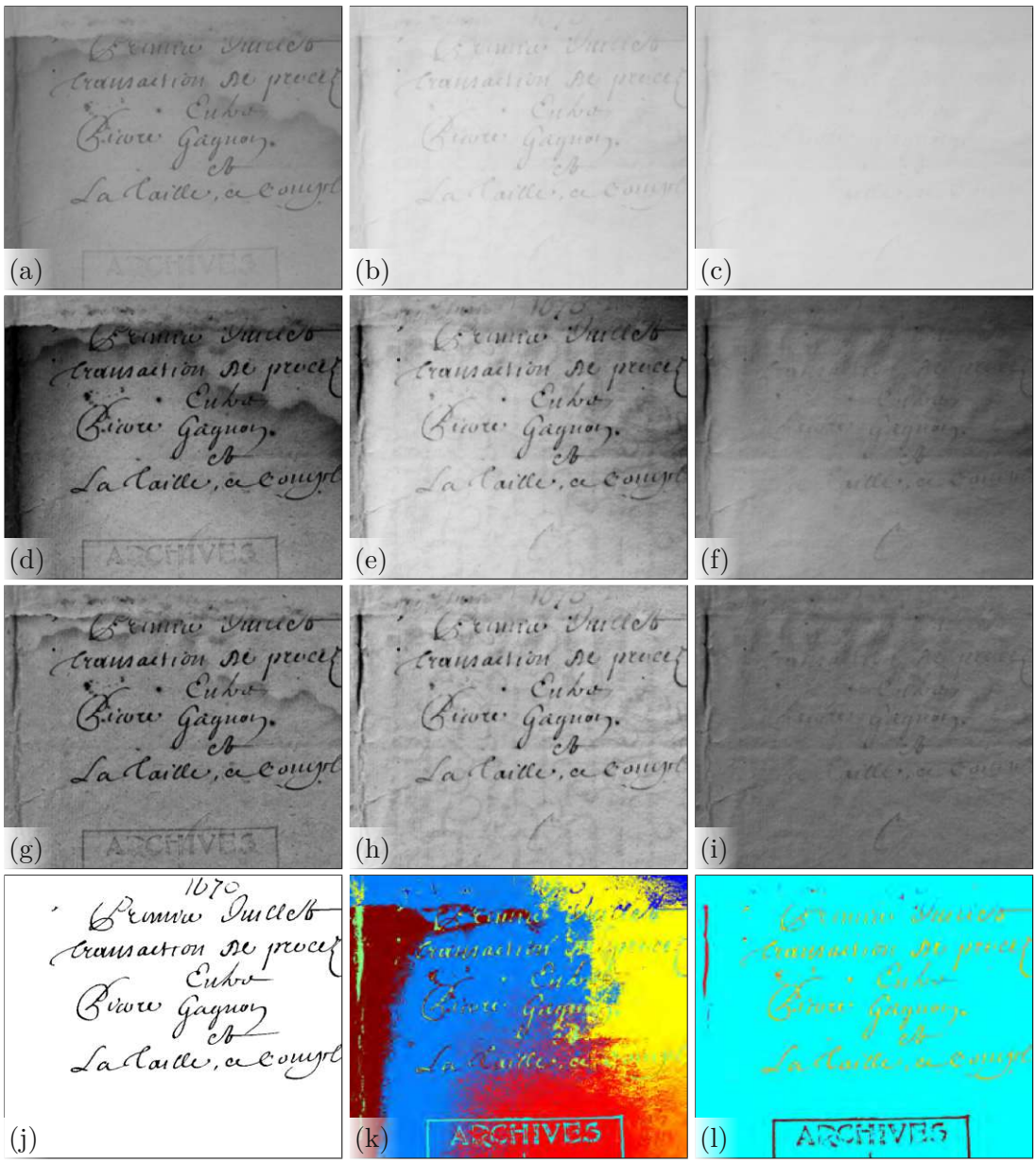


Figure 4.11: Clustering of a multispectral image. Unprocessed images acquired at (a) 500 nm, (b) 800 nm and (c) 1100 nm. (d) - (f) Images with manual contrast stretching. (g) - (i) Background compensated images with manual contrast stretching. (j) Ground truth. (k) Clustering result gained on unprocessed images. (l) Clustering result gained on preprocessed images.

Figure 4.12: Effect of the filtering of the multispectral images. (Left) GMM clustering result obtained on unprocessed data. (Right) Result obtained on images that have been filtered with a median filter. For both results the number of components is set to eight.

It was observed that by using the same covariance matrix for all components, the clustering result is less corrupted by background clutter. However, the drawback of using the same covariance matrix is that the handwriting is often modeled by more components, compared to using multiple covariance matrices. Hence, the background is better modeled by using the same covariance matrix, whereas the handwriting is better modeled by using different covariance matrices. In order to make use of these two advantages, the EM algorithm is applied twice:

First, the EM algorithm is initialized with parameters that are found by applying the k-means algorithm. In the first run, the same covariance matrix is used for all components. The parameter set θ found in the first run is then used for the initialization of the second run. This is done for all Gaussians that are identified in the EM step, except for two cases: (1) If the number of pixels that are assigned to a certain Gaussian is smaller than a predefined threshold, the component is not used in the second clustering step. This threshold is used to filter out sensor or image noise which is partially present in the MS-TEx images. The threshold used is 200px. (2) If a component is modeling mixed pixels, it is also rejected. It was found that this rejection improves the clustering in the second step, since the EM algorithm converges to a solution in which the mixed and pure spectral signatures are often modeled by the same component.

In order to identify pure and mixed pixels, the following strategy is applied: First, the binarization algorithm of Su et al. [SLT10] is applied on the channel f_{text}, which was defined in Section 4.1.1. Second, the skeleton of the binarization output (S) is computed and compared with the GMM clustering output: The class that has the highest overlap with the skeletonized image is assumed to model the pure ink class. The identification of pure pixels assumes that pure pixels are located on inner stroke regions, whereas mixed pixels are located at stroke boundaries. The pure pixels found are then used to identify mixed pixels: The binarization output of [SLT10] (S) is again analyzed. Mixed pixels are assumed to be located besides pure pixels and the CC's in S which contain at least one pixel that is labeled as pure pixel are analyzed: The class which possess the highest amount of pixels within these CC's (except for the pure pixels class) is assumed

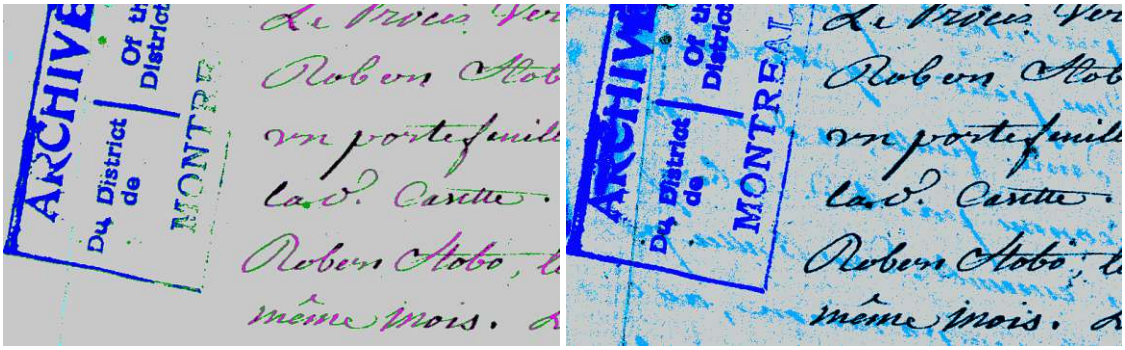


Figure 4.13: GMM clustering. (Left) Output of the first clustering step. (Right) Result of the second clustering stage.

to describe mixed pixels. These mixed pixels are rejected for the initialization of the second clustering step.

Figure 4.13 shows outputs of the first and second clustering steps. It is notable that handwriting is modeled by three components in the first clustering result. Contrary, in the second clustering step, the handwriting is modeled by one component.

It was found that it is necessary to regularize the covariance matrix: This regularization is necessary to avoid ill-conditioned covariance matrices. In order to avoid singular and thus not invertible covariance matrices, a small regularization value is added to the diagonal of the covariance matrix. This regularization does not only avoid ill-conditioned matrices, but has also an effect on the GMM found. The dependency on the regularization term is analyzed in Section 4.4.4.

4.2.4 Combination of Segmentation Results

The outputs of both clustering steps are combined in a final step. The output of the first clustering step is less sensitive to background variations, compared to the second clustering result. Therefore, the first clustering result is used to eliminate background variations: The background is found by identifying the cluster, which has the largest numbers of samples assigned. The image is inverted and the resulting mask encodes the potential foreground.

The output of the second clustering step is used to determine the definite foreground regions. Therefore, the Gaussian that models the handwriting class is found by applying the strategy for the identification of pure pixels - as described above. This Gaussian is used to create a mask that encodes the handwriting. Both masks are afterwards multiplied and the resulting image is hereafter denoted by C.

The resulting image is then combined with the output of the binarization algorithm (S) in order to recover strokes which are belonging to mixed pixels, but are not contained

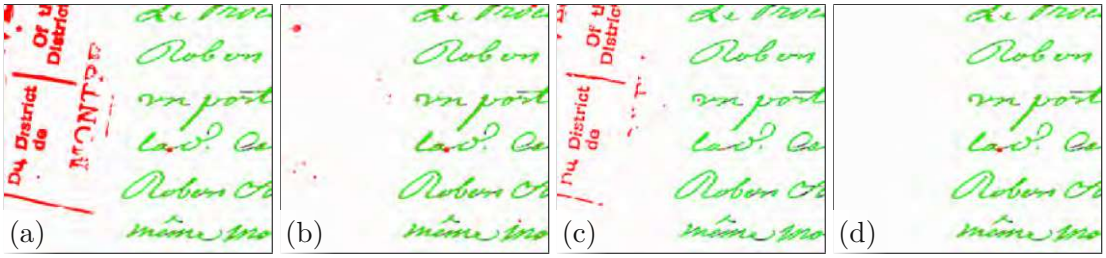


Figure 4.14: Different stages of the binarization framework. (a) Result of the first clustering step. (b) Result of the second clustering step. (c) Output of the binarization method in [SLT10]. (d) Final output.

in C. Therefore, the CC's in S are found and added to the resulting image, if they are connected with the CC's in C.

Figure 4.14 shows outputs of the different binarization steps. It can be seen that the final output in (d) contains less FN than the masks that are generated on the basis of the first (a) and second (b) clustering steps. Additionally, the number of FN is slightly decreased in the final output. This can be attributed to the circumstance that stroke endings that are present in the binarization result (c) and are connected with CC's of the multiplied masks are incorporated into the final result.

4.3 Deep Learning

Given the recent success of deep learning in the field of document image binarization [PZK+19a], we have applied a CNN for the binarization of multispectral document images in [HBS19]. Several authors have successfully applied U-Net [RFB15] inspired architectures for the task of document image binarization. The U-Net architecture is an improvement over the FCN architecture proposed by Long et al. [LSD15] for image segmentation. Long et al. [LSD15] popularized the use of FCN for semantic segmentation [OSK18]. Long et al. adapt classification networks - including AlexNet [KSH12] and VGG network [SZ15] - into FCN and "*transfer their learned representations by fine-tuning to the segmentation task*" [LSD15]. The network path that is composed of layers belonging to classification networks, is termed contracting path [RFB15] or also encoding path [OSK18]. In order to enable semantic segmentation, in-network upsampling is performed by using deconvolutional layers. The upsampling is performed in the so-called expanding [RFB15] or decoding [OSK18] path. Additionally, skip connections are introduced, which combine features from the contracting path with features from the expanding path. Thus, localization accuracy is increased by combining local information with global and semantic information [RFB15], [LSD15].

Ronneberger et al. [RFB15] extend the FCN approach of Long et al. by using a decoding path that is symmetric to the encoding path. Thus, the number of feature channels is

increased and context information is propagated to higher resolution layers [RFB15]. The network is termed U-Net, because the encoder and decoder paths are symmetric and form an U-shaped network architecture. The FCN in [LSD15] is used for semantic segmentation and is applied on the PASCAL VOC [EEG+14] dataset, which contains natural images. Contrary, the U-Net in [RFB15] is used for biomedical image segmentation, where only a few training samples are available [RFB15]. In order to compensate the limited amount of training data, Ronneberger et al. [RFB15] make use of data augmentation strategies, including shift, rotation and elastic transformation.

While Long et al. [LSD15] use pretrained networks - such as VGG16 - in the encoding path, Ronneberger et al. [RFB15] train their network from scratch. Iglovikov and Shvets show that the performance of the U-Net can be increased by using the VGG11 [SZ15] architecture in the encoding path with weights that are pretrained on the ImageNet [RDS+14] dataset. Oliveira et al. [OSK18] et al. propose to use the ResNet50 [HZRS16] architecture in the contracting path with ImageNet [RDS+14] pretrained weights. The network is used for document segmentation tasks, including page extraction, baseline extraction and layout analysis. We have applied this network for the binarization of multispectral document images and its architecture is introduced in the following section.

4.3.1 Architecture

Oliveira et al. [OSK18] use residual layers in the encoding path. Such residual layers are introduced by He et al. and the corresponding architecture is named ResNet. This architecture enables the training of deeper[2] neural networks. He et al. [HZRS16] show that the performance of previous networks (such as VGG) cannot be increased by adding more layers, because the training error increases at a certain number of layers. The authors propose to learn residual functions in order to successfully train deeper neural networks.

The overall aim of residual learning is to learn the residual function $\mathcal{F}(x) := \mathcal{H}(x) - x$ instead of the underlying function $\mathcal{H}(x)$ of certain layers. Thus, the original function becomes $\mathcal{F}(x) + x$. He et al. [HZRS16] show that it is more efficient to optimize the residual function than to optimize the original function. The learning of residuals is illustrated in Figure 4.15 (left). The residual learning is illustrated by the paths in the middle. The right path depicts the identity mapping of x and is named shortcut connection. Figure 4.15 (right) shows the residual learning in the case of deeper nets that are learnt on the ImageNet [RDS+14] dataset. The block is called bottleneck block and is used because of computational considerations. It is shown in [HZRS16] that the training and test errors are reduced by residual learning approach, compared to previous architectures.

Oliveira et al. [OSK18] use a residual net with 50 layers (ResNet50) in the encoding path of their FCN architecture. The network architecture is illustrated in Figure 4.16.

[2]Deeper networks in the work of He et al. [HZRS16] have around 100 - 150 layers or even more than 1000 layers in an extreme case.

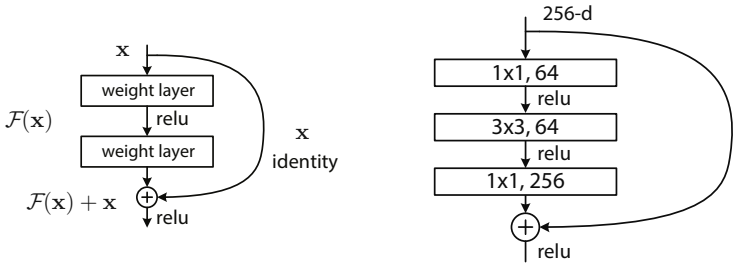


Figure 4.15: (Left) Residual block. (Right) Bottleneck block as used in ResNet50.

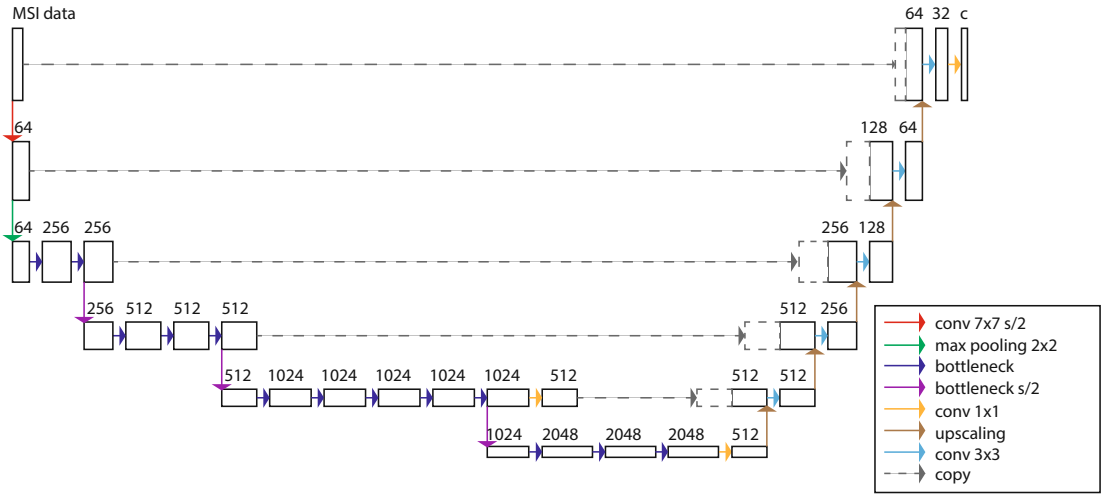


Figure 4.16: Network architecture. Adapted from [OSK18].

For comprehensibility, the bottleneck blocks are solely illustrated by red arrows. In the decoding path, the length of the feature maps is doubled and the channel number is halved. In the final step, the output image has the same size as the input image. Skip connections are depicted with dashed and gray lines. The output image is a grayscale intensity image, encoding the probability predictions of the network. The overall binarization output is obtained by applying a global threshold of 0.5. The output of the network is not further processed, since the aim of this work is to solely analyze the performance of the network.

The weights in the first layer of the network in [OSK18] are pretrained on RGB images, whereas the multispectral images contain more channels. Hence, these weights are simply adapted in order to make the network applicable on multispectral images: The weights corresponding to the RGB layers are simply repeated until the number of channels is equal to the number of multispectral channels. It was found empirically, that the training is not sensitive on the initialization of these weights. For example, a random initialization led to a network that is performing similarly.

The models have been trained using Adam [KB15] for 300 epochs. For each model, the training set was split in 80% train images and 20% validation images. The loss function used is Cross Entropy. Training a model on the MSBin dataset took approximately 12 hours on a NVIDIA TITAN X GPU.

4.3.2 Preprocessing

The dynamic ranges of the MS-TEx images are highly varying. This variation impedes the supervised binarization, because the spectral signatures of the foreground and background samples in different multispectral images are also varying. In order to compensate for this variation, the images have been preprocessed by normalizing the dynamic ranges of each channel separately. The influence of this simple preprocessing step is analyzed in the numerical evaluation that is conducted in the next section.

4.4 Experiments and Results

The binarization performance is evaluated on two different datasets: The MS-TEx and MSBin datasets. The former dataset was published in the course of the ICDAR 2015 MS-TEx contest and contains 31 images of handwritings originating between the 17th and 20th centuries. This dataset allows for a comparison of the developed methods with other binarization methods designed for multispectral document images. The latter dataset contains 120 portions of multispectral images of documents originating between the 11th and 12th centuries. The MSBin dataset is larger compared to the MS-TEx dataset. Thus it should be more appropriate for the application of deep learning-based method, which require relatively large training datasets [ZPIE17]. The datasets are briefly introduced in the following.

MS-TEx Hedjam et al. [HNM+15] have published the MS-TEx dataset in the course of the MS-TEx contest. The participants developed their methods based on a training set, which contains 21 multispectral images. The test set was made publicly available after the competition took place and it is comprised of 10 multispectral document images. Each multispectral image has 8 different channels ranging from 340 nm until 1100 nm.

The documents originate between the 17th and 20th century and the ink of the handwritings is iron gall based. Each multispectral image contains a text written with iron gall based ink. Additionally, the documents contain partially annotations, stamps or degradations [HNM+15]. These classes are not labeled in the ground truth images. Instead, the ground truth reference images encode solely the background and foreground classes. Figure 4.17 shows three different channels of a multispectral document image and the corresponding ground truth image. Further examples of the dataset are given in Figure 4.18.

