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

Suman Kumar Bera [1] · Soulib Ghosh [1] · Showmik Bhowmik [2] · Ram Sarkar [1] · Mita Nasipuri [1]

Abstract

Binarization of document images still attracts the researchers especially when degraded document images are considered. 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 without any parameter tuning. Present method is based on 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 promising results for the aforementioned datasets. In addition, this method is the winner of DIBCO 2019 competition.

Keywords Binarization · Document image · Clustering · Ensemble · DB index · DIBCO

✉ Suman Kumar Bera
   berasuman007@gmail.com

   Soulib Ghosh
   ghoshsoulib@gmail.com

   Showmik Bhowmik
   showmik@gkciet.ac.in

   Ram Sarkar
   ramjucse@gmail.com

   Mita Nasipuri
   mitanasipuri@gmail.com

[1] Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

[2] Department of Computer Science and Engineering, Ghani Khan Choudhury Institute of Engineering and Technology (GKCIET), Malda, India

1 Introduction

Document image binarization is a process which splits the pixels of an input image into two clusters, one represents the background and the other represents the foreground. A gray scale image consists of pixels, whose value ranges between 0 and 255. On the other hand, the pixel value of any binary image is either 0 or 1. Binarization is a process of converting a grayscale image into a binary image. It is an important preprocessing step that influences the performance of many Document image Analysis and Recognition (DAR) systems. Binarization of degraded document images has drawn the attention of researchers due to the challenges involved and also its extensive range of applicability in many document image processing purposes. Some of the examples are restoring historical documents, skew and slant correction, text line separation, word recognition, document layout analysis (DLA), document form processing and many other optical character recognition (OCR) systems [1, 8, 10, 14, 20, 31, 32, 36, 46, 51]. Proper binarization is always necessary to ensure the effectiveness of the later steps. In other words, poor binarization will make the later steps error prone. Thus, without an effective document binarization method, it is quite difficult to develop an efficient and complete DAR system. There are various challenges associated with this research domain. A document image generally suffers from different types of degradations such as dark spots, bleed through (seepage of ink from the other side of the page), presence of seals (in official documents) and fading of ink (due to aging and rough handling of documents or due to the use of ink of lighter shade). Some of these degradations take the form of noise in the image version of these documents. Besides that, challenges like irregular background, uneven illuminations of light and dark creases mostly appear due to erroneous digitization of documents. Even the quality and the texture of the document page itself may introduce various background level variations. Often more than one of such degradations are found in the document images. In that case, conventional binarization methods fail to provide a satisfactory output, and thus, OCR systems also become error-prone. Though the research on binarization has been started long back, finding a suitable threshold value for bimodal clustering of the entire pixel set of a degraded document image is still considered as an open challenge. Absence of a good binarization technique capable of handling all kinds of degradations and the diverse application domains has motivated us to design an effective method for such purpose.

In this paper, we have presented a new binarization technique which is an ensemble of three successors of K-means clustering algorithm [16], namely Fuzzy C-Means [7], K-Medoids [33] and K-Means++ [2]. 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 results of the other two clustering algorithms in labeling a particular pixel, the decider makes the final verdict. For the selection of the decider, we estimate the quality of each clustering result using the Davis Bouldin (DB) cluster validity index [12] and the one with the lowest value is selected as the decider. Finally, on the resultant binarized image, some 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 needs no parameter tuning, which is not so commonplace in binarization methods. With this method, we participated in the Document Image Binarization Competition (DIBCO) 2019. Since 2009, this competition is organized in every alternate year in conjunction with the International Conference on Document image Analysis and Recognition (ICDAR) to explore the recent efforts in document

image binarization. In 2019 competition of DIBCO, our method *(method 10b in* [43]*)* ranked 1st among the 24 participating methods. The contributions of our paper are as follows:

- We have proposed a novel approach to form an ensemble of three clustering algorithms based on the Davis Bouldin (DB) cluster validity index.
- The method does not require any manually tuned parameter.
- The proposed approach is validated on four most recent publicly available DIBCO datasets, namely, DIBCO 2016, DIBCO 2017, DIBCO 2018 and DIBCO 2019.
- The impressive outcomes over state-of-the-art methods empirically prove the effectiveness of the proposed method on dealing with the said challenges.

The rest of the paper is organized as follows: next section describes the existing binarization techniques along with their pros and cons. Section 3 puts forward the proposed method as a chronology of steps. Experimental results are presented and analyzed in section 4 which is followed by final epilogue.

2 Related work

Various methods have been devised to address binarization of document images since decades but all are not robust enough to handle various types of degradations / noise in a single pass. These methods can be categorized as threshold-based 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 methods herein.

Any global approach probes for a single threshold value for the entire image to classify the foreground and background pixels. Otsu [38] 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. [22] 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 [34] is the premier local thresholding based method where a pixel-wise threshold is calculated by sliding a rectangular window over the gray scale image. It is later improved by Sauvola et al. [47]. 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 [19, 21, 23] but they are computationally expensive and fail to yield promising results in complex backgrounds and faded texts. Hybrid methods undergo various steps in order to binarize the document images. The methods [5, 9, 48] which belong to the hybrid category pre-preprocess the image for reducing varied background by using inpainting or Laplacian energy map minimization. Edge-based binarization is proposed by Ramirez et al. [44] 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. Besides, there are few other approaches in the literature. In [25], Lu et al. have used a contrast enhancement aided local thresholding based binarization technique. The local threshold is determined by the mean of foreground and background gray values, which are calculated by the frequency of the gray values. In [15], Guo et al. have come up with a nonlinear edge

preserving diffusion technique for document image binarization. In [45], Rani et al. have proposed a binarization method which uses the contrast feature to compute the threshold value.

