Spatial and Spectral Based Segmentation of Text in Multispectral Images of Ancient Documents

Martin Lettner and Robert Sablatnig
Pattern Recognition and Image Processing Group
Institute of Computer Aided Automation, Vienna University of Technology, Austria
{lettner, sab}@prip.tuwien.ac.at

Abstract

In this paper we propose a character segmentation method for multispectral images of ancient documents. Due to the low quality of the images the main idea of this study is to combine the multispectral behavior and contextual spatial information. Therefore we utilize a Markov Random Field model using the spectral information of the images and stroke properties to include spatial dependencies of the characters. Since the stroke properties and the Gaussian parameters for the imaging model are evaluated automatically the proposed segmentation method requires no training phase. We compared the method to state of the art character segmentation methods and demonstrate the effectiveness of combining spectral and spatial features for the segmentation of characters in multispectral images.

1. Introduction

The application of multispectral imaging for the digitization of ancient documents offers a technique for analysis and preservation. It offers the possibility of handling obstacles like palimpsests[1] or damages caused by age [5]. Recent studies in multispectral imaging for historical manuscripts focused particularly on the revisualization of the underwritten (erased) text in palimpsests where methods like the Independent Component Analysis [16] or the Principal Component Analysis [5, 15] are applied. Others are interested in a general enhancement of the readability independent of underwritten or overwritten text [11]. In order to prepare digital documents for further computer aided analysis and to enable the use of simplified analysis techniques, e.g. feature extraction or Optical Character Recognition, usually binary images are produced [9].

This paper is concerned with the segmentation of characters in multispectral images (MSI) from ancient manuscripts. The challenges we have to deal with are caused by the documents age and include non-uniform appearance of the writing and the background, blur of the background, fading out of the ink and poor contrast, mold, water stains or humidity.

While traditional character segmentation methods for color or multispectral images usually consider the color or spectral component [12, 8] the main idea of our approach is to improve the segmentation accuracy by including contextual information in terms of pixel relations. Therefore we utilize the multispectral behavior of the MSI and combine spatial and spectral features. Recent studies exploit the combination of spectral and spatial components already but treat spectral and spatial components successive. Mancas-Thillou et al. use color clustering and subsequently a Gabor-based filter to combine color with spatial information but the method is not able to handle poor contrast images [20]. Wang et al. combines edge information, watershed transformation and a subsequent clustering for character segmentation in color images [21]. Garain et al.[6] proposes a chain of processing steps for foreground - background separation in low quality images in which a Connected Component Labeling for color images constitutes the main separation step.

In contrast to previous studies we arrange the combination of spatial and spectral components in a straightforward way. Therefore, we utilize a Markov Random Field (MRF) model which provides a probability theory for analyzing spatial or contextual dependencies [13]. Computer vision problems like color image segmentation [10], stereo matching or image denoising [7] can be elegantly expressed as MRFs [19] which have been also adopted for document image analysis of panchromatic images. For instance, Wolf and Doerman use MRF models for the binarization of low quality text [22]. Their model for the spatial relationship is defined on a neighborhood of 4×4 pixel cliques. Cao and Govindarayu use MRFs for the binarization of degraded handwritten forms [2] where the spatial relations are obtained from a training set of high quality binarized images

[1]A palimpsest is a parchment which was rewritten after the first text has been erased.

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.51

and consist of 114 representatives of shared patches.

Since there is no high quality text available for generating patches we propose to incorporate stroke characteristics to cover the spatial or contextual dependencies of characters. We tested the proposed method on MSI of an ancient Slavonic missal written on parchment. The texts in Old Church Slavonic are written in Glagolitic[2] letters. The results are compared to conventional color segmentation methods and to binarization methods applied on selected single band images.

This paper is organized as follows. Section 2 introduces the multispectral images and the camera system used for the acquisition process. Section 3 explains the MRF based segmentation methodology and in Section 4 experiments and results are presented. Section 5 concludes the paper with a conclusion and an outlook.

