Extreme Value Theory Based Text Binarization In Documents and Natural Scenes

Basura Fernando
Erasmus Mundus CIMET Master
University Jean Monnet
Saint Etienne, France
basuraf@gmail.com

Sezer Karaoglu
Erasmus Mundus CIMET Master
University Jean Monnet
Saint Etienne, France
sezerkaraolu@yahoo.com

Alain Trémeau
Laboratoire Hubert Curien, University Jean Monnet
Saint Etienne, France
alain.tremeau@univ-st-etienne.fr

***Abstract*—This paper presents a novel image binarization method that can deal with degradations such as shadows, non-uniform illumination, low-contrast, large signal-dependent noise, smear and strain. A pre-processing procedure based on morphological operations is first applied to suppress light/dark structures connected to image border. A novel binarization concept based on difference of gamma functions is presented. Next Generalized Extreme Value Distribution (GEVD) is used to find proper threshold for binarization with a significance level. Proposed method emphasizes on region of interest (with the help of morphological operations) and generates less noisy artifacts (due to GEVD). It is much simpler than other methods and works better on degraded documents and natural scene images.**

Keywords-Generalized extreme value distribution; Geodesic transform morphological reconstruction; Connected opening; Text binarization

I. INTRODUCTION

The problem of text segmentation in still images is a hard problem due to large variability of appearance of texts (font style, size), complex background, occlusions, object shadows, highlights from shiny object parts, and differences of color brightness of objects. The problem of textual image segmentation can be split into several steps, the first step consists in image binarization, and it is a crucial step. A lot of image binarization techniques [1-5], [20] have been developed by many authors.

Existing methodologies for image binarization are broadly divided under two main strategies: thresholding based, and grouping based. Thresholding based methods use global or local threshold(s) to separate text from background (e.g. see [3]). Commonly used methods are histogram based thresholding and adaptive thresholding. When the text to be detected is well contrasted with the background most of the existing algorithms work well, however these latter fail when there is no sufficient distinction between background and text. Adaptive or local binarization methods use several thresholds for each study areas of the images instead of one. The most widely used adaptive thresholding algorithms are Niblack's [15] and Sauvola [16]. These methods are more robust against uneven illumination and varying colors than global ones but suffer regarding to dependency of parametric values. Entropy based methods use the entropy of the grayscale image in order to threshold images using probability distribution of intensity values [19]. Trier and Taxt presented an evaluation of binarization methods for document images in [3].

Region based grouping methods are mainly based on spatial-domain region growing, or on splitting and merging (e.g. see [6]). They are commonly used in the field of image segmentation but these techniques are in general not well adapted to segment features such as text. To get more efficient results these methods are generally combined with scale-space approaches based on top-down cascades (high resolution to low resolution) or bottom-up cascades (low resolution to high resolution). The problem of these methods is that they depend on several parameters such as seed values; homogeneity criterion (i.e. threshold values) and initial step (i.e. start point). They are therefore not versatile and cannot produce robust results for complex urban scenes. In addition, in terms of computational time, region based grouping methods are not efficient. However, they use spatial information which groups text pixels efficiently.

Clustering based grouping methods are based on classification of intensity or color values in function of a homogeneity criterion (e.g. see [7-9]). Two main categories of clustering algorithms are histogram based and density based. Multi dimensional histogram thresholding can be used to pre-segment color images from the probability distribution of colors but 3-D histogram must be computed. Based on our experience former methods are not well-adapted for complex background images such as urban scenes. Invariance against varying color properties is the biggest advantage of these methods.

K-means algorithm had been among the main techniques used for clustering based grouping until recently. But this algorithm is not the most efficient one. Thus, Lukac et al. have proved with the ICDAR 2003 competition that the fuzzy-cmeans algorithm gives better results [10]. Recently, several studies have also shown that the Mean-Shift algorithm based density estimation [11] outperforms K-means algorithm. That is, the K-means algorithm is commonly considered as a simple way to classify color pixels through a priori fixed number of clusters. The main

ISBN: 978-1-4244-8889-6 © ICMV 2011

idea is to define k centroids, next to perform the process till all pixels belong to a cluster whose centroid is the nearest one.

