Detecting Text in Natural Scenes Based on a Reduction of Photometric Effects: Problem of Text Detection

Alain Trémeau[1], Basura Fernando[2], Sezer Karaoglu[2], and Damien Muselet[1]

[1] Laboratoire Hubert Curien, Batiment E, 18 rue Benoit Lauras, University Jean Monnet, 42000 Saint Etienne, France
{alain.tremeau,daminen.muselet}@univ-st-etienne
[2] Erasmus Mundus CIMET Master, University Jean Monnet, Batiment B, 18 rue Benoit Lauras, 42000 Saint Etienne, France

Abstract. In this paper, we propose a novel method for detecting and segmenting text layers in complex images. This method is robust against degradations such as shadows, non-uniform illumination, low-contrast, large signal-dependent noise, smear and strain. The proposed method first uses a geodesic transform based on a morphological reconstruction technique to remove dark/light structures connected to the borders of the image and to emphasize on objects in center of the image. Next uses a method based on difference of gamma functions approximated by the Generalized Extreme Value Distribution (GEVD) 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 are defined in function of the background complexity. In this paper, we show that this method is much simpler than other methods for text binarization and produces better text extraction results on degraded documents and natural scene images.

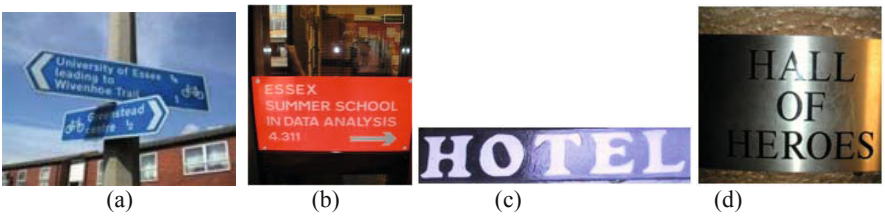
Keywords: Text binarization, Contrast enhancement, Gamma function, Photometric invariants, Color invariants.

1 Introduction

One of the most challenging tasks for color image segmentation is to be effective against image shadows, illumination variations and highlights. Several approaches based on the computation of image invariants that are robust to photometric effects have been proposed in the literature [1-3]. Unfortunately, there are too many color invariant models in the literature, making the selection of the best model and its combination with local image structures (e.g. color derivatives) quite difficult to produce optimal results [4]. In [5], Gevers et al. survey the possible solutions available to the practitioner. In specific applications, shadow, shading, illumination and highlight edges have to be identified and processed separately from geometrical edges such as corners, and T-junctions. To address the issue, Gevers et al. proposed to compute local differential structures and color invariants in a multidimensional feature space to detect salient image structures (i.e. edges) on the basis of their physical nature in [5]. In [6] the authors proposed a classification of edges into five classes, namely object

R. Schettini, S. Tominaga, and A. Trémeau (Eds.): CCIW 2011, LNCS 6626, pp. 230–244, 2011.

edges, reflectance edges, illumination/shadow edges, specular edges, and occlusion edges to enhance the performance of the segmentation solution utilized. Shadow segmentation is of particular importance in applications such as video object extraction and tracking. Several research proposals have been developed in an attempt to detect a particular class of shadows in video images, namely moving cast shadows, based on the shadow's spectral and geometric properties [7]. The problem is that cast shadow models cannot be effectively used to detect other classes of shadows, such as self shadows or shadows in diffuse penumbra [7] suggesting that existing shadow segmentations solutions could be further improved using invariant color features. The main challenge in color image segmentation is since a decade the fusion of low level image features so that image content would be better described and processed. Several researches provided some solutions to combine color derivatives features and color invariant features, color features and other low level features (e.g. color and texture, color and shape [5]), low-level features and high-level features (e.g. from graph representation [8]). However, none of the proposed solutions appear to provide the expected performance to segment complex color images unlike the human visual system which is able to take into account the semantic contents of images. Of course if some a priori information or knowledge about the segmentation task is incorporated in the process that will optimise the algorithm results. In section 2.1 we show that former solutions suffer from limitations and are useless when addressing complex illumination conditions, such as those illustrated by image (d) of Fig. 1.