MSBin The MSBin dataset is published in [HBS19]. It contains 80 training and 40 test images. The multispectral images consist of 12 different channels ranging from 365

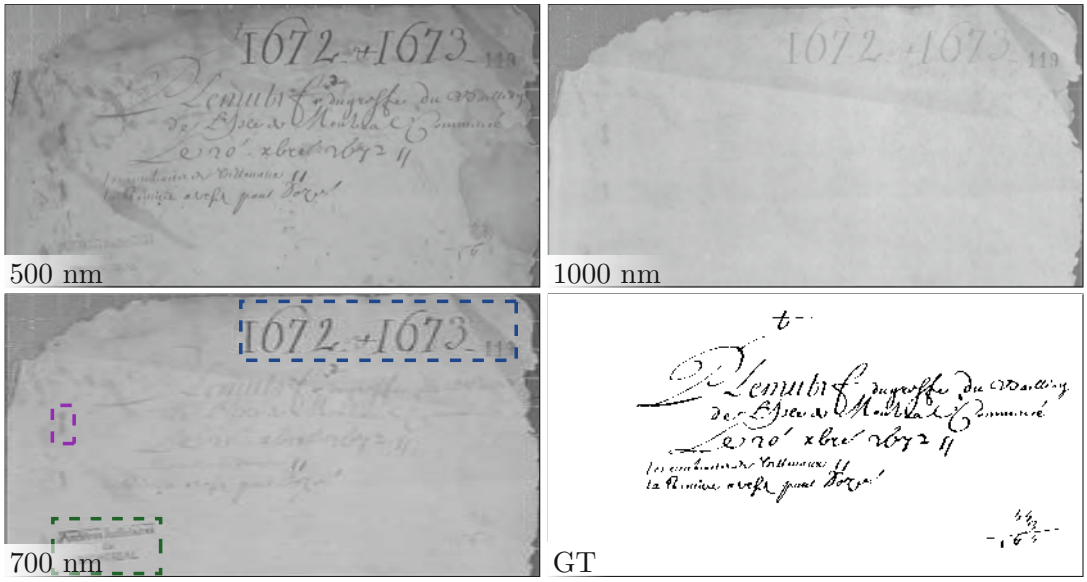


Figure 4.17: Example of the MS-TEx dataset. The dashed rectangles in the 700 nm image indicate different classes that are not annotated: Blue depicts an annotation, pink depicts a degradation, green depicts a stamp. Based on a courtesy of Hedjam et al. [HNM+15].

nm until 940 nm. The illumination used is the same as the one that was described in Section 3.1.2. The images have been acquired with a Phase One IQ260 achromatic camera that has a resolution of 60 MP. The average resolution of the acquired images is 550 Dots per Inch (dpi). For each multispectral channel a fixed exposure time was used which maximized the dynamic range within the channel. The exposure times were not changed during the acquisition. The images are taken from different folios belonging to two medieval manuscripts.

The writings in the manuscripts are written with two similar inks: The majority of the written text is written with dark - iron gall based - ink. This ink is hereafter denoted as FG 1. Additionally, a subset of the images contains characters written with red ink, denoted as FG 2. An X-Ray fluorescence investigation revealed that this red ink consists of mercury, lead and small amounts of iron. Both different inks are annotated in the ground truth images. The train set contains 30 images with FG 2 and 15 test images contain FG 2.

The test set contains additionally regions, where the human annotators were not able to clearly differentiate between foreground and background. These regions are also labeled in the ground truth as uncertain regions. For the evaluation these regions are treated as background regions, in the ground truth images as well as in the resulting images. It should be noted that only the test set contains such uncertain regions. Uncertain

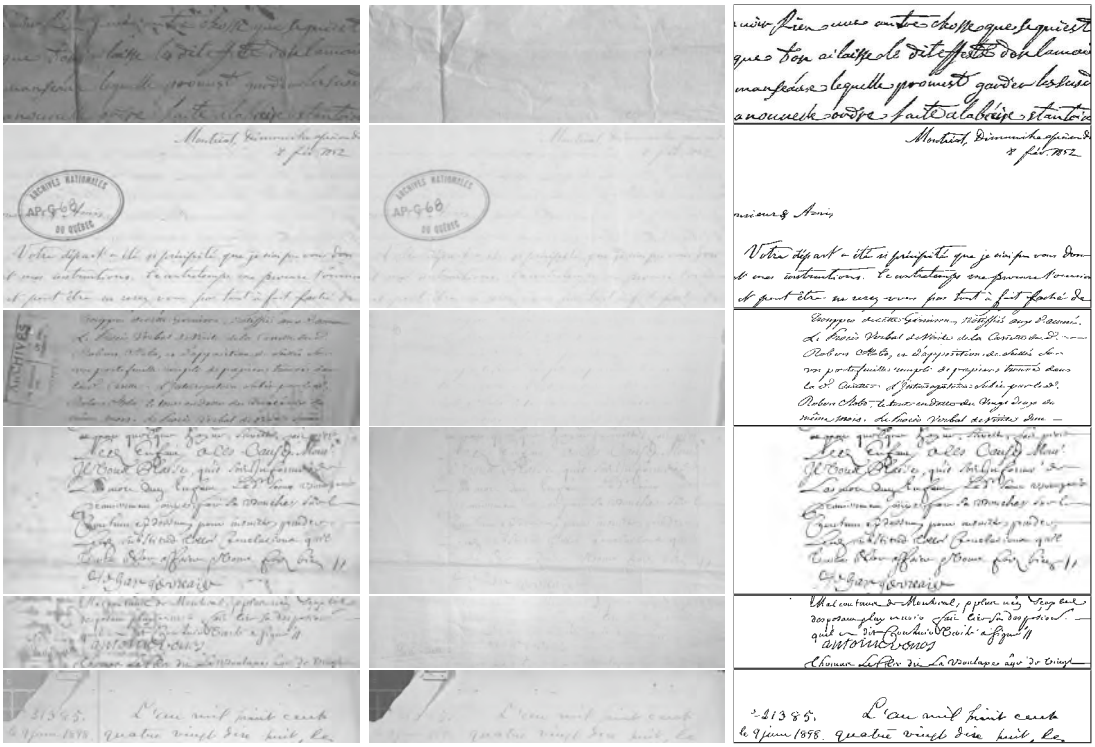


Figure 4.18: Examples of the MS-TEx dataset. (Left) Image acquired at 500 nm. (Middle) Image acquired at 1000 nm. (Right) Ground truth image.

regions are not present in training images, in order to allow for a training on entire images. Figure 4.19 shows an example of the MSBin data set. Figure 4.20 contains further exemplar images belonging to the MSBin data set.

The ACE based methods and the GMM based binarization technique assume that the iron gall ink occupies the majority of the foreground pixels. This assumption is also valid for the MSBin dataset: For each image portion, the number of pixels belonging to FG 1 is greater than the number of pixels, which are belonging to FG 2. Since the three methods mentioned are designed for a single writing, only the foreground estimation of FG 1 is evaluated for these methods. Contrary, the U-Net based approach is capable of segmenting both classes and hence the binarization of both classes is evaluated.

4.4.1 Performance Measures

The methods are evaluated by means of multiple pixel-based metrics. The selected metrics are the same metrics that have been used in recent DIBCO series. These metrics are described in the following.

FM For the calculation of the FM, the pixels in the binarization result and the corresponding ground truth image are compared and classified into the following three

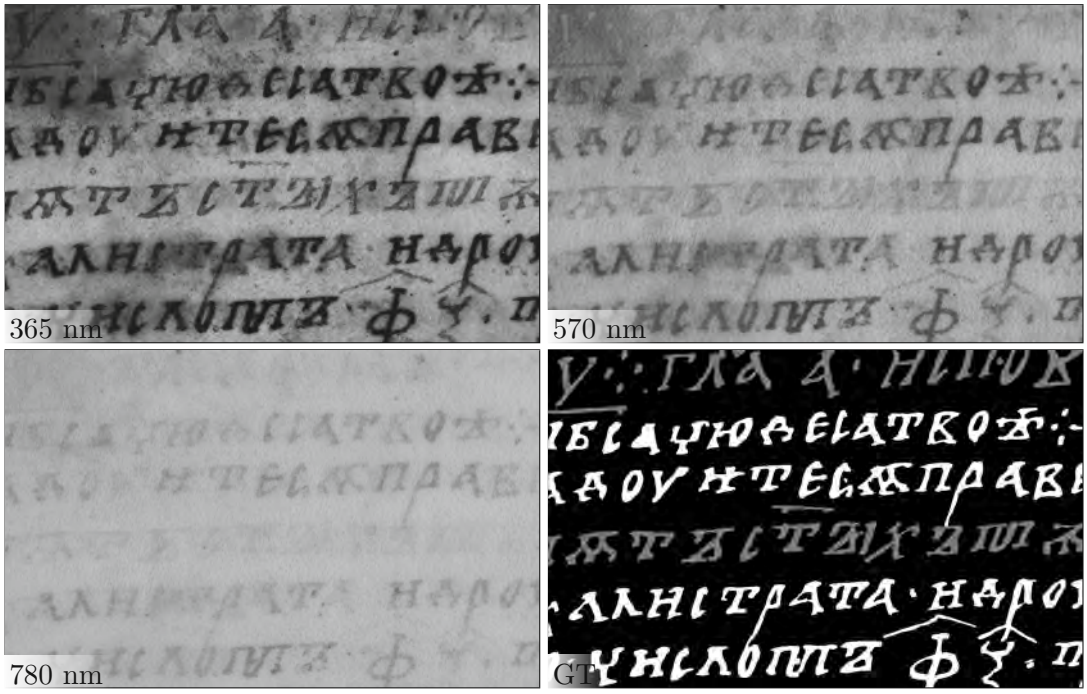


Figure 4.19: Example of the MSBin dataset. FG 1 is depicted by white and FG 2 is depicted by gray in the ground truth image.

groups:

- True Positives (TP) denotes pixels that are correctly labeled as belonging to the foreground.
- False Positives (FP) denotes pixels, which are wrongly classified as foreground pixels.
- FN denotes pixels that are wrongly labeled as belonging to the background.

The numbers of TP, FP and FN are counted and used to calculate the *Recall* and *Precision* values, which are defined by:

$$Recall = \frac{TP}{TP + FN} \tag{4.19}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.20}$$

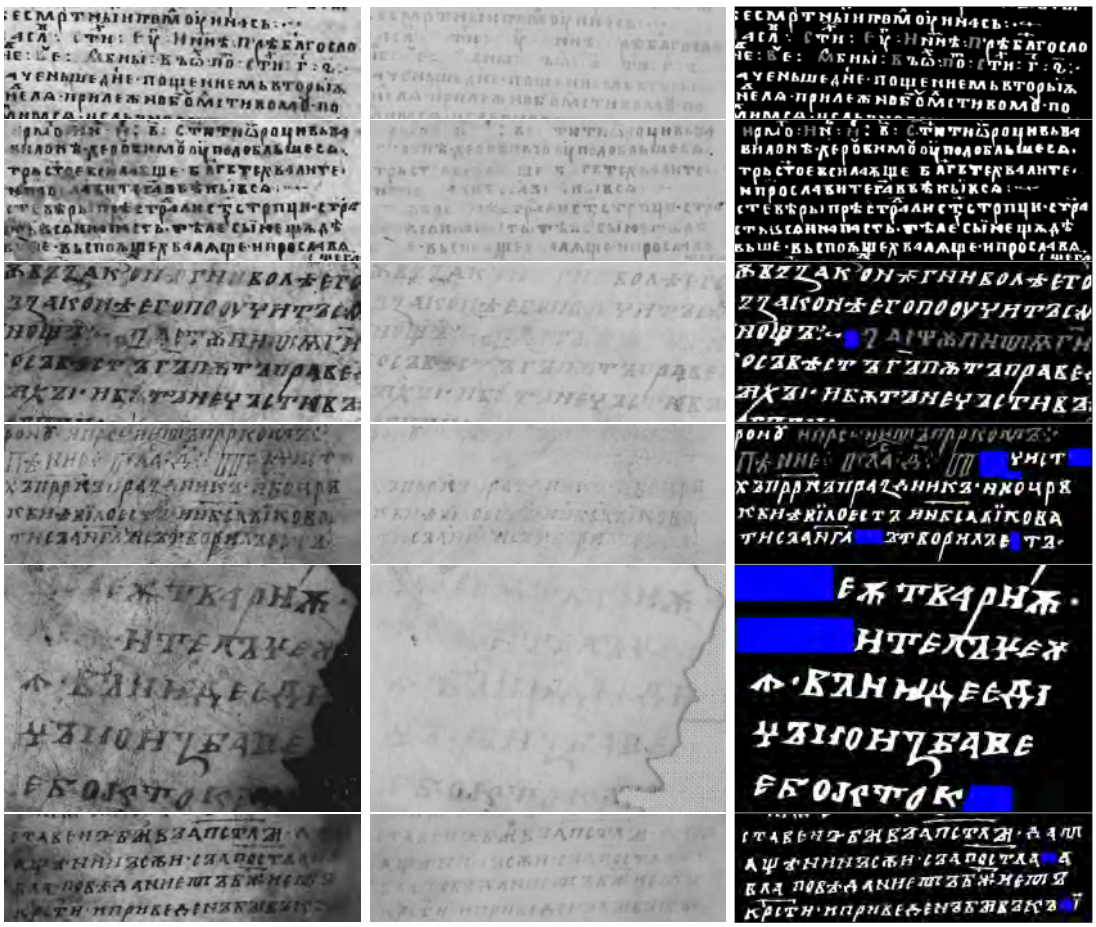


Figure 4.20: Examples of the MSBin dataset. (Left) Image acquired at 365 nm. (Middle) Image acquired at 780 nm. (Right) Ground truth image.

FM is the harmonic mean of *Recall* and *Precision*:

$$FM = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{4.21}$$

The FM performance measure has been used in every DIBCO and H-DIBCO series. According to [Ten19], it is used because the measurement penalizes binarization methods that make a disproportionate number of FP or FN.

PSNR PSNR measures the similarity between the ground truth image G and the corresponding binarization result B. It is formally defined by:

$$PSNR = 10log(\frac{C^2}{MSE}) \tag{4.22}$$

where C is the maximum pixel-wise difference between G and B. The Mean Squared Error (MSE) is given by:

$$MSE = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (G(x,y) - B(x,y))^2}{MN} \tag{4.23}$$

PSNR values are given in decibels (dB) and a high PSNR value indicates a good binarization performance. Similar to FM, PSNR is a point-based measurement and spatial relationships between pixels are not considered by the measurement. This circumstance is criticized in [NGP13] and [LKS04]. However, the PSNR metric has also been used in every DIBCO and H-DIBCO series.

DRD The DRD metric is especially designed for binary document images and is proposed by Lu et al. [LKS04]. The authors note that PSNR is not a suitable metric for binary document images, since it is a point-based metric, which does not consider mutual relations between pixels. Hence, it does not match well with the human visual perception [LKS04]. Lu et al. stress out that the distance between pixels should be considered, because it has an impact on the visual quality observed by humans. The proposed metric is named DRD metric and makes use of the reciprocal of distance to measure the distortion in binary document images [LKS04].

DRD is formally defined by:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \tag{4.24}$$

where $NUBN$ is the number of non-uniform 8×8 pixel blocks, k denotes the index of a flipped pixel and S denotes the total number of flipped pixels. A flipped pixel is either a false positive or a false negative pixel in the result image and is denoted with $B(x,y)$. DRD_k is the weighted sum of pixels in the ground truth image G within a $m \times m$ neighborhood that are different from the considered flipped pixel $B(x,y)$ in the result image. The DRD_k value is formally defined by:

$$DRD_k = \sum_{i,j} [\mathbf{D}_k(i,j) \times \mathbf{W}_{Nm}(i,j)] \tag{4.25}$$

where $i = j = 1, ..., m$ are the pixel coordinates within the considered $m \times m$ block, whereby $m = 5$. $\mathbf{D}_k(i,j)$ is the absolute difference between a pixel in the result image and a neighboring pixel in the ground truth image:

$$\mathbf{D}_k(i,j) = |G_k(i,j) - B(x,y)_k| \tag{4.26}$$

This difference $\mathbf{D}_k(i,j)$ is weighted by a normalized weighting matrix $\mathbf{W}_{Nm}(i,j)$:

$$\mathbf{W}_{Nm}(i,j) = \frac{\mathbf{W}_m(i,j)}{\sum_{i=1}^{m} \sum_{j=1}^{m} \mathbf{W}_m(i,j)} \tag{4.27}$$

The actual weights $\mathbf{W}_m(i,j)$ are indirect proportional to the (Euclidean) distance between the location (i,j) and the center of the block $(i_C, j_C), i_C = j_C = (m+1)/2$ and are defined by:

$$\mathbf{W}_m(i,j) = \begin{cases} 0, & \text{if } i = i_C \text{ and } j = j_C \\ \frac{1}{\sqrt{(i-i_C)^2+(j-j_C)^2}}, & \text{otherwise} \end{cases} \tag{4.28}$$

Thus, pixels that are closer to the considered flipped pixel $B(x,y)$ have a higher impact on the DRD_k value compared to pixels that have a larger distance to $B(x,y)$. Generally speaking, a lower DRD value indicates a better visual quality. Lu et al. present numerical results that show that DRD has a higher correlation with subjective assessments conducted by humans, compared to PSNR. The DRD metric has been used in every DIBCO series starting with DIBCO 2011.

p-FM Ntirogiannis et al. [NGP13] propose a modified FM metric named p-FM. The authors stress out that in the ordinary FM metric the location of FP and FN is neglected, although it has an impact on the readability. The authors note that text boundary localization can be ambiguous in document images, which can be mainly attributed to the document digitization process. Smith [Smi10] conducted a study, which was concerned with the manual generation of binarization ground truth: The human annotators reported that the location of text boundaries can be difficult and is a subjective matter.

In order to lower the influence of text boundary pixels, Ntirogiannis et al. [NGP13] propose a weighting scheme for recall and precision. The weighted Recall metric is named Pseudo-Recall (p-R) and is formally defined by:

$$p-R = \frac{\sum_{x,y} B(x,y) \cdot G_W(x,y)}{\sum_{x,y} G_W(x,y)} \tag{4.29}$$

where $G_W(x,y)$ denotes the weighted ground truth image. The weights in $G_W(x,y)$ are zero at stroke boundaries, while the maximum value is assigned to pixels that are located at the skeleton of the ground truth image. Thus, FN located at inner stroke regions are penalized more than FN that are located at outer stroke regions.

The weighted Precision measure is named Pseudo-Precision (p-P) and is defined by:

$$p-P = \frac{\sum_{x,y} G(x,y) \cdot B(x,y) \cdot P_W(x,y)}{\sum_{x,y} B(x,y) \cdot P_W(x,y)} \tag{4.30}$$

where $P_W(x,y)$ denotes a weight map. The weights used are dependent on the distance to the character contours and on the local stroke width:

FP that are located at the contour without seriously affecting a character topology are less penalized than FP that are altering the character topography. Figure 4.21 (a) shows

an example how FP can affect the topology of a character. The topology of the left character in Figure 4.21 (a) is only slightly modified, whereas the topology of the right character is more altered. The number of FP is the same for both binarization results and hence the corresponding FM is equal. Contrary, the p-FM considers the topography and assigns a higher value to the left result in Figure 4.21.

Additionally, p-FM penalizes FP that are connecting adjacent characters more compared to FP that are not in the near of characters, because the latter mentioned do not impede the readability. Figure 4.21 (b) shows examples for these two kinds of FP: The FP region in the middle has a higher impact on the p-P, compared to the FP region in the upper right image region.

(a) (b)

Figure 4.21: Examples of FP. TP are colored black and FP are colored red. Adapted from [NGP08].

The interested reader is referred to [NGP13] for a detailed explanation of the computation of $P_W(x, y)$ and $G_W(x, y)$.

The p-FM is finally defined as the harmonic mean of p-P and p-R:

$$p - FM = \frac{2 \times p - R \times p - P}{p - R + p - P} \tag{4.31}$$

Ntirogiannis et al. [NGP13] show exemplar images together with numerical results that indicate that the p-FM is better suited for measuring binarization performance, compared to FM, DRD and PSNR. Additionally, the metrics mentioned are compared with OCR results and it is shown that the p-FM metric yields the highest correlation with the OCR results. The p-FM metric has been used in all DIBCO series starting with the DIBCO 2013.

In the following section the four binarization methods are evaluated independently: First, the parameters that are crucial for the performance are evaluated on the training sets of the MS-TEx and MSBin dataset. The best parameter sets found are then used in the evaluation on the training and test sets.