To deal with different types of challenges researchers have also relied upon background separation technique. Many background appraisal methods have been developed in the past. For example, Lu et al. [28] estimate the document background surface through an iterative polynomial smoothing procedure. Similarly, Messaoud et al. [6] have applied a smoothening procedure for background estimation. Unlike other methods, Gatos et al. [13] utilize a background surface calculation by interpolating adjacent background intensities, using Sauvola's binarization result as the mask. Besides these methods, inpainting has emerged as an efficient technique of background estimation, which is developed by the researchers in the recent past and has undergone manifold modifications. Shen and Chan [11] develop mathematical models for local inpainting concerning non-texture images. Based on these mathematical models, Zhang et al. [52] use a shape reconstruction and background restoration method through edge detection and inpainting. They use Canny edge detection followed by morphological dilation and closing operation to procure the inpainting mask. But errors may arise 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]. In the paper [35] Ntirogiannis et al. 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 [9], which would become difficult to handle during binarization and the post-processing steps. Figure 1 demonstrates examples of such problems faced in the process. In Fig. 1(c), hollow regions containing background pixels in between the strokes are visible for the strokes, whose width exceeds 13 whereas in Fig. 1(e), stains and background noises, of width less than 60, are still included in the background-separated image.

Fig. 1 **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, **e** Niblack binarized image of **d** using w = 60 and **f** Corresponding background-separated image of **d** using **e** as mask

The window size of Niblack binarization should conform to the facts as stated by Pai et al. [39] which are: (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, thereby breaching the second condition. This effectuates in hollow regions containing background pixels in between the strokes, whose width lies above the window-size, as evident from Fig. 1. 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 the 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 Fig. 1.

In a learning-based method, a model learns from the supplied data and Ground Truth (GT) that acts as a training sample. After that, a trained model is generated which is further tested on the test data. Deep learning based approaches are also used in various other application domains such as video object tracking [27] and video object segmentation [29, 30] etc. 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 [49] for color image binarization whereas in [3] an unrolled prime dual network (PDNet) is combined with FCNN for the same purpose. Recently, Sheng et al. [17] propose DeepOtsu for binarizing the degraded documents. In [53], 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 obtain 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, most of the existing techniques still have some limitations in terms of their computational cost and/or performance. Their performance becomes more impoverished when a document gets a combined degradation over the years as shown in Fig. 2. In this paper, our objective is to design a robust binarization method that will be effective for handling combined degradations found in the document images.

Fig. 2 Sample degraded documents taken from DIBCO datasets

3 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 comprise of background separation and image normalization steps. 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. 3. In Fig. 3, D is the decider label that corresponds to the cluster label with minimum DB index value. For each pixel, L_x and L_y are two variables that indicate the labels other than the decider D, whereas P_{lbl} denotes the final pixel label obtained after using the thresholding algorithm.

3.1 Preprocessing

In this step, our objective is to obtain a normalized image from an 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 image 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 brings the randomness to the intensity distribution of the affected document image, and 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 make the further procedure of binarization much easier. For background estimation, we have used the inpainting technique as described in [35]. We have already discussed the fixed window size problem for generating a mask image. Due to that we have relied on the idea of using 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 accordingly. Given a degraded document image I_d, an adaptive thresholding is performed firstly to make its corresponding gray-scale image I_g. We have applied the Canny edge detector [4] on I_g to estimate the average stroke-width SW_{avg}. SW_{avg} is calculated as the average distance between two consecutive text-pixels in a single row. The calculation of the average stroke width is demonstrated using Eq. 1. Here, f_i defines the frequency of distances d_i and N is the number of discrete distances.
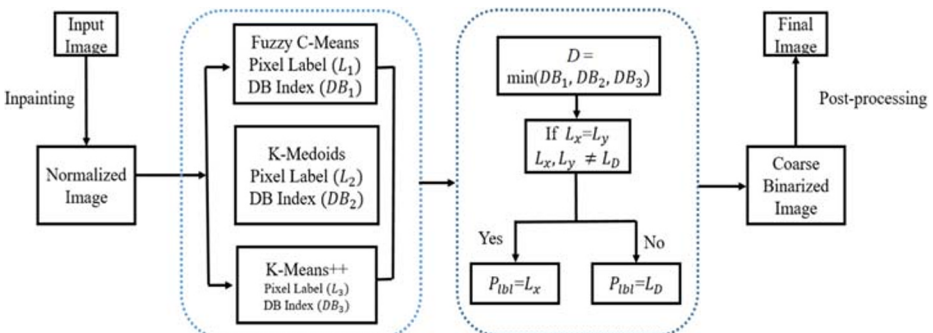

Fig. 3 Work flow of the proposed method

$$SW_{avg} = \frac{\sum_{i=1}^{N}(d_i \times f_i)}{\sum_{i=1}^{N} f_i} \qquad (1)$$

3.1.1 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 with a dynamic window W_d to create the mask image I_{mask}. Experimentally, we fix $3 * SW_{avg}$ for each side of the sliding window W_d. The algorithm calculates the pixel-wise local threshold (Th) by using the mean and standard deviation of pixel intensities of a window surrounding it where the weight w is taken -0.2 to retain all the text pixels [50]. The calculation of the threshold is formulated in Eq. 2. Here $\mu(x,y)$ *and* $\sigma(x,y)$ represent the mean and standard deviation of the pixels residing within the sliding window with (x,y) as center pixel respectively.

$$Th(x,y) = w*\sigma(x,y) + \mu(x,y) \qquad (2)$$

The resultant binarized image I_{mask} is used to generate the coarse background image (I_{bg}) with the aid of inpainting method [35]. An example of the mask image using dynamic window is shown in Fig. 4.

The method by Ntirogiannis et al. [35] estimates an approximate background surface I_{bg} from the I_g using a mask, I_{mask} . The method uses five passes with different image scanning sequences. Each of the first four passes proffers $P_p(r, c)$ where, pass, $p \in [1, 4]$. Considering pixel traversing order shown as directional arrows using left(L), right(R), top(T) and bottom(B), the first four passes traverse as ($\overrightarrow{LR}$;T↓B), ($\overrightarrow{LR}$;B↑T), (RL;T↓B) and (RL;B↑T) respectively. A non-mask pixel (i.e. background) is propagated unscathed in $P_p(r, c)$ 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 $P_p(r, c) \forall p \in [1, 4]$ to put in I_{bg} . 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 I_g and I_{bg} are approximately same. Some example of input image and its corresponding background images are put forward in Fig. 5.

