Technical Report
CVL-TR-2

On the Combination of Spatial and Spectral Features for Image Restoration

Martin Lettner and Robert Sablatnig

Computer Vision Lab
Institute of Computer Aided Automation
Vienna University of Technology

September 22, 2010

Abstract

The application of multispectral imaging is a well known method for the analysis and digitization of decayed manuscripts. The main advantage in analyzing multiple spectral ranges, including the ultraviolet and infrared range, is the additional information which is invisible to the human eye. This report focuses on two aspects for image restoration of multispectral images of degraded documents. The first problem relates to a general enhancement of the readability. Due to mold, air humidity, water, etc., parchment and text may partially be damaged and consequently hard to read. The proposed methodology is based on a spatial and spectral analysis and the main advantage of the method is that especially text regions are considered for enhancement. The second part of this work deals with a robust method for the separation of text from background. The proposed statistical framework incorporates spatial and spectral features in the context of a higher-order Markov Random Field. Spectral information is extracted from the spectral behavior of the multispectral images and the spatial dependencies are captured by means of stroke properties. Providing a strong local minimum and the potential of using arbitrary potential functions, a modified version of belief propagation is applied for the optimization of the higher order model. The proposed method requires no training and is independent of script, size, and style of characters.

Contents

1 Introduction **1**
- 1.1 Foreground-Background Separation in Multispectral Images . . . 2
- 1.2 Image Enhancement . . . 4
- 1.3 Thesis Structure . . . 5

2 Related Work **7**
- 2.1 Foreground-Background Separation and Document Image Restoration . . . 7
  - 2.1.1 Global and Adaptive Thresholding . . . 8
  - 2.1.2 Binarization Based on Color Clustering . . . 9
  - 2.1.3 Binarization Based on Spatial and Spectral Information . . . 10
  - 2.1.4 Probabilistic Approaches . . . 10
  - 2.1.5 Summary of Foreground-Background Separation in DIA . . . 12
- 2.2 Higher-Order Markov Random Fields . . . 12
- 2.3 Innovative Aspects . . . 15
- 2.4 Summary . . . 16

3 Multispectral Imaging **17**
- 3.1 Multispectral Images . . . 17
  - 3.1.1 Illumination and Electromagnetic Radiation . . . 18
  - 3.1.2 Multispectral Analysis of Ancient Manuscripts . . . 18
- 3.2 *Missale Sinaiticum (Sin. Slav. 5/N)*: Acquisition Setup . . . 20
- 3.3 Post-Processing . . . 22
  - 3.3.1 Image Registration . . . 23
  - 3.3.2 Image Enhancement . . . 24
  - 3.3.3 Image Enhancement Results . . . 28
  - 3.3.4 Discussion . . . 29
- 3.4 Summary . . . 31

4 Probabilistic Graphical Models **32**
- 4.1 Fundamentals . . . 32
- 4.2 Markov Random Fields . . . 33
  - 4.2.1 Prior Model $\Pr(\mathbf{x})$ . . . 35
  - 4.2.2 Likelihood $\Pr(\mathbf{y}|\mathbf{x})$ . . . 36
  - 4.2.3 Posterior Energy $\Pr(\mathbf{x}|\mathbf{y})$ . . . 36
- 4.3 Conditional Random Fields . . . 38

4.4 Higher-Order Models . . . 39
4.5 Summary . . . 41

5 Energy Minimization **42**
5.1 Iterated Conditional Modes . . . 42
5.2 Energy Minimization using Graph Cuts . . . 43
5.2.1 Pairwise Based Prior . . . 45
5.2.2 Solving Energies with Higher-Order Cliques . . . 47
5.3 Summary . . . 48

6 Foreground-Background Separation based on Higher-Order MRFs **49**
6.1 Higher-Order Energy Function . . . 49
6.2 Potential Functions . . . 50
6.2.1 Unary Potentials . . . 50
6.2.2 Pairwise Potential Function . . . 52
6.2.3 Higher-Order Potentials . . . 52
6.3 Belief Propagation . . . 53
6.3.1 Standard Belief Propagation for Pairwise Models . . . 54
6.3.2 Belief Propagation for Higher-Order Models: BP^n . . . 55
6.4 Summary . . . 57

7 Experiments and Results **59**
7.1 Evaluation Method . . . 59
7.2 Test Data . . . 61
7.3 Energy Function and Weighting Parameter . . . 62
7.4 Overview of the Binarization Methods . . . 62
7.5 Influence of the Higher-Order Stroke Model . . . 63
7.5.1 Missale Sinaiticum . . . 63
7.5.2 DIBCO 2009 Images . . . 69
7.5.3 General Aspects . . . 75
7.6 Synthetic Data . . . 77
7.7 General Comparison of the Methods . . . 86
7.7.1 Missale Sinaiticum . . . 86
7.7.2 DIBCO 2009 Images . . . 86
7.8 Summary of Results and Discussion . . . 95

8 Conclusion and Outlook **98**
8.1 Our Contribution . . . 100
8.2 Outlook . . . 101

A Acronyms and Symbols **103**

B List of Notation **105**

Bibliography **106**

Chapter 1

Introduction

MultiSpectral Imaging (MSI) has proven a capable technique for the analysis and preservation of ancient or damaged documents. Especially images in the UltraViolet (UV) and InfraRed (IR) light range reveal additional information such as latent texts or faded passages [23].

The multispectral images obtained are not only used for visual inspections, they also serve as input data for a computer based analysis, as is the case in enhancing the underwritten text in the *Archimedes palimpsest*[1] [24] or, for instance, to distinguish between the underwritten and the new text in the same manuscript [90].

While most of the recent studies in MSI of historical manuscripts attempt to separate the different writings in palimpsests [90, 85], the topic of this thesis considers a more general restoration process of partially damaged and decayed manuscripts. Two approaches for the examination of multispectral images are proposed in this thesis. The main topic deals with Foreground-Background Separation (FBS) in document images. FBS constitutes a major part in Document Image Analysis (DIA) and enables the use of simplified analysis techniques for subsequent algorithms like Optical Character Recognition (OCR) or page layout segmentation [45]. Therefore, we propose a robust binarization method which is especially designed for the multispectral image data of document images. The requirements for the proposed approach include a general applicability, i.e. independence of hand written or machine printed text, and a certain robustness, i.e. independence of image noise.

The second goal of this thesis follows a general legibility enhancement of damaged documents. Foundation for both methods is the simultaneously utilization of spatial and spectral features of the multispectral image data. In both cases we incorporate the full range of multispectral information.

[1] A palimpsest is a document which was rewritten after the first text had been erased. It was a common practice in medieval ecclesiastical circles to rub out or wash off the writings on parchment, in order to prepare it for new texts.

1.1 Foreground-Background Separation in Multispectral Images

An important step in DIA is image binarization, which divides page content like characters, ligatures, or decorations from the background [29]. FBS is, besides noise reduction or skew correction, a fundamental pre-processing step [17], which enables the use of simplified analysis techniques in subsequent computations [45]. The resulting binary images may serve as input data for

1. character segmentation[2] [16],
2. Optical Character Recognition (OCR) [76] or local approaches for script identification [96],
3. the restoration of broken characters [3],
4. line segmentation [68], or
5. the estimation of the stroke trace or pen trajectory [62].

Consequently, binary images are the prevalent input for consecutive steps in DIA and fast and accurate document image binarization is becoming increasingly important [98], since errors like touching or broken characters may lower, for instance, the performance of OCR systems [41].

Foreground-background separation in digital representations of degraded documents is not an easy task [29]. Even though document image binarization has been studied for many years, the thresholding of historical document images is still an unsolved problem due to the high variation within the document foreground and background [98]. Specific reasons range from shortcomings in the image acquisition setup including illumination, camera setup, etc., to degradations caused by manuscript decay. The latter may include a non-uniform appearance of the writing and the background, a blur of the background, a faded ink, mold, water stains, or humidity [74].

MSI has already provided promising results for the analysis of decayed documents [26]. Nevertheless, in most of the cases, FBS for damaged manuscripts is based on sophisticated image processing techniques applied on gray level images [95].

In this thesis, we propose a robust method for FBS in multispectral images of degraded documents. The innovation of the proposed method is the simultaneously combination of spatial and spectral information of the multispectral image data in order to improve the segmentation performance. This combination is of great importance to correct and refine segmentation errors [102]. Recent studies already exploit the combination of spatial and spectral components but treat the two components successively, e.g. [28] or [107]. In contrast to previous studies we arrange the combination of spatial and spectral components simultaneously. Therefore, we utilize a Markov Random Field (MRF) and consequently

[2]Character segmentation seeks to decompose an image of a sequence of characters into subimages of individual symbols.

a Conditional Random Field (CRF), which provide a probability theory for analyzing spatial and contextual dependencies [67].

Since high quality text is not available for the manuscript given, a specific training of the prior in the MRF is, as proposed in the study from [15], not possible. In contrast to approaches with previously learned prior models, we resort to a general adaptive prior model, making the approach independent of machine or handwritten text, as well as script, font, style, or the size of characters. Therefore, the prior model includes spatial correlations of characters and of strokes, respectively, and will be expressed within a *stroke model*. These spatial dependencies are modeled by incorporating a higher-order MRF model. Since the stroke properties and the Gaussian parameters for the imaging model are evaluated automatically, the proposed method requires no training data and is applicable as conventional binarization techniques, like Otsu's method [80] or adaptive image binarization proposed by [91].

For the optimization of the MRF energy equation, we propose to use local methods like Belief Propagation (BP) [27], which has a strong local minimum property [100] and works for arbitrary potential functions [84]. BP is generally applicable and independent of the graphical model and the form of potentials. We will show in the results, that local methods are superior to global optimization methods for the application given.

Originally designed for models with pairwise connections we adapt the standard formulation of BP and include higher-order functions [51] to incorporate the stroke properties. Following the higher-order potential functions P^n proposed by [50], the proposed algorithm will be referred to as $\mathrm{BP^n}$. The higher-order potential functions are included in a new formulation of the message update rule from the standard BP algorithm.

The proposed FBS process is based on three innovations:

1. spatial and spectral based FBS in document images,
2. higher-order MRF for incorporating stroke characteristics, and
3. local inference for local optimization problems ($\mathrm{BP^n}$).

Figure 1.1 illustrates the idea and innovative points within a diagram. The top circle of our philosophy refers to the simultaneously combination of spatial and spectral features for FBS in digital document images to correct and refine segmentation errors. Spectral features correspond to the spectral component of the multispectral image and spatial features incorporate stroke characteristics by means of the stroke width. The spatial components are incorporated within a higher-order MRF, shown in the lower left circle. The MRF model for FBS will be presented in Section 6.2. Finally, we propose to use a local inference method for the separation of characters. Therefore, we introduce an adaptation of the standard BP algorithm, which will be presented in Section 6.3.2.

The proposed method is tested on a set of folios from a historic manuscript and the results are compared to state of the art FBS methods like serialized k-means clustering [66] or adaptive binarization [91]. Furthermore, we compare the proposed method to the algorithm from [98], which is an advanced version of the best algorithm from the 2009 Document Image Binarization Contest (DIBCO 2009). This competition was organized within the framework of the Tenth International Conference on Document Analysis and Recognition (ICDAR 2009) in Barcelona, Spain.

The evaluation metric is based on an objective computer algorithm, the precision and recall rate [29]. Therefore, we generated ground truth data manually for a set of degraded documents. Further test are applied on images from the DIBCO 2009 where ground truth data is already available. Our systematic evaluation shows that the combination of spatial and spectral features offers a robust method for FBS especially in the presence of noise or when degraded documents with low contrast are given.

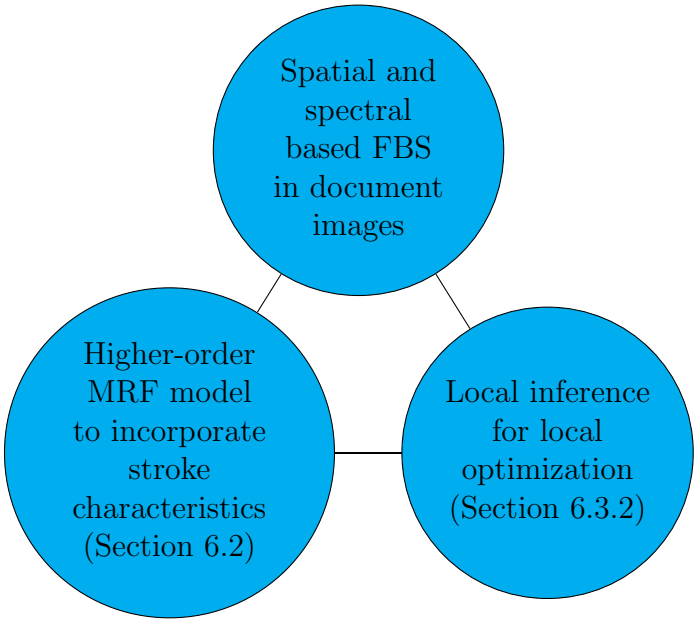


Figure 1.1: Our proposed method for FBS in document images is based on spectral observations and stroke features. Concerning the spectral features, the method can incorporate the full range of spectral information from multispectral images, but panchromatic images may also serve as input data. The second type of features, spatial characteristics or stroke properties, are based on the stroke width. Since the spatial information requires an extended neighborhood, we apply a higher-order MRF to incorporate spatial and spectral features. For statistical inference in the higher-order model, we propose to apply a local method since characters in document images are distributed locally.

1.2 Image Enhancement

To support philological studies for the transcription of partially undecipherable manuscripts, the second goal of this study is a general enhancement of the multispectral image data. In contrast to previous studies which focus on the enhancement of the underwritten texts in palimpsests, e.g. [90], our preference is a general enhancement of the readability in multispectral images of degraded manuscripts.

Multispectral image data is often highly correlated [86]. The Principal Component Analysis (PCA) is a well known method to remove this redundancy [22, 86]. For instance, [24] applied PCA to produce pseudo-colored images of the multispectral data from the *Archimedes Palimpsest*.

To enhance the legibility, we focus on the multispectral image data and the spectral signature as well as the spatial correlation in the spectral images. Multivariate Spatial Correlation (MSC) constitutes an alternative approach to PCA in order to remove this correlation. Originally designed for the analysis of geographical data, [108] showed the robustness of the method in the presence of noise. The innovation for the adaptation to image enhancement in degraded documents is the possibility to remove the correlation with a simultaneous emphasis on text regions.

In Section 3.3.2 we show some experiments and compare the output of the MSC to the results of PCA. The evaluation of the two different methods is executed on multispectral images of an medieval Slavonic manuscript, the *Missale Sinaiticum (Sin. Slav. 5/N)*, and the assessment is based on a human reader. Furthermore, we evaluate the general benefit of the multispectral images by counting the number of characters detected in the original data set (i.e. the RGB images) and in the multispectral images after enhancing with the proposed method.

1.3 Thesis Structure

Chapter 2 provides an overview of existing methods for FBS in DIA. Since previous categorizations cover especially algorithms for panchromatic images, we arrange the individual methods in a new categorization. The second topic in Chapter 2 concerns MRFs and provides an overview on recently published methods. Main focus is based on inference in higher-order models which are only present for a short time.

Chapter 3 provides background information on MSI, including physical background, acquisition methods and acquisition setup. Furthermore, we show the results of the digitization of a historic manuscript, the so called *Missale Sinaiticum*, which constitutes the main data set for our investigations including image enhancement and FBS. Post-processing methods including image registration and the proposed enhancement algorithm are presented in the third part of this chapter.

Chapter 4 explains the formulation of MRFs starting from Bayesian classification, to Maximum A Posteriori estimation and CRFs. Higher-order models are explained in the second part of the chapter.

Chapter 5 explains standard methods for probabilistic inference in MRFs including a local method and a global method.

Chapter 6 introduces the proposed approach for FBS in digital document images. The first part shows the potential functions used for unary potentials, pairwise potentials, and the higher-order functions. Main focus is set on parameter free functions to avoid training. In the second part of the chapter, we introduce an adapted version of BP to incorporate higher-order functions.

Chapter 7 presents the experiments based on three data sets: the first one includes digital representatives of the *Missale Sinaiticum* in multispectral manner, the sec-

ond set constitutes some representative images from a previously organized image binarization contest, and the third set constitutes synthetic images.

Chapter 8 concludes this dissertation and gives an outlook to some further improvements of the proposed method.

Chapter 2

Related Work

Converting digital document images into binary representations and consequently separating text from background is a fundamental step in DIA [30]. This fact is reflected in a large number of publications in this research area [28]. In the following chapter we provide an overview of methods for FBS in document images. The main focus is based on approaches using MRFs. Since the spatial information in our FBS process is modeled with higher-order random fields, we present recent studies on computationally efficient algorithms for inference in higher-order MRFs and respectively CRFs.

2.1 Foreground-Background Separation and Document Image Restoration

Efficient binarization methods for FBS in DIA have been a subject of intense research during the last several years [28]. [28] divides FBS algorithms into the following three categories :

1. global thresholding,
2. adaptive thresholding, and
3. color clustering.

Traditional methods for image binarization are primarily based on panchromatic images, in which global or adaptive thresholding methods describe efficient methods. With the widespread development of input devices for color images, documents are now digitized preserving the color information. Thus, FBS in color document images has gained considerable attention in recent years [28].

Nevertheless, the categorization above neglects two major approaches. First of all, some methods are based on sophisticated algorithms considering not only the spectral component (i.e. gray level or color information), but also spatial components, as is the case in the Gabor-based approaches from [102]. Such approaches will be referred to as spatial / spectral methods.

Secondly, methods based on a probability theory, like naive Bayes classifier or MRFs, can in turn not be assigned to one of the three categories mentioned above. Thus, an

additional category for algorithms based on a probability theorem will be added to the categorization from [28]. Including the three primarily categories, our categorization includes the two following subjects:

4. spatial / spectral approaches and

5. probabilistic approaches.

The following sections give an overview of recent work devoted to global or adaptive thresholding, color clustering, or on combined approaches for FBS. Furthermore, we provide, to the best of our knowledge, a complete overview on probabilistic approaches using MRFs for FBS.

2.1.1 Global and Adaptive Thresholding

The objective of binarization is to automatically choose a threshold that separates foreground and background information. Global and adaptive thresholding techniques use a threshold T to distinguish between foreground and background in a panchromatic image I:

$$BW(i,j) = \begin{cases} 1 & \text{if } I(i,j) \geq T \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Global thresholding selection methods assume that the gray-level histogram is bimodal [80]. In adaptive or local thresholding, the threshold values are determined locally and the threshold is computed for separate pixels or image regions based on local statistics [91, 30]. Overview papers comparing different global and adaptive methods can be found in [36, 58, 41, 29].

Depending on the quality of the original image, the resulting binary image may include gaps in lines, ragged edges on region boundaries, missing characters, or extraneous pixels in foreground or background regions [45]. Cases including such circumstances are caused, for instance, by non-uniformly illuminated pages during data capture, low contrast between text and background, or the preservation condition of historical or damaged manuscripts.

In the case of old and degraded documents local methods outperform global methods [30]. For instance, [91] developed a method for adaptive document image binarization by determining an individual threshold for each pixel. The main drawback of this window-based thresholding approach is that the thresholding performance depends heavily on the window size and hence the character stroke width [98].

For the recognition of highly degraded characters [104] use a Wavelet filtering stage for denoising, followed by an extraction of individual text lines which are separately converted into a binary image by adaptive thresholding. [7] proposed a multistage approach based on a initial global threshold which suffices for noise free characters. Then the document characters are evaluated and an accurate local method is invoked only on noisy characters. [30] developed an adaptive degraded image binarization algorithm following several distinctive steps: Wiener filtering, a rough estimation of the foreground region, the calculation of the background model, and post-processing methods like shrink and swell filtering to improve the quality. [94] presented an approach to remove the uneven

background from historical documents (background light normalization) allowing a global threshold for a complete page. To accomplish this, they use an adaptive linear function to approximate the uneven background. The approach is similar to background subtraction as specified by [58]. [70] estimated the shading of the background by fitting a least square polynomial surface to a given document image. Combining the gray values of pixels and the polynomial surface allows to directly threshold the observed image.

However, these binarization methods proposed for gray-scale documents have not been well tested or extended for color documents [28]. Moreover, most of these approaches combine different types of image information and domain knowledge and are often complex and time consuming [98, 95].

A very recent and high performing binarization method especially designed for historical document images was presented by [98]. Applied to a document image, this technique constructs a contrast image and then detects the high contrast image pixels, which usually lie around the text stroke boundary. The document text is then segmented by using local thresholds that are estimated from the detected high contrast pixels within a local neighborhood window. The technique was tested on the DIBCO 2009 dataset and showed superior performance [98]. A preliminary version of this method showed the best performance in the DIBCO 2009 competition and has beaten thirty-five other competitors [29].

2.1.2 Binarization Based on Color Clustering

Text or character extraction techniques proposed for color document images are based on clustering or color segmentation and exclusively consider the color or spectral component [28]. The methods perform clustering in the 3D feature space [71] and the segmentation process is based on the assumption that characters are printed in a visually single color as seen in most text passages in color documents [40].

For instance, [97] extract blobs consisting of similar color pixels by clustering. A subsequent Support Vector Machine (SVM) classifies these blobs into character or background patterns based on several textural features. An adaptive method based on the k-means algorithm was proposed by [66]. This approach was developed for digitized ancient manuscripts and is based on a serialization of the k-means algorithm. The clustering is applied sequentially within a sliding window and the algorithm reuses information about the clusters from previous classifications to adjusts the centers. This reutilization adapts the classifier to local modifications of colors or varying illumination conditions.

[40] propose a color segmentation for the L*a*b* color space by initially observing the color distribution in character areas. This method extracts representative colors based on a histogram analysis of the color space. [28] propose a chain of processing steps for FBS in low quality images applicable for color and panchromatic images. An initially connected component labeling captures spatially connected pixels of similar color. After the identification of the dominant background color, rectangular blocks are formed around the identified uniform background regions. Then, bicolor clustering (k-means) is performed within the rectangular blocks to extract the foreground regions.

By means of the preceding connected component labeling algorithm the method proposed by [28] analyzes spatial connectivity of similar text and background pixels addi-

tionally to color clustering. This approach is already a step forward to the combination of spatial and spectral features. Approaches which are especially based on the combination of spatial and spectral information will be highlighted in the following section.

2.1.3 Binarization Based on Spatial and Spectral Information

Although the combination of spatial and spectral components for FBS was exploited in recent studies [28, 107, 31], spatial and spectral components are examined in a successive way. [28] use a preceding connected component labeling to capture spatially connected similar color pixels. Afterwards, they divide the image into blocks to perform local clustering. [107] combined edge information, watershed transformation, and clustering for character segmentation in color images. The watershed transformation is executed on the edge images to obtain basins of uniform color which are afterwords divided into text and background by a clustering scheme. The incorporation of edge detection is also examined by [31]. The method is a combination of several techniques including edge detection, pre-processing, a combination of several state-of-the-art binarization methodologies, and post-processing resulting in a number of parameters. However, edge detection is difficult in old document images since the stroke intensity is very variable and the edges are very weak [75].

[75] propose a method for the restoration of single sided low quality document images based on a multi-level classifier. The multi-level classification provides a robust and adaptive labeling method and includes four meta levels: pixel, regional, content, and global. For the pixel level, the gray value is used for classification. The regional level determines whether a pixel belongs to a boundary in the image or not. At the content level, stroke related classifiers like the proposed stroke map or stroke profile are introduced. Finally, a global level is developed to provide an overall estimate of the background.

[102] use color clustering and subsequently a Gabor-based filter to combine color with spatial information. The Gabor-based filter simultaneously handles spatial information to locate characters in the image, and frequency information to use illumination variation to detect character edges. This method is already an attempt to combine spatial and frequency information simultaneously. But [102] note in the conclusion that images with very low resolution and poor contrast cannot be segmented by the proposed method and refer to supervised methods. However, supervised models require training data and lack of general applicability. Thus, methods with trained prior are not independent of script or characters size.

2.1.4 Probabilistic Approaches

Approaches based on a probabilistic model estimate the probability of an event on the basis of known data. Therefore, these approaches require a training phase and respective training data. Probabilistic approaches for FBS include, for instance, Independent Component Analysis (ICA) to separate different layers in digital documents [104, 90] or a Gaussian distribution and expectation-maximization for degraded document enhancement [1]. However, ICA ignores the two dimensional relations between gray levels on the image which is considered to constitute a one-dimensional signal [18].

There have also been some attempts using MRFs for FBS in DIA. For instance, one of the first approaches using MRFs for FBS in document images was presented by [20] to binarize car license plates from gray scale images using 2×2 cliques. [103] use 3×1 and 1×3 cliques to restore bimodal text from low resolution samples. Both approaches used an MRF to model the prior for noise removal and to augment small straight strokes.

In order to include more complex models, image patches modeling prior knowledge by considering spatial constraints are included [27, 15]. However, these kind of probabilistic approaches require training. Consequently, the algorithms are not independent of script, size, or font in machine printed text and cannot model handwritten text [15].

The patch based topology was introduced by [27] and divides both the image and the scene into patches. The Markov network is based on a graph based topology where nodes are connected by lines which indicate statistical dependencies. Each scene patch is connected to its corresponding image patch and to its spatial neighbors. The scene patches have to be learned from training data and are tuned to cover common shapes or details from individual characters [15, 38, 6, 59]. An example of patches for image restoration is given in Figure 4.5. The figure shows 114 representatives of shared patches obtained from clustering. For the training of these patches, completed binarized images of high quality are decomposed into 5×5 regions. Using k-means clustering, the dominant cluster centers of the patches are taken as representatives.

[55] modeled the prior as a generalized Potts model to produce smooth labeling results by considering a 4-connected neighborhood system. A more sophisticated model exploiting the properties of text characters was presented by [112]. The observation model within the Bayesian framework depends on the mean and variance of the gray value distribution of text and background, and the prior is defined by sixteen 4×4 cliques. The individual cliques and potentials are able to repair damages in characters, text curves, and serifs. However, the results do not improve existing methods and the MRF model produces also errors like extra serifs [36].

A learning based method for restoring and recognizing images of digits at the same time was presented in [37, 38]. The study covers the restoration (binarization) and recognition of blurred images of license plates, based on a multilayer MRF containing separate layers for recognition and restoration. The method requires a priori knowledge of each object category which is straightforward for the recognition of digits. The main innovation of the study is that restoration and recognition work without prior segmentation of individual digits by encoding the information on spatial relationships between patches (segmentation) and their semantic meaning (recognition). However, this scheme cannot be directly applied to unconstrained handwriting because of the larger number of classes and the low performance of existing handwriting recognition algorithms [15].

[14, 15] propose a pre-processing approach for handwritten carbon forms using MRFs. In addition to binarization using histogram thresholding, the prior, represented by an MRF under local dependence assumption, provides constraints of connectivity and smoothness. The prior probability is learned from a high quality training set of already binarized images. The observation probability density is learned from the gray level histogram of the input image. The patch based topology is based on 5×5 nonoverlapping blocks. Figure 4.5 shows the whole set of image patches which are trained from high quality images.

[79] propose an MRF approach for historic typewritten document images. In contrast to the methods stated above, the proposed approach employs an off-line estimation process and the patch initialization is data dependent. The patches are trained as proposed in [14] and the probability distribution of the foreground and the background is learned from labeled training data. Variations in paper quality or color limit the accuracy of the results.

Another approach for the binarization of seriously degraded documents based on an MRF model was presented by [59]. The model parameters are learned from training data or computed using heuristics. The prior model represents the contextual information introduced by the Markov model. Potential functions $V_d(z)$ based on simple heuristic rules are associated for each clique configuration, which is defined by the sum of two different terms $V_d^1(z)$ and $V_d^2(z)$. The first one is introduced to remove noise and the latter is used to improve character connectivity based on simple pairwise connections. A similar approach, where the text model is learned from the degraded document itself is proposed by [6] making the separation independent of script, font, and style, but require a large training set. A first stage of the algorithm involves the estimation of ideal prototype patches. In a second stage, the restoration of degraded or broken characters is based on the estimation of the most likely set of patches (from the set above) that generates the observed patches using an MRF. The requirement of a large dataset needed is disadvantageously and the method focuses on broken characters.

2.1.5 Summary of Foreground-Background Separation in DIA

An overview of the reviewed literature is given in Table 2.1. Each of the categories defined includes some representative studies, advantages and disadvantages.

As stated by [102], FBS in images with low resolution and poor contrast cannot be segmented with unsupervised methods. However, supervised models require prior knowledge and consequently training data which makes the approaches not generally applicable. Given all different types of text whether handwritten or typewritten in all possible variations of font, size, style, color, etc., it is difficult or nearly impossible to create an exact model of text which fits all virtually possible observations [112]. Hence, we restrict our method to a low level model by incorporating stroke properties. In contrast to the first order MRFs used for the patch based topology in the studies above, we resort to higher-order models to define spatial relationships of strokes. Higher-order models have been avoided for a long time due to their computational complexities [49]. However, [50] and [83] present computationally efficient methods for higher-order MRF models, which will be explained in the following section.

2.2 Higher-Order Markov Random Fields

MRFs provide a probability theory for analyzing spatial or contextual dependencies [67]. The practical use of MRF models is based on the Hammersly and Clifford theorem stating the equivalence of MRFs and Gibbs distribution [8]. The MRF-Gibbs equivalence theorem points out that the joint distribution of an MRF is a Gibbs distribution, the latter taking

Table 2.1: Summary of FBS in DIA, category 1-3 from [28].

Category	Approach with references to some important works	Pros and cons
Global thresholding (panchromatic images)	Assume histogram is bimodal [80]	Simple to implement and often effective, unable to find a global threshold in nonuniform illumination, noise, and degraded documents
Local thresholding (panchromatic images)	Compute threshold for each pixel or regions based on local statistics [91], background subtraction [58]	Several methods are computationally expensive, combination of several methods, requires individual parameters
Color clustering (color images)	Color clustering like k-means is involved [66], color segmentation based on L*a*b* color space [40]	Mostly designed for characters with similar colors, breaks in degraded documents
Spatial and spectral, composite (color images)	E.g. clustering and Gabor filter [102], combination of edge information and clustering [107], multi-level classification [75]	Computationally expensive, treats color and spatial components successively, mostly designed for characters with similar colors, combination of several (pre-processing) methods, requires individual parameters
Probabilistic approaches (particularly for panchromatic images)	MRF [15, 112], ICA [104], restoring and recognizing digits [38]	Training and high quality data required, dependency of script, size, etc., reconstruction of degraded characters possible

a simple form providing mathematically tractable means for statistical image analysis [67, 32]. Applications in computer vision like color image segmentation [46], stereo matching [101, 25], image inpainting [88], or image denoising [57] can be elegantly expressed by means of MRF models. The effectiveness of MRFs in DIA was already mentioned in the previous section.

MRFs are modeled within a probabilistic graphical framework, where the random variables are represented as nodes. Links between the nodes express probabilistic relationships. The sites are related to one another via a neighborhood system to model context dependent entities. The probability of one site in the MRF model depends on its neighbors. However, MRF priors typically exploit only small neighborhood systems, typically a standard 4-connected neighborhood systems, which limits the expressiveness of the models and only crudely captures the statistics of natural images [88]. Due to its often extreme computational demands, traditional inference algorithms are computationally expensive for higher-order cliques and their usage has long been avoided [51]. An overview of energy minimization methods for pairwise MRF is given by [100].

The most common methods for probabilistic inference are BP and Graph Cut (GC). BP is a message passing algorithm for energy minimization in graphical models. Originally designed for graphs without loops [81], [110] proves the correctness of local probability propagation in graphical models with loops. [25] propose an efficient version of the standard BP in pairwise models to compute the message update in linear time.

A rather simple approach to utilize higher-order potentials in BP is proposed for the application of scene text detection by [116]. The message update rule is expressed by the combination of the messages from two neighboring nodes to one message. The clique potential function involves only three nodes and is based on the minimum angle or maximum color distance of the observations.

The idea of formulating image priors over a large neighborhood as higher-order MRF was also proposed by [87, 88]. The main idea of their study is a framework for learning expressive image priors, in the size of 2×2 or 5×5, to capture the statistics of natural scenes. The resulting Field of Experts (FoE) models the prior probability of an image. The model is trained on a standard database of natural images and is applied to image inpainting and image denoising.

[57] propose approximation methods for BP to make inference possible in higher-order MRFs. The method is based on models using the FoE [88] framework. For higher order MRFs (2×2) they use an adaptive state space to handle the increased complexity.

[115] propose the Generalized Belief Propagation (GBP) algorithm, which considers the message passing mechanism between clusters of nodes. The problem of how to form the appropriate clusters remains. Still first-order constraints are modeled between the clusters and sometimes it is more desirable to retain the original graph structure [43]. An extension of this study is presented by [78]. They avert the regular pixel lattice of image graphs, which results in a highly connected graph with clusters and fewer nodes. Their hypergrah has clusters as nodes and hyperedges exist between the generated hypergraph. Since standard BP is not guaranteed to converge in this graph, they propose $(\mathrm{BP})^2$, which is an extension of GBP, in order to propagate messages between clusters. [43] generalize the BP algorithm to consider second-order constraints for object localization. Instead of clustering graph nodes, they extend the message variables to consider relations between

three nodes by means of hyperedges.

[2] explore a link between pairwise and higher-order models. They show how a higher-order potential energy can be efficiently transformed into polynomial form in order to subsequently reduce the higher-order polynomial to a specific quadratic function. [54] introduce a higher-order MRF optimization framework based on a master-slave message passing algorithm. It relies on the idea that higher-order MRFs can be decomposed into MRF subproblems.

Finally, [83] propose a technique to compute BP messages in time linear with respect to clique size. The graphical model is based on a factor graph, in which each potential function is represented by a factor node, see also [84]. However, a 5×5 clique model, for instance, has maximal cliques with 25 variables. Every factor node in a factor graph representation of an MRF, on which the message passing scheme is based, subsumes 25 pixels. Using major label sets makes it intractable to store the beliefs and messages [88].

Traditional inference algorithms based on BP are computationally expensive for higher-order cliques [50]. [50] provide a characterization of energy functions defined by cliques of size three (P^3) or more (P^n). They prove that the GC algorithms, α-expansion and $\alpha\beta$-swap, for this class of higher-order functions can be computed in polynomial time. They introduce a new family of higher order potential functions, referred to as the P^n Potts model, and show that the optimal α-expansion and $\alpha\beta$-swap moves can be computed by solving an s/t mincut problem.

2.3 Innovative Aspects

The content of this thesis is the development of a robust method for FBS in multispectral images of degraded documents. The approach is based on three main contributions.

FBS is applied particularly on panchromatic images and either based on spectral or spatial features. For instance, clustering algorithms work only in the spectral feature space while our segmentation model considers the spatial relationship of pixels in the image domain additionally. Only a few benefit from the combination, but utilize the combination one after another, e.g. [107, 102]. We are going to treat the combination of *spatial and spectral features* simultaneously within the framework of an MRF.

In order to incorporate spatial features, we introduce the *stroke model* of fixed shape which models the spatial correlation of strokes and respectively characters. This stroke model is incorporated within a *higher-order* MRF. The main advantage of the proposed method is that a preceding training and the requirement of high quality training data is avoided. This allows a general applicability, independent of font, style, or size of characters. The method is especially profitable when historical manuscripts are applied, where the paper or text quality may even vary within one page.

However, higher-order models have been avoided for a long time due to their computational complexities. Indeed, GC based approaches, like α-expansion and $\alpha\beta$-swap, show high performance for inference in higher-order models [51], but the optimization technique seeks for a global optimum. We propose using a local optimization method for FBS in DIA and employ BP which handles arbitrary potential functions and provides a strong local minimum. In order to prepare the standard BP algorithm for higher-order

models, we propose the BP^n algorithm to *incorporate the higher-order potentials*.

The results show that the combination of spatial and spectral features provides a robust binarization method and that the assumption of local inference for energy minimization is more appropriate for the incorporation of the stroke model than a global method.

2.4 Summary

This chapter provided an overview of state of the art methods for FBS in DIA. We divided the approaches into five categories, including global thresholding, adaptive thresholding, color clustering, approaches based on spatial and/or spectral features, and probabilistic approaches. Particular attention was given on probabilistic approaches using MRFs. Since we incorporate higher-order models, efficient methods for statistical inference in higher-order models were presented in the second part of this chapter.

Chapter 3

Multispectral Imaging

Multi- or hyperspectral imaging has high potentials for the non-destructive analysis of objects of artistic and historic value like ancient manuscripts, panel paintings, or archaeological objects [39]. Originally utilized in remote sensing applications, e.g. for earth observation and region classification [86], researchers started at the beginning of this millennium to apply MSI for the examination of historical documents [24, 35, 33, 39, 60, 95, 5]. The aim of the multispectral digitization process includes, along with the detailed examination, preservation to ensure a long term availability of the documents, information retrieval, and the potential for further computer based investigations. The main advantage of analyzing multiple and different spectral ranges, including the UV and IR light, is the ability to examine text regions which are not visible to the human eye [72].

