**Title:** A non-parametric binarization method based on ensemble of clustering algorithms

Authors:

**1. Suman Kumar Bera**

Affiliation: Dept. of Computer Science and Engineering, Jadavpur University, Kolkata, India.

Email: [berasuman007@gmail.com](mailto:berasuman007@gmail.com)

**2. Soulib Ghosh**

Affiliation: B.E., Dept. of Computer Science and Engineering, Jadavpur University, Kolkata, India

Email: [ghoshsoulib@gmail.com](mailto:ghoshsoulib@gmail.com)

**3. Showmik Bhowmik**

Affiliation: Department of Computer Science and Engineering, Ghani Khan Choudhury Institute of Engineering and Technology (GKCIET), Malda, India

Email: [showmik@gkciet.ac.in](mailto:showmik@gkciet.ac.in)

**4. Ram Sarkar, PhD**

Affiliation: Associate Professor, Dept. of Computer Science and Engineering, Jadavpur University, Kolkata, India

Email: [ramjucse@gmail.com](mailto:ramjucse@gmail.com)

ORCID: 0000-0001-8813-4086

**5. Mita Nasipuri, PhD**

Affiliation: Professor, Dept. of Computer Science and Engineering, Jadavpur University, Kolkata, India

Email: [mitanasipuri@gmail.com](mailto:mitanasipuri@gmail.com)

**Abstract**:

The field of binarization still attracts the researchers especially when we consider the degraded document images. This is evident from the recent Document Image Binarization Competition (DIBCO 2019) where we can see researchers from all over the world participated in this competition. In this paper, we present a novel binarization technique which is found to be capable of handling almost all types of degradations with no parameter tuning. Present method forms an ensemble of three classical clustering algorithms (Fuzzy C-means, K-medoids and K-means++) to group the pixels as foreground or background, after application of a coherent image normalization method. It has been tested on four publicly available datasets, used in DIBCO series, 2016, 2017, 2018 and 2019. Present method gives the promising results for the aforementioned datasets. In addition, this method is the winner of DIBCO 2019.

**Keywords**: Binarization, Document image, Clustering, Ensemble, DB index

1. **Introduction**:

Document image binarization is a process which assemblies the pixels of an input image into two clusters, one represents the background and the other represents the foreground. It is an important preprocessing step on which the performance of many Document image Analysis and Recognition (DAR) system depends. Binarization of degraded document images has drawn the courtesy of researchers due to its extensive range of applicability in many document processing and layout purposes. Some of the examples are restoring historical documents, skew and slant estimation & correction [1] [2], text line separation [3], word recognition [4], document layout analysis (DLA) [5], document form processing [6] and many other optical character recognition (OCR) systems. Thus, without an effective document binarization method, it is quite difficult to develop a complete document processing and recognition system with promising output. A document image generally suffers from different sorts of concerns such as dark spots (appear due to seeping of pen ink from other side of the page), bleed through, presence of seals (appears in many official documents) and fainting of ink (due to rough handling of old documents or due to the use of pen with lighter ink). These degradations basically take the form of noise in the image version of these documents. Besides that, challenges like uneven illuminations of light and dark creases mostly appears due to the erroneous procurement of images. Even the quality and the texture of the document itself may introduce various background level variations. Often more than one of these types of degradations is found in document images. In that case, conventional binarization methods fail to yield desired output and as a result Optical Character Recognition (OCR) also becomes error-prone. Though the research on binarization have been started since many decades, finding a suitable threshold value for bimodal clustering of entire pixel set is still considered as an open challenge for the degraded documents. Document Image Binarization Competition (DIBCO), has been organized in conjunction with International Conference on Document image Analysis and Recognition (ICDAR) to explore the recent efforts in document image binarization since 2009. In the competition of 2019, our method (method 10b in [7]) has achieved the winning place among 24 participating methods.

In this paper, we have presented an efficient binarization technique which is an ensemble of three successors of K-means clustering [8], namely Fuzzy C-Means [9], K-Medoids [10] and K-Means++ [11]. Initially a noise reduction process is employed on the input grayscale image to eliminate the background variation. The resultant image is then fed into the binarization step. The final clusters are obtained by following a voting process. However, in this work, out of three clustering results, we select one as the decider. When a conflict occurs between the other two clustering results for the labeling of a particular pixel, the decider makes the final verdict. For the selection of the decider, we estimate the quality of each clustering result using DB cluster validity index [12] and the one with the lowest value is selected as the decider. Finally, on the resultant binarized image post processing is carried out to preserve the stroke connectivity so that the quality of text regions gets improved. It is to be noted that the proposed method has no parameter to be tuned, which is not so commonplace in binarization methods.

