Journal of Universal Computer Science, vol. 15, no. 18 (2009), 3343-3363
submitted: 21/10/09, accepted: 1/12/09, appeared: 28/12/09 © J.UCS

Adaptive Binarization of Unconstrained Hand-Held Camera-Captured Document Images

Syed Saqib Bukhari
(Technical University of Kaiserslautern, Germany
bukhari@informatik.uni-kl.de)

Faisal Shafait
(German Research Center for Artificial Intelligence, Kaiserslautern, Germany
faisal.shafait@dfki.de)

Thomas M. Breuel
(Technical University of Kaiserslautern, Germany
tmb@informatik.uni-kl.de)

Abstract: This paper presents a new adaptive binarization technique for degraded hand-held camera-captured document images. The state-of-the-art locally adaptive binarization methods are sensitive to the values of free parameter. This problem is more critical when binarizing degraded camera-captured document images because of distortions like non-uniform illumination, bad shading, blurring, smearing and low resolution. We demonstrate in this paper that local binarization methods are not only sensitive to the selection of free parameters values (either found manually or automatically), but also sensitive to the constant free parameters values for all pixels of a document image. Some range of values of free parameters are better for foreground regions and some other range of values are better for background regions. For overcoming this problem, we present an adaptation of a state-of-the-art local binarization method such that two different set of free parameters values are used for foreground and background regions respectively. We present the use of ridges detection for rough estimation of foreground regions in a document image. This information is then used to calculate appropriate threshold using different set of free parameters values for the foreground and background regions respectively. The evaluation of the method using an OCR-based measure and a pixel-based measure show that our method achieves better performance as compared to state-of-the-art global and local binarization methods.

Key Words: Binarization of Document Images, Camera-Captured Document Images

Category: I.4, I.4.1, I.4.3, I.7, I.7.2

1 Introduction

Scanners are traditionally and widely used in document image capturing for document analysis systems, like optical character recognition (OCR). Scanners produce planar document images with a high resolution. From decades many novel approaches have been proposed for planar document image segmentation [Shafait et al., 2008d] and OCR [Mori et al., 1992]. Nowadays cameras are available widely at low cost and embedded with around all mobile devices, that offer fast, flexible and non-contact document imaging. On one hand, these advantages make camera a potential substitute of scanner for document capturing and on other hand open doors for many new applications, like mobile OCR, digitizing thick books, digitizing fragile historical documents, finding text-in-scene-images, etc. But the quality of unconstrained hand-held camera-captured document images is lower than the quality of scanned document images because of the degradations which are not very common in scanned images, like perspective distortions, non-uniform shading, image blurring, character smearing (due to low resolution) and lighting variations.

In the case of scanned document images, most of the state-of-the-art document analysis systems have been designed to work on binary document images [Cattoni et al., 1998]. Therefore document image binarization is an important initial step in most of the scanned document image processing tasks, such as OCR [Mori et al., 1992], page segmentation [Shafait et al., 2008d], layout analysis [Shafait et al., 2008b] etc.

In the case of camera-captured document images, current OCR systems which are designed for scanner based planar document images do not have capability to deal with geometric and perspective distortions. Therefore, current OCR systems give poor performance when applied directly to warped camera-captured document images. Designing dewarping techniques for flattening the document images is a possible solution for improving the performance of OCR systems on camera-captured document images. Over last decade, different approaches have been proposed for document image dewarping [Liang et al., 2005, Shafait and Breuel, 2007]. These approaches can be divided into two main categories based on the document capturing methodology: (i) approaches in which specialized hardware arrangement, like stereo-camera, is required for 3D shape reconstruction of warped document [Cao et al., 2003, Brown and Seales, 2004, Tan et al., 2006] and (ii) approaches in which dewarping method is designed for image which is captured using a single hand-held camera in uncontrolled environment [Zhang and Tan, 2003, Lu and Tan, 2006, Lu et al., 2005, Fu et al., 2007, Ulges et al., 2005, Gatos et al., 2007, Bukhari et al., 2009a]. Most of the monocular dewarping techniques work on binarized images.

This discussion concludes that binarization is the most important initial step for both scanned and camera-based document image analysis. But binarization of hand-held camera-captured document images is more challenging than scanned images because of one or more of the following distortions in camera-captured document images: bad shading, blurring, non-uniform illumination and low resolution.

1.1 Related Work

From decades, many different approaches have been proposed for the binarization of the grayscale document images [Otsu, 1979, White and Rohrer, 1983, Bernsen, 1986, Niblack, 1986, O'Gorman, 1994, Sauvola and Pietikainen, 2000, Kim, 2004, Gatos et al., 2006, Lu and Tan, 2007, Shafait et al., 2008c] and color images [Sobottka et al., 2000, Tsai and Lee, 2002, Badekas et al., 2006] in the literature. Additionally, grayscale binarization techniques can be applied to color documents by first converting them into grayscale. Grayscale binarization approaches can be classified into two main groups: i) global binarization methods and ii) local binarization methods.

Global binarization methods (like Otsu [Otsu, 1979]) estimate a single threshold value for the binarization of whole document. Then, based on the intensity values, each pixel is assigned either to foreground or background. Some researchers [Sezgin and Sankur, 2004, Badekas and Papamarkos, 2005] have evaluated different state-of-the-art global binarization methods and reported that Otsu binarization method [Otsu, 1979] is better than other types of global binarization methods. Global binarization methods are computationally inexpensive and perform better for typical scanned document images. However, they produce marginal noise artifacts [Shafait et al., 2008a] if grayscale document contains non-uniform illumination, which is usually present in case of scanned thick book, scanned historical document and camera-captured document images.

Local binarization methods [Bernsen, 1986, Niblack, 1986, O'Gorman, 1994, White and Rohrer, 1983, Sauvola and Pietikainen, 2000] try to overcome these problems by calculating threshold values for each pixel differently using local neighborhood information. Evaluations of local binarization methods have reported that Sauvola binarization method [Sauvola and Pietikainen, 2000] is better than other types of local binarization methods. Generally, local binarization methods perform better than global binarization methods on degraded document images but are computationally slow, sensitive to the selection of free parameter values [Rangoni et al., 2009] and do not work well for degraded camera-captured document images.

In recent years, some special global binarization and local binarization techniques [Kim, 2004, Gatos et al., 2006, Lu and Tan, 2007] have been proposed for improving the binarization of degraded historical and camera-captured document images. Gatos et. al [Gatos et al., 2006] proposed local binarization method for scanned degraded historical document images. This technique has not yet been tested on blurred and low-resolution camera-captured document images. Kim [Kim, 2004] proposed multi-window based local binarization method for camera-captured document images, which is a modification of Sauvola binarization method. This approach contains more free parameters than Sauvola binarization method. Lu and Tan [Lu and Tan, 2007] proposed global binarization method for camera-captured document images. But their method is based on the assumption that document image contains uniform illumination and uniform background, which is not usually the case.