Even if many approaches have been specifically developed for image binarization for document images most of these approaches fail when image is complex such as in natural scene images. The aim of this work is to develop a general threshold technique and to demonstrate need for such a new technique in field of document and natural scene image analysis. The main objective of our approach is to reduce noise in threshold images while keeping textual information as much as possible using substantially lesser complex processes than other well-known approaches. The noise removal is essential for later processes after binarization such as Optical Character Recognition (OCR); dealing with less number of letter candidates saves a lot of time at learning steps. Several authors used many image filtering and image enhancement techniques prior to the binarization process. As example, Wang et al. proposed in [13] to use an anisotropic filter to increase the robustness of the clustering step. Lim et al. proposed in [12] to use tensor voting and adaptive median filter iteratively to remove noise before text segmentation. B. Gatos et al. proposed in [21] to use Wiener filter as a low pass filter to reduce effects of noisy areas and smooth background during image acquisition. Nobuo Ezaki et al. in [22] proposed to use modified top hat processing to able to cope with small letters in their natural scene text detection methodology.

In this paper we propose a methodology which is robust against shadows, highlights, specular reflection, non-uniform illumination, complex background, varying text size, colors and styles.

In the proposed method first a geodesic transform based morphological reconstruction technique is used to remove objects connected to the borders and to emphasize on objects in center of the scene. After that a method based on difference of gamma functions approximated by Generalized Extreme Value Distribution (GEVD) is used to find a correct threshold for binarization. The main function of this GEVD is to find the optimum threshold value for image binarization relatively to a significance level. The significance levels can be optimized using relative background complexity of the image.

The contribution of this paper is a new binarization algorithm that use morphological connected opening based preprocessing to reduce illumination variations prior to binarization and introduction of generalized extreme value distribution to find thresholds to binarize an image. We also present a new concept of difference of gamma functions to emphasize certain regions of intensity distribution.

The remaining parts of the paper are organized as follows. The novel thresholding algorithm is presented in section II. Next, experimental results are given in section III. Lastly a conclusion is drawn in section IV.

II. PROPOSED METHODOLOGY

First, image enhancement method based on morphological reconstruction through geodesic transform is applied on the gray scale image. This step is used to remove objects connected to borders and lighter than their surroundings to emphasize on lighter objects than their surroundings in center of the scene. The rationality behind this step is that in a given document/natural-scene image, the information to be gathered should be within the image not in the border regions of the image. In this context we consider noise as any non textual regions except the background. When regions lighter than their surrounding and connected to image borders are removed, most of the noise that present in the image is removed. This operation makes it easy to deal with textual candidates which create less noisy artifacts during later processing. The intensity level is used to gather information about possible text candidates. Texts reveal useful information in documents/natural-scene images. People always give priority to text where text can take attention which results in textual regions to be more salient in the image. In these textual images the visual attention is provided by contrast issue. The text regions always contrast with their background. Nature of contrast let us build robust binarization algorithm for different lighting conditions. For instance, consider a text region lighter than its surrounding and the same region under shadow or highlight. Under different lighting conditions it will not change the fact that text region is lighter than its surrounding. This property helps us to extract textual regions even under different lighting conditions. In this paper we present a binarization algorithm which is robust to varying lighting conditions. After this preprocessing step presented in II.A, objects in the region of interests has higher intensity value compared to the background hence improves the binarization which is explained in section (II. b).

A. *Morphological reconstruction through geodesic transform*

According to Soille (see [26]) geodesic dilation of a bounded image always converges after a finite number of iterations (i.e. until the proliferation or shrinking of the marker image is totally impeded by the mask image). For this reason geodesic dilation is considered as a powerful morphological reconstruction scheme. The reconstruction by dilation $R_g^{\partial}(f)$ of a mask image (g) form a marker image (f) is defined as the geodesic dilation of (f) with respect to (g) iterated until stability as follows (see Fig. 1):

$$R_g^{\partial}(f) = \partial_g^{(i)}(f) \tag{1}$$

The stability is reached at the iteration i when: $\partial_g^{(i)}(f) = \partial_g^{(i+1)}(f)$. This reconstruction is constrained by the following conditions that both (f) and (g) images must have the same definition domain (i.e.$D_f = D_g$) and $f \leq g$.

(a) 1-D marker signal f and mask signal g

(b) Reconstruction by erosion R^{∂} of g with respect to f

(A) Algebraic Opening

Figure 1. Algebraic opening for a 1-D signal.

This reconstruction transform presents several properties: it is increasing ($g_1 \leq g_2 \Rightarrow R^{\partial}_{g1}(f) \leq R^{\partial}_{g2}(f)$), anti-extensive ($R^{\partial}_{g}(f) \leq g$), and idem-potent $R^{\partial}_{g}(R^{\partial}_{g}(f)) = R^{\partial}_{g}(f)$). This reconstruction transform corresponds to an algebraic closing of the mask image. The connected opening transformation, $\gamma_x(g)$ of a mask image (g) can be defined as:

$$\gamma_x(g) = R^{\partial}_{g}(f_x) \qquad (2)$$

Where the marker image f_x equals to zero everywhere except as x which has a value equal to that of the image g at the same position. According to Soille (see [26]) the connected opening transformation can be used to extract connected image objects having higher intensity values than their surrounding when we chose the mask image zero everywhere, except for the point **x** which has a value equal to that of the image g at the same position (see Fig. 2).