The rest of the paper is organized as follows: next section describes the existing binarization techniques along with their pros and cons. Section III puts forward the proposed method as a chronology of steps. Experimental results are presented and analyzed in section IV which is followed by final epilogue.

1. **Related Work:**

Various methods have been devised to address binarization since decades but all are not robust enough to handle various types of degradations at once. These methods, can broadly be categorized as rule and learning based methods - the former can further be sub-divided as global, local and hybrid thresholding based methods. We give a brief overview of these herein.

Any global approach probes for a single threshold value for the entire image to classify the foreground and background pixels. Otsu [13] develops a global thresholding method that maximizes the inter-class variance and minimizes the intra-class variance of the pixel intensity. Later, Kittler et al. [14] modify the Otsu’s method by minimizing the mean error rate. Though the global methods are very fast and simple to implement, these methods fail in case of diverse background. Conversely, local thresholding methods find individual threshold for each sub-region of an input image depending on its gray-level distribution. Niblack’s method [15] is initial local thresholding method where a pixel-wise threshold is calculated by sliding a rectangular window over the gray image. It is later improved by Sauvola et al. [16]. Though these methods can extract most of the foreground pixels but they fail in low-contrast images. Moreover, these methods have manually tuned parameter. Although all of these preliminary binarization methods are improved to some extend [17], [5], [18], [19] but they are computationally expensive and fail to yield promising results in complex backgrounds and fainted texts. Hybrid methods undergo various steps in order to binarize the document images. The methods [20], [21], [22] which belong to this category pre-preprocess the image for reducing varied background by using the inpainting or Laplacian energy map minimization. The edge-based binarization is proposed by Ramirez et al. [23] and improved by Lelore et al. [24]. These methods use a double-threshold edge detection tactic to find the small details of text pixels. Yet, they fail to distinguish the strong water-spilling noise and incur a high computational cost.

In a learning-based method, a model learns from the supplied data and Ground Truth (GT); thereafter performs the similar task on its own. The recent era tends to adopt deep learning (DL) based methods in binarization due to its classification prowess. A fully convolutional neural network (FCNN) is used in [25] for color image binarization whereas in [26] an unrolled prime dual network (PDNet) is combined with FCNN for the same purpose. Recently, Sheng et al. [27] propose DeepOtsu for binarizing the degraded documents. In [28], Jinyuan et al. frame the problem as image-to-image translation task and use cGAN as a two-step image binarization. Although learning-based methods get success in binarization, they mostly suffer from the availability of large dataset and GT and need to optimize a large set of parameters during training; thereby taking a long processing time.

To summarize, almost all of the existing techniques still have some limitations in terms of their computational cost or performance. Their performance becomes more impoverished when a document gets a combined degradation over the years as shown in Fig. 1. In this study, our objective is to design a robust binarization method that will be effective for handling combined degradations in document images.

|  |  |
| --- | --- |
|  |  |
|  |  |

Fig. 1: Degraded documents taken from DIBCO datasets

1. **Proposed Method:**

This section describes the proposed binarization method in detail. It consists of three sequential steps, each of which consists of further sub-steps. The first step includes pre-processing activities which is comprises of background separation and image normalization. The second section deals with the thresholding. The final section puts forward the post processing steps. The block diagram of this proposed method is displayed in Fig. 2.

|  |
| --- |
|  |
| Fig. 2: Work flow of the proposed method. |

* 1. **Preprocessing:**

In this step, our objective is to obtain a normalized image from the input degraded image by eliminating background variations to a significant extent. For that purpose, initially stroke width is calculated. After that a mask is generated which acts as a superset of the foreground. Then, the background suppressed is generated using inpainting method. Finally, the normalization is performed. The presence of degradations like, bleed through, uneven background, non-uniform illumination, perspective and geometric distortions bring the randomness to the intensity distribution of the affected document image that makes the binarization process challenging. In such cases, background estimation and image normalization deliberate a distribution with nearly bimodal nature to these images, which in turn makes the further procedure of binarization much easier. To handle the aforementioned problem, many background appraisal methods have been developed in the past. For example, Lu et al. [29] estimate the document background surface through an iterative polynomial smoothing procedure. Similarly, Messaoud et al. [30] also apply a smoothening procedure for background estimation. Unlike other methods, Gatos et al. [31] utilize a background surface calculation by interpolating adjacent background intensities, using Sauvola’s binarization result as the mask. In addition to the other methods, inpainting is such an efficient technique of background estimation, which has been developed by the researchers in a recent past and undergone manifold modifications. Bertalmio et al. [32] are among the first group to develop digital inpainting models based on high order Partial Differential Equations (PDEs). They propose an algorithm which automatically fills the gaps in a user-selected region with neighboring information, such as to totally include isophote lines arriving at the regions' boundaries. Shen and Chan [33] develop mathematical models for local inpaintings concerning non-texture images. Based on these mathematical models, Zhang et al. [34] use a shape reconstruction and background restoration based, background layer regeneration method using edge detection and inpainting techniques. They use canny edge detection with morphological dilation trailed by closing operation to procure the inpainting mask. But errors may emanate in this method, when any two isolated noisy points detected by canny edge-detector are joined by morphological filling. This fairly complex and rather time-consuming iterative inpainting method is modified by Ntirogiannis et al. [35]. They propose a relatively fast and efficient inpainting technique using Niblack’s binarization output as the inpainting mask. But for the generation of this inpainting mask, a fixed size window is used. On the other hand, a fixed window size for inpainting mask generation may give rise to some problems, which would become difficult to handle during binarization and post-processing steps. Figure 3 demonstrates an example of such problems faced in the process.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) | (b) | (c) |
|  |  |  |
| (d) | (e) | (f) |