In this paper, we deal with the binarization of degraded grayscale camera-captured document images having distortions like bad shading, blurring, low resolution and non-uniform illumination. Here, we describe a local binarization method which is less sensitive to free parameter values than well know existing methods. Instead of using the same free parameter values for all pixels in a document image, unlike other local binarization methods, we select different values of free parameters for pixels that belong to roughly estimated foreground regions and pixels that belong to background regions. Here we use a combination of multi-oriented multi-scale anisotropic Gaussian smoothing and ridges detection technique for estimating foreground regions information, which we have already reported in [Bukhari et al., 2009c, Bukhari et al., 2009d].

Part of the work presented here was published in [Bukhari et al., 2009b] for timely dissemination of this work. This paper is a substantially extended version of the previous conference publication.

The rest of this paper is organized as follows: Section 2 explains the binarization sensitivity over the selection of values of free parameters. Section 3 describes the technical details of our binarization algorithm. Section 4 deals with experimental results and Section 5 describes conclusion.

2 Local Binarization Methods Sensitivity to Selection of Free Parameters Values

Most of local binarization methods have free parameters. Suitable values of these parameters are highly dependent on the context of targeted application and type of document. For achieving high performance on heterogeneous documents manual procedures of parameters values estimation are not suitable. Some techniques have already been proposed for automatic estimation of free parameter values [Rangoni et al., 2009, Badekas and Papamarkos, 2005]. But the concern of their work is to estimate best parameters values which can be fixed for all pixels in the document image. Here, we would like to highlight the problems of using the same free parameters values, either found manually or automatically, for all pixels in a document image.

For demonstration we are using Sauvola binarization method, because it is one of the best among local binarization methods [Sezgin and Sankur, 2004, Badekas and Papamarkos, 2005]. The threshold $t(x, y)$ in Sauvola binarization method is computed using the mean $\mu(x, y)$ and standard deviation $\sigma(x, y)$ of the pixel intensities in a $w \times w$ window centered around the pixel (x, y):

$$t(x, y) = \mu(x, y) \left[1 + k \left(\frac{\sigma(x, y)}{R} - 1 \right) \right] \tag{1}$$

where R is the maximum value of the standard deviation ($R = 128$ for a grayscale document), and k is a parameter which takes positive values. The formula (Equation 1) has been designed in such a way that, the value of the threshold is adapted according to contrast in the local neighborhood of the pixel using local mean $\mu(x, y)$ and local standard deviation $\sigma(x, y)$. Because of this, it tries to estimate an appropriate threshold $t(x, y)$ for each pixel under both possible conditions: high and low contrast. In the case of high contrast region ($\sigma(x, y) \approx R$), the threshold $t(x, y)$ is nearly equal to $\mu(x, y)$. In a quite low contrast region ($\sigma << R$), the threshold goes below the mean value thereby successfully removing the relatively dark regions of the background. The parameter k controls the value of the threshold in the local window such that the higher the value of k, the lower the threshold from the local mean $m(x, y)$.

The statistical constraint in Equation 1 gives acceptable results even for degraded documents. But there is no consensus regarding the appropriate value of k in research community. Badekas et al. [Badekas and Papamarkos, 2005] experimented with different values and found that $k = 0.34$ gives the best results, but Sauvola[Sauvola and Pietikainen, 2000] and Sezgin[Sezgin and Sankur, 2004] proposed $k = 0.5$. This indicate that a suitable value of parameter k should be found experimentally for a target document collection.

We have analyzed Sauvola binarization method with different values of k (with fixed w) and different values of w (with fixed k) for degraded camera-captured document images. Some of the experimental results are shown in the Figure 1 and Figure 2 for different values of k and w respectively. As shown in Figure 1, Sauvola binarization results are sensitive to the selection of appropriate value of k. But Sauvola binarization results are not much sensitive to the value of w, as shown in figure 2. Therefore in this paper we further analyze the sensitivity of k on binarization results.

Additionally, already reported values of k, i.e $k = 0.5$ which is reported by [Sezgin and Sankur, 2004, Sauvola and Pietikainen, 2000] and $k = 0.34$ which is reported by [Badekas and Papamarkos, 2005], do not give acceptable result under blurring or non-uniform illuminations in case of degraded camera captured document images, as shown in Figure 1. However, in our experiment (Figure 1) we have noticed that, small values of k (like $k <= 0.05$) give low noise in the background but produces broken characters. On the other hand, comparatively large values of k (like $k >= 0.2$) give good results for foreground text pixels with unbroken characters but with noise in the background (Figure 1).

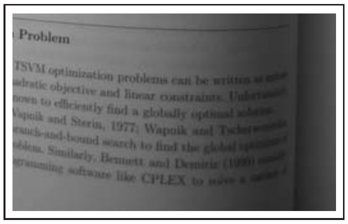
3 Foreground-Background Guided Binarization

These experiments allows us to claim that, Sauvola as well as other local binarization methods can perform better on degraded camera-captured document images if we use two different set of values of free parameters during binarization. For example, in case of Sauvola binarization small value of k is used for pixels roughly belonging to foreground regions and large value of k is used otherwise.

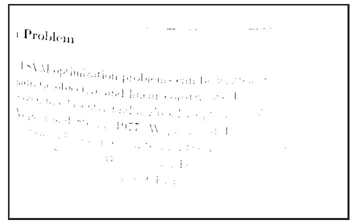
In this paper we modify Sauvola binarization method according to our approach. Our approach can also work with other types of local binarization methods. But we have selected Sauvola binarization because it is best among other types of local binarization methods, as reported by [Sezgin and Sankur, 2004, Badekas and Papamarkos, 2005]. In this section we describe the technical details of our algorithm.

3.1 Foreground Regions Detection

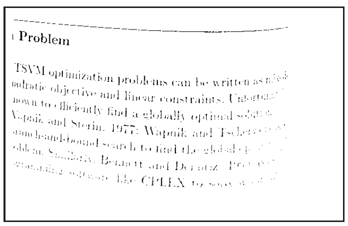
As a first step of our binarization method, we roughly estimate foreground regions from grayscale camera-captured document images. We have already de-



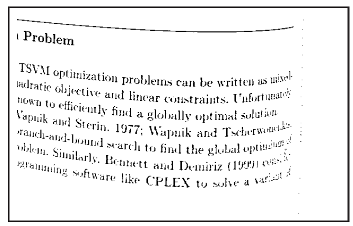
(a) Camera-captured image with non-uniform illumination.