Applied on ancient manuscripts, [24] were the first to capture and enhance the erased underscript of the *Archimedes palimpsest* using MSI. The *Archimedes palimpsest* is a manuscript which was disbound, erased, and finally rewritten with the text of a Christian prayer book [24]. By means of MSI, the erased Archimedes text was uncovered under long wave UV illumination. The revisualization of the erased text is possible due to the fluorescence of the parchment when illuminated with this wavelength. Using a constrained least square algorithm, the underwritten text was separated from the superscribed text and highlighted [24]. MSI was also applied to improve the legibility of the *Codex Sinaiticus* and the *Herculaneum papyri* [35], or, for instance, for the analysis of Caucasian palimpsests written in the V. to VIII. century [33]. It became apparent that the application of spectral images can improve the readability, especially of damaged manuscripts better than conventional color imaging procedures [85].

This chapter gives background information on MSI and describes the acquisition setup used for the digitization of the *Missale Sinaiticum*. Finally, we introduce post-processing techniques to improve the readability of decayed manuscripts.

3.1 Multispectral Images

A multispectral image is the same image of one scene in multiple spectral ranges. Applied in the spectral range from UV, via VISible light (VIS) up to the Near InfraRed (NIR) range, it combines conventional imaging and spectrometry to acquire both, spatial and

spectral information from an object. Each spectral image is a panchromatic image covering a specific spectral range. Figure 3.4 illustrates the schema of a multispectral image.

Provided an appropriate illumination and an adequate imaging system, multispectral images can be obtained in different manners:

Individual illumination An MSI device, operating as a reflectance spectrometer, records a sequence of digital images of an object illuminated with monochromatic light. This monochromatic light may irradiate from a tunable light source ranging from UV to NIR [95]. Another possibility to obtain images in different spectral ranges is the use of different Light Emitting Diodes (LEDs) as a narrow band light source as applied on investigations on the *Archimedes palimpsest* [24].

Optical filters The traditional way of obtaining multispectral images is the use of optical filters to capture specified wavelengths [39]. In order to facilitate the acquisition, one can use liquid crystal tunable filters to achieve images in very narrow ($10nm$) wavelength bands spanning the visible spectrum from roughly $400nm$ to $700nm$. Another possibility is to use optical filters mounted in front of the camera. Using a filter wheel improves the change of filters [61].

3.1.1 Illumination and Electromagnetic Radiation

Light is electromagnetic radiation of any wavelength. Electromagnetic radiation is a self-propagating wave in space with electric and magnetic components oscillating perpendicularly to each other and to the direction of propagation. A photon is the elementary particle responsible for electromagnetic phenomena and the carrier of electromagnetic radiation of all wavelengths, including gamma rays, X-rays, UV light, or VIS light.

Electromagnetic radiation is characterized by its wavelength λ, the frequency f, and the propagation speed v with $v = f\lambda$. The speed of light is fixed by definition and constitutes $c = 299.792,459km/s$ in vacuum [92]. The energy E of a photon can be calculated by Planck's equation

$$E = hf = hc/\lambda, \tag{3.1}$$

where h constitutes a physical constant called Planck's constant ($h = 6.626 * 10^{-34} Joule\ seconds$). As Equation 3.1 shows, the energy content of electromagnetic radiation is determined by its frequency and wavelength. The electromagnetic spectrum encompasses all possible wavelengths of electromagnetic radiation and extends from electric power at the long-wavelength end, to gamma radiation at the short-wavelength end, see Figure 3.1. The human eye is sensitive to electromagnetic radiation between approximately $380nm$ (corresponding to blue) and $780nm$ (corresponding to red). The range of interest for the investigation of objects of artistic and historic value ranges from long wave UV ($350nm$) to the NIR range ($2500nm$). Short wave UV might destroy objects [72].

3.1.2 Multispectral Analysis of Ancient Manuscripts

MSI applied on manuscripts allows, for instance, the visualization of underwritten text in palimpsests or legibility enhancement of decayed manuscripts [39]. Important acquisition methods are arranged by the following configurations:

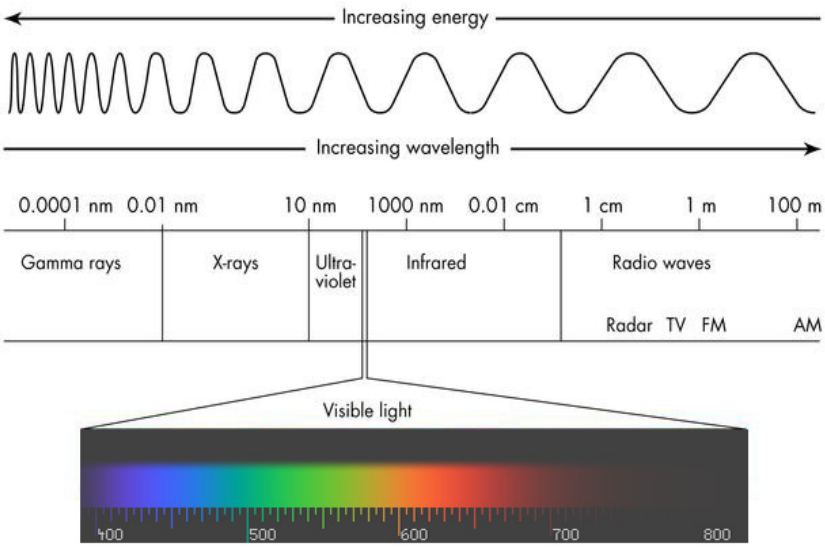


Figure 3.1: The electromagnetic Spectrum, from [4].

UV reflectography visualizes UV radiation reflected from the object. These reflections are not visible to the human eye and require cameras and sensors to be UV sensitive. In order to focus on the long wave UV light, the visible range of light has to be excluded. This is achieved by applying Short-Pass (S-P) filters and exclusively using UV light sources. UV reflectography can provide clues about material composition such as pigments or to earlier retouchings [72].

UV fluorescence visualizes UV fluorescence radiation radiated from an object. Scholars use UV illumination to read palimpsests, because the organic material in parchment fluoresces under UV illumination [23]. Fluorescence is a kind of luminescence (emission of light) by which a substance (e.g. a pigment) emits light of visible color when illuminated by e.g. UV light. UV fluorescence can be used to enhance the readability of palimpsest texts, since *old paint or varnish layers emit more fluorescence light compared to newly applied materials (repainting or retouching area)* [39]. The emitted radiation after excitation by a UV radiation source has either a shorter, a longer or equal wavelength compared to the incident wavelength [72].

A UV fluorescence facility consists of a UV illumination (e.g $\lambda = 375nm$), a barrier emission filter which cuts off wavelengths overlapping UV radiation, and a conventional camera (no UV sensitive sensor required).

IR reflectography requires a sensor which is sensitive to NIR radiation (0.8 $2\mu m$) and illumination in the NIR range. The method can be applied to the investigation of underdrawings[1] of panel paintings. IR reflectography allows to look through the covering paint layers of a painting and consequently a visualization of the underdrawing, since NIR radiation incident on carbon-based drawing materials is strongly

[1]Underdrawings are the preliminary drawings on a panel that has been prepared for painting and occur frequently in XV. and XVI. century European paintings.

absorbed [39].

3.2 *Missale Sinaiticum (Sin. Slav. 5/N)*: Acquisition Setup

Our objects of investigation consist of two parchment codices of the Old Church Slavonic canon dating from the 11$^{\text{th}}$ century, the so-called *Missale Sinaiticum (Sin. slav. 5/N)*[2] and the new part of the *Euchologium Sinaiticum (Sin. slav. 1/N)*, both in the Cyrillic and Glagolitic[3] script. Both fragments belong to the complex of new findings from St. Catherine's Monastery on Mt. Sinai in 1975. They show extensive damages like faded ink, blurring of the ink, staining due to mold or humidity, degradations of the parchment including chipping and fragmentation, or contortion of pages. The data corpus including cultural history and philological aspects is described in [74].

The multispectral digitization of the manuscripts was executed in situ in September 2007. The acquisition setup consists of a digital single-lens reflex camera and a digital high-resolution scientific camera. Color images and UV fluorescence images are captured with a Nikon D2Xs digital camera providing a resolution of 4288×2848 pixels. These images are particularly intended for visualization purposes and facsimile prints. MSI is performed with a Hamamatsu C9300-124 camera with a spectral sensitivity from UV to NIR ($330nm \quad 1000nm$) and a resolution of 4000×2672 pixels. A filter wheel mounted in front of the camera selects different spectral ranges. Figure 3.2 shows the setup of the acquisition system. For the spectral ranges selected, we use four Band-Pass (B-P) filters with peaks of $450nm$ (blue), $550nm$ (green), $650nm$ (red), and $780nm$ (NIR), two Long-Pass (L-P) filters with a cut-off frequency of $400nm$ (UV fluorescence) and $800nm$ (IR reflectography), and a short-pass filter with a cut-off frequency of $400nm$ (UV reflectography). Using different illumination (UV and VIS/NIR), we obtain nine different spectral images with a radiometric resolution of 12 bit and a spatial resolution of 565 dpi. The filters are summarized in Table 3.1 and their spectral transmittance is visualized in Figure 3.3 on the left hand side. Figure 3.4 illustrates the schema of the multispectral images and denotes the spectral components by its notation. It shows a detail from folio 41 recto from the *Missale Sinaiticum* in different spectra ranging from UV to NIR.

Since each folio is captured with both cameras, a shift of the page between the cameras is necessary. Due to the shift and the use of optical filters an image registration process is necessary to align one image to the other [13]. The registration process is summarized in Section 3.3.1.

Figure 3.5 provides an example to illustrate the condition of the manuscripts. The figure shows the RGB color image from folio 29 recto from the *Missale Sinaiticum*. It can be seen that especially the upper part of the text is highly degraded. Staining due to e.g. mold or humidity can be observed in the lower right part of the page. The folio shows degradations of the parchment itself especially on the right hand side of the page. The corresponding RGB color space is given below in Figure 3.6. It can be seen that there

[2] A missal is a liturgical book.

[3] This is the oldest known Slavic script [73].

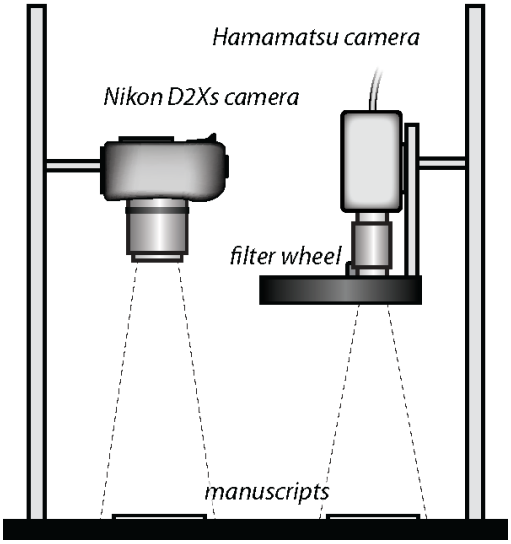


Figure 3.2: This figure illustrates the acquisition setup. The multispectral images are obtained with a Hamamatsu C9300-124 high resolution camera. A filter wheel in front of the lens including seven different filters captures nine different spectral images which are listed in Table 3.1.

Table 3.1: Description of the spectral images containing filter, type of illumination, and image. A filter type S-P denotes a short-pass, L-P is a long-pass, and B-P is a band-pass filter. Using UV light or illumination ranging from VIS to NIR, we obtain nine different spectral images.

No.	Filter type	Illumination	Image
1	S-P 400	UV	UV reflectography
2	L-P 400	UV	UV fluorescence
3	L-P 400	VIS-NIR	VIS, reduction of UV reflections
4	none	VIS-NIR	image without filter
5	B-P 450	VIS-NIR	blue
6	B-P 550	VIS-NIR	green
7	B-P 650	VIS-NIR	red
8	B-P 780	VIS-NIR	red, NIR
9	L-P 800	VIS-NIR	IR relectography

is no typical cluster for the background region nor for the text. This effect complicates simple thresholding or clustering in order to separate text from background.

Results and performance of MSI are illustrated on line 9 from folio 29 recto from the *Missale Sinaiticum*, see Figure 3.7. Note that some parts of the fragments are readable, but most of the text is hard to detect, or almost invisible, since the ink on the parchment has vanished (cf. the conventional RGB image in the first row). It can be clearly seen that wavelengths shorter than $650nm$ are highly absorbed by the ink, including the vanished parts which appear darker than in the conventional color image. Notice that the right outermost characters become transparent as the wavelength increases. This

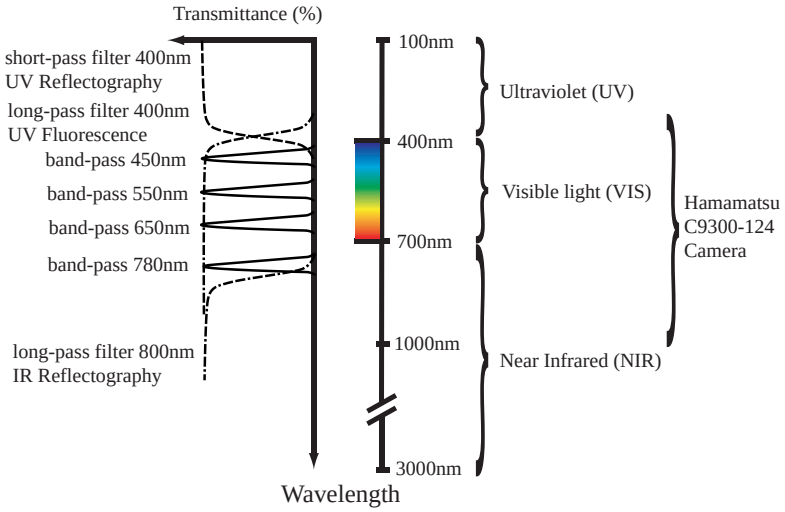


Figure 3.3: Illustration of the spectral ranges of the camera and the filters. The camera has a spectral sensitivity between 330 and 1000nm. Seven different filters in front of the lens capture specific spectral ranges: a short-pass filter for UV reflectography images, two long-pass filter for UV fluorescence and IR reflectography images, and four band-pass filters capture narrow band images in the visible range of light. Table 3.1 lists the images in more detail. Illustration from [48].

effect is generated from the parchment itself. The organic material in parchment fluoresces under UV illumination, i.e. the parchment absorbs the short UV light and re-emits longer-wavelengths, which are already in the visible region of the spectrum. The faded original text attenuates both, the incoming UV light and the exiting visible light. This double-pass attenuation enhances the visibility of the original text in images taken in blue light under UV illumination [23]. The text or ink vanishes in the spectral images of the IR range. A schematic illustration of the spectral signature of a particular ink and background pixel (blank parchment) can be seen in Figure 3.8. It shows the approximate plot of the spectral reflectance by means of gray level distribution within one spectral band. The blue line depicts the spectral reflectance of parchment and the green one corresponds to ink. The disparity between the lines is clearly visible. In the case of vanishing ink, this distinction is less sensitive and exclusively noticeable in the UV ranges.

Generally, the spectral behavior of various inks depends on the illumination and the multispectral bands [72]. In the case of the Slavonic manuscripts an X-ray fluorescence (XRF) analysis denoted only iron gall inks of various chemical compositions [74].

3.3 Post-Processing

Due to the use of optical filters the images must be registered on top of each other before further processing [13]. While measurement of the radiation in various wavelengths

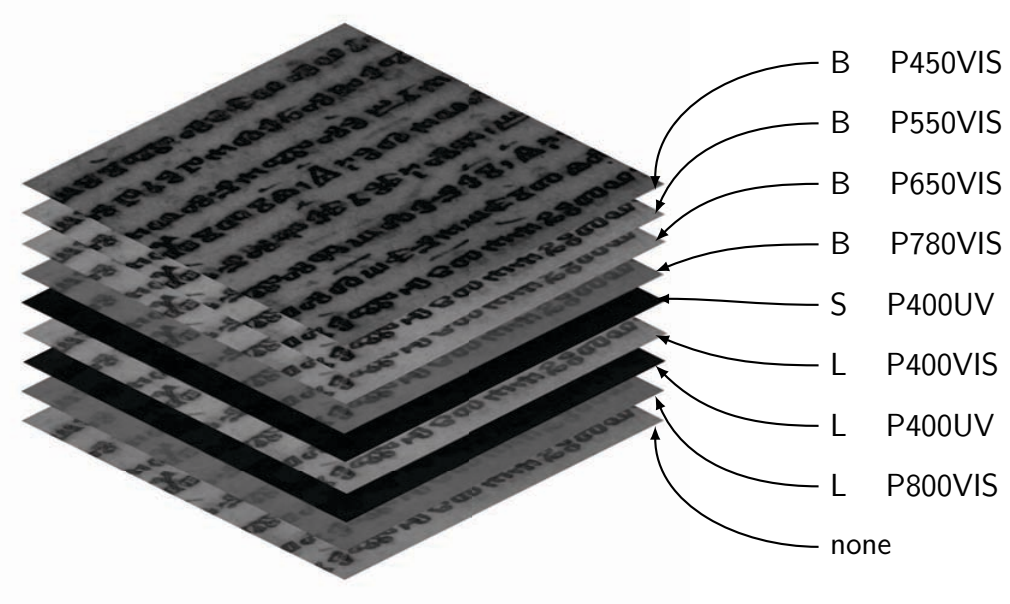


Figure 3.4: A multispectral image is the same image of one scene in multiple spectral ranges. The image shows a detail from folio 41 recto from the *Missale Sinaiticum* in different spectral ranges. Each spectral range reveals different properties of the source document. The images illuminated with UV light appear, due to its low illuminance very dark. However, due to fluorescence of the organic material in parchment under UV illumination, vanished parts of ink reveal additional information.

provides more information about the material in a scene, the resulting imagery does not lend itself to simple visual assessment [93]. Sophisticated processing of the imagery is required to extract all of the relevant information contained in the multitude of spectral bands.

In this section we summarize the process of image registration and propose an alternative to PCA to decorrelate the multispectral data [60]. The combination of several spectral bands improves considerably the readability of the *Missale Sinaiticum* when compared to conventional RGB images [74].

3.3.1 Image Registration

Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors [117]. Since the manuscript pages are repositioned between the two cameras and the use of filters in different wavelengths [13], a registration process is necessary in order to combine the spectral images and to remove distortions.

The registration process is also necessary for further image processing methods which utilize the information gained by the different spectral bands. Therefore, the images from the Nikon camera and the images from the Hamamatsu camera (they are originally rotated by 90°) are coarsely aligned on each other using rotationally invariant features [69] and an affine transformation. Afterwards, the similarity of the different images is computed by means of the normalized cross correlation. Finally, the images are accurately mapped

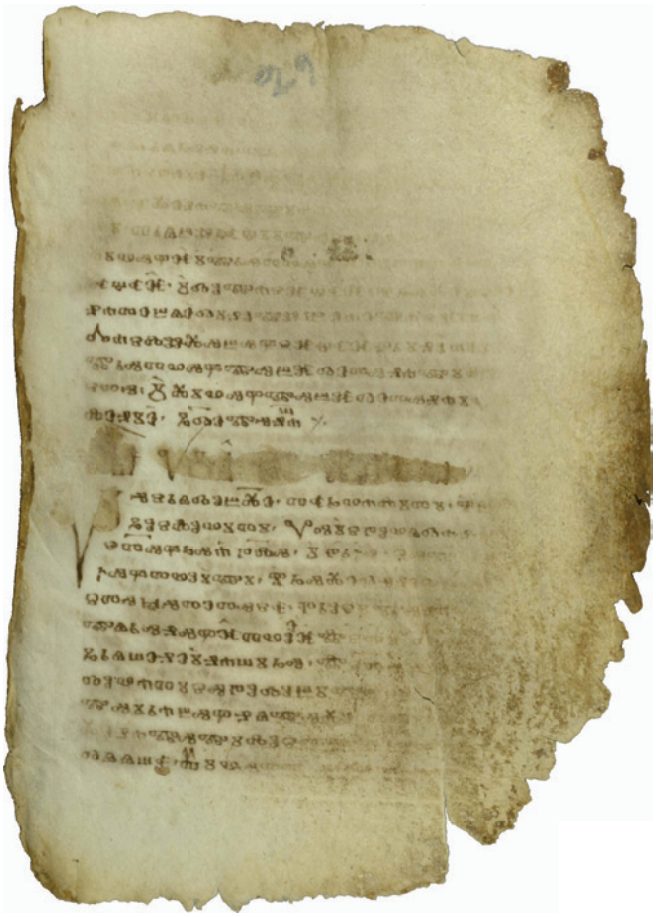


Figure 3.5: RGB color image from folio 29 recto from the *Missale Sinaiticum*.

to each other by the local weighted mean transformation. The registration process is explained in more detail in [21].

3.3.2 Image Enhancement

[24] use the PCA to produce pseudo-colored images from multispectral images of historic manuscripts. In this study, we propose an alternative approach by using a combination of spatial and spectral information of the multivariate image data to enhance the readability of the degraded text. The basis for this investigation is Multivariate Spatial Correlation (MSC) proposed by [109]. [108] applied this method to remotely sensed data and showed the effectiveness of the method in contrast to PCA. Consequently, we use MSC to enhance the readability of the varying appearance of text [60]. The benefit of this method is the possibility to consider especially text regions in document images. This spatial information is based on weighting relevant text regions. Therefore, we calculate the ruling scheme which forms the required weight factor for the MSC.

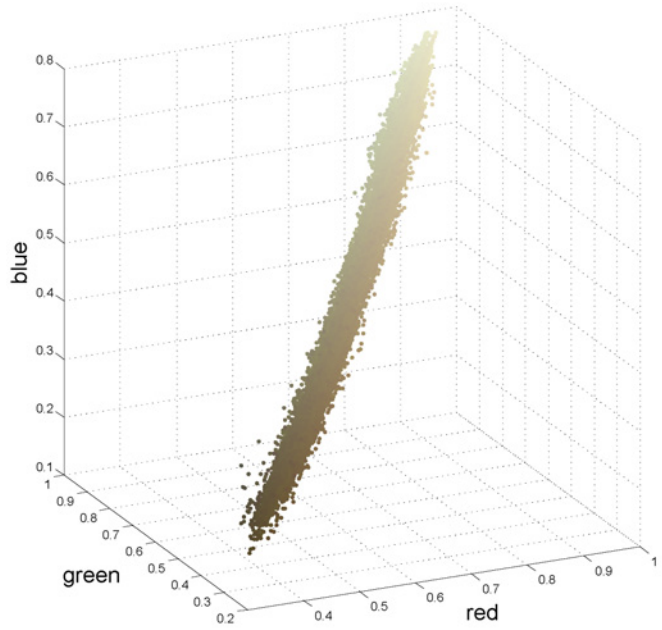


Figure 3.6: Corresponding RGB color space from folio 29 recto in Figure 3.5. It can be seen that there are no clusters for text or background which makes simple thresholding or clustering difficult. Therefore, we propose to combine spatial and spectral information for a general enhancement of the readability (Section 3.3.2) and for FBS (Chapter 6).

Multivariate Spatial Correlation

Multispectral image data is often highly correlated, i.e. they are visually similar [86]. The correlation arises through sensor band overlap and material spectral correlation. PCA removes this redundancy:

$$x' = A^t(x \quad \mu), \tag{3.2}$$

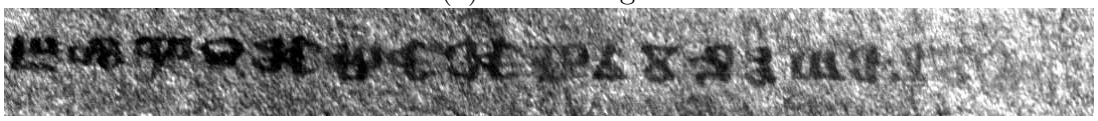
where A denotes the transformation matrix, x denotes the multivariate data and μ is the d-dimensional mean vector [22]. The transpose matrix of A is denoted by A^t . The columns of A consist of the k most valuable eigenvectors which are computed from the $d \times d$ covariance matrix.

MSC is a method for quantifying spatial autocorrelation in multi band data [108]. [109] extended a common univariate method of spatial correlation analysis for multivariate data. MSC was primarily used for geographical analysis, but [108] uses the method for the analysis of remotely sensed data and shows the robustness in the presence of noise. Compared to the results of synthetic data, the MSC matrix explained 99% of the MSC, whereas the first three components of PCA explained only 75% [108]. The spatial correlation methodology can be regarded of as a part of a generalized principal component analysis, for details see the Appendix in [109]. The MSC matrix of a d band $n \times m$ image is defined as follows:

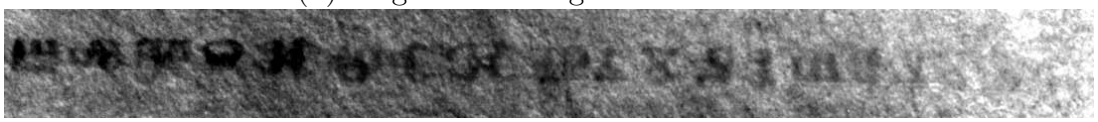
$$M = ZWZ', \tag{3.3}$$

where Z is a $d \times nm$ matrix containing the multivariate image data, W is a $nm \times nm$ weight matrix and Z' denotes the transpose of Z. The $nm \times nm$ weight matrix W is in

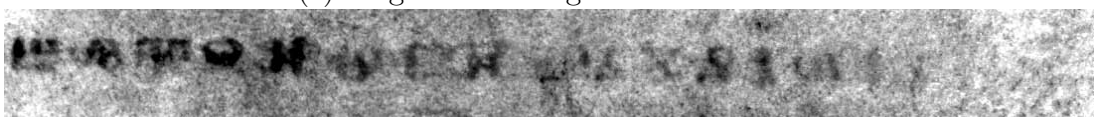
(a) RGB image.



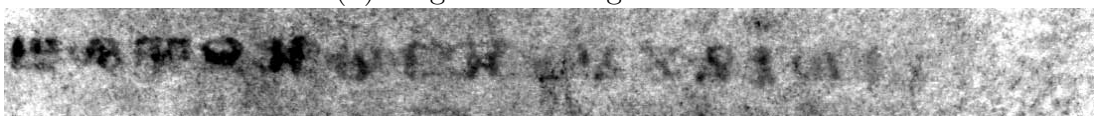
(b) Single band image S-P 400 - UV.



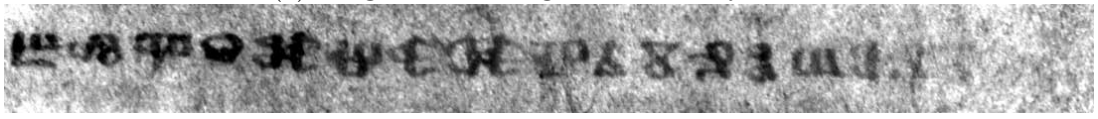
(c) Single band image L-P 400 - UV.



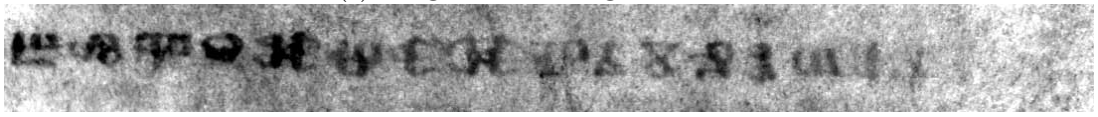
(d) Single band image L-P 400.



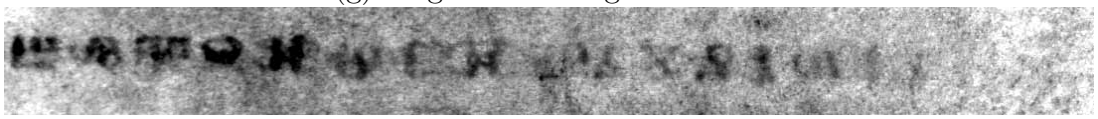
(e) Single band image without any filter.



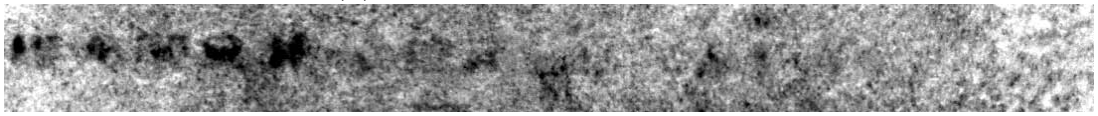
(f) Single band image B-P 450.



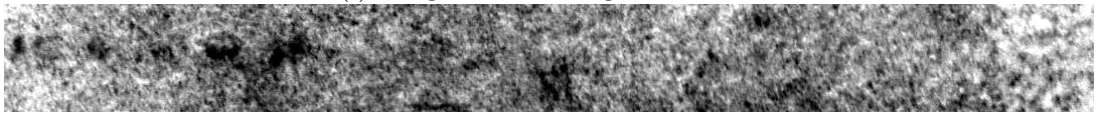
(g) Single band image B-P 550.



(h) Single band image B-P 650.



(i) Single band image B-P 780.



(j) Single band image L-P 800.

Figure 3.7: Line 9 from folio 29 recto from the *Missale Sinaiticum* in nine different spectral ranges. The contrast of each spectral image is increased with the MATLAB command `imadjust`. In accordance to Figure 3.8, the contrast of the images in the red and NIR range is low.

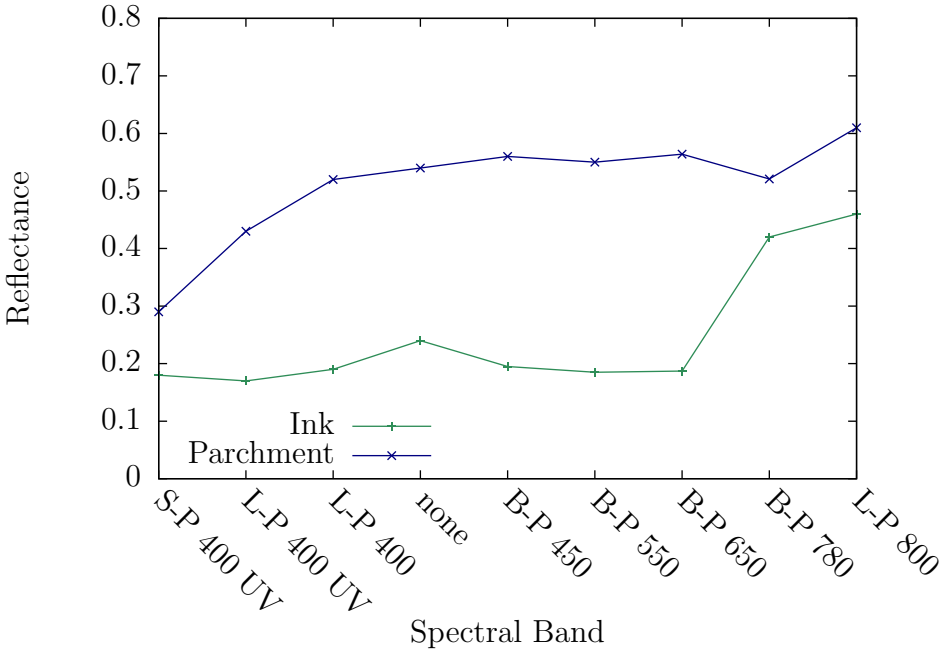


Figure 3.8: Sample for the spectral reflectance of ink and parchment. The figure shows that ink absorbs more incident light than parchment. In the NIR range of the spectrum (780 100nm) the contrast between ink and parchment is very low. When no illumination is mentioned beside the filter name, VIS-NIR light is applied.

the simplest case an adjacency matrix with $w_{ij} = 1$ if i and j are adjacent (e.g. $d_{ij} < c$ where c denotes a critical distance), otherwise $w_{ij} = 0$. The matrix is standardized so that its sum is equal to 1.

We utilize the positions of the text lines in order to generate the weight matrix for the calculation of the MSC matrix. The algorithm developed, has a pre-processing stage, which comprises a skew estimation, adaptive image binarization, and noise removal on a single band of the spectral images. After these pre-processing steps the text components (words, characters, etc.) are segmented and finally grouped to extract the text lines. The created binary mask highlights text regions and serves as the weight matrix for the calculation of the MSC. A following eigenvalue decomposition of the MSC matrix [108] and the creation of the transformation matrix similar to PCA enhances the readability. Figure 3.9 shows the mask from a detail from folio 29 recto from the *Missale Sinaiticum* in Figure 3.5. The left hand side shows the B-P 450 image detail superposed to the binary mask and the right hand side shows the binary mask itself.

The spatial correlation matrix M, which is in quadratic form, can again be decomposed into orthogonal components using eigenvector analysis [109]:

$$x'' = B^t(x \quad \mu), \tag{3.4}$$

where the columns of B consist of the eigenvectors which are computed from the $d \times d$ MSC matrix M. The components reflect the distribution of variations, comparable to PCA. In this case, the result is spatially weighted throughout the multivariate field.

Figure 3.9: Mask image for a detail from folio 29 recto from the *Missale Sinaiticum*. The mask images serves as weighting matrix for the MSC.

3.3.3 Image Enhancement Results

We demonstrate the method on folio 29 recto from the *Missale Sinaiticum* and compare the results to PCA. Since the weight matrix achieves a size of $nm \times nm$ where n and m depict the original image size, we use only fractions of the image. Figure 3.10 shows the distribution of the eigenvalues from MSC compared to PCA. It can be seen that the first eigenvalu [illegible] of the PCA ext

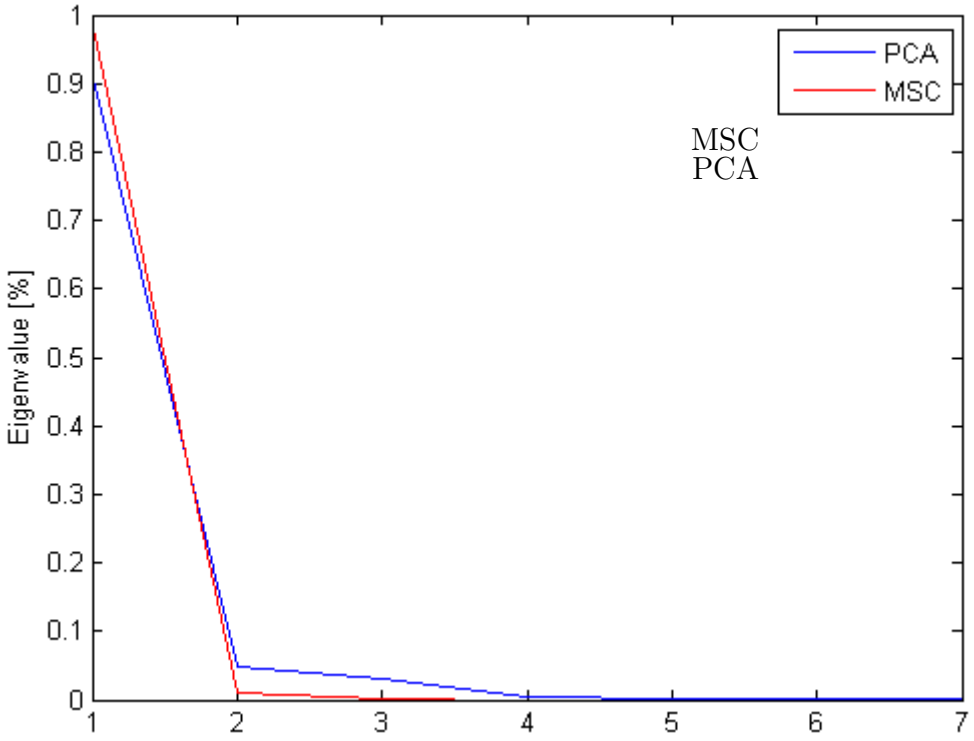


Figure 3.10: Distribution of the Eigenvalues after performing PCA and MSC. It can be seen that the first eigenvalue from the MSC matrix includes already more than 99%, while the eigenvalue of the PCA extend only 91.3% in the first and 4.8 % in the second component.

Resulting images from PCA and MSC can be seen in Figure 3.11 and Figure 3.12. The images are shown without any post-processing algorithms. Figure 3.11 shows the results

obtained with PCA. Here, the first two components contain visible characters, which in turn corresponds to the distribution of the eigenvalues from the covariance matrix. The MSC results can be seen in Figure 3.12. Regarding the second band obtained from the MSC transformation, the visibility of the characters is clearly enhanced. The third and fourth band of the MSC contain no useful information, as the distribution of the eigenvalues shows.

In a final evaluation, we demonstrate the advantage of the multispectral images compared to conventiopnal RGB images or the original manuscript. Therefore, we compared the number of characters transcribed from the original or RGB image of the whole corpus of the *Missale Sinaiticum* with the number of characters detected in the enhanced multispectral images. In the evaluation, we counted the number of characters transcribed from the RGB image, which denotes 24.448 characters. The number of additional characters detected in the enhanced images constitutes 12.459, which represents a rise of approximately 51%. The evaluation was executed in a conventional text editing program by collecting first the characters in black font referring to the original readable text, and afterwards the characters in red font which refer to the text transcribed from the enhanced images. Figure 3.13 demonstrates an example of a transcribed page.

