An Adaptive Foreground-Background Separation Method for Effective Binarization of Document Images

Bishwadeep Das[1], Showmik Bhowmik[2], Aniruddha Saha[3], Ram Sarkar [2]

[1] Department of Electronics & Communication Engg., Motilal Nehru National Institute of Technology, Allahabad, India
[2]Department of Computer Science & Engineering, Jadavpur University, India.
[3] Samsung Research Institute Bangalore, Bangalore, India

el129003@mnnit.ac.in,{showmik.cse, ani0075saha, raamsarkar }@gmail.com

Abstract. *Binarization is a process of classifying the pixels of an image as either foreground or background. Most of the binarization techniques suffer from the noise appearing in the images during acquisition such as uneven illumination. In the present work, a foreground-background separation method is developed to enhance the performance of a document image binarization method. To examine its effectiveness, it is combined with two state-of-the-art binarization methods (i.e. Otsu's method [1] and Mitianoudis' method [2]) and the performances of the combined methods are compared with the original methods. For the experiment, two standard databases viz., DIBCO 2012 and 2013 are used. The results confirm that the proposed method performs satisfactorily even if the images are considerably noisy.*

Keywords: Foreground-background separation; Binarization; Document image;

1 Introduction

Document image binarization is the process of converting a document image into a bi-level digital image, comprising text (along with graphics) and background. During the binarization process, each pixel of the original image is either classified as foreground or background. These images are used in further high level post processing tasks like Optical Character Recognition (OCR), Writer identification, Layout Analysis etc. Hence the success of such high level processing steps is strongly dependent on the quality of the binarized image.

Binarization of a document image is not an easy task to do. Because the disturbances such as, uneven illumination, quality degradation, noise and artifacts (patches, bleed through, creases etc.) are very common to document images. Presence of these affects the quality of the acquired image to a great extent. A good binarization algorithm should deal with most of these challenges effectively.

Binarization techniques which perform foreground and background separation before actual binarization are being used recently. The foreground estimation and background separation, combined with subsequent binarization can produce powerful binarization algorithms for document images. Gatos et al. [3] propose a method which involves foreground-background separation. The image is processed by a wiener filter and Sauvola's [4] technique is used to estimate the foreground roughly. The background is then obtained by using interpolation to fill the regions which have been classified as foreground. This background estimate is then compared with the original image to obtain the binarized image. Mitianoudis et al. [2] use the fact that background regions have relatively slow variations in intensity. To obtain the background they apply median filtering on a gray level representation of the document image.

Few methods have used clustering for binarization after the foreground and background separation. For example, Valizadeh and Kabir[5] partition the two dimensional feature space into a number of small regions and then classify those regions as text or background before doing pixel wise classification. Mitianoudis et al.[2] use a novel LCM (Local Co-occurrence Matrix) algorithm to perform the clustering.

In the present work, we have devised an effective foreground-background separation method inspired by the method reported in [2]. For the evaluation of our technique, it is combined with two state-of-the-art binarization methods namely, Otsu's method [1] and Mitianoudis's method [2]. The performance of the combined methods is compared with the original methods. For the experiment two standard databases viz., DIBCO-2012 and DIBCO-2013 are used. It is observed that our method can effectively eliminate some common disturbances like, noise, dark patches, few creases etc. before the actual binarization. The current paper is organized as follows: the foreground-background separation method is described in section 2. Section 3 presents the experimental results and finally the conclusion is given in section 4.