(a)   (b)

Fig. 4 **a** A sample input image from DIBCO-2016 dataset and **b** Corresponding mask image
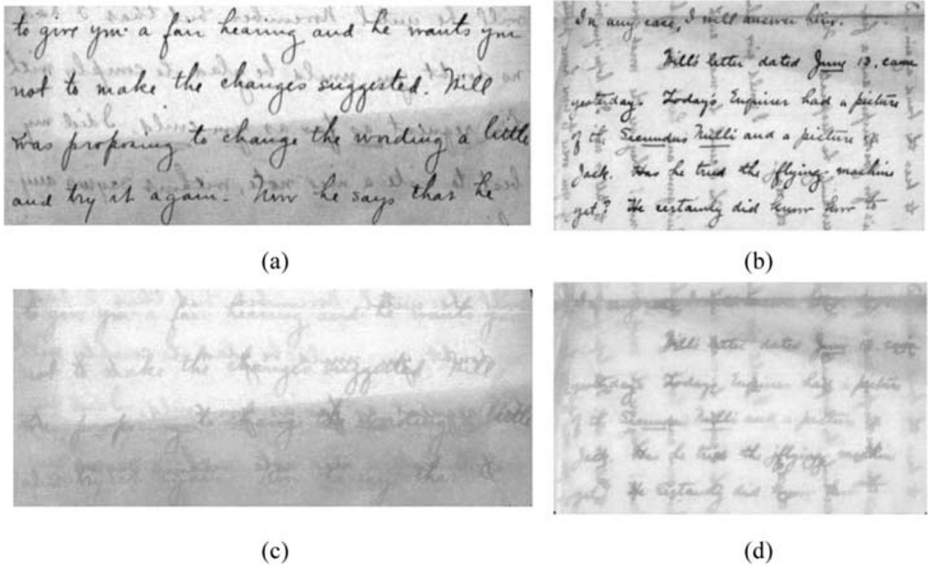
(a)   (b)

(c)   (d)

Fig. 5 **a** and **b** are input images from DIBCO dataset, whereas, **c** and **d** are their corresponding background images respectively

3.1.2 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 $I_n(x, y)$. The I_g and I_{bg} represent the gray image and the estimated background image respectively. Fig. 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.

$$I_n(x,y) = \begin{cases} \dfrac{255 * I_g(x,y)}{I_{bg}(x,y)}, & if \ I_g(x,y) < I_{bg}(x,y) \wedge I_{bg}(x,y) > 0 \\ 255, & otherwise \end{cases} \tag{3}$$

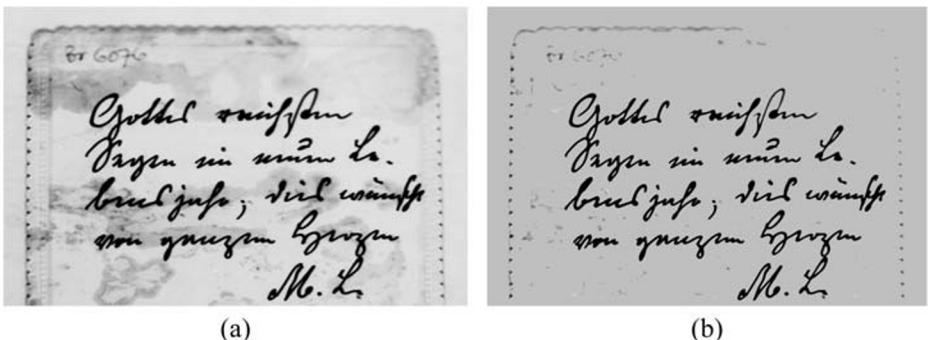(a)   (b)

Fig. 6 **a** Input image and **b** Corresponding normalized image

3.2 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 reminiscent of K-means clustering algorithm 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.

3.2.1 Pixel labeling

In this step, we briefly discuss the aforesaid clustering algorithms for the common readers. The normalized image I_n 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 is 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 (Eq. 4) where, $w_{ij} \in [0,1]$ represents the degree to which each element x_i belongs to the cluster c_j. FCM can be formulated according to Eq. 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 α is considered. If the degree of membership is greater than the threshold, then the data point is assigned to that cluster. The objective function is formulated in Eq. 4. In Eq. 4, x_i is the ith data point and c_j indicates cluster center of jth cluster. Besides, n and c denote the number of data points and number of clusters respectively. w_{ij}^m denotes the weight associated for each data point. In the case of Fuzzy C-means clustering algorithm, we obtain a membership value for each data point. Equation 5 provides the procedure to associate each data point to a particular cluster based on the membership value μ_{i1}, which is the membership value of x_i for belonging to cluster 1. If μ_{i1} is greater than some threshold α, then the data point belongs to cluster 1 else cluster 2.

$$\arg\,min \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^m \left|\left|x_i - c_j\right|\right|^2 \tag{4}$$

$$x_i \in \begin{cases} \text{cluster}_1, & \mu_{i1} \geq \alpha \\ \text{cluster}_2, & \mu_{i1} < \alpha \end{cases} \tag{5}$$

K-medoids: In contrast to the K-means algorithm, K-medoids chooses data points as cluster 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 shown in Eq. 6:

$$\text{arg}min \sum_{i=1}^{n} \sum_{j=1}^{c} d_{ij} * z_{ij} \tag{6}$$

where, d_{ij} denotes the dissimilarity between objects x_i and x_j, $z_{ij} \in [0,1]$ is equal to 1 if and only if x_j is assigned to cluster of which x_i 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 shown in Eq. 7:

$$\text{minimize} \sum_{i=1}^{n} \sum_{j=1}^{c} \left|\left|x_i - c_j\right|\right|^2 \tag{7}$$

Let, the corresponding binarized images are I_{cc}, I_{km} and I_{kp} obtained by using the Fuzzy C-Means, K-mediods and K-means algorithms respectively. Next, we discuss our decision making process.