The parameter evaluation is performed by means of FM. The binarization results that are gained by the best parameter set are afterwards evaluated by means of FM, p-FM, DRD and PSNR. Since the parameter selection is based on the FM metric it is also

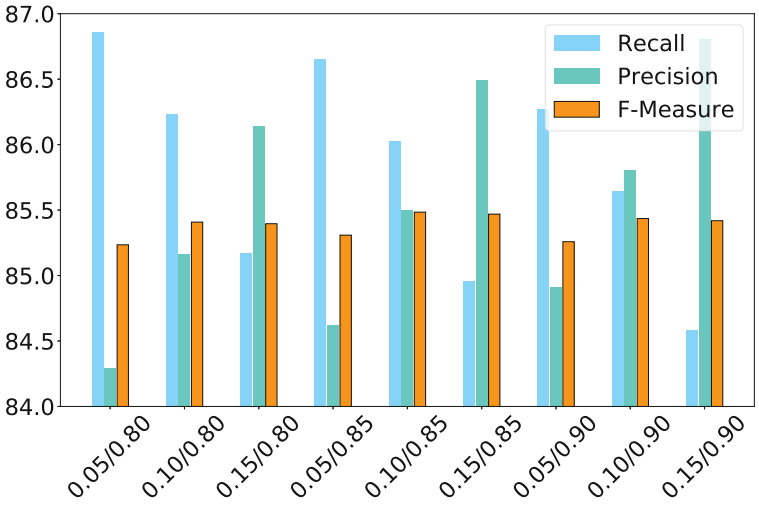


Figure 4.22: Performance gained on the MS-TEx training database. The different scores have been achieved by using varying combinations of t_{low} and t_{high}, which are provided on the x-axes.

optimized with respect to FM. Using for instance the p-FM for the parameter selection would probably lead to a higher overall p-FM. However, the p-FM was not used for the parameter selection because its computation is very time consuming: The computation for the MSBin training set takes approximately 1.5 hours[3].

4.4.2 Evaluation of the Base ACE Method

The ACE base method is first evaluated on the MS-TEx dataset. Afterwards the binarization method is evaluated on the MSBin dataset.

Parameter Evaluation - MS-TEx First, the performance of the base method is analyzed. The performance dependence on the parameters t_{low} and t_{high} (see Equation 4.9) was evaluated. The performance gained on the MS-TEx training dataset is shown in Figure 4.22. Recall, precision and F-Scores for nine different parameter sets are depicted by the bar plot. The highest FM values - namely 85.48% - is gained by using a parameter set of $t_{low} = 0.1$ and $t_{high} = 0.85$. It can be seen that by increasing the t_{low} value, the recall is decreased, which can be attributed to the increased number of false negatives. Contrary, by increasing the t_{high} value, the number of false positives is decreased, which results in an increased precision.

[3]The p-FM is computed with the official executable provided by the DIBCO organizers.

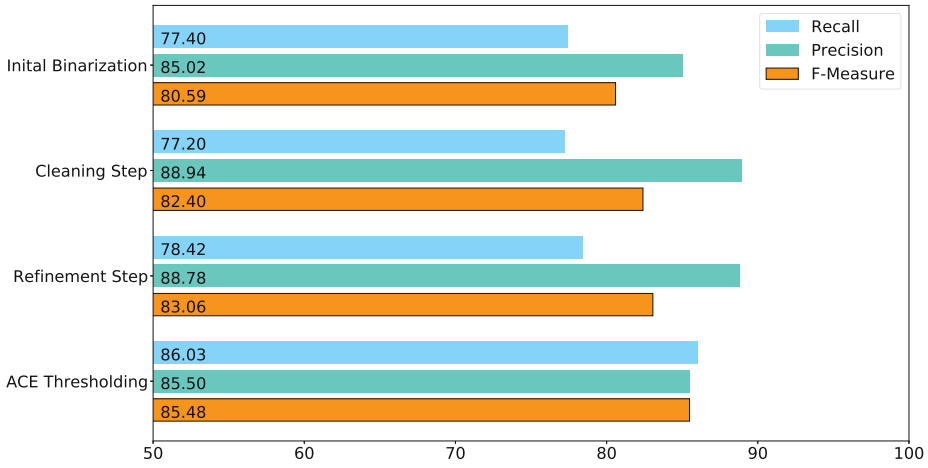


Figure 4.23: Evaluation of different stages on the MS-TEx training set.

Stages Evaluation - MS-TEx The evaluated thresholding step is the last step in the binarization method, which consists of four stages (see Section 4.1.1). Figure 4.23 shows the recall, precision and F-Score values that are gained by these four different steps. The initial binarization depicts the application of the method of Su et al. [SLT10] on a single channel of the multispectral image (taken at 500 nm). This grayscale binarization approach gains an FM of 80.59%. The ensuing cleaning step is used for the removal of false positives: Thus, the precision is increased, whereas the recall is only slightly decreased. This leads to an increased FM of 82.40%. The subsequent refinement step recovers thin and elongated strokes that are not found in the initial binarization step. This step increases the recall by 0.66%. Finally, the thresholding step with the selected thresholds of $t_{low} = 0.1$ and $t_{high} = 0.85$ leads to an overall FM of 85.48%. Thus, the incorporation of spectral information leads to a performance increase of 4.89%, compared to the output of the initial binarization.

Example outputs gained by the four stages are shown in Figure 4.24. The image in (a) is used as an input image for the initial binarization (c), which is gained by applying the algorithm of Su et al. [SLT10]. This result is refined by incorporating the target detection result that is shown in (b). The final binarization result in (f) contains less false positives than the result of Su et al. [SLT10]. It should be noted that the number '1677' in the upper image half in (f) is falsely labeled as belonging to the foreground. The target detection image exhibits that the spectral signature of the number is different compared to spectral signature of the remaining text. However, the method is only partially capable of identifying these false positives in the initial binarization result.

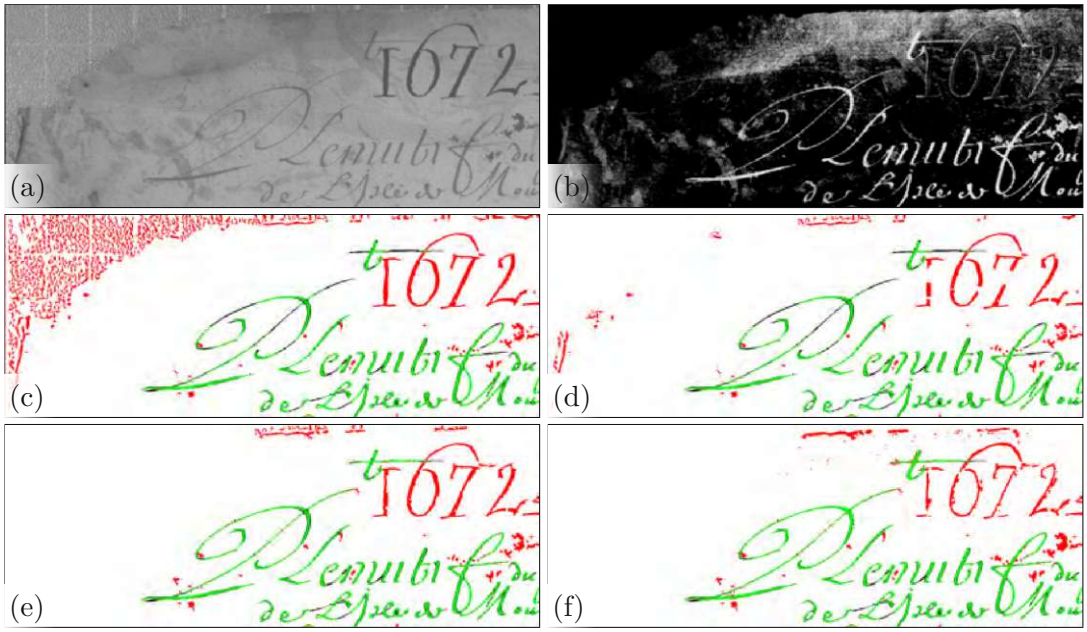


Figure 4.24: Binarization of an image belonging to the MS-TEx training data set. (a) Image acquired at 500 nm. (b) ACE target detection image. (c) Initial binarization result. (d) Cleaning step result. (e) Refinement step result. (f) ACE thresholding result.

DIBCO Metrics - MS-TEx The performances in terms of FM, p-FM, PSNR and DRD are given in Table 4.1. It can be seen that method performs worse on the test set.

	FM	p-FM	PSNR	DRD
Train	85.48	85.90	17.95	3.74
Test	82.24	83.69	17.19	4.81

Table 4.1: Performance on the MS-TEx set.

Parameter Evaluation - MSBin The parameter evaluation of the two thresholds is also carried out on the MSBin training dataset. The performance in terms of FM is shown in Figure 4.25. The highest F-Score, namely 87.65%, is gained by using a parameter set of $t_{low} = 0$ and $t_{high} = 0.45$. Using a value for t_{low} that is greater than zero causes a performance decrease for every evaluated t_{high} value. This circumstance can also be seen in Figure 4.25, where recall, precision and F-Score values for different parameter sets are shown. Using a t_{low} value that is greater than zero increases the number of false positives and results in decreased recall values. It is notable that the best parameters for the MSBin dataset ($t_{low} = 0, t_{high} = 0.45$) are lower than the best thresholds for the MS-TEx dataset ($t_{low} = 0.1, t_{high} = 0.85$). This can be mainly attributed to the following two reasons: First, the images in the MSBin dataset have a higher spatial

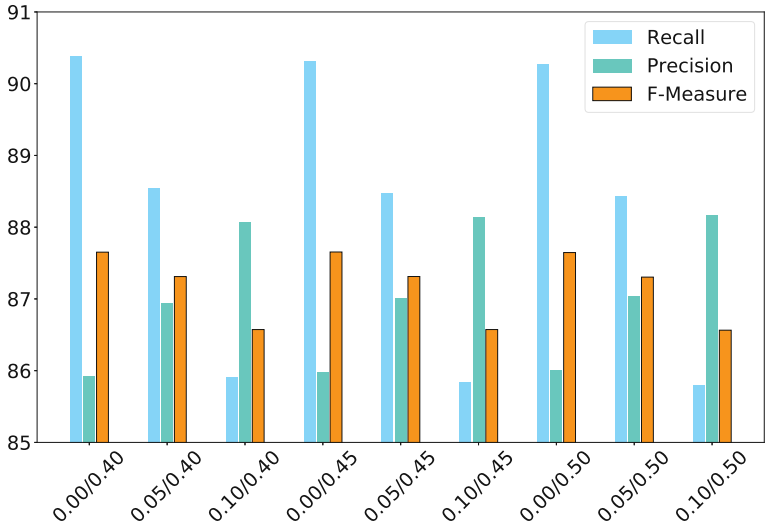


Figure 4.25: Performance gained on the MSBin training database. The different scores have been achieved by using varying combinations of t_{low} and t_{high}, which are provided on the x-axes.

resolution than the images in the MS-TEx dataset. Hence, the character strokes in the MSBin dataset have a larger stroke width and the pixel intensities within the strokes are more varying. Second, the MSBin dataset contains also more characters that are partially faded-out or interrupted. Due to these two reasons, the spectral signatures within the stroke regions are more varying in the MSBin dataset. Hence, the foreground regions within the ACE images are less homogeneous and the target detection values are decreased, compared to the MS-TEx dataset.

Stages Evaluation - MSBin The influence of the four different binarization stages is shown in Figure 4.26. The initial binarization of the grayscale images results in an average FM of 86.57%. The ensuing cleaning step reduces the number of false positives and leads to an increased FM of 87.69%. The subsequent refinement step slightly increases the recall, but leads to a decreased precision value. Thus, the average F-Score is reduced to 87.64%. The final ACE thresholding step has only a minor effect on the binarization performance, since the FM value of 87.65% is only slightly larger than in the previous step.

It is notable that only the cleaning step is capable of improving the binarization performance. The ensuing refinement step is designed for thin and elongated strokes that are not segmented by the initial binarization step. However, such strokes are less present in the MSBin dataset, compared to the MS-TEx dataset. Hence, the refinement step does not improve the performance. Instead the performance is slightly decreased.

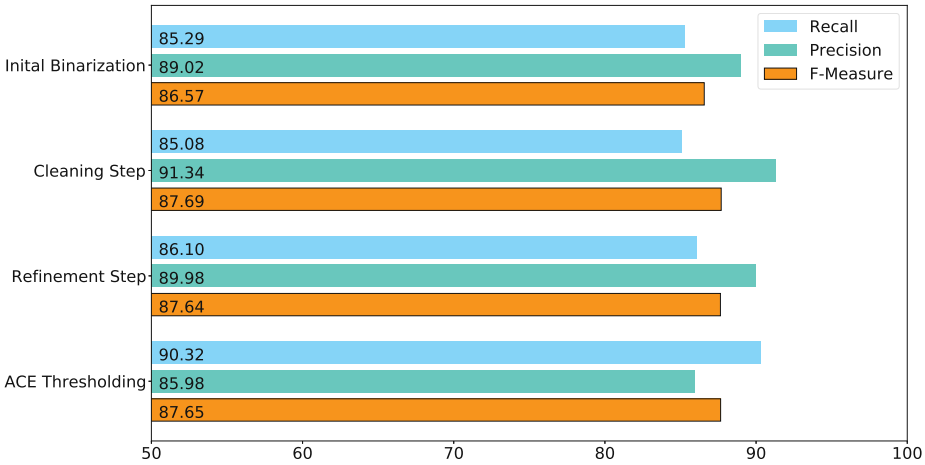


Figure 4.26: Performance gained on the MSBin training set by the different binarization steps.

The binarization method was designed for the MS-TEx dataset, whereas the MSBin dataset was not existing at this time. Since the last two stages of the method do not improve the F-Score, it has to be concluded that the method does not generalize well on the MSBin dataset. Nevertheless, the method is still capable of using spectral information in order to improve binarization results that are gained by the binarization method [SLT10], which is designed for grayscale images. The average performance increase in terms of FM is 1.08%.

Figure 4.27 shows an example binarization output gained on the MSBin training dataset. The input image for the initial binarization step is given in (a). The middle text line contains four characters that are written with red ink. The spectral signature of these characters is different than the spectral signature of the remaining characters, as can be seen in the target detection image (b). The characters written with red ink are falsely labeled as foreground pixels in the initial binarization step (shown in (c)). The majority of these false positives are removed in the cleaning step (d).

DIBCO Metrics - MSBin The performances in terms of FM, p-FM, PSNR and DRD are given in Table 4.2. It can be seen that for all four metrics, the performance on the test set is significantly lower, compared to the training set. This performance drop is also notable in the results that are gained by the remaining methods.

The decrease of the numerical results indicates that the images in the test set are at least partially different from the training set: The test set contains in fact more images which are strongly degraded by faded-out ink, bleeding artifacts and background variations. It

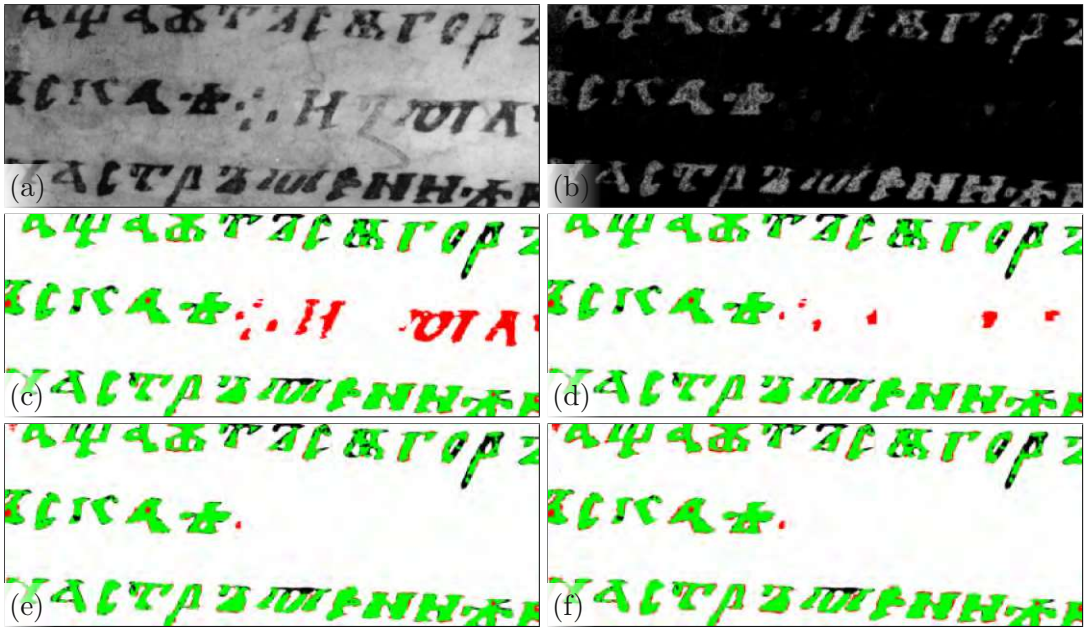


Figure 4.27: Binarization of an image belonging to the MSBin training data set. (a) UV fluorescence image acquired at 365 nm. (b) ACE target detection image. (c) Initial binarization result. (d) Cleaning step result. (e) Refinement step result. (f) ACE thresholding result.

was described above that the test set contains image regions, which are excluded from the numerical evaluation because the human annotators were not capable of differentiating between foreground and background. The images which contain such regions are generally in a worse condition, compared to images that are not comprised of such regions. These unclear regions are not present in the training set in order to allow for a training on entire image patches. The training set was compiled with the aim to make it as large as possible in order to be suitable for deep learning-based approaches. Since the overall amount of candidate patches was limited, patches with unknown regions had to be included in the test set. Otherwise, the test set would have been too small for an exhaustive evaluation.

	FM	p-FM	PSNR	DRD
Train	87.66	87.75	14.27	13.04
Test	81.28	81.18	13.28	22.03

Table 4.2: Performance on the MSBin set.

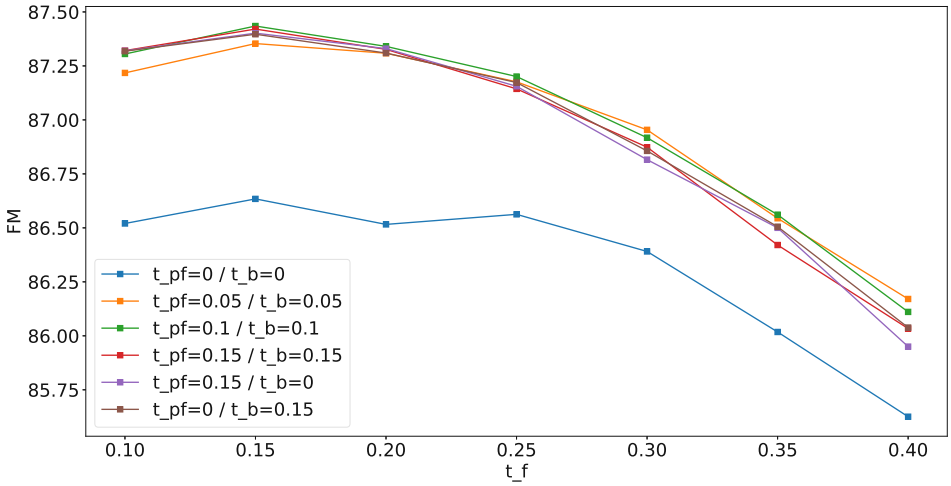


Figure 4.28: Parameter evaluation on the MSTEx dataset.

4.4.3 Evaluation of the GrabCut Based ACE Method

The second ACE based method is evaluated in the following. First, the evaluation is conducted on the MS-TEx dataset. The evaluation on the MSBin dataset follows afterwards.

Parameter Evaluation - MS-TEx The performance of the GrabCut based method relies mainly on the selection of the thresholds that are used for the mask initialization of the GrabCut mask. The influence of these parameters on the binarization performance is discussed in the following. These parameters are t_f, t_{pf} and t_b and they are used for labeling pixels as foreground, potential foreground and background pixels. The results for varying parameter combinations are shown in Figure 4.28[4].

The performance in terms of FM is given on the vertical axes. The horizontal axes exhibits increasing t_f values and the 6 graphs show combinations of different t_b and t_{pf} values. The worst performance - namely 85.63% is gained by using a parameter set of $t_f = 0.4$, $t_{pf} = 0$ and $t_b = 0$. The blue graph depicts a combination of $t_{pf} = 0$ and $t_b = 0$ with varying t_f values. The FM values that are depicted by this blue graph are significantly lower than the F-Scores gained by the remaining graphs, independent of the t_f values. This can be attributed to the fact that by using $t_{pf} = 0$ the more pixels are labeled as potential foreground, compared to higher t_{pf} values. This initialization results in a higher number of foreground pixels in the overall binarization result. The resulting images are over-segmented compared to binarization results that are gained by

[4]For comprehensibility, only selected resulting graphs are shown. This includes the best and worst performances gained.

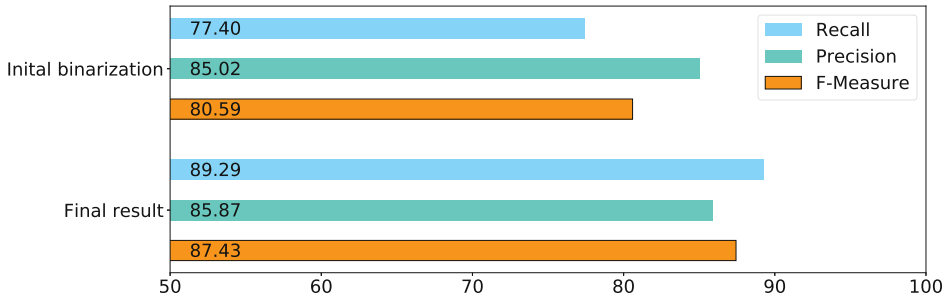


Figure 4.29: Performance gained on the MS-TEx training set.

an initialization with higher t_{pf} values. The best performance is 87.43%, which is gained by a combination of $t_f = 0.15$, $t_{pf} = 0.1$ and $t_b = 0.1$.

