Engineering Applications of Artificial Intelligence 92 (2020) 103672

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Modified Sauvola binarization for degraded document images

Amandeep Kaur [a], Usha Rani [b,*], Gurpreet Singh Josan [c]

[a] Department of Computer Science and Technology, Central University of Punjab, Bathinda, Punjab, India
[b] University College Ghanaur, Patiala, Punjab, India
[c] Department of Computer Science, Punjabi University, Patiala, Punjab, India

ARTICLE INFO

Keywords:
Binarization techniques
Degraded document images
Historical documents
Stroke width transform

ABSTRACT

The binarization of historical documents is a difficult job due to the presence of many degradations. Many existing local binarization techniques use certain manually adjusted parameters. The output of these techniques is much dependent on the value of these parameters. One of such parameters is window size which is kept fixed for the whole text image. The fixed window size will not be able to perform well for images having variable stroke widths and text sizes. The proposed binarization technique (Modified Sauvola) is the modification of state of art Sauvola's binarization technique. It automatically computes window size dynamically across the image pixel to pixel using the stroke width transform (SWT). This led to reduction in number of manually adjusted parameters. The results are compared with the nine existing techniques using the quantitative measures: FM, PSNR, NRM, MPM, and DRD. The results show that the proposed method outperforms existing methods for images having variable stroke widths and text sizes.

1. Introduction

Historical documents contain vast knowledge in various fields such as medicine, politics, science and literature, etc. These documents are localized in some libraries, museums etc. and need to make approachable for everyone so that people from different fields can be benefited. Also, these documents get deteriorated with passage of time so some way to preserve these documents is needed. With the advent of information technology, it became possible to digitize these documents. The benefits we get with digitization are less need of physical storage, better preservation, and easy access to all people anywhere, at any time using internet facility. Today, libraries all over the world are busy in digitizing books, journals or other readable materials. Digitization of the documents is done using some acquisition system such as scanner or camera. The images we get after digitations are not in the machine-readable form. This huge collection of images will not be of much use if the information they contain cannot be edited, searched or indexed. It will become very difficult to retrieve the desired information from the huge collection of images. Document image analysis system is the set of algorithms and techniques that enable document images into machine-readable text. Optical Character Recognition (OCR) is well known document image analysis system. Binarization is one of the most important stage in such systems. It is the process that transforms grayscale or colored images into bi-level form, generally background is represented in white color and foreground in black color. Binarization of the images is done to reduce the computational overload and increase the system efficiency. Binarization itself depends upon the quality of document images. There is not any problem to binarize good quality document images, but in case of degraded document images it is very difficult to obtain proper binary image. Historical documents degrade with passage of time due to environmental factors, bad storage conditions, aging and chemical properties of paper and ink. Due to this, documents have wrinkles, yellow coloring of paper, ink bleed through, fading of text, smear, and stains type of degradations. The results of next stages of the data analysis system that is segmentation and recognition are affected by the binarization quality. The image binarization has been studied for many years and several binarization techniques have been proposed and developed in the literature. The image binarization methods are mainly categorized as global and local. Global methods use a single threshold value to binarize the whole image. Global methods are fast and perform well for images having bimodal histogram. The documents containing uneven illumination, stains, ink bleed, and uneven contrast are not well processed by global methods because they do not have bimodal histogram. Otsu (1979), Pun (1980, 1981), Johannsen and Bille (1982), Kapur et al. (1985), Kittler and Illingworth (1985), Abutaleb (1989) and Brink and Pendock (1996) are well known global methods.

The local binarization methods (Bernsen, 1986; Niblack, 1986; Wellner, 1993; Sauvola and Pietaksinen, 2000; Yang and Yan, 2000; Wolf and Jolion, 2003; Do et al., 2005; Gatos et al., 2006; Bardley and Roth, 2007; Shafait et al., 2008; Khurshid et al., 2009; Zhou et al., 2009; Kawano et al., 2009; Lu and Tan, 2010; Singh et al., 2011,

* Corresponding author.
E-mail addresses: aman_k2007@hotmail.com (A. Kaur), usha_gupta7@yahoo.co.in (U. Rani), josangurpreet@pbi.ac.in (G.S. Josan).

https://doi.org/10.1016/j.engappai.2020.103672
Received 23 December 2019; Received in revised form 6 March 2020; Accepted 17 April 2020
Available online xxxx
0952-1976/