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

Fig. 1. Color changes due to shading (a), local variation in the intensity of the illumination (b), specularities (c), and specularities and inter-reflections (d)

In a first paper (see [9]) we have demonstrated that none illuminant-invariant model is sufficiently robust to complex photometric effects to solve the issue of text detection in complex natural scenes. To solve this issue, in this paper we propose to use another strategy, more robust to photometric effects, based on the computation of the difference of gamma functions to detect text layers in complex scenes.

Most of existing text segmentation approaches assume that text layers are of uniform color and fail when this is not the case. Furthermore, the background may also be multicolor consequently the assumption according with it is the largest area of (almost) uniform color in the image does not necessarily hold [10]. Lastly, most of existing text segmentation approaches assume that there is a high contrast between text and background in the image this is unfortunately not always the case in real images. Furthermore, many approaches assume that in segmenting the highest peak in the lightness histogram we can deduce if text layers are of lower or a higher lightness

than the background region, this information may be helpful to segment text layers, but this is once again not always the case in real images.

In this paper we propose to use a new text segmentation method robust to photometric effects. The proposed method, introduced in [11-12] (see flowchart in Fig. 2), first uses a geodesic transform based morphological reconstruction technique to remove dark/light structures connected to the borders of the image and to emphasize on objects in center of the image. Next uses a method based on difference of gamma functions approximated by the Generalized Extreme Value Distribution (GEVD) 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. This approach is based on a new concept of difference of gamma functions used to emphasize certain regions in function of to their intensity distribution. The novel thresholding algorithm is presented in sections 2.3 and 2.4. Next, experimental results are given in section 3. In order to assess text detection methods we use two datasets (ICDAR 2003 and DIBCO 2009) used for competitions [13-15]. Lastly a conclusion is drawn in section 4.

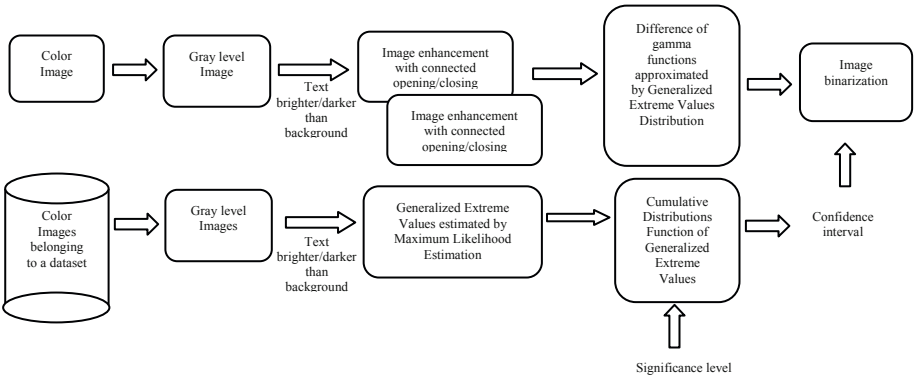


Fig. 2. Flow Chart of the method proposed

2 Text Segmentation

2.1 Image Enhancement Methods

Existing text segmentation approaches 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 [16]. Commonly used methods are histogram based thresholding and adaptive thresholding. 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 had been proposed by Niblack [17] and Sauvola [18].These methods are more robust against uneven illumination and varying colors than global ones but suffer regarding to dependency of parametric values. Trier and Taxt presented an evaluation of binarization methods for

document images in [19]. Most of existing text segmentation approaches assume that text layers and background regions are of uniform color and that there is a high contrast between text layers and background regions in the image this is unfortunately not always the case in real images. For example Karatzas proposed in [10] a text segmentation method which first splits the image in regions that are perceptually different in color next merges connected components having the highest overlapping degree. The splitting process in based on histogram analysis. Peaks are identified by locating minima and maxima on lightness histogram next on hue histogram then a tree structure of layers is created. This method is based on the hypothesis that the intra-class variances are low and the inter-class variances are high but this is not always the case for complex images.

Region based grouping methods are mainly based on spatial-domain region growing, or on splitting and merging. 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 natural scenes. In addition, in terms of computation 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. 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. These methods are not well-adapted for complex natural scenes such as urban scenes with complex background. To be effective these methods require that the intra-class variances are low and the inter-class variances are high. The K-means algorithm and the fuzzy-cmeans algorithm were until recently two of the main techniques used for clustering based grouping. Recently, several studies have also shown that the mean-shift algorithm based density estimation outperforms K-means algorithm [20]. 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 idea is to define k centroids, next to perform an iterative process till all pixels belong to a cluster whose centroid is the nearest one.