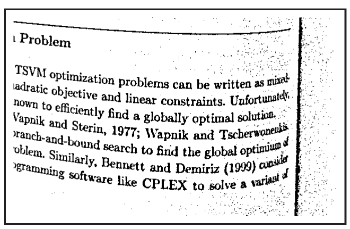
(b) $k = 0.5$.



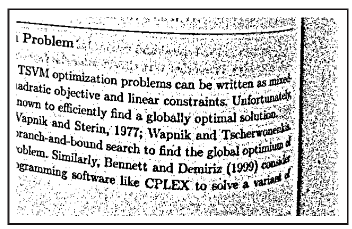
(c) $k = 0.34$.



(d) $k = 0.2$.

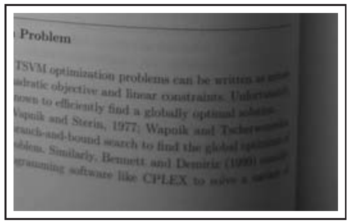


(e) $k = 0.05$.

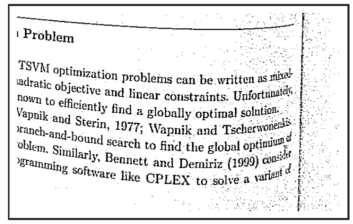


(f) $k = 0.02$.

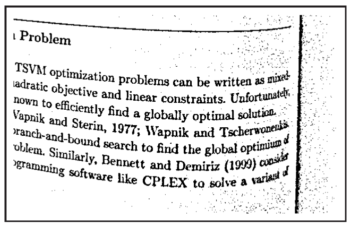
Figure 1: Sauvola binarization results for different values of k, with fixed $w = 15$. $k = 0.5$ is reported by Sauvola[Sauvola and Pietikainen, 2000] and Sezgin[Sezgin and Sankur, 2004]. $k = 0.34$ is used by Badekas et al. [Badekas and Papamarkos, 2005]. We have also added some more values, like $k = 0.2$, $k = 0.05$ and $k = 0.02$. With $k >= 0.2$, results have cleaned-background and broken-foreground-characters. And with $k <= 0.05$ results have uncleaned-background and unbroken-foreground-characters.



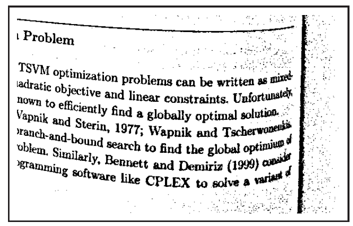
(a) Camera-captured image with blurring.



(b) $w = 7$.



(c) $w = 15$.



(d) $w = 21$.

Figure 2: Sauvola binarization results for different values of w with fixed $k = 0.05$.

scribed (foreground) textline detection techniques for grayscale camera-captured document images using multi-oriented multi-scale anisotropic Gaussian smoothing and ridges detection [Bukhari et al., 2009c, Bukhari et al., 2009d]. Detected ridges represent the central lines structure of foreground objects. In this paper, we use same technique for finding foreground regions. For the completeness of this paper we describe this method [Bukhari et al., 2009c] here.

Foreground regions detection method is divided into two steps: (i) image smoothing using multi-oriented multi-scale anisotropic Gaussian smoothing and then (ii) ridges detection. Following sections discuss these steps in detail.

3.1.1 Image Smoothing

As a first step we need to smooth document image in order to find foreground regions, especially textline regions. Gaussian filter is used for image smoothing. Basic isotropic Gaussian smoothing formula is given in Equation 2, where σ is standard deviation. In document images, textlines are usually horizontal in nature and can be enhanced well by selecting different standard deviations for width (σ_x) and height (σ_y) in Gaussian filter, with σ_x is greater than σ_y. Therefore, anisotropic Gaussian filter (given in Equation 3) is better than isotropic Gaussian filter for document image smoothing or enhancement. Apart from this, camera-captured document images usually contain curled and skewed textlines structure because of geometric and perspective distortions respectively. Therefore we use oriented anisotropic Gaussian filter for camera-captured document image smoothing, given in Equation 4, where σ_x is x-axis standard deviation, σ_y is y-axis standard deviation and θ is the orientation of Gaussian filter. Anisotropic Gaussian smoothing is a slow operation, therefore here we use fast implementation of anisotropic Gaussian filtering proposed by Lampert and Wirjadi [Lampert and Wirjadi, 2006].

$$g(x, y; \sigma) = \frac{1}{2\pi\sigma^2} exp\{-\frac{1}{2}\frac{(x^2 + y^2)}{\sigma^2}\} \tag{2}$$

$$g(x, y; \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\{-\frac{1}{2}(\frac{x^2}{{\sigma_x}^2} + \frac{y^2}{{\sigma_y}^2})\} \tag{3}$$

$$g(x, y; \sigma_x, \sigma_y, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} exp\{-\frac{1}{2}(\frac{(xcos\theta + ysin\theta)^2}{{\sigma_x}^2} + \frac{(-xsin\theta + ycos\theta)^2}{{\sigma_y}^2})\} \tag{4}$$

But a camera-camera document image may contain different directions of curl/skew with different font sizes. Therefore fixed values of σ_x, σ_y and θ for Gaussian smoothing for a complete document image can not produce reasonable enhanced textlines structure. Matched filter bank approach has been used for enhancing the structure of multi-oriented blood vessels [Chaudhuri et al., 1989] and finger prints [Gorman, 1988]. We have described multi-oriented multi-scale anisotropic Gaussian smoothing based on matched filter bank approach for enhancing textlines structure in [Bukhari et al., 2009d, Bukhari et al., 2009c]. In this paper, we use multi-oriented multi-scale anisotropic Gaussian smoothing for enhancing curled textlines structure, where a set of filters is generated from different combinations of σ_x, σ_y and θ. The values of σ_x, σ_y and θ are selected from their predefined ranges. Similar range can be selected for both σ_x and σ_y with some small step size. Similarly, a suitable range for θ is from -45 to 45 degrees with some small step size. Then from these ranges of σ_x, σ_y and θ a set of

Gaussian filters is generated for all possible combinations of σ_x, σ_y and θ. This set of filters is applied to each pixel of grayscale image and then maximum resulting value is selected for resulting smoothed image. Multi-oriented multi-scale anisotropic Gaussian smoothing is not much sensitive to the ranges of σ_x, σ_y and θ. One can select these ranges from reasonably small to large values with small step size, which depends upon the targeted result. For example if one would like to enhance vertically written textlines as well as drawing structures than one should select the ranges for σ_x, σ_y and θ appropriately. Generally large set of filters takes long execution time as compared to small set of filters. In our case we have given more focus to horizontal nature of textlines and chosen following ranges: σ_x from 15 to 30 pixels with step size of 3 pixels, σ_y from 3 to 15 pixels with step size of 3 pixels and θ from -20 to +20 degrees with step size of 5 degrees. Figures 3(a) and 3(b) show the input and smoothed images respectively.