2012) are better approaches for degraded document images which estimate a different threshold value for each pixel depending upon its neighborhood. These techniques perform well for badly degraded images, but are slow because computation of image features from local neighborhood is done for each image pixel. In the literature, most of the local adaptive techniques are sliding window based, which calculate the threshold for every pixel by moving a rectangular window pixel to pixel over the whole image. Bernsen (1986), Niblack (1986), Wellner (1993), Sauvola and Pietaksinen (2000), Wolf and Jolion (2003), Bardley and Roth (2007), Shafait et al. (2008), Khurshid et al. (2009), Singh et al. (2011, 2012) and Natarajan and Sreedevi (2017) are local binarization methods which utilize mean, standard deviation or local contrast in the neighborhood of the pixel under consideration to determine the threshold value. To reduce computational complexity, Bardley and Roth (2007), Shafait et al. (2008) and Singh et al. (2011) utilized the integral image concept to determine statistical measures such as mean, standard deviation and variance. Using the concept of integral images, the time to compute mean or standard deviation over a rectangular area becomes independent of the size of that area. Local techniques need to determine the size of the local window manually for each image under consideration to get better binarization result. There are local adaptive binarization techniques which compute window size automatically using the features of the image under consideration. Do et al. (2005) proposed adaptive Niblack and adaptive Sauvola methods which estimate parameters k and local window size in original methods adaptively. The window size is estimated from the height of characters in both the adaptive Niblack and adaptive Sauvola methods. In the adaptive Niblack, the parameter k is computed from the ratio of number of black pixels to the number of white pixels in small rectangular blocks. For the adaptive Sauvola, the value of k is estimated from the local and global values of k computed using Niblack method. Gatos et al. (2006), Zhou et al. (2009) and Kawano et al. (2009) techniques based on background estimation and subtraction, compute the local window size using the height and width of the characters in the image. Gatos et al. (2006) estimated background surface by interpolating the background intensities in to the foreground areas which are calculated by using Sauvola's binarization method (Sauvola and Pietaksinen, 2000). Zhou et al. (2009) computed the background using the Laplacian-Gauss algorithm. Kawano et al. (2009) used the median filter to estimate background. Yang and Yan (2000), Lu and Tan (2010), Hejdam et al. (2011), Su et al. (2010, 2013), Van and Lee (2014) and Shukla et al. (2014) are local adaptive methods that used the stroke width of the characters in the input image to compute the local window size. The local window size in these techniques remain fixed for the whole image to compute the threshold value. Hadjadj et al. (2016) proposed the binarization technique named ISauvola that combines Su et al. (2010) method and Sauvola's method (Sauvola and Pietaksinen, 2000). They computed high contrast pixel image using (Su et al., 2010) and bilevel image of the input image using (Sauvola and Pietaksinen, 2000). The final binarized image is obtained by sequential combination of high contrasted pixel image and the bilevel image. Kim et al. (2002), Oh et al. (2005) and Valizadeh and Kabir (2013) are water flow model based local binarization techniques. Moghaddam and Cheriet (2010) introduced adaptive Otsu and provided multiscale framework for binarization which can be used along with any adaptive thresholding technique. Farahmand et al. (2017) introduced kernel fuzzy c-means (KFMC) method which performs noise removal and binarization simultaneously by clustering image pixels into foreground, background and noise with respect to proper features. Sehad et al. (2019) used the Gabor filter to extract text from the degraded text image.

1.1. Issues with most of the local binarization techniques

(1) In most of the local binarization techniques, the binary output is dependent upon the parameters which are to be set manually to produce optimal result. One most important parameter in local methods is the size of the neighborhood (widow size) from which desired features are extracted to get the threshold value. In these methods, the window size is kept fixed for the whole image.

(2) For different images, same value of window size will not work well as images may contain the lower or greater density of information in different regions. Stroke width and text size may also vary within the same document. The fixed window size can perform well for one size of text but not for others.

This paper presents the local binarization technique that is the modification of state of art Sauvola and Pietaksinen (2000) binarization technique. It automatically computes window size dynamically across the image pixel to pixel using the stroke width transform. The size of window from which features are extracted to compute threshold value using intensity values of pixels will vary for every pixel depending upon the stroke width transform matrix. This led to reduction in the number of manually adjusted parameters. The rest of the paper is organized as follows: The proposed work is discussed in Section 2. The experimental results are presented in Section 3. Section 4 gives the conclusions.

2. Proposed work

This section discusses the proposed binarization technique (Modified Sauvola) for degraded document images. This method computes the neighborhood size automatically and dynamically across input image using stroke width transform (SWT). Our proposed technique produces improved results than Sauvola's method for all types of degraded images. It outperforms many other existing binarization techniques in case of degraded document images having variable stroke widths and text sizes.

2.1. Sauvola and Pietaksinen (Sauvola) binarization method (Sauvola and Pietaksinen, 2000)

This method computes local threshold $Th(i,j)$ using mean and standard deviation of the pixel values within the neighboring window of fixed size as follows:

$$Th(i,j) = \mu_{ij}\left[1 + k \times \left(1 - \frac{\sigma_{ij}}{R} - 1\right)\right], \tag{1}$$

Here, R is equal to 128 for gray images. μ_{ij} and σ_{ij} are the local mean and local standard deviation. The local window size (w) and k are two manually adjusted parameters, the binarization results are very much dependent upon these parameters. The main problem with this technique is the need to set the correct parameter values. The value of k and window size is recommended equal to 0.5 and 15. In this method, the window size and value of k remains fixed for whole image. As we have discussed above that in case of document image having different characteristics in different regions of the image or varying text sizes, fixed window size will not be able to produce good results.