Even if many approaches have been specifically developed for text layers segmentation based on image binarization most of these approaches fail when image is complex such as in natural scene images. The aim of this work is to present a new strategy to segment text layers in complex images. The main objective is to be robust against photometric effects such as shadows, highlights, specular reflection, non-uniform illumination, complex background, varying text size, colors and styles. The second objective is to reduce noise while enhancing contrast between text layers and background regions using substantially lesser complex processes than other well-known approaches. Noise removal is essential not only for text segmentation but also for other processes such as Optical Character Recognition (OCR).

2.2 Morphological Reconstruction Based on Geodesic Transform

In order to suppress lighter objects (e.g. text layers) than their surroundings and connected to border of the image, another strategy consists to use a morphological reconstruction transform based on geodesic dilation.

According to Soille [21] 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. 3):

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

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

$$\gamma_x(g) = R_g^{\partial}(f_x) \tag{2}$$

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.

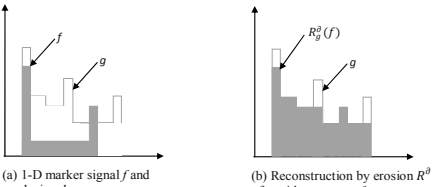
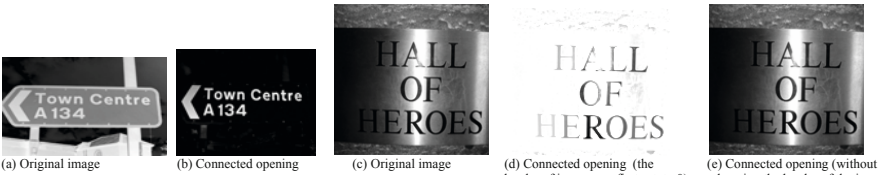


Fig. 3. Algebraic opening for a 1-D signal

According to Soille [21] 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. 4). 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 layer candidates which leads us to emphasize more on region of interest of the image (see Fig. 4 (b)). Especially in our experiments we have seen that this process reduce the background intensity variations and enhance the text layers of the image.

In order to suppress darker objects (e.g. text layers) than their surroundings and connected to border of the image the connected closing transformation can be used. The first shortcoming of this morphological transformation and of the former (i.e. closing and opening) is that we must first estimate if the background is lighter or darker than the text layers, i.e. we must first extract the background of the image. The second shortcoming of these two transformations is that they work quite fine when text layers are only darker or whiter than the background but do not perform well when text layers are darker and whiter than their surrounding local background in the image. Lastly, these transforms do not work well when the border of the image has the same intensity than text layers, such as in image (d) of Fig. 1. That why, to enhance this image we set its borders to zero before applying the connected closing transform (see Fig. 4 (d)) otherwise this transform is inefficient (see Fig. 4 (e)).



(a) Original image (b) Connected opening (c) Original image (d) Connected opening (the border of image was first set to 0) (e) Connected opening (without changing the border of the image)

Fig. 4. (b) Connected image objects having higher intensity values than their surrounding can be extracted by the connected opening transform. (d) Connected image objects having darker intensity values than their surrounding can be extracted by the connected closing transform.

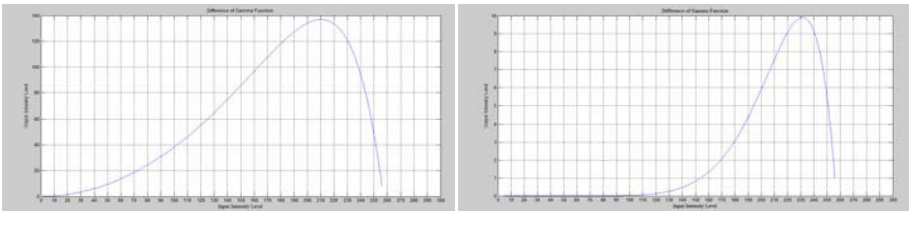


Fig. 5. (a) Difference of gamma functions for ($\gamma 1 = 2$, $\gamma 2 = 4$). (b) Difference of gamma functions for ($\gamma 1 = 9$, $\gamma 2 = 10$).

2.3 Background Estimation Based on Difference of Gamma Functions

In order to classify pixels belonging to the foreground (e.g. text layers) of those belonging to the background, we propose here an additional step based on the computation of the difference of gamma functions.