3.1.2 Ridges Detection

Multi-oriented multi-scale anisotropic Gaussian smoothing enhances the foreground structure well, which is clearly visible in Figure 3(b). Now the task is to find the foreground regions information. Ridges detection technique has been used for producing rich description of significant features from smoothed grayscale images [Horn, 1970] and speech-energy representation in time-frequency domain [Riley, 1987]. Ridges detection over a smoothed image can produce central lines structure of foreground textlines/images. In this paper, the Horn-Riley [Horn, 1970, Riley, 1987] based ridges detection approach is used. This approach is based on the information of local direction of gradient and second derivatives as a measure of curvature. From this information, which is calculated by Hessian matrix, ridges are detected by finding the zero-crossing of the appropriate directional derivatives of the smoothed image. Detected Ridges over the smoothed image of Figure 3(b) are shown in Figure 3(c). It is clearly visible in the Figure 3(c) that detected ridges cover the central line structure of foreground objects.

3.2 Foreground-Background Guided Local Binarization

We have already discussed in Section 2 that no single value of parameter k in Sauvola method is suitable for different types of degraded camera-captured documents. But according to our experiment (Figure 1), small values of k (like $k <= 0.05$) give better results for foreground textlines with background noise and comparatively large values of k (like $k >= 0.2$) gives noise free background with broken characters. As shown in Figure 3(c), ridges are present near the foreground pixels. Therefore, instead of using a fixed value of k for all pixels, we use different values of k for foreground and background pixels to improve the binarization result. We redefine Sauvola binarization method, such that:

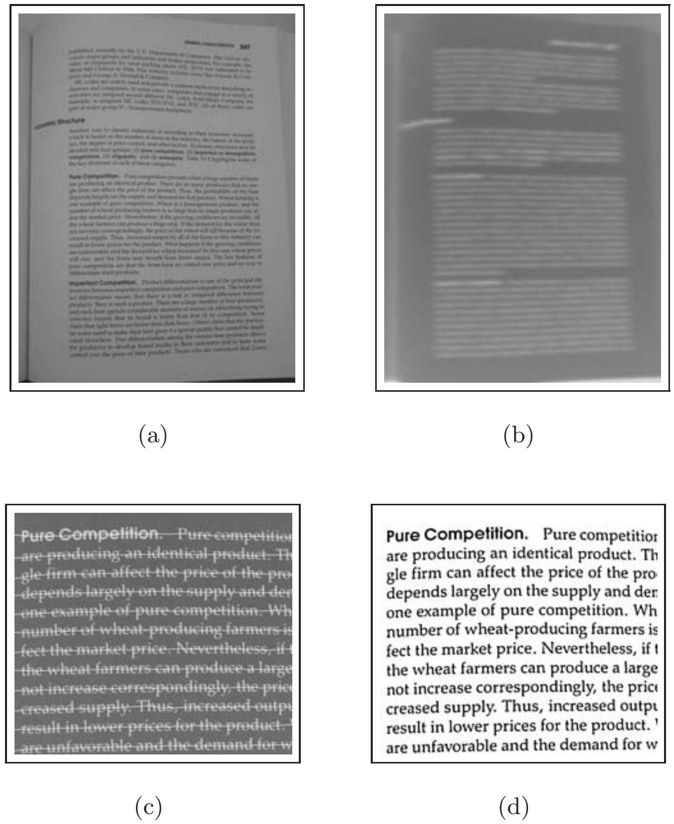
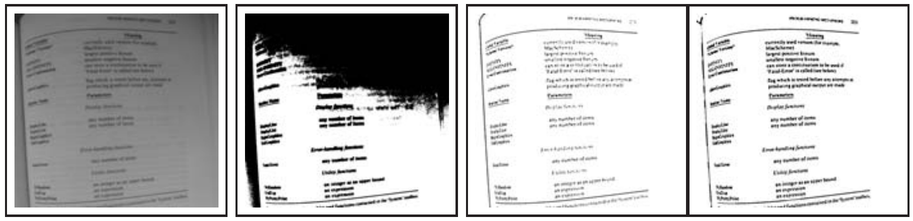


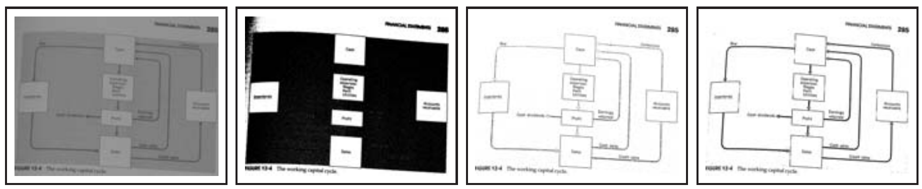
Figure 3: Binarization algorithm snapshots. (a) Input Image, (b) Smoothed Image generated by using match filter bank approach, (c) Horn-Riley method [Horn, 1970, Riley, 1987] is used for detecting ridges, which are visible in the zoom area of document image, (d) Result of foreground/background guided Sauvola binarization (zoomed-in area).

$$t(x,y) = \mu(x,y)\left[1 + k(x,y)\left(\frac{\sigma(x,y)}{R} - 1\right)\right] \tag{5}$$

where k(x,y) is equal to a small value of k if a ridge found in the local neighborhood window, otherwise equal to comparatively large value. After thresholding, median filter can also be applied to further remove the salt and pepper noise. Binarization results based on foreground/background guided Sauvola method are shown in Figures 3(d). The results of Otsu, Sauvola and foreground-background guided Sauvola binarization methods on some example documents images are shown in Figures 4 and 7.



(a) Input (b) Otsu (c) Sauvola (d) Our method



(e) Input (f) Otsu (g) Sauvola (h) Our method

Figure 4: Binarization results of Otsu, Sauvola and our Guided-Binarization. Note that Otsu results have large amount of noise. For Sauvola binarization we have manually selected the appropriate parameter values $w = 15$ and $k = 0.15$ for given dataset (subset of CBDAR-2007). Sauvola results ($w = 15$, $k = 0.15$) have broken-characters for blured images. Our proposed guided binarization method shows better results in the presence of degradations, like blurring.

4 Experiments and Results

We tested the performance of our binarization approach on both low and high resolution degraded camera-captured document images. We conducted two experiments for evaluating our binarization approach:

– for high resolution degraded camera-captured document images we perform **OCR-based evaluation**.
– for low resolution degraded camera-captured images we perform **pixel-based evaluation**.

One can compare the quality of high and low resolution of grayscale camera-captured document images in Figure 5. For these experiments, we have used $k = 0.05$ for pixels near roughly estimated foreground region and $k = 0.2$ otherwise. But we have also shown the robustness of our method over two different values of k for foreground and background regions respectively in pixel-based evaluation Section 4.2.