Figure 4.29 shows the performance in terms of recall, precision and F-Score of the best parameter combination. Additionally, the performance of the initial binarization step is shown. It can be seen that the overall FM is 87.43%, whereas the initial binarization gains an average F-Score of 80.59%.

DIBCO Metrics - MS-TEx The performances in terms of FM, p-FM, PSNR and DRD are given in Table 4.3. Similar to the base method, the GrabCut based method performs worse on the test set. It is also notable that the GrabCut based method gains performance values that are superior to the results gained by the base method. For example, the GrabCut based method gains an p-FM of 84.26% on the test set, whereas the base method gains an average p-FM of 83.69%.

	FM	p-FM	PSNR	DRD
Train	87.43	86.46	18.48	3.18
Test	84.61	84.26	17.62	4.33

Table 4.3: Performance on the MS-TEx set.

Parameter Evaluation - MSBin The parameter dependency is also evaluated on the MSBin training dataset and is shown in Figure 4.30. The blue graph in Figure 4.30 depicts parameters with varying t_f values, whereas t_{pf} and $t_b = 0$ are both set to zero. It is notable that the performances gained by these combinations are better than the remaining combinations. The highest F-Score, namely 87.54% is gained by a combination of $t_f = 0.1$, $t_{pf} = 0$ and $t_b = 0$. Compared to the MS-TEx dataset, the best parameters values are decreased. This can be attributed to the fact that the target detection values within foreground regions are lower in the MSBin dataset. Nevertheless, the application of the GrabCut algorithm increases the performance in terms of FM: Figure 4.31 shows the performance of the initial binarization step and the performance that is gained by

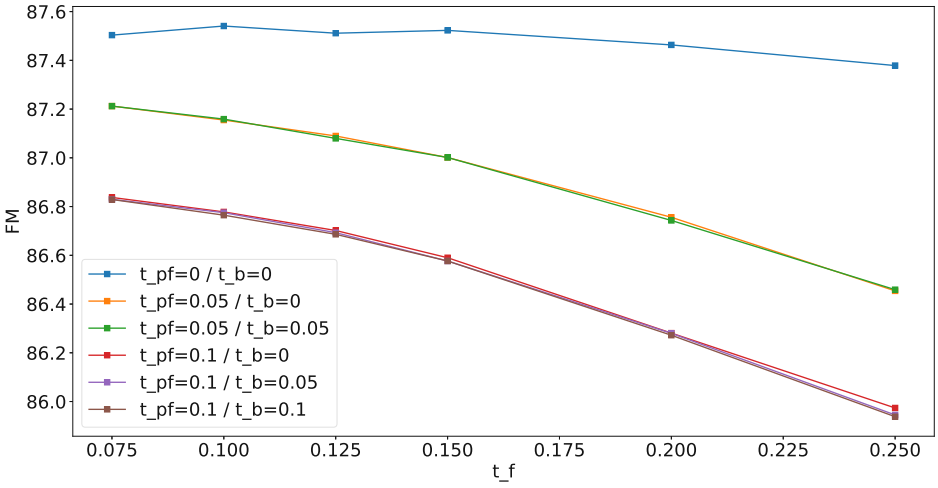


Figure 4.30: Parameter evaluation on the MSBin training dataset.

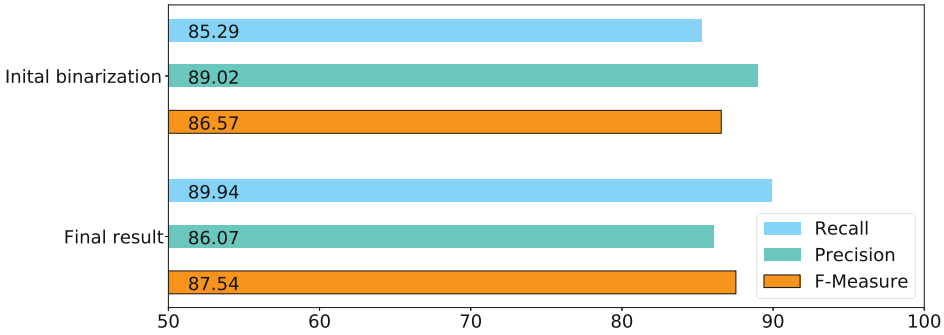


Figure 4.31: Performance gained on the MSBin training set.

applying the GrabCut segmentation. It is notable that the incorporation of the spectral information leads to a performance increase of approximately 1%.

DIBCO Metrics - MSBin The performance in terms of recent DIBCO metrics is given in Table 4.4. Similar to the base method, the performance is significantly worse on the test set. This can be again attributed to the circumstance that the test set contains more degraded panels than the train set. While the GrabCut based method clearly outperforms the base method on the MS-TEx dataset, the GrabCut based technique is not capable of achieving the expected superior performance on the MSBin data set. The results on the training set are even slightly inferior: For example, the GrabCut based method gains an FM of 87.54% on the training set, whereas the base method achieves an

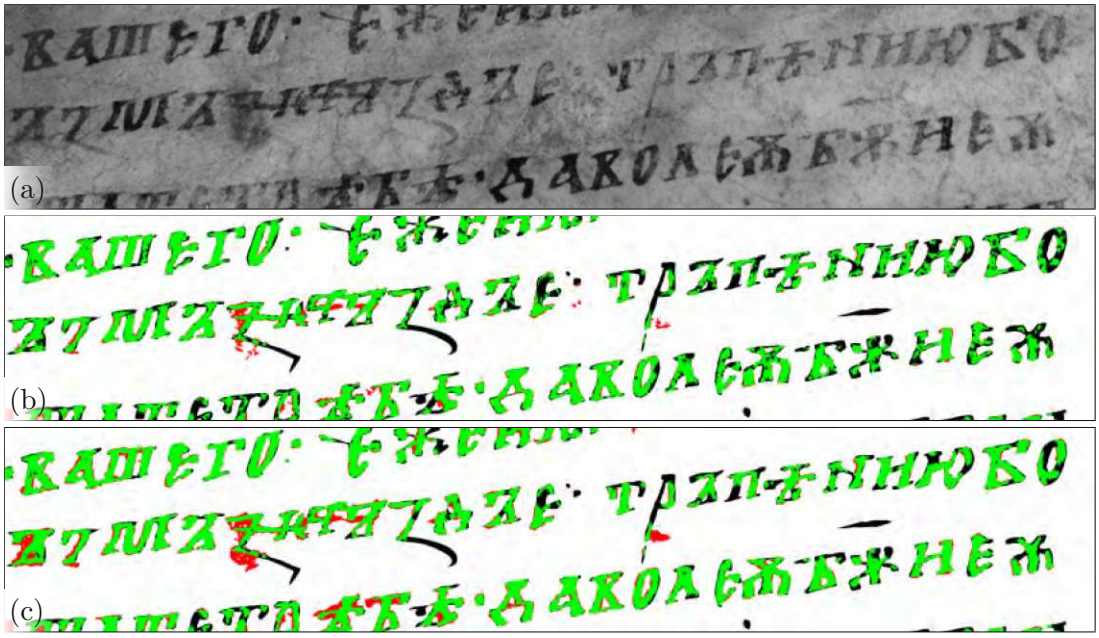


Figure 4.32: Influence of the GrabCut segmentation. (a) UV fluorescence image. (b) Initial segmentation gained by the method in [SLT10]. (c) GrabCut based binarization output.

FM of 87.66%. The results on the test set are relatively similar: The GrabCut based method gains an average FM of 81.25% and an p-FM of 81.36%. Contrary the base method achieves an FM of 81.28% and an p-FM of 81.16%.

The GrabCut based method is not capable of gaining the expected superior performance on the MSBin dataset. This can be mainly attributed to the following circumstance: The application of the GrabCut leads partially to a reinforcement of FP, which are stemming from background variations or from document boundaries. These degradations are less present in the MS-TEx dataset, for which the method was designed. The performance of the GrabCut based method could eventually be increased by using a different pseudo-color image as an input for the GrabCut mask. However, this step is out of the scope of the current work and is left for future work. One example for the reinforcement of background variations is shown in Figure 4.32. Note that the left image region in the GrabCut result (Figure 4.32 (c)) contains a larger amount of FP than the initial segmentation result (Figure 4.32 (b)).

Evaluation on DIBCO datasets Although the method is designed for MSI data, it is also evaluated on the datasets of H-DIBCO 2016, H-DIBCO 2018, DIBCO 2017 and DIBCO 2019. This evaluation shows the performance that is achieved on RGB images. Table 4.5 summarizes the results gained. For each competition, the winning method is

	FM	p-FM	PSNR	DRD
Train	87.54	87.83	14.35	13.53
Test	81.25	81.36	13.27	20.80

Table 4.4: Performance on the MSBin set.

listed in the first row. Additionally, the method of Sauvola [SP00] is provided in the second row. This method serves as a baseline method in the DIBCO contests. The ACE based method is listed in the third row. The method proposed performs worst on the DIBCO 2019 average dataset. This dataset can be divided into two groups: Track A and Track B. The performance on the latter set is significantly lower compared to the performance gained on Track A. Track B is comprised of papyri, which are in varying conditions. This dataset is more challenging compared to the remaining datasets, which results in a decreased performance of the method proposed.

Author	FM	p-FM	PSNR	DRD
H-DIBCO 2016				
Kligler and Tal [KKT18]	87.61	**91.28**	18.11	5.21
Sauvola [SP00]	82.52	86.85	16.42	7.49
Proposed	**88.84**	90.39	**18.42**	**4.01**
DIBCO 2017				
[BIN19]	**91.04**	**92.86**	**18.28**	**3.40**
Sauvola [SP00]	77.11	84.10	14.25	8.85
Proposed	87.50	90.06	16.63	4.99
H-DIBCO 2018				
Xiong et al. [XJX+18]	**88.34**	**90.24**	**19.11**	**4.92**
Sauvola [SP00]	67.81	74.08	13.78	17.69
Proposed	75.41	77.90	14.93	12.47
DIBCO 2019 - Average				
Bera et al. [BGB+21]	**72.86**	**72.15**	**14.48**	**16.26**
Sauvola [SP00]	42.52	39.76	7.71	112.40
Proposed	68.29	64.88	13.27	27.76
DIBCO 2019 - Track A				
Bera et al. [BGB+21]	**77.76**	**76.42**	**16.81**	**5.60**
Sauvola [SP00]	61.70	58.78	12.64	15.44
Proposed	75.59	71.33	15.84	6.75
DIBCO 2019 - Track B				
Dang et al. (unpublished)	**76.41**	**77.66**	**15.27**	**11.16**
Sauvola [SP00]	23.33	20.74	2.78	209.36
Proposed	61.00	58.43	10.70	48.77

Table 4.5: Results on the last two DIBCO and H-DIBCO datasets.

Figure 4.33: Input images and corresponding ground truth images. From top to bottom row: DIBCO 2009, H-DIBCO 2018, DIBCO 2019, DIBCO 2019 papyri dataset.

Figure 4.33 shows exemplar input images and the corresponding results gained by the proposed method. The third row shows a limitation of the method: The input image contains a green underline, whereas the main handwriting is written with black ink. The method proposed wrongly classifies the underline as belonging to the background, because of the different ink type. The fourth row shows that the method is sensitive to background variations contained in the DIBCO 2019 papyri dataset (Track B).

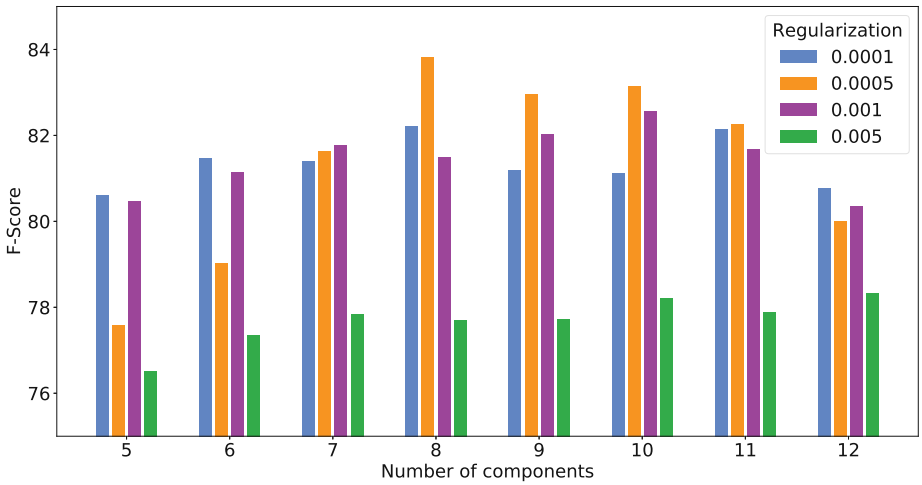


Figure 4.34: Parameter evaluation on the MS-TEx dataset.

4.4.4 Evaluation of the GMM Method

This section contains an evaluation of the GMM based method. First, the parameter dependency is evaluated. Afterwards, it is evaluated how the method performs on preprocessed and unprocessed multispectral images. Additionally, it is evaluated if the GMM based clustering can be successfully replaced by a k-means clustering step. Similar to the evaluation provided above, the experiments are first conducted on the MS-TEx dataset. Afterwards, an evaluation on the MSBin dataset is given.

Parameter Evaluation - MS-TEx The method is mainly dependent on the number of Gaussians and on the regularization value. This parameter dependency is first analyzed on the MS-TEx training set. Eight different number of components have been evaluated together in combination with four regularization values. A regularization value of zero is not evaluated, since the resulting covariance matrices are partially not invertible. The gained F-Scores are shown in Figure 4.34.

The highest FM value is 83.83%, which is gained by eight Gaussians and a regularization value of 0.0005. It is notable that models with five, six or twelve components gain a relatively poor performance, which is below 80%. Additionally, it can be seen that the performance is sensitive on the chosen regularization value. A regularization value of 0.005 leads to F-Scores that are below 80%, regardless of the number of components. In this case, the relatively large regularization value prevents an appropriate estimation of the Gaussian distributions.

Clustering Experiment - MS-TEx In this experiment, the influence of the background compensation (see Section 4.2.2) step is analyzed. Therefore, the binarization

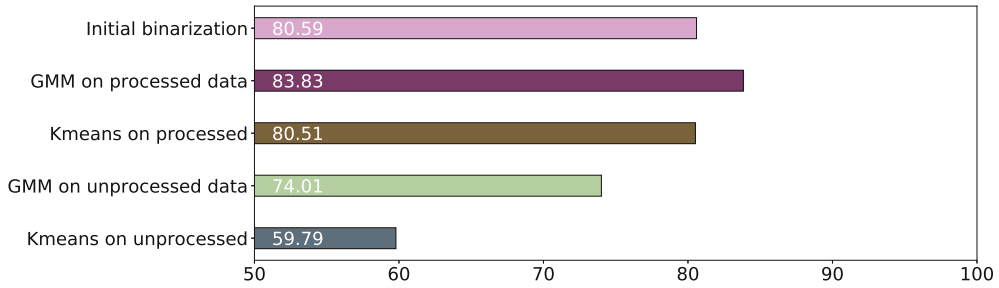


Figure 4.35: Dependency on clustering method and image processing.

method is applied on background compensated images as well as on unprocessed images. The resulting F-Scores are shown in Figure 4.35: The purple bar depicts the performance that is gained by applying the GMM based method on preprocessed images. The GMM is modeled by eight Gaussians with a regularization value of 0.0005 and gains an average F-Score of 83.83%. The same parameters have been used for the training of a GMM that is applied on the original MSI data. The FM gained by this GMM is significantly worse, namely 74.01%.

It was described in Section 4.2.3 that the GMM's are first initialized with parameters that are found by the k-means algorithm. It was also evaluated how the k-means method performs: Therefore, the GMM based clustering was replaced by a k-means clustering step. The best k-means results are gained by three clusters. The resulting performance measures are 80.51% and 59.79% for preprocessed and unprocessed images respectively. It can be concluded that the background compensation method has a strong impact on the clustering performance and that the GMM based clustering outperforms the k-means based clustering.

The pink bar depicts the F-Score that is gained by solely applying the binarization method of Su et al. [SLT10]. By combining this approach with the GMM based clustering the performance is increased by 3.24%.

DIBCO Metrics - MS-TEx Table 4.6 summarizes the performance in terms of DIBCO metrics. Similar to the ACE based methods, the performance drops on the test set: For example, the FM and p-FM values are reduced by 2.8% and 1.6% respectively.

	FM	p-FM	PSNR	DRD
Train	83.83	81.28	17.22	4.26
Test	81.75	79.68	16.63	5.04

Table 4.6: Performance on the MS-TEx set.

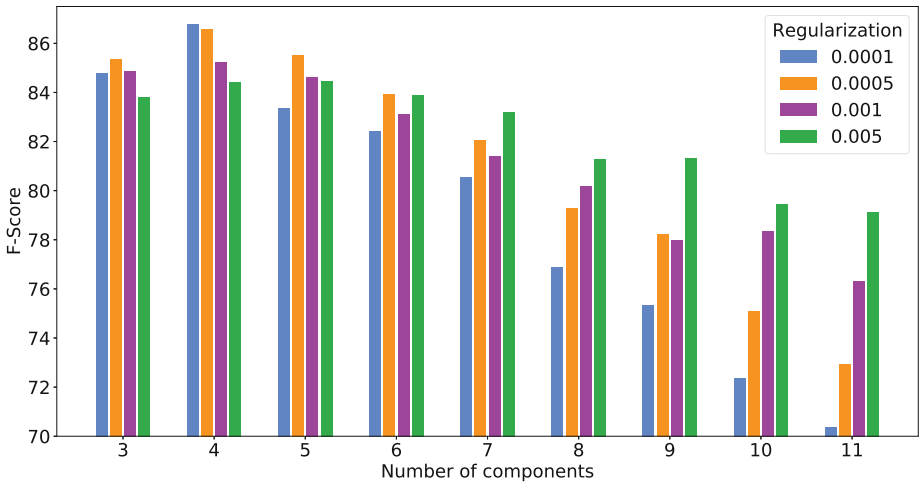


Figure 4.36: Parameter evaluation on the MSBin dataset.

Parameter Evaluation - MSBin The parameter evaluation is also carried out on the MSBin training dataset. Figure 4.34 shows the F-Scores that are gained by varying combinations of regularization values and number of Gaussians. The best performance is gained by using four Gaussians with a regularization value of 0.0001. The corresponding performance in terms of FM is 86.79%. The best performance is gained by four Gaussians, whereas the best performance on the MS-TEx dataset was gained by eight Gaussians.

Figure 4.37 shows two clustering outputs that are gained by using different numbers of Gaussians: The image in Figure 4.37 (c) is the result of using four Gaussian. The foreground is modeled by two different Gaussians. The method is designed for dealing with two different clusters that are resulting from *pure* and *mixed* foreground signatures. The *pure* class in Figure 4.37 (c) is depicted by yellow and the *mixed* class is depicted by red. The corresponding binarization result is shown in Figure 4.37 (b). Figure 4.37 (d) shows the clustering output that is gained by using eight Gaussians. It is notable that the foreground and background are modeled by multiple Gaussians. This impedes an adequate binarization and the resulting F-Score is significantly lower when using eight Gaussians, compared to using four clusters.

Clustering Experiment - MSBin Figure 4.38 shows several clustering results: The best performance (86.79%) is gained on background compensated images by the GMM based method. The method is also applied on unprocessed images, whereby the same parameters are used as for the processed data (four Gaussians and a regularization value of 0.0001). This leads to a decreased F-Score of 84.88%. The GMM based clustering was also replaced by a k-means clustering step. The k-means clustering leads also to reduced F-Scores of 84.72% and 84.00% for preprocessed and unprocessed MSI data.

Figure 4.37: Binarization of an image belonging to the MSBin test set. (a) White light image. (b) Binarization output gained by using four Gaussians. (c) GMM output by using four Gaussians. (d) GMM output by using four Gaussians.

