Combining Spectral and Spatial Features for Robust Foreground-Background Separation

Martin Lettner and Robert Sablatnig

Computer Vision Lab
Institute of Computer Aided Automation
Vienna University of Technology, Austria
{martin.lettner,robert.sablatnig}@tuwien.ac.at

Abstract—Foreground-background separation in multispectral images of damaged manuscripts can benefit from both, spectral and spatial information. Therefore, we incorporate a Markov Random Field which provides a powerful tool to combine both features simultaneously. Higher order models enable the inclusion of spatial constraints based on stroke characteristics. We apply belief propagation for inference and include the higher order potentials by upgrading the message update. The proposed segmentation method requires no training and is independent of script, size, and style of characters. We will demonstrate the robust performance on a set of degraded documents and on synthetic images.

Keywords—Document Image Analysis; Markov Random Fields; Binarization

I. INTRODUCTION

Multispectral imaging has proven a capable technique for the analysis and preservation of ancient or damaged documents. Especially images in the UltraViolet (UV) and Near-InfraRed (NIR) light range reveal additional information like latent texts or faded passages [1]. An important step in Document Image Analysis (DIA) is image binarization dividing page content (characters, ligatures, decoration) from background [2]. Nevertheless in most of the cases Foreground-Background Separation (FBS) for damaged manuscripts is based on sophisticated image processing techniques applied on gray level images [3]. The main idea of this paper is to combine spectral and contextual spatial information to improve the segmentation performance for FBS in multispectral images of degraded or ancient documents. Therefore we utilize a higher order Markov Random Field (MRF) model using the spectral component of the images and stroke properties to include spatial dependencies.

A widely used approach in the context of MRFs is to use energy functions composed of unary and pairwise terms. The unary terms represent the data likelihood and pairwise terms (standard 4-connected neighborhood system) encode a prior over labellings to include smoothness constraints. In order to include the spatial characteristics of strokes and to avoid the limited expressiveness of only pairwise potential functions [4] we incorporate higher order potentials to include stroke characteristics. In a previous study on FBS, we used Iterated Conditional Modes (ICM) for inference [5] and have shown that the utilization of a higher order model improves the accuracy of the results. However, since ICM proved to be inefficient [6], we apply an improved version of Loopy Belief Propagation (LBP) to solve the higher order MRF model. The main advantage is the less computation time of the algorithm [6]. The higher order potential functions are included in a new formulation of the message update rule in BP.

A. Related Work

One of the first approaches using MRF for document images was concerned in the extraction of binarized car license plates from gray scale images by using $2 \times 2$ cliques [7]. Lelore et al. presented an approach for the binarization of seriously degraded documents. The MRF model parameters are learned from training data or computed using heuristics. Two potential functions for the prior are introduced to remove noise [8]. A similar approach is proposed in [9] where the text model is learned from the degraded document itself, making the approach independent of script, font and style. Related approaches in which the restoration procedure is based on trained image patches are presented in [10], [11], [12].

Disadvantageous, the MRF models need a training phase which is on the one hand time consuming and on the other, it requires an adequate amount of training data which is hardly or not available for degraded ancient documents. Another serious limitation is that only pairwise interactions are considered which are often insufficient to capture the full statistics of the problem [13].

B. Our Approach

The proposed algorithm requires no training data since the Gaussian parameters for the foreground and the background model are computed automatically. Furthermore the method is independent of script, size, style etc. and works for arbitrary documents regardless of machine printed or handwritten texts. The main parameter of the proposed method is the average stroke width of the characters. However, this
parameter can be obtained automatically after, for instance, a pre-binarization of the input image [14].

The data set consists of a set of multispectral images of a degraded document dating from the 11th century, the so-called Missale Sinaiticum. It shows extensive damages like faded ink, blurring of the ink, or staining due to mold or humidity. The data set consists of nine spectral images ranging from UV to NIR (350−1100 nm) with a radiometric resolution of 12 bit and spatial resolution of 565 dpi. On a second data set of synthetic images, we demonstrate the robustness of the proposed method compared to adaptive thresholding. The rest of the paper is organized as follows. Section 2 introduces the MRF model and the potential functions. In Section 3 we introduce the BP algorithm for higher order MRFs. Experiments and results are presented in Section 4 and Section 5 gives a conclusion and an outlook.