Figure 3: (a,d) Portion of original images (b) Niblack binarized image of (a) using w = 13. (c) Corresponding background-separated image of (a) using (b) as mask. Hollow regions containing background pixels in between the strokes are visible for the strokes, whose width exceeds 13 (e) Niblack binarized image of (d) using w = 60. (f) Corresponding background-separated image of (d) using (e) as mask. Stains and background noises, of width less than 60, are still included in the background-separated image.

The window size of Niblack binarization should conform to the facts as stated by Pai et al. [36] that (i) it manifests the local illumination condition of the image accurately, and (ii) it refrains itself from including foreground pixels exclusively, i.e. at the minimum best, a few background pixels must be confined by the window. A smaller window size for a document image with large stroke width, may position the sliding window entirely within the potential foreground region and thus breaching the second condition. This effectuates in hollow regions containing background pixels in between the strokes, whose width lie above the window-size, as evident from Figure 3. On a counter note, a relatively large window size for a document image with smaller stroke width may include the larger noise stains within the bounds of window. This disregards the first condition and thus local lighting level gets varied within the sliding window, resulting in inclusion of these stains and noises in the background-separated image as evident from Figure 3. For the purpose of inpainting, we propose a dynamic mask generation method based on stroke-width, which takes into consideration the most reasonable stroke-width of a document image, and customize the window-size according to that. Given a degraded document image, an adaptive thresholding is performed firstly to make its corresponding gray-scale image. We have applied canny edge detector [37] on to estimate the average stroke-width . is calculated as the average distance between two consecutive text-pixels in a single row by the following equation (1). Here, defines the frequency of distances and is the number of discrete distances.

|  |  |
| --- | --- |
|  | (1) |

* + 1. **Background Estimation:**

Background estimation and image normalization produce a near-bimodal histogram of the images, thus simplifying the subsequent processes. It reduces the background variation of the input image. Due to the high recall value, we choose the Niblack’s binarization method [9] with a dynamic window to create the mask image. Experimentally, we fix for each side of sliding window. The algorithm calculates the pixel-wise local threshold by using the mean and standard deviation of pixel intensities of a window surrounding it where the weightis taken -0.2 to retain all the text pixels.

|  |  |
| --- | --- |
|  | (2) |

The resultant binarized image is used to generate the coarse background image *()* with the aid of inpainting method [35]. An example of the mask image using dynamic window is shown I n figure 4.

|  |  |
| --- | --- |
|  |  |
| (a) | (b) |

Figure 4: (a) – denotes an input image from DIBCO-2016 dataset. (b) – represents its corresponding mask image.

The method by Ntirogiannis et al. [35] estimates an approximate background surface from the using a mask, . The method uses five passes with different image scanning sequences. Each of the first four passes proffers where, pass, . Considering pixel traversing order shown as directional arrows using left(L), right(R), top(T) and bottom(B), the first four passes traverse as (;T↓B), (;B↑T), (;T↓B) and (;B↑T) respectively. A non-mask pixel (i.e. background) is propagated unscathed in whereas a mask pixel’s (potential foreground) intensity is computed as an average of at most four of its non-mask neighbors. The mask pixel is considered as non-mask in subsequent computation of the same pass. The consolidation is done in fifth pass by taking pixel intensity as minimum in to put in . This helps choosing the intensity value with highest potential to be considered as noise in that neighborhood. Thus, in the areas where there is noisy background, the intensity of corresponding pixels in and are approximately same. Some example of input image and its corresponding background images are put forward in figure 5.

|  |  |
| --- | --- |
|  |  |
| (a) | (b) |
|  |  |
| (c) | (d) |

Figure 5: (a) and (b) are input images from DIBCO dataset, whereas, (c) and (d) are their corresponding background images respectively.

