# Markov Random Field Based Binarization for Hand-held Devices Captured Document Images

Xujun Peng[*]
CUBS, Department of CSE
University at Buffalo, SUNY
Amherst, NY 14228, USA
xpeng@buffalo.edu

Srirangaraj Setlur
CUBS, Department of CSE
University at Buffalo, SUNY
Amherst, NY 14228, USA
setlur@buffalo.edu

Venu Govindaraju
CUBS, Department of CSE
University at Buffalo, SUNY
Amherst, NY 14228, USA
govind@buffalo.edu

Ramachandrula Sitaram
HP Labs India
Bangalore 560030, India
sitaram@hp.com

## ABSTRACT

In this paper, a novel Markov random fields (MRF) based binarization algorithm is proposed to segment foreground text from document images captured using hand-held devices (such as cell-phone or digital camera). In the MRF based framework, an edge potential feature is extracted to preserve the strokes of foreground text and to remove isolated noise and an intensity feature is used to smooth the entire document image. Prior to binarization, we use a nonlinear function to enhance the quality of document images which suffer from insufficient or uneven illumination. Experimental results show that our method outperforms other state-of-the-art approaches.

## Keywords

Binarization, Document, MRF

## 1. INTRODUCTION

After decades of research and development, document processing systems have attained reasonable success, in that large volumes of paper documents can be digitized and processed quickly to facilitate various levels of search and retrieval based on the textual content of the documents. The underlying components of most document processing systems, such as document layout analysis and optical character recognition (OCR), rely on high quality binarized document images. Hence, pre-processing steps such as binarization play a critical role in the success of these systems. The traditional and most popular device for document image acquisition has been the scanner which provides the most reliable digitizing result as users can control the capture environment. However, the advent of small, high resolution digital cameras and hand-held devices such as mobile phones with imaging capabilities have led to hand-held devices being widely used by lay users to capture document images of interest for future use.

Despite the convenience provided by hand-held devices, there are a number of factors that hinder downstream document processing tasks on images captured by these devices. The first obstacle is a lack of control of lighting conditions which causes insufficient, non-uniform or over illumination of document images. This causes most global thresholding based binarization algorithms, such as Otsu's [14] method, to be ineffective. To overcome the limitations of single threshold based methods, researchers have tried mapping the original document image to a new feature domain prior to binarization. Valizadeh et al. [20] used a sigmoid function to enhance the document image before Otsu thresholding. Lu and Tan [10] employed two rounds of polynomial smoothing procedure to normalize gray scale images and binarized them using a single threshold. Similarly, a linear plane estimation method was proposed by Shi and Govindaraju [18] to segment text from degraded historical documents. Pilu and Pollard [15] used a retinex algorithm to estimate the reflectance surface on which they applied their binarization.

For images with a wide variation in intensities across the document image, researchers have used many local or adaptive thresholding methods for binarization. The earliest attempt at using local thresholds can be traced back to Niblack's method [13] which utilizes mean value and variance within a small window to determine the property of centered pixels and is formulated as:

$$T(i) = \mu(i) + k \times \sigma(i) \tag{1}$$

where $\mu(i)$ and $\sigma(i)$ are mean value and standard variance in a small window centered on pixel $i$, $k$ is a parameter less than 0. The potential problem of this method is that large amount of noise is introduced in pure blank background areas and it is sensitive to the window size. An improvement of Niblack's method was described by Sauvola et al [17] using

[*] Corresponding author

*ICVGIP* '10, December 12-15, 2010, Chennai, India

the formula:

$$T(i) = \mu(i) + \left[1 + k\left(\frac{\sigma(i)}{R} - 1\right)\right] \tag{2}$$

where $\mu(i)$ and $\sigma(i)$ have the same meaning as Equation 1, $k$ takes positive values and $R$ is 128 for a grayscale document.

Both Niblack's algorithm and Sauvola's algorithm are sensitive to the parameter $k$ and window size $s$ which limit their performances, so Bukhari et al. [3] suggested an adaptive method to estimate free parameters according to local ridge properties. By using Otsu method locally, Moghaddam and Cheriet [12] calculated the optimal window size based on stroke width and text line height within the document image.