2.2. Stroke width transform

"The stroke width transform (SWT) is a local image operator which computes per pixel the width of most likely stroke containing the pixel". Epshtein et al. (2010) has introduced stroke width transform (SWT) to detect text in natural scene images. The output of the stroke width transform is a matrix (SW) of size same as the input image where each element contains the stroke width related to that pixel.

To find the stroke width transform matrix, firstly in the matrix of same size as the input image, all elements are initialized with ∞. Then edge map of the input image is generated using canny edge detector. The edges are stroke boundaries, and we need to find the width of these strokes. Then the gradient direction at each edge pixel is computed. For any edge pixel u, the gradient direction g_u will be normal to the orientation of the stroke boundary. For each edge pixel u, the path $r = u + n * g_u (n > 0)$ is followed in the gradient direction until we find

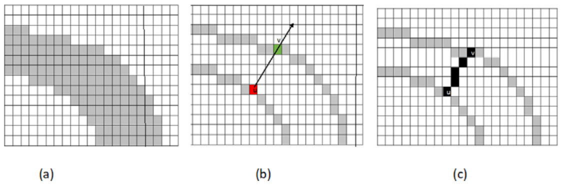

Fig. 1. Implementation of SWT. (a) A typical stroke, (b) u is pixel on the stroke boundary. v is pixel on the other side of the stroke boundary in the gradient direction at u, (c) each pixel along the path is assigned minimum of its current value and the found width of the stroke (Epshtein et al., 2010).

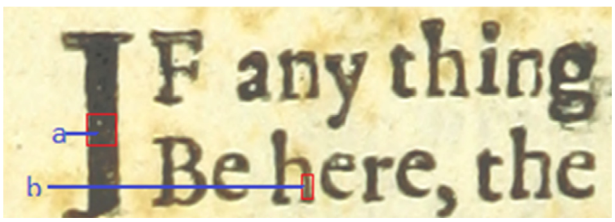

Fig. 2. Document image having variable stroke width.

another edge pixel v. If the gradient direction g_v at v is roughly opposite to g_u, then each pixel in the path is assigned the distance between u and v as their stroke width, unless its current value is less than computed value. If, however, an edge pixel v is not found, or g_v is not opposite to g_u, the path is discarded. The algorithm is applied twice, once with the gradient direction g_u and once with $-g_u$ to take into account both light text on a dark background and dark text on a light background. Fig. 1 shows the typical stroke and procedure to find out the stroke width. Tables 1 and 2 gives the stroke width transform matrix values for part a and part b in Fig. 2 respectively.

2.3. Modified Sauvola's method

In the Sauvola's method (Sauvola and Pietaksinen, 2000), parameters window size (W) and k are fixed and it is essential to set the correct values of these two variables for a particular document image. But it is difficult to manually compute and set exact values of these parameters for each and every document image. Also, a document image may contain the text of varying sizes and stroke width. The single window size can perform well for one size of text but not for others. The proposed method named Modified Sauvola's automatically computes the neighborhood size for every pixel based on the stroke width that is computed using stroke width transform. The algorithm is explained in the following steps:

Table 2
Stroke width transform matrix values for the subpart b in the Fig. 2.

10.77	9.75	9.75	9.75	10.00	10.00	10.82	10.15	10.15	10.15
10.77	9.75	9.75	9.75	9.75	9.75	9.59	9.75	9.75	9.95
10.77	9.75	9.90	9.90	9.90	9.75	9.75	9.75	9.75	9.75
10.77	9.75	9.90	9.90	9.90	9.90	9.75	9.75	9.75	9.75
10.77	9.75	9.90	9.90	9.90	9.75	9.75	9.75	9.75	10.00
10.77	10.00	10.00	10.00	10.00	10.00	9.75	10.00	10.00	10.00
10.77	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.77	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.77	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.77	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
10.34	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.20
10.34	10.00	10.00	10.00	10.00	10.00	9.75	10.34	10.34	10.34
10.34	9.75	10.00	10.00	10.00	9.75	9.59	10.34	10.82	10.82
10.34	9.59	9.75	10.00	9.75	9.59	9.54	14.87	14.87	14.87

Step 1:
If input document image is colored, convert it to gray level image.
Step 2:
Stroke Width Transform (SWT) of the input image is computed (as explained in Section 2.2). It generates SWT matrix.
Step 3:
This method computes the window size automatically for each pixel using the SWT matrix as follows:

$$W(i,j) = 4 \times SW(i,j) + 1 \tag{2}$$

Experimentally, the window size given by Eq. (2) produces the best results. The window size $W(i,j) \times W(i,j)$ for the pixel at location (i,j) is used for the computation of threshold and it will be different for each pixel in the image.
Step 4:
The threshold value for a pixel at location (i,j) is estimated using the window size for that pixel from step 3 and Eq. (1) (same as Sauvola's method). The values of R and k will be taken same as the Sauvola's method.