3.2.2 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 the case of C-means, there is a weight for every data point in the process of rearrangement. 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 allocates the initial clusters' centers, though, initial cluster centers effect the final outcomes. In the case of K-means++, a probabilistic approach is used to predict the initial cluster centers properly, but is not robust and ineffective 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 algorithms are the modified versions of the K-means algorithm, and they individually take care of the different challenges, so these algorithms provide complementary information. Hence, there is a scope for applying a 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 clustering evaluation scheme, where the validation of how well the clustering has been done is made using the quantities and the features inherent to the dataset. Lower value of DB index indicates a good cluster. Mathematically, if A_{ij} is the separation between i^{th} and j^{th} clusters which ideally has to be very large and S_i is the within cluster scatter for i^{th} cluster, then DB index can be defined by Eq. (8).

$$DB = \frac{1}{C} \sum_{i=1}^{C} d_i, \text{where } d_i = \max\left\{\frac{S_i + S_j}{A_{ij}}, i \neq j\right\} \tag{8}$$

Firstly, we determine the best cluster based on their DB index value. The best cluster is denoted as the decider. In this work, we have proposed a majority voting scheme based on a decider whose vote gets more weight in decision making. To label a pixel as background or foreground, if the other two clustering algorithms conflict with each other, the decision of the decider is taken as the final label. In this case, the decider gets the higher weight. On the other hand, if the other two algorithms provide same result but conflict with the decider, then the complement of the decider is taken as the 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 images (i.e., outputs of the three clustering algorithms) as input and yields a coarse binarized image I_{cb}. In Algorithm 1, D is the decider label that corresponds to the cluster label with minimum DB index value. L_x and L_y are two variables that indicate the labels other than the decider label. P_{lbl} denotes the final pixel label obtained using this algorithm. I_{cm}, I_{km} and I_{kp} are the images obtained after applying the Fuzzy C-means, K- Mediods and K-means++ algorithms respectively.

ALGORITHM 1: DECISION-MAKING

Input: I_{cc}, I_{km} and I_{kp}

Output: Coarse binarized image (I_{cb})

Step 1: Initialize I_{cb} as blank image of input image size.

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

   a. DB_1 = DB index value of Fuzzy C-Means
   b. DB_2= DB index value of K-Mediods
   c. DB_3= DB index value of K-Means++

Step 3: For each pixel P_i of I_{cb} do

   a. Find corresponding pixel labels for P_i
     i) L_1= label of I_{cc}
     ii) L_2= label of I_{km}
     iii) L_3= label of I_{kp}
   b. Decider $D = \min(DB_1, DB_2, DB_3)$
   c. Let, L_D=label of D and $x, y \in 1,2,3$
   d. If $L_x=L_y$ and $L_x, L_y \neq L_D$ then
       $P_{lbl}=L_x$
    Else
       $P_{lbl}=L_D$
    End if

    End off
    End for

3.3 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. For each foreground pixel, a $w_{sh} \times w_{sh}$ sliding window is considered, where w_{sh} is odd so as to make the definition of a unique

center-pixel is possible for the window. Let N_{back} be the number of background pixels in the sliding window with a foreground pixel as the center-pixel. If $N_{back} > t_{sh}$, i.e., at least t_{sh} pixels are background pixels in the concerned window, then the foreground pixel is changed to a background pixel. t_{sh} is the threshold for shrink filter and its value is 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 $w_{sw} \times w_{sw}$ sliding window is considered, where w_{sw} is odd so as to make the definition of a unique center-pixel possible for the window. Let N_{fore} be the number of foreground pixels in the sliding window with a background pixel as the center-pixel. If $N_{fore} > t_{sw}$, i.e. at least t_{swell} pixels are foreground pixels in the concerned window, then the background pixel is changed to a foreground pixel. t_{sw} is the threshold for swell filter and its value can be determined experimentally. Shrink filter as shown in Eq. (9) is used to remove isolated noise scattered over the background of I_{cb} and swell filter as shown in Eq. (10) is used to fill possible discontinuities, gaps or holes in the foreground of a binarized image, where N_{back} and N_{fore} are the numbers of white and black pixels respectively in the sliding window W_d. I_b is the final binarized image of our method.

$$I_b(x,y) = 1 \ \forall \ N_{back} > t_{sh} \tag{9}$$

$$I_b(x,y) = 0 \ \forall \ N_{fore} > t_{sw} \tag{10}$$

Though there are few parameters in the post processing step, these are not manually tuned. The parameters are specific throughout the experimentation. Besides, the parameters are not dataset specific. The parameters used at this step depend on the average character height, which are calculated using the connected component analysis as mentioned in [13].

3.4 Experimental section

The entire experimental section is divided into four subsections. The first subsection includes a detailed explanation of the datasets used in this work, whereas, the second subsection deals with the evaluation metrics. Third subsection contains results obtained using the proposed method. The final subsection reports the comparison of our method with some past methods.

3.5 Dataset description

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 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 proposed method is validated with four standard datasets which are H-DIBCO 2016 [40], DIBCO 2017 [41], H-DIBCO 2018 [42] and DIBCO 2019 [43]. In each year DIBCO provides various degraded document images along with their ground truths. In 2016, they have published 10 handwritten images consisting of many difficult challenges like bleed through, faint foreground etc. Whereas, there are 20 images in 2017 that contain 10 handwritten and 10 printed document images. In the 2018 dataset, there are a total 10 handwritten document images, and in 2019 dataset there are 20 document images. The datasets consist of 60 handwritten and printed degraded document images. Among them, few sample images and their corresponding binarized forms obtained using proposed method, are shown in Figs. 8, 9, 10 and 11.

3.6 Evaluation metric

To assess 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 (F_{ps}),* (c) *Peak Signal-to-Noise Ratio (PSNR), and* (d) *Distance Reciprocal Distortion (DRD).* To further assess the performance of the proposed method, we have considered additional metrics that include precision, recall, true positive, true negative, false positive and false negative [37]. A *true positive* (*TP*) is an outcome where the model correctly predicts the positive class. Similarly, *true negative* (*TN*) is an outcome where the model correctly predicts the negative class. A *false positive* (*FP*) is an outcome where the model incorrectly predicts the positive class and *false negative* (*FN*) is an outcome where the model *incorrectly* predicts the *negative* class. A diagram is given in Fig. 7 to show this concept in a tabular way.