Another drawback of a document image captured using a hand-held devices is that the low cost lenses that are typically used can cause out of focus blur and down sampling blur and produce lots of noise [19, 21, 11, 4].

In this paper, we focus on enhancing the quality of the binarized image and suppressing noise due to uneven or insufficient lighting. In section 2, a non-linear function based segmentation surface is proposed which is followed by a novel Markov random field based relabeling method described in section 3. The experimental setup and results are described in section 4 and we present our conclusions in section 5.

## 2. INITIAL BINARIZATION

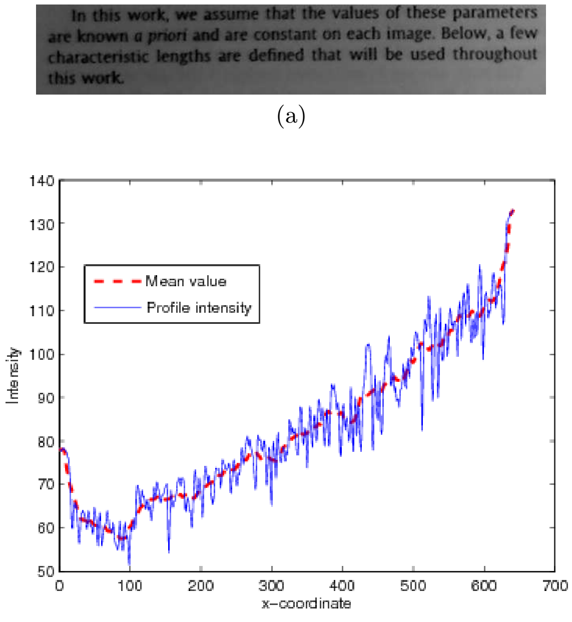
Due to uneven or bad illumination, the intensity of background may not be consistent within a camera captured document image. Fig.1(a) shows an example document image captured under uneven lighting condition whose background is darker on the left part of image than on the right side. Fig.1(b) shows the vertical profile intensity of this document image and it is apparent that any single threshold based binarization method will not work since the intensity varies gradually from left to right. Ideally, the thresholding should be an adaptive curve that tracks the intensity. The mean value of intensity in a window around a given pixel is a likely threshold which is the basis of many adaptive thresholding binarization techniques [13, 17]. As shown in Fig. 1(b), the average of profile intensity which is calculated using a 10 sized window crosses the intensity curve and can be used as a starting point to estimate an optimal local threshold.

In our work, prior to estimating the adaptive threshold surface, we compute the mean value $\mu_i = \sum_{x,y} I(x,y)/(m \times n)$ and variance $\delta_i = \sqrt{\sum_{x,y}(I(x,y) - \mu_i)^2/(m \times n)}$ for a given pixel $i$ using a $m \times n$ sized window centered on this pixel. The maximum variance $\delta_{max}$ and minimum variance $\delta_{min}$ are also obtained over the entire document image. The adaptive threshold for each pixel is calculated according to a logistic function as described in Equation 3:

$$\mathcal{O}(i) = \mu_i \left\{ \frac{1-k}{\left(1 + e^{-B\left(\frac{\delta_i - \delta_{min}}{\delta_{max} - \delta_{min}} - M\right)}\right)^{1/\nu}} + k \right\} \tag{3}$$

where $B$ controls the growth rate of the logistic curve, $M$ and $\nu$ affect the time for which maximum growth occurs and $k$ is the minimum value the curve can achieve.

The idea behind Equation 3 is that text areas have larger values of variance and mean value of given pixel $i$ can be a candidate threshold. On the other hand, the variance of background pixels is small and the threshold should be smaller than the mean value to avoid noise. Then the free parameters of Equation 3 are chosen such that the threshold is close to mean value $\mu_i$ if variance $\delta_i$ is big and the threshold is smaller than the mean value if variance $\delta_i$ is small.