2. MSI Images of Ancient Documents

Multispectral imaging applied in the spectral range from ultra violet (UV), visible light range (VIS) up to the near infrared (NIR) range combines conventional imaging and spectrometry to acquire both spatial and spectral information from an object. For the acquisition of the manuscripts we use a Hamamatsu C9300-124 (spectral response: $330-1000nm$) gray scale camera. To obtain multispectral data a set of 7 optical filters is used to select specific ranges from the spectrum. The set consist of three bandpass filters with peak transmittance at 450, 550, 650 and 780 nm (BP 450, BP 550, BP 650, BP 780), one low pass filter with cutoff frequency of 400 nm (LP 400) and two high pass filters with a cutoff frequency of 400 and 800 nm (HP 400, HP 800). The filters are mounted within a filter wheel in front of the gray scale camera. We obtain two images with UV illumination (SP 400, LP 400) and 7 with VIS-NIR illumination (LP 400, BP 450, BP 550, BP 650, BP 780, LP 800, no filter). Due to the use of filters in different wavelengths a registration process is necessary in order to combine the spectral images [3].

3. Markov Random Field Model for Character Segmentation

Our MSI is defined over a finite rectangular lattice $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ in which $1 \ldots n$ indicate the pixel index and the observations $\vec{f_s} \in \mathcal{F}$ for each pixel s represent the multispectral feature vectors used for partitioning the image. The goal is to find a labeling $\omega \in \Omega$ which separates the characters or text pixels from the background. Given a set of feature vectors $\mathcal{F}$ and the set Ω of all possible segmentations our purpose is to find the segmentation $\hat{\omega}$ with the highest probability. A widely accepted standard is to construct this probability measure within a Bayesian framework [10]:

$$\Pr(\omega|\mathcal{F}) \propto \Pr(\mathcal{F}|\omega) \Pr(\omega), \tag{1}$$

where $\Pr(\omega)$ depicts the prior probability and $\Pr(\mathcal{F}|\omega)$ is the likelihood. $\hat{\omega}$ which maximizes the posterior probability $\Pr(\omega|\mathcal{F})$ can be found via the the *Maximum A Posteriori* (MAP) estimate [7]:

$$\hat{\omega} = \arg\max \Pr(\mathcal{F}|\omega) P(\omega). \tag{2}$$

3.1 Prior Model $\Pr(\omega)$

The prior $\Pr(\omega)$ represents the fact that the segmentation is locally homogeneous [10]. According to the *Hammersly-Clifford theorem* [7], a random field is a MRF if $\Pr(\omega)$ follows a Gibbs distribution

$$\Pr(\omega) = \frac{1}{Z} \exp(-U(\omega)) = \frac{1}{Z} \exp\left(-\sum_{c \in \mathcal{C}} V_C(\omega_c)\right), \tag{3}$$

where $Z = \sum_{\omega \in \Omega} \exp\left(-U(\omega)\right)$ is a normalizing constant and $U(\omega) = \sum_{c \in \mathcal{C}} V_C(\omega_c)$ is an energy function. c is a set of pixels within a neighbor set $\mathcal{N}$, called *clique*, and V_c denotes the potential function or clique potential of clique $c \in \mathcal{C}$ having the label configuration ω_c. For a regular lattice $\mathcal{S}$, the neighbor set $\mathcal{N}$ of i is defined as the set of nearby sites within a radius r [13]

$$\mathcal{N}_i = \{i' \in \mathcal{S} | \left[dist(s_{i'}, s_j)\right]^2 \leq r, i \neq j\} \tag{4}$$

where $dist(s_{i'}, s_j)$ denotes the Euclidean distance between two pixels $s_{i'}$ and s_j. A first order MRF involves only the 4 directly connected pixels, as shown in Figure 1a. The numbers $n = 1 \ldots 5$ indicate the neighboring sites in a n-th order neighborhood system.

We propose to use stroke properties to cover the spatial dependencies of the characters. Thus $\mathcal{C}$ is the set of all pixels s within a radius r, where r corresponds to the mean stroke width. The mean stroke width follows the set of all foreground pixels S and the set of all border pixels D from characters and is computed as $w = 2\frac{N_S}{N_D}$ where N_S and N_D are the number of pixels in S and D [14]. Our experiments for manually segmented characters devoted a mean stroke width of 5 pixels. Thus the prior considers a neighborhood set of at least 4th order. Figure 1b shows one character from our data set. The white circle corresponds to a neighborhood system of 4th order.