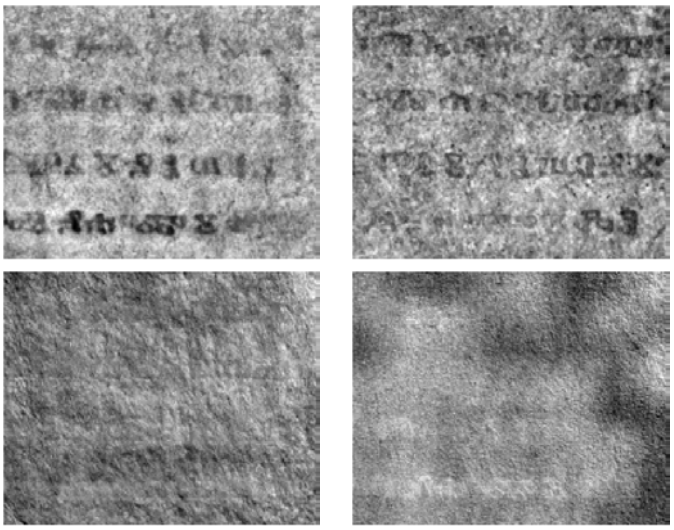


Figure 3.11: PCA results. The first row shows the first and second component and the second row displays the third and fourth component.

3.3.4 Discussion

MSI supports the investigation of ancient manuscripts where the text is hardly visible in conventional RGB images or for the human eye. A drawback of the method is the assemblage of highly correlated image data and the need for registration. The PCA is a method to reduce the spectral image data and to produce pseudo-colored images. In this section, we presented an alternative approach for the enhancement of the readability in ancient documents. The MSC matrix includes spatial and spectral image data to

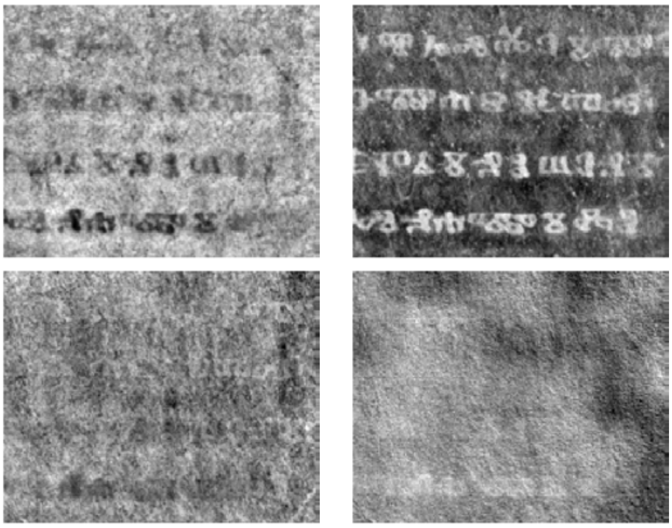


Figure 3.12: Example for MSC results. In the images given, the second component highlights the vanished characters best. Compared to PCA, especially the characters in the first row, e.g. Ⱁ, Ⰵ, and Ⰴ, are easier to detect.

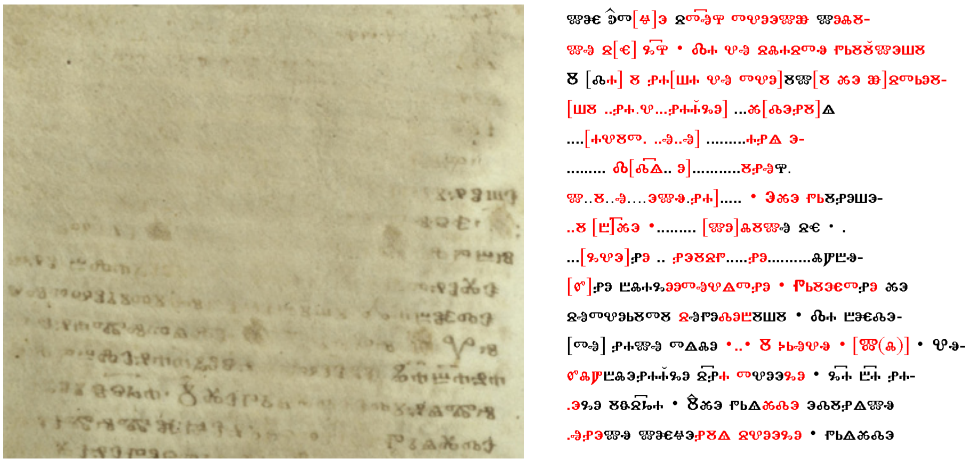


Figure 3.13: Transcription results of a detail from folio 16 verso from the *Missale Sinaiticum*. The left hand side gives the RGB image from the Nikon camera and the right hand side shows a transcription from philologists specialized in the objects given. Black characters represent information gathered from the original manuscript (cf. also the RGB image on the left hand side) and red characters are detected in the enhanced image with the proposed algorithm for readability enhancement. An evaluation of the transcription from the original data set and the enhanced images results in a rise of identifiable characters by approximately 51%.

remove spectral correlation. The benefit of the MSC based approach is that especially the text regions are considered for the enhancement. The experiments demonstrated the performance of combining spatial and spectral information for contrast enhancement.

3.4 Summary

The benefit of MSI for the investigation of ancient manuscripts was presented in this chapter. MSI has a high potential for the non-destructive analysis of old and degraded manuscripts. Multispectral analytical techniques benefit from a larger number of wavebands. The main advantage is additional information when using UV and IR bands additionally. In this chapter we explained multispectral images with background information on illumination and radiation, and we have shown three configurations for analyzing objects of artistic and historic value : UV reflectography, UV fluorescence, and IR reflectography. Afterwards, we explained the digitization of the *Missale Sinaiticum* which constitutes the main data set for our investigations and finally, we summarized the image registration process and proposed an approach for the enhancement of the readability in degraded document images.

Chapter 4

Probabilistic Graphical Models

MRFs and CRFs are probabilistic graphical models providing a probability theory for analyzing spatial and contextual dependencies [67]. Probabilistic graphical models allow the combination of graph theory and probability theory where nodes or vertices in a given graph represent random variables, and the links or edges express probabilistic relationships between these variables [44]. They provide a simple way to visualize the structure of a probabilistic model and can be used to design new models [9].

As already mentioned, our main motivation for using MRF and CRF for FBS, is the potential to simultaneously model spatial characteristics of strokes and contextual constraints by means of their spectral signature. MRF models usually used for object segmentation are characterized by energy functions defined on unary and pairwise potentials [49]. Unary potentials cover, for instance, color or texture features [46] and the pairwise potentials consider spatial dependencies typically in a regular 4-connected neighborhood. In addition, we incorporate higher-order potentials, for instance, regions of 5×5 pixels, to include spatial characteristics of strokes. This higher-order model allows major spatial formulations compared to traditional 4-connected graphs to enforce label consistency within local regions [50].

In standard MRFs the observations are assumed to be conditionally independent given the labels. However, this assumption limits its modeling ability since spatial relations depend only on the labels of neighboring pixels but not on their observation [114]. A different non-generative approach is to model the conditional probability of labels from given images within a CRF [56].

This chapter gives basic knowledge in graphical models and demonstrates the functionality of MRFs and CRFs, respectively. The focus is based on higher-order models. Inference methods for optimization are presented in Chapter 5.

4.1 Fundamentals

Probabilistic graphical models are graphs in which nodes represent random variables and links between the nodes probabilistic relationships. A graphical model is a graph $G = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is a set of nodes in one-to-one correspondence with a set of random variables $\mathbf{X}$ and $\mathcal{E}$ is a set of edges or links connecting the nodes. The edges $e \in \mathcal{E}$

of the graph can either be directed or undirected. An example for directed graphical models are Bayesian networks which are able to represent induced dependencies. Induced dependencies cannot be represented in undirected models, which in return are able to represent cyclic dependencies. MRFs are representatives of undirected models. Directed and undirected models allow functions to be defined on a set of several variables to be expressed as a product of factors over variables. Factor graphs make this decomposition explicit by introducing additional nodes for the factors themselves in addition to the nodes representing the variables [9]. Such sets of variables, or in a broader sense factors, allow the formulation of local relations which is the concept of MRFs. Figure 4.1 shows a directed, an undirected, and a factor graph.

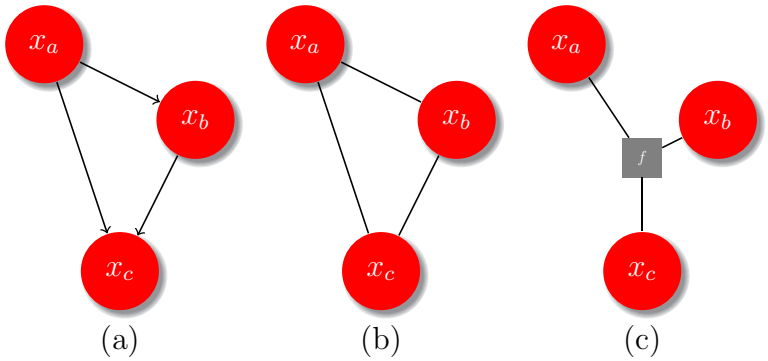


Figure 4.1: Directed (a), undirected (b) and a factor graph (c) with an additional node f.

4.2 Markov Random Fields

Let $\mathbf{X}$ be a random field defined over a finite rectangular lattice $\mathcal{V} = \{1, 2, \ldots, N\}$. Each random variable X_i in the random field $\mathbf{X}$ is associated with a lattice point $i \in \mathcal{V}$. A labeling or configuration $\mathbf{x}$ assigns a label l_k from the label set $\mathcal{L}$ to each of the random variables X_i, i.e. $\mathbf{x}$ takes values from the set $\mathbf{L} = \mathcal{L}^N$. The probability of any labeling $\mathbf{x}$, denoted as $\Pr(\mathbf{X} = \mathbf{x})$, will be referred to as $\Pr(\mathbf{x})$.

What we want to infer in FBS of digital document images is a binarized version of the input image $\mathbf{y}$, where the text consisting of characters and initials is separated from the background. For a given image data $\mathbf{y} = \{y_1, \ldots, y_N\}$, we seek to estimate the underlying scene by estimating a labeling $\mathbf{x} \in \mathbf{L}$ which separates the character or text pixels from the background. The label set is given as $\mathcal{L} = \{l_t, l_b\}$, where l_t depicts text and l_b is the label for background.

A widely accepted approach is to cast this labeling problem within a Bayesian framework [46]:

$$\Pr(\mathbf{x}|\mathbf{y}) = \frac{\Pr(\mathbf{y}|\mathbf{x})\Pr(\mathbf{x})}{\Pr(\mathbf{y})} \propto \Pr(\mathbf{y}|\mathbf{x})\Pr(\mathbf{x}), \tag{4.1}$$

where $\Pr(\mathbf{x}|\mathbf{y})$ is the posterior probability of a labeling $\mathbf{x}$ given the observations $\mathbf{y}$, $\Pr(\mathbf{y}|\mathbf{x})$ is the likelihood function of $\mathbf{x}$, $\Pr(\mathbf{x})$ denotes the prior probability of a labeling, and $\Pr(\mathbf{y})$

is the evidence given by:

$$\Pr(\mathbf{y}) = \sum_{i \in N} \Pr(\mathbf{x}|y_i) \Pr(y_i). \tag{4.2}$$

Given an image $\mathbf{y}$ we want to estimate the labeling $\mathbf{x}$ with the highest probability, i.e. the labeling which approximates the true labeling $\mathbf{x}^\star$ best. The labeling $\hat{\mathbf{x}}$ which maximizes the posterior probability $\Pr(\mathbf{x}|\mathbf{y})$ can be found via the Maximum A Posterior (MAP) estimate:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x} \in \mathbf{L}} \Pr(\mathbf{y}|\mathbf{x}) \Pr(\mathbf{x}). \tag{4.3}$$

Smoothness assumes that physical properties in a neighborhood of space present some coherence and generally do not change abruptly [67]. For instance, in the domain of document images, neighboring character or background pixels have similar properties like intensity or color. Thus, in the presence of spatial context, the labels are mutually dependent. MRFs are a probabilistic model which capture such spatial constraints by defining a neighborhood system $\mathcal{N}$ in the random field $\mathbf{X}$.

Definition 1 *A random Field* $\mathbf{X}$ *is said to be an* MRF on the sites $i \in \mathcal{V}$ with respect to a neighborhood system $\mathcal{N}$, if and only if the two following conditions are satisfied:

$$\Pr(x_i) > 0, \forall i \in \mathcal{V} \qquad (positivity), \tag{4.4}$$

$$\Pr(x_i \mid x_{\mathcal{V}\setminus\{i\}}) = \Pr(x_i \mid x_{\mathcal{N}_i}), \forall i \in \mathcal{V} \qquad (Markovianity), \tag{4.5}$$

where $x_{\mathcal{V}\setminus\{i\}}$ denotes the set of labels at the sites $\mathcal{V}$ without $\{i\}$ and $x_{\mathcal{N}_i}$ is the set of labels neighboring i.

Condition 4.5 implies that the label of a pixel i depends on its neighbors $\mathcal{N}_i$. The neighborhood system $\mathcal{N}$ in the random field $\mathbf{X}$ is defined as

$$\mathcal{N} = \{\mathcal{N}_i | \forall i \in \mathcal{V}\}, \tag{4.6}$$

where $\mathcal{N}_i$ denotes the set of all neighbors of the variable X_i. For a regular lattice $\mathcal{V}$, the neighbor set $\mathcal{N}_i$ of i is defined as the set of nearby sites within a radius r:

$$\mathcal{N}_i = \{i' \in \mathcal{V} | \left[dist(x_i, x_{i'})\right]^2 \leq r, i' \neq i\}, \tag{4.7}$$

where $dist(x_i, x_{i'})$ denotes the Euclidean distance between two pixels x_i and $x_{i'}$ [67]. The neighborhood of a first order MRF involves the four directly connected pixels in both the horizontal and vertical direction (standard 4-connected neighborhood system), a second order MRF involves its eight neighbors, and so on. Figure 4.2 illustrates the neighborhood system or the order of an MRF. The numbers $n = 1 \ldots 5$ indicate the neighboring sites of a n-th order neighborhood system.

A clique c is defined as a set of random variables x_c which are conditionally dependent on each other. A clique can consist of a single site, a pair of neighboring sites, a triple and so on:

$$x_c = \{x_i | i \in c\}. \tag{4.8}$$

5	4	3	4	5
4	2	1	2	4
3	1	x	1	3
4	2	1	2	4
5	4	3	4	5

Figure 4.2: Neighborhood system and order of MRFs.

4.2.1 Prior Model $\Pr(\mathbf{x})$

Having defined a neighborhood system and cliques, we can pass over to the realization of MRFs and the *Hammersley-Clifford* theorem. This theorem proves that a random field $\mathbf{X}$ is an MRF if the prior distribution $\Pr(\mathbf{x})$ follows a Gibbs distribution [32]:

$$\Pr(x) = \frac{1}{Z} \exp(\quad \frac{1}{T} U(x)), \tag{4.9}$$

where $Z = \sum_{x \in \mathbf{X}} \exp(\quad U(x))$ is a normalizing constant called the partition function, T is a constant called temperature, which shall be assumed to be 1, and $U(x)$ is the energy function defined as:

$$U(x) = \sum_{c \in \mathcal{C}} \psi_c(x_c). \tag{4.10}$$

The energy function $U(x)$ is the sum of clique potentials $\psi_c(x_c)$ over all possible cliques $\mathcal{C}$ and $\psi_c(x_c)$ denotes the potential function or clique potential of clique c having the label configuration x_c. Since the distribution of Gibbs Random Fields is equivalent to the distribution of MRF we can use Gibbs distribution to calculate the prior $\Pr(\mathbf{x})$ in the MRF [67].

The prior $\Pr(\mathbf{x})$ in the MRF represents the fact that the segmentation is locally homogeneous. It can be interpreted as the probability of a labeling x_i given the surrounding neighbors of pixel i. For pairwise connections, the clique potentials ψ_c follow the Potts model and favors similar configurations of neighboring pixels $x_i, x_j \in \mathcal{N}_i$:

$$\psi_c = \delta(x_i, x_j) = \begin{cases} 1 & \text{if } x_i \neq x_j \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

Incorporating Equation 4.11 into Equation 4.9, the full prior for pairwise MRFs is given by:

$$\Pr(x) = \frac{1}{Z} \exp\left(\sum_{i,j \in \mathcal{C}} \delta(x_i, x_j) \right). \tag{4.12}$$

4.2.2 Likelihood $\Pr(\mathbf{y}|\mathbf{x})$

The likelihood is based on the observation model which is assumed conditionally independent even though the underlying observation model is not as simple [67]. This yields to the following simplified form of the likelihood:

$$\Pr(\mathbf{y}|\mathbf{x}) = \prod_{i \in \mathcal{V}} \Pr(y_i|x_i). \tag{4.13}$$

The observation model or image process $\mathbf{y}$ can be formalized as follows: $\Pr(\mathbf{y}|\mathbf{x})$ follows a normal distribution $\mathcal{N}(\mu, \Sigma)$ [46]. Each class l_t and l_b (text and background) is represented by its mean vector μ_l and covariance matrix Σ_l

$$\mathcal{N}(\mu_l, \Sigma_l) = \frac{1}{\sqrt{(2\pi)^d \left|\Sigma_l\right|}} \exp\left(\quad \frac{1}{2} \left(y \quad \mu_l\right) \Sigma_l{}^{1} \left(y \quad \mu_l\right)^T \right), \tag{4.14}$$

where d is the dimension of the MSI. Thus, the likelihood is given by

$$\Pr(\mathbf{y}|\mathbf{x}) = \prod_{i \in \mathcal{V}} \Pr(y_i|x_i) = \prod_{i \in \mathcal{V}} \frac{1}{\sqrt{(2\pi)^d \left|\Sigma_{x_i}\right|}} \exp\left(\quad \frac{1}{2} (y_i \quad \mu_{x_i}) \Sigma_{x_i}{}^{1} (y_i \quad \mu_{x_i})^T \right). \tag{4.15}$$

The entities are modeled by a Gaussian mixture model (GMM). Given a GMM, the goal is to maximize the likelihood function with respect to the parameters μ and Σ. An elegant and powerful method for finding maximum likelihood solutions for models with latent variables is the Expectation Maximization (EM) algorithm [46, 22]. Applying EM on the multispectral image data, we obtain μ_t and Σ_t for the characters as well as μ_b and Σ_b for the background.

4.2.3 Posterior Energy $\Pr(\mathbf{x}|\mathbf{y})$

In the MAP estimation of an MRF, denoted as MAP-MRF framework, the optimal configuration is the maximum of the posterior $\Pr(\mathbf{x}|\mathbf{y})$ or equivalently of the joint probability $\Pr(\mathbf{x}, \mathbf{y})$. The posterior $\Pr(\mathbf{x}|\mathbf{y})$ can be simplified by including the contribution of the likelihood term via the singletons, i.e. the pixel sites $i \in \mathcal{V}$. The singleton energies reflect the probabilistic modeling of labels without spatial context and doubleton potentials express relationships between neighboring pixel labels [46]. Thus, after dropping the normalization constant, we get

$$\Pr(\mathbf{x}|\mathbf{y}) \propto \exp(\quad U(\mathbf{x}, \mathbf{y})) = \exp\left(\quad \left(\sum_{i \in \mathcal{V}} V_i(x_i, y_i) + \beta \sum_{i,j \in \mathcal{C}} \delta(x_i, x_j) \right) \right), \tag{4.16}$$

where $\beta > 0$ is a weighting parameter controlling the prior, i.e. the influence of the neighborhood connectivity, and $V_i(x_i, y_i)$ are the singleton potentials. The singleton potentials of pixel sites i are obtained from Eq. 4.15 by

$$V_i(x_i, y_i) = \ln(\sqrt{(2\pi)^d |\Sigma_{xi}|}) + \frac{1}{2} (y_i \quad \mu_{x_i}) \Sigma_{x_i}{}^{1} (y_i \quad \mu_{x_i})^T. \tag{4.17}$$

Now, the energy function $U(\mathbf{x}, \mathbf{y})$ of the MRF model has the following form:

$$U(\mathbf{x}, \mathbf{y}) = \sum_{i \in \mathcal{V}} \left(\ln(\sqrt{(2\pi)^d |\Sigma_{xi}|}) + \frac{1}{2}(y_i \quad \mu_{x_i}) {\Sigma_{x_i}}^{1} (y_i \quad \mu_{x_i})^T \right) + \beta \sum_{i,j \in \mathcal{C}} \delta(x_i, x_j). \quad (4.18)$$

This energy function can be shortly expressed as

$$U(\mathbf{x}, \mathbf{y}) = \sum_{i \in \mathcal{V}} V_i(x_i, y_i) + \beta \sum_{i,j \in \mathcal{C}} \delta(x_i, x_j). \quad (4.19)$$

Maximizing the posterior probability is equivalent to minimizing the energy, see Eq. 4.3, and the highest probability $\hat{\mathbf{x}}$ can be found via

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x} \in \mathbf{L}} \Pr(\mathbf{x}|\mathbf{y}) = \arg\min_{\mathbf{x} \in \mathbf{L}} U(\mathbf{x}, \mathbf{y}). \quad (4.20)$$

[11] formulated the energy function $E(\mathbf{x}, \mathbf{y})$ as

$$E(x_1, x_2, \ldots, x_N, y_1, y_2, \ldots, y_N) = \sum_i D_i(x_i, y_i) + \sum_{i,j} V(x_i, x_j), \quad (4.21)$$

where the functions $V(.)$ and $D(.)$ are energy functions. Generally, Equation 4.19 and Equation 4.21 can be expressed for pairwise connections in first order MRFs as:

$$E(\mathbf{x}, \mathbf{y}) = \sum_{i \in \mathcal{V}} \psi_i(x_i, y_i) + \sum_{i,j \in \mathcal{E}} \psi_{ij}(x_i, x_j), \quad (4.22)$$

where $\psi_i(x_i, y_i)$ is the local evidence for node i and the negative log of the likelihood of a label being assigned to pixel i. Generally, ψ_i denotes the unary compatibility function or data cost using learned foreground and background models. This data energy measures how well the label x_i fits to pixel i given the observation y_i. Data penalties ψ_i indicate individual label preferences of pixels, based on observed intensities and a pre-specified likelihood function [12]. We can write $\psi_i(x_i)$ as shorthand for $\psi_i(x_i, y_i)$.

The pairwise terms or smoothness term ψ_{ij} modeled via the prior, denotes pairwise compatibility functions where (i, j) indicates neighboring nodes i and j. As already mentioned, the prior represents the fact that the segmentation is locally homogeneous and labels depend on each other within a local neighborhood.

The Markov network topology can be seen in Figure 4.3. Each scene x_i is connected to its neighbors and to its underlying observation y_i. This topology provides the information about the observed data at any position i, because x_i has the only link to y_i, and gives information about nearby scenes neighbors.

However, the assumption for conditional independence of observations (cf. Equation 4.13) limits its modeling ability since spatial relations depend only on the labels of neighboring pixels but not on the observation [114]. A very different non-generative approach is to directly model the conditional probability of labels given images [42]. This is the key idea of a CRF. A CRF relaxes the strong independence assumption and captures contextual dependencies along observations [114].

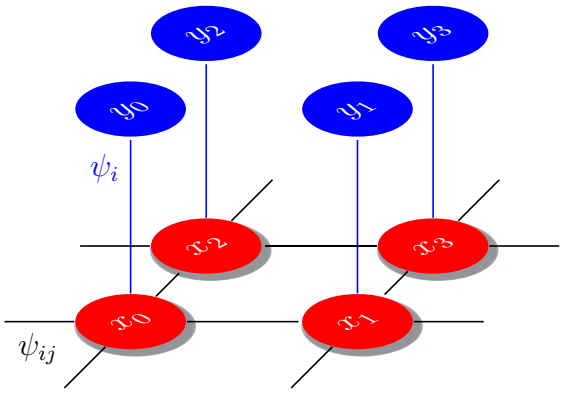


Figure 4.3: Markov network for vision problems. The nodes in the network describe observations $\mathbf{y}$ or hidden variables $\mathbf{x}$. Each observation y_i has a underlying scene explanation x_i. Links between the nodes express statistical dependencies. Blue connections indicate unary costs ψ_i and black connections pairwise costs ψ_{ij}.

4.3 Conditional Random Fields

CRFs are a discriminative probabilistic approach which models the posterior distribution $\Pr(\mathbf{x}|\mathbf{y})$ directly as an MRF without modeling the prior and the likelihood individually. Thus, a CRF is a random field globally conditioned on the observations $\mathbf{y}$. This approach allows one to capture arbitrary dependencies between the observations without resorting to any model approximations [42, 49].

Definition 2 *Let $G = (\mathcal{V}, \mathcal{E})$ be a graph such that $\mathbf{X} = \{X_i | i \in \mathcal{V}\}$, i.e. $\mathbf{X}$ is indexed by the vertices of G. Then $(\mathbf{x}, \mathbf{y})$ is a* CRF in case, when conditioned on $\mathbf{y}$, the random variables X_i obey the Markov property with respect to the graph: $\Pr(x_i \mid \mathbf{y}, x_{\mathcal{V}\setminus\{i\}}) = \Pr(x_i \mid \mathbf{y}, x_{\mathcal{N}_i}), \forall i \in \mathcal{V}$ [56].

In a CRF, the posterior distribution $\Pr(\mathbf{x}|\mathbf{y})$ over the configurations is a Gibbs distribution and can be written as:

$$\Pr(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \exp\left(\sum_{c \in \mathcal{C}} \psi_c(x_c) \right), \tag{4.23}$$

where $\psi_c(x_c)$ are potential functions defined over the variables $x_c = \{x_i, i \in c\}$, $\mathcal{C}$ is the set of cliques c, and Z is a normalizing constant [49].

Taking the log of Equation 4.23 the corresponding Gibbs energy is defined as:

$$E(\mathbf{x}, \mathbf{y}) = \quad \log \Pr(\mathbf{x}|\mathbf{y}) \quad \log Z = \sum_{c \in \mathcal{C}} \psi_c(x_c), \tag{4.24}$$

which can be expressed as first order model composed of unary and pairwise cliques:

$$E(\mathbf{x}, \mathbf{y}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i,j \in \mathcal{E}} \psi_{ij}(x_i, x_j). \tag{4.25}$$

In contrast to an MRF, the unary potential ψ_i of a CRF at site i is a function of all the observation data $\mathbf{y} = \{y_1, \ldots, y_N\}$ as well as that of the label $\mathbf{x}_i$ and not only of the observation y_i. Furthermore, the pairwise potential ψ_{ij} for each pair of sites i and j is independent of the observation in an MRF. In a CRF ψ_{ij} is also a function of all observations $\{y_1, y_2, , \ldots, y_N\}$ as well as that of the labels $\mathbf{x}_i$ and $\mathbf{x}_j$.

4.4 Higher-Order Models

Due to computational complexity, the priors in MRFs typically exploit pairwise-connected models in a standard 4-connected neighborhood system [49] and the resulting energy function is constructed of unary and pairwise cliques potentials, cf. Equation 4.25. However, pairwise interactions are often insufficient to capture the full spatial statistics of an image [84]. Higher-order clique potentials have the capability to model complex interactions of random variables and thus could overcome this problem [50]. Even, a prior model of natural images using 2×2 MRF cliques outperforms pairwise-connected models in the study from [84].

However, the use of higher-order models has been quite limited due to the lack of efficient algorithms for minimizing the resulting energy functions [49]. Recently, [50] introduced move making algorithms for minimizing energy functions involving higher-order cliques and presented a set of potential functions based on higher-order cliques to enforce label consistency [49].

As already stated, the stroke characteristics are modeled by incorporating higher-order cliques forming the proposed stroke model. In order to solve higher-order functions, several studies incorporate additional and more extensive connections by means of highly connected graphs, e.g. [99, 89]. Such energy functions on graphs with larger connectivity are becoming increasingly important in vision [52]. They typically arise in stereo vision, when two images need to be matched. Pixels (or features) in one image can potentially match to many pixels (features) in the other image, which yields to a highly connected graph structure [52]. However, the energy function of highly connected graphs is still based on pairwise connections as defined in Equation 4.25.

In contrast, higher-order models consider several pixels within a clique c in order to incorporate spatial probabilities in an extended range. Researchers recognized this fact and have used higher-order models to improve the expressive power of MRF and CRF frameworks [57, 88, 49].

The difference between pairwise, highly connected and higher order connections is illustrated in Figure 4.4. The illustration on the left hand side shows a graph with pairwise connections. Each random variable X_i is connected to its 4 directly horizontal and vertical neighbors. In the highly connected graph, the random variables can be connected to arbitrary sites in the graph, in the case of the illustrated graph, x_i is linked to its 8 neighbors. Higher-order cliques exploit segments of pixel sites or pixels within a clique c to analyze mutual behavior.

In expansion to Equation 4.24 and Equation 4.25 the Gibbs energy for higher-order

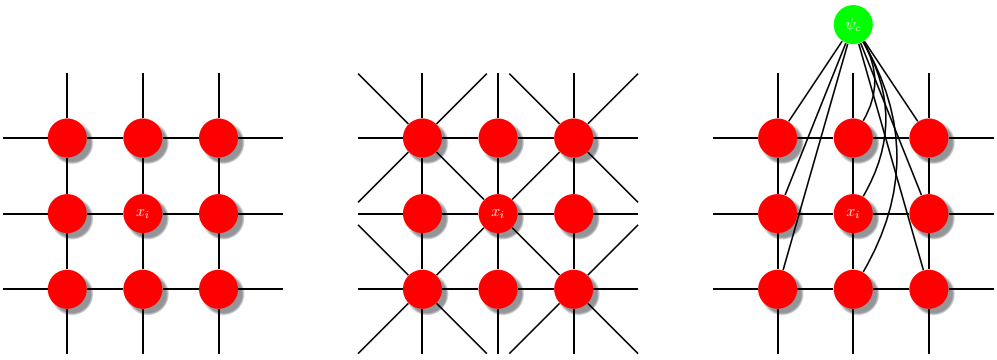


Figure 4.4: Graph construction for pairwise, highly connected, and higher-order models. In the pairwise model on the left hand side, x_i is connected to its 4 directly horizontal and vertical neighbors. Potentials between the nodes are based on pairwise functions ψ_{ij}. The highly connected graph in the middle illustrates an MRF of second order, where x_i depends on its 8 neighbors. The potentials between the nodes are still based on the pairwise potential function ψ_{ij}. Higher order cliques x_c include an arbitrary number of sites within an additional potential function ψ_c.

random fields can be written as [49]:

$$E(\mathbf{x}, \mathbf{y}) = \sum_{i \in \mathcal{V}} \psi_i(x_i) + \sum_{i,j \in \mathcal{E}} \psi_{ij}(x_i, x_j) + \sum_{c \in \mathcal{S}} \psi_c(x_c), \tag{4.26}$$

where $\mathcal{S}$ refers to image segments and ψ_c are higher-order potentials defined on these segments. The robust $\mathcal{P}^n$ model for higher-order cliques proposed in [49] is defined as:

$$\psi_c(x_c) = \begin{cases} \gamma_k & \text{if } x_i = l_k, \forall i \in c \\ \gamma_{max} & \text{otherwise,} \end{cases} \tag{4.27}$$

where γ_k and γ_{max} are learned parameters with $\gamma_{max} \geq \gamma_k$.

Some studies for FBS in DIA use the patch based topology introduced by [27]. For instance, [15] and [38] use 5×5 or 4×4 cliques learned from training data. These patches can also be interpreted as an extended spatial context to model the prior. An example of patches can be seen in Figure 4.5. The figure shows 114 patches obtained for the restoration of handwritten characters. The patches are trained from high quality document images and represent the most frequent occurrence of these patterns [15].

The topology can be interpreted in that way, that the nodes x_i are 5×5 scene patches and the observations are 5×5 image patches. The topology of the MRF model is the same as for the pairwise model in Figure 4.3. However, training is time consuming, it requires an adequate amount of training data, and omits a general use in FBS.

In order to omit learning we prefer a more general approach without training. Instead, we incorporate a model of the stroke characteristics and not of characters themselves. Therefore, we utilize a higher-order MRF and CRF to model spatial dependencies of strokes.

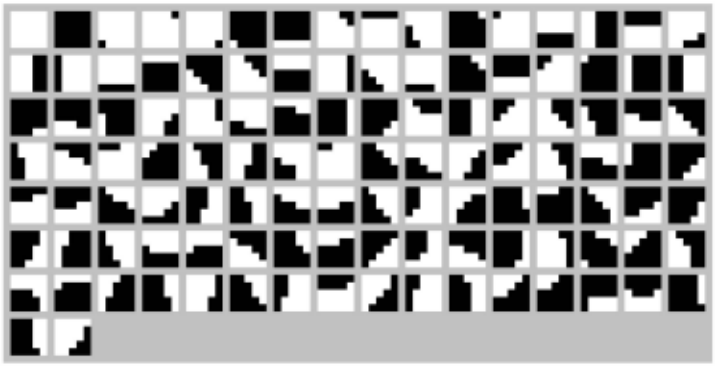


Figure 4.5: Examples for learning a prior model by means of image patches. The image patches are trained from high-quality binarized handwriting images. Therefore, the images are subdivided into 5×5 image patches and the most representatives are learned by clustering. The remaining 114 patches serve as the hidden nodes for the binarization of handwritten reports, from [15].

4.5 Summary

This chapter explained the concept of MRFs and CRFs and showed the framework in detail. We started with fundamentals of probabilistic graphical models and explained the details of MRFs and CRFs in the subsequent section. The main motivation to model our FBS approach by means of an MRF, is the possibility of simultaneously incorporating spatial and spectral features of the multispectral image data. Spatial characteristics are modeled within higher-order MRFs where efficient inference algorithms have been developed in recent years. The main difference of higher-order models in contrast to pairwise and highly connected ones is the additional potential function for separate image cliques. Potential functions for FBS and efficient inference algorithms for higher-order models will be described in the next two chapters.

Chapter 5

Energy Minimization

Probabilistic inference serves to find the solution of the maximum posterior probability $\Pr(x|y)$ in the MRF framework. As we have seen in Section 4.2, finding the MAP estimate is equivalent to minimizing the energy function $E(\mathbf{x}, \mathbf{y})$ in Equation 4.22. In general, the loopy structure of the underlying MRF graph makes exact inference NP-hard [88]. Therefore, methods which approximate the solution must be used.

Energy optimization to find MAP-MRF solutions can be done either by local or global methods. Popular inference methods include Iterated Conditional Modes (ICM) [8], Simulated Annealing (SA) [47], Belief Propagation (BP) [27], Tree-Reweighted Message Passing (TRW) [106], or Graph Cut (GC) based algorithms [11]. An overview of energy minimization methods for MRF with smoothness based priors is given in [100].

Inference in higher-order models is, due to the larger size of the cliques, particularly demanding. Thus, only pairwise interactions have been used for a long time. However, since higher-order models offer advantages compared to pairwise connections, some efficient methods for higher-order models have been published recently, e.g. [51] or [84].

Submodular set functions play an important role in energy minimization as they can be minimized in polynomial time [50]. However, common applicable methods can handle arbitrary potential functions [84] and need not consider the submodularity.

In this chapter, we describe two approaches for inference in MRFs. Since we have to deal with local minima, we prefer local energy minimization methods and compare in our experiments two of them, ICM and BP, with GC. We describe the two most popular GC algorithms, α-expansion and $\alpha\beta$-swap [100] as well as the adaptation from [50] for the minimization of higher-order potentials. BP and the proposed extension for higher-order models is presented in the next chapter.

5.1 Iterated Conditional Modes

[8] proposed a deterministic method which maximizes local conditional probabilities sequentially. ICM uses a deterministic "greedy" strategy to find a local maximum. Redrafted to energy minimization (see Equation 4.22), ICM starts with an estimate configuration $\mathbf{x}^0$ and iteratively selects a label for each pixel i which gives the largest decrease of the energy function. The process is repeated until it converges.

The calculation of the local posterior $\Pr(x_i|\mathbf{y}, x_{\mathcal{V}\setminus i})$ is based on two assumptions. The first one is the Markovianity (cf. Equation 4.5), i.e. x_i depends on the labels in the local neighborhood, and the second one is that variables $y_1, \ldots, y_N$ are conditionally independent given $\mathbf{y}$, and each y_i has the same known conditional density function $\Pr(y_i|x_i)$, cf. Equation 4.13.

Given the observation y_i and the neighboring configurations $x^k_{\mathcal{N}_i}$, ICM sequentially updates each x^k_i into x^{k+1}_i by minimizing $E(x_i|\mathbf{y}, x_{\mathcal{N}_i})$, the conditional posterior energy w.r.t. x_i. The energy function in Equation 4.22 is solved by iteratively minimizing the function with respect to each pixel i. Algorithm 1 illustrates the process flow of ICM. The first step in the inner loop calculates the local energy $E(x_i)$ for a certain pixel x_i in relation to its neighborhood $\mathcal{N}_i$ and selects the label with the minimal corresponding energy. The process is repeated until a steady state is obtained, i.e. $E(\mathbf{x}, \mathbf{y})^{k+1} \quad E(\mathbf{x}, \mathbf{y})^k \leq T$, where T is a given threshold.