* + 1. **Gray Image Normalization:**

The image normalization is a process of compensation of noises so that the background variations get eliminated. Equation (3) helps here to achieve the normalized image. The and represent the gray image and the estimated background image respectively. Figure 6 displays the normalized image corresponding to the input image. It has been noticed that, the low contrast edge pixels and faded characters in the original image are restored in the normalized image.

|  |
| --- |
|  |

|  |  |
| --- | --- |
|  |  |
| (a) | (b) |
| Fig. 6: (a) Input image and (b) Normalized image | |

* 1. **Thresholding:**

After background separation and image normalization, the resultant image is now devoid of local illumination variations, due to shadows, highlights and noise. As a consequence, the image is henceforth improved, before a thresholding algorithm is applied. The final thresholding for binarization is done with the help of three clustering algorithms: Fuzzy C-means, K-medoids and K-means++. All these algorithms are the reminiscent of K-means clustering with some significant improvements. The basic objective of all clustering algorithms is breaking up the dataset into groups and all of them attempt to minimize the intra cluster distance and maximize inter cluster distance. The thresholding is done here by two steps: pixel labeling and then decision making.

* + 1. **Pixel Labeling:**

In this step, we shall briefly discuss the aforesaid clustering algorithms for the common readers. The normalized image is passed through the three clustering algorithms independently. The resultant binarized images are then fed into a decision-making module to finalize the labeling of the entire pixel set. A brief explanation of the three clustering algorithms are provided.

**Fuzzy C-means (FCM):** It allows data points to be assigned into more than one cluster where each data point has a degree of membership of belonging to each cluster. Alike K-means algorithm, a good cluster does not break in each successive iteration in FCM. FCM aims to minimize the following objective function (equation 4) where, represents the degree to which each element belongs to the cluster. FCM can be formulated according to equation 4. In case of C-means clustering algorithm, we get a degree of membership of a data point to belong in a particular class. A threshold is considered. If the degree of membership is greater than the threshold, then the data point is assigned to that cluster. This is formulated in equation 5.

|  |  |
| --- | --- |
|  | (4) |

|  |  |
| --- | --- |
|  | (5) |

**K-medoids:** In contrast to the K-means algorithm, K-medoids chooses data points as centers (medoids) and can be used with arbitrary distances, while in K-means, the center of a clusters is not necessarily one of the input data points (it is the mean of all the points in the cluster). It is more robust to noise and outliers as compared to K-means because it minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances. The optimization function of the K-medoids can be written as:

|  |  |
| --- | --- |
|  | (6) |

where, denotes the dissimilarity between objects and , is equal to 1 if and only if is assigned to cluster of which is the medoid.

**K-means++:** It solves the initial seeding problem of K-means by allocating the initial cluster center efficiently that increases the chances of forming good clusters. In K-means++ algorithm, first cluster centers are chosen uniformly from the data points, after which each subsequent cluster centers are chosen from the remaining data points with probability proportional to its squared distance from the cluster center closest to that data point. Subsequently, the algorithm follows a similar approach same as K-means which aims to optimize the following function:

|  |  |
| --- | --- |
|  | (7) |

Let, the corresponding binarized images are, and obtained by using the Fuzzy C-Means, K-mediods and K-means algorithms respectively.

* + 1. **Decision Making:**

The three clustering algorithms show their uniqueness to label each pixel. But each of them suffers from a certain limitation. It may lead to a wrong estimation if we rely on only one clustering technique. In case of C-means, there is a weight for every data point in the process of rearrangements. So, the good clusters remain intact throughout the process whereas, only the bad clusters are reallocated for constructing a better one. K-means also suffers from initializing problem. It randomly allocated the initial clusters centers, though, initial cluster centers effect the final outcomes. In case of K-means++, a probabilistic approach is used to predict the initial cluster centers properly. K-means algorithm is not robust and very effective toward outliers. This problem is solved by K-medoids by minimizing the pairwise distances instead of minimizing the total variance. K-medoids is also robust and invariant towards outliers. As the aforementioned three clustering algorithm are modifications of K-means in different aspects and they individually take care of different challenges, so, these algorithms provides complement information. Hence, there is a scope of applying clustering ensemble technique. It is because of the fact that the diversification of the base models is the key essence of forming an ensemble. First, we check the Davis Bouldin (DB) clustering validity index value to set a priority to a clustering algorithm which eventually solves the conflict of labeling a pixel. DB index is an internal evaluation scheme, where the validation of how well the clustering has been done is made using quantities and features inherent to the dataset. Lower value of DB index indicates a good cluster. Mathematically, if is the separation between and clusters which ideally has to be very large and is the within cluster scatter for cluster, then DB index can be defined by equation (8). Firstly, we determine the best cluster based on their DB index value. The best cluster is denoted as decider. In this work, we have proposed an updated version of majority voting based on a decider whose vote gets more weight. To label a pixel as background or foreground, if the other two clustering algorithms conflicts with each other, the decision of the decider is taken as final label. In this case the decider gets the higher weight. On the other hand, if the other two algorithms provide same result but conflicts with the decider, then the complement of the decider is taken as final label. Here, the majority wins over the higher weight. The entire decision making process is described in Algorithm 1 which takes the three binary image (Which are output of the three clustering algorithms) as input and outputs a coarse binarized image .