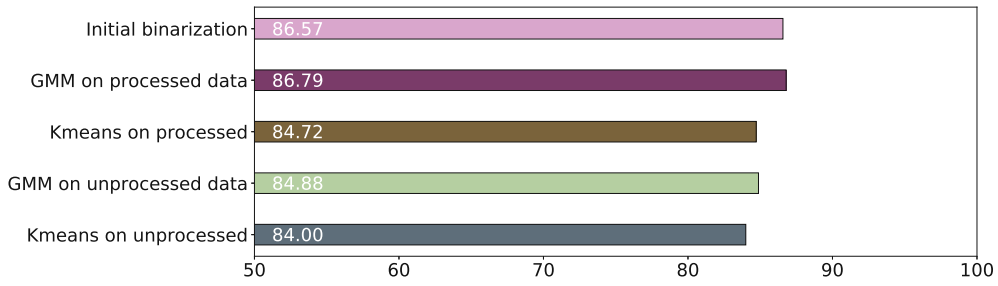


Figure 4.38: Dependency on clustering method and preprocessing.

Figure 4.38 shows also the performance that is gained by solely applying the initial binarization method of Su et al. [SLT10]. This method gains an average F-Score of 86.57%, whereas the entire method proposed gains an FM of 86.79%. Thus, the performance is only slightly increased by 0.21%. Contrary, the performance gain on the MS-TEx training set is 3.24%. The GMM based method was designed for the MS-TEx dataset and could not reproduce the performance increase on the MSBin training set.

DIBCO Metrics - MSBin The performances in terms of FM, p-FM, PSNR and DRD are summarized in Table 4.7. It is notable that method gains FM and p-FM scores that are approximately 6% worse on the test set, compared to the training set.

	FM	p-FM	PSNR	DRD
Train	86.80	86.94	14.05	13.39
Test	80.00	80.28	13.18	20.35

Table 4.7: Performance on the MSBin set.

Net	Train	Test
Original data		
U-Net	58.8	42.6
ResNet50	79.1	76.6
Normalized dynamic range		
U-Net	81.3	78.5
ResNet50	**86.6**	**78.9**

Table 4.8: Average F-Measures gained on the original MS-TEx dataset and on the preprocessed dataset.

4.4.5 Evaluation of the FCN

The performance of the FCN is evaluated in the following. First, the influence of the preprocessing step is evaluated on the MS-TEx dataset. Afterwards, it is analyzed if the preprocessing step is also effective on the MSBin dataset. Finally, two further experiments are conducted on the MSBin dataset.

Evaluation of the Preprocessing - MS-TEx In the first experiment, the preprocessing step is evaluated on the MS-TEx dataset. The FCN is applied on unprocessed and preprocessed multispectral images. These two kinds of images were also used for the training of two original U-Nets, which do not make use of pretrained weights. Table 4.8 summarizes the results in terms of FM. It is notable that the preprocessing step increases the performance of both models and that the performance of the ResNet50 based architecture is superior to the performance gained by the U-Net.

DIBCO Metrics - MS-TEx The performance in terms of DIBCO metrics is given in Table 4.9. Only the performance of the ResNet50 on preprocessed images is given due to its superior performance. It is notable that the performance on the test set is significantly decreased. It is shown in Section 4.4.6 that this weak performance can be mainly attributed to a single multispectral image contained in the test set.

	FM	p-FM	PSNR	DRD
Train	86.63	82.47	17.72	3.58
Test	78.92	78.14	17.02	4.82

Table 4.9: Performance on the MS-TEx set.

Net	Train	Test
Original data		
U-Net	89.5	86.3
ResNet50	92.15	89.4
Normalized dynamic range		
U-Net	90.4	83.6
ResNet50	93.3	89.4

Table 4.10: Average FM for FG 1.

Net	Train	Test
Original data		
U-Net	80.1	75.2
ResNet50	90.3	84.2
Normalized dynamic range		
U-Net	79.6	77.1
ResNet50	91.3	83.2

Table 4.11: Average FM for FG 2

Evaluation of the Preprocessing - MSBin The performance on unprocessed and preprocessed images is also evaluated on the MSBin dataset. First the binarization of the iron gall based ink (FG 1) is evaluated. The results gained by the ResNet50 architecture and by the U-Net architecture are given in Table 4.10. It can be seen that again the ResNet50 clearly outperforms the U-Net. The performance of the ResNet50 on the unprocessed and preprocessed test images is similar: For both image types, an FM of 89.4% is achieved.

For the segmentation of FG 2 only a subset of the images is evaluated: The training set contains 30 images with FG 2 and the test set is comprised of 15 images containing FG 2. Only these images are used in the numerical evaluation, because in the remaining images the number of TP is zero and hence the corresponding FM is also zero. The results for the segmentation of FG 2 are given in Table 4.11. It can be seen that the U-Net is again outperformed by the ResNet50. The ResNet50 based FCN gains the highest performance, namely 84.2%, on unprocessed images, whereas an FM of 83.2% is achieved on preprocessed images.

The normalization of the dynamic range led to a significant performance increase on the MS-TEx dataset. Contrary, the results gained on the MSBin test are more ambiguous: For FG 1 the FM is nearly identical for preprocessed and original multispectral images. The preprocessing of FG 2 led to a performance decrease of 1%.

The dynamic ranges of the images in the MSBin dataset are not varying. This can be attributed to the circumstance that the exposure times were fixed for each channel. These exposure times were chosen in order to maximize the dynamic range and they were not

altered during the acquisition. This measurement makes the normalization of the dynamic ranges obsolete, which is also indicated by the numerical results gained. Therefore, in the following only results are provided, which are gained on the unprocessed MSBin dataset. The normalization is only advantageously for the MS-TEx dataset, because the dynamic ranges within the images are highly varying. It is generally preferably to avoid such a normalization step by using similar acquisition settings or by applying adequate calibration techniques (as for instance described in [Ber19]).

DIBCO Metrics - MSBin The performance in terms of FM, p-FM, PSNR and DRD for the segmentation of FG 1 is given in Table 4.12. The results for the segmentation of FG 2 are given in Table 4.13.

	FM	p-FM	PSNR	DRD
Train	92.15	94.24	15.97	6.87
Test	89.39	90.89	15.17	9.91

Table 4.12: Performance on FG 1 on the MSBin set.

	FM	p-FM	PSNR	DRD
Train	90.25	92.86	28.30	9.00
Test	84.17	84.53	27.62	15.77

Table 4.13: Performance on FG 2 on the MSBin set.

Dependency on the Size of the Training Set The aim of the following experiment is to determine the impact of the training size on the binarization performance. Six different FCN's have been trained with a varying number of training images, ranging from 10 to 60 images. The training images are randomly selected for each FCN. Figure 4.39 shows the F-Scores that are gained by the different FCN's on the test set. The validation set consists of 20 images and was not changed within the experiment. It is notable that the worst performance (85.68%) is gained by the model that is trained on 20 images. The performance is worse than the performance of the FCN, which is trained on 10 images. This can be attributed to the random selection of the training images and to the random initialization of the network. The best performance - namely 89.39% - is achieved by the model that is trained on the entire training set (60 images). The performance is nearly similar to the performance of the FCN, which is trained on 50 images. This FCN gains an F-Score of 89.27%.

It is notable that the number of training images has a relatively large impact on the performance: For instance, using 40 training images results in a performance decrease of more than 1%, compared to using 50 or 60 training images.

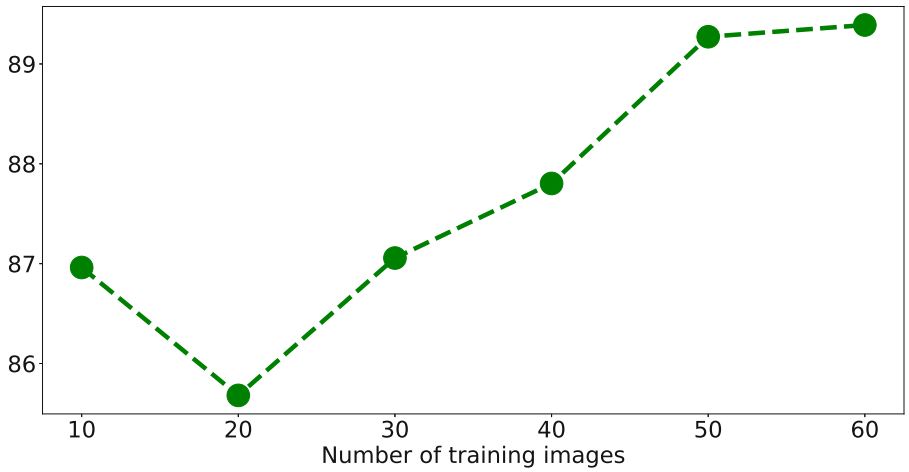


Figure 4.39: Influence of the training size.

Spectral Ranges In the following experiment it is evaluated which spectral ranges are most suitable for the segmentation of the different inks. Twelve different models have been trained on different single channels belonging to the multispectral images. Figure 4.40 shows the F-Scores gained by the twelve different models. Additionally, the performance is shown, which is gained by the model that is trained on all twelve channels. This model achieves the best binarization performance for both ink types: For FG 1 an average FM value of 89.39% is gained and for FG 2 an F-Score of 84.17% is obtained. FG 1 is written with iron gall based ink, which is best visible within the UV fluorescence images. Hence, the model that is trained on these UV illuminated (365 nm) images gains the best performance compared to the remaining models that are trained on single channels. The model gained an average F-Score of 86.76% on FG 1. The performances for FG 2 is lower than the one for FG 1: The best performance - 55.08% - on single channels is gained by the model that is applied on white light images. By making use of all spectral ranges the performance is significantly increased: The corresponding model gains an overall F-Score of 84.17%.

Figure 4.42 shows a multispectral image that contains both ink types: Both ink types are best distinguishable in the RGB image shown in Figure 4.42. This image is provided for visualization purposes, but is not contained within the MSBin dataset, because of its lower spatial resolution. The remaining images in Figure 4.42 are instead contained within the dataset. It is notable that the red ink is not visible under longer wavelengths, because it is reflected in the red and NIR spectral ranges. Contrary, the iron gall based ink is partially visible within these ranges. The FCN that is applied on all channels makes use of this spectral behavior and hence it is capable of correctly identifying both ink types.

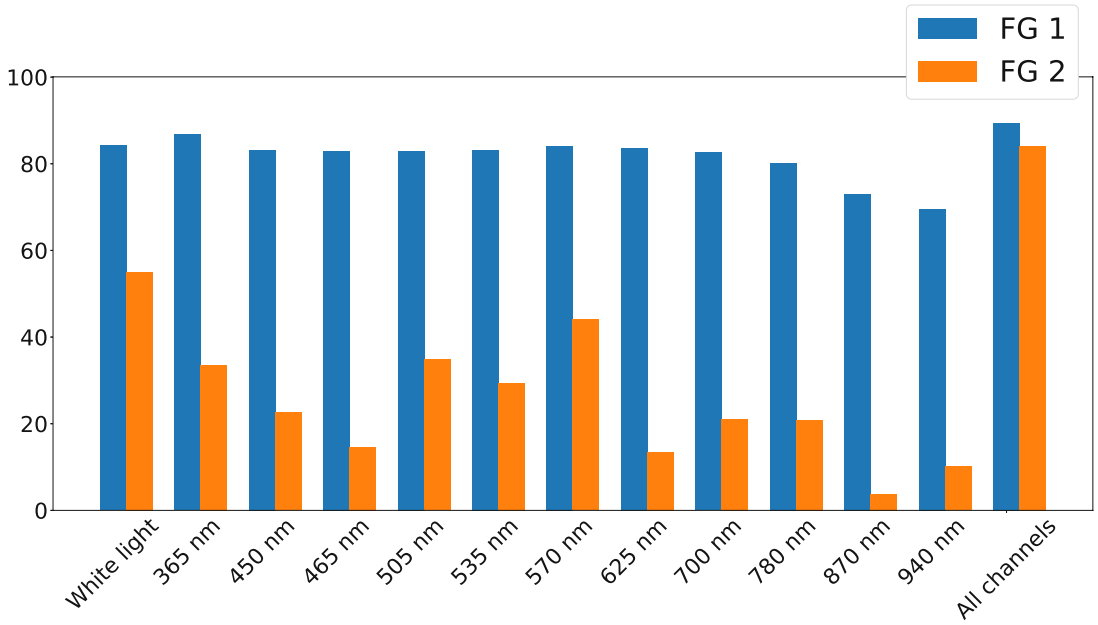


Figure 4.40: Evaluation of models trained on single channels and of a model trained on all channels.

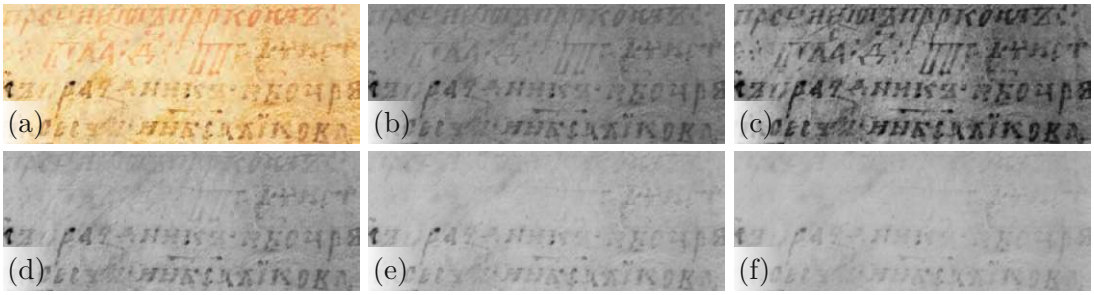


Figure 4.41: Multispectral image belonging to the MSBin test set. (a) White light RGB image. (b) White light channel. (c) 365 nm channel. (d) 570 nm channel. (e) 700 nm channel. (f) 780 nm channel.

4.4.6 Comparison of the Results

In this section the results gained by the four methods are compared to each other. First, the results gained by the methods on the MS-TEx dataset are compared. Afterwards, the performances are compared against the performances, which are gained by other binarization techniques. This involves also the results, which were obtained within the MS-TEx competition. Finally, the performances gained on the MSBin dataset are

Figure 4.42: Segmentation results. (a) Ground truth image. Blue regions are ambiguous regions that are excluded from the numerical evaluation. (b) Result gained on the white light channel. (c) Result achieved on all spectral ranges.

compared. Unfortunately, the source codes of the related approaches are not published. Hence, it is only possible to compare the performance of the four methods, which are introduced above.

Comparison of the Methods - MS-TEx First, the performances in terms of DIBCO metrics are given in Table 4.14. It is notable that the GrabCut based method performs best on the training and on the test set. The FCN based method performs worst on the test. For example, it gains an p-FM that is approximately 6% worse than the p-FM which is gained by the GrabCut based method. The GMM based method gains a higher p-FM, but it is still approximately 4% less than the p-FM (83.69%) that is scored by the ACE base method. The GrabCut based method gains the highest p-FM of 84.26%.

Method	FM	p-FM	PSNR	DRD
Train				
ACE v1	85.48	85.90	17.95	3.74
ACE v2	**87.43**	**86.46**	**18.48**	**3.18**
GMM	83.83	81.28	17.22	4.26
FCN	86.63	82.47	17.72	3.58
Test				
ACE v1	82.24	83.69	17.19	4.81
ACE v2	**84.61**	**84.26**	**17.62**	**4.33**
GMM	81.75	79.68	16.63	5.04
FCN	78.92	78.14	17.02	4.82

Table 4.14: DIBCO metrics gained on the MS-TEx dataset

Figure 4.43 shows exemplar binarization results that are gained on two images of the MS-TEx test set. Note that the stamp in the first image is partially visible in the ACE based results and in the output of the FCN. It is also notable that the elongated horizontal handwriting strokes are best retained in the FCN output. The last image in Figure 4.43 shows the image on which the FCN performs worst. It can also be seen that the GMM based result is over-segmented, since the horizontal line and the digits are binarized although they are not belonging to the iron gall-based handwriting.

Comparison to State-of-the-Art Methods The results, which are obtained on the MS-TEx test set, are compared to results that are gained by state-of-the-art methods. The comparison is given in Table 4.15. The performance metrics used are average FM, NRM and DRD. The last three columns show results that are obtained by excluding the worst binarization output for each of the methods. The results gained by the proposed methods are provided in the first group of Table 4.15. It can be seen that the GrabCut based method gains the best mean FM, DRD and NRM values, whereas the FCN based binarization achieves the worst average performance scores. The FCN based method gains an average FM value of 84.2%, if the worst binarization result is excluded from the test set. This F-Score is close to the corresponding F-Score of 85.48% that is gained by the GrabCut based method. This can be attributed to the circumstance that the test set contains one multispectral image, which has no similar counterpart in the training set. The corresponding spectral signature of the ink is highly varying and the preprocessing step is not capable of adapting the spectral signature in an adequate manner. Since the test set is relatively small (i.e. ten images), the weak performance on the single image has a huge impact on the overall average scores.

The second group in Table 4.15 shows the performance that is gained by the MS-TEx participants. The first two rows contain results that are gained by the two ACE based methods, which participated in the contest. It can be seen that the performance of these two submitted methods is worse than the corresponding performance that is given in the first group. The decreased performance of the submitted methods can be attributed to the fact that the conducted parameter optimization was not as exhaustive as the one in the current work. However, both submitted methods gain a better performance than the remaining binarization techniques, whereby the GrabCut based algorithm is the winning method of the MS-TEx competition with an average F-Score of 83.33%.

The third group in Table 4.15 summarizes the performance of other binarization methods that have been applied on the MS-TEx dataset. The highest performance is gained by the method of Abderrahmane et al. [AAC19]. The gained F-Score of 89.54% is significantly higher than the F-Scores gained by the remaining methods. However, this performance is questionable, since it is probably gained on six images instead of ten images - as was already explained in Section 2.3. The results of Abderrahmane et al. [AAC19] are written in italic in order to indicate that they are probably not gained on the entire test set and are not comparable to the remaining results. The source separation-based method of Salehani et al. [SAR+20] is a follow-up paper of [AAC19] and gains an average F-Score of 85.51% if the sources are manually selected. These results are written in italic in Table 4.15 in order to indicate that they are obtained with a manual processing step. In the case of automated source selection an average FM value of 82.33% is obtained, whereby the parameters in [SAR+20] are optimized on the test set.

Finally, the last group in Table 4.15 shows binarization results that are gained by two methods [How12], [LB13], which are designed for grayscale images. It can be seen that these methods achieve a lower performance than the methods, which make use of spectral

information.

Method	FM	NRM	DRD	FM*	NRM*	DRD*
Proposed						
ACE v1	82.24	9.28	4.81	84.69	8.80	4.11
ACE v2	**84.61**	6.93	**4.33**	**85.48**	6.54	**3.67**
GMM	81.75	7.60	5.04	82.73	6.61	4.59
FCN	78.92	8.97	4.82	84.63	**5.30**	4.02
MSTEx competition results						
ACE v1 [HDS15]	81.87	10.05	4.793	83.29	9.64	4.07
ACE v2 [DHS16]	83.33	9.25	4.241	84.87	8.70	3.56
Zhang and Liu	79.09	12.58	5.084	80.14	11.41	4.53
Wu et al.	76.57	14.31	5.548	78.49	12.77	5.02
Raza	73.14	10.26	9.325	74.38	9.77	8.59
MSI binarization methods						
Mogghadam [MC15]				73.64	6.87	9.56
Salehani [SAR+20] - manual	85.51	5.71	3.41			
Salehani [SAR+20] - automated	82.33	**6.55**	4.56			
Abderrahmane [AAC19]	89.54	0.65	2.23			
Grayscale binarization methods						
Howe	70.35	12.09	8.60	75.96	8.5	7.81
Lelore	67.16	6.97	15.27	69.37	6.50	12.36

Table 4.15: Comparison to other binarization methods.

Comparison of the Methods - MSBin The results gained on the MSBin dataset are shown in Table 4.16. The FCN based method gains the best scores. It was already described above that the GrabCut based method was not capable of achieving the expected superior performance in comparison to the base method. Instead, both ACE based methods achieve similar results. It is also notable that the GMM based method performs worst. The ACE and GMM based methods have been especially designed for the MS-TEx dataset. These methods are clearly outperformed by the FCN based method on the MSBin dataset. Table 4.16 shows also the performance that is gained by applying the algorithm of Su et al. [SLT10] on a single multispectral channel. This method is used the ACE based and GMM based method. It can be seen that these methods improve the outcome of the grayscale binarization method by incorporating spectral information. Figure 4.44 shows outputs of the methods that are gained on two multispectral images, which are belonging to the MSBin test set.