3. Experimental results

The experiments are performed on degraded document images selected from the DIBCO benchmark datasets DIBCO-2009 (Anon, 0000a), HDIBCO-2010 (Anon, 0000b), DIBCO-2011 (Anon, 0000c), HDIBCO-2012 (Anon, 0000d), HDIBCO-2016 (Anon, 0000e), and DIBCO-2017 (Anon, 0000f). These datasets include the variety of real degraded document images and their corresponding semi-automatically generated ground truth. The quantitative and OCR based results of the experiments are discussed in this section. The proposed method is compared with nine existing local binarization techniques: Otsu (1979),

Table 1
Stroke width transform matrix values for the subpart a in the Fig. 2.

15.07	15.07	15.07	15.07	15.07	15.46	15.46	15.46	15.46	15.46	15.46	15.46
15.07	15.07	15.07	15.43	15.43	15.43	15.43	15.43	15.43	15.43	15.43	15.43
12.81	14.87	14.87	14.87	15.36	15.36	15.36	15.36	15.36	15.36	15.36	15.36
13.75	14.97	14.97	14.97	14.97	14.97	14.97	15.36	15.36	15.36	15.36	15.36
13.75	14.42	14.97	14.97	14.97	14.97	14.97	14.97	14.97	14.97	15.36	15.36
13.75	14.42	14.42	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00
13.75	14.42	14.42	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00
14.39	14.39	14.39	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00
14.39	14.39	14.39	14.87	14.87	14.87	14.87	14.87	14.87	14.87	14.87	14.87
14.28	14.28	14.97	14.97	14.97	14.97	14.97	14.97	14.97	14.97	14.97	14.97
15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07
15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07
15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.07	15.33	15.33
15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.33	15.33	15.33	15.33	15.33
15.00	15.00	15.00	15.00	15.00	15.33	15.33	15.33	15.33	15.33	15.33	15.33
14.14	14.87	14.87	15.36	15.36	15.36	15.36	15.36	15.36	15.36	15.36	15.36
14.14	15.36	15.36	15.36	15.36	15.36	15.36	15.36	15.36	15.36	15.36	15.36
10.00	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.46
9.75	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.33	15.49

Table 3
Quantitative comparison of Sauvola's and proposed method using different datasets.

	DIBCO-2009 PR		DIBCO-2009HW		DIBCO-2011PR		DIBCO-2011 HW		HDIBCO-2016		DIBCO-2017	
	Sauvola	Proposed	Sauvola	Proposed	Sauvola	Proposed	Sauvola	Proposed	Sauvola	Proposed	Sauvola	Proposed
FM%	68.69	85.39	51.13	64.68	67.76	77.39	62.62	67.75	73.74	85.74	39.25	52.31
Recall %	53.77	75.47	40.84	54.94	52.19	64.74	49.00	58.24	61.26	85.445	30.69	43.88
Prec.%	99.32	98.84	97.41	95.21	99.70	98.96	98.00	95.01	98.77	87.44	83.63	95.21
PSNR	11.89	14.51	15.48	16.34	12.48	13.85	14.70	15.59	16.33	17.34	12.27	12.81
NRM	23.15	12.34	29.60	22.60	23.91	17.66	25.00	21.06	19.41	7.81	34.70	27.14
MPM (×1000)	2.57	1.05	0.45	0.38	0.88	0.63	1.50	1.48	1.26	1.01	5.00	1.40
DRD	12.06	7.18	10.95	8.63	9.82	6.57	9.14	8.40	8.99	8.28	14.13	10.44

Bernsen (1986), Niblack (1986), Wellner (1993), Sauvola and Pietaksinen (2000), Wolf and Jolion (2003), Bardley and Roth (2007), Khurshid et al. (2009) and Singh et al. (2012). Otsu is a global method, while others are local binarization methods. All these local methods use two manually adjusted parameters similar to the Sauvola's method, and in all binarization results are much affected by the values of these parameters.

3.1. Statistical results

The statistical results are evaluated using the quantitative measures: F-measure (FM), Peak Signal to Noise Ratio (PSNR), Negative Rate Metric (NRM), Misclassification penalty metric (MPM), and Distance Reciprocal Distortion Metric (DRD) adopted from the ICDAR benchmark evaluation measures, proposed in the International Document Image Binarization Contest (Gatos et al., 2009; Pratikakis et al., 2016, 2017). These metrics define the similarity percentage between the resulting binarized image and ground truth image.

FM combines precision (PR) and recall (RC) to determine overall binarization accuracy. PR represents the binarization noise, and RC gives text body accuracy. High values of these three measures indicate more closeness between resulting binarized image (I_B) and ideal binary image (I_{GT}).

$$FM = \frac{2 \times RC \times PR}{RC + PR}, \quad (3)$$

where $PR = \frac{N_{TT}}{N_{FT}+N_{TT}}$, $RC = \frac{N_{TT}}{N_{FNT}+N_{TT}}$

Here, N_{TT} is the number of true text pixels, N_{FT} is the number of false text pixels and N_{FNT} is the number of false non-text pixels.