(a) Original image (b) Connected opening

Figure 2. Connected opening visual sample.

In order to suppress lighter objects than their surroundings and connected to border of the image, we choose the marker image zero everywhere except the border of the image. At the border of the image we chose the pixel value of marker the same as mask pixel value at the same position. Once we get the connectivity information with the help of morphological reconstruction based on geodesic transform, we suppress these lighter objects connected to image border. After this preprocess step most of the non text regions are reduced and kept only most probable text candidates which leads us to emphasize more on region of interest of the image (See Fig. 2.b). Especially we have seen that this process reduce the background intensity variations and enhance the text regions of the image.

By this way the image is enhanced before being analyzed by binarization step. After this step of image enhancement the binarization algorithm based on difference of gamma function approximated by GEVD is applied. The next section explains this algorithm.

B. Difference of gamma for background estimation

Different image enhancement algorithms can be used to improve the appearance of an image such as its contrast in order to make the image interpretation, understanding, and analysis easier. Various contrast enhancement algorithms have been developed to modify the appearance of images by highlighting certain features while suppressing others. A widely used approach for contrast enhancement is based on the use of a power law response equation such as follows (see Fig. 3):

$$s = cr^{\gamma} \qquad (3)$$

Generally c and γ are positive constants; r, s are the input, output intensity levels respectively (see [27]). (3) is widely known as gamma contrast enhancement function.

In the proposed method, two corresponding gamma contrast enhancement functions are defined as follows:

$$g_1(r) = c_1 r^{\gamma_1}, \quad g_2(r) = c_2 r^{\gamma_2} \qquad (4)$$

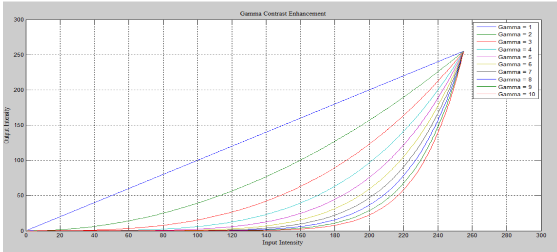


Figure 3. Influence of the parameter gamma on the contrast of the output image.

Here r is the intensity level of the input image, M is the maximum intensity value (i.e. $0 \leq r \leq M$, Ex. For 8-bit image $M = 255$) and $c = M^{(1-\gamma)}$ and gamma values $\gamma 1, \gamma 2$ ($\gamma 1 < \gamma 2$).

These two contrast enhancement functions defined in (4) can be applied to image $f(x, y)$ to obtain two enhanced images $f_1(x, y)$ and $f_2(x, y)$. Then the difference of gamma functions $diff_{f1,f2}(x, y)$ is given by (5) as (see Fig. 3):

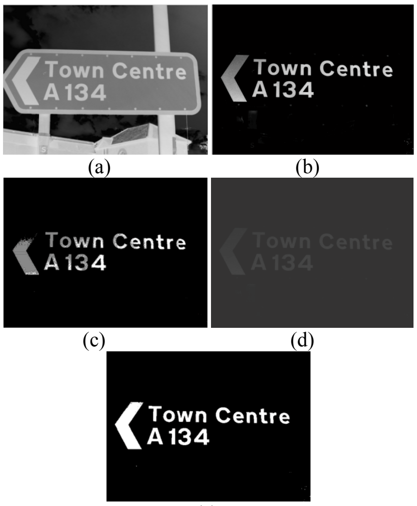
$$diff_{f1,f2}(x,y) = |f_1(x,y) - f_2(x,y)| \qquad (5)$$

Next, in order to classify pixels belonging to the foreground or to the background (see Fig. 4) we propose to apply the following rule on the image corresponding to the difference of gamma functions.

$$\forall (x,y) \in f(x,y) \text{ if } diff_{f1,f2}(x,y) > T \Longrightarrow (x,y) \in \text{foreground}$$

$$\text{otherwise } (x,y) \in \text{background}$$

We apply above rule because we know that the enhanced image from previous step consists of middle level pixels as text regions and low level pixels as background regions. As it can be seen in Fig. 3, different gamma functions suppress different intensity ranges. As in Fig. 5.a and Fig. 5.b different gamma values yields different suppression ranges for (5).