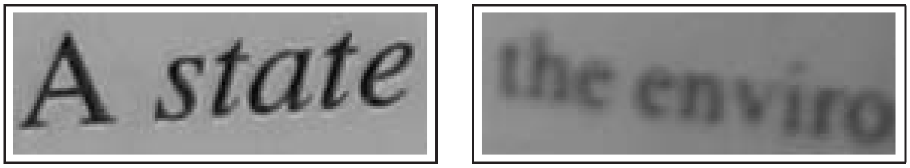


Figure 5: High vs Low Resolution Image Comparison: (left) 6 mega-pixels high resolution camera-captured image and (right) 2 mega-pixels low resolution camera-captured image.

Pixel-based evaluation has been inspired from Document Image Binarization COntest (DIBCO-2009) [Gatos et al., 2009] in which the binarization result of an algorithm is compared with semi-automatically generated binary image ground truth. DIBCO dataset consists of 10 scanned images with distortions like smudge, bleed-through, show-through and shadows. As compared to degraded scanned document images, camera-captured document images contain different types of degradations like non-uniform illumination, blurring, smearing of characters at low resolution and bad-shading. Therefore, we have used our own small datasets of camera-captured document images which are representative of above mentioned degradations for both pixel-based and OCR-based evaluation.

4.1 OCR-based Evaluation

OCR-based evaluation is very important for a comparison of reported algorithm with different state-of-the-art binarization methods. OCR-based evaluation can also be considered as goal-oriented evaluation, because at the end we need better OCR results in most of the document analysis tasks.

Here, we evaluate our binarization approach on hand-held camera-captured document images dataset used in CBDAR 2007 for document image dewarping contest [Shafait and Breuel, 2007] having resolution of 6 mega-pixels. For this purpose, we have selected 10 degraded documents from the dataset. State-of-the-art Otsu and Sauvola binarization methods are used for OCR-based comparative evaluation.

We compare the OCR error rate of all three binarization methods for 10 selected documents. As mentioned earlier in the introduction, after binarization we can not apply OCR engine directly. First, we have to dewarp all binarized images. We have already reported a dewarping method for binarized document images [Bukhari et al., 2009a]. We apply this dewarping algorithm on the results of all three binarization methods. Then dewarped documents of all methods are processed through a commercial OCR system ABBYY Fine Reader 9.0. After obtaining text from the OCR software, the block edit distance[1] with the ASCII

[1] http://sites.google.com/site/ocropus/release-notes

ground-truth has been used as the error measure. Table 1 shows the comparative results of all methods with respect to mean edit distance and the number of documents for each algorithm on which it has the lowest edit distance (in case of tie, all algorithms having the lowest edit distance are scored for that document). It is shown in the Table 1 that our algorithm achieved lowest mean edit distance as well as performed binarization better than other methods on a large number of document images.

Table 1: OCR error rates of different binarization algorithms on subset of dataset of CBDAR 2007 Document Image Dewarping Contest using ABBYY Fine Reader 9.0.

Algorithm	**Mean Edit Distance %**	**Number of documents**[a]
Otsu Binarization	6.96	2
Sauvola Binarization[b]	4.92	3
Guided-Binarization	**4.62**	**5**

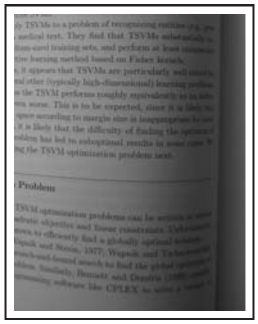
[a] Number of documents for each algorithm on which it has the lowest edit distance.
[b] manually selected: ($w = 15$, $k = 0.15$); tested different values for k in between 0.1 to 0.5 and found 0.15 is the best for the given dataset.

4.2 Pixel-based Accuracy Evaluation

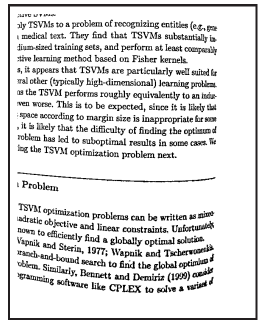
Similar to OCR-based evaluation, here we also compare our algorithm with different state-of-the-art global (Otsu binarization [Otsu, 1979]) and local (Sauvola binarization [Sauvola and Pietikainen, 2000]) binarization methods using pixel-based binarization accuracy. First, we explain about the dataset, ground truth generation and evaluation measure. Then we analyze and compare our binarization results with Otsu and Sauvola binarization results.

4.2.1 Dataset

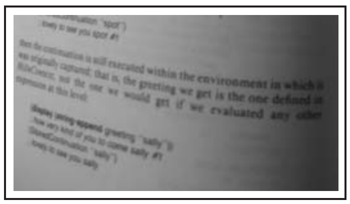
We have selected small portions of degraded text having distortions like bad-shading, non-uniform illumination, blurring and smearing of characters at low resolution from four different camera-captured document images. Furthermore, these document images have been captured at low resolution of 2 mega-pixels as compared to document images captured at high resolution of 6 mega-pixels for OCR-based evaluation (Section 4.1). The dataset and corresponding ground-truth images are shown in Figure 6.



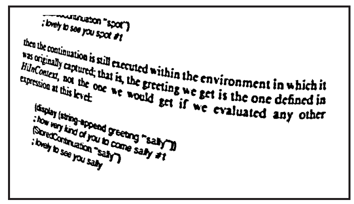
(a) Image-1



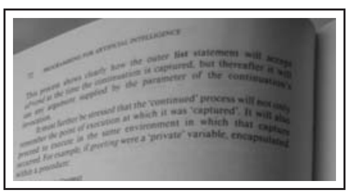
(b) Ground-truth



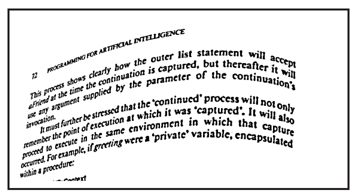
(c) Image-2



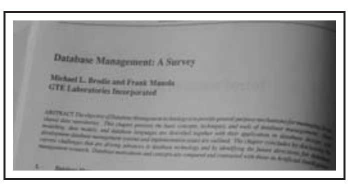
(d) Ground-truth



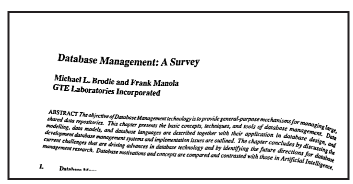
(e) Image-3



(f) Ground-truth



(g) Image-4



(h) Ground-truth

Figure 6: Dataset and Ground-Truth: 4 image-portions have been selected from low resolution (2 mega-pixels) camera-captured images, which contain degradations like, blurring, non-uniform illumination and smearing. Binary ground-truth generatrion process is described in Section 4.2.2

4.2.2 Ground-Truth Generation

We have generated ground-truth binarized images using a semi-automatic process. In this process, we have manually compared different binarized results generated using Sauvola binarization method with different combinations of parameter values of k and w. The results show that for the given dataset $k = 0.02$ and $w = 15$ preserve character strokes at foreground regions better than other values of k and w. However, this combination of k and w produces too much noise in the background regions. Therefore, we have generated binary image ground-truth in two steps: first, we applied Sauvola binarization method with $k = 0.02$ and $w = 15$. Then, we manually removed noise from the background regions. Semi-automatically generated binary ground-truth images are shown in Figure 6 with their corresponding grayscale images.