4.5 Summary

The binarization of multispectral document images was covered in this chapter. Three methods have been designed for the MS-TEx dataset. It was shown that the GMM

Method	FM	p-FM	PSNR	DRD
Train				
ACE v1	87.66	87.75	14.27	13.04
ACE v2	87.54	87.83	14.35	13.53
GMM	86.80	86.94	14.05	13.39
FCN	**92.15**	**94.24**	**15.97**	**6.87**
Su et al.	86.58	88.09	14.24	13.80
Test				
ACE v1	81.28	81.18	13.28	22.03
ACE v2	81.25	81.36	13.27	20.80
GMM	80.00	80.28	13.18	20.35
FCN	**89.39**	**90.89**	**15.17**	**9.91**
Su et al.	78.68	79.83	12.53	25.77

Table 4.16: DIBCO metrics gained on the MSBin dataset

based method can at least compete with other state-of-the-art methods. The ACE based methods were ranked first and second in the MS-TEx competition. The use of the GrabCut algorithm led to a significant performance increase, compared to the base method. Additionally, it is shown that the FCN based binarization is outperformed by the three methods in terms of average FM. However, this can be mainly attributed to the bad performance that is gained on a single multispectral image. The supervised approach is not capable of successfully segmenting the handwriting that is different to the writings in the training set. If this image is excluded from the test set, the FCN based method can at least compete with the state-of-the-art.

The methods were also evaluated on a second dataset, named MSBin. This data set consists of a larger training set, which enables the successful training of the FCN. It was shown that the FCN based method clearly outperforms the remaining three methods. These methods are especially designed for the MS-TEx data set and they are only partially capable of handling the degradations contained in the MSBin data set. Contrary, the FCN based approach achieved a superior performance. Additionally, it was shown that the method is capable of differentiating between the two ink types that are present in the MSBin dataset.

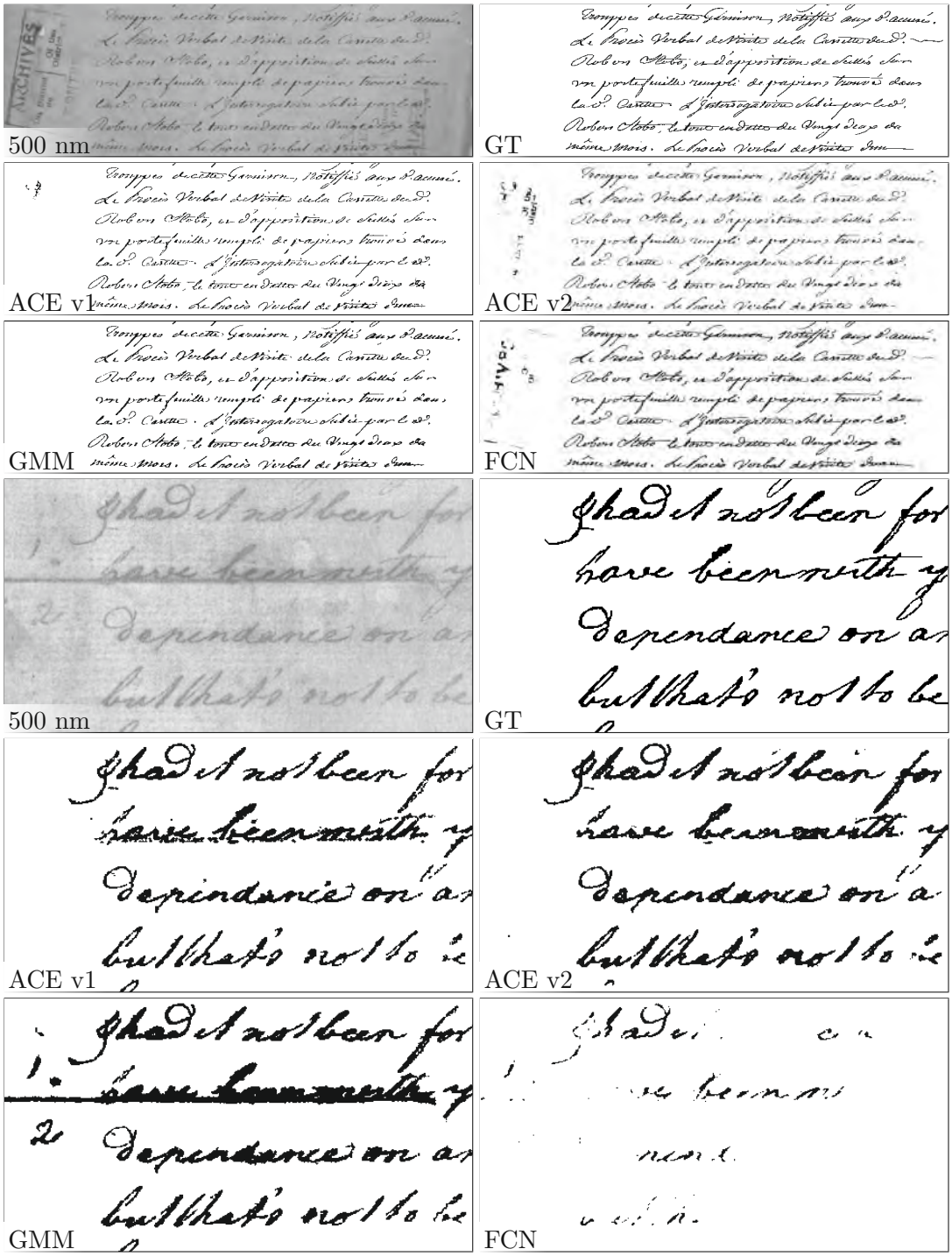


Figure 4.43: MS-TEx test set binarization examples.

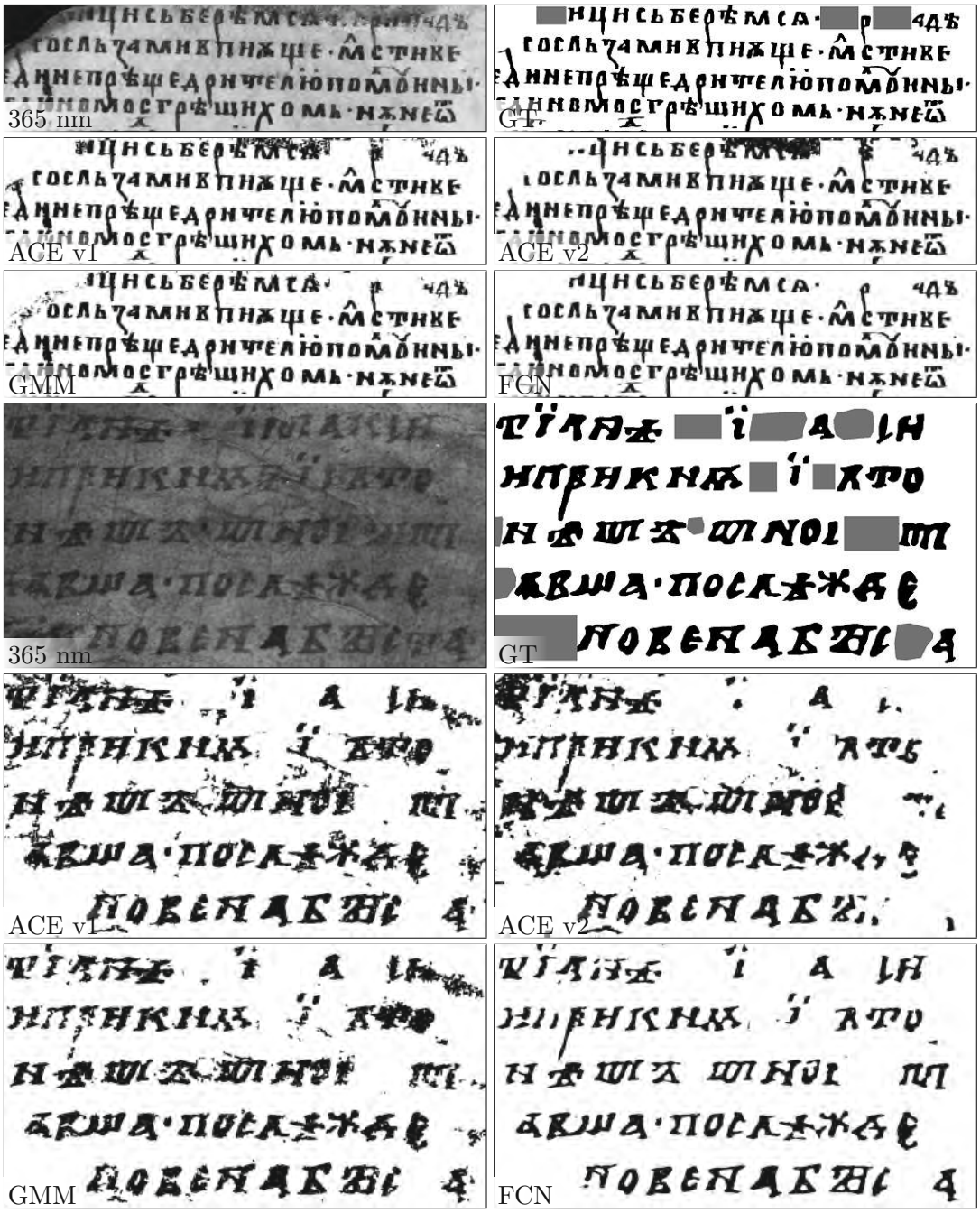


Figure 4.44: MSBin test set binarization examples. The gray regions in the ground truth image (GT) are uncertain regions. These regions are treated as background regions in the numerical results and are visualized similar in the binarization results.

CHAPTER 5

Conclusion and Future Work

In this thesis, restoration techniques for multispectral document images have been presented. Two different kinds of restoration methods have been proposed: An enhancement method which increases the visibility was introduced as well as multiple binarization methods for multispectral document images. The related work was presented first in Chapter 2. The underlying theory of MSI was discussed in Chapter 3. MSI has proven to be a valuable tool for the non-invasive investigation of ancient documents. The portable MSI system is especially designed for acquiring images of historical manuscripts. It was shown that especially UV fluorescence imaging can be used to increase the legibility of faded-out writings that are written on parchment, since parchment is a luminescent material whereas iron gall based ink is not fluorescent. Additionally, it was shown that NIR reflectography can be useful for other supporting materials - as was shown for a fragment that is written on leather.

While MSI is used to support the work of scholars, the manual investigation of such multispectral images is still a tedious task, since it is partially not known a-priori in which channel the writing is best visible. Additionally, the visibility within the multispectral channels is still limited. In order to overcome these drawbacks, an enhancement method was introduced that projects the multispectral samples on a one-dimensional space by applying a LDA based transformation. Since the LDA classifier requires labeled training samples, a labeling strategy is proposed that is applied on PCA images. The main contribution of the method is that it was shown that the LPP based text line detection allows for an automated labeling of strongly degraded handwritings within a from coarse to fine scheme. The text line detection is applied first, since a binarization is error-prone, because the writings are barely visible within the PCA images. The method was evaluated in a qualitative analysis, which was conducted by philologists, where its resulting images were compared to resulting images of PCA and ICA. The scholars assigned in five out of seven test images the highest score to the LDA based resulting images.

The method was also used as a preprocessing step for an OCR system. The method had to be adapted for non-degraded regions, because the original method gains no visibility increase on non-degraded image portions. Despite this adaption, the method was not capable of increasing the performance on eleven test images, which are not degraded: The method gained an average character accuracy of 0.83 on this test set, whereas the best performance, namely an accuracy of 0.87, was achieved on unprocessed UV fluorescence images. Contrary, the results on eleven degraded test images are promising: The method gained the highest character accuracy of .60.

A drawback of the text line detection based labeling is its dependency on regular text line structures. Therefore, other document layout analysis methods have to be incorporated in order to detect for instance ornamental elements or text with irregular structures. This involves especially deep learning-based approaches, because of the expected superior performance. As a further future work, it has to evaluated if the performance can be increased by using other classifiers for the dimension reduction. This involves SVM's and neural networks amongst others.

Various approaches have been introduced in this work that are designed for the binarization of multispectral document images: First, a method was proposed that combines ACE based target detection with a traditional binarization approach that is applied on a single grayscale channel. The target signature is directly estimated on the multispectral image and hence no supervised training is required. The binarization is fulfilled within a rule-based system.

The approach was improved by adding a spatial segmentation step. The segmentation is based on the GrabCut algorithm and replaces the rule-based system, which is used in the base method. The GrabCut is guided by information that is gained by the target detection step and by the initial binarization step.

The third binarization approach is based on GMM based clustering. It was shown that this clustering is sensitive to background variations, because these variations are partially modeled by multiple Gaussians. In order to overcome this drawback, a background compensation step is performed on each multispectral channel. It was shown that this background compensation leads to a significant performance increase.

Finally, an FCN was utilized for the binarization task. The FCN is an U-Net based approach, whereby the encoding path follows the ResNet-50 architecture. Since the MS-TEx dataset is comprised of a relatively small number of training images, a further dataset was introduced in order to enable a successful application of the deep learning-based method. The FCN serves as a baseline approach for the dataset introduced. In a future work, the performances of other deep learning-based segmentation methods have to be analyzed, including for instance Global Convolutional Networks [PZY+17] or networks that are based on atrous convolution [CZP+18].

The evaluation on the MS-TEx test set showed that the GrabCut based method gains the highest FM (84.61%) and p-FM (84.26%) scores. The base ACE method gained the second-best performance (FM = 82.24% and p-FM = 83.69). The GMM based method

and FCN were outperformed by the ACE based methods, since they gained FM scores of 81.75% and 78.92%, respectively. The ACE based methods participated in the MS-TEx competition, where they were ranked first and second. The GMM and FCN based method can at least compete with the remaining state-of-the-art approaches. The weak performance of the supervised method can be mainly attributed to the inappropriate binarization of a single test image.

Contrary, the FCN gained the highest performance scores on the MSBin dataset: The method achieved an average F-Score of 89.39% and a p-FM of 90.89. The ACE based methods gained in contrast FM and p-FM scores of approximately 81% and the GMM based method achieved corresponding scores of about 80%. The superior performance of the FCN can be on the hand attributed to the superior performance of deep learning-based binarization methods in general - as it was for example shown in the DIBCO 2017 contest. On the other hand, the ACE and GMM based methods have been developed and finetuned on the MS-TEx dataset and are partially not capable of handling the degradations contained in the MSBin dataset.

The performance of the unsupervised methods could be improved by multiple measurements: For example, the rule-based system in the ACE based method makes use of global thresholds, which could be replaced by adaptive thresholds. The GrabCut is developed for RGB images and therefore, the dimensionality of the multispectral images is reduced by creating the pseudo-color images. Modelling the GMM's on entire multispectral images would increase the information that is provided to the GrabCut and thus the performance could be increased. The GMM based method is exemplar-based and hence the rule-based system is required to identify pure and mixed pixels. By comparing the spectral signatures of the test and training set in a supervised manner, the rule-based system could eventually be replaced. Such a comparison could also be used to determine the target signature of the ink. Thus, the initial binarization step could be removed from the ACE based binarization methods.

Although these measurements might increase the performance, it can be expected that deep learning-based binarization methods gain still a better performance, given a sufficiently large training set is available. While the MSBin dataset introduced offers a sufficient number of training images, the FCN was outperformed by the remaining methods on the MS-TEx test set, since one particular test image deviates from the training set. This shows the importance of sufficiently large training and test sets, particularly for deep learning-based methods. The training of deep learning-based methods could be improved by synthetically generating more training images. This could be achieved by conditioning an GAN based approach on ground truth images, as it was suggested in [TBSM19] for grayscale document images. Another possibility for synthesizing new data is spectral reconstruction [ATBS+20] based on RGB images. Thus, multispectral datasets could be synthesized from existing datasets published in the DIBCO series. Generally speaking, the number of datasets for multispectral document binarization is very limited. As a future work, new datasets have to be created, since such databases are required for exhaustive evaluations and further developments of binarization algorithms.

Bibliography

[AAC19] Rahiche Abderrahmane, Bakhta Athmane, and Mohamed Cheriet. Blind source separation based framework for multispectral document images binarization. *2019 International Conference on Document Analysis and Recognition, ICDAR 2019*, 2019.

[ABAmO19] Younes Akbari, Alceu S. Britto, Somaya Al-maadeed, and Luiz S. Oliveira. Binarization of degraded document images using convolutional neural networks based on predicted two-channel images. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, sep 2019.

[ABD+13] Anila Anitha, Andrei Brasoveanu, Marco Duarte, Shannon Hughes, Ingrid Daubechies, Joris Dik, Koen Janssens, and Matthias Alfeld. Restoration of x-ray fluorescence images of hidden paintings. *Signal Processing*, 93(3):592–604, mar 2013.

[Abu89] Ahmed S. Abutaleb. Automatic thresholding of gray-level pictures using two-dimensional entropy. *Computer Vision, Graphics, and Image Processing*, 47(1):22–32, 1989.

[APA+16] Corneliu T. C. Arsene, Peter E. Pormann, Naima Afif, Stephen Church, and Mark Dickinson. High performance software in multidimensional reduction methods for image processing with application to ancient manuscripts. *CoRR*, abs/1612.06457, 2016.

[APS+15] Muhammad Zeshan Afzal, Joan Pastor-Pellicer, Faisal Shafait, Thomas M. Breuel, Andreas Dengel, and Marcus Liwicki. Document image binarization using LSTM: A sequence learning approach. In *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing, HIP@ICDAR 2015, Nancy, France, August 22, 2015*, pages 79–84. ACM, 2015.

[Arc] The Archimedes Palimpsest. `https://archimedespalimpsest.net/Data/120v-121r/`. Accessed: March 7, 2021.

[ATBS+20] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, Graham Finlayson, Shai Givati, Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li,

Fei Liu, Zhiqiang Lang, Wei Wei, Lei Zhang, Jiangtao Nie, Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, Tingyu Lin, Youngjung Kim, Changyeop Shin, Kyeongha Rho, Sungho Kim, Zhiyu ZHU, Junhui HOU, He Sun, Jinchang Ren, Zhenyu Fang, Yijun Yan, Hao Peng, Xiaomei Chen, Jie Zhao, Tarek Stiebel, Simon Koppers, Dorit Merhof, Honey Gupta, Kaushik Mitra, Biebele Joslyn Fubara, Mohamed Sedky, Dave Dyke, Atmadeep Banerjee, Akash Palrecha, Sabarinathan sabarinathan, K Uma, D Synthiya Vinothini, B Sathya Bama, and S M Md Mansoor Roomi. NTIRE 2020 challenge on spectral reconstruction from an RGB image. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, jun 2020.

[AV07] David Arthur and Sergei Vassilvitskii. k-means++: the advantages of careful seeding. In Nikhil Bansal, Kirk Pruhs, and Clifford Stein, editors, *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January 7-9, 2007*, pages 1027–1035. SIAM, 2007.

[BAS+16] Umair Muneer Butt, Sheraz Ahmed, Faisal Shafait, Christian Nansen, Ajmal Saeed Mian, and Muhammad Imran Malik. Automatic signature segmentation using hyper-spectral imaging. In *15th International Conference on Frontiers in Handwriting Recognition, ICFHR 2016, Shenzhen, China, October 23-26, 2016*, pages 19–24. IEEE Computer Society, 2016.

[Ber86] John Bernsen. Dynamic thresholding of gray-level images. In *Eighth International Conference on Pattern Recognition, ICPR 1986*, pages 1251–1255, 1986.

[Ber19] Roy S. Berns. Color and material-appearance measurement. In *Billmeyer and Saltzman's Principles of Color Technology*, pages 111–144. John Wiley & Sons, Inc., mar 2019.

[Bes86] Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.

[BGB+21] Suman Kumar Bera, Soulib Ghosh, Showmik Bhowmik, Ram Sarkar, and Mita Nasipuri. A non-parametric binarization method based on ensemble of clustering algorithms. *Multim. Tools Appl.*, 80(5):7653–7673, 2021.

[BHK97] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997.

[BIL19] Manuel Bouillon, Rolf Ingold, and Marcus Liwicki. Grayification: A meaningful grayscale conversion to improve handwritten historical documents analysis. *Pattern Recognit. Lett.*, 121:46–51, 2019.

[BIN19] Pavel Bezmaternykh, Dmitry Ilin, and Dmitry Nikolaev. U-net-bin: hacking the document image binarization contest. *Computer Optics*, 43(5):825–832, oct 2019.

[Bis07] Christopher M. Bishop. *Pattern recognition and machine learning, 5th Edition*. Information science and statistics. Springer, 2007.

[BJ01] Yuri Boykov and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *Proceedings of the Eighth International Conference On Computer Vision (ICCV-01), Vancouver, British Columbia, Canada, July 7-14, 2001 - Volume 1*, pages 105–112. IEEE Computer Society, 2001.

[BK04] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(9):1124–1137, 2004.

[BKC17] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[BM20] Galal M. Binmakhashen and Sabri A. Mahmoud. Document layout analysis. *ACM Computing Surveys*, 52(6):1–36, jan 2020.