II. MARKOV RANDOM FIELDS

According to the Hammersly-Clifford theorem [15], a random field X is a MRF if the posterior distribution \( \Pr(x|y) \) over the labellings follows a Gibbs distribution:

\[
\Pr(x|y) = \frac{1}{Z} \exp \left( - \sum_{c \in C} \psi_c(x_c) \right),
\]

(1)

where \( \psi_c(x_c) \) are potential functions defined over the variables \( x_c = \{x_i, i \in c\}, C \) is the set of cliques \( c \), and \( Z \) is a normalizing constant [16]. A clique \( c \) is a set of random variables \( x_c \) which are conditionally dependent on each other.

A widely used approach is to use energy functions in the context of Maximum A Posteriori (MAP) estimation of MRFs [6]. The energy function is defined on a set of discrete random variables \( X = \{X_1, X_2, \ldots, X_N\} \) taking a value from the label set \( \mathcal{L} = \{l_1, l_2\} \) (text and background). A labeling \( x \in \mathcal{L}^N \) is a possible assignment of labels to the variables. Energy functions in the context of MRFs are usually defined on unary and pairwise potentials, \( \psi_i \) and \( \psi_{ij} \) [6]:

\[
E(x) = \sum_{i \in V} \psi_i(x_i) + \sum_{(i,j) \in E} \psi_{ij}(x_i, x_j).
\]

(2)

where each random variable \( x_i \) is associated with a lattice point \( i \) in the rectangular lattice \( V = 1, 2, \ldots, N \). The neighborhood system \( E \) is the set of edges connecting variables in the random field (4-connected neighborhood system). Kohli et al. [16] extended the pairwise energy function by means of higher order potentials:

\[
E(x) = \sum_{i \in V} \psi_i(x_i) + \sum_{(i,j) \in E} \psi_{ij}(x_i, x_j) + \sum_{c \in \mathcal{S}} \psi_c(x_c).
\]

(3)

In the case of FBS for DIA we refer to image segments \( \mathcal{S} \) based on stroke properties and \( \psi_c \) are higher order potentials defined on these segments.

A. Unary Potentials

The observation model follows a normal distribution \( \mathcal{N}(\mu, \Sigma) \) where each class \( l_i \) and \( l_b \) is represented by its mean vector \( \mu \) and covariance matrix \( \Sigma \). \( \psi_i(x_i) \) is a two dimensional vector measuring how well a label fits an observation \( y_i \):

\[
\psi_i(x_i) = \mathcal{N}(y_i|\mu, \Sigma).
\]

(4)

Due to partially low contrast of the ink caused through fading or water stains our approach is based on local calculations of \( \mu \) and \( \Sigma \). The entities are modeled by a Gaussian Mixture Model (GMM) and \( \mu \) and \( \Sigma \) are found via Expectation-Maximization.

B. Pairwise Potentials

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

\[
\psi_{ij}(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j, \\ \rho_1(\nabla I) & \text{otherwise,} \end{cases}
\]

(5)

where the function \( \rho_1(\nabla I) \) is defined in terms of the image gradient between the pixels \( i \) and \( j \), resulting in a low penalty when two neighbored pixels \( x_i \) and \( x_j \) have similar observations.

C. Higher Order Potentials

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

\[
\psi_c(x_c) = \begin{cases} 0 & \text{if } x_i = l_k, \forall i \in c \\ |y_i - \mu_{N_i}| & \text{otherwise,} \end{cases}
\]

(6)