Precision (*Pre*) is defined as the ratio of true positives and the number of true positives plus the number of false positives. Whereas, *recall (Rec)* is ratio of true positives and the number of true positives plus the number of false negative. Recall expresses the ability to find all relevant instances in a dataset, and precision expresses the proportion of the data points our model says was relevant actually were relevant. Precision and recall are formulated in Eqs. 11 and 12 respectively.

Fig. 7 Illustration of true positive, false positive, true negative and false negative

$$Pre = TP/(FP + TP) \tag{11}$$

$$Rec = TP/(FN + TP) \tag{12}$$

The $F-measure$ (FM) is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test. F-measure is expressed in Eq. 13.

$$FM = 100 * (2 * Pre * Rec)/(Pre + Rec) \tag{13}$$

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. In the binarization domain, the noise of the output binarized image is calculated with respected to the ground truth. PSNR is formulated in Eq. 14. In the Eq. 14, MSE and R represent the mean square error and the maximum fluctuation in the input image data type respectively.

$$PSNR = 10 \, log_{10} \left(\frac{R^2}{MSE} \right) \tag{14}$$

Finally, DRD measures the visual distortion among the output binarized image and the ground truth. DRD is formulated in Eq. 15. Where, $NUBN$ is the number of the non-uniform (not all black or white pixels) 8×8 blocks in the ground truth image, and DRD_k is the distortion of the $k-th$ flipped pixel. For better understanding of the calculation of DRD, we refer to paper [26].

$$DRD = \frac{\sum_{k=1}^{s} DRD_k}{NUBN} \tag{15}$$

Higher values of PSNR, F-measure and pseudo F-measure indicate better outcomes, whereas, low value of DRD represents superior result.

4 Results

For the entire experimentations, we have used the evaluation tool provided by the DIBCO 2019 competition, publicly available on their website. Few examples of the output binary images after applying the proposed approach are shown in Figs. 8, 9, 10 and 11. Figures 8, 9, 10 and 11 display the output binary images taken from the year 2016, 2017, 2018 and 2019 respectively.

A meticulous visual inspection of the output binarized images (in Figs. 8, 9, 10 and 11) obtained by the proposed technique reveals that it can eliminate almost all the background noises and retrieve the faint characters as well. For better assessment of the proposed approach, we have added outcomes in terms of true positive, true negative, false positive and false negative. Year wise results are represented in Figs. 12(a)–(d). The values represent the average values obtained over the entire images of dataset of a particular year. It is to be noted that the true negative in all the cases are significantly higher than the others. The reason is in a typical document image; the proportion of background pixels is much higher compared to foreground pixels. The method is able to capture both foreground and background pixels. However, due to the large amount of background pixels, the volume of detected background pixels are much higher than the foreground pixels. As a result, the value of the true negative is very high in all the cases.
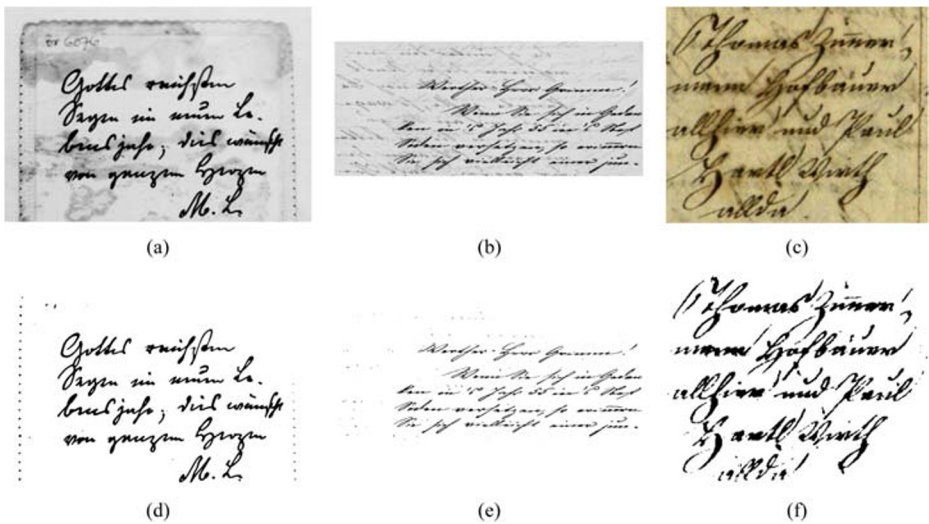
Fig. 8 **a**, **b** and **c** are sample images from DIBCO'16 dataset, whereas, **d**, **e** and **f** are their corresponding binary output images respectively

Moreover, precision, recall, F-measure, Pseudo F-measure, PSNR and DRD values obtained by the proposed approach are provided in Table 1. The table contains year wise values for the aforementioned metrics. In the context of binarization, recall indicates a sensitivity index that measures the amount of foreground pixels retrieved in the binarization process. High recall value indicates that the method is able to capture the foreground pixels properly. On the other hand, low precision value indicates that the method has considered some background pixels as foreground. So, a good binarization method must have both the precision and recall value very

Fig. 9 **a**, **b** and **c** are sample images from DIBCO'17 dataset, whereas, **d**, **e** and **f** are their corresponding binary output images respectively
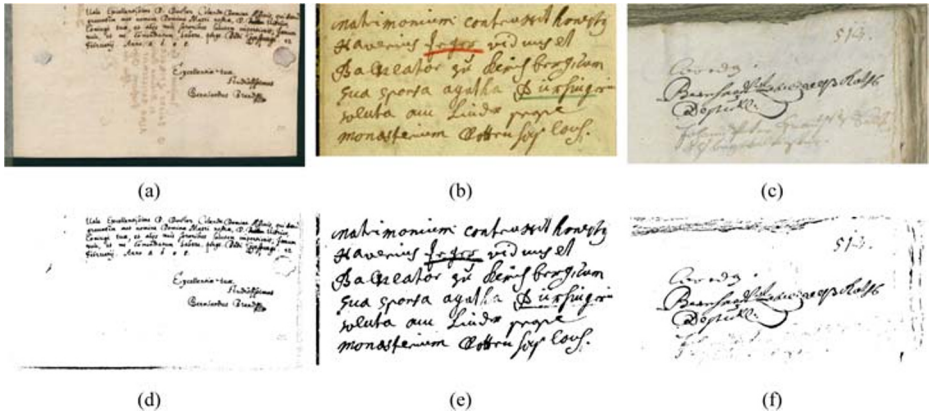
Fig. 10   **a**, **b** and **c** are sample images from DIBCO'18 dataset, whereas, **d**, **e** and **f** are their corresponding binary output images respectively

high (close to 1). Keeping in mind the extreme challenges of the document images, it is evident from Table 1 that the proposed method has achieved outstanding performance in most of the datasets. In the DIBCO 2019 competition, there were 24 methods developed by the researchers from the different corners of the globe. It is worth mentioning that the proposed approach outperformed all the other participating methods and stood first in that competition.