|  |  |  |
| --- | --- | --- |
|  | where | (8) |

**Algorithm 1: Decision-making**

**Input:** , and

**Output:** Coarse binarized image ()

**Step 1:** Initialize as blank image of input image size.

**Step 2:** Calculating DB index value of the following clustering algorithm.

1. = DB index value of Fuzzy C-Means
2. = DB index value of K-Mediods
3. = DB index value of K-Means++

**Step 3:** For each pixel of do

1. Find individual pixel label for
   1. = label of
   2. = label of
   3. = label of
2. Decider = min(
3. If = then

=

Else

=

End if

|  |
| --- |
|  |

* 1. **Post Process:**

The post-processing is a fairly important process after binarization of document images because it suffices to eliminate noise, ameliorate the quality of the text-regions and also preserve the stroke-connectivity by removal of all isolated pixels and filling of potential breaks, gaps or holes. In this work, we have used shrink and swell filters to exclude persisting noises and enhance the quality of the text-regions. To serve this purpose, entire binary image is scanned and each foreground pixel is scrutinized. A sliding window is considered, where is odd so as to make the definition of a unique center-pixel is possible for the window. Let be the number of background pixels in the sliding window with a foreground pixel as the center-pixel. If , i.e. at least pixels are background pixels in the concerned window, then the foreground pixel is changed to a background pixel. is the threshold for shrink filter and its value can be determined experimentally. A swell filter is used to fill possible discontinuities, gaps or holes in the foreground of a binarized image. It also improves the quality of character strokes. It is functionally similar to the shrink filter. To serve this purpose, the shrinked image is scanned and each background pixel is scrutinized. A sliding window is considered, where is odd so as to make the definition of a unique center-pixel is possible for the window. Let be the number of foreground pixels in the sliding window with a background pixel as the center-pixel. If , i.e. at least pixels are foreground pixels in the concerned window, then the background pixel is changed to a foreground pixel. is the threshold for swell filter and its value can be determined experimentally. Shrink filtering equation (9) is used to remove isolated noise scattered over the background of and swell filtering equation (10) is used to fill possible discontinuities, gaps or holes in the foreground of a binarized image. where and are the numbers of white and black pixels respectively in the sliding window . is the final binarized imaged of our method.

|  |  |
| --- | --- |
|  | (9) |
|  | (10) |

1. **Experimental Section:**

Document Image Binarization Contest (H-DIBCO) is an established and popular forum in the field of document image binarization. Since 2009, DIBCO and H-DIBCO provide benchmarking dataset not only to the participants in the contest, but also to the research community for the purpose of research. These datasets provide handwritten as well as printed document images providing challenging assignments with noises and distortions to be barely readable in its as-is form. Stroke disconnections, imperceptible characters, uneven background, background ink-stains, smudges, non-uniform illumination, perspective and geometric distortions, degradation by imaging artifacts, etc. are some of the such challenges that the DIBCO datasets offer to the researchers. In addition, DIBCO datasets for the contest provide a common platform of comparison yardsticks with ground truth, as created by a forum of human referees after applying different binarization techniques on the image dataset. The datasets considered for the present work are H-DIBCO 2016 [38], DIBCO 2017 [39], H-DIBCO 2018 [40] and DIBCO 2019 [7]. Few selected images of DIBCO 2019 dataset and their corresponding binarized forms obtained using proposed method, are shown in Fig. 10. *Moreover, this method is the winner of DIBCO 2019 competition.* To asses and compare our method with the state-of-the-art techniques, we have used four standard evaluation metrics: (a) *F-Measure (FM),* (b) *pseudo F-Measure (),* (c) *Peak Signal-to-Noise Ratio (PSNR), and* (d) *Distance Reciprocal Distortion (DRD)*. The F measure is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test. PSNR is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Finally, DRD measures the visual distortion among two images. We have used the evaluation tool provided by the DIBCO 2019 competition. Our proposed method is compared with the winner method of the individual competitions along with Otsu’s [13], Sauvola’s [16], Niblack’s [15] and recently published K-Means [41] based binarization methods. The detailed comparison study is shown in Table I. Some output of the proposed method is also shown from figure 7-10. These figure contains outputs of DIBCO 2016, 2017, 2018 and 2019 respectively.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) | (b) | (c) |
|  |  |  |
| (d) | (e) | (f) |