4.2.3 Evaluation Measure

We use one of the evaluation measure mentioned in [Gatos et al., 2009] for the comparison of different binarization algorithms. The main reason of using only one evaluation measure is to simplify the analysis of different binarization algorithms. Here, we use 'F-measure' [Gatos et al., 2009] for evaluation purpose, which is described below in Equations 6, 7 and 8, where TP, FP, and FN represent the true-positive (total number of matched foreground pixels), false-positive (total number of misclassified foreground pixels in binarization result as compared to ground-truth) and false-negative (total number of misclassified background pixels in binarization result as compared to ground-truth) values respectively.

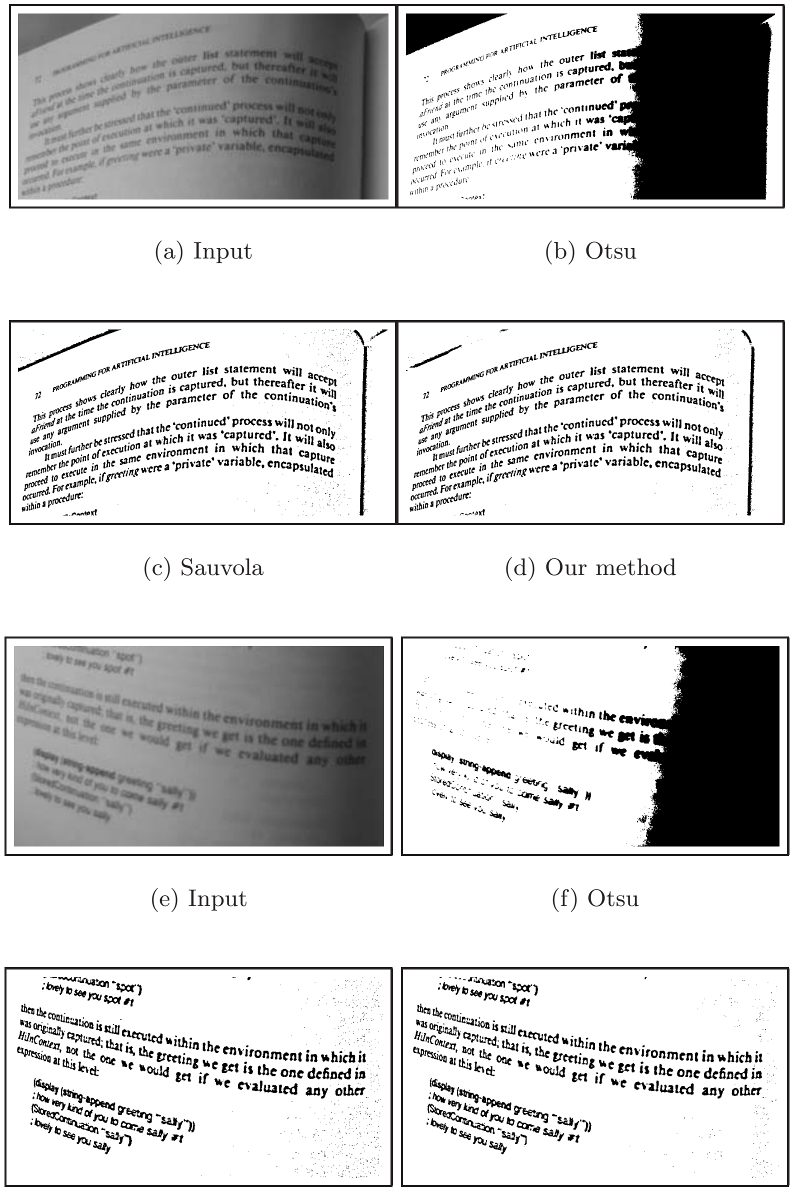
$$\mathrm{FMeasure} = \frac{2 \times \mathrm{Recall} \times \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}} \tag{6}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{7}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{8}$$

4.2.4 Analysis

Based on the above mentioned setup for pixel-based evaluation, comparative results of Otsu binarization, Sauvola binarization and our guided-binarization methods are shown in Table 2. For Sauvola binarization method, we have tested different combinations of k and w and found $k = 0.05$ and $w = 15$ is the best for given dataset. Similarly, for our guided-binarization we fixed $w = 15$ and chose $k = 0.02$ if a ridge is found within the neighborhood region, otherwise $k = 0.2$. It is mentioned in our algorithm (Section 3.2) that we can apply median-filter after binarization. But for pixel-based evaluation we use raw results of our algorithm



(a) Input (b) Otsu

(c) Sauvola (d) Our method

(e) Input (f) Otsu

(g) Sauvola (h) Our method

Figure 7: Binarization results of different algorithms on the low resolution dataset mentioned in Figure 6. For Sauvola and our guided-binarization method in this figure, best parameter values for k have been selected manually which give good compromise between character-strokes and noise.

Table 2: Pixel-based performance evaluation of different binarization methods using low resolution dataset mentioned in Figure 6.

	FMeasure (%)				
	Image-1	Image-2	Image-3	Image-4	Average
Otsu Binarization	32.71	22.41	27.64	54.79	**34.39**
Sauvola Binarization[a]	90.33	89.77	87.82	93.55	**90.37**
Guided-Binarization[b]	90.74	93.10	90.66	92.19	**91.67**

[a]manually selected: ($w = 15$ and $k = 0.05$); tested different values for widow-size and k and found ($w = 15$ and $k = 0.05$) is the best for the given dataset.

[b]manually selected: ($w = 15$ and $k = 0.02$ in the presence of ridge(s) otherwise $k = 0.2$)

to do a fair comparison. Comparative binarization results on some of the images from dataset (Figure 6) are shown in Figure 7.