The last column of Table 1 contains the execution time of the proposed approach in various datasets. We have provided the execution time of our algorithm for all the datasets. DIBCO
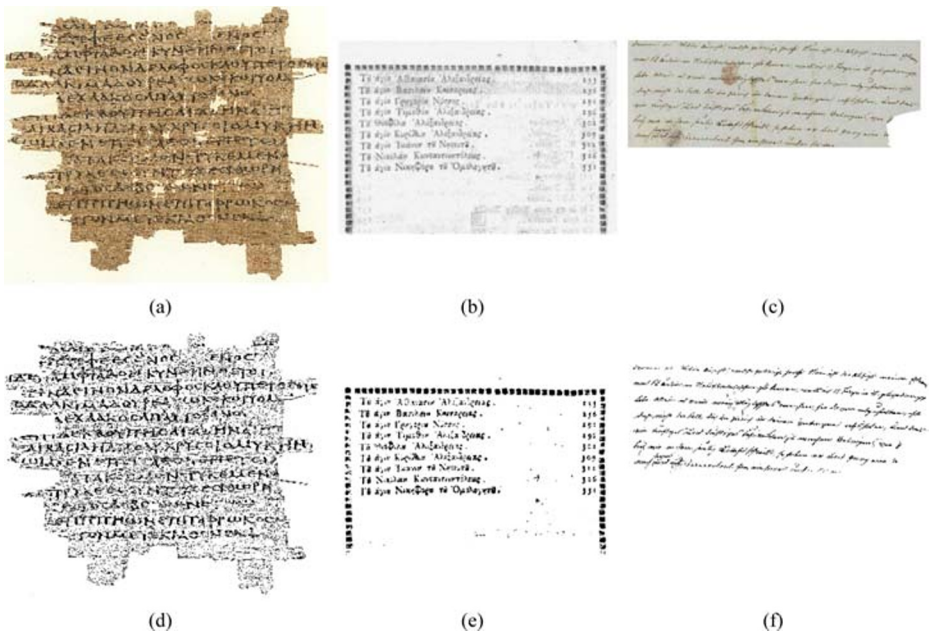

Fig. 11   **a**, **b** and **c** are sample images from DIBCO'19 dataset, whereas, **d**, **e** and **f** are their corresponding binary output images respectively
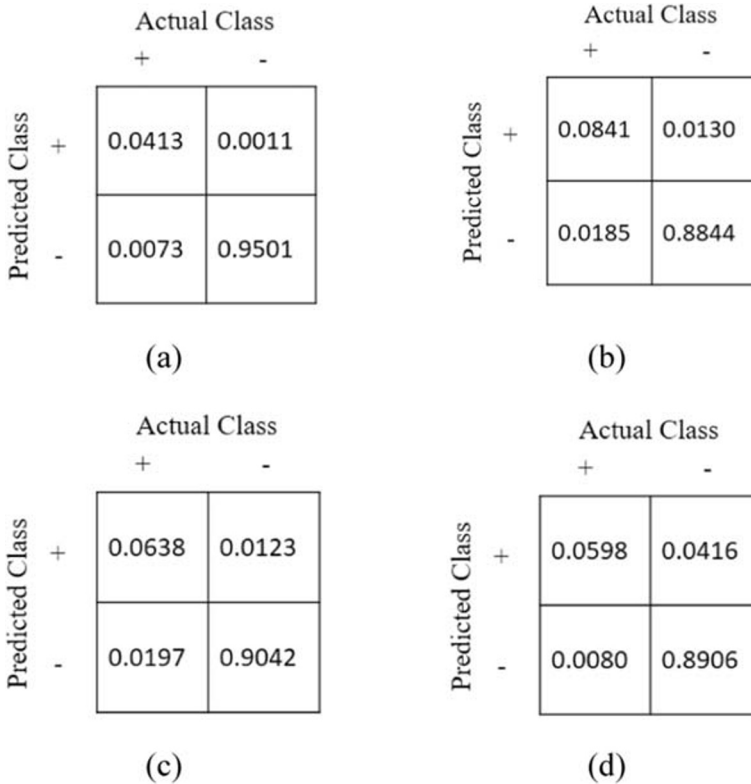
Fig. 12 Year wise true positive, true negative, false positive and false negative value obtained by our proposed binarization technique. (**a**), (**b**), (**c**) and (**d**) represent results for the year 2016, 2017, 2018 and 2019 respectively. In the context of binarization, positive sign signifies foreground and negative indicates background

2016 and 2018 datasets contain 10 images whereas, DIBCO 2017 and 2019 datasets have 20 images each. Due to this, execution time for DIBCO 2017 and 2019 is higher than the other two. The machine specification is as follows: Processor – Intel(R) Core (TM) i5-7200U CPU @ 2.50 GHz 2.71 GHz, RAM – 8 GB and System Type – 64-bit Operating System, ×64-based processor.