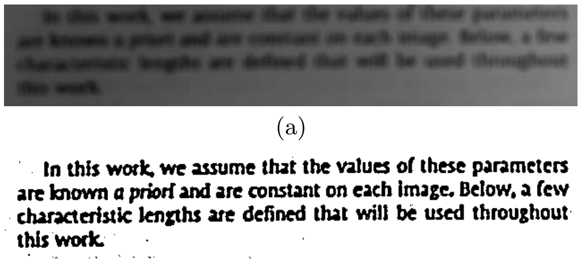


In this work, we assume that the values of these parameters are known *a priori* and are constant on each image. Below, a few characteristic lengths are defined that will be used throughout this work.

(a)

(b)

**Figure 1: A sample image and its vertical profile**



In this work, we assume that the values of these parameters are known a priori and are constant on each image. Below, a few characteristic lengths are defined that will be used throughout this work.

(a)

In this work, we assume that the values of these parameters are known *a priori* and are constant on each image. Below, a few characteristic lengths are defined that will be used throughout this work.

(b)

**Figure 2: Adaptive threshold surface and initial binarization result**

Fig. 2 shows the adaptive threshold surface calculated using Equation 3 and the initial binarization result of Fig. 1(a) using this surface. From Fig. 2 we can see that the intensity of the threshold surface changes gradually along with the intensity of the document image and the result of binarized image has consistent text stroke width which is not affected by the variation in intensities caused by uneven lighting.

## 3. MRF BASED RELABELING

Normally, noise in the background and holes within text cannot be avoided by using adaptive threshold based binarization only as shown in Fig. 2 (b). In this paper, we propose a Markov random fields based relabeling procedure to remove noise and holes from the initial binarized document image.

## 3.1 MRF and Gibbs Model

In recent years, Markov Random Fields (MRF) based image restoration algorithms have attracted interest from researchers in document processing. Considering the ideal document image to be a binary image which is down-sampled and blurred to a gray-scale image by adding Gaussian noise, document binarization can be looked at as a special restoration problem. Lelore and Bouchara [8] proposed a MRF model for binarization which can remove noise and improve the character connectivity. Lettner et al. [9] used a similar framework but defined the Gibbs distribution in a different form to binarize the degraded documents. To binarize unevenly illuminated documents, Kuk et al. [7, 6] initially segmented the entire document image using mean filter response as features and relabeled the pixels using a minimization technique which can be looked at as a variant of Gibbs model. As proved by Hammersley and Clifford [5], Markov random fields can be considered equivalent to Gibbs fields. In the following sections, we use these two terms interchangeably.

The relabeling of the initial binarized document image which assigns one of two labels (black or white) to each pixel of the document can be modeled as a maximum a *posteriori* Markov random field (MAP-MRF) estimation of ideal binarized document image $X$ given only the degraded camera captured image $Y$. In our MRF framework, the observed degraded image $Y$ is the original gray-scale document image along with its initial binarized image and our task is to calculate an optimal configuration of $X$ which maximizes the posteriori:

$$\begin{aligned}\overline{X} &= \arg\max_X P(X|Y) \\ &= \arg\max_X \prod_{i=1}^{n} P(x_i|y_i, x_{N-\{i\}}) \end{aligned} \tag{4}$$

where $y_i$ is the observed feature and $x_i$ is hidden configuration of pixel $i$ respectively, and $N-\{i\}$ denotes all pixels in document image except pixel $i$.

By taking Bayesian rule and considering Markov property, the MAP-MRF estimation can be re-written as:

$$\begin{aligned}\overline{X} &= \arg\max_X \prod_{i=1}^{n} \frac{P(y_i|x_i, x_{N-\{i\}})P(x_i|x_{N-\{i\}})}{P(y_i|x_{N-\{i\}})} \\ &= \arg\max_X \prod_{i=1}^{n} P(y_i|x_i, x_{N-\{i\}})P(x_i|x_{N-\{i\}}) \\ &= \arg\max_X \prod_{i=1}^{n} P(y_i|x_i)P(x_i|x_{N(i)}) \\ &= \arg\min_X \left[ -\sum_{i=1}^{n} logP(y_i|x_i) - \sum_{i=1}^{n} logP(x_i|x_{N(i)}) \right] \end{aligned} \tag{5}$$

where likelihood $P(y_i|x_i)$ represents the dependency of observations on hidden configuration and prior $P(x_i|x_{N(i)})$ shows the influence from immediate neighbors $N(i)$ to centered pixel $i$. In our work, we use 4-connectivity lattice system where each pixel has four neighbors.