Figure 7: (a), (b) and (c) are sample images from DIBCO’16 dataset, whereas, (d), (e) and (f) are their corresponding binary output respectively.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) | (b) | (c) |
|  |  |  |
| (d) | (e) | (f) |

Figure 8: (a), (b) and (c) are sample images from DIBCO’17 dataset, whereas, (d), (e) and (f) are their corresponding binary output respectively.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) | (b) | (c) |
|  |  |  |
| (d) | (e) | (f) |

Figure 9: (a), (b) and (c) are sample images from DIBCO’18 dataset, whereas, (d), (e) and (f) are their corresponding binary output respectively.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) | (b) | (c) |
|  |  |  |
| (d) | (e) | (f) |

Figure 10: (a), (b) and (c) are sample images from DIBCO’19 dataset, whereas, (d), (e) and (f) are their corresponding binary output respectively.

A meticulous visual inspection of the output binarized images (in Fig. 7-10) obtained by the proposed technique reveals that the proposed method can eliminate almost all the background noises and retrieve the faint characters as well. The corresponding quantitative comparison shown in Table I proves that proposed method keeps up well with the persistent challenges of different datasets when compared to the other methods. Furthermore, the present method is the winner of ICDAR 2019, DIBCO competition. Thus, we can safely claim that proposed method is robust enough to provide reliable binarization results for the combined degraded documents.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table I: Performance comparison of the proposed method with several methods on the datasets of DIBCO series | | | | | | | |
| **Method** | **Evaluation Metrics** | | | | | | |
|  |  | | | |  |  |
| **H-DIBCO 2016** | | | | | | | |
| Rank 1st | 87.61 | 91.28 | | | | 18.11 | 5.21 |
| Otsu | 86.61 | 88.67 | | | | 17.80 | 5.56 |
| Sauvola | 82.52 | 86.85 | | | | 16.42 | 7.49 |
| Niblack | 48.57 | 48.68 | | | | 8.01 | 82.24 |
| K-Means | 89.08 | 90.26 | | | | 18.48 | 4.47 |
| **Proposed** | **90.43** | **91.66** | | | | **18.94** | **3.51** |
| **DIBCO 2017** | | | | | | | |
| Rank 1st | **91.04** | | **92.86** | | **18.28** | | **3.40** |
| Otsu | 82.52 | | 86.85 | | 16.42 | | 7.49 |
| Sauvola | 86.61 | | 88.67 | | 17.80 | | 5.56 |
| Niblack | 51.17 | | 51.47 | | 7.72 | | 59.49 |
| K-Means | 77.82 | | 80.14 | | 13.87 | | 15.37 |
| **Proposed** | 83.38 | | 89.43 | | 15.45 | | 6.71 |
| **H-DIBCO 2018** | | | | | | | |
| Rank 1st | **88.34** | | **90.24** | | **19.11** | | **4.92** |
| Otsu | 51.45 | | 53.03 | | 9.74 | | 59.06 |
| Sauvola | 67.81 | | 74.08 | | 13.78 | | 17.69 |
| Niblack | 41.18 | | 41.39 | | 6.79 | | 99.46 |
| K-Means | 51.52 | | 53.54 | | 9.75 | | 59.09 |
| **Proposed (2nd)** | 76.84 | | 83.58 | | 15.31 | | 9.58 |
| **DIBCO 2019** | | | | | | | |
| Rank 1st **(Proposed)** | **72.87** | | | **72.15** | **14.48** | | **16.24** |
| Otsu | 53.77 | | | 54.42 | 05.01 | | 35.42 |
| Sauvola | 47.42 | | | 48.53 | 07.44 | | 27.12 |
| Niblack | 5151 | | | 53.86 | 10.54 | | 31.05 |
| K-Means | 65.05 | | | 66.55 | 09.04 | | 24.38 |
| AKMB | 63.59 | | | 67.40 | 13.42 | | 23.58 |