[BSA08] Johannes Brauers, Nils Schulte, and Til Aach. Multispectral filter-wheel cameras: Geometric distortion model and compensation algorithms. *IEEE transactions on image processing*, 17(12):2368–2380, 2008.

[BSF94] Yoshua Bengio, Patrice Y. Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Networks*, 5(2):157–166, 1994.

[BTG06] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: speeded up robust features. In Ales Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision - ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part I*, volume 3951 of *Lecture Notes in Computer Science*, pages 404–417. Springer, 2006.

[BW02] Thomas Brox and Joachim Weickert. Nonlinear matrix diffusion for optic flow estimation. In Luc Van Gool, editor, *Pattern Recognition, 24th DAGM Symposium, Zurich, Switzerland, September 16-18, 2002, Proceedings*, volume 2449 of *Lecture Notes in Computer Science*, pages 446–453. Springer, 2002.

[BYBS10] Mauro Birattari, Zhi Yuan, Prasanna Balaprakash, and Thomas Stützle. F-race and iterated f-race: An overview. In Thomas Bartz-Beielstein, Marco

Chiarandini, Luís Paquete, and Mike Preuss, editors, *Experimental Methods for the Analysis of Optimization Algorithms.*, pages 311–336. Springer, 2010.

[Can86] John F. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.

[CG19] Jorge Calvo-Zaragoza and Antonio-Javier Gallego. A selectional auto-encoder approach for document image binarization. *Pattern Recognition*, 86:37–47, 2019.

[Cha] Chein-I Chang. Spectral information divergence for hyperspectral image analysis. In *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS 99 (Cat. No.99CH36293)*. IEEE.

[CHHH11] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1548–1560, 2011.

[CM11] Moshe Caine and Michael Magen. Pixels and parchment: The application of RTI and infrared imaging to the dead sea scrolls. In Stuart Dunn, Jonathan P. Bowen, and Kia Ng, editors, *Electronic Visualisation and the Arts, EVA 2011, London, UK, 6-8 July 2011*, Workshops in Computing. BCS, 2011.

[Coh60] Jacob Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, apr 1960.

[CS05] Adam P. Cisz and John R. Schott. Performance comparison of hyperspectral target detection algorithms in altitude varying scenes. In Sylvia S. Shen and Paul E. Lewis, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*. SPIE, jun 2005.

[CSS98] Mohamed Cheriet, Joseph N. Said, and Ching Y. Suen. A recursive thresholding technique for image segmentation. *IEEE Trans. Image Processing*, 7(6):918–921, 1998.

[CZP+18] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII*, volume 11211 of *Lecture Notes in Computer Science*, pages 833–851. Springer, 2018.

[DFO20] Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020.

[DGH19] Binu Melit Devassy, Sony George, and Jon Y Hardeberg. Comparison of ink classification capabilities of classic hyperspectral similarity features. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, volume 8, pages 25–30. IEEE, 2019.

[DHS12] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[DHS16] Markus Diem, Fabian Hollaus, and Robert Sablatnig. MSIO: multispectral document image binarization. In *12th IAPR Workshop on Document Analysis Systems, DAS 2016, Santorini, Greece, April 11-14, 2016*, pages 84–89. IEEE Computer Society, 2016.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via theEMAlgorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, sep 1977.

[DLS07] Markus Diem, Martin Lettner, and Robert Sablatnig. Registration of Multi-Spectral Manuscript Images. In *VAST: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, pages 133–140. The Eurographics Association, 2007.

[DS10] Markus Diem and Robert Sablatnig. Are characters objects? In *International Conference on Frontiers in Handwriting Recognition, ICFHR 2010, Kolkata, India, 16-18 November 2010*, pages 565–570, 2010.

[Eas] The Archimedes Palimpsest. `https://archimedespalimpsest.net/Data/120v-121r/`. Accessed: March 7, 2021.

[EEG+14] Mark Everingham, S. M. Ali Eslami, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, jun 2014.

[EN10] Roger L Easton and William Noel. Infinite possibilities: Ten years of study of the archimedes palimpsest. *Proceedings of the American Philosophical Society*, 154(1):50–76, 2010.

[EOW10] Boris Epshtein, Eyal Ofek, and Yonatan Wexler. Detecting text in natural scenes with stroke width transform. In *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, pages 2963–2970. IEEE Computer Society, 2010.

[FHGS13] Stefan Fiel, Fabian Hollaus, Melanie Gau, and Robert Sablatnig. Writer identification on historical glagolitic documents. In Bertrand Coüasnon and Eric K. Ringger, editors, *Document Recognition and Retrieval XXI*. SPIE, dec 2013.

[FK06] Christian Fischer and Ioanna Kakoulli. Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications. *Studies in Conservation*, 51(sup1):3–16, jun 2006.

[GB10] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In Yee Whye Teh and D. Mike Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010, Chia Laguna Resort, Sardinia, Italy, May 13-15, 2010*, volume 9 of *JMLR Proceedings*, pages 249–256. JMLR.org, 2010.

[GD07] Mark Grundland and Neil A. Dodgson. Decolorize: Fast, contrast enhancing, color to grayscale conversion. *Pattern Recognit.*, 40(11):2891–2896, 2007.

[GG84] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721–741, 1984.

[GH15] Sony George and Jon Yngve Hardeberg. Ink classification and visualisation of historical manuscripts: Application of hyperspectral imaging. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, aug 2015.

[GJG07] Maya R. Gupta, Nathaniel P. Jacobson, and Eric K. Garcia. OCR binarization and image pre-processing for searching historical documents. *Pattern Recognition*, 40(2):389–397, 2007.

[GNP09] Basilios Gatos, Konstantinos Ntirogiannis, and Ioannis Pratikakis. ICDAR 2009 document image binarization contest (DIBCO 2009). In *10th International Conference on Document Analysis and Recognition, ICDAR 2009, Barcelona, Spain, 26-29 July 2009*, pages 1375–1382. IEEE Computer Society, 2009.

[GNP11] Basilios Gatos, Konstantinos Ntirogiannis, and Ioannis Pratikakis. DIBCO 2009: document image binarization contest. *IJDAR*, 14(1):35–44, 2011.

[GPP06] Basilios Gatos, Ioannis Pratikakis, and Stavros J. Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39(3):317–327, 2006.

[HBS05] Hatem Hamza, Abdel Belaïd, and Eddie Smigiel. Neural based binarization techniques. In *Eighth International Conference on Document Analysis and Recognition (ICDAR 2005), 29 August - 1 September 2005, Seoul, Korea*, pages 317–321. IEEE Computer Society, 2005.

[HBS19] Fabian Hollaus, Simon Brenner, and Robert Sablatnig. CNN based binarization of MultiSpectral document images. In *2019 International Conference*

on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019, pages 533–538. IEEE, 2019.

[HC11a] Rachid Hedjam and Mohamed Cheriet. Combining statistical and geometrical classifiers for text extraction in multispectral document images. In Bill Barrett, Michael S. Brown, R. Manmatha, and Jake Gehring, editors, *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing, HIP ICDAR 2011, Beijing, China, September 16-17, 2011*, pages 98–105. ACM, 2011.

[HC11b] Rachid Hedjam and Mohamed Cheriet. Novel data representation for text extraction from multispectral historical document images. In *2011 International Conference on Document Analysis and Recognition, ICDAR 2011, Beijing, China, September 18-21, 2011*, pages 172–176. IEEE Computer Society, 2011.

[HC13a] Rachid Hedjam and Mohamed Cheriet. Ground-truth estimation in multispectral representation space: Application to degraded document image binarization. In *12th International Conference on Document Analysis and Recognition, ICDAR 2013, Washington, DC, USA, August 25-28, 2013*, pages 190–194, 2013.

[HC13b] Rachid Hedjam and Mohamed Cheriet. Historical document image restoration using multispectral imaging system. *Pattern Recognition*, 46(8):2297–2312, 2013.

[HCK14] Rachid Hedjam, Mohamed Cheriet, and Margaret Kalacska. Constrained energy maximization and self-referencing method for invisible ink detection from multispectral historical document images. In *22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August 24-28, 2014*, pages 3026–3031. IEEE Computer Society, 2014.

[HDS14] Fabian Hollaus, Markus Diem, and Robert Sablatnig. Improving OCR accuracy by applying enhancement techniques on multispectral images. In *22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August 24-28, 2014*, pages 3080–3085. IEEE Computer Society, 2014.

[HDS15] Fabian Hollaus, Markus Diem, and Robert Sablatnig. Binarization of multispectral document images. In George Azzopardi and Nicolai Petkov, editors, *Computer Analysis of Images and Patterns - 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II*, volume 9257 of *Lecture Notes in Computer Science*, pages 109–120. Springer, 2015.

[HDS18] Fabian Hollaus, Markus Diem, and Robert Sablatnig. Multispectral image binarization using gmms. In *16th International Conference on Frontiers in Handwriting Recognition, ICFHR 2018, Niagara Falls, NY, USA, August 5-8, 2018*, pages 570–575, 2018.

[HGS12] Fabian Hollaus, Melanie Gau, and Robert Sablatnig. Multispectral image acquisition of ancient manuscripts. In Marinos Ioannides, Dieter Fritsch, Johanna Leissner, Rob Davies, Fabio Remondino, and Rossella Caffo, editors, *Progress in Cultural Heritage Preservation - 4th International Conference, EuroMed 2012, Limassol, Cyprus, October 29 - November 3, 2012. Proceedings*, volume 7616 of *Lecture Notes in Computer Science*, pages 30–39. Springer, 2012.

[HGS13] Fabian Hollaus, Melanie Gau, and Robert Sablatnig. Enhancement of multispectral images of degraded documents by employing spatial information. In *12th International Conference on Document Analysis and Recognition, ICDAR 2013, Washington, DC, USA, August 25-28, 2013*, pages 145–149, 2013.

[HJB+12] Mattias P. Heinrich, Mark Jenkinson, Manav Bhushan, Tahreema Matin, Fergus V. Gleeson, Sir Michael Brady, and Julia A. Schnabel. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423–1435, oct 2012.

[HK13] Nathan Hagen and Michael W. Kudenov. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):090901, sep 2013.

[HNKC15] Rachid Hedjam, Hossein Ziaei Nafchi, Margaret Kalacska, and Mohamed Cheriet. Influence of color-to-gray conversion on the performance of document image binarization: Toward a novel optimization problem. *IEEE Trans. Image Process.*, 24(11):3637–3651, 2015.

[HNM+15] Rachid Hedjam, Hossein Ziaei Nafchi, Reza Farrahi Moghaddam, Margaret Kalacska, and Mohamed Cheriet. ICDAR 2015 contest on multispectral text extraction (ms-tex 2015). In *13th International Conference on Document Analysis and Recognition, ICDAR 2015, Nancy, France, August 23-26, 2015*, pages 1181–1185. IEEE Computer Society, 2015.

[HO00] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430, 2000.

[How11] Nicholas R. Howe. A laplacian energy for document binarization. In *2011 International Conference on Document Analysis and Recognition, ICDAR 2011, Beijing, China, September 18-21, 2011*, pages 6–10. IEEE Computer Society, 2011.

[How12] Nicholas R. Howe. Document binarization with automatic parameter tuning. *International Journal on Document Analysis and Recognition (IJDAR)*, 16(3):247–258, sep 2012.

[HS14] Fabian Hollaus and Robert Sablatnig. Enhancement of multispectral images of ancient manuscripts. In *2014 Eurographics Workshop on Graphics and Cultural Heritage, GCH 2014, Darmstadt, Germany, October 6-8, 2014*, pages 45–53. Eurographics Association, 2014.

[HS17] Fabian Hollaus and Robert Sablatnig. MultiSpectral imaging for the analysis of historical handwritings and forgery detection. In *Die getäuschte Wissenschaft*, pages 233–246. V&R unipress, may 2017.

[HS19] Sheng He and Lambert Schomaker. Deepotsu: Document enhancement and binarization using iterative deep learning. *Pattern Recognition*, 91:379–390, 2019.

[HSB+99] Jon Yngve Hardeberg, Francis Schmitt, Hans Brettel, Jean-Pierre Crettez, and Henri Maître. Multispectral image acquisition and simulation of illuminant changes. In *Colour imaging: vision and technology*, number 145–161. Citeseer, 1999.

[HW12] William R. Hendee and Peter N.T. Wells. *The Perception of Visual Information*. Springer, 2012.

[HWS15] Sheng He, Marco A. Wiering, and Lambert Schomaker. Junction detection in handwritten documents and its application to writer identification. *Pattern Recognit.*, 48(12):4036–4048, 2015.

[HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016.

[Jai89] Anil K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.

[JCK11] Roger L. Easton Jr., William A. Christens-Barry, and Keith T. Knox. Spectral image processing and analysis of the archimedes palimpsest. In *Proceedings of the 19th European Signal Processing Conference, EUSIPCO 2011, Barcelona, Spain, August 29 - Sept. 2, 2011*, pages 1440–1444. IEEE, 2011.

[JSH+18] Fuxi Jia, Cunzhao Shi, Kun He, Chunheng Wang, and Baihua Xiao. Degraded document image binarization using structural symmetry of strokes. *Pattern Recognition*, 74:225–240, feb 2018.

[JTHW06] Wen Jin, Anthony K. H. Tung, Jiawei Han, and Wei Wang. Ranking outliers using symmetric neighborhood relationship. In Wee Keong Ng, Masaru Kitsuregawa, Jianzhong Li, and Kuiyu Chang, editors, *Advances in Knowledge Discovery and Data Mining, 10th Pacific-Asia Conference, PAKDD 2006, Singapore, April 9-12, 2006, Proceedings*, volume 3918 of *Lecture Notes in Computer Science*, pages 577–593. Springer, 2006.

[Kö14] Raphael Kögel. *Die Photographie historischer Dokumente nebst den Grundzügen der Reproduktionsverfahren*. University Library Heidelberg, 1914.

[Kö20] Raphael Kögel. *Die Palimpsestphotographie : (Photographie der radierten Schriften) in ihren wissenschaftlichen Grundlagen und praktischen Anwendungen*. Halle (Saale), 1920.

[KB15] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[KDB11] Seon Joo Kim, Fanbo Deng, and Michael S. Brown. Visual enhancement of old documents with hyperspectral imaging. *Pattern Recognition*, 44(7):1461–1469, 2011.

[KDG16] Nal Kalchbrenner, Ivo Danihelka, and Alex Graves. Grid long short-term memory. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[KDS11] Florian Kleber, Markus Diem, and Robert Sablatnig. Scale space binarization using edge information weighted by a foreground estimation. In *2011 International Conference on Document Analysis and Recognition*. IEEE, sep 2011.

[KKS19] Muhammad Jaleed Khan, Khurram Khurshid, and Faisal Shafait. A spatio-spectral hybrid convolutional architecture for hyperspectral document authentication. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, sep 2019.

[KKT09] Pushmeet Kohli, M. Pawan Kumar, and Philip H. S. Torr. P^3 & beyond: Move making algorithms for solving higher order functions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(9):1645–1656, 2009.

[KKT18] Netanel Kligler, Sagi Katz, and Ayellet Tal. Document enhancement using visibility detection. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 2374–2382. IEEE Computer Society, 2018.

[KLB+93] F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2-3):145–163, may 1993.

[KLD+08] Florian Kleber, Martin Lettner, Markus Diem, Maria Vill, Robert Sablatnig, Heinz Miklas, and Melanie Gau. Multispectral acquisition and analysis of ancient documents. In *Proceedings of the 14th International Conferenceon Virtual Systems and MultiMedia (VSMM 2008), Dedicated to Cultural Heritage- Project Papers*, pages 184–191, 01 2008.

[KP06] Zoltan Kato and Ting-Chuen Pong. A markov random field image segmentation model for color textured images. *Image Vision Comput.*, 24(10):1103–1114, 2006.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[KSK18] Ronald Kemker, Carl Salvaggio, and Christopher Kanan. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS journal of photogrammetry and remote sensing*, 145:60–77, 2018.

[KSM13] Zohaib Khan, Faisal Shafait, and Ajmal S. Mian. Hyperspectral imaging for ink mismatch detection. In *12th International Conference on Document Analysis and Recognition, ICDAR 2013, Washington, DC, USA, August 25-28, 2013*, pages 877–881. IEEE Computer Society, 2013.

[KSM15] Zohaib Khan, Faisal Shafait, and Ajmal S. Mian. Automatic ink mismatch detection for forensic document analysis. *Pattern Recognition*, 48(11):3615–3626, 2015.

[KTB07] Sagi Katz, Ayellet Tal, and Ronen Basri. Direct visibility of point sets. *ACM Trans. Graph.*, 26(3):24, 2007.

[KYAK18] Muhammad Jaleed Khan, Adeel Yousaf, Asad Abbas, and Khurram Khurshid. Deep learning for automated forgery detection in hyperspectral document images. *J. Electronic Imaging*, 27(05):053001, 2018.

[LB13] Thibault Lelore and Frédéric Bouchara. FAIR: A fast algorithm for document image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):2039–2048, 2013.

[LBE04] Yann Leydier, Frank Le Bourgeois, and Hubert Emptoz. Serialized unsupervised classifier for adaptative color image segmentation: Application to digitized ancient manuscripts. In *17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, August 23-26, 2004*, pages 494–497. IEEE Computer Society, 2004.

[Lev66] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710, 1966.

[LG13] Guillaume Lazzara and Thierry Géraud. Efficient multiscale sauvola's binarization. *International Journal on Document Analysis and Recognition (IJDAR)*, 17(2):105–123, jul 2013.

[Li09] Stan Z. Li. *Markov Random Field Modeling in Image Analysis*. Advances in Pattern Recognition. Springer, 2009.

[LKS04] Haiping Lu, Alex C. Kot, and Yun Q. Shi. Distance-reciprocal distortion measure for binary document images. *IEEE Signal Processing Letters*, 11(2):228–231, feb 2004.

[LKSM08] Martin Lettner, Florian Kleber, Robert Sablatnig, and Heinz Miklas. Contrast enhancement in multispectral images by emphasizing text regions. In *2008 The Eighth IAPR International Workshop on Document Analysis Systems*. IEEE, sep 2008.

[LMS07] Qi Li, Nikolaos Mitianoudis, and Tania Stathaki. Spatial kernel k-harmonic means clustering for multi-spectral image segmentation. *IET Image Processing*, 1(2):156, 2007.

[Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110, 2004.

[LS09] Martin Lettner and Robert Sablatnig. Spatial and spectral based segmentation of text in multispectral images of ancient documents. In *10th International Conference on Document Analysis and Recognition, ICDAR 2009, Barcelona, Spain, 26-29 July 2009*, pages 813–817. IEEE Computer Society, 2009.

[LS10] Martin Lettner and Robert Sablatnig. Higher order MRF for foreground-background separation in multi-spectral images of historical manuscripts. In David S. Doermann, Venu Govindaraju, Daniel P. Lopresti, and Premkumar Natarajan, editors, *The Ninth IAPR International Workshop on Document Analysis Systems, DAS 2010, June 9-11, 2010, Boston, Massachusetts, USA*, ACM International Conference Proceeding Series, pages 317–324. ACM, 2010.

[LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 3431–3440. IEEE Computer Society, 2015.

[LSM15] Zhipei Luo, Faisal Shafait, and Ajmal S. Mian. Localized forgery detection in hyperspectral document images. In *13th International Conference on Document Analysis and Recognition, ICDAR 2015, Nancy, France, August 23-26, 2015*, pages 496–500. IEEE Computer Society, 2015.

[LST10] Shijian Lu, Bolan Su, and Chew Lim Tan. Document image binarization using background estimation and stroke edges. *International Journal on Document Analysis and Recognition (IJDAR)*, 13(4):303–314, oct 2010.

[Mai00] Franz Mairinger. The infrared examination of paintings. *Radiation in Art and Archeometry Eds. DC Creagh, DA Bradley, Elsevier*, pages 40–55, 2000.

[MAS+15] Muhammad Imran Malik, Sheraz Ahmed, Faisal Shafait, Ajmal Mian, Christian Nansen, Andreas Dengel, and Marcus Liwicki. Hyper-spectral analysis for automatic signature extraction. In *Proceedings of the 17th Biennial Conference of the International Graphonomics Society (IGS)*, volume n/a, pages 1–4. International Graphonomics Society (IGS), 2015.

[MBC+14] Wing-Kin Ma, José M. Bioucas-Dias, Tsung-Han Chan, Nicolas Gillis, Paul D. Gader, Antonio J. Plaza, Arul-Murugan Ambikapathi, and Chong-Yung Chi. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Process. Mag.*, 31(1):67–81, 2014.

[MC10] Reza Farrahi Moghaddam and Mohamed Cheriet. A multi-scale framework for adaptive binarization of degraded document images. *Pattern Recognition*, 43(6):2186–2198, 2010.