The clique potentials favor similar classes within a neighborhood $\mathcal{N}$

$$V_C = \delta(\omega_s, \omega_r) = \begin{cases} +1 & \text{if } \omega_s \neq \omega_r \\ -1 & \text{otherwise} \end{cases} \tag{5}$$

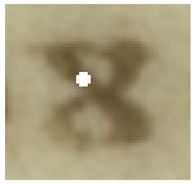
[2]This is the oldest known Slavic script.

Thus, the full prior is given as follows:

$$\Pr(\omega) = \frac{1}{Z} \exp\left(- \sum_{\{s,r\} \in \mathcal{C}} \delta(\omega_s, \omega_r) \right) \tag{6}$$

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

(a)



(b)

Figure 1. (a) Neighborhood system [13] and (b) Glagolitic character with marked neighborhood system of 4th order.

3.2 Observation Model $\Pr(\mathcal{F}|\omega)$

The observation model or image process $\mathcal{F}$ can be formalized as follows: $\Pr(\mathcal{F}|\omega)$ follows a normal distribution $N(\mu, \Sigma)$ [10]. Since the distortions of the document images vary the probability densities for the foreground and the background change over an image. We use background light normalization [18] to avoid these distortions. Each class ω_t and ω_b (text and background) is represented by its mean vector μ_ω and covariance matrix Σ_ω by $\mathcal{N}(\mu_\omega, \Sigma_\omega)$:

$$\frac{1}{\sqrt{(2\pi)^n \left|\Sigma_\omega\right|}} \exp\left(-\frac{1}{2} \left(\vec{f} - \vec{\mu_\omega} \right) \Sigma_\omega^{-1} \left(\vec{f} - \vec{\mu_\omega} \right)^T \right), \tag{7}$$

n is the dimension of the MSI. 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 [10, 4]. Applying EM on the MSI we obtain μ_t and Σ_t for the characters as well as μ_b and Σ_b for the background.

3.3 Posterior Energy $P(\omega|\mathcal{F})$

According to [10] the posterior probability can be simplified by including the contribution of the likelihood term via the singletons (i.e. pixel sites $s \in \mathcal{S}$)

$$\Pr(\omega|\mathcal{F}) \propto \exp\left(-U(\omega, \mathcal{F})\right) \tag{8}$$

$$\propto \exp\left(- \left(\sum_{s \in \mathcal{S}} V_s(\omega_s, \vec{f_s}) + \beta \sum_{\{s, r \in \mathcal{C}\}} \delta(\omega_s, \omega_r) \right) \right) \tag{9}$$

where $\beta > 0$ is a weighting parameter controlling the prior, i.e. the influence of the neighborhood connectivity. The singleton potentials $V_s(\omega_s, \vec{f_s})$ of pixel sites s are obtained from Eq. 7 by $V_s(\omega_s, \vec{f_s})$ which equals

$$\ln(\sqrt{(2\pi)^n |\Sigma_{\omega s}|}) + \frac{1}{2}(\vec{f_s} - \vec{\mu}_{\omega_s})\Sigma_{\omega_s}^{-1}(\vec{f_s} - \vec{\mu}_{\omega_s})' \tag{10}$$

Now, the energy function $U(\omega, \mathcal{F})$ of the MRF image segmentation model has the following form:

$$\sum_{s \in \mathcal{S}} \left(\ln(\sqrt{(2\pi)^n |\Sigma_{\omega s}|}) + \frac{1}{2}(\vec{f_s} - \vec{\mu}_{\omega_s})\Sigma_{\omega_s}^{-1}(\vec{f_s} - \vec{\mu}_{\omega_s})' \right) + \beta \sum_{\{s, r \in \mathcal{C}\}} \delta(\omega_s, \omega_r)$$

It is clear that Eq. 2 is equivalent to the following energy minimization problem:

$$\hat{\omega} = \arg\min U(\omega, \mathcal{F}) \tag{11}$$

Energy optimization in finding MAP-MRF solutions can be done by either local methods, like Iterated Conditional Modes (ICM) [1], or global methods like simulated annealing [13]. We used ICM to solve this global energy which are a good tradeoff between quality and computing time [10]. ICM uses a deterministic strategy to find local minimums. It starts with an estimate and then selects a label for each pixel which gives the largest decrease of the energy function. The process is repeated until it converges.