1. **Conclusion:**

In this paper, we propose a non-parametric binarization technique that takes the advantages of three clustering algorithms namely, Fuzzy C-means, K-medoids and K-means++. It initially performs a noise reduction step on the input grayscale image to reduce the background variation. Though, the normalized image is then coarsely binarized separately by three said clustering algorithms but the final pixel-labeling is decided by a voting process based on their DB index values. A suitable post-processing is finally applied to preserve the stroke width of binarized image. This harmonization proves to be a reasonably satisfactory adaptive binarization technique which can deal with severe challenges and is robust to diverse situations in an image. Our technique proves its mettle in handling severe instances of various kinds of degradations in a generic domain of document images. We have used the most recent dataset of images provided by DIBCO, containing benchmarking and challenging images with an amalgamation of various types of degradations. The resultant images are substantiated with outcomes of other traditional and classic techniques in literature, leading competing techniques in the contest along with the ground-truth of the original images, in terms of an ensemble of measures. Although the scope of the proposed technique handles amalgamation of various types of degradations quite efficiently, yet it has some limitations. The proposed technique efficiently handles images possessing non-distinguishable text and background intensity, but its efficiency is somewhat curtailed in case of images possessing very small isolated foreground dots, whose size is comparable with salt-and-pepper noises in the image. These dots remain on the verge of becoming eliminated in the final binarized image and thus results in a relatively low score. This can, however, be improved if a more efficient post-processing step is involved.

**Reference:**

[1] A. A. Mascaro, G. D. C. Cavalcanti, and C. A. B. Mello, “Fast and robust skew estimation of scanned documents through background area information,” *Pattern Recognit. Lett.*, vol. 31, no. 11, pp. 1403–1411, Aug. 2010, doi: 10.1016/j.patrec.2010.03.016.

[2] E. Kavallieratou, N. Fakotakis, and G. Kokkinakis, “Slant estimation algorithm for OCR systems,” *Pattern Recognit.*, vol. 34, no. 12, pp. 2515–2522, Dec. 2001, doi: 10.1016/S0031-3203(00)00153-9.

[3] E. Ahmadi, Z. Azimifar, M. Shams, M. Famouri, and M. J. Shafiee, “Document image binarization using a discriminative structural classifier,” *Pattern Recognit. Lett.*, vol. 63, pp. 36–42, Jul. 2015, doi: 10.1016/j.patrec.2015.06.008.

[4] S. Bhowmik, S. Malakar, R. Sarkar, and M. Nasipuri, “Handwritten Bangla word recognition using elliptical features,” in *2014 International Conference on Computational Intelligence and Communication Networks*, 2014, pp. 257–261.

[5] R. F. Moghaddam and M. Cheriet, “AdOtsu: An adaptive and parameterless generalization of Otsu’s method for document image binarization,” *Pattern Recognit.*, vol. 45, no. 6, pp. 2419–2431, 2012.

[6] S. Ghosh, R. Bhattacharya, S. Majhi, S. Bhowmik, S. Malakar, and R. Sarkar, “Textual Content Retrieval from Filled-in Form Images,” in *Communications in Computer and Information Science*, 2019, vol. 1020, pp. 27–37, doi: 10.1007/978-981-13-9361-7\_3.

[7] I. Pratikakis *et al.*, “ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019 ),” *2019 15th IAPR Int. Conf. Doc. Anal. Recognit.*, no. November 2019, 2019, doi: 10.1109/ICDAR.2019.00249.

[8] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A k-means clustering algorithm,” *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 28, no. 1, pp. 100–108, 1979.

[9] J. C. Bezdek, R. Ehrlich, and W. Full, “FCM: The fuzzy c-means clustering algorithm: Computer s & Geosciences, 10, no. 2–3, 191–203.” 1984.

[10] R. T. Ng and J. Han, “Efficient and Effective Clustering Methods for Spatial Data Mining,” *Proc. 20th Int. Conf. Very Large Data Bases*, pp. 144–155, 1994.

[11] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027–1035.

[12] D. L. Davies, “L., Bouldin, DW,(1979): A cluster separation measure,” *IEEE Trans. Pattern, Anal. Mach. Inetell*, vol. 1, pp. 224–227.

[13] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[14] J. Kittler and J. Illingworth, “On threshold selection using clustering criteria,” *IEEE Trans. Syst. Man. Cybern.*, no. 5, pp. 652–655, 1985.

[15] W. Niblack, *An introduction to digital image processing*. Strandberg Publishing Company, 1985.

[16] J. Sauvola and M. Pietikäinen, “Adaptive document image binarization,” *Pattern Recognit.*, vol. 33, no. 2, pp. 225–236, 2000.

[17] K. Khurshid, I. Siddiqi, C. Faure, and N. Vincent, “Comparison of Niblack inspired Binarization methods for ancient documents,” in *Document Recognition and Retrieval XVI*, 2009, vol. 7247, p. 72470U.

[18] G. Lazzara and T. Géraud, “Efficient multiscale Sauvola’s binarization,” *Int. J. Doc. Anal. Recognit.*, vol. 17, no. 2, pp. 105–123, 2014, doi: 10.1007/s10032-013-0209-0.

[19] P. Jana, S. Ghosh, R. Sarkar, and M. Nasipuri, “A Fuzzy C-Means Based Approach Towards Efficient Document Image Binarization,” in *2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR)*, 2017, pp. 1–6.