The pairwise clique potentials ψ_{ij} are based on the Potts model and favor similar classes within a neighborhood $\mathcal{N}$:

$$\psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1 & \text{if } \mathbf{x}_i \neq \mathbf{x}_j \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

In order to realize higher-order models, the algorithm is adopted to a highly connected graph and the algorithm iteratively scans all labels in the local neighborhood. The additional connections are based on the pairwise model.

The quality of the result strongly depends on the initial labeling as a result of the high number of local minima [100]. A conventional way for the initial configuration is the maximum likelihood estimator, which ignores spatial dependence of one pixel to the others. To save computation time, we initialized the image with the result of the adaptive local thresholding algorithm proposed by [91]. In order to include stroke properties, we use a highly connected graph to cover all sites within a local neighborhood [65].

Algorithm 1 Iterated Conditional Modes (ICM).

1: Start a good initial configuration $\mathbf{x}^0$ and set $k = 0$
2: **repeat**
3: **for all** $i \in \mathcal{V}$ **do**
4: **for all** $l \in \mathcal{L}$ **do**
5: $x^k_i = \arg\min_{x_i \in \mathbf{L}} \left(\psi_i(x_i) + \sum_{\forall j \in \mathcal{N}_i} \psi_{ij}(x_i, x_j) \right)$
6: $k = k + 1$
7: **end for**
8: **end for**
9: **until** $E(\mathbf{x}, \mathbf{y})^{k+1} \quad E(\mathbf{x}, \mathbf{y})^k \leq T$

5.2 Energy Minimization using Graph Cuts

[11] present two algorithms based on GC [34] that efficiently find a local minimum with respect to two large moves called α-expansion and $\alpha\beta$-swap. They are very efficient and

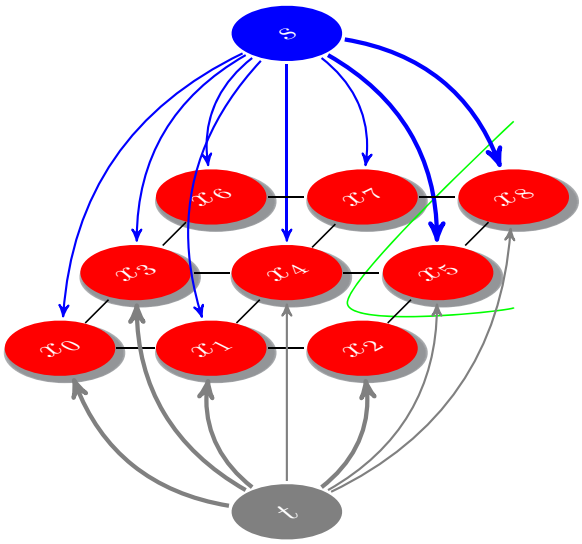


Figure 5.1: Graph construction for GC and possible cut (green line). Edge costs are reflected by thickness and the terminal nodes s and t are associated with labels. In the given binary decision problem, the cut yields some configuration in which the nodes x_5 and x_8 are segmented as the label given by s, and the remaining nodes as the label given by t.

are applied to a number of problems [100]. However, their results are limited to only pairwise functions. [51] provide a characterization of energy functions involving higher-order cliques, i.e. cliques of size 3 and beyond, for which the optimal moves can be computed in polynomial time.

Move making algorithms minimize the energy starting with an initial labeling and making a series of changes or moves to decrease the energy iteratively. At each step, the optimal move to decrease the energy is computed in polynomial time. In contrast to ICM, which allow only one pixel to change its label, α-expansion and $\alpha\beta$-swap allow several sites to change their configuration. Convergence is achieved when the energy cannot be minimized further.

Let $G = (\mathcal{V}, \mathcal{E})$ be a graph of the random field with nodes $\mathcal{V}$ and edges $\mathcal{E}$. The node set contains two additional terminal nodes which are called source s and sink t. Figure 5.1 shows a simple example of the construction. An edge is called t $link$ if it connects a non-terminal with a terminal node and n $link$ if it connects two non-terminal nodes. Here, t $links$ refer to the unary costs and n $links$ to the pairwise costs. The configuration in Figure 5.1 can only solve a two label problem. Assume two vertices, the source s and the sink t, have been distinguished. An s/t cut is a partitioning of the nodes in the graph into two disjoint subsets $\mathcal{S}$ and $\mathcal{T}$. The minimum cut problem is to find a cut that has the minimum of cost among all cuts [10]. Figure 5.1 shows the graph with the random variables x_i and the terminal nodes s and t. Edge costs are reflected by thickness and the green line illustrates a possible cut which assigns the nodes two different labels.

Computing the optimal move in polynomial time can only be done for a specific class of energy functions [53]. The optimal move t^* to get the exact optimum of a binary labeling can be computed in polynomial time if the energy function $E_m(\mathbf{t})$ of a move $\mathbf{t}$ is

submodular:

$$E_m^p(0,0) + E_m^p(1,1) \leq E_m^p(1,0) + E_m^p(0,1), \forall p \in \mathcal{V} \times \mathcal{V}. \tag{5.2}$$

The α-expansion and $\alpha\beta$-swap algorithms are represented as a vector of binary variables $\mathbf{t} = \{t_i, \forall i \in \mathcal{V}\}$. A transformation function $T(\mathbf{x}, \mathbf{t})$ takes the current labeling $\mathbf{x}$ and a move $\mathbf{t}$, and returns a new labeling $\mathbf{x}'$. The energy of a move is defined as the energy of the labeling $\mathbf{x}'$, the move induces: $E_m(\mathbf{t}) = E(T(\mathbf{x}, \mathbf{t}))$.

5.2.1 Pairwise Based Prior

[11] show that if the pairwise function ψ_{ij} defines a metric, the energy function can be approximately minimized using α-expansion and if the pairwise potential function defines a semimetric, the energy function can be approximately minimized using a $\alpha\beta$-swap move.

Conditions for $\alpha\beta$-swaps

The $\alpha\beta$-swap is an iterative GC algorithm applicable to situations when the smoothness term ψ_{ij} is semimetric.

Semimetric A potential function $\psi(a, b)$ for a pairwise clique of two random variables is semimetric, if for all $a, b \in \mathcal{L}$, it satisfies

$$\begin{aligned} \psi_{ij}(a,b) &= 0 \Leftrightarrow a = b \text{ (identity of indiscernibles)}, & (5.3)\\ \psi_{ij} &\geq 0 \text{ (nonnegativity)}, & (5.4)\\ \psi_{ij}(a,b) &= \psi_{ij}(b,a) \text{ (symmetry)}. & (5.5) \end{aligned}$$

If this condition is satisfied, the $\alpha\beta$-swap is defined as:

$$T_{\alpha\beta}(x_i, t_i) = \begin{cases} \alpha, & \text{if } x_i = \alpha \text{ or } \beta \text{ and } t_i = 0, \\ \beta, & \text{if } x_i = \alpha \text{ or } \beta \text{ and } t_i = 1. \end{cases} \tag{5.6}$$

Conditions for α-expansion

The α-expansion algorithm assumes that the smoothness prior term ψ_{ij} is a metric, such that the submodularity condition in Equation 5.2 which is weaker than the aforementioned is satisfied.

Metric The potential function is metric if in addition to the constraints above it also satisfies the triangle inequality:

$$\psi_{ij}(a,d) \leq \psi_{ij}(a,b) + \psi_{ij}(a,d), \forall a, b, d \in \mathcal{L}. \tag{5.7}$$

If condition 5.7 is satisfied, the α-expansion move is defined as:

$$T_\alpha(x_i, t_i) = \begin{cases} x_i, & \text{if } t_i = 0, \\ \alpha, & \text{if } t_i = 1. \end{cases} \tag{5.8}$$

Algorithm 2 and 3 illustrate the functionality of α-expansion and $\alpha\beta$-swap.

Algorithm 2 α-expansion algorithm.

1: Start with an arbitrary labeling $\mathbf{x}$
2: Set flag = $false$
3: For all label $\alpha \in \mathcal{L}$
4: Find $\hat{x} = \arg\min E(x')$ among x' within one α-expansion of x
5: **if** $E(x') < E(x)$ **then**
6: $x = x'$ and flag= $true$
7: **end if**
8: **if** flag= $true$ **then**
9: goto 2
10: **end if**

Algorithm 3 $\alpha\beta$-swap algorithm.

1: Start with an arbitrary labeling $\mathbf{x}$
2: Set flag = $false$
3: For each pair $\{\alpha\beta\} \in \mathcal{L}$
4: Find $\hat{x} = \arg\min E(x')$ among x' within one $\alpha\beta$-swap of x
5: **if** $E(x') < E(x)$ **then**
6: $x = x'$ and flag= $true$
7: **end if**
8: **if** flag= $true$ **then**
9: goto 2
10: **end if**

5.2.2 Solving Energies with Higher-Order Cliques

[51] extend the definition of α-expansion and $\alpha\beta$-swap for the minimization of energy functions whose clique potentials form a robust $\mathcal{P}^n$ model:

$$\psi_c(x_c) = \begin{cases} \gamma_k & \text{if } x_i = l_k, \forall i \in c \\ \gamma max & \text{otherwise.} \end{cases} \tag{5.9}$$

where γ_k and γ_{max} are learned parameters with $\gamma_{max} \geq \gamma_k$. The expansion and swap moves for any energy function composed of these potentials can be found by minimizing a submodular function [49]. Furthermore, the optimal move for the higher-order potentials P^n can be found by solving a s/t mincut problem.

The P^n functions are defined on cliques having a size of at most 2. The clique potentials take the form:

$$\psi_c(\mathbf{x}_c) = f_c(\mathcal{Q}_c(\oplus, \mathbf{x}_c)), \tag{5.10}$$

where f_c is an arbitrary function of $\mathcal{Q}_c$ and the clique inconsistency function $\mathcal{Q}_c(\oplus, \mathbf{x}_c)$ is a functional defined as:

$$\mathcal{Q}_c(\oplus, \mathbf{x}_c) = \oplus_{i,j \in c} \phi_c(x_i, x_j). \tag{5.11}$$

$\phi_c(x_i, x_j)$ is a pairwise function defined on all pairs of pixels in the clique c, and $\oplus$ is an operator applied to these functions $\phi_c(x_i, x_j)$. [51] characterize some higher-order potentials for which the optimal swap and expansion move can be computed in polynomial time. Therefore, they consider the sum form $\oplus = \sum$ and the max form $\oplus = \max$:

$$\mathcal{Q}_c(\mathbf{x}_c) = \sum_{i,j \in c} \phi_c(x_i, x_j), \tag{5.12}$$

$$\mathcal{Q}_c(\mathbf{x}_c) = \max_{i,j \in c} \phi_c(x_i, x_j). \tag{5.13}$$

Theorem 1 *The optimal $\alpha\beta$-swap move for any $\alpha, \beta \in \mathcal{L}$ can be computed in polynomial time if the potential function $\psi_c(\mathbf{x}_c)$ defined on the clique c is of the form 5.10, where $f_c(\cdot)$ is a concave non-decreasing function, $\oplus = \sum$ and $\phi_c(\cdot, \cdot)$ satisfies the constraints:*

$$\phi_c(a, b) = \phi_c(b, a) \qquad \forall a, b \in \mathcal{L} \tag{5.14}$$

$$\phi_c(a, b) \geq \phi_c(d, d) \qquad \forall a, b, d \in \mathcal{L} \tag{5.15}$$

Theorem 2 *The optimal α-expansion move for any $\alpha \in \mathcal{L}$ can be computed in polynomial time if the potential function $\psi_c(\mathbf{x}_c)$ defined on the clique c is of the form 5.10, where $f_c(\cdot)$ is a increasing linear function, $\oplus = \sum$ and $\phi_c(\cdot, \cdot)$ is a metric.*

If the conditions described are satisfied, the $\alpha\beta$-swap and α-expansion are defined as follows:

$$\psi_c(T_{\alpha\beta}(\mathbf{x}_c, \mathbf{t}_c)) = \begin{cases} \gamma_\alpha & \text{if } t_i = 0, \forall i \in c, \\ \gamma_\beta & \text{if } t_i = 1, \forall i \in c, \\ \gamma_{\max} & \text{otherwise.} \end{cases} \tag{5.16}$$

$$\psi_c(T_\alpha(\mathbf{x}_c, \mathbf{t}_c)) = \begin{cases} \gamma & \text{if } t_i = 0,\ \forall i \in c, \\ \gamma_\alpha & \text{if } t_i = 1,\ \forall i \in c, \\ \gamma_{\max} & \text{otherwise,} \end{cases} \tag{5.17}$$

where $\gamma = \gamma_\beta$ if $x_i = \beta$ for all $i \in c$ and $\gamma = \gamma_{\max}$ otherwise.

5.3 Summary

In this chapter, we presented two popular algorithms for statistical inference: ICM and GC. The focus was based on solving higher-order models, since the stroke properties in the application incorporate more extensive connections in the graphical model.

In the first section of this chapter, we explained ICM, a well known method used until the late 1990s, but with poor performance [100]. However, the method is easy to understand and for demonstrative purposes, we included this method also in our experiments. Nevertheless, the method shows beside its computationally shortcomings good performance for the separation of text from background.

In the second part, we reviewed two modern energy minimization algorithms based on GC that efficiently find a local minimum with respect to two large moves, namely, α-expansion and $\alpha\beta$-swap. The algorithms perform well on a variety of computer vision problems such as image restoration, stereo, and motion. In the comparative study from [100] expansion performs best along different benchmarks (such as image denoising and inpainting, binary image segmentation, or stereo matching). In terms of runtime, α-expansion was the winner among ICM, BP, and TRW.

However, inference in higher-order models was due to the larger size of the cliques neglected and only pairwise interactions have been used for a long time. Since higher-order models offer advantages compared to pairwise connections, [50] recently proposed expansion and swap moves for higher-order models.

For the experiments, we use on the one side the α-expansion implementation from [50], and on the other, ICM. Inference based on ICM is based on a highly connected graph. The advantage of local inference algorithms (ICM and BP), in contrast to global minimization methods such as GC, is the stronger property for a local minimum [100].

Chapter 6

Foreground-Background Separation based on Higher-Order MRFs

In the previous two chapters we provided theoretical background on MRFs and CRFs, as well as on state of the art methods for statistical inference. MRFs, CRFs, and consequently the optimization methods build the theoretical framework for the proposed FBS algorithm. In this chapter we introduce the unary, pairwise and higher-order potentials which build the energy function of the MRF. These functions collect the spatial and spectral features of the multispectral image data.

In the second part of this chapter, we present an adaptation of the standard BP in order to optimize higher-order energy functions. Since the complexity of BP is exponential in the size of the largest clique [84], we introduce a new message update rule to incorporate higher-order functions and to keep computational efficiency [63].

6.1 Higher-Order Energy Function

The final formulation of the energy function for the proposed FBS approach in digital document images is defined as (cf. Section 4.4):

$$E(\mathbf{x}, \mathbf{y}) = \alpha \sum_{i \in \mathcal{V}} \psi_i(x_i) + \beta \sum_{i,j \in \mathcal{E}} \psi_{ij}(x_i, x_j) + \gamma \sum_{c \in \mathcal{S}} \psi_c(x_c), \tag{6.1}$$

where α, β, and γ are weighting parameters for the individual potentials:

- The unary term, also called data term, ψ_i describes how an individual observation y_i matches a label l_k. The unary potential is based on the multispectral behavior of individual observations and models the spectral component in our approach.
- The pairwise term or smoothness energy ψ_{ij} represent the fact that a segmentation is locally homogeneous. The pairwise terms consider pairwise connected observations and are based on the image gradient. We have a high penalty if two neighboring sites have different labels and low costs if two connected nodes have the same configuration. Pairwise connections treat the first part of the spatial component of the proposed approach.

- The second part of the spatial constraints are modeled by the higher-order term ψ_c, which represents the proposed stroke model. These potentials are defined on cliques c of fixed shape. The higher-order term provides a function to enforce label consistency within predefined clique sizes, i.e. same site configurations within a clique obtain low penalties and different configurations receive higher penalties.

In the following section, we define the function for each potential. We start with the unary potentials, defined by means of a foreground and background model, then we present the pairwise terms, and finally the higher-order clique potentials and the proposed stroke model are presented.

6.2 Potential Functions

Potential functions provide the capability to assign predefined values for the configuration of individual, pairwise, or a set of sites. Since single sites and pairwise connections are, according to their definition, cliques we can specify the clique potential as the energy associated with a clique configuration. Generally, we assign a low penalty to a preferred clique configuration and a high penalty to an undesirable configuration. The following sections define the potentials which form the higher-order energy function in Equation 6.1.

6.2.1 Unary Potentials

The observation model on the pixel level can be estimated from the distribution of gray-scale densities of pixels [112]. Thus, the unary potentials ψ_i are obtained from the observations y_i. The multivariate normal density is typically an appropriate model for most classification problems where the feature vectors y_i for a given class l are continuous valued, mildly corrupted versions of a single mean vector μ_l [46]. Most approaches use simple models, e.g. Gaussian noise with zero mean and variance σ_n^2.

Observations from this type of distribution tend to cluster about the mean, and the extent to which they spread around the mean depends on the variance, cf. the RGB color space of a single page from the *Missale Sinaiticum* in Figure 3.6. It can be seen that similar colors are clustered in the RGB metric. This property makes the color space a good choice when using Gaussian modelization. However, there are no typical clusters for the foreground region nor the text region since the color distribution merge, at least in a global observation of the whole page.

It can be observed that the probability densities for text and background change over an image while the intensity of the background is changing. Therefore, we use local calculations within windows of size e.g. 64×64. The resulting unary potential $\psi_i(x_i)$ is a two dimensional vector measuring how well a label fits an observation y_i.

The unary potential follows a normal distribution $\mathcal{N}(\mu, \Sigma)$ and each pixel class is represented by its mean vector μ_l and covariance matrix Σ_l

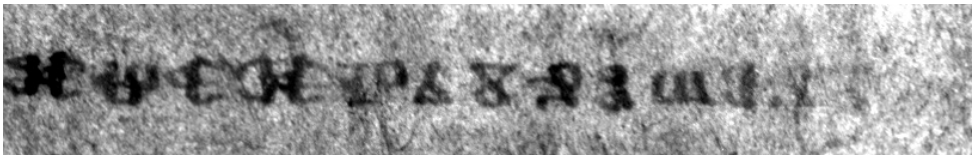
$$\psi_i(x_i) = \mathcal{N}(y_i|\mu_l, \Sigma_l). \tag{6.2}$$

We estimate the densities by modeling each class l_t and l_b, for text and background, as a two-Gaussian Mixture Model (2GMM) using the Expectation Maximization (EM) algorithm. Each class l_t and l_b is represented by its mean vector μ and covariance matrix Σ:

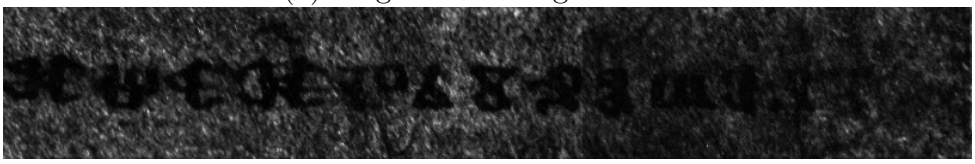
$$\mathcal{N}(\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d \left|\Sigma\right|}} \exp\left(\frac{1}{2} \left(y_i \quad \mu\right) \Sigma^{-1} \left(y_i \quad \mu\right)^T \right), \tag{6.3}$$

where d is the number of the spectral images.

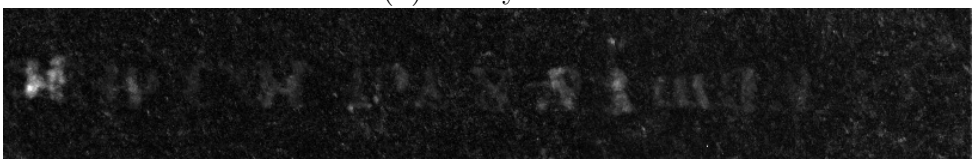
In this study, the observation model for the background and the text is based on local calculations of μ and Σ in order to avoid changes in the background, like water stains or mold. Given a Gaussian Mixture Model (GMM), the goal is to maximize the likelihood function with respect to the parameters μ and Σ. An elegant and powerful method for finding maximum likelihood solutions for models with latent variables is the EM algorithm [46, 22]. Applying EM on the multispectral image data, we obtain local estimates for μ and Σ for the text and the background. Figure 6.1 demonstrates the output of the potential functions for a detail of folio 29 recto from the *Missale Sinaiticum* and shows the unary potentials for text (a) and background (b). It can be seen that the text model, i.e. unary text, has high penalties within the background region, i.e. a label l_t is preferred, and for the background, i.e. unary back, we obtain high penalties for regions with text.



(a) Single band image BP450.



(b) Unary text.



(c) Unary back.

Figure 6.1: Single band image B-P 450 from a detail from folio 29 recto from the *Missale Sinaiticum*. The second and third row show the output of the unary potential function ψ_i for text (unary text) and background (unary back).

6.2.2 Pairwise Potential Function

The pairwise terms $\psi_{ij}(x_i, x_j)$ corresponds to the matching cost computation between nodes x_i and x_j. The energy function takes the form of a pairwise Potts model:

$$\psi_{ij}(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j, \\ \rho_I(\nabla I) & \text{otherwise,} \end{cases} \tag{6.4}$$

where the function $\rho_I(\nabla I)$ is defined in terms of the image gradient between the pixels i and j [101]. Thus, when two neighboring pixels x_i and x_j have similar values we have a low penalty and in border regions, e.g. background to text, we have an high penalty in the energy function [11].

6.2.3 Higher-Order Potentials

Since only pairwise potential functions limit the expressiveness of the models, [49] defined the $\mathcal{P}^n$ Potts model for higher-order cliques of size n as:

$$\psi_c(x_c) = \begin{cases} \gamma_k & \text{if } x_i = l_k, \forall i \in c \\ \gamma_{max} & \text{otherwise,} \end{cases} \tag{6.5}$$

where $\gamma_{max} \geq \gamma_k, \forall l_k \in \mathcal{L}$. In the experiments from [51] on a set of natural images, the set of higher-order cliques consists of all segments of multiple segmentations of an image. The image segments are obtained using an unsupervised image segmentation algorithm such as mean-shift [19].

For the separation of text in digital documents, we propose to use stroke properties which cover spatial dependencies of characters. For the application of FBS the size n of the higher-order cliques x_c comprises the set of all pixels i within a diameter $\varnothing$, where $\varnothing$ corresponds to the mean diameter of the strokes on one text page. This predefined prior model of fixed shape will be referred to as the stroke model. The mean stroke width can be extracted automatically after a preceding binarization of the input image. Then, the mean stroke diameter $\varnothing$ follows the set of all foreground pixels S and the set of all border pixels D from characters in BW and can be obtained by

$$\varnothing = 2\frac{|S|}{|D|}, \tag{6.6}$$

where $|S|$ and $|D|$ are the number of pixels in S and D [82]. Experiments devoted a mean width of 5 pixels for the data set given in the corpus of the *Missale Sinaiticum*. The proposed stroke model is illustrated in Figure 6.2. It shows the Glagolitic character  including a white circle which corresponds to the average diameter of the stroke. Since the stroke width of the character given is approximately five pixels, the circle in the figure corresponds to a fourth order neighborhood system, cf. also Figure 4.2. Thus, the prior of the proposed stroke model considers a neighborhood set of at least 4th order.

Several potential functions have been proposed in the literature. For instance, [116] suggested to use the maximum color distance of observations which can be expressed as

$$\psi_c(x_c) = \|y_{c\max} \quad y_{c\min}\|, \tag{6.7}$$

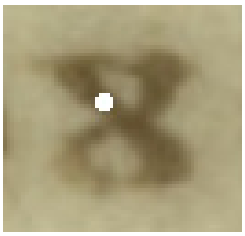


Figure 6.2: Glagolitic character with marked neighborhood system of 4th order. The white circle corresponds to the proposed stroke model with fixed shape.

where $\|y_{c\max} \quad y_{c\min}\|$ depicts the distance from the highest to the lowest value of all observations y_c in a clique c. In this model we obtain high penalties when the observations y_i are different and low penalties for similar observations.

The clique potentials in [51] are defined such that they form a P^n Potts model:

$$\psi_c(x_c) = \begin{cases} \gamma_3 G(c,s) & \text{if } x_i = s,\ \forall i \in c \\ \gamma_4 & \text{otherwise.} \end{cases} \tag{6.8}$$

Here, $G(c,s)$ is the minimum difference between the RGB values of a patch D_c and all patches belonging to the dictionary P_s. Note that the above energy function encourages the patch D_c which is similar to a patch in P_s to take the label s. Since the clique potentials form a P^n Potts model, they can be minimized using the $\alpha\beta$-swap and α-expansion algorithms [50]. The proposed parameter setting is $\gamma_3 = 0.6$ and $\gamma_4 = 6.5$.

For the separation of text from background we prefer potentials which characterize the differences from a given observation y_i with its neighboring observations $y_{\mathcal{N}_i}$ and we compare the observation y_i with the pixels in its local neighborhood

$$\psi_c(x_c) = \begin{cases} 0 & \text{if } x_i = l_k,\ \forall i \in c \\ |y_i \quad \mu_{\mathcal{N}_i}| & \text{otherwise,} \end{cases} \tag{6.9}$$

where $\mu_{\mathcal{N}_i}$ is the mean value of the observations in $\mathcal{N}_i$ and $|y_i \quad \mu_{\mathcal{N}_i}|$ is the absolute value of $y_i \quad \mu_{\mathcal{N}_i}$. The higher-order potential can be interpreted as the deviation from observation y_i to its surrounding sites in $\mathcal{N}_i$.

6.3 Belief Propagation

Originally designed for graphs without loops [81], BP uses the idea of passing local messages around nodes. BP provides an exact solution when there are no loops in a graph, but it has also been tried on loopy graphs providing approximate solutions [115, 111]. When applied on graphs with loops, the messages must be updated iteratively and the algorithm is called Loopy Belief Propagation (LBP). LBP is not guaranteed to converge, but achieves outstanding empirical results in several studies [84] and has a strong local minimum property [100].

The main advantage of BP is that it works for arbitrary kinds of potential functions, including non-regular potential functions, which are not available to GC [84]. A drawback

is that BP is slow, since it requires messages from each node to its neighboring nodes. Just like ICM, the number of computations for the messages in higher-order models is exponential in the number of neighbors, i.e. the runtime complexity increases exponentially with the size of the largest clique in the random field.

6.3.1 Standard Belief Propagation for Pairwise Models

The standard BP for pairwise grid connections has two versions for the message update rule, the sum-product and the max-product rule. The max-product formulation has often been used for pixel labeling problems, whereas the sum-product formulation is more appropriate for interpolation problems with non-integer solutions [57].

BP works by iteratively passing messages around the graph. Let $m_{ij}^t(x_j)$ be the message sent from node x_i to its neighbor x_j at iteration t. Each message m_{ij} is a vector of the dimension given by the number of labels, with each component being proportional to how likely node i thinks it is having the same state as node j. All entries are initialized to zero and at each iteration, new messages are computed.

The message update for the sum-product BP is given by:

$$m_{ij}^t(x_j) \leftarrow \sum_{x_i} \Big(\psi_i(x_i) \psi_{ij}(x_i, x_j) \Big) \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}(x_i), \tag{6.10}$$

and the messages for the max-product algorithm are computed in the following way:

$$m_{ij}^t(x_j) \leftarrow \max_{x_i} \Big(\psi_i(x_i) \psi_{ij}(x_i, x_j) \Big) \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}(x_i), \tag{6.11}$$

where $\mathcal{N}_i \setminus j$ denotes the neighbors of i without j. Figure 6.3 illustrates the message update rule from node x_i to x_j. The incoming messages $m_{ki}(x_i)$ are passed through node x_i to its neighbor x_j. An equivalent computation can be performed with negative log probabilities, where the max-product becomes a min-sum:

$$m_{ij}^t(x_j) \leftarrow \min_{x_i} \Big(\psi_i(x_i) \psi_{ij}(x_i, x_j) \Big) \sum_{k \in \mathcal{N}_i \setminus j} m_{ki}(x_i). \tag{6.12}$$

After t iterations, a belief $b_i(x_i)$ for each node is computed. The beliefs are an approximation of the marginal probability $\Pr_i(x_i)$ of a node i to be labeled x_i. The belief of a node i is proportional to the product of the local potential $\psi_i(x_i)$ and all incoming messages $m_{ki}(x_i)$:

$$b_i(x_i) = \psi_i(x_i) \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}(x_i). \tag{6.13}$$

The configuration x_i which minimizes $b_i(x_i)$ individually at each node is selected. We use the max-product algorithm to compute the MRF-MAP estimate, because it is less sensitive to numerical artifacts [25] and it directly corresponds to the formulation of the energy function.

Algorithm 4 explains the processing steps of BP in pseudo code. The first nested `for`-loop initializes all messages to zero. The second expression manages the message updates

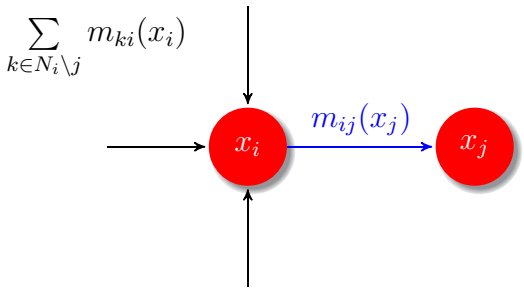


Figure 6.3: Illustration of message passing in the BP algorithm. The message m_{ij} from node x_i to x_j contains the messages from its neighboring nodes m_{ki}, where k denotes all nodes in the neighborhood of x_i except x_j, i.e. $k \in \mathcal{N}_i \setminus j$.

m_{ij} iteratively for T loops. Finally, the beliefs and the corresponding configurations are calculated in the last two expressions.

Several variants of the BP algorithm exist. For instance, [25] propose an efficient version of the standard BP to compute the message update in linear time. The standard implementation for pairwise connections runs in $\mathcal{O}(NK^2T)$ time, where N is the number of nodes, K is the number of labels for each pixel and T is the number of iterations, i.e. it takes $\mathcal{O}(K^2)$ time to compute each message and there are $\mathcal{O}(N)$ messages to be computed in each iteration. For higher-order models, the runtime complexity increases exponentially with the size of the largest clique in the random field. BP for higher models has a complexity of $\mathcal{O}(K^n)$ where n is the number of neighbors of an image site i.

6.3.2 Belief Propagation for Higher-Order Models: BP^n

An intuitive way to incorporate higher-order models is to define messages that propagate between groups of nodes rather than just single nodes. This is the intuition in GBP [27]. The graph is split into clusters and a hierarchy of regions and sub-regions is created. GBP propagates messages across clusters of nodes an not only between nodes as in standard BP. It reduces to standard BP when the clusters consist of only two nodes. GBP provides accurate solutions on highly connected graphs, however the clusters are hard to generate [78].

An alternative is to combine messages from two neighboring nodes to one message [116] or to represent the potential functions $\psi_i(x_i)$ via additional nodes in a factor graph [84]. However, using major label sets or clique sizes makes it intractable to store the beliefs and messages in a factor graph [88].

[43] generalized the message-passing process for second order BP as

$$m_{ijk}(x_i) \leftarrow \sum_{x_j} \sum_{x_k} \psi_j(x_j)\psi_k(x_k)\psi_{ijk}(x_i, x_j, x_k) \prod_{s \in \mathcal{N}_j \setminus i} m_{jis}(x_j) \prod_{s \in \mathcal{N}_k \setminus i} m_{kis}(x_k), \qquad (6.14)$$

in which second order constraints are modeled. Here, $m_{ijk}(x_i)$ represents the message passed from node j and k to node i.

Algorithm 4 Loopy Belief Propagation (LBP).

for all $(i, j) \in \mathcal{E}$ **do**
  for all $l \in \mathcal{L}$ **do**
    set $m_{ij}^0 = 0$
  end for
end for
for $t = 1 \dots T$ **do**
  for all m_{ij}^t & $l \in \mathcal{L}$ **do**
    $m_{ij}^t(x_j) \leftarrow \min_{x_i} \Big(\psi_i(x_i)\psi_{ij}(x_i, x_j) \Big) \sum_{k \in \mathcal{N}_i \setminus j} m_{ki}^{t\ 1}(x_i)$
  end for
end for
for all $i \in \mathcal{V}$ **do**
  for all $l \in \mathcal{L}$ **do**
    $b_i(x_i) = \psi_i(x_i) \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}(x_i)$
  end for
end for
for all $i \in \mathcal{V}$ **do**
  $x_i = \arg\max b_i(x_i)$
end for

An additional model for higher-order was proposed by [116]. This three node potential clique feature vector is described by the minimum angle, the consistency of the region inter-distance, the maximum color distance, and the height consistence of the characters. The pairwise potentials are characterized by three nodes and are denoted as ψ_{ijk}:

$$p(x, y) = \frac{1}{Z} \prod_{ijk} \psi_{ijk}(x_i, x_j, x_k, y_{ijk}) \psi_i(x_i, y_i). \tag{6.15}$$

For the separation of text from background, we propose to incorporate a local optimization problem. Thus, we apply BP [27] for inference which has a strong local minimum property [100] and works for arbitrary potential functions [84]. In order to include higher-order potentials $\psi_c(x_c)$, we introduce BP for higher-order MRF models, abbreviated as BP^{n}, as an extension of the standard BP for pairwise connections. In contrast to the standard formulation, we update the rule for the message updates m_{ij} and include the higher-order potentials ψ_c. The model for higher-order MRF is based on Equation 4.18, which incorporates unary costs ψ_i observed from the image, pairwise or smoothness costs ψ_{ij}, which afford that the segmentation is locally homogeneous, and higher-order costs ψ_c, which enforce label consistency within predefined cliques. The cliques may overlap and each pixel x_i receives the higher-order penalty from its surrounding pixels in $\mathcal{N}_i$. These clique-node messages integrate label properties of cliques. The resulting network topology can be seen in Figure 6.4. Each node x_i receives its unary potential $\psi_i(x_i)$, the pairwise costs $\psi_{ij}(x_i, x_j)$ from its neighbors in the 4-connected grid, and the higher-order potential $\psi_c(x_c)$, which is composed of the observations y_c. Since BP has no feedback

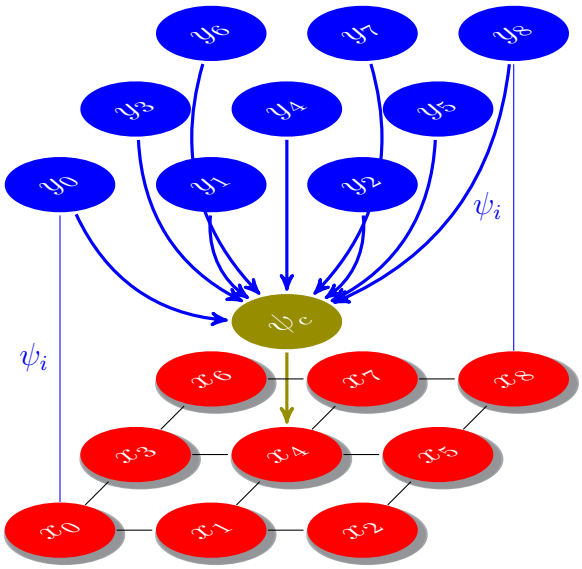


Figure 6.4: Network topology for the proposed higher order BP^n algorithm. As in the standard case, each node x_i obtains its unary potentials ψ_i and pairwise potentials ψ_{ij}. Additionally, x_i receives the higher order potential ψ_c, which is based on the observations y_i within a clique c.

of the configurations of the pixels in the corresponding clique, the potential function for higher oder potentials is based on the similarity of the observations given, see Equation 6.9.

We use the max-product algorithm to compute the MAP estimate of the MRF. Each message is a vector of dimension given by the number of labels. The message update $m_{ij}^t(x_j)$ passed from node x_i to its neighbor x_j at iteration t is composed by the unary, the pairwise, and the higher-order potential:

$$m_{ij}^t(x_j) \leftarrow \max_{x_i}\Big(\psi_i(x_i)\psi_{ij}(x_i,x_j)\psi_c(x_{c_j})\Big)\prod_{k\in\mathcal{N}_i\setminus j} m_{ki}(x_i), \tag{6.16}$$

where $k \in \mathcal{N}_i \setminus j$ denotes the neighbors of i without j and x_{c_j} is the clique x_c around node j. After t iterations, the belief for each node is computed as:

$$b_i(x_i) = \psi_i(x_i)\prod_{k\in\mathcal{N}_i} m_{ki}(x_i). \tag{6.17}$$

The label $l_k \in \mathcal{L}$ which minimizes $b_i(x_i)$ individually at each node is selected.