Let us now consider two gamma contrast enhancement functions defined as follows:

$$g_1(r)=c_1 r^{\gamma_1}, g_2(r)=c_2 r^{\gamma_2} \quad (3)$$

where r represents the input intensity levels, $g_1(r)$ and $g_2(r)$ represent the output intensity levels for gamma values $\gamma 1, \gamma 2$ ($\gamma 1 < \gamma 2$). M is the maximum intensity value (i.e. $0 \leq r \leq M$, e.g. with a 8-bit image M = 255). The constant c is defined by $c = M^{(1-\gamma)}$.

The two functions defined by eq. (4) can be applied to image $f(x,y)$ to obtain two enhanced images $f_1(x,y)$ and $f_2(x,y)$. Then, we can compute the difference of gamma functions as follows (see Fig. 5):

$$diff_{f1,f2}(x,y)= |f_1(x,y)-f_2(x,y)| \quad (4)$$

Then, in order to classify pixels belonging to the foreground (e.g. text layers) of those belonging to the background (see Fig. 6), we propose to apply the following rule.

$$\forall (x,y) \in f(x,y) \text{ if } diff_{f1,f2}(x,y)>T \implies (x,y) \in \text{foreground} \quad (5)$$

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

where $diff_{f1,f2}(x,y)$ is the image corresponding to the difference of gamma functions.

Fig. 6. (a) Original Images. (b) Gamma correction applied to connected opening (resp. closing) enhanced image. (c) Gamma correction applied to connected opening (resp. closing) enhanced image. (d) Difference between gamma corrected images. (e) Thresholded image.

The above rule makes sense only if, in the enhanced image $diff_{f1,f2}(x,y)$, higher values correspond to text layers and lower values correspond to background regions. As it can be seen in Fig.5, the choice of gamma values affects the suppression of some intensity ranges on the resulting image, so they play a major role in the classification process. As example, the gamma values used in Fig. 6 (b1) and (c1) yield to the suppression of low intensity ranges in Fig. 6 (d1) (see Eq. (6)).

As the background is always either darker or lighter than the surround we consider that there is always a contrast issue between them. When the background is lighter than the foreground, such as in Fig. 6 (a2), rather than using two power law functions (i.e. two negative gamma coefficients) we propose to apply the above difference of gamma functions on the inverse of the image. In the following we consider that the background is darker than the foreground.

When $\gamma_1 < \gamma_2$ the second gamma function $f_2(x, y)$ suppresses more background intensities and enhances more contrasts of foreground intensities than the first $f_1(x, y)$. Unlike other binarization techniques which generate noise artifacts, especially in relatively homogeneous areas such as the background, when we take the difference between two gamma-corrected images we do not generate noisy artifacts in the background. The function $diff_{f1,f2}(x, y)$ presents two main advantages; firstly it better contrasts middle range intensity values, secondly it suppresses lower and higher intensities (see Fig. 5).

By thresholding the resulting image by a value very close to zero, we obtain a perfect separation of foreground and background (see Fig. 6 (e)). As mentioned earlier, different gamma values for γ_1 and γ_2 yields suppression of different intensity ranges. Depending on gamma values γ_1 , γ_2 and threshold value T we obtain different binarization results. In order to improve the classification process of foreground pixels and background pixels, we propose now to use a non-supervised process whose objective is to optimize the choice of gamma values for γ_1, γ_2 and of threshold parameter T. Fig.5 shows that the suppression of some intensity ranges depends on the value of γ_1 and γ_2. In the following we use the following notations to define the difference of gamma functions:

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

It can be seen on Fig. 5 that $\Delta f_{2,4}$ has a lower power of suppression of intensity ranges compared to $\Delta f_{9,10}$. Let us now consider an arbitrary threshold corresponding to an output value of 2, then for $\Delta f_{2,4}$ the suppression concerns a range of intensity values less than 10, meanwhile for $\Delta f_{9,10}$ the suppression concerns a range of intensity values 10 times larger. When we use the function $\Delta f_{9,10}$ with a threshold value T = 2 then the corresponding threshold is 100 for the input image. When we use the function $\Delta f_{2,4}$ with a threshold value T = 2 then the corresponding threshold is 10 for the input image.