(a) (b) (c) (d) (e)

Figure 4. (a) The original Image (b) Gamma correction γ=2 applied to connected opening enhanced image (c) Gamma correction γ=4 applied to connected opening enhanced image (d) Difference between gamma corrected images (contranst enhanced by 20%) (e) Thresholded

In order to better classify pixels belonging to the foreground from pixels belonging to the background, we propose to apply the following process which try to compute the optimum values for γ_1, γ_2 and T. Knowing that the background is either darker or lighter than the surround there is always a contrast issue between them. When background is lighter we have to deal with the inverse of the image. When $\gamma_1 < \gamma_2$ the second gamma corrected function $f_2(x, y)$ suppresses the image background intensities more than $f_1(x, y)$. For example in Fig. 3 compare when $\gamma = 3$ and $\gamma = 10$. As a result $f_{2\ (\gamma = 10)}(x, y)$ appears more contrasted than $f_{1\ (\gamma = 3)}(x, y)$. Both $f_1(x, y)$ and $f_2(x, y)$ suppress the background. Now we compute the difference of gamma function $\mathrm{diff}_{f1,f2}(x,y)$. Unlike other binarization techniques which generate some noise artifacts especially in relatively homogeneous areas such as the background, when we take the difference between two corrected images (i.e. $\mathrm{diff}_{f1,f2}(x,y)$ since both images remove the background substantially we do not generate noisy artifacts in the background. Image generated by the difference of gamma function has the desirable property of emphasizing on middle range intensity values while suppressing the lower and higher intensities (see Fig. 5). By thresholding the resulting image (by a value very close to zero, as shown in Fig. 4(e)) we obtain a perfect separation of foreground and background. As mentioned earlier, different gamma values for γ_1 and γ_2 yields different suppression ranges. Depending on γ_1, γ_2 and the threshold T we will obtain different binarization outputs. Now the challenge is to find appropriate (optimized) values for γ_1 and γ_2 and T. Fig. 5 shows that the suppression of some values of intensity depends on the value of γ. We can rewrite the difference of gamma functions as follows:

$$\Delta f_{\gamma_1, \gamma_2}(x) = M^{(1-\gamma_1)} x^{\gamma_1} - M^{(1-\gamma_2)} x^{\gamma_2} \quad (6)$$

To illustrate the idea of selecting proper γ_1, γ_2 and thresholds consider the examples given in Fig. 5. For instance as it can be seen from Fig. 5, $\Delta f_{2,4}$ has a lower suppression range compared to $\Delta f_{9,10}$. Let us consider an arbitrary threshold corresponding to output intensity level of 2, then for $\Delta f_{2,4}$ the suppression range concerns input intensity values less than 10, meanwhile for $\Delta f_{9,10}$ the suppression range concerns input values lower 100. In other words if we use $\Delta f_{9,10}$ function with $T = 2$ on a particular image for binarization then the corresponding global binarization threshold is 100. If we use $\Delta f_{2,4}$ with $T = 2$ the corresponding global threshold for binarization is 10. We are interested in finding proper γ_1, γ_2 and T values to binarizes the image.

As discussed in the introduction, the main problem of text extraction is to find correct thresholds to remove background in order to separate textual visual information from background. As pointed out in the current section, the difference of gamma functions with proper gamma values and thresholds can achieves binarization with good properties such as less noisy artifacts. It is clearly seen from Fig. 5 that depending on the gamma values, difference of gamma suppression range will vary. Here the problem occurs how to arrange appropriate gamma values because we do not know the pixel distribution of each different image. Due to the fact that there is no scientific method to find corresponding gamma values that perfectly binarizes the image, we suggest looking at the problem from a different perspective. During our experiments we observed that most of the significant visual information in textual images resides in the middle of the distribution of pixel intensities. When we look at the pixel distribution of gamma difference of images with varying gamma values, they are identical. Even though the intervals vary they keep the identity of the shape (see Fig. 5). To solve this issue we propose to compute image statistics from a dataset of text images and to use these statistics to model this distribution.

Extreme value theory is a well-known statistical tool that deals with extreme events. This theory is based on the assumption that three types of distributions are needed to model the maximum or minimum of a collection of random observations from a unique distribution. These three distributions are called Gumbel, Fréchet, and Weibull distributions [23]. We propose to use the generalized extreme value distribution model [23] to find the best thresholds (i.e. the optimized ones) for our problem.