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

III. BELIEF PROPAGATION FOR HIGHER ORDER MRFs

In order to implement BP for higher order MRFs, Zhang et al. managed the message update rule for utilizing higher order potentials by combining messages from two neighboring nodes to one message [17]. Potetz et al. proposed a technique to compute the BP messages in time linear with respect to clique size by using a factor graph [13]. In contrast to these studies, we include the higher order potentials \( \psi_c(x_c) \) in the LBP messages \( m_{ij} \) to incorporate higher order models. 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 \( m_{ij}^{l}(x_j) \) 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:

$$m_{t,j}^{(j)}(x_j) \leftarrow \max_{x_i} \psi_i(x_i) \psi_{ij}(x_i,x_j) \psi_c(x_{c_j}) \prod_k m_{kj}(x_i),$$

where $k \in N_i \setminus j$ 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:

$$b_t(x_i) = \psi_i(x_i) \prod_{k \in N_i} m_{ki}(x_i).$$

The label $l_k \in L$ which minimizes $b_t(x_i)$ individually at each node is selected.

IV. EXPERIMENTS AND RESULTS

The proposed algorithm is tested on synthetic images and on four images of the Missale Sinaiticum with manually tagged Ground Truth (GT) data. We evaluate the influence of higher order MRFs and compare the results to pairwise connections for the proposed BP algorithm and Graph Cuts (GC) for higher order potentials [16]. Both algorithms use the same potentials $\psi_i$, $\psi_{ij}$, and $\psi_c$. The proposed method of combining spatial and spectral features is compared to Adaptive Binarization (AB) [18] applied on a single band image. Since UV and the adjacent visible illumination enhances the visibility of faded or washed original text on parchment [1], AB is applied on the BP450 image which depicts an image filtered by a band-pass with center frequency $\lambda = 450 nm$. An example can be seen in the first row of Figure 2.

We use the $F_1$ score which follows $F_1 = \frac{2P \cdot R}{P + R}$ to measure the accuracy ($P$ constitutes the precision and $R$ the recall rate). To limit the influence of the higher order potential $\psi_c(x_c)$ we use an weighting parameter $\beta = 0.3$. Since the mean stroke width in both data set corresponds to five, we use image cliques which include pixels in a diameter $\varnothing = 5$.

Table 1 shows precision $P$, recall $R$ and $F_1$ score for pairwise models (BP-P, GC-P), higher order models (BP-HO, GC-HO), and AB. The $F_1$ score is calculated from the average values from the four images. The results for the $F_1$ score reflect that BP achieves better results for the local character segmentation problem than GC. Furthermore, the proposed spatial characteristics of strokes integrated in higher order models outperform the pairwise models. The influence between pairwise and higher order models of GC is not as crucial as for BP since this inference method finds a global optimum. Both methods outperform AB.

Results imaging from the experiments on the Missale Sinaiticum can be seen in Figure 2. The first row shows the BP450 image where especially the outermost right characters are hard to detect. The output of higher order BP and GC, detect these characters, however the noise in the background increases in this region due to the low contrast. Compared to adaptive binarization the higher order MRF approaches produce less noise within the characters and the background and detect even highly degraded characters by simultaneously using spectral and spatial information.

In a second experiment we test the algorithms on a synthetic image with varying text color. The image is degraded in ten steps by adding Gaussian noise with varying variance $\sigma$ between 0.00 and 0.04. The proposed higher order BP with $r = 5$ shows a best performance. The $F_1$ measure for AB drops already below 0.50 with $\sigma > 0.015$ for AB. Best results for $F_1$ can be obtained with BP-HO for the proposed stroke model ($\varnothing = 5$).

![Figure 1. $F_1$ for a synthetic image with varying text color and varying Gaussian noise ($\sigma = 0.00 – 0.04$).](image)

V. CONCLUSION AND OUTLOOK

This paper outlined three contributions for foreground-background separation in degraded manuscripts. First of all, the combination of spatial and spectral features improves the segmentation accuracy especially in faded regions with low contrast. Second, in order to include spatial constraints of strokes, we use higher order MRFs to enforce label consistency resulting in less foreground and background noise. And third, local inference methods like belief propagation achieve better results for local optimization problems than global methods. We have used a very simple function for the higher order potential based on the maximum color
distance. Since belief propagation works for arbitrary potential functions it is possible to use more sophisticated functions to simulate the spatial arrangement of stroke characteristics in order to avoid broken characters and noise in the background.

ACKNOWLEDGMENT

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

REFERENCES