6.4 Summary

In this chapter we introduced the individual potential function to form the posterior energy of the MRF. The unary and pairwise costs are traditional functions and incorporated in several studies [100, 11]. To incorporate character properties by means of stroke features, we add higher-order potentials to the traditional formulation of pairwise MRFs. The

potential function for these higher-order costs are based on the P^n Potts model and aim to favour similar configurations within individual cliques.

Since we prefer local methods for statistical inference for the application of FBS in DIA, we proposed a new formulation of the standard definition of BP, resulting in $\mathrm{BP^n}$. With the $\mathrm{BP^n}$ formulation we are able to incorporate the higher-order potentials which incorporates the stroke model.

Chapter 7

Experiments and Results

This chapter presents the evaluation of the proposed method for FBS. The experiments are executed on a set of three different types of document images. The first test set includes ten selected images from the corpus of the *Missale Sinaiticum* (cf. Section 3.2). To evaluate the performance and to compare the results to related methods, we generated the Ground Truth (GT) data manually. Further experiments are carried out on a set of representative samples from the DIBCO 2009 test set [29]. This set consists of gray scale and color images of machine printed or handwritten text. The GT data is already available and provided by the organizers of the contest. Finally, we created synthetic data assembled of several artificial spectral images to highlight the robustness of the proposed method.

The proposed FBS method based on higher-order MRF models is compared to adaptive document image binarization proposed by [91]. This algorithm showed good performance in several comparative studies [91, 41]. Furthermore, we compare our approach to the binarization method propsed by [98] which describes an improved version of algorithm no. 26 from the DIBCO 2009. This algorithm showed the best performance in the contest and has beaten thirty-five other competitors.

We use the *precision* and *recall* rate and the corresponding F_1 measure to quantify the accuracy and to rank the performance of the different methods. We start the experiments with an evaluation of the influence of different selections of higher-order cliques in $\psi_c(x_c)$. The impact of the proposed BP^{n} optimization algorithm is compared to the inference methods proposed in Chapter 5. The second experiment in Section 7.6 evaluates the robustness of the proposed MRF approach on synthetic images. Finally, a general comparison of the methods is given in the third part of the experiments in Section 7.7.

7.1 Evaluation Method

The evaluation of FBS algorithms in DIA can be accomplished in several ways. These efforts can be divided into four categories [77]. An intuitive approach is a visual inspection and evaluation of the results by one or more human experts in order to grade visual criteria like broken symbols or noise [105]. Another possibility is to use the binarization results as an input for an OCR machine and to evaluate the results with respect to character or

word accuracy [41]. A popular method is to compare the results to previously generated GT data [29]. Finally, the last category uses a combination of human-oriented evaluation and OCR accuracy [113].

Since a visual inspection is too inaccurate to distinguish the influence of different higher-orders models and commercial OCR software is not available for our purpose on medieval languages, the performance is quantified and compared to previously generated GT data.

To measure the accuracy of the proposed algorithm and to compare the results to related methods we use the F_1 score. The F_1 score (also called F score or F measure) is an accuracy measure considering the precision and recall rate of a test. Therefore, we have to count the True Positive (TP), False Positive (FP), and the False Negative (FN) pixels of the results when compared to the GT data.

A pixel is classified as true positive if it is ON in both, the GT and the result, hence it is identified correctly. It is classified as false positive if it is ON only in the result, i.e. pixels which are detected as text, however they belong to the background. Finally, a pixel is classified as false negative if it is ON only in the GT image, i.e. false negatives are those pixels which are not detected as foreground or text in the result.

An example of an evaluated image with marked TP, FP, and FN pixels is illustrated in Figure 7.1. The real image or the GT can be seen on the left hand side, a possible result is shown in the middle, and the right hand side shows the evaluation including the labels TP, FP, and FN. In the example given, the resulting image can be interpreted in that effect, that the left hand side of the character "A" is under-segmented, and the right hand side is over-segmented.

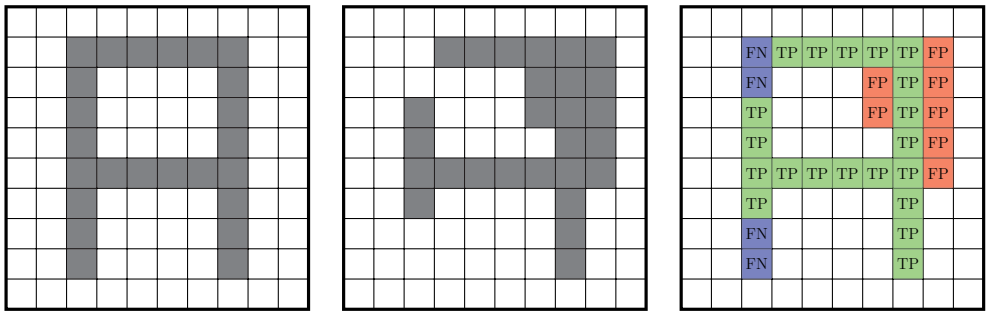


Figure 7.1: Illustration of true positive, false positive, and false negative classifications. The GT data is given in (a), a possible result is shown in (b), and (c) shows the evaluated image with the labels TP, FP and FN. Green pixels correspond to TP, blue pixels to FN, and red pixels show FP classifications.

Let $|TP|$ be the number of true positives, $|FP|$ the number of false positives, and $|FN|$ the number of pixels labeled as false negatives. The precision P and recall rate R

are given as follows:

$$P = \frac{|TP|}{|FP| + |TP|}, \tag{7.1}$$

$$R = \frac{|TP|}{|FN| + |TP|}. \tag{7.2}$$

The precision decreases when the algorithm detects too many pixels additionally, for instance, image noise, filled character holes, or filled character gaps. On the other side, the recall rate drops when the algorithm detects fewer pixels from the foreground, i.e. when parts of characters or even characters are missing.

The F_1 measure, as an overall metric to compare the individual methods, is calculated as:

$$F_1 = \frac{2RP}{R + P}. \tag{7.3}$$

7.2 Test Data

In this section we describe the digital document images on which the proposed algorithm is tested.

Missale Sinaiticum Images

The first category for the evaluation contains a set of selected images from the corpus of the *Missale Sinaiticum* with manually tagged GT data. The test set consist of folio 17 recto, 27 verso, 27 recto, 29 recto, 30 verso, 38 verso, 40 verso, 41 recto, 44 verso, and 53 verso. The generation of GT data was supported by philological experts in the field, by using a conventional graphics editing program and repainting each character on a separated layer which depicts the generated binary GT image.

The digital document images consist of 9 spectral images in each case as described in Section 3.2. The images have a radiometric resolution of 12 bit by a spatial resolution of 565 dpi. The data corpus including cultural history and philological aspects, material investigations, and details of the digitization process is described in [74].

DIBCO 2009 Images

The second test contains representative samples from the DIBCO 2009[1] which was arranged by [29] in the framework of the Tenth International Conference on Document Analysis and Recognition in Barcelona, Spain in 2009. The contest focused on the evaluation of document image binarization methods using a variety of scanned machine-printed and handwritten documents. The data material consists of gray scale and color images and includes five machine-printed and five handwritten images. The GT data was created following a semi-automatic procedure based on [77] and is provided by the organizers of the contest. These human segmented images act as the benchmark to evaluate the accuracy of the designed algorithms. The selection of the images in the dataset was made

[1] http://users.iit.demokritos.gr/ bgat/DIBCO2009/

so that it contains representative degradations as appear frequently in the real world, e.g. variable background intensity, shadows, smear, smudge, low contrast, bleed-through and show-through. An example can be seen in Figure 7.9(a).

Synthetic Data

The third data set consist of two synthetically generated RGB images. The synthetic images are based on a color image of a blank page from the *Missale Sinaiticum* with machine printed text. The images can be seen in Figure 7.21 (a) and (b). The images consists of three spectral bands (red, green and blue) in each case whereas the first image contains pure black text and the second one text in three different colors. To evaluate the methods at different degradations we add Gaussian white noise with zero mean and varying variance (0.005 0.05) to each spectral component.

7.3 Energy Function and Weighting Parameter

The energy function of the higher-order MRF model is composed of unary, pairwise, and higher-order potentials:

$$E(x) = \alpha \sum_{i \in \mathcal{V}} \psi_i(x_i) + \beta \sum_{i,j \in \mathcal{E}} \psi_{ij}(x_i, x_j) + \gamma \sum_{c \in \mathcal{S}} \psi_c(x_c). \tag{7.4}$$

Unary potentials depict the likelihood of the posterior function and the entities are modeled by a GMM where μ and Σ for foreground and background are found via EM. Pairwise potentials are based on the image gradient and the higher-order potential is based on a similarity measure of the pixels considered, see Chapter 6. For generality, each potential function is weighted with a weighting parameter. The unary potentials are weighted by α which corresponds throughout the experiments to 1.

The energy function solved by ICM is realized as a highly connected graph without higher-order potentials $\psi_c(x_c)$. The influence of the stroke model is weighted within pairwise connections with the weighting parameter β. For the inference of the higher-order models with BP and GC, the pairwise connections are weighted with $\beta = 1$ and the influence of the higher-order potential $\psi_c(x_c)$ is weighted with γ.

7.4 Overview of the Binarization Methods

As stated in the previous section, the higher-order energy function is solved with the proposed BP^{n} optimization method for higher-order models and α-expansion for robust P^n potentials[2] as representative method for GC. In the following, the method will be abbreviated with GC^{n}, where the superscript n denotes the order of the higher-order potentials. The third method for the minimization of the energy function is based on ICM

[2]The software library which implements the α-expansion for robust P^n potentials is available at http://research.microsoft.com/en-us/um/people/pkohli/code.html.

which is realized as a highly connected graph with pairwise connections. In addition, all three methods are performed on the standard pairwise models, i.e. a first order MRF.

We compare our method to two local thresholding algorithms, Adaptive Binarization (AB) [91] and binarization of historical document images using the Local Maximum and Minimum (LMM) proposed by [98]. The parameters are all set according to the recommendations within the reported papers. Furthermore, the proposed method based on spatial and spectral components is compared to a local performing k-means clustering algorithm proposed by [66]. Table 7.1 summarizes the algorithms used for our experiments.

Table 7.1: Overview of the algorithms applied in the experiments.

Method	Description and Reference
ICM	standard implementation of ICM [64]
BP	BP for first order MRF [25]
BP^n	BP for higher-order cliques [63]
GC	α-expansion for pairwise models [11]
GC^n	α-expansion for robust P^n potentials [49]
AB	adaptive document image binarization [91]
LMM	binarization using local max and min [98]
k-means	serialized k-means clustering [66]

7.5 Influence of the Higher-Order Stroke Model

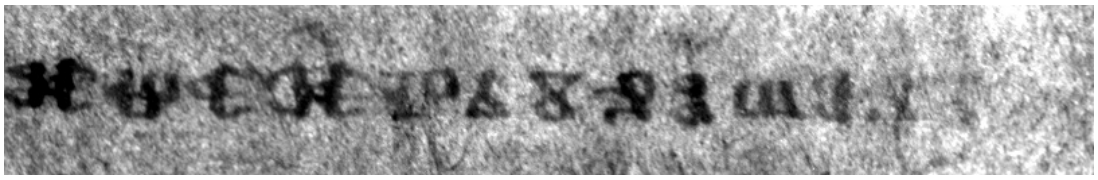
The goal of the first experiment is to analyze the behavior of the influence of the higher-order stroke model compared to the standard pairwise model. In the following, we compare the influence of the order n of the MRF and varying weighting parameters β and γ. As already stated, pairwise connections are used in the majority of the studies but show limited expressiveness [88]. We show the performance of the proposed method on a degraded folio of the *Missale Sinaiticum* and on an image from the DIBCO 2009 test set.

7.5.1 Missale Sinaiticum

When applied on the multispectral images of the *Missale Sinaiticum* the proposed algorithm can utilize its capability of incorporating spatial and spectral features simultaneously. Figure 7.2 shows an exemplary image of the corpus to illustrate the quality of the images. The image in the first row shows line 9 from folio 29 recto, represented by the

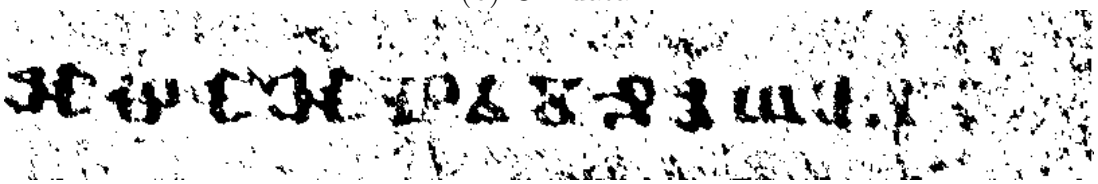
spectral image B-P 450. This spectral component tends to have the best contrast between text and background within the full set of the spectral images, cf. also Figure 3.5 and Figure 3.7 in Chapter 3. The detail of folio 29 recto in Figure 7.2(a) shows 13 characters whereas the last two are already hard to detect. The second row shows the manually tagged GT data and the third row shows the output of the AB method presented by [91] when applied on the spectral image B-P 450.

It can be seen that the binary image contains background noise which affects the precision rate ($P = 0.63$) and the two right outermost characters are due to its low contrast missing. These false negatives decrease the recall rate R to 0.60. Neglected and additional pixels have approximately the same rate and the corresponding F_1 measure yields 0.61.



(a) Single band image B-P 450.

(b) GT data.



(c) AB, $P = 0.63$, $R = 0.60$.

Figure 7.2: Detail from folio 29 recto: B-P 450, GT data, and output of AB. It can be seen that the output of AB contains noise and the right outermost character is missing.

ICM

Table 7.2 shows the precision and recall rate for the result of the posterior energy when minimized with ICM. The results depend on the weighting parameter β and the influence of the order n. The values for β range within $\{0.07, 0.1, 0.2, 0.3\}$ and n is between 1 and 5. Pairwise connections correspond to $n = 1$ and $n = 4$ refers to the stroke model since experiments in the test set of the *Missale Sinaiticum* showed an average stroke width of five pixels ($\varnothing = 5$), cf. also Figure 6.2.

The value for the precision is low for a small order of n and increases for higher-order connections. This low precision rate especially for $n \leq 3$ results primarily from the background noise in the binary image. The output of the MRF model with pairwise connections is similar to the result of AB. The stroke model is not considered in this case.

When using pairwise connections influenced by $\beta = 0.07$, the F_1 measure constitutes 0.70 with a precision rate of 0.63 and a recall rate of 0.78.

On the other side, when the order n and its influence by means of β increase ($n = 5, \beta = 0.3$), the precision rate rises to 0.82 due to less background noise, but the recall rate drops synchronously to 0.53 due to missing characters. The reason is in the influence of the pixels environment, which is too influential and characters with low contrast or narrow strokes fail the segmentation.

Table 7.2: Precision, recall, and F_1 for the output of ICM with varying order n and β for folio 29 recto.

β		1	2	3	4	5
0.07	P	0.63	0.65	0.66	0.71	0.70
	R	0.78	0.78	0.78	0.77	0.78
	F_1	0.70	0.71	0.72	0.74	0.74
0.1	P	0.64	0.66	0.67	0.76	0.71
	R	0.78	0.78	0.78	0.75	0.78
	F_1	0.70	0.72	0.72	0.75	0.74
0.2	P	0.64	0.66	0.68	0.80	0.80
	R	0.79	0.79	0.79	0.72	0.69
	F_1	0.70	0.72	0.73	0.76	0.74
0.3	P	0.69	0.75	0.79	0.81	0.82
	R	0.77	0.76	0.73	0.68	0.53
	F_1	0.73	0.75	0.76	0.74	0.64

Figure 7.3 shows resulting images illustrating the influence of the order n and the weighting parameter β. The first row shows the output of a pairwise connected random field with low influence ($\beta = 0.07$). The opposite adjustment can be seen in the third row. Here, the influence of surrounding pixels n equals 5 and its weighting parameter is too high which in fact results in less background noise, but characters with low contrast are already missing. The best solution is obtained with $n = 4$, which is in accordance to the proposed stroke model, and a weighting parameter $\beta = 0.2$ with $F_1 = 0.76$. The output of the model adjusted with $n = 4$ and $\beta = 0.2$ can be seen in Figure 7.3(b). The background noise has in contrast to the result in the first row vanished and the character on the right hand side are partially detected. Disadvantageously, the stroke width increased and some character gaps merged.

The experiments show, that the smaller the considered neighborhood system, the more noise emerges in the background, and on the other side, a neighborhood set considering too much pixels (i.e. when n exceeds the stroke diameter) leads to missing characters or to closed character gaps and closed character holes. The influence of β is likewise. The smaller β the more noise we have and values chosen too big causes missing characters. This situation is reflected in a 3D plot in Figure 7.4 which illustrates the precision and recall rate and the F_1 measure against the influence of n and β. The diagram on the left hand side shows the precision with reference to the order n on the x-axis and β on the y-axis. It can be seen that precision rises for $n \geq 4$ and $\beta \geq 0.2$. On the other side the

recall rate drops. The corresponding diagram for the F_1 measure is given below. Best performance is obtained with $n = 4$ and $\beta = 0.2$. The F_1 rate decreases for values with $n \geq 5$ and $\beta \geq 0.3$, i.e. when n exceeds the stroke diameter.

Belief Propagation

The results for the optimization of the posterior energy with BP and the proposed message update rule for higher-order models, BP^n, are given in Table 7.3. The influence of the order and its weighting parameter is not as crucial as for ICM, but still improves the accuracy. Comparable to the findings of the influence from n and β for inference based on ICM, the precision rate rises with increasing n and γ, but the recall rate drops simultaneously.

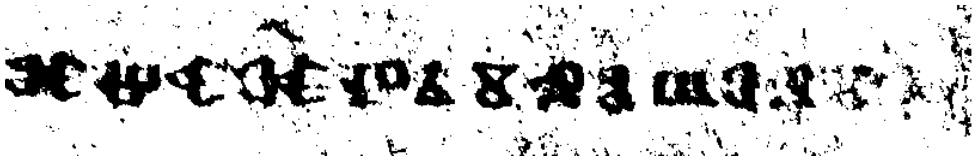
According to the stroke model, the best results are obtained when considering all pixels within $n = 4$. Then, the precision rate P consititues 0.83 and the recall rate $R = 0.69$ with a corresponding F_1 measure of 0.75 (with $\gamma = 0.3$). Resulting images can be seen in Figure 7.5. Compared to the images from ICM it can be seen that more background noise is present, but the right outermost character ℧ is segmented more exactly. The 3D plots in Figure 7.6 illustrate again the influence of n and γ. The findings are in turn comparable to the ICM based results. The precision rises with increasing n and γ, and vice versa the recall rate drops. Again, the best value for F_1 is obtained with the proposed stroke model.

Table 7.3: Precision, recall, and F_1 for the output of BP^n with varying order n and γ for folio 29 recto.

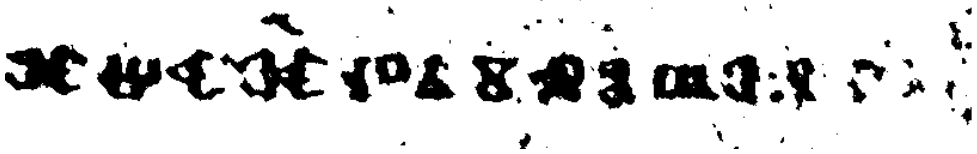
γ		1	2	3	4	5
0.1	P	0.65	0.73	0.74	0.75	0.76
	R	0.75	0.72	0.70	0.70	0.69
	F_1	0.70	0.72	0.72	0.72	0.72
0.2	P	0.66	0.75	0.76	0.77	0.78
	R	0.75	0.70	0.69	0.71	0.66
	F_1	0.70	0.72	0.72	0.74	0.72
0.3	P	0.68	0.79	0.81	0.83	0.84
	R	0.75	0.65	0.66	0.69	0.58
	F_1	0.72	0.71	0.73	0.75	0.68
0.4	P	0.68	0.82	0.84	0.86	0.87
	R	0.75	0.61	0.57	0.53	0.50
	F_1	0.72	0.70	0.68	0.65	0.64

Graph Cuts

Finally, Table 7.4 presents the output when minimizing the energy of the random field with the α-expansion move (GC^n). Since the method finds a global minimum, the influence of the neighborhood system is negligible. This can be especially seen in the 3D plots in Figure 7.8 when compared n and γ against the evaluation metrics. However, the precision rises and the recall rate drops with increasing values for n and γ.



(a) ICM; $n = 1$, $\beta = 0.07$; $P = 0.63$, $R = 0.78$.



(b) ICM; $n = 4$, $\beta = 0.2$; $P = 0.80$, $R = 0.72$.

(c) ICM; $n = 5$, $\beta = 0.3$; $P = 0.82$, $R = 0.53$.

Figure 7.3: Folio 29 recto: resulting images after the ICM based FBS: (a) shows the output of pairwise connections with $n = 1$ and $\beta = 0.07$, (b) shows the output with the best F_1 score with $n = 4$ and $\beta = 0.2$, consider that even the right outermost characters are due to their low contrast partially segmented, and (c) shows the result when the influence of n and β exceeds the stroke diameter ($n = 5$ and $\beta = 0.3$).

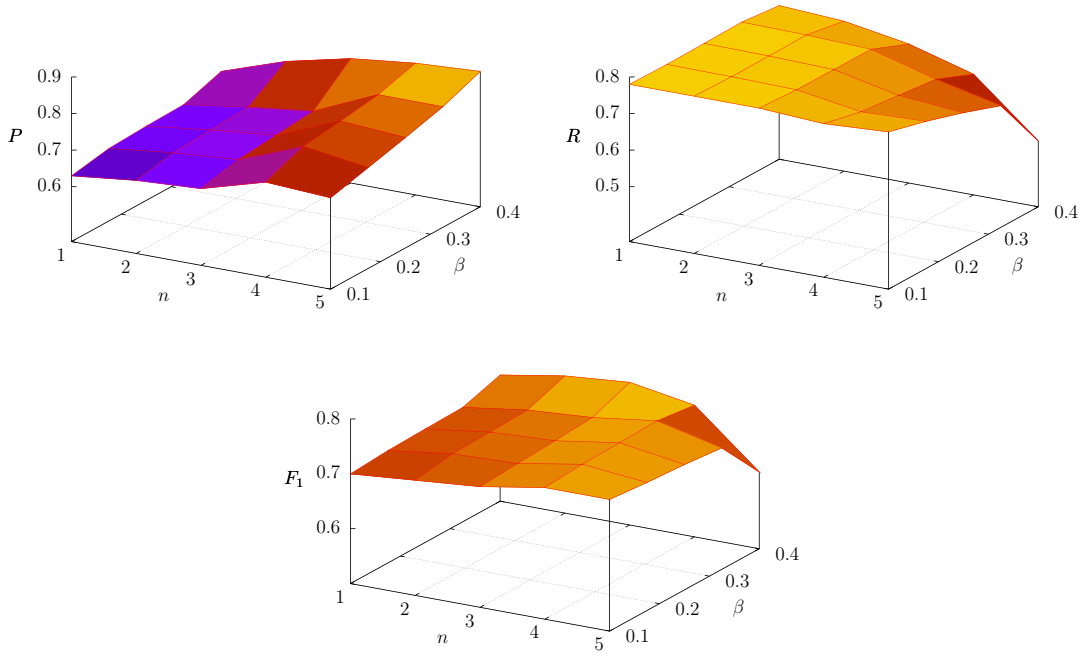
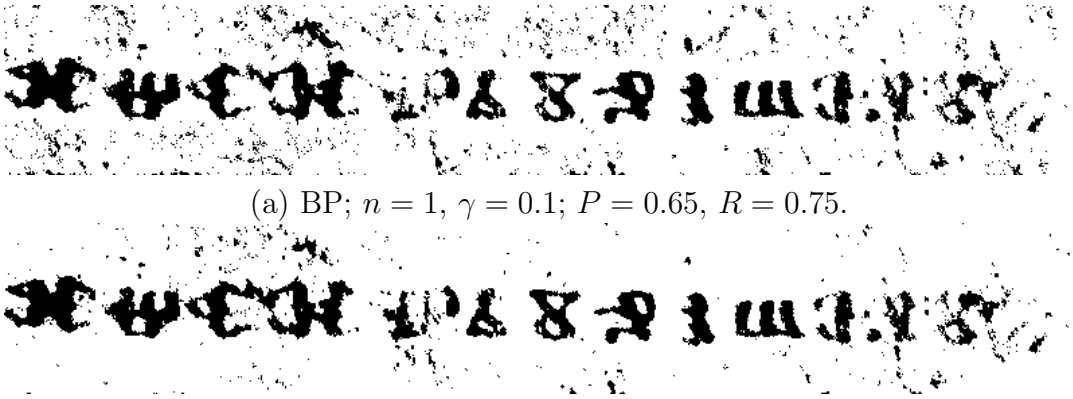


Figure 7.4: Precision P, recall R, and F_1 for the results with ICM on folio 29 recto. The precision rises with increasing n and β, and vice versa the recall rate drops. As the diagram for the F_1 measure shows, best result are obtained with $n = 4$. Since the stroke width of the *Missale Sinaiticum* averages five, the observations are in conformity with the proposed stroke model.



(a) BP; $n = 1$, $\gamma = 0.1$; $P = 0.65$, $R = 0.75$.

(b) BP$^{\mathrm{n}}$; $n = 4$, $\gamma = 0.3$; $P = 0.83$, $R = 0.69$.

Figure 7.5: Folio 29 recto: resulting images after the proposed BP$^{\mathrm{n}}$ based FBS: (a) shows $n = 1$ and $\gamma = 0.1$ and (b) shows the result for $n = 4$ and $\gamma = 0.3$. The use of higher-order cliques results in turn in a more accurate segmentation.

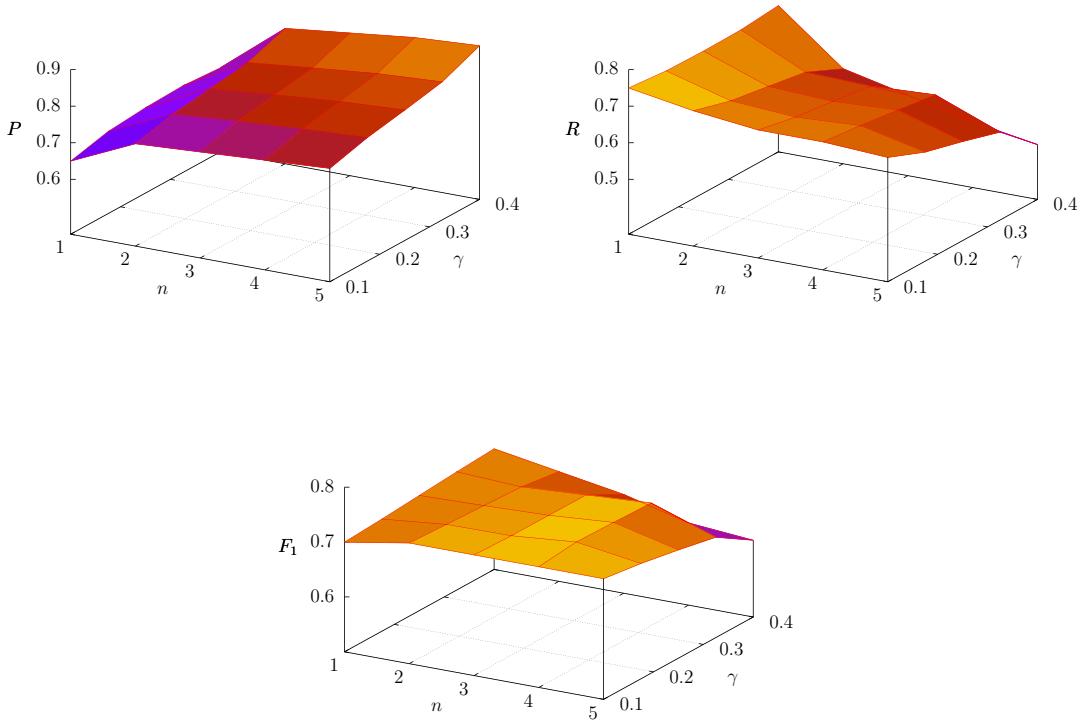


Figure 7.6: Precision P, recall R, and F_1 for the results with BP$^{\mathrm{n}}$ on folio 29 recto. Again, the precision rises with increasing n and γ, and vice versa the recall rate drops. The diagram from the F_1 measure shows, that the best result is in turn obtained with $n = 4$.

Concerning the precision and recall rate, we obtain 0.78 for the precision, 0.65 for the recall rate, and the corresponding F_1 measure yields 0.71 for the proposed stroke model. Notice, that the values vary particularly only in the 3rd decimal place, though the values are rounded for the table given. Some results are shown in Figure 7.7. Noticeable, the right outermost character is clearly segmented and even the outline of an adjacent character is noticeable.

Table 7.4: Precision, recall, and F_1 for the output of GC^n with varying order n and γ for folio 29 recto.

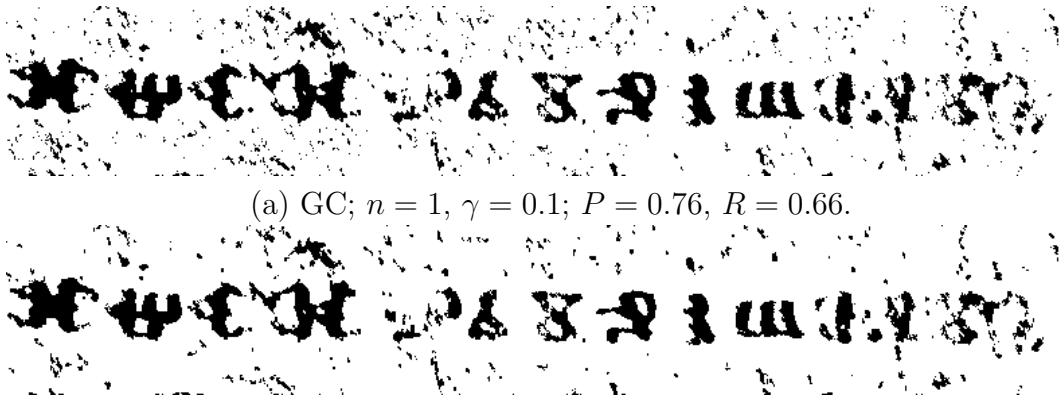
γ		1	2	3	4	5
0.1	P	0.76	0.77	0.78	0.78	0.78
	R	0.66	0.65	0.65	0.65	0.65
	F_1	0.71	0.71	0.71	0.71	0.71
0.2	P	0.76	0.77	0.78	0.78	0.78
	R	0.66	0.65	0.65	0.65	0.65
	F_1	0.71	0.71	0.71	0.71	0.71
0.3	P	0.76	0.77	0.78	0.78	0.78
	R	0.66	0.65	0.65	0.65	0.65
	F_1	0.71	0.71	0.71	0.71	0.71
0.4	P	0.76	0.77	0.78	0.78	0.78
	R	0.66	0.65	0.65	0.65	0.65
	F_1	0.71	0.71	0.71	0.71	0.71

7.5.2 DIBCO 2009 Images

In this section of the experiments, we evaluate the influence of the stroke model and its weighting factor on an image from the DIBCO 2009 test set. In contrast to images of the *Missale Sinaiticum*, this set contains conventional RGB color and panchromatic images and the quality of the legibility is considerably better. An example can be seen in Figure 7.9(a). It shows a RGB image with German text, containing document ink bleed-through from the recto page between the lines. The corresponding GT can be seen in the second row and the third row shows the output of AB. The binarized image contains noise between the lines caused through document ink bleed-through. The precision rate P constitutes 0.93 and the recall rate R of the AB method yields 0.85 due to missing character gaps and narrow strokes. The F_1 measure constitutes 0.89.

ICM

Table 7.5 shows the evaluation metrics for the optimization of the MRF energy function with ICM with varying n and β. Due to the quality of the images, the precision and recall rates are better than for the images of the *Missale Sinaiticum*. However, the findings are comparable to the previous experiments on the *Missale Sinaiticum*. The F_1 measure shows the highest rate for $n = 3$ and $\beta = 0.3$, but the differences are not as big as



(a) GC; $n = 1$, $\gamma = 0.1$; $P = 0.76$, $R = 0.66$.

(b) GC$^{\mathrm{n}}$; $n = 4$, $\gamma = 0.3$; $P = 0.78$, $R = 0.65$.

Figure 7.7: Folio 29 recto: resulting images after the GC$^{\mathrm{n}}$ based FBS: (a) shows $n = 1$ and $\gamma = 0.1$ and (b) shows the output of the proposed stroke model with $n = 4$ and $\gamma = 0.3$.

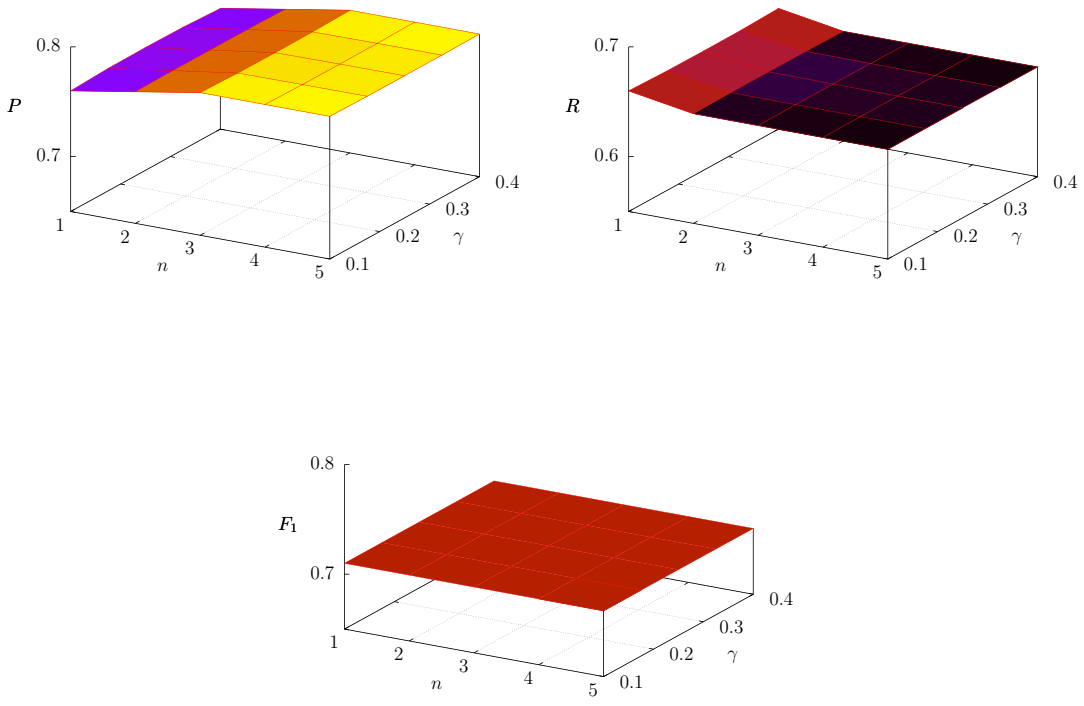
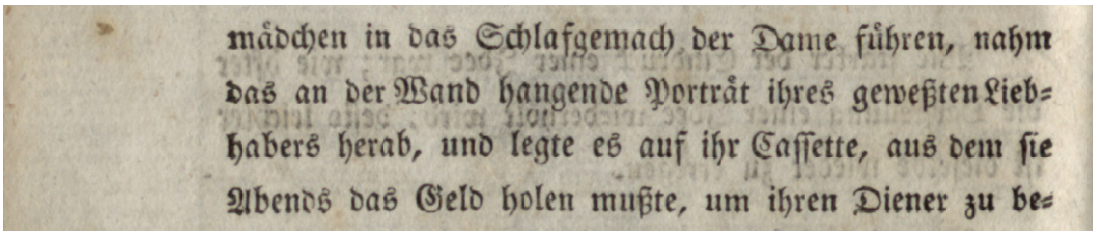
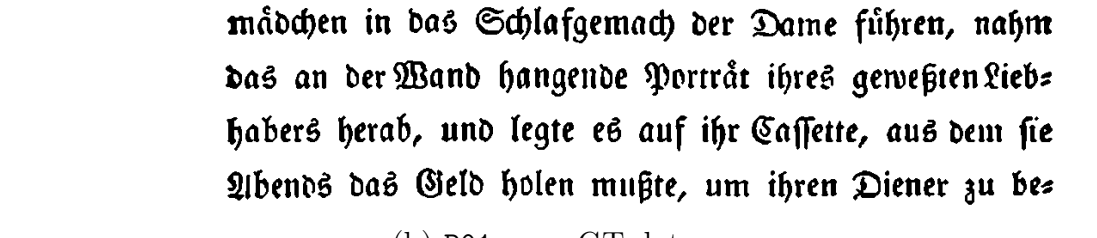


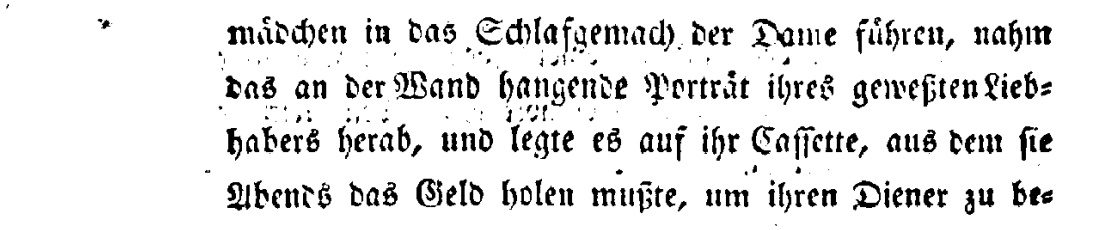
Figure 7.8: Precision P, recall R, and F_1 for the results with GC$^{\mathrm{n}}$ on folio 29 recto. With the global based inference, the influence of n and γ is not as crucial as for the local based methods ICM and BP.