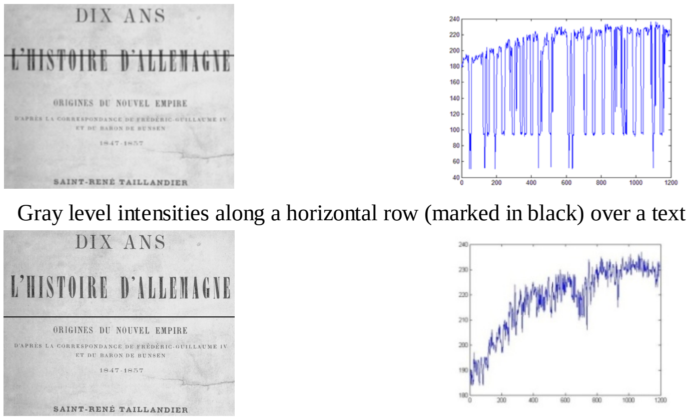
2 Foreground-Background Separation

In this work, the input images are first converted to their gray scale version and then the foreground-background separation is performed. For document images, the foreground refers to the regions of the image which contain text along with some relevant information. The foreground is further processed to obtain the binarized output whereas the background is generally not needed to be processed any further. This overall process can be broadly divided into two integral sub-processes viz., Background approximation and background elimination. In the first stage, the background of the image is approximated using a combination of adaptive median filtering and max-filtering. Whereas, in the next stage, the background is eliminated through a process of hybrid multi-level thresholding followed by an adaptive method loosely based on the Signal-to-Noise Ratio (SNR). Both of these processes are explained in detail in the subsequent sections.

2.1 Background Approximation

Typically the background of a document image is the portion of the image which does not contain text, graphics or meaningful information. The information in the text region represents higher variations in space, whereas the background can be treated as the part of the image having low variation in spatial frequency (see Fig 1). Hence, a way of extracting or approximating the background would be to obtain all the low spatial variations in the image while separating out the high spatial variations. This is precisely the job of a low-pass spatial filter. As images are two dimensional spatial digital signals, filters can be applied on them effectively.

A median filter is a type of order statistic filter where the middle value from this ordered set is chosen. A median filter is very effective at eliminating outliers and it does not have to compute new values at each pixel. Hence, wherever there is an abrupt change in spatial intensity, they can smoothen out the variations. A smoothened signal represents low spatial variation and hence a median filter converts a signal, spatial or temporal, into its smoothened or low-pass version. We use median filters, likewise, to extract the regions having low spatial variation *viz.* the background. We start with the grayscale image of the document image $I_G(x,y)$. This image is filtered continually using adaptive median filtering [2]. After that, a *max-filter* of rectangular size dxd is used where,d is the final size of the median filtering window. A max-filter, like a median filter is an order statistic filter which picks the maximum value from a set of ordered values.



(a) Gray level intensities along a horizontal row (marked in black) over a text region

(b) Gray level intensities along a horizontal row (marked in black) over background region

Fig. 1. Illustrates the spatial variation in gray level intensity. The plots on the right show the variation of intensity(Y axis) along the row pixels(X axis) for the row shown on the left image with a black line.

2.2 Background Elimination

Mitianoudis et al. [2] have separated the foreground from the background on the basis of a practically obtained global threshold. A difference image $I_D(i,j)$ is obtained by subtracting the background image from the original gray level image. The difference image is then separated into foreground and background on the basis of a global threshold which is estimated from its histogram.

We have opted for a hybrid thresholding approach, where we have divided the difference image I_D into non-overlapping PxP patches. The local threshold T_{local} for each patch is computed using Otsu's threshold [1], whereas the global threshold T_{global} is already obtained as per Mitianoudis et al. [2]. The average of these thresholds T_{patch} is computed and applied on each patch.

$$T_{patch} = \frac{T_{global} + T_{local}}{2} \tag{1}$$

Here, $T_{patch}$is the threshold computed for each patch. The thresholded image thus obtained is called I_T. Thresholded image generation process is given in Algorithm 1.

Algorithm 1: Computation of primarily threshold image I_T

```
For each patch of size $PxP$in the difference image $I_D$
    Step1:    Calculate $T_{patch} = \frac{T_{global}+T_{local}}{2}$
    Step2:    Calculate the thresholded image$I_T$as
              $I_T(i,j) = \begin{cases} I_G(i,j). & if I_D(i,j) < T_{patch} \\ 255. & otherwise \end{cases}$
End
```

It is observed that the local thresholds for the regions containing text are much higher than the global threshold which is having low value throughout the experiment. Hence, the average would be midway and this essentially states that upon thresholding, the regions containing text and their surroundings can be kept intact as the foreground. However for patches which correspond to the background regions in the difference image, the Otsu threshold is low and close to the global threshold. Hence, thresholding creates an output which resembles a very noisy patch. This brings us towards an approach inspired by SNR (Signal-to-Noise Ratio), which is defined in eqn (2).

$$SNR = \frac{Signal\ Power}{Noise\ Power} \tag{2}$$

SNR is the measure of the quality of an analog or a digital signal which indicates how much the original signal is ruined by noise. Digital images are also a type of signal in the spatial domain. Hence, the concept of SNR applies equally to digital images as well. Noise is random in nature and variance is often a good measure of it.