[MC12] Reza Farrahi Moghaddam and Mohamed Cheriet. AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization. *Pattern Recognition*, 45(6):2419–2431, jun 2012.

[MC15] Reza Farrahi Moghaddam and Mohamed Cheriet. A multiple-expert binarization framework for multispectral images. In *13th International Conference on Document Analysis and Recognition, ICDAR 2015, Nancy, France, August 23-26, 2015*, pages 321–325. IEEE Computer Society, 2015.

[MGC+13] Lindsay W. MacDonald, Alejandro Giacometti, Alberto Campagnolo, Stuart Robson, Tim Weyrich, Melissa Terras, and Adam P. Gibson. Multispectral imaging of degraded parchment. In Shoji Tominaga, Raimondo Schettini, and Alain Trémeau, editors, *Computational Color Imaging - 4th International Workshop, CCIW 2013, Chiba, Japan, March 3-5, 2013. Proceedings*,

volume 7786 of *Lecture Notes in Computer Science*, pages 143–157. Springer, 2013.

[MGK+08] Heinz Miklas, Melanie Gau, Florian Kleber, Markus Diem, Martin Lettner, Maria Vill, Robert Sablatnig, Manfred Schreiner, Michael Melcher, and E. Hammerschmid. St. Catherine's monastery on Mount Sinai and the Balkan-Slavic manuscript-tradition. *Slovo. Towards a Digital Library of South Slavic Manuscripts. Sofia: Bulgarian Academy of Science, Institute of Literature., 286 (Res.).*, pages 13–36, 2008.

[Mik03] Heinz Miklas. Die slavischen Schriften: Glagolica und Kyrillica. In Wilfried Seipel, editor, *Der Turmbau zu Babel. Ursprung und Vielfalt von Sprache und Schrift. Ausstellungskatalog des Kunsthistorischen Museums*, volume 3a, pages 243–249, Wien, 2003.

[MMA14] Rafael G. Mesquita, Carlso A.B. Mello, and L.H.E.V. Almeida. A new thresholding algorithm for document images based on the perception of objects by distance. *Integrated Computer-Aided Engineering*, 21(2):133–146, Mar 2014.

[MMZ+11] Emilio Marengo, Marcello Manfredi, Orfeo Zerbinati, Elisa Robotti, Eleonora Mazzucco, Fabio Gosetti, Greg Bearman, Fenella France, and Pnina Shor. Development of a technique based on multi-spectral imaging for monitoring the conservation of cultural heritage objects. *Analytica Chimica Acta*, 706(2):229–237, nov 2011.

[MP14a] Nikolaos Mitianoudis and Nikolaos Papamarkos. Local co-occurrence and contrast mapping for document image binarization. In *2014 14th International Conference on Frontiers in Handwriting Recognition*. IEEE, sep 2014.

[MP14b] Nikolaos Mitianoudis and Nikolaos Papamarkos. Multi-spectral document image binarization using image fusion and background subtraction techniques. In *2014 IEEE International Conference on Image Processing, ICIP 2014, Paris, France, October 27-30, 2014*, pages 5172–5176. IEEE, 2014.

[MP15] Nikolaos Mitianoudis and Nikolaos Papamarkos. Document image binarization using local features and gaussian mixture modeling. *Image and Vision Computing*, 38:33–51, jun 2015.

[MS07] Nikolaos Mitianoudis and Tania Stathaki. Pixel-based and region-based image fusion schemes using ICA bases. *Information Fusion*, 8(2):131–142, 2007.

[MSMM15] Rafael G. Mesquita, Ricardo M.A. Silva, Carlos A.B. Mello, and Péricles B.C. Miranda. Parameter tuning for document image binarization using a racing algorithm. *Expert Systems with Applications*, 42(5):2593–2603, apr 2015.

[MTP+14] Dimitris Manolakis, Eric Truslow, Michael Pieper, Thomas Cooley, and Michael Brueggeman. Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms. *IEEE Signal Processing Magazine*, 31(1):24–33, jan 2014.

[NGP08] Konstantinos Ntirogiannis, Basilios Gatos, and Ioannis Pratikakis. An objective evaluation methodology for document image binarization techniques. In Koichi Kise and Hiroshi Sako, editors, *The Eighth IAPR International Workshop on Document Analysis Systems, DAS 2008, September 16-19, 2008, Nara, Japan*, pages 217–224. IEEE Computer Society, 2008.

[NGP13] Konstantinos Ntirogiannis, Basilios Gatos, and Ioannis Pratikakis. Performance evaluation methodology for historical document image binarization. *IEEE Trans. Image Processing*, 22(2):595–609, 2013.

[NGP14] Konstantinos Ntirogiannis, Basilis Gatos, and Ioannis Pratikakis. A combined approach for the binarization of handwritten document images. *Pattern Recognition Letters*, 35:3–15, 2014.

[NH98] Radford M. Neal and Geoffrey E. Hinton. A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, pages 355–368. Springer Netherlands, 1998.

[Nib85] Wayne Niblack. *An Introduction to Digital Image Processing*. Strandberg Publishing Company, Birkeroed, Denmark, Denmark, 1985.

[NMC14] Hossein Ziaei Nafchi, Reza Farrahi Moghaddam, and Mohamed Cheriet. Phase-based binarization of ancient document images: Model and applications. *IEEE Trans. Image Processing*, 23(7):2916–2930, 2014.

[O'G94] Lawrence O'Gorman. Experimental comparisons of binarization and multithresholding methods on document images. In *12th IAPR International Conference on Pattern Recognition, Conference B: Patern Recognition and Neural Networks, ICPR 1994, Jerusalem, Israel, 9-13 October, 1994, Volume 2*, pages 395–398. IEEE, 1994.

[OSK18] Sofia Ares Oliveira, Benoit Seguin, and Frederic Kaplan. dhSegment: A generic deep-learning approach for document segmentation. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, aug 2018.

[Ots79] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[PBZ+15] Joan Pastor-Pellicer, Salvador España Boquera, Francisco Zamora-Martínez, Muhammad Zeshan Afzal, and María José Castro Bleda. Insights on the use of convolutional neural networks for document image binarization. In Ignacio

Rojas, Gonzalo Joya Caparrós, and Andreu Català, editors, *Advances in Computational Intelligence - 13th International Work-Conference on Artificial Neural Networks, IWANN 2015, Palma de Mallorca, Spain, June 10-12, 2015. Proceedings, Part II*, volume 9095 of *Lecture Notes in Computer Science*, pages 115–126. Springer, 2015.

[PCN17] Xujun Peng, Huaigu Cao, and Prem Natarajan. Using convolutional encoder-decoder for document image binarization. In *14th IAPR International Conference on Document Analysis and Recognition, ICDAR 2017, Kyoto, Japan, November 9-15, 2017*, pages 708–713. IEEE, 2017.

[PG01] Ernst Pringsheim and Otto Gradenwitz. Photographische Reconstruction von Palimpsesten. *Jahrbuch für Photographie und Reproduktionstechnik 15*, 1901.

[PGN10] Ioannis Pratikakis, Basilios Gatos, and Konstantinos Ntirogiannis. H-DIBCO 2010 - handwritten document image binarization competition. In *International Conference on Frontiers in Handwriting Recognition, ICFHR 2010, Kolkata, India, 16-18 November 2010*, pages 727–732. IEEE Computer Society, 2010.

[Pri79] Claus Priesner. *Neue Deutsche Biographie 12*, chapter Kögel, Gustav, page 295. 1979.

[Pun81] Thierry Pun. Entropic thresholding, a new approach. *Computer Graphics and Image Processing*, 16(3):210 – 239, 1981.

[PWC19] Xujun Peng, Chao Wang, and Huaigu Cao. Document binarization via multi-resolutional attention model with DRD loss. In *2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019*, pages 45–50. IEEE, 2019.

[PZBB13] Joan Pastor-Pellicer, Francisco Zamora-Martínez, Salvador España Boquera, and María José Castro Bleda. F-measure as the error function to train neural networks. In Ignacio Rojas, Gonzalo Joya Caparrós, and Joan Cabestany, editors, *Advances in Computational Intelligence - 12th International Work-Conference on Artificial Neural Networks, IWANN 2013, Puerto de la Cruz, Tenerife, Spain, June 12-14, 2013, Proceedings, Part I*, volume 7902 of *Lecture Notes in Computer Science*, pages 376–384. Springer, 2013.

[PZK+19a] Ioannis Pratikakis, Konstantinos Zagoris, Xenofon Karagiannis, Lazaros T. Tsochatzidis, Tanmoy Mondal, and Isabelle Marthot-Santaniello. ICDAR 2019 competition on document image binarization (DIBCO 2019). In *2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019*, pages 1547–1556. IEEE, 2019.

[PZK+19b] Ioannis Pratikakis, Konstantinos Zagoris, Xenofon Karagiannis, Lazaros T. Tsochatzidis, Tanmoy Mondal, and Isabelle Marthot-Santaniello. ICDAR 2019 competition on document image binarization (DIBCO 2019). In *2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019*, pages 1547–1556. IEEE, 2019.

[PZY+17] Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, and Jian Sun. Large kernel matters - improve semantic segmentation by global convolutional network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1743–1751. IEEE Computer Society, 2017.

[RB05] Konstantinos Rapantzikos and Costas Balas. Hyperspectral imaging: potential in non-destructive analysis of palimpsests. In *Proceedings of the 2005 International Conference on Image Processing, ICIP 2005, Genoa, Italy, September 11-14, 2005*, pages 618–621, 2005.

[RDH18] Cristian Costa Rocha, Hilda Deborah, and Jon Yngve Hardeberg. Ink bleed-through removal of historical manuscripts based on hyperspectral imaging. In Alamin Mansouri, Abderrahim Elmoataz, Fathallah Nouboud, and Driss Mammass, editors, *Image and Signal Processing - 8th International Conference, ICISP 2018, Cherbourg, France, July 2-4, 2018, Proceedings*, volume 10884 of *Lecture Notes in Computer Science*, pages 473–480. Springer, 2018.

[RDS+14] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Fei-Fei Li. Imagenet large scale visual recognition challenge. *CoRR*, abs/1409.0575, 2014.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells III, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015.

[RKB04] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[RS08] Alejandro Ribes and Francis Schmitt. Linear inverse problems in imaging. *IEEE Signal Processing Magazine*, 25(4):84–99, jul 2008.

[SAR+20] Yaser Esmaeili Salehani, Ehsan Arabnejad, Abderrahmane Rahiche, Athmane Bakhta, and Mohamed Cheriet. MSdB-NMF: MultiSpectral document image binarization framework via non-negative matrix factorization approach. *IEEE Transactions on Image Processing*, 29:9099–9112, 2020.

[SCC14] Abdenour Sehad, Youcef Chibani, and Mohamed Cheriet. Gabor filters for degraded document image binarization. In *2014 14th International Conference on Frontiers in Handwriting Recognition*. IEEE, sep 2014.

[SCM+18] Daniel Stromer, Vincent Christlein, Andreas K. Maier, Patrick Zippert, Eric Helmecke, Tino Hausotte, and Xiaolin Huang. Non-destructive digitization of soiled historical chinese bamboo scrolls. In *13th IAPR International Workshop on Document Analysis Systems, DAS 2018, Vienna, Austria, April 24-27, 2018*, pages 55–60. IEEE Computer Society, 2018.

[SKB08] Faisal Shafait, Daniel Keysers, and Thomas M. Breuel. Efficient implementation of local adaptive thresholding techniques using integral images. In Berrin A. Yanikoglu and Kathrin Berkner, editors, *Document Recognition and Retrieval XV, part of the IS&T-SPIE Electronic Imaging Symposium, San Jose, CA, USA, January 29-31, 2008. Proceedings*, volume 6815 of *SPIE Proceedings*, page 681510. SPIE, 2008.

[SKB12] Toufik Sari, Abderrahmane Kefali, and Halima Bahi. An MLP for binarizing images of old manuscripts. In *2012 International Conference on Frontiers in Handwriting Recognition, ICFHR 2012, Bari, Italy, September 18-20, 2012*, pages 247–251. IEEE Computer Society, 2012.

[SLT10] Bolan Su, Shijian Lu, and Chew Lim Tan. Binarization of historical document images using the local maximum and minimum. In David S. Doermann, Venu Govindaraju, Daniel P. Lopresti, and Premkumar Natarajan, editors, *The Ninth IAPR International Workshop on Document Analysis Systems, DAS 2010, June 9-11, 2010, Boston, Massachusetts, USA*, ACM International Conference Proceeding Series, pages 159–166. ACM, 2010.

[SLT13] Bolan Su, Shijian Lu, and Chew Lim Tan. Robust document image binarization technique for degraded document images. *IEEE Transactions on Image Processing*, 22(4):1408–1417, apr 2013.

[Smi10] Elisa H. Barney Smith. An analysis of binarization ground truthing. In David S. Doermann, Venu Govindaraju, Daniel P. Lopresti, and Premkumar Natarajan, editors, *The Ninth IAPR International Workshop on Document Analysis Systems, DAS 2010, June 9-11, 2010, Boston, Massachusetts, USA*, ACM International Conference Proceeding Series, pages 27–34. ACM, 2010.

[SP00] Jaakko J. Sauvola and Matti Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.

[SPH+14] Carolina S. Silva, Maria Fernanda Pimentel, Ricardo S. Honorato, Celio Pasquini, José M. Prats-Montalbán, and Alberto Ferrer. Near infrared hyperspectral imaging for forensic analysis of document forgery. *The Analyst*, 139(20):5176–5184, jul 2014.

[STB07] Emanuele Salerno, Anna Tonazzini, and Luigi Bedini. Digital image analysis to enhance underwritten text in the archimedes palimpsest. *IJDAR*, 9(2-4):79–87, 2007.

[Stu07] Barbara H. Stuart. *Analytical Techniques in Materials Conservation*. John Wiley & Sons, Ltd, feb 2007.

[SZ15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[TBS09] Anna Tonazzini, Gianfranco Bianco, and Emanuele Salerno. Registration and enhancement of double-sided degraded manuscripts acquired in multispectral modality. In *10th International Conference on Document Analysis and Recognition, ICDAR 2009, Barcelona, Spain, 26-29 July 2009*, pages 546–550. IEEE Computer Society, 2009.

[TBSM19] Chris Tensmeyer, Mike Brodie, Daniel Saunders, and Tony Martinez. Generating realistic binarization data with generative adversarial networks. In *2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019*, pages 172–177. IEEE, 2019.

[TC19] Marvin Teichmann and Roberto Cipolla. Convolutional crfs for semantic segmentation. In *30th British Machine Vision Conference 2019, BMVC 2019, Cardiff, UK, September 9-12, 2019*, page 142, 2019.

[Tel04] Alexandru Telea. An image inpainting technique based on the fast marching method. *J. Graphics, GPU, & Game Tools*, 9(1):23–34, 2004.

[Ten19] Christopher Alan Tensmeyer. *Deep Learning for Document Image Analysis*. PhD thesis, Brigham Young University, 2019.

[TFF07] James Theiler, Bernard R. Foy, and Andrew M. Fraser. Beyond the adaptive matched filter: nonlinear detectors for weak signals in high-dimensional clutter. In Sylvia S. Shen and Paul E. Lewis, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIII*, volume 6565, pages 26 – 37. International Society for Optics and Photonics, SPIE, 2007.

[TJ95] Øivind Due Trier and Anil K. Jain. Goal-directed evaluation of binarization methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(12):1191–1201, 1995.

[TM17] Chris Tensmeyer and Tony Martinez. Document image binarization with fully convolutional neural networks. In *14th IAPR International Conference on Document Analysis and Recognition, ICDAR 2017, Kyoto, Japan, November 9-15, 2017*, pages 99–104. IEEE, 2017.

[TM20] Chris Tensmeyer and Tony R. Martinez. Historical document image binarization: A review. *SN Comput. Sci.*, 1(3):173, 2020.

[VJ04] Paul A. Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[War85] Daniel Wartenberg. Multivariate spatial correlation: A method for exploratory geographical analysis. *Geographical Analysis*, 17(4):263–283, sep 1985.

[WJC02] Christian Wolf, Jean-Michel Jolion, and Françoise Chassaing. Text localization, enhancement and binarization in multimedia documents. In *16th International Conference on Pattern Recognition, ICPR 2002, Quebec, Canada, August 11-15, 2002.*, pages 1037–1040. IEEE Computer Society, 2002.

[WLG18] Florian Westphal, Niklas Lavesson, and Håkan Grahn. Document image binarization using recurrent neural networks. In *13th IAPR International Workshop on Document Analysis Systems, DAS 2018, Vienna, Austria, April 24-27, 2018*, pages 263–268. IEEE Computer Society, 2018.

[WNRA16] Yue Wu, Premkumar Natarajan, Stephen Rawls, and Wael Abd-Almageed. Learning document image binarization from data. In *2016 IEEE International Conference on Image Processing, ICIP 2016, Phoenix, AZ, USA, September 25-28, 2016*, pages 3763–3767. IEEE, 2016.

[XJX+18] Wei Xiong, Xiuhong Jia, Jingjing Xu, Zijie Xiong, Min Liu, and Juan Wang. Historical document image binarization using background estimation and energy minimization. In *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 20-24, 2018*, pages 3716–3721. IEEE Computer Society, 2018.

[YHKD09] Itay Bar Yosef, Nate Hagbi, Klara Kedem, and Its'hak Dinstein. Line segmentation for degraded handwritten historical documents. In *10th International Conference on Document Analysis and Recognition, ICDAR 2009, Barcelona, Spain, 26-29 July 2009*, pages 1161–1165. IEEE Computer Society, 2009.

[ZPIE17] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, oct 2017.

Appendix

Curriculum Vitae

M.Sc. Fabian Hollaus
Technische Universität Wien
Computer Vision Lab
Favoritenstr. 9-11/193-1
1040 Wien, Österreich
Tel.: +43 1 58801 – 193182
Mail: holl@cvl.tuwien.ac.at

Personal Dates:

Nationality:	Austria
Date of Birth:	25.10.1983
Place of Birth:	Mittersill

Education:

2007-2011	TU Wien, Graduation M.Sc., Computer graphics and digital image processing, Thesis title: Automated Inpainting of Unknown Palimpsest Regions
2003-2007	TU Wien Media Informatics (Bachelor)
1994-2002	Bundesrealgymnasium Zell am See

Work Experience:

Since 2011	TU Wien, Computer Vision Lab Project assistant in the projects: 'The Enigma of the Sinaitic Glagolitic Tradition' ‚Centre of Image and Material Analysis in Cultural Heritage' 'Detection and Visualization of Ordnance Risks' 'Digitization and Information processing for Digital Humanities'

Publications

Major peer reviewed publications (sorted by year of publication):

[HGS12] Fabian Hollaus, Melanie Gau, and Robert Sablatnig. Multispectral image acquisition of ancient manuscripts. In Progress in Cultural Heritage Preservation - 4th International Conference, *EuroMed 2012*, Limassol, Cyprus, October 29 - November 3, 2012. Proceedings, volume 7616 of Lecture Notes in Computer Science, pages 30–39. Springer, 2012.

[HGS13] Fabian Hollaus, Melanie Gau, and Robert Sablatnig. Enhancement of multispectral images of degraded documents by employing spatial information. In 12th International Conference on Document Analysis and Recognition, ICDAR 2013, Washington, DC, USA, August 25-28, 2013, pages 145–149, 2013.

[HDS14] Fabian Hollaus, Markus Diem, and Robert Sablatnig. Improving OCR accuracy by applying enhancement techniques on multispectral images. In 22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August 24-28, 2014, pages 3080–3085. IEEE Computer Society, 2014.

[HDS15] Fabian Hollaus, Markus Diem, and Robert Sablatnig. Binarization of multispectral document images. In George Azzopardi and Nicolai Petkov, editors, Computer Analysis of Images and Patterns - 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II, volume 9257 of Lecture Notes in Computer Science, pages 109–120. Springer, 2015.

[DHS16] Markus Diem, Fabian Hollaus, and Robert Sablatnig. MSIO: multispectral document image binarization. In 12th IAPR Workshop on Document Analysis Systems, DAS 2016, Santorini, Greece, April 11-14, 2016, pages 84–89. IEEE Computer Society, 2016.

[HDS18] Fabian Hollaus, Markus Diem, and Robert Sablatnig. Multispectral image binarization using GMM's. In 16th International Conference on Frontiers in Handwriting Recognition, ICFHR 2018, Niagara Falls, NY, USA, August 5-8, 2018, pages 570–575, 2018.

[HBS19] Fabian Hollaus, Simon Brenner, and Robert Sablatnig. CNN based binarization of MultiSpectral document images. In 2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, September 20-25, 2019, pages 533–538. IEEE, 2019.