(a) `P01.png`



(b) `P01.png`, GT data.



(c) AB; $P = 0.93$, $R = 0.85$.

Figure 7.9: Image `P01.png` from the DIBCO 2009 test set and the corresponding GT data in (b), (c) shows the output of the AB. Due to the low contrast of the image, the characters are thinner in the resulting binary image and the bleed through from the recto page causes noise.

for the degraded manuscripts of the *Missale Sinaiticum*. This results from the more discriminative likelihood model.

Compared to AB, particularly the precision rate increased to $P = 0.94$, due to less noise in the background. The bleed through of the ink is suppressed, except for the middle part between line two and three. The recall rate equals approximately 0.90 throughout the experiments.

As already mentioned, we obtain the best F_1 score for $n = 3$ and $\beta = 0.3$ which is again coherent with the stroke width, which averages approximately 4 pixels for the characters in image `P01.png`. Some results can be seen in Figure 7.10. The first row presents the results with $\beta = 0.3$ and $n = 3$. In Figure 7.10(b), the order of n exceeds the average stroke width of the image which lowers especially the recall rate to $R = 0.84$ ($n = 5$ and $\beta = 0.5$). Parts of character with low contrast are already missing, cf. the "W" or "g" in the fourth and fifth word of the second row. The suppression of background noise and bleed through with increasing n and β is neglected by filled character holes, obtainable within the characters "s" and "e".

The visualization of the influence of n and β to precision and recall is similar to the findings of the *Missale Sinaiticum*. Again, the recall drops with increasing n and β and

Table 7.5: Precision, recall, and F_1 for the output of ICM with varying order n and β for `P01.png`.

β		1	2	3	4	5
0.1	P	0.93	0.93	0.93	0.93	0.93
	R	0.90	0.90	0.90	0.90	0.90
	F_1	0.91	0.91	0.91	0.91	0.92
0.2	P	0.93	0.93	0.93	0.93	0.94
	R	0.90	0.90	0.90	0.90	0.90
	F_1	0.91	0.91	0.91	0.92	0.92
0.3	P	0.93	0.93	0.94	0.94	0.94
	R	0.90	0.90	0.90	0.89	0.89
	F_1	0.91	0.92	0.92	0.92	0.91
0.4	P	0.93	0.94	0.94	0.94	0.94
	R	0.90	0.90	0.89	0.88	0.87
	F_1	0.91	0.92	0.92	0.91	0.90
0.5	P	0.93	0.94	0.94	0.94	0.94
	R	0.90	0.89	0.89	0.87	0.84
	F_1	0.91	0.92	0.92	0.90	0.89

the precision shows contrary behavior, see Figure 7.11.

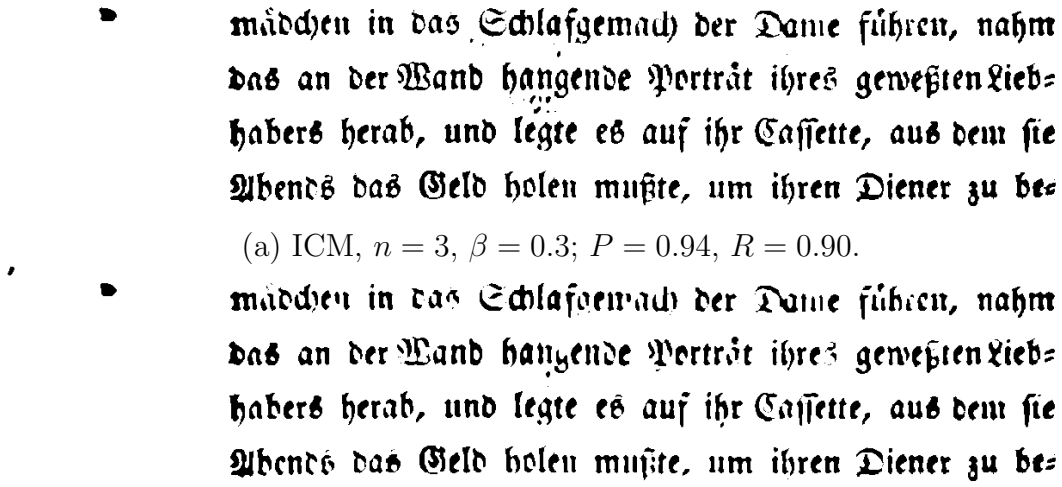
Belief Propagation

Table 7.6 presents the result for the same image when applied the proposed BP^{n} for inference. Best performance is obtained for $n = 3$ and $\gamma = 0.3$ with $F_1 = 0.91$. Again, the differences between the varying values for n and γ are less crucial than for the highly degraded documents of the *Missale Sinaiticum*. The tendency of increasing precision and decreasing recall rate when rising n and γ is again observable. The precision rises up to 0.97 and the recall rate drops to 0.74 when n exceeds the mean stroke width. Figure 7.12 shows resulting images for a preferred configuration in the first row and a less optimal configuration with too much influence of n and γ in the second row. The influence of n and γ with respect to precision, recall, and F_1 is plotted in Figure 7.13. The progress of precision, recall, and F_1 is again comparable to the previous findings.

Graph Cuts

Table 7.7 shows the results from the global α-expansion move for minimization. The results are similar to AB. The values for P, R, and F_1 are relatively constant and drop with increasing n and γ, c.f. the diagrams in Figure 7.15.

Resulting images can be seen in Figure 7.14. The first row shows the output for a preferred configuration with $n = 3$ and $\gamma = 0.1$ and the second row a suboptimal results with parameters $n = 5$ and $\gamma = 0.4$. It can be seen that some characters and even words merge, particularly visible in the fifth word of the second row and the first character in



(a) ICM, $n = 3$, $\beta = 0.3$; $P = 0.94$, $R = 0.90$.

(b) ICM, $n = 5$, $\beta = 0.5$; $P = 0.94$, $R = 0.84$.

Figure 7.10: `P01.png`, resulting images after the ICM based FBS: (a) shows a preferred configuration with $n = 3$ and $\beta = 0.3$ and (b) shows an resulting image when n exceeds the mean stroke width ($n = 5$ and $\beta = 0.5$).

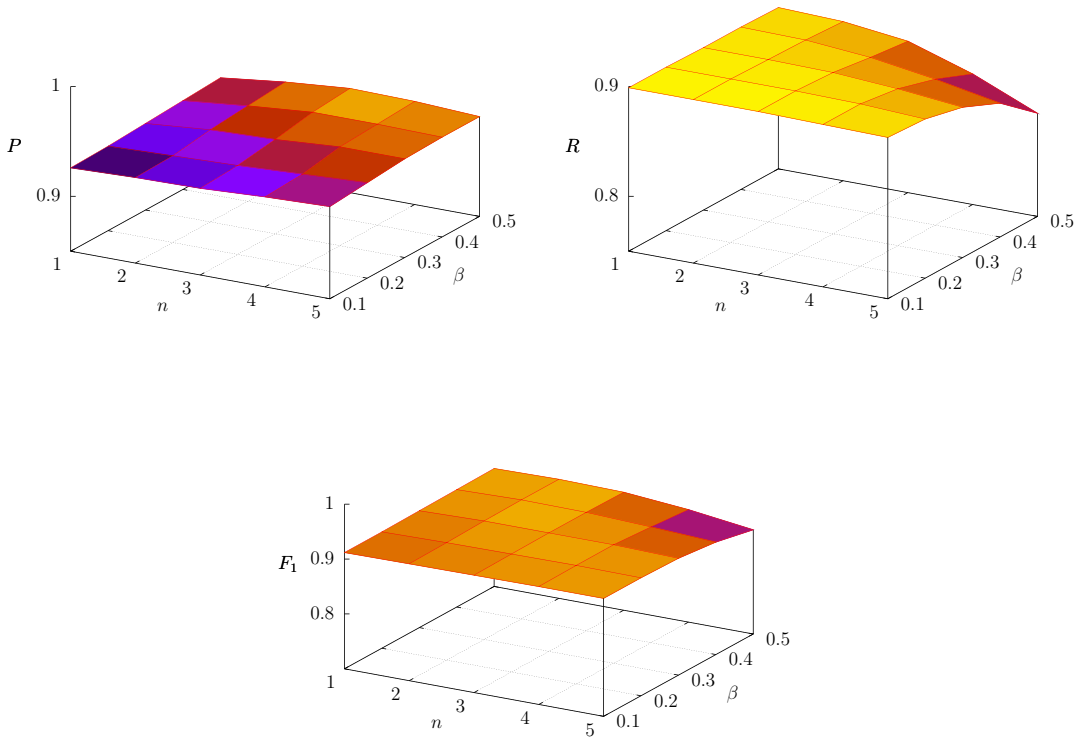
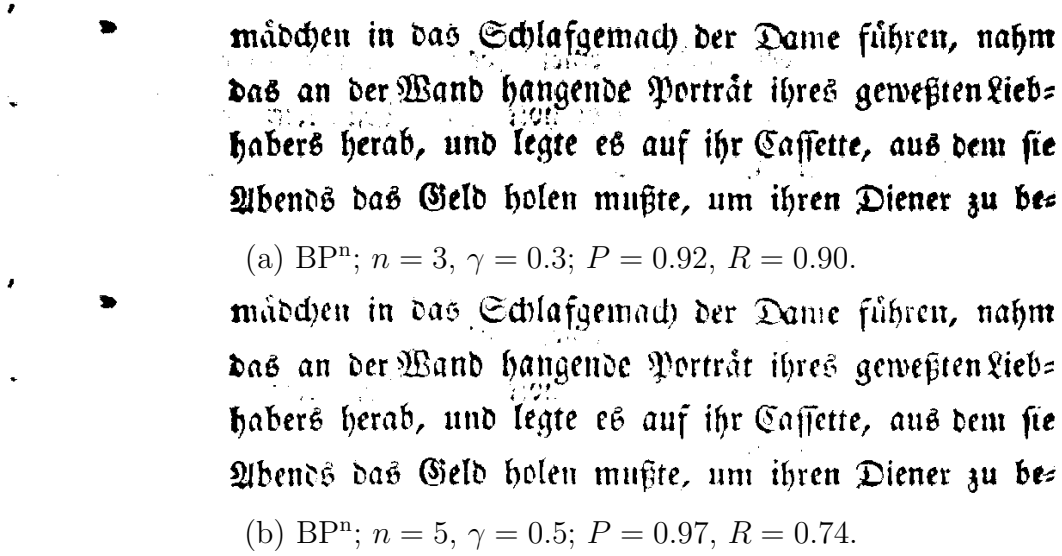


Figure 7.11: Precision P, recall R, and F_1 for the results with ICM on `P01.png`. The precision rises with increasing n and β, and the recall rate drops.



(a) BP^{n}; $n = 3$, $\gamma = 0.3$; $P = 0.92$, $R = 0.90$.

(b) BP^{n}; $n = 5$, $\gamma = 0.5$; $P = 0.97$, $R = 0.74$.

Figure 7.12: `P01.png`: resulting images after the BP^{n} based FBS: (a) shows $n = 3$ and $\gamma = 0.3$ and (b) shows the result when n exceeds the mean stroke width ($n = 5$ and $\gamma = 0.5$).

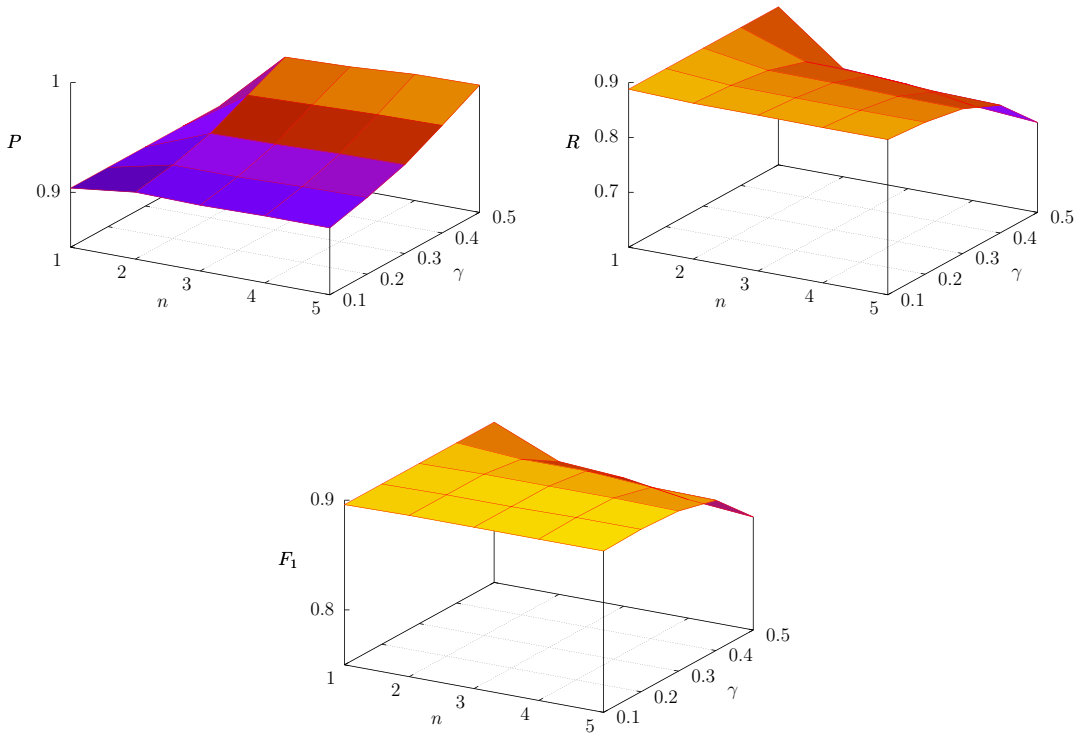


Figure 7.13: Precision P, recall R, and F_1 for the results with BP^{n} on `P01.png`. Again, the precision rises with increasing n and γ, and the recall rate drops.

Table 7.6: Precision, recall, and F_1 for the output of BP^n with varying order n and γ for `P01.png`.

γ		1	2	3	4	5
0.1	P	0.90	0.91	0.91	0.91	0.91
	R	0.89	0.88	0.88	0.88	0.88
	F_1	0.90	0.90	0.90	0.90	0.90
0.2	P	0.90	0.92	0.92	0.92	0.92
	R	0.89	0.88	0.88	0.88	0.88
	F_1	0.90	0.90	0.90	0.90	0.90
0.3	P	0.90	0.92	0.92	0.93	0.93
	R	0.89	0.89	0.90	0.87	0.86
	F_1	0.90	0.90	0.91	0.90	0.90
0.4	P	0.90	0.94	0.94	0.95	0.95
	R	0.89	0.85	0.84	0.84	0.83
	F_1	0.90	0.89	0.89	0.89	0.89
0.5	P	0.90	0.96	0.96	0.96	0.97
	R	0.89	0.80	0.79	0.77	0.74
	F_1	0.90	0.87	0.87	0.86	0.84

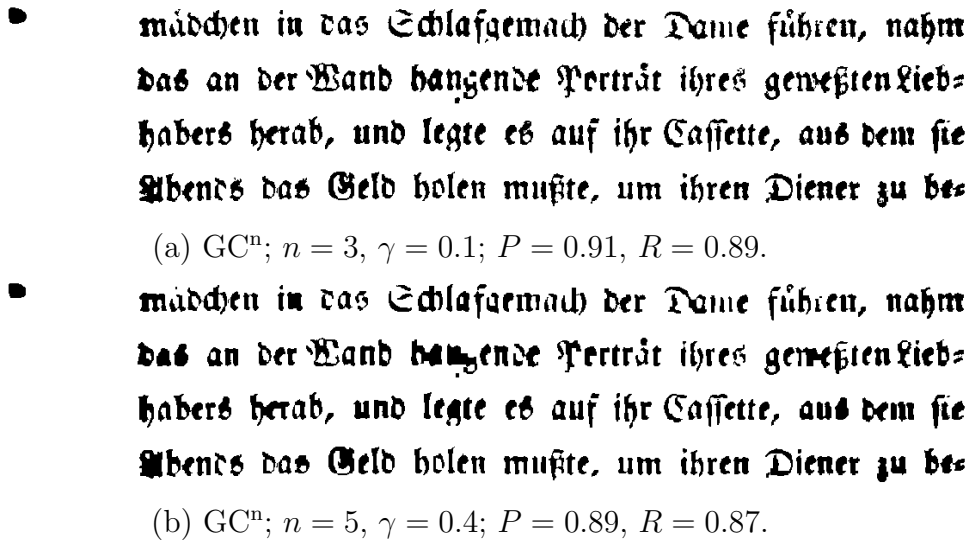
the last row. The recall rate for this experiment averages approximately 0.90 except for increasing n and γ where the rate decreases.

7.5.3 General Aspects

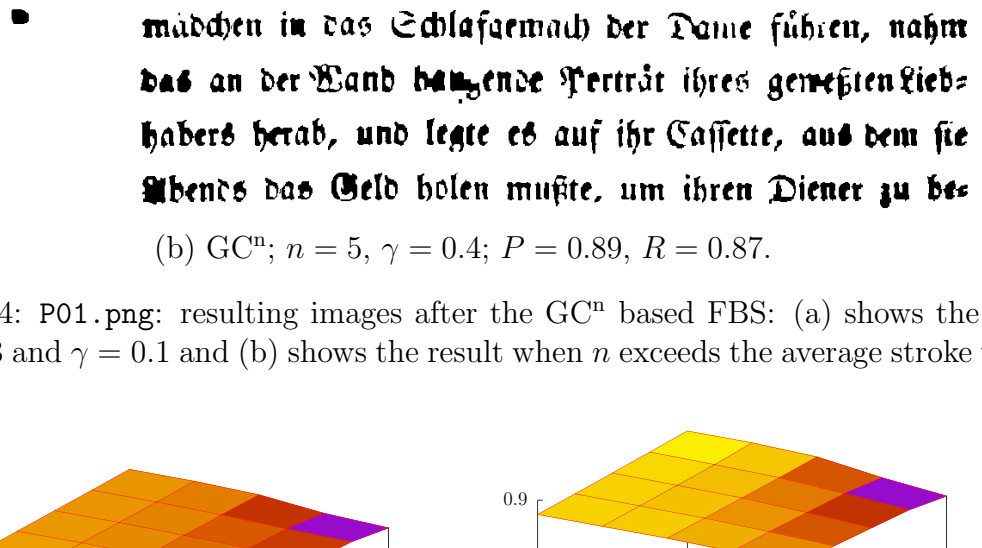
The results show that the smaller the considered neighborhood system or the order n of the MRF (e.g. the pairwise model with $n = 1$), the more noise emerges in the background and on the other side, a neighborhood system exceeding the mean stroke width leads, for instance, to missing characters, to closed character gaps, or to merged holes in characters. The influence of β and γ is likewise. The smaller β or γ the more noise we have and values chosen too big cause missing characters. Thus, due to the consideration of spatial context by means of the prosed stroke model we obtain less noise within the background and text regions. Furthermore, characters with very low contrast have been partially segmented with the proposed stroke model where AB detects only noise.

The influence of n and respectively β and γ is illustrated in Figure 7.16. It shows that the precision increases with rising n or β/γ and the recall rate decreases. The higher the stroke radius considered, the higher is the possibility that background labels or remote text pixels are included in the current clique which affects the configuration of the current pixel.

The influence of n and β/γ becomes noticeable particularly in the output of the ICM based minimization of the posterior energy. This effect is due to the local inference and the label-feedback after each iteration, i.e. each pixel can request the state of its surrounding pixels. BP estimates the configuration of individual pixels at the end of the iterative message passing. The main disadvantage of ICM is its poor computational performance.



(a) GC$^{\mathrm{n}}$; $n = 3$, $\gamma = 0.1$; $P = 0.91$, $R = 0.89$.



(b) GC$^{\mathrm{n}}$; $n = 5$, $\gamma = 0.4$; $P = 0.89$, $R = 0.87$.

Figure 7.14: `P01.png`: resulting images after the GC$^{\mathrm{n}}$ based FBS: (a) shows the result with $n = 3$ and $\gamma = 0.1$ and (b) shows the result when n exceeds the average stroke width.

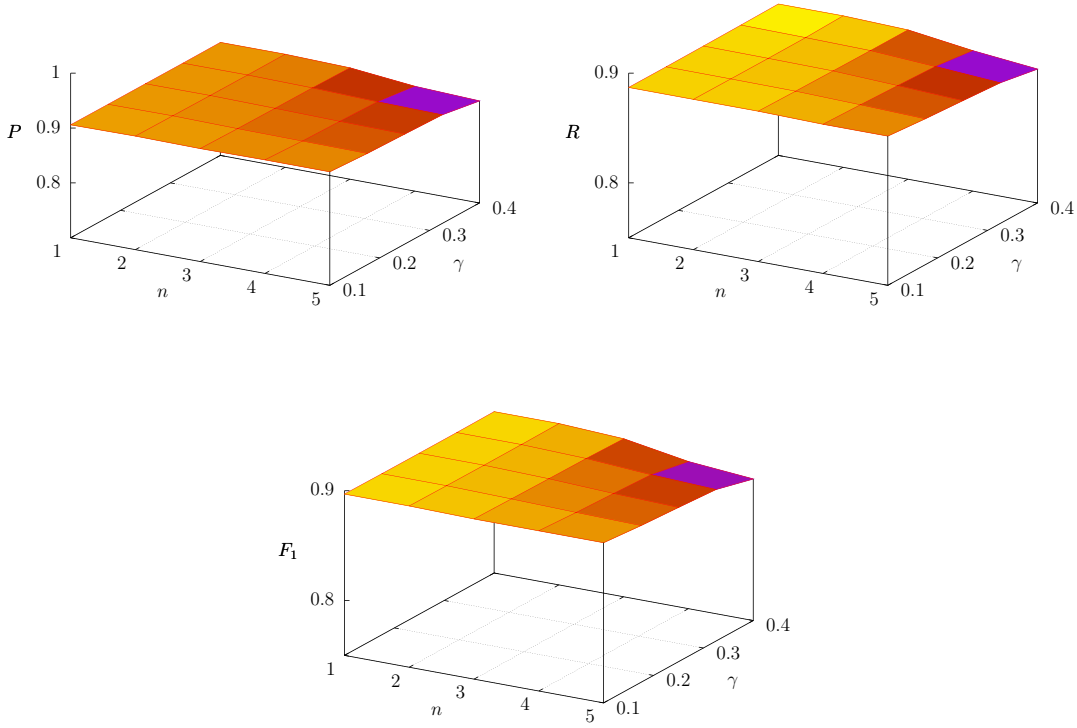


Figure 7.15: Precision P, recall R, and F_1 for the results with GC$^{\mathrm{n}}$ on `P01.png`. With the global based method for statistical inference, the influence of n and γ is not as crucial as for the local based methods ICM and BP.

Table 7.7: Precision, recall, and F_1 for the output of GC^n with varying order n and γ for `P01.png`.

γ		1	2	3	4	5
0.1	P	0.91	0.91	0.91	0.91	0.91
	R	0.89	0.89	0.89	0.89	0.89
	F_1	0.90	0.90	0.90	0.90	0.90
0.2	P	0.91	0.91	0.91	0.90	0.90
	R	0.89	0.89	0.89	0.89	0.88
	F_1	0.90	0.90	0.90	0.90	0.89
0.3	P	0.91	0.91	0.90	0.90	0.90
	R	0.89	0.89	0.89	0.88	0.88
	F_1	0.90	0.90	0.90	0.89	0.89
0.4	P	0.91	0.91	0.90	0.89	0.89
	R	0.89	0.89	0.89	0.88	0.87
	F_1	0.90	0.90	0.89	0.88	0.88

BP is in fact also a local inference algorithm but has no label feedback of neighboring pixels after each iteration. However, the influence of the proposed stroke model improves the accuracy and the method has a better computational performance. In the case of FBS for text separation, α-expansion could not outperform BP.

7.6 Synthetic Data

For the experiments on the synthetic images we apply the algorithms listed in Table 7.1 to two different images. The MRF based approaches are applied in pairwise manner and with the proposed higher-order stroke model, i.e. n equals the mean stroke width of the images. The input images can be seen in Figure 7.21(a) and (b). In the following we degrade the original image in in ten steps by adding Gaussian noise with varying variance σ. For the synthetic image with black text, σ varies between 0.01 and 0.05, for the image with varying color, σ varies between 0.00 and 0.04. A degraded version of both images is given in Figure 7.21(c) and (d).

The results for the image with black text achieve nearby 100% throughout the varying noise rate for the MRF based approaches, see Table 7.8. The MRF approaches with the proposed stroke model outperform the local thresholding methods and k-means clustering and the higher-order models show better results than pairwise models. The results for the recently proposed LMM method [98] averages 0.97 for the precision, 1.00 for recall, and 0.98 for the F_1 measure. AB shows minor results with an F_1 measure of 0.93 and k-means shows an F_1 measure of 0.99. An overview of the average values for precision, recall, and F_1 can be seen in Figure 7.19.

The progress of precision, recall, and F_1 with increasing noise is shown in the diagrams in Figure 7.17. Due to their sensitivity to noise, the local thresholding methods AB and LMM show a decreasing precision with increasing noise, see 7.17(a). As already stated, the MRF based approaches show throughout the increasing noise rate approximately

100% for precision and recall. Thus, the incorporation of spatial and spectral information is robust against noise. Noise in the background is suppressed and the proposed stroke model which considers not only the color of individual pixels, but also the neighboring labels, shows advantages. Some resulting images can be seen in Figure 7.21(e) and (g). While the proposed BP^n method suppresses the noise in the background, AB fails in the presence of noise.

When we consider the synthetic image with varying text color, the influence of noise is more crucial and the performance is, as expected, not that high. However, the MRF based approaches still obtain a good performance compared to AB, LMM, and k-means. Table 7.9 shows the results for a synthetic image with varying text color and varying Gaussian noise (σ varies between 0.00 and 0.04). The best performance is given by the highly connected ICM based approach with $n = 5$. The proposed BP^n with $n = 5$ shows a similar performance, however the computational complexity is significantly better. The progress of precision, recall, and F_1 can be seen in the diagrams in Figure 7.18. It is observable that the precision for ICM with $n = 5$ and the proposed BP^n outperform the other methods. The recall rate is very high for LMM and k-means, but the results are useless since the precision is low.

The F_1 measure for AB and LMM drops already below 0.50 with $\sigma > 0.01$ for LMM and $\sigma > 0.04$ for AB. Best results for F_1 can be obtained with ICM with $n = 5$ and the BP^n algorithm for the proposed stroke model. The overview of the average values for precision, recall, and F_1 can be seen in Figure 7.20.

Summarized, the approaches based on MRF and especially the proposed stroke model outperform AB, LMM, and k-means. The two local minimization methods (ICM and BP) show superior results compared to inference based on GC. Since AB and LMM estimate a threshold in local regions the methods fail for Gaussian noise in the synthetic image.

Table 7.8: Precision, recall, and F_1 for a synthetic image with black text and added Gaussian noise. The stroke width of the synthetic text averages 6 pixels.

σ	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045	0.050	AVG
ICM, $n = 1$										
P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
R	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00
F_1	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00
ICM, $n = 5$										
P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
F_1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
BP										
P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
R	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	1.00
F_1	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00
$\mathrm{BP^n}$, $n = 5$										
P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
R	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	1.00
F_1	1.00	1.00	1.00	1.00	1.00	1.0	0.99	0.99	0.99	1.00
GC										
P	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00
R	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.99
F_1	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00
$\mathrm{GC^n}$, $n = 5$										
P	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
R	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99
F_1	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	1.00
LMM										
P	1.00	1.00	1.00	1.00	0.99	0.98	0.95	0.95	0.88	0.97
R	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	1.00
F_1	1.00	1.00	1.00	1.00	0.99	0.99	0.97	0.97	0.93	0.98
AB										
P	0.99	0.97	0.94	0.90	0.86	0.83	0.80	0.80	0.77	0.87
R	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.98	0.99
F_1	1.00	0.98	0.97	0.95	0.92	0.90	0.88	0.88	0.86	0.93
k-means										
P	1.00	1.00	1.00	0.99	0.99	0.98	0.96	0.96	0.95	0.98
R	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F_1	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.98	0.97	0.99

Table 7.9: Precision, recall, and F_1 for a synthetic image with varying text color and added Gaussian noise. The stroke width of the synthetic text averages 6 pixels.

σ	0.000	0.005	0.010	0.015	0.020	0.025	0.030	0.035	0.040	AVG
ICM, $n = 1$										
P	1.00	0.98	0.91	0.84	0.78	0.73	0.68	0.64	0.61	0.80
R	1.00	0.99	0.94	0.87	0.80	0.75	0.71	0.67	0.64	0.82
F_1	1.00	0.99	0.93	0.86	0.79	0.74	0.69	0.66	0.63	0.81
ICM, $n = 5$										
P	1.00	1.00	0.99	0.98	0.96	0.95	0.93	0.90	0.88	0.95
R	1.00	1.00	0.97	0.92	0.86	0.80	0.74	0.69	0.64	0.85
F_1	1.00	1.00	0.98	0.95	0.91	0.87	0.83	0.78	0.74	0.89
BP										
P	1.00	0.97	0.89	0.81	0.76	0.71	0.68	0.65	0.62	0.79
R	1.00	0.99	0.93	0.86	0.80	0.75	0.71	0.68	0.65	0.82
F_1	1.00	0.98	0.91	0.84	0.78	0.73	0.70	0.66	0.63	0.80
$\mathrm{BP^n}$, $n = 5$										
P	1.00	0.99	0.95	0.93	0.91	0.89	0.87	0.86	0.85	0.92
R	1.00	0.98	0.89	0.81	0.74	0.69	0.65	0.61	0.58	0.77
F_1	1.00	0.98	0.92	0.87	0.82	0.78	0.74	0.71	0.69	0.83
GC										
P	1.00	0.96	0.86	0.78	0.71	0.66	0.62	0.58	0.55	0.75
R	1.00	0.99	0.92	0.85	0.78	0.74	0.70	0.66	0.64	0.81
F_1	1.00	0.97	0.89	0.81	0.75	0.69	0.66	0.62	0.59	0.78
$\mathrm{GC^n}$, $n = 5$										
P	1.00	0.97	0.88	0.80	0.73	0.68	0.63	0.60	0.57	0.76
R	1.00	0.99	0.92	0.85	0.78	0.74	0.70	0.67	0.64	0.81
F_1	1.00	0.98	0.90	0.82	0.76	0.71	0.66	0.63	0.60	0.78
LMM										
P	0.90	0.49	0.27	0.25	0.25	0.24	0.23	0.23	0.23	0.34
R	0.66	0.98	0.96	0.95	0.93	0.92	0.91	0.90	0.89	0.90
F_1	0.76	0.65	0.43	0.40	0.39	0.38	0.37	0.36	0.36	0.46
AB										
P	0.99	0.98	0.93	0.86	0.78	0.71	0.66	0.60	0.56	0.79
R	0.97	0.55	0.49	0.45	0.45	0.42	0.42	0.42	0.41	0.51
F_1	0.98	0.70	0.64	0.59	0.57	0.53	0.51	0.50	0.47	0.61
k-means										
P	1.00	0.97	0.87	0.73	0.62	0.55	0.49	0.46	0.43	0.64
R	1.00	1.00	1.00	0.99	0.98	0.98	0.97	0.96	0.96	0.98
F_1	1.00	0.98	0.93	0.84	0.76	0.70	0.65	0.62	0.59	0.76

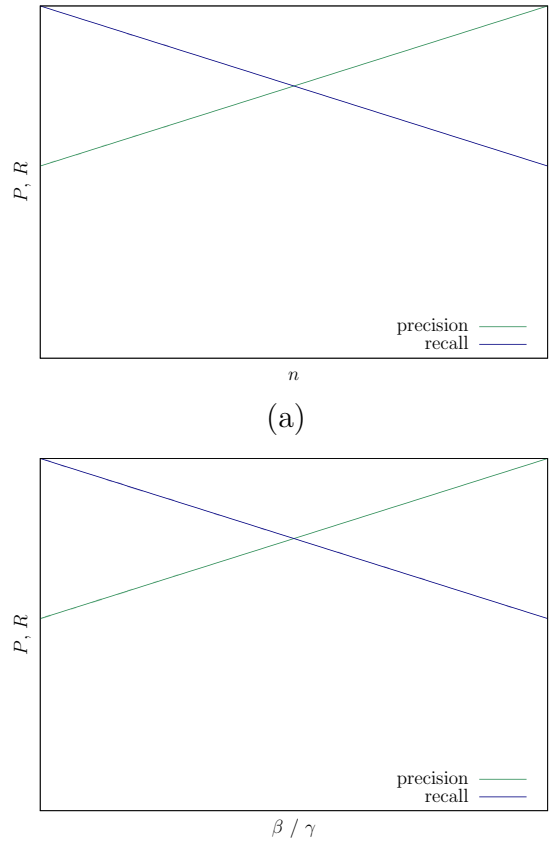


Figure 7.16: Schematic illustration of the influence of the order n of the MRF and the weighting parameters β, γ with respect to precision and recall. The experiments in this section have shown, that an increasing order n rises the precision while the recall rate drops. Equivalently, the precision rises while the recall rate drops with increasing impact of the weighting parameter. Best results are obtained with the proposed stroke model.

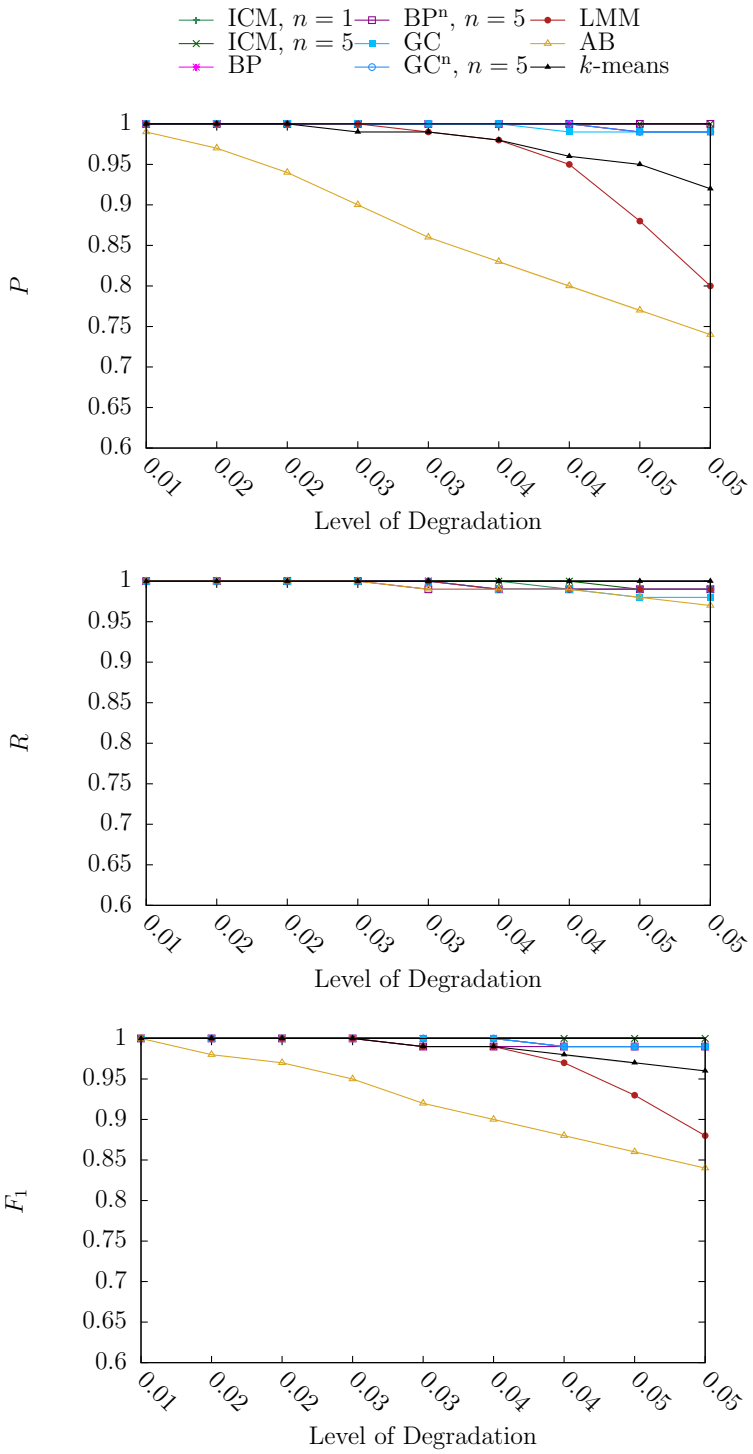


Figure 7.17: Progress of P, R, and F_1 for increasing Gaussian noise in a synthetic image with black text. The MRF based approaches (ICM, BP, GC) keep the precision and recall near 1. The local thresholding methods (BP and LMM) and k-means clustering show a decreasing precision with increasing noise.