PSNR is another performance metric that measures closeness of resulting binary output to the ideal binary image. A higher value of PSNR gives better binarization quality.

$$PSNR = 10 \times log\left(\frac{D^2}{MSE}\right), \quad (4)$$

where D is the contrast of the image. For binary images the value of D is taken as 1. MSE denotes the mean square error.

$$MSE = \sum_{i=1}^{M}\sum_{j=1}^{N}\frac{\left(I_B(i,j) - I_{GT}(i,j)\right)^2}{M.N} \quad (5)$$

$I_B(i,j)$, the binarized image pixel value and $I_{GT}(i,j)$ is the ideal binary image intensity at the same pixel position. **NRM** measures the pixel dissimilarity rate between I_{GT} and the I_B. The lower NRM value indicates better binarization.

$$NRM = \frac{P+Q}{2} \quad (6)$$

where

$$P = \frac{N_{FNT}}{N_{FNT} + N_{TT}}, \quad (7)$$

$$Q = \frac{N_{FT}}{N_{FT} + N_{TNT}} \quad (8)$$

MPM is defined as follows:

$$MPM = \frac{\sum_{i=1}^{C_{FNT}} d_{FNT}^{i} + \sum_{j=1}^{C_{FT}} d_{FT}^{j}}{2D} \quad (9)$$

Fig. 3. (a) Input image (P02.bmp from DIBCO-2009), (b) Sauvola's output (F-measure = 77.14), (c) Proposed (F-measure = 91.72).

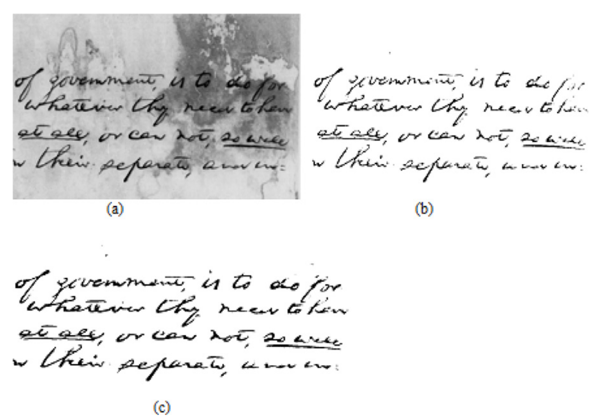

Fig. 4. (a) Input image (H04.bmp from DIBCO-2009), (b) Sauvola's output (F-measure = 73.15), (c) Proposed (F-measure = 85.79).

Here, d_{FNT}^{i} denote the distance of the ith false non-text and d_{FT}^{j} is the distance of the jth false text pixel from the contour of the text in the ground truth image. The normalization factor D is the sum over all the pixel-to-contour distances of the ground truth object. This metric measure how well the output binary image represents the contour of ground truth image. Smaller the value of MPM, better is the quality of the algorithm.

DRD measures the visual distortion in binary document images, and it was introduced by Lu et al. (2004). This measure correlates the performance of method with the human visual perception of an image.

The average quantitative results of the proposed method and Sauvola's method using all the above-mentioned datasets is recorded in Table 3. The quantitative results in Table 3 and visual results in Figs. 3, 4, and 5 indicate that for all types of degraded images (machine printed as well as handwritten) proposed method produces better results than the Sauvola's method. The experiment is also performed

Fig. 5. (a) Part of image PR4.png from DIBCO-2011 dataset (Heavy and variable text stroke), (b) Ground truth, (c) Otsu, (d) Niblack, (e) Bernsen, (f) Wellner, (g) Sauvola, (h) Wolf, (i) Bradley & Roth, (j) NICK, (k) Singh et al. (2012), (l) Proposed.

on printed document images having varying stroke widths and text sizes. Ten degraded document images PR1.png, PR3.png, PR4, PR6 from DIBCO-2011, P02.bmp, P03.bmp from DIBCO-2009, and 13.bmp, 14.bmp, 15.bmp, 20.bmp from DIBCO-2017 datasets are selected and average statistical measures for the selected existing techniques and the proposed method are recorded in Table 4. The results show that proposed method outperforms other methods for such type of images. The visual results of different methods (Fig. 5) for degraded document image having variable stroke width and text size also confirms the same. The results in Fig. 5 show that Niblack (1986) and Bernsen (1986) method produce black noise in non-text region. The Wellner (1993), Sauvola and Pietaksinen (2000), Wolf and Jolion (2003), Bardley and Roth (2007), Khurshid et al. (2009) and Singh et al. (2012) methods are not able to retrieve text pixels completely from heavy and variable size text images. The proposed method gives the best results for such images. The experiments on low contrast images show that like Sauvola's method, the proposed method is not able to produce good binarization results for such images. The F-measure values in Fig. 6 show that in comparison to Sauvola's method, the proposed method retrieves more true pixels for such types of images also.