2.4 Background Estimation Based on Generalized Extreme value Distribution

As discussed in the introduction, the main problem of text extraction is to find correct thresholds to remove background in order to separate text layers from background. The main problem we have to face here is to find appropriate gamma values and threshold value to obtain a relevant binarization. Fig. 5 shows clearly that depending on the gamma values, the suppression of intensity ranges by function $\Delta f_{\gamma_1, \gamma_2}$ varies significantly. How to find appropriate gamma values without a prior knowledge of the image studied?

To solve this issue we propose to compute image statistics from a dataset of text images and to use these statistics to model the distribution of intensities of text images. This proposal is justified by the fact that in natural scene images most of pixels

belonging to text layers reside in the middle range of the distribution of pixel intensities. We propose to use the Generalized Extreme Value Distribution (GEVD) model [22] to find the best thresholds (i.e. the optimized ones) to separate text layers from background. 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 necessary to model the maximum and the minimum values of a collection of random observations from a unique distribution. These three distributions are called Gumbel, Fréchet, and Weibull distributions [22]. Extreme value theory is an excellent tool to deal with the modeling of sparse data. It is a useful tool to face the thresholding problem in the field of image binarizartion.
Generalized Extreme Value Distribution can be written as:

$$\text{For } k \neq 0 \qquad f(x)=\left\{\frac{1}{\sigma}\exp^{\left(-(1+kz)^{-1/k}\right)} . (1+kz)^{-1-(1/k)}\right. \tag{7}$$

$$\text{For } k = 0 \qquad f(x)=\left\{\frac{1}{\sigma}\exp^{\left(-z-\exp(-z)\right)}\right. \tag{8}$$

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, different methods can be used, such as [23-24]. Pickands showed in [25] that if X is a random variable and F (x) is its Probability Distribution Function (PDF) then under certain conditions, $F(x|u) = P(X \leq u + x \,|\, X > u)$ can be approximated by a Generalized Pareto Distribution (GPD) [26]. In other words, GPD can be used to find the thresholds of an identical distribution. Let $X = \{X_1, X_2, X_3, \ldots X_n\}$ be independent random variables with identical distribution F. Next, suppose that $D_n = \max(X)$, then it can be shown that for a large value of n:

$$P(D_n<x) \approx f(x) \tag{9}$$

here f(x) corresponds to the Generalized Extreme Value Distribution (GEVD) and u represents the threshold over which the observations {X} exceed. Then, u can be modeled by GPD.

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 propose to compute these significance levels to find proper thresholds for binarization. From our experimentations, we have empirically considered that a significance level of 10% is sufficient to detect simple backgrounds; (see Fig. 7 and 8) meanwhile a significance level of 35-40% is necessary to detect complex backgrounds in natural scenes (see Fig. 9 (a) and (c)).

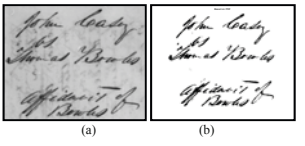


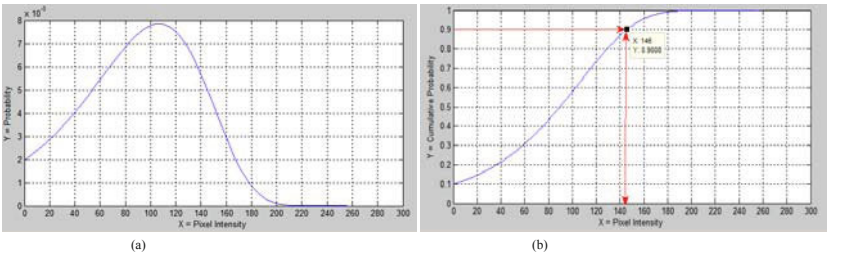
Fig. 7. (a) Input Image with simple background. (b) Threshold image (nb. a significance level of 10% corresponds here to an input intensity value of 146).

To remove both background and over exposed regions we have to define a confidence interval. We assume that foreground intensities lie in a given range:

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

here U_{t1} and U_{t2} are, respectively, the lower and the upper thresholds of this interval.

To find t_1 and t_2, we propose to compute for $U_{t1}$the cumulative probability of P_1 and for $U_{t2}$the cumulative probability of P_2, in function of the significance level desired. According to our experiments done from 500 images belonging to the ICDAR 2003 dataset [13] and DIBCO 2009 dataset [15], the GEV cumulative probabilities $P_1 = 0.7$ and $P_2 = 0.99$ are sufficient to remove most of overexposed regions and backgrounds [12]. As example, see Fig. 9 (c) and (d).