4.1 Comparison

Our proposed method is compared with the winner method of the individual competitions along with Otsu's [38], Sauvola's [47], Niblack's [34] and recently published K-Means [18]

Table 1 Binarization results obtained using the proposed method on various the datasets

Metric Dataset	Precision	Recall	F-measure	Pseudo F-Measure	PSNR	DRD	Execution time (in Sec)
DIBCO 2016	0.9651	0.8508	90.43	91.66	18.94	3.51	401.99
DIBCO 2017	0.8578	0.8289	83.38	89.43	15.45	6.71	941.17
DIBCO 2018	0.8091	0.7521	76.84	83.58	15.31	9.58	476.89
DIBCO 2019	0.6527	0.8771	72.87	72.15	14.48	16.24	1720.21

based binarization methods. Otsu is a global thresholding method that aims to find a global threshold for the entire image by reducing intra class variation and increasing inter class variance. On the other hand, Sauvola and Niblack are local thresholding methods that probe for a threshold value for a sub image. The method proposed in [18] has adopted a binarization technique based K-means clustering algorithm followed by a background separation technique. Our proposed method is additionally compared with two other techniques in DIBCO 2016 dataset. The two techniques are based on edge preserving diffusion Guo et al. [15] and contrast based approach Rani et al. [45]. The detailed comparison study is shown in Table 2. The corresponding quantitative comparison shown in this table proves that the proposed method keeps up well with the persistent challenges of different datasets when compared to the other methods. As the proposed method outperforms the method described in [18], we can claim that the proposed ensemble technique performs better than K-means based method. *Furthermore, the present method is the winner of ICDAR 2019, DIBCO competition among the 24 participating methods across the globe.* Thus, we can safely claim that the proposed method is robust enough to provide reliable binarization results on the degraded document images.

Table 2 Performance comparison of the proposed method with different methods on the datasets of DIBCO series

Method	Evaluation Metrics			
	FM	F_{ps}	PSNR	DRD
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
Rani et al. [45]	83.5	–	**19.38**	–
Guo et al. [15]	88.51	90.46	18.42	4.13
Proposed	**90.43**	**91.66**	18.94 (2nd)	**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	51.51	53.86	10.54	31.05
K-Means	65.05	66.55	09.04	24.38
AKMB [18]	63.59	67.40	13.42	23.58

5 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 the 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.

References

1. Ahmadi E, Azimifar Z, Shams M, Famouri M, Shafiee MJ (Jul. 2015) Document image binarization using a discriminative structural classifier. Pattern Recogn Lett 63:36–42. https://doi.org/10.1016/j.patrec.2015.06.008
2. Arthur D, Vassilvitskii S (2007) k-means++: the advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp 1027–1035
3. Ayyalasomayajula KR, Malmberg F, Brun A (2019) PDNet: semantic segmentation integrated with a primal-dual network for document binarization. Pattern Recogn Lett 121:52–60
4. Bao P, Zhang L, Wu X (2005) Canny edge detection enhancement by scale multiplication. IEEE Trans Pattern Anal Mach Intell 27(9):1485–1490
5. Bataineh B, Abdullah SNHS, Omar K (2011) An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows. Pattern Recogn Lett 32(14):1805–1813
6. Ben Messaoud I, Amiri H, El Abed H, Märgner V (2011) New binarization approach based on text block extraction. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR). IEEE, Beijing, China, pp 1205–1209. https://doi.org/10.1109/ICDAR.2011.243
7. Bezdek JC, Ehrlich R, Full W (1984) FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences 10(2–3):191–203
8. Bhowmik S, Malakar S, Sarkar R, Nasipuri M (2014) Handwritten Bangla word recognition using elliptical features. In: 2014 International Conference on Computational Intelligence and Communication Networks, pp 257–261
9. Bhowmik S, Sarkar R, Das B, Doermann D (2019) GiB: a game theory inspired binarization technique for degraded document images. IEEE Trans Image Process 28(3):1443–1455. https://doi.org/10.1109/TIP.2018.2878959

10. Bera SK, Chakrabarti A, Lahiri S, Barney Smith EH (2019) Sarkar R (2019) Normalization of unconstrained handwritten words in terms of Slope and Slant Correction. Pattern Recognit Lett 128:488–495. https://doi.org/10.1016/j.patrec.2019.10.025

11. Chan TF, Shen J (Jul. 2002) Mathematical models for local nontexture inpaintings. SIAM J Appl Math 62(3):1019–1043. https://doi.org/10.1137/S0036139900368844

12. Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans Pattern, Anal Mach Inetell 1: 224–227

13. Gatos B, Pratikakis I, Perantonis SJ (2006) Adaptive degraded document image binarization. Pattern Recogn 39(3):317–327. https://doi.org/10.1016/j.patcog.2005.09.010

14. Ghosh S, Bhattacharya R, Majhi S, Bhowmik S, Malakar S, Sarkar R (2019) Textual Content Retrieval from Filled-in Form Images. Communications in Computer and Information Science 1020:27–37. https://doi.org/10.1007/978-981-13-9361-7_3

15. Guo J, He C, Zhang X (2019) Nonlinear edge-preserving diffusion with adaptive source for document images binarization. Appl Math Comput 351:8–22. https://doi.org/10.1016/j.amc.2019.01.021

16. Hartigan JA, Wong MA (1979) Algorithm AS 136: a k-means clustering algorithm. J R Stat Soc: Ser C: Appl Stat 28(1):100–108

17. He S, Schomaker L (2019) DeepOtsu: document enhancement and binarization using iterative deep learning. Pattern Recogn 91:379–390

18. Jana P, Ghosh S, Bera SK, Sarkar R (2018) Handwritten document image binarization: An adaptive K-means based approach. In: 2017 IEEE Calcutta Conference, CALCON 2017 - Proceedings, vol. 2018-Janua. https://doi.org/10.1109/CALCON.2017.8280729

19. Jana P, Ghosh S, Sarkar R, Nasipuri M (2018) A fuzzy C-means based approach towards efcient document image binarization. 2017 9th International Conference on Advances in Pattern Recognition (ICAPR). IEEE, Bangalore, India, pp 332–337

20. Kavallieratou E, Fakotakis N, Kokkinakis G (2001) Slant estimation algorithm for OCR systems. Pattern Recogn 34(12):2515–2522. https://doi.org/10.1016/S0031-3203(00)00153-9

21. Khurshid K, Siddiqi I, Faure C, Vincent N (2009) Comparison of Niblack inspired Binarization methods for ancient documents. In: Document Recognition and Retrieval XVI, vol 7247, p 72470U