We have also analyzed the sensitivity of Sauvola binarization method and robustness of our guided-binarization method with respect to the different values of k. We have conducted this experiment on the same dataset mentioned in Figure 6. Figure 8 shows the pixel-based accuracy of Sauvola binarization method for different values of k. Similarly Figure 9 shows the pixel-based accuracy of our guided-binarization method for different values of pair of k. Note that Sauvola uses a single value of k while our guided method uses two values of k i.e. (k_r, k_nr). Therefore, the range of appropriate values of k is much larger for Sauvola's method than for our method. It can be concluded from Figure 8 that Sauvola binarization method is sensitive to the parameter selection for k in the presence of degradations in camera-captured document images, whereas our guided binarization method is robust against parameter selection for the pair (k_r, k_nr).

5 Conclusion

In this paper we have explored the sensitivity of fixing free parameters values for all pixels of a camera-captured document image. We have demonstrated that no matter how to find the free parameters values (either manually or automatically), some range of values of free parameters gives better binarization results for foreground (text-area) document image regions and some other range of values gives better binarization result for background regions. We overcome this sensitivity by introducing the idea of not using the constant values of free parameters for all pixels, but use different values of free parameters for pixels belong to roughly estimated foreground and background regions. For this purpose, we have presented the idea of using multi-oriented multi-scale anisotropic

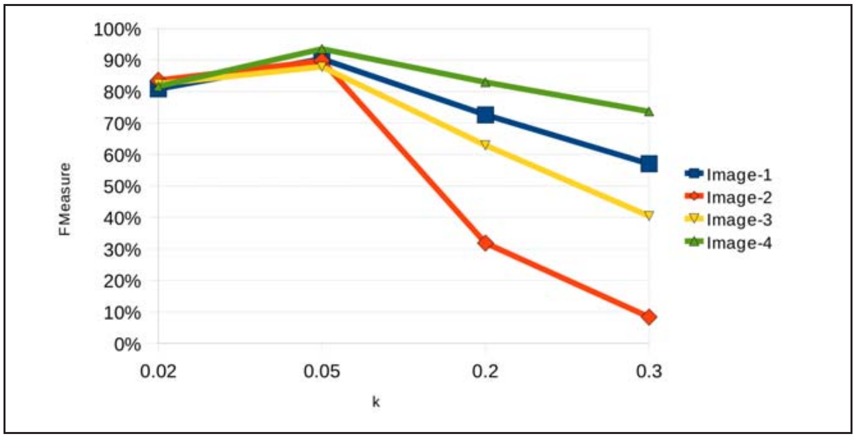


Figure 8: Analysis of the sensitivity of Sauvola binarization method with respect to different values of k over degraded low resolution camera-captured document images shown in Figure 6.

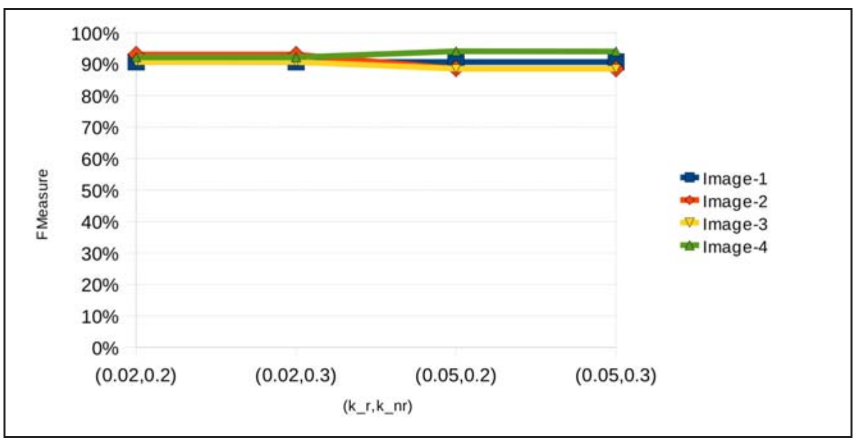


Figure 9: Analysis of the robustness of our guided binarization method with respect to different values of pair of k over degraded low resolution camera-captured document images shown in Figure 6 (Note: k_r: value of k in the presence of ridges; k_nr: value of k in the absence of ridges).

Gaussian smoothing and ridges detection for roughly estimating foreground regions from grayscale document image. Execution time of our method is quite slow as compared to other locally adaptive binarization methods because of the approximation of foreground regions before applying local binarization method. Memory cost is approximately similar to other local binarization methods, like Sauvola. We have performed OCR-based and pixel-based comparative experimental evaluation of our reported method with other state-of-the-art Otsu and Sauvola binarization methods. We have shown an improvement over Sauvola binarization method by selecting different values of parameter k for foreground and background regions respectively. Our idea of foreground-background guided binarization is also adaptable with other types of local binarization methods.

Acknowledgments

This work was partially funded by the BMBF (German Federal Ministry of Education and Research), project PaREn (01 IW 07001).

References

[Badekas et al., 2006] Badekas, E., Nikolaou, N., and Papamarkos, N. (2006). Text binarization in color documents. *International Journal of Imaging Systems and Technology*, 16(6):262–274.

[Badekas and Papamarkos, 2005] Badekas, E. and Papamarkos, N. (2005). Automatic evaluation of document binarization results. In *Proceedings 10th Iberoamerican Congress on Pattern Recognition*, pages 1005–1014, Havana, Cuba.

[Bernsen, 1986] Bernsen, J. (1986). Dynamic thresholding of gray level images. In *Proceedings 8th International Conference on Pattern Recognition*, pages 1251–1255.

[Brown and Seales, 2004] Brown, M. S. and Seales, W. B. (2004). Image restoration of arbitrarily warped documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1295–1306.

[Bukhari et al., 2009a] Bukhari, S. S., Shafait, F., and Breuel, T. M. (2009a). Dewarping of document images using coupled-snakes. In *Proceedings of First International Workshop on Camera-Based Document Analysis and Recognition*, pages 34–41, Barcelona, Spain.

[Bukhari et al., 2009b] Bukhari, S. S., Shafait, F., and Breuel, T. M. (2009b). Foreground-background regions guided binarization of camera-captured document images. In *Proceedings of First International Workshop on Camera-Based Document Analysis and Recognition*, pages 18–25, Barcelona, Spain.

[Bukhari et al., 2009c] Bukhari, S. S., Shafait, F., and Breuel, T. M. (2009c). Ridges based curled textline region detection from grayscale camera-captured document images. In *Proc. The 13th International Conference on Computer Analysis of Images and Patterns*, volume 5702 of *Lecture Notes in Computer Science*, pages 173–180, Muenster, Germany.

[Bukhari et al., 2009d] Bukhari, S. S., Shafait, F., and Breuel, T. M. (2009d). Script-independent handwritten textlines segmentation using active contours. In *Proceedings 10th International Conference on Document Analysis and Recognition*, pages 446–450, Barcelona, Spain.