4. Experiments and Results

The proposed segmentation method has been tested on a varied set of MSI of ancient documents. The test set consists of MSI of 3 folios (17 *recto*, 29 *recto* and 30 *verso*) for which we generated the ground truth (GT) data by manually segmenting the characters. This step was supported by philologists which are experts in analyzing the documents given. Figure 2 shows details from folio 29 *recto* including a single band of the MSI with highest contrast (BP450) (a) and the corresponding GT data (b). The MSI consists of 9 spectral images as described in Table 1 and they have a radiometric resolution of 12 bit by a spatial resolution of 565 dpi.

Our approach is compared to character segmentation approaches developed especially for ancient, low contrast or noisy document images. The serialization of the $k$means algorithm by Leydier et al. [12] is the first method and the Sauvola-Niblack binarization method with the proposed parameters [17] constitutes the second method. The binarization method is performed on the BP450 single band image which represents the best contrast within the MSI data. We use the *precision* and *recall* rate to measure the accuracy metric and to rank the performance of the different methods:

$$precision = \frac{\#\text{ of text pixels correctly classified}}{\#\text{ of text pixels detected}} \quad (12)$$

$$recall = \frac{\#\text{ of text pixels correctly classified}}{\#\text{ of text pixels in GT}} \quad (13)$$

The first experiment aimed to analyze the behavior of the neighborhood system $\mathcal{N}$ and the weighting parameter β. We compared different orders n of MRF along them a first order MRF which is used for instance in [10, 2]. Table 2 shows the *precision* and *recall* rate for $\beta = 0.1$ and $\beta = 0.3$ and a neighborhood system $n = 1 \dots 5$. It can be seen that the *recall* values are very low for $n = 1 \dots 3$. This results from the background noise. The best solution is obtained with $n = 4$ and $\beta = 0.1$ which is in concordance to the proposed stroke characteristics. Generally it can be said that the smaller the considered neighborhood system, the more noise emerges in the background. On the other side, a neighborhood set considering to much pixels leads to missing characters or to closed character gaps or holes. The influence of β is likewise. The smaller β the more noise we have and values chosen too big cause missing characters ($\beta \geq 5$).

Table 1. Evaluation of neighborhood and β.

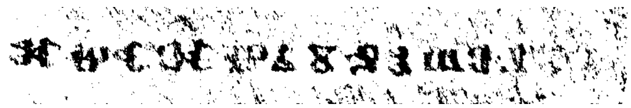
$\beta = 0.1$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
precision	0.81	0.81	0.82	0.82	0.80
recall	0.62	0.65	0.65	0.78	0.70

$\beta = 0.3$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
precision	0.80	0.79	0.83	0.79	0.75
recall	0.67	0.73	0.66	0.80	0.66

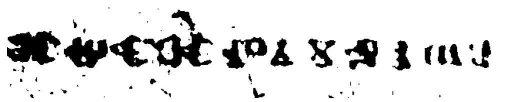
Compared to $k$means clustering and the Sauvola-Niblack binarization technique the proposed method showed the best results. Figure 2 shows a detail from folio 29 *recto* after the segmentation with the Sauvola binarization method (c), $k$means clustering (d), and the MRF approach (e) with parameters $\beta = 0.1$ and a neighborhood of 4th order. It can be seen that especially the thresholding image contains a lot of noise in the background and even within the characters. The $k$means method performs better but the rightmost character was not segmented and others are broken. The result of the MRF approach segments even the rightmost character which has very low contrast and is even hard to detect by experts. The *precision* and *recall* rate for the evaluation with the GT data can be seen in Table 3. Especially for folio 17 *recto* and 30 *verso* the $k$means shows suboptimal results due to low contrast. The locally performing thresholding method shows better results but has low *precision*. The MRF method has the best performance with a mean *precision* rate of 0.88 and a mean *recall* rate of 0.70.

(a) Single band image BP450.

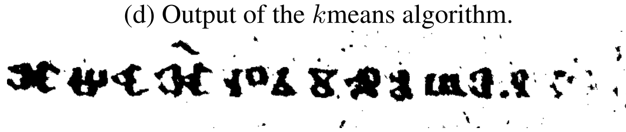
(b) GT data.