Generally, the optimal configuration $X$ of Equation 5 can be achieved by minimizing energy function [2, 16]:

$$E(X) = \sum_{i\in\mathcal{V}} U_i(x_i) + \sum_{(i,j)\in\mathcal{E}} V_{i,j}(x_i, x_j) \tag{6}$$

where $\mathcal{V}$ is the vertex corresponding to pixels in the image and $\mathcal{E}$ is the edge connection between pixels, $U_i(x_i)$ denotes the unary energy which is derived from $-logP(y_i|x_i)$ and pairwise energy $V_{i,j}(x_i, x_j)$ is derived from $-logP(x_i|x_j)$ in Equation 5 respectively. The unary energy $U_i(x_i)$ tends to force hidden configuration $x_i$ to have a value which is compatible with its observation $y_i$ and pairwise energy $V_{i,j}(x_i, x_j)$ forces $x_i$ to be smoothly connected with its neighbors.

## 3.2 Edge potentials

Unlike other MRF based relabeling algorithms which only use the intensity difference between neighboring pixels and smooth the entire document image [6, 9], we explore a stroke width related feature which preserves the edge of strokes and removes noise from the document.

For each connected text component of the binarized document image obtained from section 2, we compute the shortest distance from foreground pixels to the background which is denoted as $e_n(i)$ for pixel $i$ in connected component $n$. The maximum distance from the foreground pixel within a connected component $n$ to the background is represented as $\hat{e}_n$. To measure the potential of a foreground pixel to be on the edge, the distance from the foreground pixels to the innermost pixel within a connected component is calculated as:

$$s_n(i) = \hat{e}_n - e_n(i) \tag{7}$$

Fig. 3(a) shows the edge potential of an example character **a** where the brighter area in the character corresponds to a low edge potential and the darker area corresponds to higher edge potential. The velocity vectors which indicate the direction and strength of edge potential for each foreground pixel are shown in Fig. 3(b), from which we can see that the edge potential decreases gradually from edge pixels to inner pixels within the character.

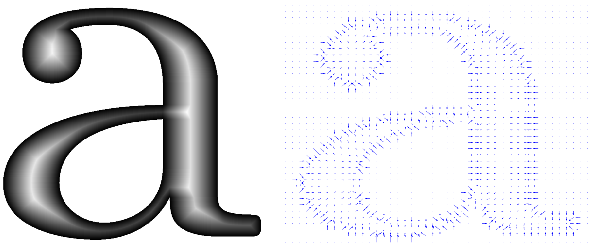

(a) (b)

**Figure 3: Edge Potential**

## 3.3 Likelihood and Prior

The goal of the MRF based relabeling procedure is to remove noise and smooth the entire document image. As described in section 3.1, $-logP(y_i|x_i)$ can be approximately represented by an unary energy function $U_i(x_i)$ which forces the label $x_i$ of pixel $i$ to be close to its observation $y_i$. We define $U_i(x_i)$ as:

$$U_i(x_i) = \lambda\sqrt{(y_i - x_i)^2} \tag{8}$$

where $y_i$ is the gray-scale value for pixel $i$ and $x_i$ takes value of 255 for background and 0 for foreground.