We can treat the noisy patches obtained in the thresholded image as a form of noisy signal. For any MxN deterministic image I, power of the signal is expressed as

$$P = \sum_{i=1,j=1}^{i=M,j=N} I(i,j) * I(i,j) \tag{3}$$

For a pure noise signal $n(i,j)$, the power is given by σ_n^2, which is the variance of the distribution from which the samples $n(i,j)$ are derived. It can be calculated from the variance of a sufficient number of image pixels or from the histogram of a corrupted image directly if the noise is additive in nature.

For a deterministic image corrupted by additive noise, the observed values of the image samples are drawn from a shifted version of the distribution of the noise. Hence, for observed images, the variance is also a measure of the power present in the image. The variance of an MxN patch of such an image I is calculated as

$$var = \frac{1}{MN} \sum_{i=1,j=1}^{i=M,j=N} (I(i,j) - \mu)^2 \tag{4}$$

where μ is the sample mean given by

$$\mu = \frac{1}{MN} \sum_{i=1,j=1}^{i=M,j=N} I(i,j) \tag{5}$$

The thresholded image is similar to a noisy image although it is not analogous to noise. For each PxP non-overlapping patch$k$, the variance var_T^k is calculated for the thresholded image to get a measure of its behavioral similarity to noise. For each corresponding patch in the difference image, the variance var_D^k is calculated as given by eqns.(4) and (5). The variance from the difference image of the patches is high in the text regions and is very low in the background. Hence, the ratio $r(k) = {var_D^k}/{var_T^k}$would be low in the background regions and relatively higher in regions having textual information. This justifies the use of ratio of variances $r(k)$.The ratios from all the patches are obtained along with the mean value. A practical threshold T_p is selected which is fraction t_r times the mean of the ratios. This is done for all the patches. The adaptive hybrid background separation based on variance ratio is summarized in Algorithm 2.

Algorithm 2: Separation of background

Let there be n patches of size PxP in the difference image I_D and the thresholded image I_T. Let r be a vector of size n.

```
Step1:      For each patch $k$
              Compute the variances $var_D^k$ and $var_T^k$ using eq. (4).
              Calculate$r[k] = {var_D^k}/{var_T^k}$
            End
Step2:      Calculate $\mu_r$the mean of$r$ using eq.(5).
Step3:      Set $T_p = t_r \times \mu_r$
Step4:      For each patch $k$ in $I_T$
                $while(\ r(k) \leq T_p)$
                    $T_{final} = \frac{T_{global} + T_{patch}}{2}$
                    $if\left(\left|T_{final} - T_{global}\right| \leq 1\right)$
                        $Consider\ it\ as\ a\ background\ patch$
                        Exit
                    $else$
                        $temp = T_{final}$
                        $T_{final} = \frac{T_{global} + T_{final}}{2}$
                        $if(T_{final} == temp)$
                            Exit
                        $else$
                            $I_T(i,j) = \begin{cases} I_T(i,j). & if I_T(i,j) < T_{final} \\ 255. & otherwise \end{cases}$
                            Update $var_T^k$
                            Calculate $r(k)$
                        End
                    End
                End
            End
```

3 Experiment

In this section, the proposed binarization method is evaluated. The handwritten images have been obtained from the DIBCO datasets. It is worth mentioning that one important aspect of the DIBCO dataset is that it covers most of the degradations commonly seen in document images. The DIBCO dataset contains the actual images along with the ground truth images for evaluation purposes. We have used the H-DIBCO 2012 [6] and DIBCO 2013 handwritten images [7]. Some of the successful binarization results obtained by the present technique are shown in Fig 2 and Fig. 3

While describing the proposed technique, we have mentioned about certain parameters such as $\mathrm{t_r}$and P. The practical value of $\mathrm{t_r}$ is taken to be 0.3 which is set experimentally. The value of $\mathrm{t_r}$ is kept low to avoid data loss, at the cost of less effective background separation. The value of P is taken as 50, as both larger and smaller values of P produced unsatisfactory results.