[Cao et al., 2003] Cao, H., Ding, X., and Liu, C. (2003). Rectifying the bound document image captured by the camera: a model based approach. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, pages 71–75, Edinburgh, Scotland.

[Cattoni et al., 1998] Cattoni, R., Coianiz, T., Messelodi, S., and Modena, C. M. (1998). Geometric layout analysis techniques for document image understanding: a review. Technical report, IRST, Trento, Italy.

[Chaudhuri et al., 1989] Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., and Goldbaum, M. (1989). Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transaction on Medical Imaging*, 8(3):263–269.

[Fu et al., 2007] Fu, B., Wu, M., Li, R., Li, W., and Xu, Z. (2007). A model-based book dewarping method using text line detection. In *Proceedings 2nd International Workshop on Camera Based Document Analysis and Recognition*, pages 63–70, Curitiba, Barazil.

[Gatos et al., 2009] Gatos, B., Ntirogiannis, K., and Pratikakis, I. (2009). Coupled snakelet model for curled textline segmentation of camera-captured document images. In *Proceedings 10th International Conference on Document Analysis and Recognition*, Barcelona, Spain.

[Gatos et al., 2007] Gatos, B., Pratikakis, I., and Ntirogiannis, K. (2007). Segmentation based recovery of arbitrarily warped document images. In *Proceedings 9th International Conference on Document Analysis and Recognition*, pages 989–993, Curitiba, Brazi.

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

[Gorman, 1988] Gorman, L. O. (1988). Matched filter design for fingerprint image enhancement. In *Proceedings International Conference on Acoustics, Speech, and Signal Processing*, pages 916–919, New York, NY, USA.

[Horn, 1970] Horn, B. K. P. (1970). Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. *PhD Thesis, MIT*.

[Kim, 2004] Kim, I.-J. (2004). Multi-window binarization of camera image for document recognition. In *Proceedings 9th International Workshop on Frontiers in Handwriting Recognition*, pages 323–327, Washington, DC, USA.

[Lampert and Wirjadi, 2006] Lampert, C. and Wirjadi, O. (2006). An optimal nonorthogonal separation of the anisotropic gaussian convolution filter. *IEEE Transactions on Image Processing*, 15(11):3501–3513.

[Liang et al., 2005] Liang, J., Doermann, D., and Li, H. (2005). Camera-based analysis of text and documents: a survey. *International Journal of Document Analysis and Recognition*, 7(2-3):84–104.

[Lu and Tan, 2007] Lu, S. and Tan, C. L. (2007). Thresholding of badly illuminated document images through photometric correction. In *Proceedings 2007 ACM symposium on Document engineering*, pages 3–8, Winnipeg, Manitoba, Canada.

[Lu et al., 2005] Lu, S. J., Chen, B. M., and Ko, C. C. (2005). Perspective rectification of document images using fuzzy set and morphological operations. *Image and Vision Computing*, 23:541–553.

[Lu and Tan, 2006] Lu, S. J. and Tan, C. L. (2006). The restoration of camera documents through image segmentation. In *Proceedings 7th IAPR workshop on Document Analysis Systems*, pages 484–495, Nelson, New Zealand.

[Mori et al., 1992] Mori, S., Suen, C., and Yamamoto, K. (1992). Historical review of OCR research and development. *Proceedings of the IEEE*, 80(7):1029–1058.

[Niblack, 1986] Niblack, W. (1986). *An Introduction to Image Processing*. Prentice-Hall, Englewood Cliffs, NJ.

[O'Gorman, 1994] O'Gorman, L. (1994). Binarization and multithresholding of document images using connectivity. *Graphical Model and Image Processing*, 56(6):494–506.

[Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions Systems, Man and Cybernetics*, 9(1):62–66.

[Rangoni et al., 2009] Rangoni, Y., Shafait, F., and Breuel, T. M. (2009). OCR based thresholding. In *Proceedings IAPR Conference on Machin Vision Applications*, Yokohama, Japan.

[Riley, 1987] Riley, M. D. (1987). Time-frequency representation for speech signals. *PhD Thesis, MIT*.

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

[Sezgin and Sankur, 2004] Sezgin, M. and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–165.

[Shafait et al., 2008a] Shafait, F., Beusekom, J. V., Keysers, D., and Breuel, T. M. (2008a). Document cleanup using page frame detection. *Int. Jour. on Document Analysis and Recognition*, 11(2):81–96.

[Shafait et al., 2008b] Shafait, F., Beusekom, J. V., Keysers, D., and Breuel, T. M. (2008b). Structural mixtures for statistical layout analysis. In *Proceedings 8th International Workshop on Document Analysis Systems*, pages 415–422, Nara, Japan.

[Shafait and Breuel, 2007] Shafait, F. and Breuel, T. M. (2007). Document image dewarping contest. In *Proceedings 2nd International Workshop on Camera Based Document Analysis and Recognition*, pages 181–188, Curitiba, Brazil.

[Shafait et al., 2008c] Shafait, F., Keysers, D., and Breuel, T. M. (2008c). Efficient implementation of local adaptive thresholding techniques using integral images. In *Proceedings 15th International Conference on Document Recognition and Retrieval*, volume 6815 of *SPIE Electronic Imaging*, page 81510, San Jose, CA, USA.

[Shafait et al., 2008d] Shafait, F., Keysers, D., and Breuel, T. M. (2008d). Performance evaluation and benchmarking of six page segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):941–954.

[Sobottka et al., 2000] Sobottka, K., Kronenberg, H., Perroud, T., and Bunke, H. (2000). Text extraction from colored book and journal covers. *International Journal on Document Analysis and Recognition*, 2(4):163–176.

[Tan et al., 2006] Tan, C. L., Zhang, L., Zhang, Z., and Xia, T. (2006). Restoring warped document images through 3d shape modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):195–208.

[Tsai and Lee, 2002] Tsai, C. and Lee, H. (2002). Binarization of color document images via luminance and saturation color features. *IEEE Transactions on Image Processing*, 11(4):434–451.

[Ulges et al., 2005] Ulges, A., Lampert, C. H., and Breuel, T. M. (2005). Document image dewarping using robust estimation of curled text lines. In *Proceedings 8th International Conference on Document Analysis and Recognition*, pages 1001–1005, Seoul, Korea.

[White and Rohrer, 1983] White, J. M. and Rohrer, G. D. (1983). Image thresholding for optical character recognition and other applications requiring character image extraction. *IBM Journal of Research and Development*, 27(4):400–411.

[Zhang and Tan, 2003] Zhang, Z. and Tan, C. L. (2003). Correcting document image warping based on regression of curved text lines. In *Proceedings 7th International Conference on Document Analysis and Recognition*, pages 589–593, Edinburgh, Scotland.