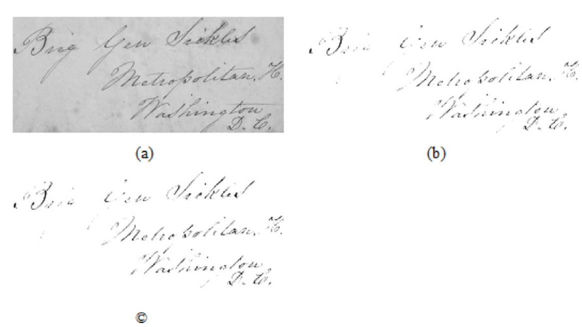

Fig. 6. (a) Input image (Low contrast image H06.tif from HDIBCO-2010), (b) Sauvola's output (F-measure = 31.13), (c) Proposed (F-measure = 39.68)(a).

3.2. OCR based evaluations

The performance of a binarization directly affects the recognition process in the Optical Character Recognition (OCR) system. The quality of the proposed method over other existing methods is evaluated on OCR results by calculating Levenshtein distance. "The Levenshtein distance is a measure of the similarity between two strings. The distance is the number of deletions, insertions, or substitutions required to transform one string into another" (Levenshtein, 1966). This method can be used in the evaluation of machine-printed documents only since handwritten OCR yields no satisfactory results. The binary results of the proposed method and the existing techniques are analyzed using free online OCR (Anon, 0000g). For better quality binarization method, the Levenshtein distance between the OCR result of the ground truth image, and the OCR result of the output image of the binarization method should be less. The OCR results and Levenshtein distances of the proposed method and other methods for part of an image selected from the considered datasets is shown in Fig. 7. The results show the best performance of the proposed method for images having variable stroke widths and text sizes.

Table 4
Quantitative comparison of different methods for selected images having heavy and variable stroke width text.

	Recall (%)	Prec. (%)	FM (%)	PSNR	NRM	MPM (×1000)	DRD
Otsu (1979)	95.23	68.54	74.14	12.79	13.41	104.55	51.96
Bernsen (1986)	89.99	28.69	41.99	5.14	22.10	213.38	98.01
Niblack (1986)	85.35	30.75	43.53	5.69	21.82	185.99	84.43
Wellner (1993)	45.52	91.83	59.04	11.52	27.49	4.91	16.37
Sauvola and Pietaksinen (2000)	59.27	98.41	73.11	13.17	20.43	2.24	10.74
Wolf and Jolion (2003)	70.01	95.21	79.56	14.79	15.22	2.45	7.28
Bardley and Roth (2007)	79.68	80.17	78.50	13.32	12.36	12.47	11.77
Khurshid et al. (2009)	70.57	87.86	76.51	13.21	15.28	6.65	11.42
Singh et al. (2012)	58.78	81.68	65.12	11.57	21.59	11.42	19.24
Proposed	75.71	96.68	**84.81**	**15.01**	**12.30**	**1.63**	**6.79**