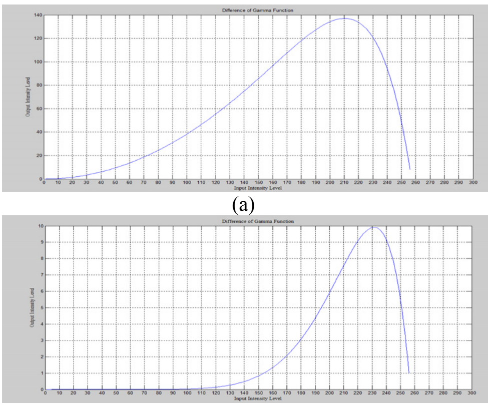
Generalized extreme value distribution can be written as:

For $k \neq 0$

$$f(x) = \left\{ \frac{1}{\sigma} \exp^{\left(-(1+kz)^{-1/k}\right)} . (1+kz)^{-1-(1/k)} \right. \quad (7)$$

For $k = 0$

$$f(x) = \left\{ \frac{1}{\sigma} \exp^{\left(-z-\exp(-z)\right)} \right. \quad (8)$$



(a)

(b)

Figure 5. (a) Difference of gamma functions ($\gamma1= 2$, $\gamma2=4$)
(b) Difference of gamma functions ($\gamma1= 9$, $\gamma2=10$)

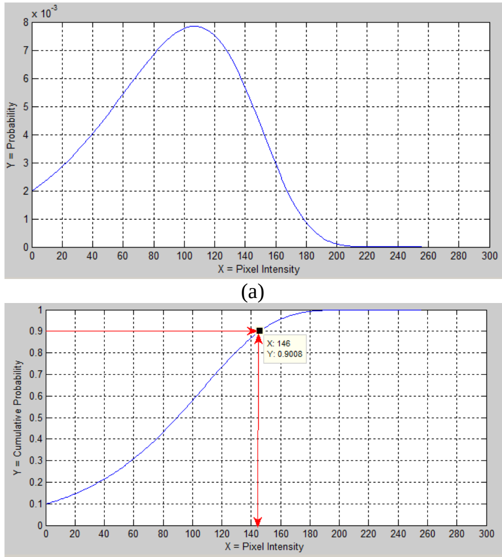
Where $z = \frac{x-\mu}{\sigma}$, x is the variable under study (e.g. the intensity), k is a shape parameter which is 1 for our case (Gumbel), σ is a scale parameter and μ is a location parameter.

We propose to use the maximum likelihood estimation (MLE) method to estimate the function f(x). To find parameters of the GEVD using MLE we used the method proposed by Lawless in [24]. Prescott in [31] proposed a new method for parameter estimation. Pickands in [17] showed that, if X is a random variable and F(x) is its probability distribution function (PDF), then under certain conditions, $F(x|u) = P(X \leq u + x \mid X > u)$ can be approximated by a Generalized Pareto Distribution (GPD) [25]. In other words GPD can be used to find the thresholds of an identical distribution. Let $X = \{X_1, X_2, X_3, \dots X_n\}$ be independent random variables with identical distribution F. Next, suppose that $D_n = \max(X)$ then it can be shown that for a large n:

$$P(D_n < x) \approx f(x) \qquad (9)$$

Here f(x) is the generalized extreme value distribution (GEVD). Therefore, u is the threshold over which these observations {X} exceeds, can be modeled by GPD. This shows that u can be found with a significance level if GPD is known. As the theory suggests GEV is an excellent tool to deal with the thresholding problem in image binarization.

Extreme value theory has been used in many fields such as engineering, oceanography, environment, actuarial sciences and economics, among others. In such areas, the main problem is the scarcity of data or, more specifically, modeling with a fairly small amount of observations. We propose here to use the cumulative distributions function (CDF) of the GEV to define the significance levels which best describe the distributions studied. Next we use these significance levels to find proper thresholds for binarization. Our experiments suggested that a significance level of 10% is sufficient to detect simple backgrounds; (see Fig. 6, 7) meanwhile a significance level of 35-40% is necessary to detect complex backgrounds and scenes (see Fig. 8.a and Fig. 8.b).



(a)

(b)

Figure6. (a) PDF of the GEVD of the image of Fig. 7(a) ,
(b) CDF of the GEVD of the image of Fig. 7(a)

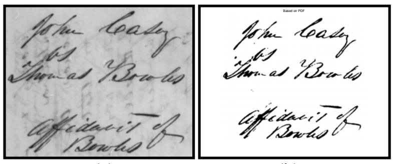
To remove both background and the over exposed regions we have therefore to use a confidence interval. Here we assume that foreground intensities lie in certain range:

$$\Pr(U_{t1} < X < U_{t2}) \qquad (10)$$

Here U_{t1} is the lower threshold and the U_{t2} the upper threshold. To find t_1 and t_2 cumulative probability of P_1 for lower threshold and P_2 for higher threshold can be selected depending on the statistical desired significance level. According to our experiments based on 500 images from ICDAR2003 dataset, $P_1 = 0.7$ and $P_2 = 0.99$ in GEV cumulative probabilities are sufficient to remove overexposed regions and background. (As example, see Fig. 8.c and Fig. 8.d). During the binarization step we found appropriate significance levels to find the proper thresholds experimentally and our results suggest that these significance values generalize to both ICDAR2003[29] dataset and DIBCO[30] datasets.