22. Kittler J, Illingworth J (1985) On threshold selection using clustering criteria. IEEE Trans. Syst. Man. Cybern. 5:652–655

23. Lazzara G, Géraud T (2014) Efficient multiscale Sauvola's binarization. Int J Doc Anal Recognit 17(2): 105–123. https://doi.org/10.1007/s10032-013-0209-0

24. Lelore T, Bouchara F (2013) Fair: a fast algorithm for document image restoration. IEEE Trans Pattern Anal Mach Intell 35(8):2039–2048

25. Lu D, Huang X, Sui LX (2018) Binarization of degraded document images based on contrast enhancement. Int J Doc Anal Recognit 21(1–2):123–135. https://doi.org/10.1007/s10032-018-0299-9

26. Lu H, Kot AC, Shi YQ (2004) Distance-reciprocal distortion measure for binary document images. IEEE Signal Process Lett 11(2):228–231

27. Lu X, Ma C, Ni B, Yang X, Reid I, Yang MH (2018) Deep regression tracking with shrinkage loss. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol 11218. LNCS, pp 369–386. https://doi.org/10.1007/978-3-030-01264-9_22

28. Lu S, Su B, Tan CL (2010) Document image binarization using background estimation and stroke edges. Int J Doc Anal Recognit 13(4):303–314. https://doi.org/10.1007/s10032-010-0130-8

29. Lu X, Wang W, Ma C, Shen J, Shao L, Porikli F (2019) See more, know more: Unsupervised video object segmentation with co-attention siamese networks. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol 2019-June, pp 3618–3627. https://doi.org/10.1109/CVPR.2019.00374

30. Lu X, Wang W, Shen J, Tai YW, Crandall D, Hoi SCH (2020) Learning video object segmentation from unlabeled videos. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 8960–8970

31. Kundu S, Paul S, Bera SK, Abraham A (2020) Sarkar R (2020) Text-line extraction from handwritten document images using GAN. Expert Systems with Applications 140:112916

32. Moghaddam RF, Cheriet M (2012) AdOtsu: an adaptive and parameterless generalization of Otsu's method for document image binarization. Pattern Recogn 45(6):2419–2431

33. Ng RT, Han J (1994) Efficient and Effective Clustering Methods for Spatial Data Mining. In: Proc. 20th Int. Conf. Very Large Data Bases, pp 144–155

34. Niblack W (1986) An introduction to digital image processing. Prentice Hall, pp 115–116

35. Ntirogiannis K, Gatos B, Pratikakis I (2014) A combined approach for the binarization of handwritten document images. Pattern Recogn Lett 35(1):3–15. https://doi.org/10.1016/j.patrec.2012.09.026

36. Olszewska JI (2015) Active contour based optical character recognition for automated scene understanding. Neurocomputing 161:65–71

37. Olszewska JI (2019) Designing transparent and autonomous intelligent vision systems. In: ICAART 2019 - Proceedings of the 11th International Conference on Agents and Artificial Intelligence, vol 2, pp 850–856. https://doi.org/10.5220/0007585208500856
38. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66
39. Pai YT, Chang YF, Ruan SJ (Sep. 2010) Adaptive thresholding algorithm: efficient computation technique based on intelligent block detection for degraded document images. Pattern Recogn 43(9):3177–3187. https://doi.org/10.1016/j.patcog.2010.03.014
40. Pratikakis I, Zagoris K, Barlas G, Gatos B (2016) ICFHR2016 handwritten document image binarization contest (H-DIBCO 2016). In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp 619–623
41. Pratikakis I, Zagoris K, Barlas G, Gatos B (2017) ICDAR2017 competition on document image binarization (DIBCO 2017). In: 2017 14th IAPR international conference on document analysis and recognition (ICDAR), vol 1, pp 1395–1403
42. Pratikakis I, Zagoris K, Kaddas P, Gatos B (2018) ICFHR 2018 competition on handwritten document image binarization (HDIBCO 2018)
43. Pratikakis I et al (2019) "ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019)," 2019 15th IAPR Int. Conf. Doc. Anal. Recognit., no. November 2019. https://doi.org/10.1109/ICDAR.2019.00249
44. Ramírez-Ortegón MA, Tapia E, Ramírez-Ramírez LL, Rojas R, Cuevas E (2010) Transition pixel: a concept for binarization based on edge detection and gray-intensity histograms. Pattern Recogn 43(4):1233–1243
45. Rani U, Kaur A, Josan G (2020) A New Contrast Based Degraded Document Image Binarization. In: Cognitive Computing in Human Cognition, pp 83–90. https://doi.org/10.1007/978-3-030-48118-6_8
46. Sahlol AT, Abd Elaziz M, Al-Qaness MAA, Kim S (2020) Handwritten Arabic optical character recognition approach based on hybrid whale optimization algorithm with neighborhood rough set. IEEE Access 8:23011–23021
47. Sauvola J, Pietikäinen M (2000) Adaptive document image binarization. Pattern Recogn 33(2):225–236
48. Su B, Lu S, Tan CL (2010) Binarization of historical document images using the local maximum and minimum. In: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, pp 159–166
49. Tensmeyer C, Martinez T (2017) Document image binarization with fully convolutional neural networks. In: 2017 14th IAPR international conference on document analysis and recognition (ICDAR), vol 1, pp 99–104
50. Trier OD, Jain AK (1995) Goal-directed evaluation of Binarization methods. IEEE Trans Pattern Anal Mach Intell 17(12):1191–1201. https://doi.org/10.1109/34.476511
51. Zahoor S, Naz S, Khan NH, Razzak MI (2020) Deep optical character recognition: a case of Pashto language. J Electron Imaging 29(2):23002
52. Zhang L, Yip AM, Brown MS, Tan CL (Nov. 2009) A unified framework for document restoration using inpainting and shape-from-shading. Pattern Recogn 42(11):2961–2978. https://doi.org/10.1016/j.patcog.2009.03.025
53. Zhao J, Shi C, Jia F, Wang Y, Xiao B (2019) Document image binarization with cascaded generators of conditional generative adversarial networks. Pattern Recogn 96:106968