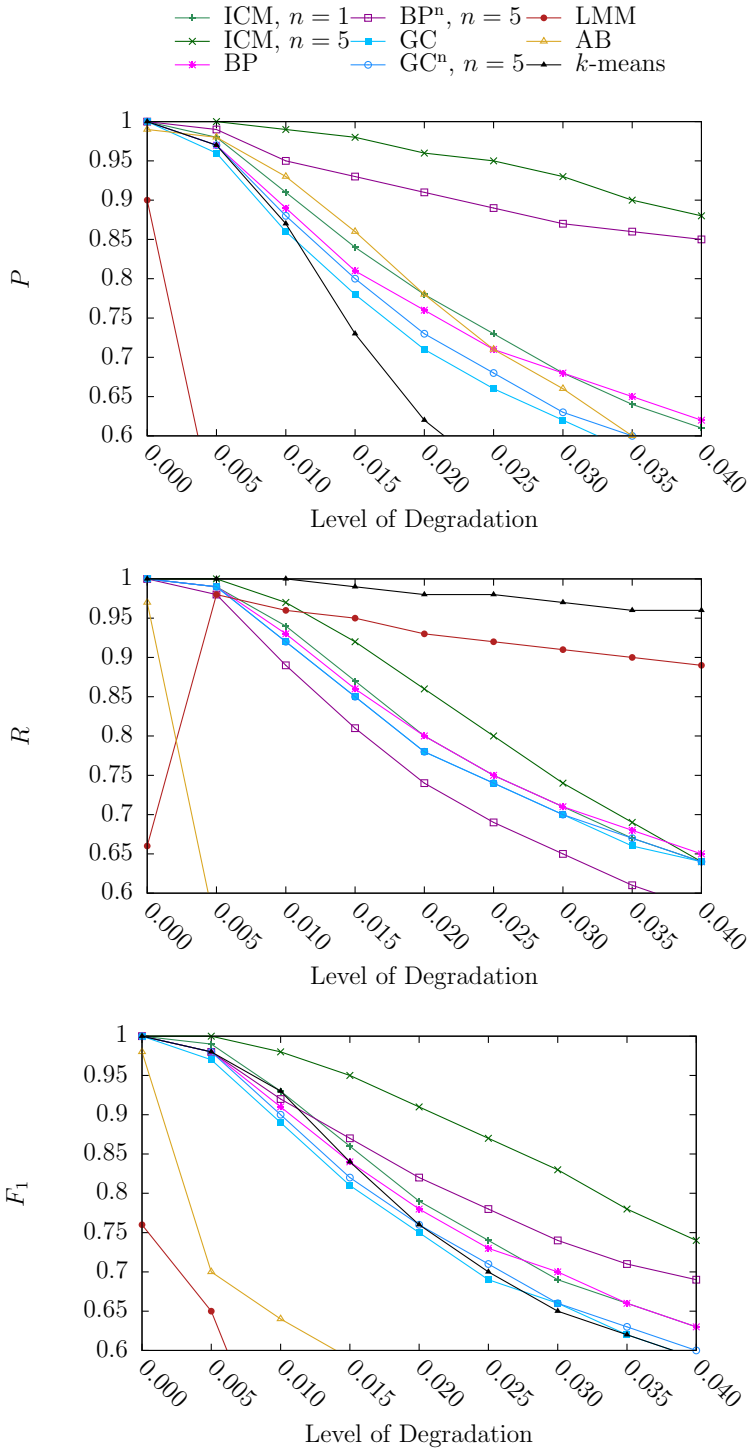


Figure 7.18: Progress of P, R, and F_1 for increasing Gaussian noise in a synthetic image with varying text color. The local based methods on the higher-order models (ICM and BP^{n} with $n = 5$) show the best performance. The low precision from LMM is a result from the low contrast of the characters in the synthetic image. The F_1 measure for LMM and AB shows only suboptimal values.

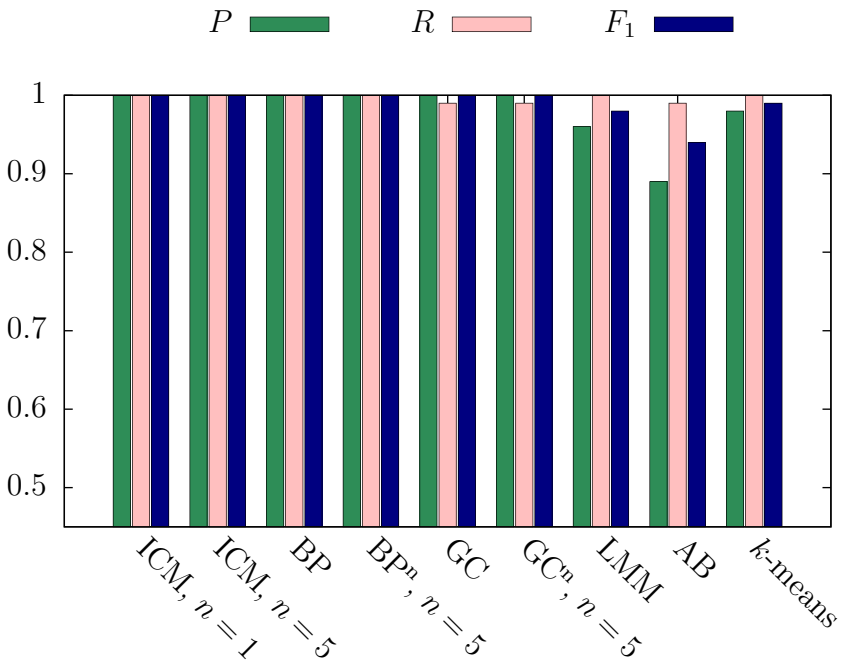


Figure 7.19: Average values for P, R, and F_1 for a synthetic image with black text and varying Gaussian noise ($\sigma = 0.01 \quad 0.05$).

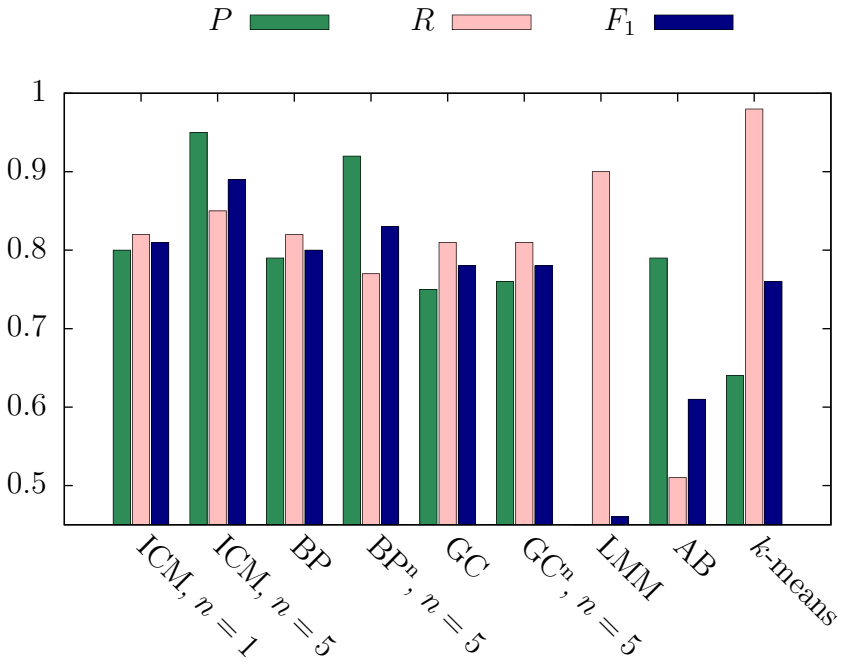


Figure 7.20: Average values for P, R, and F_1 for a synthetic image with varying text color and varying Gaussian noise ($\sigma = 0.00 \quad 0.04$).

Figure 7.21: Synthetic image with black text in (a) and varying text color in (b). An image with Gaussian noise ($\sigma = 0.03$) added is given in (c) for the image with black text, and in (d) for the image with varying text color. Results from the noisy image with $\sigma = 0.03$ are given in (e) and (f) for the proposed BP^5 algorithm, and for AB in (g) and (h). It can be seen that the proposed method has less noise in the background and foreground in both images, the image with black text and the image with varying text color.

7.7 General Comparison of the Methods

In this section we evaluate ICM, BP^n, and GC^n based minimization of the posterior energy of the MRF and compare the results to AB, LMM, and k-means clustering. Concerning the MRF approaches we use the stroke model for the spatial connections with parameters $\beta = 0.2, \gamma = 0.3$. The methods are applied on ten pages from the corpus of the *Missale Sinaiticum* and on ten pages from the DIBCO 2009 test set.

7.7.1 Missale Sinaiticum

Table 7.10 shows the results for ten folios of the *Missale Sinaiticum* when applied ICM, BP^n, GC^n, AB, LMM, and k-means for FBS. The average values for P, R, and F_1 can be seen in the last column for each method. We added the average stroke width of each image in the first row of the table. The average stroke width is calculated with Equation 6.6 after the binarization with AB and equals approximately 5 pixels for all images.

The MRF based approaches show better performance than AB and LMM. The proposed BP^n has the best performance with $F_1 = 0.79$. The recall rate averages 0.93 for k-means, but the value is useless since the precision is 0.49. For that case, the majority of characters is not segmented.

Two examples are illustrated in the following. The first example is folio 27 recto, the spectral band B-P 450 can be seen in Figure 7.24(a). The results for AB, LMM, and the proposed BP^n approach are shown in Figure 7.24(b)-(d). It can be seen that the result from AB contains more noise than the results from LMM and BP^n.

For the results of the MRF based FBS, it can be seen that even the right outermost characters in the first row are due to their low contrast separated from the background. The results from the MRF approaches are similar, ICM has the best precision, and GC^n the best recall rate. The F_1 measure constitutes approximately 0.81 for BP^n and GC^n. For comparison, AB has 0.79 and LMM 0.80.

Another example is given in Figure 7.25. Figure 7.25(a) shows B-P 450 from folio 41 recto and the results of the individual methods can be seen in Figure 7.25(b)-(d). Again, the output from AB contains noise and fails in regions with low contrast which results in $P = 0.87$ and $R = 0.80$. Best performance is obtained with the BP^n approach resulting in $P = 0.83$ and $R = 0.90$.

Finally, Figure 7.22 shows a diagram with the average results of the six methods for the *Missale Sinaiticum*. It can be seen that the MRF approaches have similar results, BP^5 has along them the best F_1 score.

7.7.2 DIBCO 2009 Images

In this section we show the performance of the individual methods on the DIBCO 2009 images. As already stated in Section 7.5 the differences between the proposed MRF stroke model and AB is due to the quality of the images minor. Two results are shown: image `H03.png` and `P01.png`. Figure 7.26 shows the input image and the results for `H03.png` and Figure 7.27 shows the input image and the results for `P02.png`. The input images

Table 7.10: Precision, recall, and F_1 for ten images from the corpus of the *Missale Sinaiticum*. The average stroke width for each image is denoted by $\varnothing$.

Image	17r	27v	27r	29r	30v	38v	40v	41r	44v	53v	AVG
$\varnothing$	5	5	5	5	5	5	5	5	5	5	
ICM, $n = 4$											
P	0.69	0.94	0.87	0.80	0.79	0.73	0.75	0.86	0.84	0.73	0.80
R	0.79	0.71	0.74	0.72	0.75	0.72	0.70	0.78	0.87	0.85	0.76
F_1	0.73	0.81	0.80	0.76	0.77	0.72	0.72	0.82	0.86	0.79	0.78
BP^{n}, $n = 4$											
P	0.69	0.87	0.81	0.83	0.82	0.73	0.74	0.83	0.82	0.79	0.79
R	0.80	0.77	0.82	0.69	0.75	0.70	0.74	0.90	0.93	0.81	0.79
F_1	0.74	0.82	0.81	0.75	0.78	0.72	0.74	0.86	0.87	0.80	0.79
GC^{n}, $n = 4$											
P	0.80	0.93	0.80	0.78	0.79	0.72	0.74	0.79	0.81	0.77	0.79
R	0.65	0.69	0.83	0.65	0.80	0.73	0.75	0.90	0.95	0.85	0.78
F_1	0.72	0.79	0.81	0.71	0.80	0.73	0.74	0.84	0.87	0.80	0.78
LMM											
P	0.77	0.85	0.71	0.88	0.81	0.76	0.79	0.88	0.84	0.81	0.81
R	0.74	0.75	0.92	0.55	0.70	0.64	0.63	0.75	0.89	0.79	0.74
F_1	0.75	0.80	0.80	0.68	0.76	0.69	0.70	0.81	0.87	0.80	0.77
AB											
P	0.68	0.85	0.82	0.63	0.74	0.70	0.73	0.87	0.82	0.69	0.75
R	0.73	0.73	0.77	0.60	0.79	0.69	0.71	0.80	0.80	0.83	0.75
F_1	0.70	0.78	0.79	0.61	0.76	0.70	0.72	0.83	0.81	0.75	0.75
k-means											
P	0.37	0.78	0.54	0.26	0.39	0.46	0.47	0.65	0.69	0.27	0.49
R	0.94	0.89	0.93	0.86	0.87	0.93	0.92	0.96	0.99	0.99	0.93
F_1	0.54	0.83	0.68	0.40	0.54	0.62	0.62	0.78	0.81	0.42	0.62

can be respectively seen in (a), AB is given in (b), (c) shows the result from the LMM, (d), (e), and (f) shows the output for ICM, BP^{n}, and GC^{n}.

The results for all ten images of the DIBCO 2009 test set are given in Table 7.11. The precision is especially for the first five images (`H01 - H05`) lower than the recall rate. This is a result from partial ink bleed through and the neighborhood system considered which induces over-segmented characters. As Figure 7.26 shows, the handwritten text appears more bulky than the GT data. Since the characters in the second part of the images (`P01 - P05`) have a wider stroke width, we have the same findings as above: less background noise and a good coherence of the individual characters. The average values for P, R, and F_1 can be seen in Figure 7.23. LMM achieves the best performance for this data set ($F_1 = 90$) followed by the MRF based approaches. They show approximately the same results.

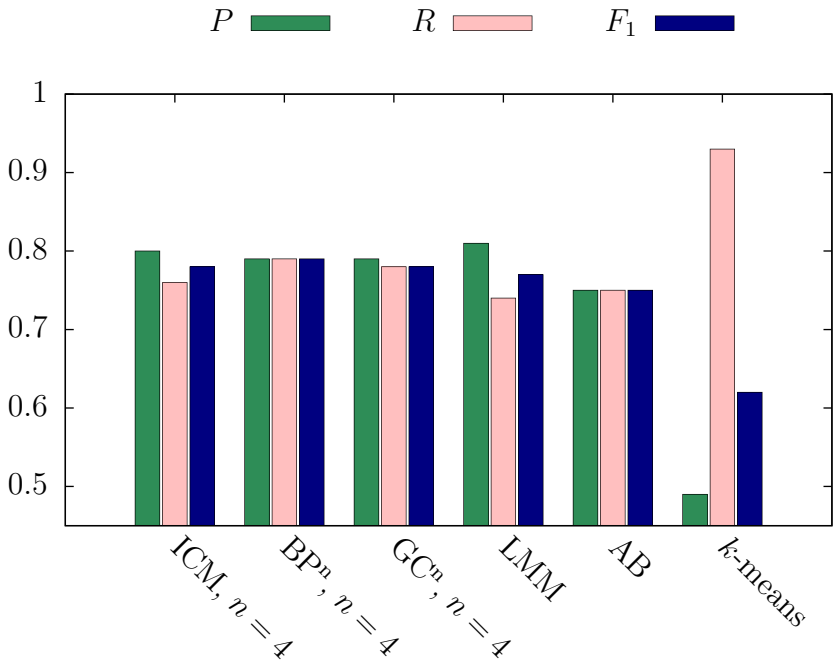


Figure 7.22: Average values for P, R, and F_1 for ten images from the *Missale Sinaiticum*. The diagram shows the average values for P, R, and F_1 from the results in Table 7.10. The best result for F_1 is obtained with the proposed BP^{n} algorithm, followed by GC^{n}, ICM with $n = 4$, LMM, and AB. The clustering method shows only suboptimal results.

Table 7.11: Precision, recall, and F_1 for ten images from the DIBCO 2009 test. The average stroke width for each image is denoted by $\varnothing$.

Image	H01	H02	H03	H04	H05	P01	P02	P03	P04	P05	AVG
$\varnothing$	4.11	4.49	4.47	5.09	5.13	3.56	7.26	3.44	4.69	3.50	
ICM, $n = 4$											
P	0.87	0.74	0.81	0.79	0.48	0.94	0.99	0.99	0.78	0.93	0.83
R	0.95	0.95	0.94	0.86	0.80	0.90	0.93	0.95	0.88	0.78	0.90
F_1	0.91	0.84	0.87	0.82	0.60	0.92	0.96	0.97	0.83	0.85	0.86
$\mathrm{BP^n}$, $n = 4$											
P	0.87	0.72	0.84	0.82	0.53	0.92	0.99	0.99	0.78	0.92	0.84
R	0.95	0.94	0.92	0.83	0.79	0.90	0.90	0.93	0.87	0.79	0.88
F_1	0.91	0.82	0.88	0.83	0.64	0.91	0.94	0.96	0.82	0.85	0.85
$\mathrm{GC^n}$, $n = 4$											
P	0.86	0.70	0.84	0.82	0.53	0.90	0.99	0.99	0.78	0.91	0.83
R	0.95	0.94	0.92	0.83	0.81	0.89	0.92	0.93	0.87	0.80	0.89
F_1	0.91	0.80	0.88	0.83	0.64	0.90	0.95	0.96	0.82	0.85	0.85
LMM											
P	0.95	0.96	0.91	0.97	0.89	0.95	0.98	0.97	0.95	0.99	0.95
R	0.93	0.85	0.93	0.83	0.87	0.91	0.94	0.75	0.88	0.72	0.86
F_1	0.94	0.90	0.92	0.89	0.88	0.93	0.96	0.85	0.91	0.83	0.90
AB											
P	0.97	0.67	0.89	0.91	0.88	0.93	0.98	0.97	0.92	0.91	0.90
R	0.85	0.90	0.87	0.87	0.84	0.85	0.90	0.36	0.88	0.80	0.81
F_1	0.91	0.77	0.88	0.89	0.86	0.89	0.94	0.52	0.90	0.85	0.84
k-means											
P	0.94	0.79	0.74	0.26	0.16	0.85	0.97	0.98	0.73	0.89	0.73
R	0.88	0.94	0.97	0.99	0.96	0.96	0.96	0.95	0.96	0.90	0.95
F_1	0.91	0.86	0.84	0.41	0.28	0.90	0.97	0.97	0.83	0.89	0.79

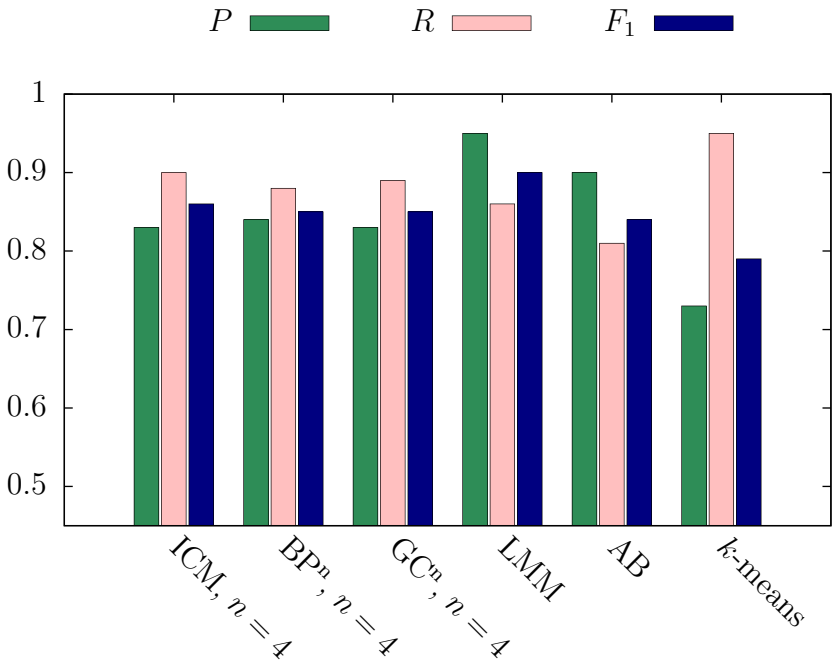
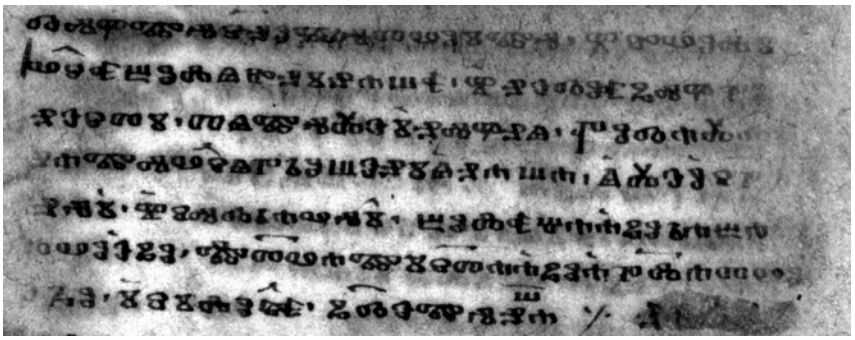
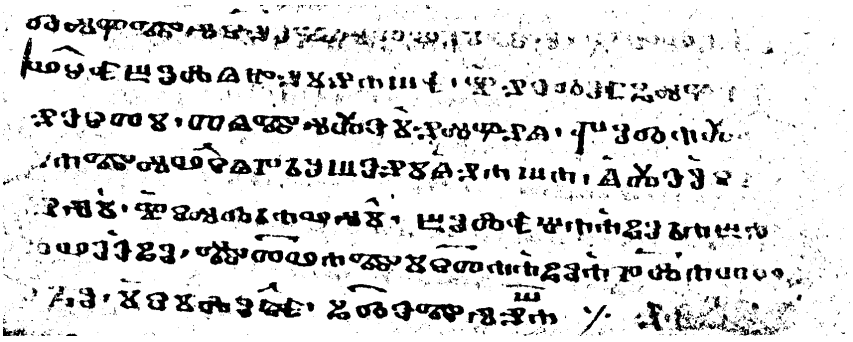


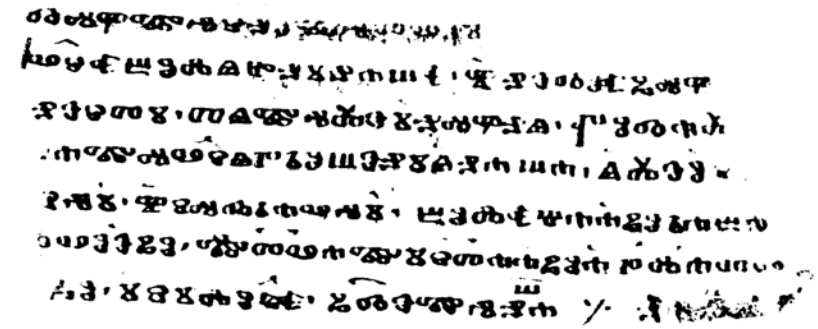
Figure 7.23: P, R, and F_1 for DIBCO 2009 images. The diagram shows the average values for P, R, and F_1 from the results in Table 7.11. As already in the contest itself, LMM shows the best performance on the DIBCO 2009 test set, followed by the MRF based approaches and AB. The clustering method shows again less performance.



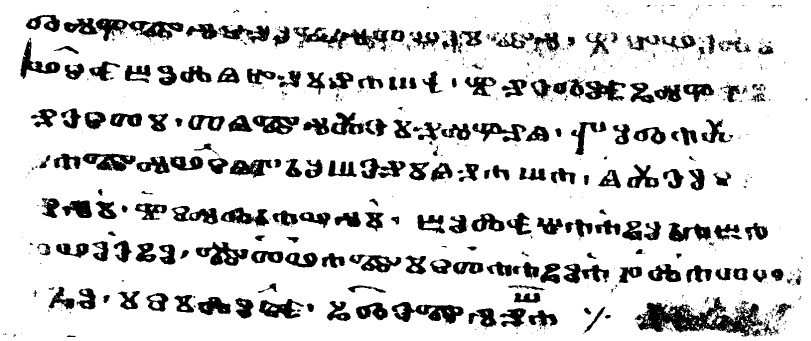
(a) Detail from folio 27 recto: B-P 450.



(b) AB; $P = 0.82$, $R = 0.77$.

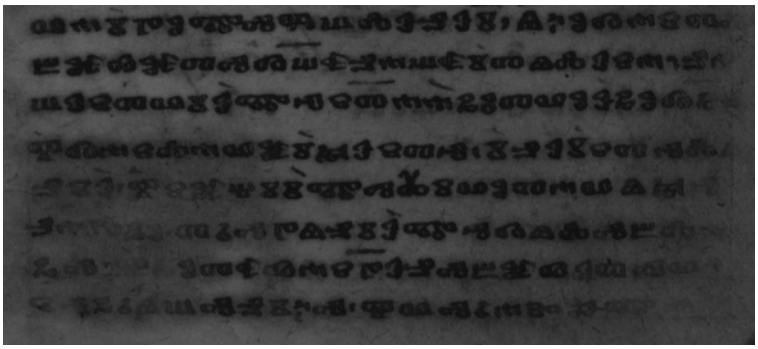


(c) LMM; $P = 0.71$, $R = 0.92$.

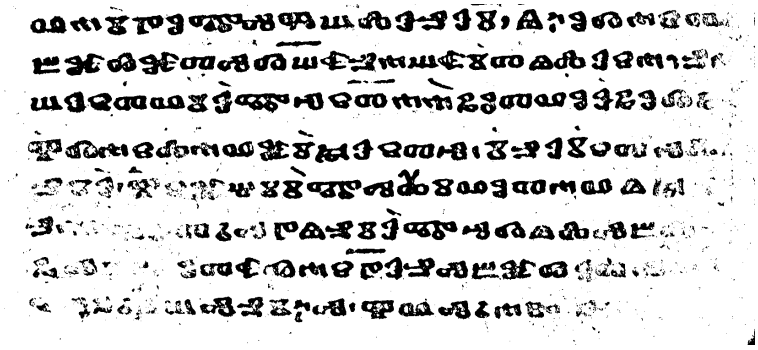


(d) BP^n; $n = 4$, $\gamma = 0.3$; $P = 0.81$, $R = 0.82$.

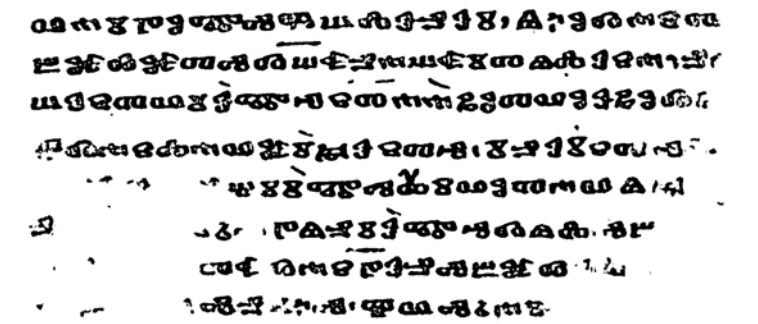
Figure 7.24: Results for folio 27 recto from the *Missale Sinaiticum*: original image (a), AB (b), LMM (c), and BP^n (d). The incorporation of spatial and spectral features in the MRF based approach results in less background noise and a more accurate separation of characters.



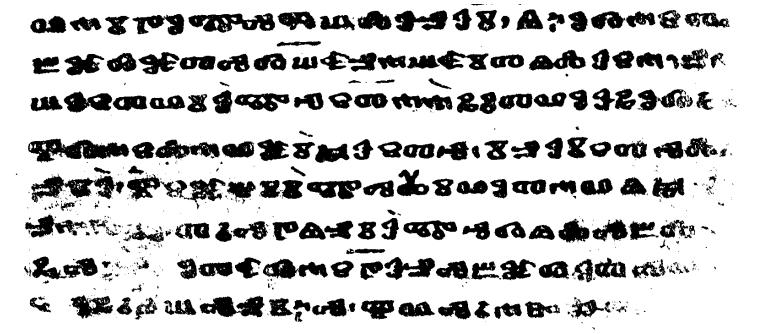
(a) Detail from folio 41 recto: B-P 450.



(b) AB; $P = 0.87$, $R = 0.80$.



(c) LMM; $P = 0.88$, $R = 0.75$.



(d) BP^{n}; $n = 4$, $\gamma = 0.3$; $P = 0.83$, $R = 0.90$.

Figure 7.25: Results for folio 41 recto from the *Missale Sinaiticum*: original image (a), AB (b), LMM (c), and BP^{n}, $n = 4$ (d). Similar to folio 27 recto, the proposed method results in less background noise and a more accurate separation of characters.

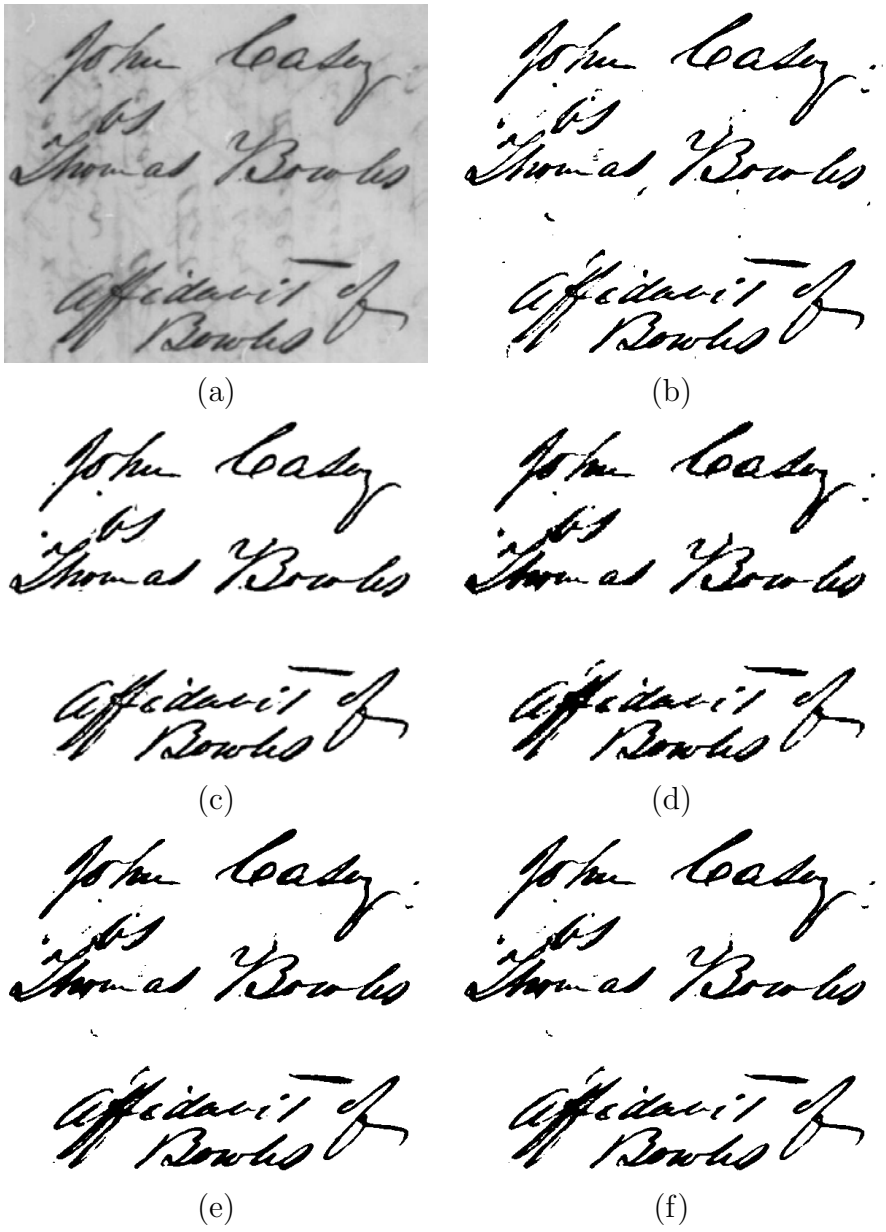
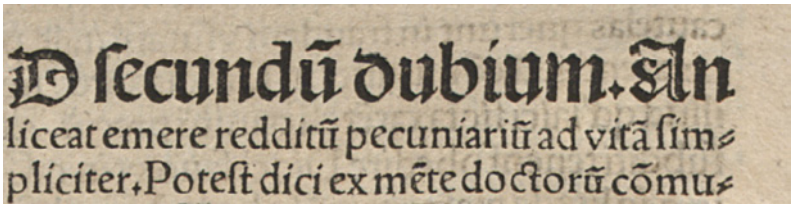
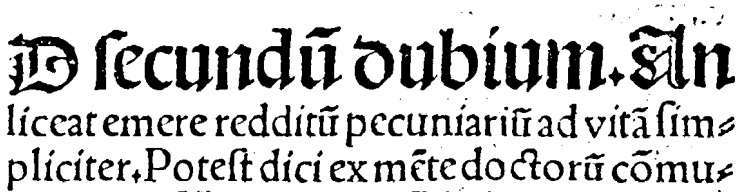


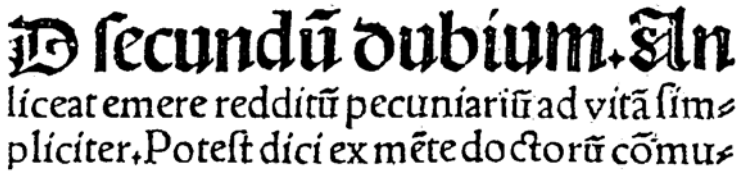
Figure 7.26: `H03.png`: (a) input image, (b) AB, (c) LMM, (d) ICM, (e) BP^n, $n = 4$, (f) GC^n, $n = 4$. All MRF based approaches include the proposed stroke model. For the ICM based inference, the stroke model is too crucial which leads to merged character gaps.



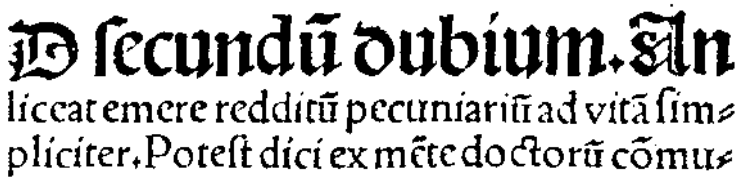
(a)



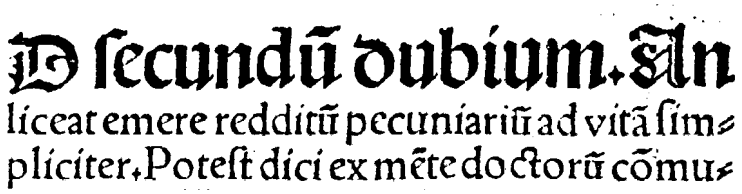
(b)



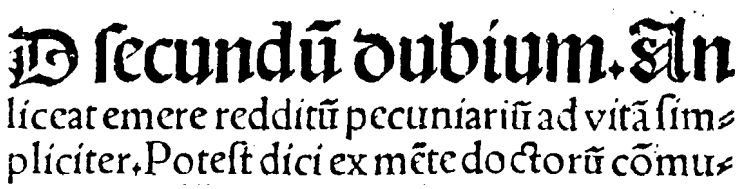
(c)



(d)



(e)



(f)

Figure 7.27: `P02.png`: (a) input image, (b) AB, (c) LMM, (d) ICM, (e) BP^n, $n = 4$, (f) GC^n, $n = 4$. These results are very similar, except for the result from AB, which shows noise in the upper right corner.