Pairwise energy function $V_{i,j}(x_i, x_j)$ or $-logP(x_i, x_j)$ is defined in Equation 9 using edge potential features along with gray-scale value for each pair of pixels,

$$V_{i,j}(x_i, x_j) = \alpha \exp\left(\frac{(-1)^{|x_i - x_j|}}{|[s(i) - s(j)]^2 - (sw/2)^2| + 1}\right) + \beta \exp\left(\frac{(-1)^{|x_i - x_j|}|y_i - y_j|}{256}\right) \quad (9)$$

where $x_i$ and $x_j$ are the hidden configuration (0 for foreground and 255 for background) for pixel $i$ and $j$, $s(i)$ and $s(j)$ are edge potentials for pixel $i$ and $j$ using Equation 7, $sw$ is the mean stroke width within the document image, and $y_i$ and $y_j$ are the gray-scale values for two neighboring pixels.

The underlying principle of Equation 9 is that if two neighboring pixels are from the same source, e.g. both of them are from foreground text, they should have similar edge potentials and gray-scale values which cause the pairwise energy to be low.

To achieve the global minimum energy corresponding to the optimal configuration of MRF, We use a graph $G = \langle V, E \rangle$ which contains two special terminals called *source* and *sink* to model the image and the graph cuts algorithm [2] is used in our experiment repeatedly until the binarized result is stable. Prior to the iteration, the original document image is binarized using the adaptive threshold surface described in section 2. During each iteration, the edge potential of each pixel is computed on the binarized image prior to the calculation of unary energy and pairwise energy according to Equation 8 and Equation 9. The unary energy of each pixel is then assigned to the edge which connects the *source*/*sink* and the corresponding node. The pairwise energy is assigned to the edge which connects corresponding neighboring nodes in the graph $G$ respectively. Finally, the initial binarized image is refined using graph cuts algorithm in each iteration and is used as the initial input binarized image to the next iteration until the difference between two images is smaller than a pre-defined threshold. The overall procedure of our algorithm is described in Fig. 4.

```
MRF Relabeling Algorithm
Input: Gray-scale document image $I$.
Initial Binarization:
  1: Estimate adaptive threshold surface using Equation 3;
  2: Segment text from background using adaptive threshold.
  3: Initialize difference threshold $t$ and maximum iteration number $N$.
Relabeling:
  4: Extract edge potential feature on binarized document image for each pixel according to Equation 7;
  5: Compute the unary energy of each pixel for the entire document image using Equation 8;
  6: Calculate the pairwise energy of each pixel for the entire document image using Equation 9;
  7: Use graph cuts algorithm to get the optimal relabeling of the binarized image;
  8: Calculate the difference $\varepsilon^{(n)}$ between the relabeled image and the previous binarized image:
     $\varepsilon^{(n)} = \sum_i \sqrt{(x_i^{(n)} - x_i^{(n-1)})^2}$
     where $x^{(n)}$ is the configuration after relabeling for pixel $i$ and $x^{(n-1)}$ is the older label for pixel $i$;
  9: If $\varepsilon^{(n)} < t$ or $n > N$, go to next step, otherwise, $n = n + 1$ and go back to step 4;
Output: binarized result.
```