So in summary first connected opening transformation is applied on the gray scale image to emphasize on the region of interest in the image. Then image statistics are collected and by use of maximum likelihood estimation as proposed in [24] generalized extreme value distribution is computed. Then certain significance levels are used to find the corresponding threshold as explained by the (9) and (10). These corresponding thresholds are used to binarize the image.

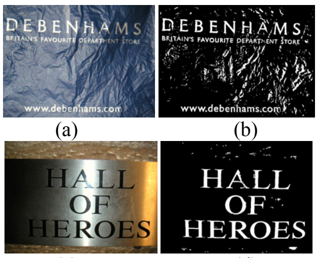
III. EXPERIMENTAL RESULTS



(a) (b)

Figure 7. (a) Input Image with simple background (b) Threshold image (Significance level of 10% corresponds here to an intensity value of 137).

Fig. 7, 9, 10, 11 and Table 1 illustrate the experimental results that we get with the DIBCO2009 dataset [30]. The main interest of the DIBCO2009 dataset is that we know the ground truth of the binarization of each image with evaluation performance measures. Most of the images belonging to this database are not overexposed or subjected to shadows, but their background is moderately complex (see Fig. 9). Consequently, we have used a significance level of 10%. Precision (PR), recall (RC), F-measure (FM) and peak signal to noise ratio (PSNR) values have been computed for each image of the DIBCO2009 dataset. These values have been computed from the ground truths provided by DIBCO2009. During these experiments, most common parameter values have been used for Niblack [15], Sauvola [16] and Otsu [18] algorithms to analyze the performance of their binarization method.



(a) (b)

(c) (d)

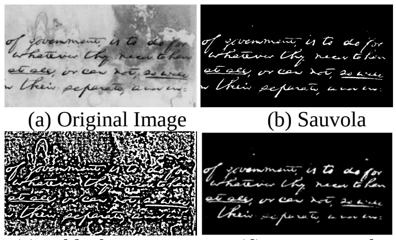
Figure 8. (a) Complex background input image (b) Threshold image (c) Input image with over exposed region (d) Threshold image

(a) Original Image (b) Sauvola

(c) Niblack (d) Our approach

Figure 9. Output image corresponding to complex image belonging to the DIBCO2009 dataset [30]

The DIBCO2009 results can be found in [14]. All performance calculations for DIBCO dataset have been computed according to the definitions provided by DIBCO2009 competition. Comparison of results for DIBCO dataset can be found in the following (see Tables 1). As it can be seen from Table 1, proposed method has the best F measure which is equal to 88.49 with higher PSNR value of 17.20. Niblack has a very poor PSNR value because of noisy artifacts it generates. Sauvola has a very low recall while Otsu has a very low precision.



(a) Original Image (b) Sauvola

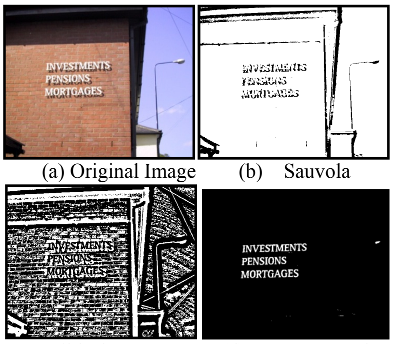
(c) Niblack (d) our approach

Figure 10. Output image corresponding to a second image (with complex background) belonging to the DIBCO2009 dataset [30].

Fig. 8, 12, 13 and 14 results are based on ICDAR2003 dataset [29]. These images are highly complex, subject to shadows and over exposed (see Fig. 12(a), 13 (a) and 14 (a)). For these images a significance level of 35% is used for binarization. As shown in Fig. 12, our results do not suffer from noise and are robust to uneven illumination and shadows. Niblack suffer from a lot of noise and takes a long time to perform binarization. Our algorithm seems to be a more robust candidate for text extraction and localization. Likewise, we can see on Fig. 13 that our results do not suffer from uneven hue variation changes. Both Sauvola and Niblack suffer from hue variations and specular reflections for the image in Fig. 13. Lastly, we can see on Fig. 14 some results based on ICDAR 2003 dataset. These images correspond to one of the most difficult images from ICDAR2003 dataset. As it can be seen from Fig. 14, proposed algorithm is robust against uneven illumination; shadowing and specular reflections. No ground truth has been provided for the ICDAR2003 dataset for thresholding evaluation. As a result we cannot provide any evaluation performance measures for the images belonging to this dataset to assess the robustness of our binarization algorithm.