(a) (b)

Fig. 8. (a) PDF of the GEVD of the image of Fig. 6 (a). (b) CDF of the GEVD of Fig. 6 (a).

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

Fig. 9. (a) Input image with a complex background. (b) Threshold image. (c) Input image with over exposed background. (d) Threshold image.

3 Experimental Results

Fig.7 and 10, and Table 1 illustrate some experimental results that we get with the DIBCO2009 dataset [15]. The main interest of the DIBCO2009 dataset is that the ground truth of the binarization of each image is provided and that evaluation performance measures are also provided. Let us note that most of the images belonging to this database are not overexposed nor subjected to shadows, in other words their background is moderately complex (see Fig. 10). Consequently, to binarize these images we have used a significance level of 10%. For each of these images we have computed the Precision (PR), recall (RC), F-measure (FM) and peak signal to noise ratio (PSNR) values relatively to the ground truths provided. To analyze the performance of our binarization method we have compared the results that we get with those obtained by Niblack [17], Sauvola [18] and Otsu [27] algorithms.

The results of the DIBCO2009 competition can be found in [15]. All performance calculations based on DIBCO dataset have been computed according to the definitions provided for DIBCO2009 competition. The comparison of results computed from the DIBCO dataset is done in Table 1. As it can be seen the proposed method has the best F measure (value equal to 88.49) and the higher PSNR (value equal to 17.20). Niblack has a very poor PSNR value because it generates noisy artifacts. Sauvola has a very low recall while Otsu has a very low precision.

Fig. 9, 11 and 12 illustrate some experimental results that we get with the ICDAR 2003 dataset [13]. These images are highly complex, subject to shadows or overexposed (see Fig. 11 (a) and 12 (a)). For these images a significance level of 35% is used for binarization. As shown in Fig. 11 (d) and Fig. 12 (b), our results do not suffer from noise and are robust to uneven illumination and shadows. Niblack suffers from a lot of noise and takes a long time to perform binarization. Our algorithm seems to be more robust for text extraction and segmentation.


(a) Original Images

(b) Sauvola

(c) Niblack

(d) Our approach

Fig. 10. Output images corresponding to three handwritten images (with moderately complex background) belonging to the DIBCO2009 dataset [15]

Furthermore, we can see on Fig. 11 (d2) that our algorithm does not suffer from uneven hue variation changes. Both Sauvola and Niblack suffer from hue variations and specular reflections, such as those shown in Fig. 11 (a2). Lastly, in Fig. 12, we have selected some of the most difficult images of ICDAR 2003 dataset. As it can be seen from Fig. 12 (b), the proposed algorithm is robust against uneven illumination; shadowing and specular reflections. Unfortunately, no ground truth has been provided for the ICDAR 2003 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.

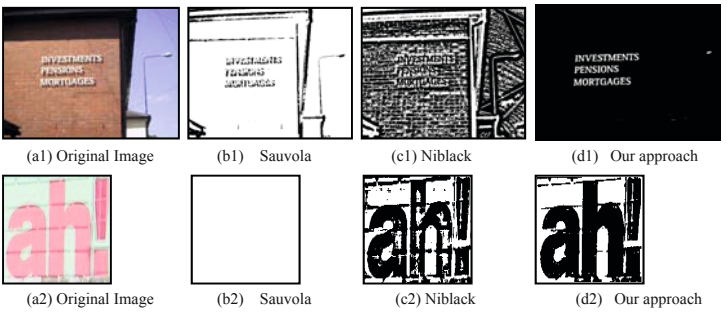


Fig. 11. Output images corresponding to two outdoor images (with complex background and uneven illumination) belonging to the ICDAR 2003 dataset [13]

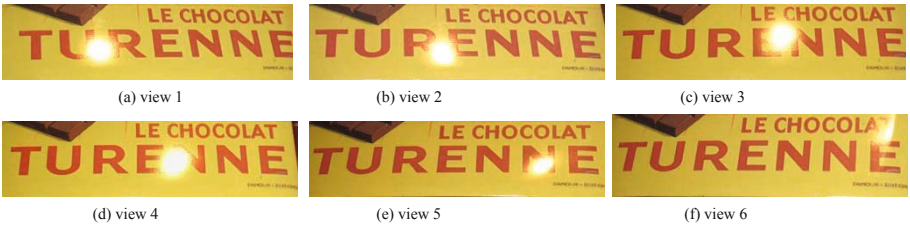
Table 1. Summary of experimental results