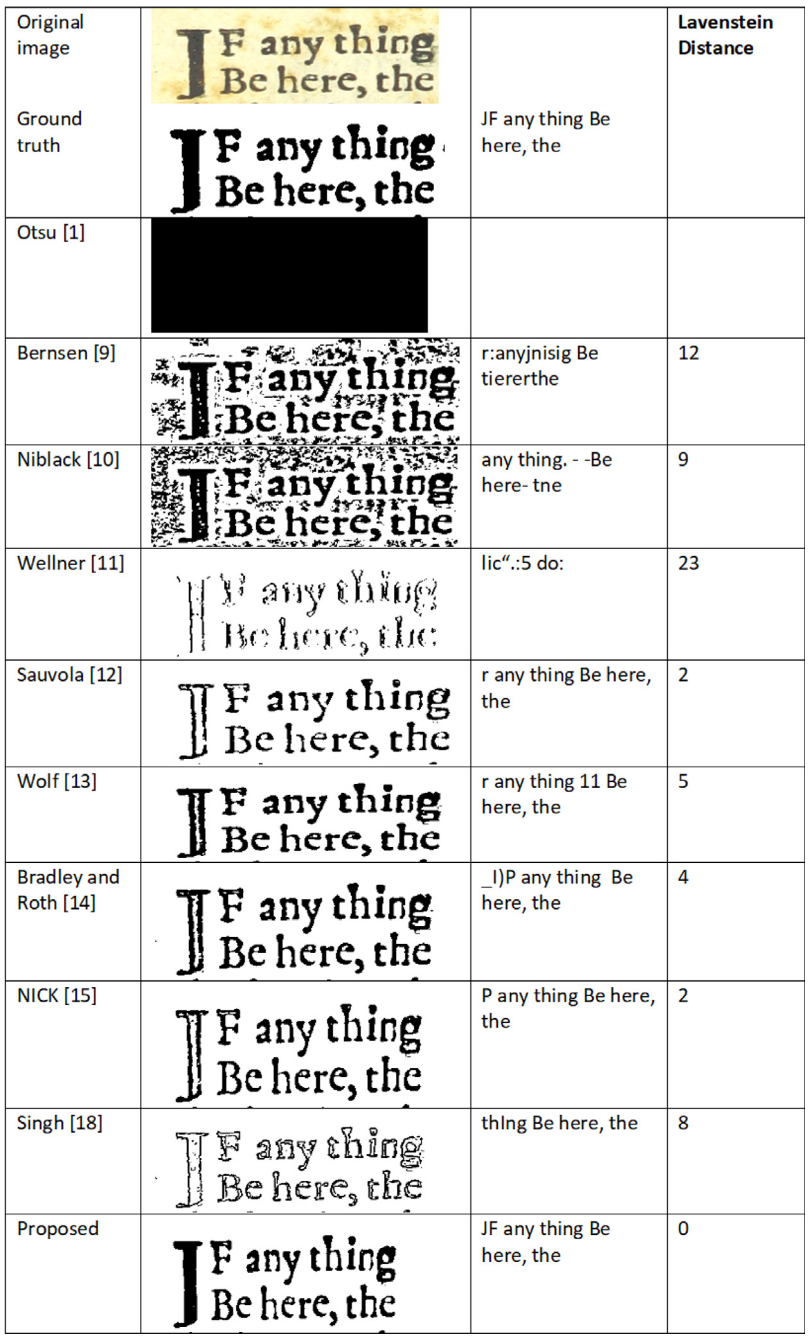

Fig. 7. The OCR results and Levenshtein distances for the part of image PR3.png from DIBCO-2011 dataset.

4. Conclusions

This paper presents the local binarization technique that is the modified form of state of art binarization technique Sauvola. In the Sauvola's method, window size parameter is needed to manually set by the user and it remains fixed for whole image irrespective of the varying characteristics of text in different regions. The proposed technique computes the window size dynamically using the stroke width transform matrix. The visual and quantitative results are presented which shows the improved performance of the proposed method over Sauvola's for printed as well for handwritten images. For images having variable stroke widths and text sizes, the proposed method outperforms the Otsu, Niblack, Bernsen, Wellner, Sauvola, Wolf, Bradley and Roth, NICK, and Singh binarization methods.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Amandeep Kaur: Resources, Methodology, Supervision. **Usha Rani:** Conceptualization, Writing - original draft. **Gurpreet Singh Josan:** Resources, Supervision.