TABLE 1. EXPERIMENTAL RESULTS – SUMMARY

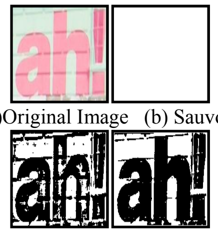
Method	RC	PR	FM	PSNR
Niblack	0.94	0.31	43.75	6.50
Sauvola	0.58	0.98	69.54	14.73
Otsu	0.96	0.16	78.48	15.17
Our Method	0.88	0.89	88.49	17.20



(a) Original Image (b) Sauvola

(c) Niblack (d) Our approach

Figure 12. Output image corresponding to one image (with complex background and uneven illumination) belonging to the ICDAR203 dataset [29].



(a)Original Image (b) Sauvola

(c) Niblack (d) our approach

Figure 13. Output image corresponding to one image (with complex background) belonging to the ICDAR2003 dataset [29].

(a) Input (b) output

(c) Input (d) Output

(e) Input (f) Output

(g) Input (h) Output

(i) Input

(j) Output

Figure 14. Output image corresponding to a third image (with uneven illumination, hue variations, and specular reflection) belonging to the ICDAR2003 dataset [29].

IV. CONCLUSION

From the results we obtained from ICDAR2003 [29] and DIBCO2009 [30] datasets we can conclude that the novel binarization algorithm proposed in this paper performs well on images with shadows, non-uniform illumination, low-contrast, large signal-dependent noise, smear and strain. The use of connected opening prior to binarization step significantly reduces the illumination variations and specular reflections. Since it emphasizes the region of interest it makes the binarization based on image statistics more reliable due to lack of variations after the preprocessing step. In comparison to other methods mentioned in DIBCO2009 [14], the proposed method is much simpler. Moreover, the F-measure (FM) results are very close to the best results reported in 2009, our PSNR values are higher. Lack of noise in the threshold image, good and robust performance results (as recall, precision), and low complexity time are of paramount importance when performing optical character recognition in degraded documents and text extraction from natural scenes applications. The experimental results that we have obtained show that the proposed method enables to reach this objective to greater extent.

The proposed methodology is based on the computation of the difference of gamma functions and on an approximation of these differences by image statistics. The main advantage of this novel algorithm is that it is not necessary to provide external parameters to tune the image results. Also proposed algorithm has the advantage of preprocessing to succeed in binarization step. Shadowing, reflection and uneven illumination problems can be solved substantially due to the fact that the Generalized Extreme Value Distribution (GEVD) is a very relevant statistical model which performs very well in the approximation to the difference of gamma functions. Also GEVD is capable of finding proper extreme values based on image statistics allowing us to deal with extreme conditions like shadows, high illuminations and reflections. Based on our experience, proposed algorithm is very fast and easy to implement.

To the best of our knowledge we are the first to present a binarization algorithm based on difference of gamma functions approximated by generalized extreme value distribution. Also we are the first to use connected opening operation as a preprocessing tool to emphasize on the region of interest prior to binarization. We believe our work in this regard is significant and in future experiments we will try to use this technique in color image segmentation.

REFERENCES

[1] J. Ohya, A. Shio, and S. Akamatsu. Recognizing characters in scene images. IEEE Trans. Pattern Anal. Mach. Intell., 16(2), 1994, pp.214–220.

[2] Y. Zhong, K. Karu, and A.K. Jain. Locating text in complex color images. Pattern Recognition, 28(10) , 1995, pp.1523–1535.

[3] O. D. Trier and T. Taxt, Evaluation of binarization methods for document images, IEEE Trans. Pattern Anal. Machine Intell., vol. 17, Mar. 1995, pp. 312-315.

[4] A.T. Abak, U. Baris, and B. Sankur, "The Performance Evaluation of Thresholding Algorithms for Optical Character Recognition", ICDAR 97, Ulm, Germany, 1997, pp. 697-700.

[5] Mehmet Sezgin, Survey over image thresholding techniques and quantitative performance evaluation, Journal of electronic imaging, 13, 146 ,2004, doi:10.1117/1.1631315.