**Figure 4: Overall Procedure of Binarization**

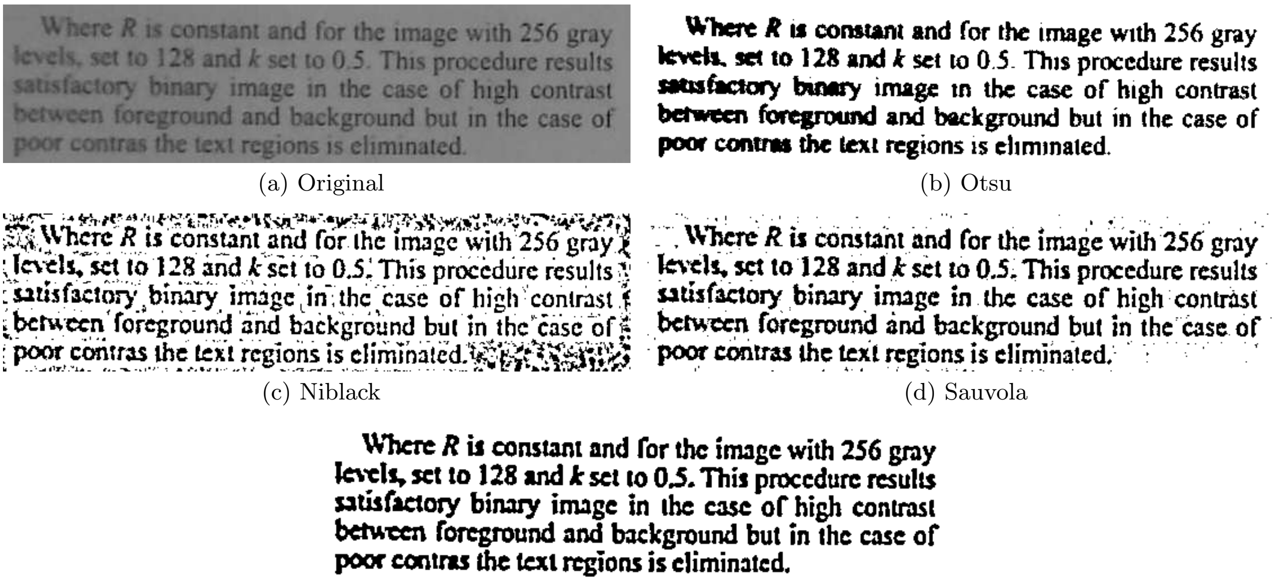
## 4. EXPERIMENTAL RESULTS

In our experiment, we use a data set of 28 pages of document images which were captured using a hand-held cellphone camera with a resolution of 3.8 mega pixels. The document pages in our data set are research papers in two column style and the text occupies at least 95% of the page area. All images were captured in an indoor office environment and our experiment was carried out using image portions with insufficient or uneven illumination along with out of focus blur. Fig. 5(a) shows an example image in our data set.

We used both qualitative judgement in the form of visual inspection as well as a quantitative metric based on OCR performance to evaluate the proposed binarization algorithm.

Fig. 4 shows an example of visual comparison of the binarization results with Otsu method, Niblack method, Sauvola method and the proposed MRF based method. Fig. 5(b) is the binarized result of Otsu method where strokes on the left portion of the image are thicker than those on the right portion because of uneven illumination. Although the stroke width is consistent in the result from the Niblack method, a lot of noise is introduced in pure background areas as shown in Fig. 5(c). The Sauvola method has better performance as shown in Fig. 5(d) where noise is restrained and strokes are retained. The last Fig. 5(e) shows the result of the proposed MRF based binarization method which not only removes all isolated noise, but enhances the quality of strokes.

The goal of most binarization algorithms is to provide a reliable binarized image for further document processing such as Optical Character Recognition (OCR). So, we compared OCR results on the binarized images generated from the four different binarization methods considered in our experiments. The OCR experiment was carried out using the open source OCR software Tesseract [1] without deskew or other pre-processing procedures. The OCR accuracy was measured using the F-Score which is defined in Equation 10, 11, and 12:



Where $R$ is constant and for the image with 256 gray levels, set to 128 and $k$ set to 0.5. This procedure results satisfactory binary image in the case of high contrast between foreground and background but in the case of poor contras the text regions is eliminated.

(a) Original

(b) Otsu

(c) Niblack

(d) Sauvola

(e) Proposed method

**Figure 5: Binarization result of uneven illuminated document image using proposed algorithm compared with other methods. (a) The original grayscale document image, (b) Otsu binarization result, (c) Niblack binarization result with $k = -0.3$ and $s = 11$, (d) Sauvola binarization result with $k = 0.02$ and $s = 11$, (e) Proposed MRF based binarization result.**

$$FScore = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

where $TP$ is the number of the words that appeared in both ground truth and result, $FN$ is the number of words which are only in the ground truth and $FP$ is the number of words which are only in the OCR results.

As can be seen from Table 1, the proposed MRF based binarization algorithm provides better OCR accuracy than the other three methods whereas the Niblack method has the worst OCR performance since it introduces more noise into the image.