References

Abutaleb, A.S., 1989. Automatic thresholding of gray-level pictures using two-dimensional entropy. Comput. Vis. Graph. Image Process. 47 (1), 22–32.
Anon, users.iit.demokritos.gr/~bgat/DIBCO2009/benchmark.
Anon,.
Anon, utopia.duth.gr/~ipratika/DIBCO2011/resources.
Anon, utopia.duth.gr/ipratika/HDIBCO2012/benchmark.
Anon, https://vc.ee.duth.gr/h-dibco2016/benchmark.
Anon, https://vc.ee.duth.gr/dibco2017.
Anon, https://www.onlineocr.net.
Bardley, D., Roth, G., 2007. Adaptive thresholding using the integral image. J. Graph. Tools 12 (2), 13–21.
Bernsen, J., 1986. Dynamic thresholding of gray level images. In: Proceedings IEEE International Conference on Pattern Recognition, pp. 1251–1255.
Brink, A.D., Pendock, N.E., 1996. Minimum cross entropy threshold selection. Pattern Recognit. 29, 179–188.
Do, J., He, Q.D.M., Downton, A.C., Kim, J.H., 2005. A comparison of binarization methods for historical archive documents. In: Eighth International Conference on Document Analysis and Recognition (ICDAR'05), Vol. 1, Seoul, South Korea, pp. 538–542.
Epshtein, B., Ofek, E., Wexler, Y., 2010. Detecting text in natural scenes with stroke width transform. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, pp. 2963–2970.
Farahmand, A., Sarrafzadeh, A., Shanbehzadeh, J., 2017. Noise removal and binarization of scanned document images using clustering of features. In: Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS2017), Vol. 1, pp. 410–414.
Gatos, B., Ntirogiannis, K., Pratikakis, I., 2009. ICDAR 2009 document image binarization contest (DIBCO 2009). In: 10th International Conference on Document Analysis and Recognition, Barcelona, pp. 1375–1382.
Gatos, B., Pratikakis, I., Perantonis, S.J., 2006. Adaptive degraded document image binarization. Pattern Recognit. 39 (3), 317–327.
Hadjadj, Z., Meziane, A., Cherfa, Y., Cheriet, M., Setitra, I., 2016. ISauvola: Improved Sauvola's Algorithm for Document Image Binarization. Springer International Publishing, Switzerland, pp. 737–745.
Hejdam, R., Moghaddam, R.F., Cheriet, M., 2011. A spatially adaptive statistical method for the binarization of historical manuscripts and degraded document images. Pattern Recognit. 44, 2184–2196.
Johannsen, G., Bille, J., 1982. A threshold selection method using information measures. In: ICPR'82: Proc. 6th Intl. Conf. Patt. Recogn., pp. 140–143.
Kapur, J.N., Sahoo, P.K., A.K.C., Wong., 1985. A new method for gray-level picture thresholding using the entropy of the histogram. Comput. Vis. Graph. Image Process. 29 (3), 273–285.
Kawano, H., Oohama, H., Maeda, H., Okada, Y., Ikoma, N., 2009. Degraded document image binarization combining local statistics. In: IEEE International Joint Conference (ICROS-SICE). pp. 439–443.
Khurshid, K., Siddiqi, I., Faure, C., Vincent, N., 2009. Comparison of Niblack inspired methods for ancient document. In: Proceedings 16th IEEE International Conference on Document Recognition and Retrieval, pp. 1–10.
Kim, I.K., Jung, D.W., Park, R.H., 2002. Document image binarization based on topographic analysis using a water flow model. Pattern Recognit. 3, 265–277.
Kittler, J., Illingworth, J., 1985. Threshold selection using clustering criteria. IEEE Trans. Syst. Man Cybern. 15, 652–655.
Levenshtein, V.I., 1966. Binary codes capable of correcting deletions, insertions and reversals. Sov. Phys. Dokl. 10 (8), 707–710.
Lu, H., Kot, A.C., Shi, Y.Q., 2004. Distance-reciprocal distortion measure for binary document images. IEEE Signal Process. Lett. 11 (2), 228–231.
Lu, S., Tan, C.L., 2010. Document image binarization using background estimation and stroke edges. Int. J. Doc. Anal. Recogn. 13 (4), 303–314.
Moghaddam, R.F., Cheriet, M., 2010. A multi-scale framework for adaptive binarization of degraded document images. Pattern Recognit. 43 (6), 2186–2198.
Natarajan, J., Sreedevi, I., 2017. Enhancement of ancient manuscripts images by log based binarization technique. Int. J. Electron. Commun. 75, 15–22.
Niblack, W., 1986. An Introduction to Image Processing. Prentice-Hall, Englewood Cliffs, pp. 115–116.
Oh, H.H., Lim, K.T., Hien, S.I., 2005. An improved binarization algorithm based on a water flow model for document image with inhomogeneous backgrounds. Pattern Recognit. 38, 2612–2625.
Otsu, N., 1979. A threshold selection method from gray level histograms. IEEE Trans. Syst. Man Cybern. 9 (1), 62–66.
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), Shenzhen, pp. 619–623.
Pratikakis, I., Zagoris, K., Barlas, G., Gatos, B., 2017. ICDAR2017 competition on document image binarization (DIBCO 2017). In: 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, pp. 1395–1403.
Pun, T., 1980. A new method for gray-level picture threshold using the entropy of the histogram. Signal Process. 2 (3), 223–237.
Pun, T., 1981. Entropic thresholding: a new approach. Comput. Graph. Image Process. 16, 210–239.
Sauvola, J., Pietaksinen, M., 2000. Adaptive document image binarization. Pattern Recognit. 33, 225–236.
Sehad, A., Chibani, Y., Hedjam, R., Cheriet, M., 2019. Gabor filter-based texture for ancient degraded document image binarization. Pattern Anal. Appl. 22, 1–22.
Shafait, F., Keysers, D., Bruel, T.M., 2008. Efficient implementation of local adaptive thresholding techniques using integral images. In: Proceedings of SPIE 6815 on Document Recognition and Retrieval XV, 6815.
Shukla, S., Sonawane, A., Topale, V., Tiwari, P., 2014. Improving degraded document images using binarization technique. Int. J. Sci. Technol. Res. 3 (5), 333–338.
Singh, T.R., Roy, S., Singh, O.I., Sinam, T., Singh, K.M., 2011. A new local adaptive thresholding technique in binarization. Int. J. Comput. Sci. 8 (6), 271–277.
Singh, O.I., Sinam, T., James, O., Singh, T.R., 2012. Local contrast and mean based thresholding technique in binarization. IJCSI Int. J. Appl. 51 (6), 5–10.
Su, B., Lu, S., Tan, C.L., 2010. Binarization of historical handwritten document images using local maximum and minimum filter. In: International Workshop on Document Analysis Systems. pp. 159–165.
Su, B., Lu, S., Tan, C.L., 2013. A robust document image binarization technique for degraded document images. IEEE Trans. Image Process. 22 (4), 1408–1417.
Valizadeh, M., Kabir, E., 2013. An adaptive water flow model for binarization of degraded document images. IJDAR 16 (2), 165–176.
Van, L.T.K., Lee, G., 2014. A stroke width-based contrast feature for document image binarization. J. Inf. Process. Syst. 10 (1), 55–68.
Wellner, P., 1993. Adaptive Thresholding for the Digital Desk. Xerox, EPC1993-110.
Wolf, C., Jolion, J.M., 2003. Extraction and recognition of artificial text in multimedia documents. Pattern Anal. Appl. 6 (4), 309–326.
Yang, Y., Yan, H., 2000. An adaptive logical method for binarization of degraded document images. Pattern Recognit. 33 (5), 787–807.
Zhou, S., Liu, C., Cui, Z., Gong, S., 2009. An improved adaptive document image binarization method. In: Proceedings of 2nd IEEE International Conference on Image Signal Processing, pp. 1–5.