We have also combined the foreground-background separation technique proposed with the binarization methods discussed in [1] and [2]. The performances of the combined methods are compared with their respective original methods.

3.1 Evaluation protocol

The binarized output $I_b(i,j)$ is compared with $I_g(i,j)$ and the following evaluation metrics are calculated to measure the effectiveness of the proposed technique.

Mean Square Error (MSE)

$$MSE = \frac{1}{PQ}\sum_{i=1}^{P}\sum_{j=1}^{Q}(\mathrm{I_g(i,j) - I_b(i,j)})^2 \qquad \textbf{(6)}$$

Picture Signal to Noise Ratio (PSNR)

$$PSNR = 10 \log\frac{255^2}{MSE} \qquad \textbf{(7)}$$

For classification of the pixels, some standard parameters like *True Positive*(TP), *True negative* (TN), *False Positive*(FP) and *False Negative*(FN) are also calculated. These values are combined to calculate the Recall, Precision, F-Measure (FM) and Negative Rate Measurement (NRM)

$$Recall = \frac{TP}{TP + FN} \qquad \textbf{(7)}$$

$$Precision = \frac{TP}{TP + FP} \qquad \textbf{(8)}$$

$$FM = \frac{2 \times Recall \times Precision}{Recall + Precision} \qquad \textbf{(9)}$$

$$NRfn = \frac{FN}{FN + TP} \qquad \textbf{(10)}$$

$$NRfp = \frac{FP}{FP + TN} \quad (11)$$

$$NRM = \frac{NRfn + NRfp}{2} \quad (12)$$

3.2 Results and Discussion

The binarized output is compared with the original results of Otsu [1] and LCM [2] on the basis of the evaluated parameters. It is observed that the traditional Otsu method fails to handle the noise appearing in the images during acquisition like poor illumination but when it is combined with the proposed method it responds well (see Fig 3).

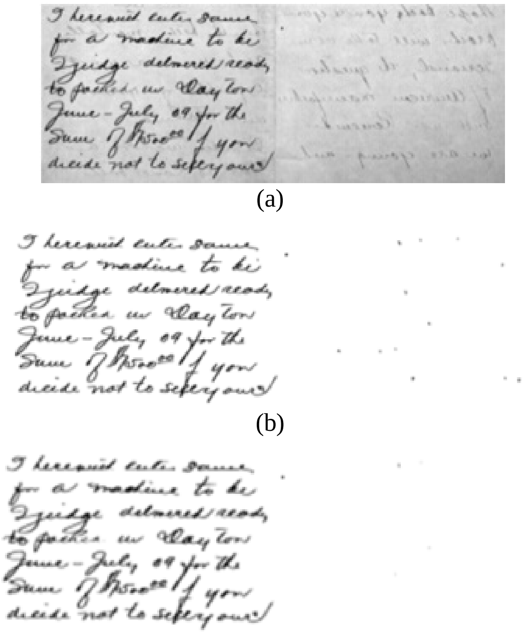
Though the results for the combined LCM are more or less similar to the method reported in [2], the inclusion of the proposed method results in the reduction of data points for the clustering process. LCM is basically a clustering based binarization technique, thus the reduction in number of data points cause reduction in computational time which is one of the key advantages of our method (see Table 2).