(c) Output of the Sauvola-Niblack algorithm.



(d) Output of the $k$means algorithm.



(e) Output of the proposed MRF algorithm.

Figure 2. Detail from folio 29 *recto*: BP450, GT data, and comparison of results.

Table 2. Precision and Recall.

Folio		$k$means	MRF	Sauvola
17 *recto*	precision	0.58	0.93	0.67
	recall	0.34	0.64	0.72
29 *recto*	precision	0.67	0.82	0.63
	recall	0.72	0.78	0.61
30 *verso*	precision	0.54	0.90	0.73
	recall	0.75	0.68	0.72
Average	precision	0.59	0.88	0.68
	recall	0.60	0.70	0.68

Concerning the boundary characteristics of the methods Figure 3 shows examples for the $k$means and the MRF solution. The BP450 image on the left shows three characters. The MRF model based approach shows through the consideration of contextual information in terms of stroke properties a smoothly segmented boundary (Figure 3b). The result of the $k$means method is even rougher and shows artifacts within the characters (Figure 3c).

5. Conclusions and Outlook

This paper demonstrated the effectiveness of combining spatial and spectral features for text - background segmentation in MSI of ancient manuscripts. The method is based on a MRF model and we use stroke characteristics to incor-

(a) (b) (c)

Figure 3. Differences within the boundary characteristics: (a) BP450, (b) MRF segmentation, and (c) $k$means segmentation.

porate contextual spatial information. Compared to methods like thresholding or clustering of spectral features the proposed algorithm was even able to detect low contrast characters where other methods failed. Furthermore, due to the consideration of spatial context we obtain lower noise within the background and within the text regions.

Indeed dark stained areas within the folios affect the segmentation already during the EM algorithm. Such stains are detected as foreground, even when there is enough contrast to read the characters. A solution is to remove these areas before learning the observation model $\Pr(\mathcal{F}|\omega)$. Further future work includes the usage of recently proposed energy minimization algorithms like graph cuts to improve the performance of the energy minimization by means of time and accuracy [19].

6 Acknowledgment

This work was supported by the Austrian Science Fund (FWF) under grant P19608-G12.

References

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

[2] H. Cao and V. Govindaraju. Handwritten carbon form preprocessing based on markov random field. In *Intl. Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, Minneapolis, Minnesota, USA, 2007.

[3] M. Diem, M. Lettner, and R. Sablatnig. Registration of multi-spectral manuscript images. In *8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, pages 133–140, Brighton, UK, 2007.

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

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

[6] 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.

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

[8] H. Hase, M. Yoneda, S. Tokai, J. Kato, and Y. Suen. Color segmentation for text extraction. *Int. J. Doc. Anal. Recognit.*, 6(4):271–284, 2003.

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

[10] Z. Kato and T.-C. Pong. A markov random field image segmentation model for color textured images. *Image and Vision Computing*, 24:1103–1114, 2006.

[11] 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.

[12] Y. Leydier, F. L. 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.

[13] S. Li. *Markov Random Field Modeling in Computer Vision*. Springer, 1995.

[14] V. Pervouchine, G. Leedham, and K. Melikhov. Handwritten character skeletonisation for forensic document analysis. In *Proc. ACM Symposium on Applied Computing (SAC)*, pages 754–758, Santa Fe, New Mexico, USA, 2005.

[15] K. Rapantzikos and C. Balas. Hyperspectral Imaging: Potential in Non-Destructive Analysis of Palimpsests. In *IEEE International Conference on Image Processing*, volume 2, pages 618–621, 2005.

[16] 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.

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

[18] 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.

[19] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(6):1068–1080, 2008.

[20] C. M. Thillou and B. Gosselin. Spatial and color spaces combination for natural scene text extraction. In *Int. Conference on Image Processing ICIP06*, pages 985–988, October 2006.

[21] 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 (ICDAR01)*, pages 210–214, Washington, DC, USA, September 2001.

[22] C. Wolf and D. S. Doermann. Binarization of low quality text using a markov random field model. In *Proceedings of the International Conference on Pattern Recognition (ICPR02)*, Quebec, Canada, August 2002.