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

Invariance to	Lightness change(s) &/or shift(s) +	Color change(s) &/or shift(s) +	Color & Lightness change(s) &/or shift(s) +
Examples of image	(a)	(b)	(c)
Results obtained	(d)	(e)	(f)
Examples of image	(g)	(h)	(i)
Results obtained	(j)	(k)	(l)

Fig. 12. Output images corresponding to images with uneven illumination, hue variations, specular reflections. Images (a), (b), (g) and (h) belongs to the ICDAR 2003 dataset [13]. Invariance to lightness &/or color change(s) &/or shift(s) is indicated with '+' and lack of invariance with '-'.

Lastly, Fig. 12 shows the color invariance of the proposed algorithm to different types of lighting changes [28, 9].



(a) view 1 (b) view 2 (c) view 3

(d) view 4 (e) view 5 (f) view 6

Fig. 13. According to the angle of observation some letter are hidden by highlights

4 Conclusion

From the results we obtained from ICDAR 2003 and DIBCO 2009 datasets we can conclude that the binarization algorithm proposed in this paper performs well on images with shadows, non-uniform illumination, low-contrast, large signal-dependent noise, smear and strain. Several examples have been given in this paper to show this invariance. In comparison to other methods mentioned in DIBCO 2009 and in H-DIBCO 2010 [29], the proposed method is much simpler. Moreover, the F-measure (FM) results are very close to the best results reported in 2009 meanwhile 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 (OCR) 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. Users have only to indicate if the image is a complex image or a moderately complex image, in the former case the significance level is put to 35%, otherwise to 10%. Rather than asking to the user if the image is complex or not, it would be desirable to use an automatic process, but how to differentiate a complex image of a moderately complex image? We have shown in this study that this seems very complicated given the different photometric effects and cases of study that may exist in natural images.

To solve the issue of shadowing, reflection and uneven illumination, we have shown that the Generalized Extreme Value Distribution (GEVD) is a very relevant model to approximate differences of gamma functions. Indeed 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. To solve the problem of low contrast between text and background, we have shown that the difference of gamma functions is a very relevant model to enhance contrast between text layers and background regions while reducing noise, using substantially lesser complex processes than other well-known approaches. Furthermore, this tool is robust to photometric effects. In the proposed paper, users have to indicate if the text is darker or lighter than the background. Different methods could be used to automatically estimate the intensity level of the background, but we have shown in this study that without any prior knowledge it is very complicated to characterize complex backgrounds in natural image. Lastly, the proposed algorithm is very fast and easy to implement.

In the future, we aim to address the challenging problem of text detection and segmentation with multi-view images. The idea will be to combine the text information extracted from different views, as example see Fig. 13. We aim also to address the challenging problem of text detection and segmentation in video sequences. The idea will be to take into account the evolution of text information over time under the hypothesis that the camera moves in the scene, in other words that the angles of observation changes during the video and then local photometric effects can be temporally compensated. The interest of exploiting temporal redundancy is that it can increase the probability of detecting text layers since the same text may appear under varying lighting conditions from frame to frame [16]. Consequently, missed texts in individual frames can be interpolated. It can also remove false alarms in individual frames since they are usually not stable over time.