**Table 1: OCR Results from Tesseract on binarized images**

| Method | OCR Result(F-Score) |
|---|---|
| Otsu [14] | 35.7% |
| Niblack [13] | 5.8% |
| Sauvola [17] | 42.5% |
| Proposed method | 56.0% |

## 5. CONCLUSIONS

In this paper, we propose a novel MRF based algorithm to binarize document images captured using a hand-held device such as a mobile phone. Prior to binarization, a non-linear transformation function is used to estimate an adaptive threshold surface on which the document image is initially segmented. To remove noise and retain text strokes, a novel edge potential feature is employed in our MRF framework. The experimental results show that the proposed method provides a visually superior binarized output which also results in better OCR performance than the other techniques.

## 6. REFERENCES

[1] Tesseract-ocr. *http://code.google.com/p/tesseract-ocr/.*

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Recognition and Machine Intelligence*, 26(9):1124–1137, Sept 2004.

[3] S. S. Bukhari, F. Shafait, and T. M. Breuel. Adaptive binarization of unconstrained hand-held camera-captured document images. *Journal of Universal Computer Science*, 15(18):3343–3363, 2009.

[4] D. Doermann, J. Liang, and H. Li. Progress in camera-based document image analysis. In *Proc. IEEE 7th ICDAR*, volume 1, pages 606–616, 2003.

[5] J. M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. 1971.

[6] J. G. Kuk and N. I. Cho. Feature based binarization of document images degraded by uneven light condition. In *Proc. IEEE 10th ICDAR*, pages 748–752, July 2009.

[7] J. G. Kuk, N. I. Cho, and K. M. Lee. MAP-MRF approach for binarization of degraded document image. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 2612–2615, Oct. 2008.

[8] T. Lelore and F. Bouchara. Document image binarisation using markov field model. In *Proc. IEEE 10th ICDAR*, pages 551–555, July 2009.

[9] M. Lettner and R. Sablatnig. Spatial and spectral based segmentation of text in multispectral images of ancient documents. In *Proc. IEEE 10th ICDAR*, pages 813 –817, July 2009.

[10] S. Lu and C. L. Tan. Binarization of badly illuminated document images through shading estimation and compensation. In *Proc. IEEE 9th ICDAR*, volume 1, pages 312–316, Sept. 2007.

[11] T. A. Mahmoud and S. Marshall. Document image sharpening using a new extension of the aperture filter. *Signal, Image and Video Processing*, 3(4):403–419, 2009.

[12] R. F. Moghaddam and M. Cheriet. A multi-scale framework for adaptive binarization of degraded document images. *Pattern Recognition*, 43:2186–2198, 2010.

[13] W. Niblack. *An Introduction to digital image processing.* Prentice Hall, Englewood Cliffs, NJ, USA, 1986.

[14] N. Otsu. A threshold selection method from gray level histograms. *IEEE Trans. Systems, Man and Cybernetics*, 9:62–66, Mar 1979.

[15] M. Pilu and S. Pollard. A light-weight text image processing method for handheld embedded cameras. In *The 13th British Machine Vision Conference*, 2002.

[16] A. Raj and R. Zabih. A graph cut algorithm for generalized image deconvolution. In *Proc. IEEE 10th ICCV*, volume 2, pages 1048 –1054, Oct 2005.

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

[18] Z. Shi and V. Govindaraju. Historical document image segmentation using background light intensity normalization. In *Document Recognition and Retrieval XII*, volume 5676, pages 167–174. SPIE, 2005.

[19] M. J. Taylor and C. R. Dance. Enhancement of document images from cameras. In *Document Recognition V*, volume 3305, pages 230–241. SPIE, 1998.

[20] M. Valizadeh, N. Armanfard, M. Komeili, and E. Kabir. A novel hybrid algorithm for binarization of badly illuminated document images. In *Proc. IEEE 14th International CSI Computer Conference*, pages 121–126, 2009.

[21] A. Zandifar, R. Duraiswami, A. Chahine, and L. S. Davis. A video based interface to textual information for the visually impaired. In *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*, pages 325 – 330, 2002.