[6] P. K. Loo and C. L. Tan. Adaptive Region Growing Color Segmentation for Text Using Irregular Pyramid. Document Analysis Systems VI Lecture Notes in Computer Science, Volume 3163/2004, 2004, pp. 103-106, DOI: 10.1007/978-3-540-28640-0_25.

[7] C.C. Fung, R. Chamchong, "A Review of Evaluation of Optimal Binarization Technique for Character Segmentation in Historical Manuscripts," Third International Conference on Knowledge Discovery and Data Mining, 2010 ,pp.236-240.

[8] T.N. Dinh, J. Park, G. Lee, Korean Text Detection and Binarization in Color Signboards, ALPIT ,2008, pp. 235-240.

[9] P. Stathis, E. Kavallieratou, N. Papamarkos, An evaluation survey of binarization algorithms on historical documents, ICPR, 2008, pp.1-4

[10] S.M. Lucas, A. Panaretos, L. Sosa, A. Tang, S.Wong, et al. Icdar 2003 robust reading competitions :entries, results, and future directions.International Journal on Document Analysis and Recognition, 7(2-3) , July 2005,pp. 105–122.

[11] D. Comanicu, P. Meer: "Mean shift: A robust approach toward feature space analysis." IEEE Trans. Pattern Anal. Machine Intell., 24, May 2002,pp. 603-619.

[12] J. Lim, J. Park and G. G. Medioni, Text segmentation in color images using tensor voting, Image and Vision Computing Volume 25, Issue 5, 1 May 2007, pp. 671-685.

[13] K. Wang and J.A. Kangas. Character location in scene images from digital camera. Pattern Recognition, 36(10), October 2003, pp. 2287–2299.

[14] B. Gatos, K. Ntirogiannis and I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO 2009)" DOI 10.1109/ICDAR.2009.246

[15] W. Niblack, An Introduction to Image Processing, Prentice-Hall, Englewood Cliffs, NJ 1986 , pp. 115–116.

[16] J. Sauvola and M. Pietaksinen, "Adaptive document image binarization," Pattern Recogn. 33, 2000 , pp. 225–236.

[17] Pickands, J. Statistical inference using extreme order statistics. The Annals of Statistics, 3, 1975,pp. 119-131.

[18] N. Otsu, "A threshold selection method from graylevel histograms", IEEE Trans. Systems Man Cybernet. Vol. 9 (1), 1979, pp. 62–66.

[19] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram," Computer Vision, Graphics and Image Processing 29, Mar. 1985, pp. 273-285.

[20] Pavlos Stathis, Ergina Kavallieratou, Nikos Papamarkos, "An Evaluation Technique for Binarization Algorithms", Journal of Universal Computer Science, vol. 14, no. 18 ,2008, pp. 3011-3030.

[21] B. Gatos, I. Pratikakis, K. Kepene, S.J. Perantonis, Text detection in indoor/outdoor scene images, in: Proc. First Workshop of Camera-based Document Analysis and Recognition, 2005, pp. 127-132.

[22] N. Ezaki, M. Bulacu, L. Schomaker, Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons, Proc. of 17th IEEE Int. Conf. on Pattern Recognition (ICPR 2004), vol. II, 2004, pp. 683-686.

[23] Coles, Stuart (2001). An Introduction to Statistical Modeling of Extreme Values, Springer-Verlag. ISBN 1-85233-459-2, pp. 45-50, pp75-78.

[24] J.F. Lawless, Statistical Models and Methods for Lifetime Data, Wiley, New York,1982, pp. 211- 255.

[25] C. N. Behrens, H. F. Lopes ,D. Gamerman "Bayesian Analysis of Extreme Events with Threshold Estimation", Statistical Modelling, Vol. 4, No. 3, 2004,pp. 227-244.

[26] P.Soille, "Morphological Image Analysis: Principles and Applications", Springer, 2003, pp. 182-198.

[27] R.C. Gonzalez , R.E. Woods "Digital Image Processing", 2008, ISBN 9780131687288

[28] J.B. Broadwater, R. Chellappa, "Adaptive Threshold Estimation via Extreme Value Theory", IEEE Trans. on Signal Processing, vol. 58, no. 2, Feb. 2010, pp. 490 – 500.

[29] http://www.cvc.uab.es/icdar2009/ (June 2010)

[30] http://users.iit.demokritos.gr/~bgat/DIBCO2009/ (June 2010)

[31] P. Prescott, "Parameter estimation for the generalized extreme value distribution", Journal of Statistical Computation and Simulation, Vol. 16, Issue 3 & 4, Jan.1983, pp. 241 –250.