Table 1. Performance comparison between original Otsu, LCM and the Otsu, LCM combined with the proposed foreground background separation method

Mean Results on DIBCO 2012 handwritten dataset						
Method	MSE	PSNR	Recall	Precision	F Measure	NRM
Mitianoudis et al.[2]	0.0142	18.7265	0.9077	0.8922	0.8971	0.0502
LCM with our method	0.0155	22.139	0.8919	0.9043	0.8942	0.0601
Otsu et al.[1]	0.0660	20.7649	0.9080	0.6820	0.7434	0.0778
Otsu with our method	0.0435	20.9361	0.8968	0.6796	0.7430	0.0695
Mean Results on DIBCO 2013 handwritten dataset						
Method	MSE	PSNR	Recall	Precision	F Measure	NRM
Mitianoudis et al.[2]	0.0075	21.5742	0.8948	0.9492	0.9209	0.0540
LCM with our method	0.0163	22.15	0.6820	0.9447	0.7705	0.1602
Otsu et al.[1]	0.0178	18.4383	0.7714	0.8844	0.7769	0.1183
Otsu with our method	0.0434	21.284	0.9249	0.6603	0.7294	0.0589

Table 2. Performance comparison in terms of average running time (in second) between original LCM and combined LCM

Method	**DBCO- 2013**	**HDBCO- 2012**
LCM with our method	1998.204 sec.	624.375 sec
Mitianoudis et al.[2]	2875.572 sec	900.318 sec

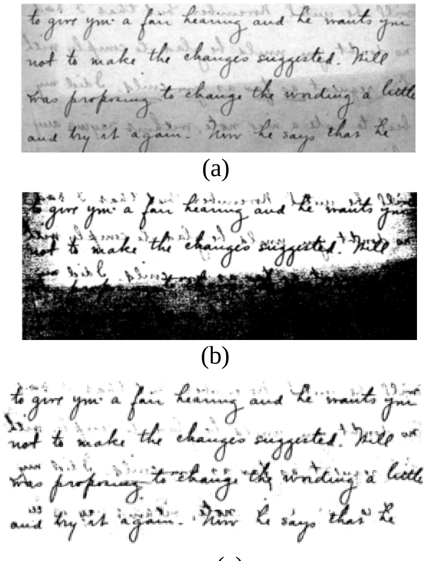


(a)

(b)

(c)

Fig. 2. Shows (a) the original image, (b) LCM output, (c) output of LCM with separation (or combined LCM) respectively.



(a)

(b)

(c)

Fig. 3. Shows (a) original image, (b) Otsu output, (c) output of Otsu with separation (or combined Otsu) respectively.

4 Conclusion

Although binarization of the document images is a challenging task, it helps enormously to design a reliable document image analysis system. Binarization deals with several degradations due to natural and artificial causes. Background estimation as well as elimination is an elementary step which is beneficial for binarization techniques. In this paper, we have presented a binarization technique strongly dependent on background separation for handwritten document images. The proposed background separation technique is based on the ratio of variances which is adaptive in nature. The technique can effectively deal with the problems like uneven illumination and patches. However, it encounters difficulties when the document images have significant amount of bleed through. The comparison of the present technique with some state-of-the-art binarization methodologies reflects the effectiveness of the same. In future, we will look at pixel level robust feature estimation technique which can then be clustered effectively as either foreground or background.

5 References

1. Otsu, Nobuyuki. "A threshold selection method from gray-level histograms."*Automatica* 11.285-296 (1975): 23-27.
2. Mitianoudis, Nikolaos, and Nikolaos Papamarkos. "Document image binarization using local features and Gaussian mixture modeling." *Image and Vision Computing* 38 (2015): 33-51.
3. Gatos, Basilios, Ioannis Pratikakis, and Stavros J. Perantonis. "Adaptive degraded document image binarization." *Pattern recognition* 39.3 (2006): 317-327.
4. Sauvola, Jaakko, and Matti Pietikäinen. "Adaptive document image binarization." *Pattern recognition* 33.2 (2000): 225-236.
5. Valizadeh, Morteza, and Ehsanollah Kabir. "Binarization of degraded document image based on feature space partitioning and classification."*International Journal on Document Analysis and Recognition (IJDAR)* 15.1 (2012): 57-69.
6. Pratikakis, Ioannis, Basilis Gatos, and Konstantinos Ntirogiannis. "ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012)." *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*. IEEE, 2012.
7. Pratikakis, Ioannis, Basilis Gatos, and Konstantinos Ntirogiannis. "ICDAR 2013 document image binarization contest (DIBCO 2013)." *2013 12th International Conference on Document Analysis and Recognition.* IEEE, 2013