References

1. van de Weijer, J., Gevers, T., Geusebroek, J.M.: Edge and corner detection by photometric quasi-invariants. IEEE Trans. on Pattern Analysis and Machine Intelligence 27(4), 625–630 (2005)
2. Li, B., Xue, X., Fan, J.: A robust incremental learning framework for accurate skin region segmentation in color images. Pattern Recognition 40(12), 3621–3632 (2007)
3. Moreno-Noguer, F., Sanfeliu, A., Samaras, D.: Integration of deformable contours and a multiple hypotheses Fischer color model for robust tracking in varying illuminant environments. Image and Vision Computing 25, 285–296 (2007)
4. Trémeau, A., Tominaga, S., Plataniotis, K.: Color in Image and Video Processing: most recent trends and future research directions. EURASIP Journal on Image and Video Processing 2008, article ID 581371, 26 p. (2008)
5. Gevers, T., van de Weijer, J., Stokman, H.: Color feature detection. In: Color Image Processing: Methods and Applications Book, ch. 9, pp. 203–226. CRC press, Boca Raton (2007)
6. Koschan, A., Abidi, M.: Detection and classification of edges in color images. IEEE Signal Processing Magazine, 64–73 (January 2005)
7. Salvador, E., Cavallaro, A., Ebrahimi, T.: Cast shadow segmentation using invariant color features. Computer Vision and Image Understanding 95, 238–259 (2004)
8. Dong, G., Xie, M.: Color clustering and learning for image segmentation based on neural networks. IEEE Trans. on Neural Networks 16, 925–936 (2005)
9. Trémeau, A., Godau, C., Karaoglu, S., Muselet, D.: Detecting text in natural scenes based on a reduction of photometric effects: problem of color invariance. In: Schettini, R., Tominaga, S., Trémeau, A. (eds.) CCIW 2011. LNCS, vol. 6626, pp. 234–248. Springer, Heidelberg (2011)
10. Karatzas, D., Antonacopoulos, A.: Colour text segmentation in web images based on human perception. Image and Vision Computing 25, 564–577 (2007)
11. Fernando, B., Karaoglu, S., Trémeau, A.: Extreme value theory based text binarization in documents and natural scenes. In: Proceedings of IEEE, ICMV, Hong-Kong (to be published)
12. Karaoglu, S., Fernando, B., Trémeau, A.: A Novel Algorithm for Text Detection and Localization in Natural Scene Images. In: Proceedings of IEEE, DICTA 2010, Sydney, Australia, December 1-3 (2010) (to be published)

13. ICDAR 2003 robust reading competitions. In: Proc. of 7th Intl. Conf. on Document Analysis and Recognition, pp. 682–687 (2003)
14. ICDAR 2003 text locating competition results. In: Proc. of 8th Intl. Conf. on Document Analysis and Recognition, pp. 80–84(1) (2005)
15. Document Image Binarization Contest (DIBCO 2009) in the framework of ICDAR2009. In: Proc. of 10th Intl. Conf. on Document Analysis and Recognition, pp. 1375–1382 (2009)
16. Lienhart, R., Wernickle, A.: Localizing and segmenting text in images and videos. IEEE Trans. on Circuits and Systems for Video Technology 12(4), 256–268 (2002)
17. Niblack, W.: An Introduction to Image Processing, pp. 115–116. Prentice-Hall, Englewood Cliffs (1986)
18. Sauvola, J., Pietaksinen, M.: Adaptive document image binarization. Pattern Recogn. 33, 225–236 (2000)
19. Trier, O.D., Taxt, T.: Evaluation of binarization methods for document images. IEEE Trans. Pattern Anal. Machine Intell. 17, 312–315 (1995)
20. Lim, J., Park, J., Medioni, G.G.: Text segmentation in color images using tensor voting. Image and Vision Computing 25, 671–685 (2007)
21. Soille, P.: Morphological Image Analysis: Principles and Applications, pp. 182–198. Springer, Heidelberg (2003)
22. Coles, S.: An Introduction to Statistical Modeling of Extreme Values, pp. 45–50, 75-78. Springer, Heidelberg (2001) ISBN 1-85233-459-2,
23. Lawless, J.F.: Statistical Models and Methods for Lifetime Data, pp. 211–255. Wiley, New York (1982)
24. Prescott, P.: Parameter estimation for the generalized extreme value distribution. Journal of Statistical Computation and Simulation 16(3&4), 241–250 (1983)
25. Pickands, J.: Statistical inference using extreme order statistics. The Annals of Statistics 3, 119–131 (1975)
26. Behrens, C.N., Lopes, H.F., Gamerman, D.: Bayesian Analysis of Extreme Events with Threshold Estimation. Statistical Modeling 4(3), 227–244 (2004)
27. Otsu, N.: A threshold selection method from graylevel histograms. IEEE Trans. Systems Man Cybernet. 9(1), 62–66 (1979)
28. Álvarez, J.M., Gevers, T., López, A.M.: Learning Photometric Invariance for Object Detection. Int. J. Comput. Vis. 90, 45–61 (2010)
29. Pratikakis, I., Gatos, B., Ntirogiannis, K.: H-DIBCO 2010 - Handwritten Document Image Binarization Competition. In: Proceedings of the 12th International Conference on Frontiers in Handwriting Recognition, pp. 727–732 (2010)