7.8 Summary of Results and Discussion

The first experiment in this chapter demonstrated the performance of incorporating a higher-order model in the MRF. We have shown that the proposed higher-order stroke model of the MRF shows superior results when compared to conventional pairwise connections. Table 7.12 presents the final ranking of the different inference methods for spatial and spectral based FBS. The table shows the best result for each method by means of the order n and the weighting parameters β and γ. Best performance is obtained with ICM with $n = 4$. Second place takes the proposed algorithm for BP in higher-order models (BP^n) followed by the GC based optimized posterior energy in higher-order models. The order n matches in each case the average stroke width of the image. For the case of the images from the *Missale Sinaiticum*, the stroke diameter averages five pixels which corresponds to cliques of fourth order to cover all pixels within this diameter, cf. Figure 4.2. However, computations with ICM in higher-order models are exponential in the number of neighbors, i.e. $\mathcal{O}(K^n)$, where K is the number of labels and n the number of variables in the local neighborhood. Thus, the runtime complexity increases exponentially with the size of the largest clique in the random field. In contrast, BP has significantly advances concerning the runtime [100, 25].

Table 7.12: Ranking of the results from Section 7.5, folio 29 recto from the corpus of the *Missale Sinaiticum* and image `P01.png` from the DIBCO 2009 data set.

Image	Method	n	β, γ	P	R	F_1
29 recto	ICM	4	0.2	0.80	0.72	0.76
	BP^n	4	0.3	0.83	0.69	0.75
	GC^n	4	0.2	0.78	0.65	0.71
`P01.png`	ICM	3	0.3	0.94	0.90	0.92
	BP^n	3	0.3	0.92	0.90	0.91
	GC^n	3	0.2	0.91	0.89	0.90

In the second experiment in Section 7.6, we demonstrated the robustness of the MRF based approaches in noisy images. Therefore, we generated two synthetic images with varying Gaussian noise. When the document image contains plain black text, the MRF based approaches average approximately 100% for precision and recall. The local optimization methods ICM and BP^n show better performance than GC. The higher-order model outperformed the standard formulation of MRFs. The k-means clustering method takes fourth place and, since they are very sensitive to noise, the local thresholding methods, AB and LMM, follow. The results from LMM depend heavily on the high contrast image pixels. As a result, it may introduce errors if the background contains certain amount of pixels that are dense and at the same time have a high contrast [98].

For the results of the synthetic image with varying text color, the local thresholding methods AB and LMM show only suboptimal results with an average F_1 score of 0.61 for AB and 0.46 for LMM. Especially in regions with low contrast between text and background, the influence of noise is too crucial and causes false positive and false negative classifications. Concerning the MRF based approaches, the higher-order models show

better results than first order MRFs. Both outperform the local thresholding methods and k-means clustering. Table 7.13 shows the final ranking of the methods when applied on the synthetic images.

Table 7.13: Ranking of the results from a synthetic image with added Gaussian noise (Section 7.6).

Text	Rank	Method	n	β, γ	P	R	F_1
black	1	ICM	5	0.3	1.00	1.00	1.00
	1	BP^n	5	0.3	1.00	1.00	1.00
	3	GC^n	5	0.2	1.00	0.99	1.00
	4	k-means			0.98	1.00	0.99
	5	LMM			0.97	1.00	0.98
	6	AB			0.87	0.99	0.93
colored	1	ICM	5	0.3	0.95	0.85	0.89
	2	BP^n	5	0.3	0.92	0.77	0.83
	3	GC^n	5	0.3	0.76	0.81	0.78
	4	k-means			0.64	0.98	0.76
	5	AB			0.79	0.51	0.61
	6	LMM			0.34	0.90	0.46

Finally, a general comparison of the methods was executed on ten images from the *Missale Sinaiticum* and on ten images from the DIBCO 2009 test set. Table 7.14 lists the ranking from the results in Section 7.7. For the DIBCO 2009 test set, LMM has already shown the best performance in the competition itself and shows again the best performance with $F_1 = 0.90$. The MRF based approaches reach 0.86 for ICM based optimization and 0.85 for the GC based optimization and the proposed BP^n algorithm.

On the test set of the *Missale Sinaiticum*, the proposed BP^n algorithm shows the best performance with $F_1 = 0.79$. ICM and GC^n average 0.78. The order n of the MRF constitutes 4 in each case. The F_1 measure for LMM results in 0.77, AB 0.75 and k-means clustering averages 0.62.

The experiments show that the proposed FBS method based on spatial and spectral features obtains better results than conventional thresholding or clustering methods. The proposed stroke model implemented within a higher-order MRF shows superior performance than traditional pairwise connections in MRFs. Especially for degraded documents with low contrast or image noise, the consideration of spatial cliques shows advantages, since individual outliers are compensated due to the configurations from the labels in the local neighborhood.

The higher-order MRF model has two parameters. The first one is the approximate stroke width and denotes the size of the higher-order cliques, i.e. the diameter of the stroke model. This parameter can be estimated automatically, e.g. through a preliminary binarization and the calculation of the mean stroke width as Equation 6.6 shows. The second parameter concerns the influence of the stroke model. The experiments show, that the smaller the stroke model or its weighting parameters β and γ, the more noise remains in the background and within characters. On the other side, when the size of

Table 7.14: Final ranking of the results after general experiments on the *Missale Sinaiticum* and the DIBCO 2009 test set (Section 7.7).

Images	Rank	Method	P	R	F_1
Missale Sinaiticum	1	BP^n	0.79	0.79	0.79
	2	ICM	0.80	0.76	0.78
	3	GC^n	0.79	0.78	0.78
	4	LMM	0.81	0.74	0.77
	5	AB	0.75	0.75	0.75
	6	k-means	0.49	0.93	0.62
DIBCO 2009	1	LMM	0.95	0.86	0.90
	2	ICM	0.83	0.90	0.86
	3	BP^n	0.84	0.88	0.85
	4	GC^n	0.83	0.89	0.85
	5	AB	0.90	0.81	0.84
	6	k-means	0.73	0.95	0.79

the higher-order cliques rises the average stroke width or when the influence by means of β and γ is too high, neighboring characters may merge or gaps in characters close. Furthermore, characters with low contrast or narrow stroke width may vanish, since the majority of pixels may be labeled as background in cliques rising the stroke width (even when the current observation is text). Then, the text pixel obtains a high penalty from the spatial correlation and is assigned as background too.

Concerning the boundary characteristics of the methods, Figure 7.28 shows examples for the k-means and the MRF based solution. The B-P 450 image on the left hand side shows three characters (a). The MRF based approach shows, through the consideration of contextual information in terms of stroke properties, a more smoothly segmented boundary, see Figure 7.28(b), than the result from the k-means algorithm in Figure 7.28(c). Here, the boundary is even rougher and shows artifacts within the characters as a result from the vanished boundary of the characters [64].

(a) (b) (c)

Figure 7.28: Differences within the boundary characteristics: (a) Original image (B-P 450), (b) output of the MRF based method (based on ICM with $n = 5$), and (c) k-means segmentation.

Chapter 8

Conclusion and Outlook

In this dissertation we have presented two approaches for the restoration of multispectral images of degraded documents. The first topic of this thesis covered the legibility enhancement of multispectral images. We have presented a method similar to Principal Component Analysis (PCA) which enables the enhancement of the highly correlated multispectral image data. The second topic we have discussed is a robust method for Foreground-Background Separation (FBS) in multispectral images.

Generally, MultiSpectral Imaging (MSI) supports the investigation of ancient manuscripts in which text is hardly visible in conventional color images or for the human eye. Since MSI has proven to be a capable technique for the digitization of decayed manuscripts, our motivation for the restoration of document images is to incorporate the full range of multispectral information for enhanced readability and FBS.

In Chapter 3 we presented background knowledge on MSI and showed the benefits for the investigation of ancient manuscripts. In the first part of the chapter we presented the image acquisition system used for the digitization of an Old Church Slavonic manuscript, the so called *Missale Sinaiticum (Sin. Slav. 5/N)*. In contrast to other studies on MSI which aim at enhancing the readability of the underwritten text in palimpsests, our focus is a general enhancement of text in multispectral images of degraded manuscripts. A drawback of MSI is the assemblage of highly correlated image data. The PCA is a method to reduce the spectral image data and to produce pseudo-colored images. In the second part of the third chapter, we presented an alternative approach to PCA. In order to improve the readability, we use the Multivariate Spatial Correlation (MSC) matrix, which includes spatial and spectral image data to remove spectral correlation. The benefit of an MSC based approach is that especially the text regions are considered for the enhancement. The experiments demonstrated the performance of combining spatial and spectral information for contrast enhancement. When compared to PCA, resulting images show that already in one of the first orthogonal components after an eigenvector analysis, the text appears clearly enhanced.

An evaluation of the philological transcription of the *Missale Sinaiticum* yields to an improvement of approximately 51% of the content. In the evaluation, we first counted the number of characters transcribed from the color image, and afterwards, the number of additional characters detected in the enhanced images.

The second part of the thesis covered the main topic, a robust method for FBS in

multispectral images of historical manuscripts. In contrast to the reviewed literature in Chapter 2 which use either spectral or spatial components to separate text from background, our approach simultaneously combines both, spatial and spectral features, in one framework.

Therefore, we incorporate both features in a higher-order Markov Random Field (MRF) which provides a probability theory for analyzing spatial and contextual dependencies. The spatial model of the MRF incorporates stroke characteristics and contextual dependencies of the MRF consider spectral observations of the multispectral image data. Since pairwise connections in graphical models cannot accomplish the extended spatial stroke model, we use a higher-order MRF model which is more expressive than the standard ones. The resulting posterior energy function, which forms the basis for the proposed FBS algorithm, is constructed over unary, pairwise, and higher-order potentials.

Inference in higher-order models was, due to the larger size of the cliques, neglected for a long time and only pairwise interactions have been used. Since higher-order models offer advantages compared to pairwise connections, [50] recently proposed expansion and swap moves for higher-order models. We reviewed two popular algorithms for statistical inference: Iterated Conditional Modes (ICM) and the above mentioned expansion and swap move as Graph Cut (GC) based algorithms. The focus was based on solving higher-order models, since the proposed stroke model requires extensive connections in the graphical model.

ICM are a well known method used until the late 1990s. However, inference based on ICM is based on a highly connected graph without higher-order cliques and the method has a weak computational performance. The method was applied for demonstrative purposes, but showed, due to its strong local minimum property, a good segmentation performance. The second method, we reviewed are energy minimization algorithms based on GC, namely, α-expansion moves and $\alpha\beta$-swap moves. Since the local energy minimization method ICM showed better results than the global GC methods, we introduced an adapted version of the well known Belief Propagation (BP) algorithm, which is also a local method for statistical inference but shows better computational performance.

The proposed algorithm for FBS was presented in Chapter 6. In this chapter, we have presented the individual potential functions forming the higher-order MRF energy and the proposed algorithm for BP in higher-order models. The unary or data potentials consider the spectral behavior of the observations, pairwise potentials cover neighboring observations, and the higher-order potentials include the proposed stroke model. Pairwise potentials represents the fact that the segmentation is locally homogeneous and labels depend on each other within a local neighborhood. The higher-order potentials consider observations within cliques of fixed shape. These cliques cover the approximate diameter of the strokes and are denominated as the stroke model.

In the experiments in Chapter 7 we have shown an extensive evaluation of the proposed approach for FBS. The aim of the first experiment was to analyze the behavior of the higher-order stroke model compared to pairwise formulations. Therefore, we evaluated the influence of different cliques sizes or MRF order n and the appropriate weighting parameter for the proposed model. It turned out that the incorporation of spatial probabilities improves the results. The resulting images have less noise in the background and even characters with very low contrast are separated from the background due to the

consideration of the full range of the multispectral information.

For the minimization of the posterior energy we compared three approaches: two local minimization methods and one global. It turned out, that local methods are better suited for the application given, since they have a strong local minimum property. Best results are obtained with ICM and the proposed BP^n algorithm. ICM obtained due to its label feedback slightly better results, but has a high computational complexity. The influence of the higher-order model has a low effect for the GC based optimization method.

In the second and third part of the experiments we compared the performance of the proposed method to state of the art binarization methods, among Adaptive Binarization (AB) [91], k-means clustering, and binarization of historical document images using the Local Maximum and Minimum (LMM). The k-means clustering was proposed by [66] for FBS in digitized ancient manuscripts and LMM proposed by [98] is an improved version of the best performing algorithm from a Document Image Binarization Contest (DIBCO 2009).

The approaches were executed on synthetic images with Gaussian noise and the experiments have shown that the simultaneously consideration of spatial and spectral information is robust against noise. Finally, in a general comparison of the methods on a set of multispectral images from the *Missale Sinaiticum* and on a set of images provided by the organizers from the DIBCO 2009, the proposed method showed high performance again. The results have shown that the extended spatial context refines the segmentation performance of individual characters and outperforms existing methods especially when multispectral images of degraded documents are applied.

8.1 Our Contribution

The main motivation for this thesis was the development of a robust method for FBS in multispectral images of degraded documents. Therefore, our approach is based on three contributions.

The basic idea of the proposed FBS method is the *simultaneous combination of spatial and spectral features*. In the reviewed literature, only a few benefit from this combination, but utilize the combination one after another. In our study we treat the combination simultaneously within the framework of an MRF.

In order to incorporate spatial features of strokes, we introduced a *stroke model* which models the spatial correlation of strokes. The model considers cliques of fixed shape and covers approximately the average stroke width of characters in a document image given. The approximate stroke width can be obtained automatically, in our case after a preceding conventional binarization step. This stroke model is incorporated within a higher-order MRF since pairwise connections are not sufficient. A main advantage of the proposed method is that a preceding training of the model and the requirement of high quality training data is avoided. This allows a general applicability and the method is independent of font, style, or size of characters.

However, higher-order models have been avoided for a long time due to their computational complexities. For statistical inference in the higher-order MRF, we proposed to use a local optimization method for FBS, since the influence of the stroke model has

more impact to the results than in global optimization methods. Therefore, we employed BP which handles arbitrary potential functions and provides a strong local minimum. In order to prepare the standard BP algorithm for higher-order models, we proposed the BP^n algorithm, to *incorporate the higher-order potentials* in the message updates.

The results have shown, that the combination of spatial and spectral features provides a robust binarization method and that our assumption of local inference for energy minimization is more appropriate for the incorporation of the stroke model than global inference methods.

The higher-order model has great impact to the result and is influenced by two parameters. The first one is the approximate stroke width and the second parameter concerns the influence of the stroke model. Our experiments have shown, that the smaller the neighborhood or its weighting parameter, the more noise remains in the background and within characters. On the other side, when the cliques of the stroke model exceed the average stroke width given in the document, or when the influence by means of the weighting factors is too high, neighboring characters may merge, gaps in characters close, or characters with low contrast may vanish completely. Best results are obtained when the stroke parameter coincides the real stroke width of a document image.

Our study has shown that the combination of spatial and spectral features improves the segmentation accuracy, especially in faded regions with low contrast or in the presence of image noise. In order to include spatial constraints of strokes, we use higher-order MRFs to enforce label consistency. This spatial information results in less noise in the foreground and the background. Furthermore, local inference methods like ICM or BP achieve better results for the application given than global methods.

8.2 Outlook

The proposed method for FBS has a few limitations. Designed for low contrast regions, ink bleed through from the background leads to errors since it is detected as foreground text. Here, pre-processing methods might produce more accurate results. Furthermore, the local values for the mean and covariance are calculated within rectangular boxes of constant size. However, the probability density of text and background changes over an image while the intensity of the background is changing. In this study we have used a fixed size of the observation windows. A model which is fitted to the background conditions might improve the results. However, not only the background changes, but also the stroke width of characters may change, for instance, in ligatures. Thus, an adaptation of the stroke model for alternating stroke conditions may improve the results by means of less broken characters, or touching characters.

We have used a simple potential function for the higher-order model based on the variance within a clique. Since BP works for arbitrary potential functions it is possible to use more sophisticated functions to simulate the spatial arrangement of stroke characteristics. It would be interesting to analyze the influence of different potential functions. Such functions could aim to avoid broken characters or noise in the background.

The proposed method is generally applicable, a comparison to supervised methods would be of interest. Furthermore, a more comprehensive comparison of related methods

is a major goal of the author, for instance, in the course of a Document Image Binarization Contest.

Appendix A

Acronyms and Symbols

AB	Adaptive Binarization
BP	Belief Propagation
B-P	Band-Pass
CRF	Conditional Random Field
DIA	Document Image Analysis
DIBCO 2009	Document Image Binarization Contest
EM	Expectation Maximization
FBS	Foreground-Background Separation
GMM	Gaussian Mixture Model
GBP	Generalized Belief Propagation
GC	Graph Cut
GT	Ground Truth
ICA	Independent Component Analysis
ICM	Iterated Conditional Modes
IR	InfraRed
L-P	Long-Pass
LBP	Loopy Belief Propagation
LMM	binarization of historical document images using the Local Maximum and Minimum
MAP	Maximum A Posterior

MSC	Multivariate Spatial Correlation
MSI	MultiSpectral Imaging
MRF	Markov Random Field
NIR	Near InfraRed
OCR	Optical Character Recognition
PCA	Principal Component Analysis
S-P	Short-Pass
TRW	Tree-Reweighted Message Passing
UV	UltraViolet
VIS	VISible light

Appendix B

List of Notation

α	weighting parameter for unary potentials
β	weighting parameter for pairwise potentials
γ	weighting parameter for higher-order potentials
$\mathcal{C}$	set of cliques
c	clique
$E(\mathbf{x}, \mathbf{y})$	energy function
i	index of a site
K	number of labels
$\mathcal{L}$	set of labels
l_b, l_t	label for background and text
$\mathcal{N}$	neighborhood system
$\mathcal{N}_i$	set of sites neighboring site i
N	number of variables in $\mathbf{X}$
n	order of the MRF, higher-order model, size of cliques
ψ	potential function
$U(\mathbf{x}, \mathbf{y})$	energy function for MRF posterior distribution
$\mathcal{V}$	set of sites
$\mathbf{X}$	random field
X_i	random variable
$\mathbf{x}$	MRF configuration of $\mathbf{X}$
x_i	configuration of X_i
$\mathbf{y}$	observed data
y_i	single observation

Bibliography

[1] G. Agam, G.Bal, G. Frieder, and O. Frieder. Degraded Document Image Enhancement. In *Proceedings of SPIE, Document Recognition and Retrieval XIV*, pages 65000C.1–65000C.11, San Jose, California, USA, January 2007.

[2] Asem M. Ali, Aly A. Farag, and Georgy L. Gimel'Farb. Optimizing Binary MRFs with Higher Order Cliques. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 98–111, Berlin, Heidelberg, 2008. Springer-Verlag.

[3] B. Allier, N. Bali, and H. Emptoz. Automatic accurate broken character restoration for patrimonial documents. *International Journal on Document Analysis and Recognition*, 8(4):246–261, 2006.

[4] T. Arny. *Explorations: An Introduction to Astronomy*. The McGraw-Hill Companies, 2003.

[5] C. Balas, V. Papadakis, N. Papadakis, A. Papadakis, E. Vazgiouraki, and G. Themelis. A novel hyper-spectral imaging apparatus for the non-destructive analysis of objects of artistic and historic value. *Journal of Cultural Heritage*, 4(1):330–337, January 2003.

[6] J. Banerjee, A.M. Namboodiri, and C.V. Jawahar. Contextual restoration of severely degraded document images. In *International Conference on Computer Vision and Pattern Recognition*, pages 517–524, Los Alamitos, CA, USA, 2009.

[7] I. Bar-Yosef. Input sensitive thresholding for ancient Hebrew manuscript. *Pattern Recognition Letters*, 26(8):1168–1173, 2005.

[8] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, 48(3):259–302, 1986.

[9] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[10] Y. Boykov and O. Veksler. *Graph Cuts in Vision and Graphics: Theories and Applications*, chapter 5, pages 79–96. Springer, 2005.

[11] Y. Boykov, O. Veksler, and R. Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[12] Yuri Boykov and Vladimir Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[13] J. Brauers, N. Schulte, and T. Aach. Multispectral Filter-Wheel Cameras: Geometric Distortion Model and Compensation Algorithms. *IEEE Trans. on Image Processing*, 17(2):2368–2380, December 2008.

[14] Huaigu Cao and Venu Govindaraju. Handwritten carbon form preprocessing based on markov random field. In *International Conference on Computer Vision and Pattern Recognition*, volume 0, pages 1–7, Minneapolis, Minnesota, USA, 2007.

[15] Huaigu Cao and Venu Govindaraju. Preprocessing of Low-Quality Handwritten Documents Using Markov Random Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(7):1184–1194, 2009.

[16] R.G. Casey and E. Lecolinet. A Survey of Methods and Strategies in Character Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):690–706, July 1996.

[17] B. B. Chaudhuri. *Digital Document Processing: Major Directions and Recent Advances*. Springer-Verlag New York, Inc., 2006.

[18] Mohamed Cheriet and Reza Farrahi Moghaddam. Low Quality Image Processing for DIAR. Issues and Directions. In *Proc. 16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, August 2008.

[19] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002.

[20] Y. T. Cui and Q. Huang. Character extraction of license plates from video. In *International Conference on Computer Vision and Pattern Recognition*, pages 502–507, Washington, DC, USA, 1997.

[21] Markus Diem, Martin Lettner, and Robert Sablatnig. Registration of multi-spectral manuscript images. In *Proceedings of the 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST07*, pages 133–140, November 2007.

[22] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience, New York, 2nd edition, 2001.

[23] R.L. Easton and K.T. Knox. Digital Restoration of Erased and Damaged Manuscripts. In *Proceedings of the 39th Annual Convention of the Association of Jewish Libraries*, Brooklyn, NY, June 2004.

[24] R.L. Easton, K.T. Knox, and W.A. Christens-Barry. Multispectral Imaging of the Archimedes Palimpsest. In *32nd Applied Image Pattern Recognition Workshop*, pages 111–118, Washington, DC, October 2003.

[25] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1):41–54, 2006.

[26] Christian Fischer and Ionna Kakoulli. Multispectral and hyperspectral imaging technologies in conservation : current research and potential applications. *Reviews in Conservation*, 2006(7):3–16, 2006.

[27] W.T. Freeman, E.C. Pasztor, and O.T. Carmichael. Learning low-level vision. *International Journal of Computer Vision*, 40(1):25–47, 2000.

[28] U. Garain, T. Paquet, and L. Heutte. On foreground - background separation in low quality document images. *International Journal on Document Analysis and Recognition*, 8(1):47–63, 2006.

[29] B. Gatos, K. Ntirogiannis, and I. Pratikakis. ICDAR 2009 Document Image Binarization Contest (DIBCO 2009). In *10th International Conference on Document Analysis and Recognition*, pages 1375 – 1382, Barcelona, Spain, July 2009.

[30] B. Gatos, I. Pratikakis, and I.J. Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39(3):317–327, 2006.

[31] B. Gatos, I. Pratikakis, and S. J. Perantonis. Efficient binarization of historical and degraded document images. In *Proceedings of the Eighth IAPR International Workshop on Document Analysis Systems*, pages 447–454, Nara, Japan, September 2008. IEEE Computer Society.

[32] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[33] Jost Gippert. The application of multispectral imaging in the study of caucasian palimpsests. *Bulletin of the Georgian Nationl Acadamy of Sciences*, 175(1):168–179, 2007.

[34] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51(2):271–279, 1989.

[35] C.W. Griffin. Digital imaging: Looking toward the future of manuscript research. *Currents in Biblical Research*, 5(1):58–72, 2006.

[36] 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.

[37] Mithun Das Gupta, Shyamsundar Rajaram, Nemanja Petrovic, and Thomas S. Huang. Models for patch based image restoration. In *International Conference on Computer Vision and Pattern Recognition Workshop*, June 2006.

[38] Mithun Das Gupta, Shyamsundar Rajaram, Nemanja Petrovic, and Thomas S. Huang. Models for patch-based image restoration. *EURASIP Journal on Image and Video Processing*, 2009:1–12, 2009.

[39] M. Hain, J. Bartl, and V. Jacko. Multispectral analysis of cultural heritage artefacts. *Measurement Science Review*, 3(3):9–12, 2003.

[40] H. Hase, M. Yoneda, S. Tokai, J. Kato, and Y. Suen. Color segmentation for text extraction. *International Journal on Document Analysis and Recognition*, 6(4):271–284, 2003.

[41] J. He, Q. D. M. Do, A. C. Downton, and J. H. Kim. A Comparison of Binarization Methods for Historical Archive Documents. In *8th International Conference on Document Analysis and Recognition*, pages 538–542, Washington, DC, USA, 2005. IEEE Computer Society.

[42] Xuming He, Richard S. Zemel, and Miguel A. Carreira-Perpinan. Multiscale conditional random fields for image labeling. *International Conference on Computer Vision and Pattern Recognition*, 2:695–702, 2004.

[43] Yong-Dian Jian and Chu-Song Chen. Second-order belief propagation and its application to object localization. In *International Conference on Systems, Man and Cybernetics*, pages 3955–3960, Taipei, ROC, October 2006. IEEE Computer Society.

[44] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, November 1999.

[45] R. Kasturi, L. O'Gorman, and V. Govindaraju. Document image analysis: A primer. *Sadhana*, 27(1):3–22, February 2002.

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

[47] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[48] Florian Kleber, Martin Lettner, Markus Diem, Maria Vill, Robert Sablatnig, Heinz Miklas, and Melanie Gau. Multispectral Acquisition and Analysis of Ancient Documents. In M. Ioannides, A. Addison, A. Georgopoulos, and L. Kalisperis, editors, *Proc. of the 14th International Conference on Virtual Systems and MultiMedia (VSMM 2008), Dedicated to Cultural Heritage - Project Papers*, pages 184–191, Limassol, Cyprus, 2008. Archaeolingua.

[49] P. Kohli, M. Kumar, and P. Torr. P^3 & beyond: Move making algorithms for solving higher order functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1645–1656, 2009.

[50] P. Kohli, M.P. Kumar, and P.H.S. Torr. P^3 & beyond: Solving energies with higher order cliques. In *International Conference on Computer Vision and Pattern Recognition*, pages 1–8, Minneapolis, Minnesota, USA, June 2007.

[51] P. Kohli, L. Ladický, and P.H. Torr. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 82(3):302–324, 2009.

[52] Vladimir Kolmogorov and Carsten Rother. Comparison of energy minimization algorithms for highly connected graphs. In *9th European Conference on Computer Vision*, pages 1–15, 2006.

[53] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? In *Proceedings of the 7th European Conference on Computer Vision*, volume 3, pages 65–81, London, UK, 2002. Springer-Verlag.

[54] N. Komodakis and N. Paragios. Beyond pairwise energies: Efficient optimization for higher-order mrfs. *International Conference on Computer Vision and Pattern Recognition*, 0:2985–2992, 2009.

[55] Jung Gap Kuk, Nam Ik Cho, and Kyoung Mu Lee. Map-mrf approach for binarization of degraded document image. In *International Conference on Image Processing*, pages 2612–2615, San Diego, California, U.S.A., October 2008.

[56] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the 18th International Conference on Machine Learning*, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[57] Xiangyang Lan, Stefan Roth, Daniel P. Huttenlocher, and Michael J. Black. Efficient belief propagation with learned higher-order markov random fields. In *9th European Conference on Computer Vision*, pages 269–282, Graz, Austria, 2006.

[58] Graham Leedham, Chen Yan, Kalyan Takru, Joie Hadi Nata Tan, and Li Mian. Comparison of some thresholding algorithms for text/background segmentation in difficult document images. In *7th International Conference on Document Analysis and Recognition*, page 859, Washington, DC, USA, 2003.

[59] Thibault Lelore and Frédéric Bouchara. Document image binarisation using markov field model. In *International Conference on Document Analysis and Recognition*, pages 551–555, Barcelona, Spain, 2009.

[60] M. Lettner, M. Diem, R. Sablatnig, and H. Miklas. Registration and Enhancing of Multispectral Manuscript Images. In *Proc. 16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, August 2008.

[61] M. Lettner, F. Kleber, R. Sablatnig, and H. Miklas. Contrast Enhancement in Multispectral Images by Emphasizing Text Region. In *8th IAPR International Workshop on Document Analysis Systems*, pages 225–232, Nara, Japan, September 2008.

[62] M. Lettner and R. Sablatnig. Estimating the original drawing trace of painted strokes. In *IS&T SPIE Electronic Imaging*, volume 6810, San Jose, California, USA, January 2008.

[63] M. Lettner and R. Sablatnig. Higher Order MRF for Foreground-Background Separationin Multispectral Images of Historical Manuscripts. 9th IAPR International Workshop on Document Analysis Systems, June 2010. to be published.

[64] Martin Lettner and Robert Sablatnig. Document image binarization in multispectral images using markov random fields. In *33rd Workshop of the Austrian Association for Pattern Recognition*, pages 85–96, Stainz, Austria, May 2009. OCG.

[65] 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)*, pages 813–817, Barcelona, Spain, July 2009.

[66] Y. Leydier, F. Le Bourgeois, and H. Emptoz. Serialized Unsupervised Classifier for Adaptative Color Image Segmentation: Application to Digitized Ancient Manuscripts. In *17th International Conference on Pattern Recognition*, pages 494–497, Cambridge, UK, 2004.

[67] S.Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer, 3 edition, 2009.

[68] Laurence Likforman-Sulem, Abderrazak Zahour, and Bruno Taconet. Text line segmentation of historical documents: a survey. *International Journal on Document Analysis and Recognition*, 9(2):123–138, 2007.

[69] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[70] Shijian Lu and Chew Lim Tan. Thresholding of Badly Illuminated Document Images Through Photometric Correction. In *ACM Symposium on Document Engineering*, pages 3–8, 2007.

[71] L. Lucchese and S.K. Mitra. Color image segmentation: a state-of-the-art survey. In *Proceedings of the Indian National Science Academy*, pages 207–221, 2001.

[72] F. Mairinger. *Strahlenuntersuchung an Kunstwerken*. Bücherei des Restaurators, Band 7. E.A. Seemann, 2003.

[73] H. 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.

[74] H. Miklas, M. Gau, F. Kleber, M. Lettner, M. Vill, R. Sablatnig, M. Schreiner, M. Melcher, and G. Hammerschmid. St. Catherine' s Monastery on Mount Sinai and the Balkan-Slavic Manuscript-Tradition. In H. Miklas and A. Miltenova, editors, *Slovo: Towards a Digital Library of South Slavic Manuscripts*, pages 13–36, 2008.

[75] R.F. Moghaddam and Mohamed Cheriet. RSLDI: Restoration of single-sided low-quality document images. *Pattern Recognition Letters*, 42(12):3355–3364, 2009.

[76] Shunji Mori, Hirobumi Nishida, and Hiromitsu Yamada. *Optical character recognition*. John Wiley & Sons, Inc., New York, NY, USA, 1999.

[77] Konstantinos Ntirogiannis, B. Gatos, and I. Pratikakis. An objective evaluation methodology for document image binarization techniques. In *Proceedings of the Eighth IAPR International Workshop on Document Analysis Systems*, pages 217–224, Nara, Japan, September 2008. IEEE Computer Society.

[78] Ifeoma Nwogu and Jason J. Corso. $(BP)^2$: Beyond pairwise Belief Propagation labeling by approximating Kikuchi free energies. In *International Conference on Computer Vision and Pattern Recognition*, pages 1–8, Alaska, USA, June 2008.

[79] Tayo Obafemi-Ajayi, Gady Agam, and Ophir Frieder. Efficient MRF approach to document image enhancement. In *19th International Conference on Pattern Recognition*, pages 1–4, Tampa, Florida, USA, December 2008.

[80] N. Otsu. A threshold selection method from gray-level. *IEEE Transactions on Systems, Man, and Cybernetics*, 9:62–66, 1979.

[81] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[82] V. Pervouchine, G. Leedham, and K. Melikhov. Handwritten character skeletonisation for forensic document analysis. In *Proceedings of the ACM Symposium on Applied Computing*, pages 754–758, Santa Fe, New Mexico, USA, 2005.

[83] Brian Potetz. Efficient belief propagation for vision using linear constraint nodes. *International Conference on Computer Vision and Pattern Recognition*, 0:1–8, 2007.

[84] Brian Potetz and Tai Sing Lee. Efficient belief propagation for higher-order cliques using linear constraint nodes. *Computer Vision and Image Understanding*, 112(1):39–54, 2008.

[85] K. Rapantzikos and C. Balas. Hyperspectral imaging: potential in non-destructive analysis of palimpsests. In *IEEE International Conference on Image Processing*, volume 2, pages 618–21, 2005.

[86] J.A. Richards and X. Jia. *Remote Sensing Digital Image Analysis : An Introduction*. Springer, 1999.

[87] S. Roth and M.J. Black. Fields of experts: A framework for learning image priors. In *Conference on Computer Vision and Pattern Recognition*, pages 860–867, San Diego, CA, USA, June 2005.

[88] Stefan Roth and Michael J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205–229, 2009.

[89] Carsten Rother, Sanjiv Kumar, Vladimir Kolmogorov, and Andrew Blake. Digital tapestry. In *International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 589–596, Washington, DC, USA, 2005. IEEE Computer Society.

[90] E. Salerno, A. Tonazzini, and L. Bedini. Digital image analysis to enhance underwritten text in the Archimedes palimpsest. *International Journal on Document Analysis and Recognition*, 9(2-4):79–87, 2007.

[91] J. Sauvola and M. Pietikainen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.

[92] Raymond A. Serway and John W. Jewett. *Physics for Scientists and Engineers*. Brooks/Cole, 6 edition, 2004.

[93] Gary A. Shaw and Hsiao-hua K. Burke. Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14(1):3–28, 2003.

[94] Z. Shi and V. Govindaraju. Historical document image enhancement using background light intensity normalization. In *17th International Conference on Pattern Recognition*, pages 473–476, Washington, DC, USA, 2004.

[95] Patrick Shiel, Malte Rehbein, and John Keating. The ghost in the manuscript: Hyperspectral text recovery and segmentation. In Malte Rehbein and Patrick Sahle und Torsten Schaßan, editors, *Codicology and Palaeography in the Digital Age*, chapter 2, pages 159–174. Norderstedt: Books on Demand, 2009.

[96] Lu Shijian and Chew Lim Tan. Script and language identification in noisy and degraded document images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(1):14–24, 2008.

[97] Keiichiro Shirai, Masanori Wakabayashi, Masayuki Okamoto, and Hiroaki Yamamoto. A study for high performance character extraction from color scene images. *IAPR International Workshop on Document Analysis Systems*, 0:293–298, 2008.

[98] B. Su, S. Lu, and C.T. Tan. Binarization of Historical Document Images Using the Local Maximum and Minimum. 9th IAPR International Workshop on Document Analysis Systems, June 2010. to be published.

[99] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003.

[100] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M.F. Tappen, and C. Rother. A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

[101] Marshall F. Tappen and William T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *International Conference on Computer Vision and Pattern Recognition*, page 900, Washington, DC, USA, 2003.

[102] C. Mancas Thillou and B. Gosselin. Spatial and color spaces combination for natural scene text extraction. In *International Conference on Image Processing*, pages 985–988, October 2006.

[103] P. Thouin, Y. Du, and C. Chang. Low resolution expansion of gray scale text images using gibbs-markov random field model. In *Symposium on Document Image Understanding Technology*, pages 41–47, April 2001.

[104] A. Tonazzini, L. Bedini, and E. Salerno. Independent component analysis for document restoration. *International Journal on Document Analysis and Recognition*, 7(1):17–27, March 2004.

[105] Oivind Due Trier and Torfinn Taxt. Evaluation of binarization methods for document images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):312–315, 1995.

[106] Martin Wainwright, Tommi Jaakkola, and Alan Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. *IEEE Transactions on Information Theory*, 51:3697–3717, 2002.

[107] K. Wang, J. A. Kangas, and W. Li. Character segmentation of color images from digital camera. In *6th International Conference on Document Analysis and Recognition*, pages 210–214, Washington, DC, USA, September 2001.

[108] T.A. Warner. Analysis of Spatial Patterns in Remotely Sensed Data Using Multivariate Spatial Correlation. *Geocarta International*, 14(1):59–65, 1999.

[109] D. Wartenberg. Multivariate Spatial Correlation: A Method for Exploratory Geographical Analysis. *Geographical Analysis*, 17(4):263–283, 1985.

[110] Yair Weiss. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 12(1):1–41, 2000.

[111] Yair Weiss and William T. Freeman. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *IEEE Transactions on Information Theory*, 47(2):736–744, 2001.

[112] Christian Wolf and David S. Doermann. Binarization of low quality text using a markov random field model. In *International Conference on Pattern Recognition*, volume 3, pages 160–163, Quebec, Canada, August 2002.

[113] Victor Wu, Raghavan Manmatha, and Edward M. Riseman, Sr. Textfinder: An automatic system to detect and recognize text in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1224–1229, 1999.

[114] X. Wu and S. Shah. A Conditional Random Field Model for Cell Segmentation Using Multispectral Data. In *Proceedings of MICCAI Workshop on Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy*, London, September 2009.

[115] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Generalized belief propagation. *Neural Information Processing Systems*, 13:689–695, 2000.

[116] Dong-Qing Zhang and Shih-Fu Chang. Learning to detect scene text using a higher-order mrf with belief propagation. In *International Conference on Computer Vision and Pattern Recognition*, pages 101–108, Washington, DC, USA, 2004.

[117] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.