[20] S. Bhowmik, R. Sarkar, B. Das, and D. Doermann, “GiB: A game theory inspired binarization technique for degraded document images,” *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1443–1455, 2019, doi: 10.1109/TIP.2018.2878959.

[21] B. Su, S. Lu, and C. L. Tan, “Binarization of historical document images using the local maximum and minimum,” in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, 2010, pp. 159–166.

[22] B. Bataineh, S. N. H. S. Abdullah, and K. Omar, “An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows,” *Pattern Recognit. Lett.*, vol. 32, no. 14, pp. 1805–1813, 2011.

[23] M. A. Ramírez-Ortegón, E. Tapia, L. L. Ramírez-Ramírez, R. Rojas, and E. Cuevas, “Transition pixel: A concept for binarization based on edge detection and gray-intensity histograms,” *Pattern Recognit.*, vol. 43, no. 4, pp. 1233–1243, 2010.

[24] T. Lelore and F. Bouchara, “Fair: A fast algorithm for document image restoration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2039–2048, 2013.

[25] C. Tensmeyer and T. Martinez, “Document image binarization with fully convolutional neural networks,” in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 1, pp. 99–104.

[26] K. R. Ayyalasomayajula, F. Malmberg, and A. Brun, “PDNet: Semantic segmentation integrated with a primal-dual network for document binarization,” *Pattern Recognit. Lett.*, vol. 121, pp. 52–60, 2019.

[27] S. He and L. Schomaker, “DeepOtsu: Document enhancement and binarization using iterative deep learning,” *Pattern Recognit.*, vol. 91, pp. 379–390, 2019.

[28] J. Zhao, C. Shi, F. Jia, Y. Wang, and B. Xiao, “Document image binarization with cascaded generators of conditional generative adversarial networks,” *Pattern Recognit.*, vol. 96, p. 106968, 2019.

[29] S. Lu, B. Su, and C. L. Tan, “Document image binarization using background estimation and stroke edges,” *Int. J. Doc. Anal. Recognit.*, vol. 13, no. 4, pp. 303–314, Dec. 2010, doi: 10.1007/s10032-010-0130-8.

[30] I. Ben Messaoud, H. Amiri, H. El Abed, and V. Märgner, “New binarization approach based on text block extraction,” in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 2011, pp. 1205–1209, doi: 10.1109/ICDAR.2011.243.

[31] B. Gatos, I. Pratikakis, and S. J. Perantonis, “Adaptive degraded document image binarization,” *Pattern Recognit.*, vol. 39, no. 3, pp. 317–327, 2006.

[32] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image inpainting,” in *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, 2000, pp. 417–424, doi: 10.1137/1.9780898717877.ch6.

[33] T. F. Chan and J. Shen, “Mathematical models for local nontexture inpaintings,” *SIAM J. Appl. Math.*, vol. 62, no. 3, pp. 1019–1043, Jul. 2002, doi: 10.1137/S0036139900368844.

[34] L. Zhang, A. M. Yip, M. S. Brown, and C. L. Tan, “A unified framework for document restoration using inpainting and shape-from-shading,” *Pattern Recognit.*, vol. 42, no. 11, pp. 2961–2978, Nov. 2009, doi: 10.1016/j.patcog.2009.03.025.

[35] K. Ntirogiannis, B. Gatos, and I. Pratikakis, “A combined approach for the binarization of handwritten document images,” *Pattern Recognit. Lett.*, vol. 35, no. 1, pp. 3–15, 2014, doi: 10.1016/j.patrec.2012.09.026.

[36] Y. T. Pai, Y. F. Chang, and S. J. Ruan, “Adaptive thresholding algorithm: Efficient computation technique based on intelligent block detection for degraded document images,” *Pattern Recognit.*, vol. 43, no. 9, pp. 3177–3187, Sep. 2010, doi: 10.1016/j.patcog.2010.03.014.

[37] P. Bao, L. Zhang, and X. Wu, “Canny edge detection enhancement by scale multiplication,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1485–1490, 2005.

[38] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, “ICFHR2016 handwritten document image binarization contest (H-DIBCO 2016),” in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2016, pp. 619–623.

[39] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, “ICDAR2017 competition on document image binarization (DIBCO 2017),” in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 1, pp. 1395–1403.

[40] I. Pratikakis, K. Zagoris, P. Kaddas, and B. Gatos, “ICFHR 2018 competition on handwritten document image binarization (HDIBCO 2018),” 2018.

[41] P. Jana, S. Ghosh, S. K. Bera, and R. Sarkar, “Handwritten document image binarization: An adaptive K-means based approach,” in *2017 IEEE Calcutta Conference, CALCON 2017 - Proceedings*, 2018, vol. 2018-Janua, doi: 10.1109/CALCON.2017.8280729.