IJDAR (2011) 14:35–44
DOI 10.1007/s10032-010-0115-7

ORIGINAL PAPER

DIBCO 2009: document image binarization contest

B. Gatos · K. Ntirogiannis · I. Pratikakis

Received: 23 November 2009 / Revised: 26 March 2010 / Accepted: 28 March 2010 / Published online: 13 May 2010

Abstract DIBCO 2009 is the first International Document Image Binarization Contest organized in the context of ICDAR 2009 conference. The general objective of the contest is to identify current advances in document image binarization using established evaluation performance measures. This paper describes the contest details including the evaluation measures used as well as the performance of the 43 submitted methods along with a short description of the top five algorithms.

Keywords Binarization · Performance evaluation · Document image preprocessing

1 Introduction

Document image binarization is an important step in the document image analysis and recognition pipeline. Therefore, it is imperative to have a benchmarking dataset along with an objective evaluation methodology in order to capture the efficiency of current document image binarization practices. To this end, we organized the first International Document Image Binarization Contest (DIBCO 2009) in the context of ICDAR 2009 conference. In this contest, we focused on the evaluation of document image binarization methods using a variety of scanned machine-printed and handwritten documents for which we created the binary image ground truth following a semi-automatic procedure based on Ref. [1]. The authors of submitted methods registered in the competition and downloaded representative samples along with the corresponding ground truth. At a next step, all registered participants were required to submit their binarization executable. After the evaluation of all candidate methods, the testing dataset (5 machine-printed and 5 handwritten images with the associated ground truth) became publicly available (http://www.iit.demokritos.gr/~bgat/DIBCO2009/benchmark).

The remainder of the paper is structured as follows: All the participants are listed in Sect. 2. The evaluation measures are detailed in Sect.3. The experimental results are shown in Sect. 4. In Sect. 5, the top five methods are detailed while in Sect. 6 conclusions are drawn.

B. Gatos (✉) · K. Ntirogiannis · I. Pratikakis
Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
National Center for Scientific Research "Demokritos",
153 10 Agia Paraskevi, Athens, Greece
e-mail: bgat@iit.demokritos.gr

K. Ntirogiannis
e-mail: kntir@iit.demokritos.gr

I. Pratikakis
e-mail: ipratika@iit.demokritos.gr

2 Participants

Thirty-five (35) research groups have participated in the competition with forty-three (43) different algorithms (several participants submitted more than one algorithm). Table 1 presents all the participants sorted by the date of submission. A brief description of each participating method is given in Ref. [2].

3 Evaluation measures

The evaluation measures used comprise an ensemble of measures that have been widely used for evaluation purposes. These measures consist of (i) F-measure; (ii) PSNR; (iii) negative rate metric and (iv) misclassification penalty metric.

Table 1 Methods submitted to DIBCO 2009 sorted by the date of submission

No.	Research group
1	The Generations Network, Inc. USA (*D. Curtis*)
2	Meisei University, Japan (*Y. Shima*)
3	Democritus University of Thrace, Greece (*M. Makridis, N. Papamarkos*)
4	South University of Toulon-Var, France (*F. Bouchara, T. Lelore*)
5	University of the Aegean, Greece (*E. Kavallieratou*)
6	University of Groningen, The Netherlands (*A. Brink*)
7	Institute of Space Technology, Pakistan – (*K. Khurshid*)
8	East China Normal University, China (*G. Gu*)
9	Université de Lyon, INSA, France (*C. Wolf, Jean-Michel Jolion*)
10	Tsinghua University, China, (*X. Shen*)
11	Centre de Morphologie Mathématique, France (*B. Marcotegui, J. Hernández*)
12	Google R D Bangalore, India (*A. Jain*)
13	University of Sfax, Tunisia (*M. Chakroun, M. Charfi, M. A. Alimi*)
14	Université Pierre et Marie Curie CMM, France (*J. Fabrizio, B. Marcotegui*)
15	Freie Universität Berlin, Germany (*M. Block, R. Rojas*)
16	Universidade Federal de Pernambuco, Brazil (*D. M. Oliveira, R. D. Lins*)
17	University of Joensuu, Finland (*M. Chen, Q. Zhao, T. Kinnunen, R. Saeidi, P. Franti*)
18	Centre de morphologie mathématique, France (*J. Hernández, B. Marcotegui)*
19	Freie Universität Berlin, Germany (*M. Ramirez, E. Tapia and R. Rojas*)
20	University of Quebec, Canada (*R. Hedjam, R. F. Moghaddam and M. Cheriet*)
21	Universidade Federal de Pernambuco, Brazil (*R. D. Lins, J. M. M. da Silva*)
22	The Neat Company, PA, USA (*H. Ma*)
23	University of Sfax, Tunisia (*F. Drira, F. LeBourgeois*)
24	University of Quebec, Canada (*D. Rivest-Hénault, R. F. Moghaddam and M. Cheriet*)
25	University of Quebec, Canada (*R. F. Moghaddam*)
26	Institute for Infocomm Research, Singapore (*S. Lu, C.L. Tan*)
27	University of Sfax, Tunisia (*A. Bougacha, W. Boussellaa, A. M. Alimi*)
28	Google R D Bangalore, India (*K. Chaudhury, A. Jain, S. Thirthala, V. Sahasranaman, S. Saxena and S. Mahalingam*)
29	University of Malta (*A. Bonnici, K. P. Camilleri*)
30	University at Buffalo, SUNY, USA (*Z. Shi, S. Setlur, V. Govindaraju*)
31	Pune Institute of Computer Technology, India(*S. D. Shelke*)
32	Universitat Autonoma de Barcelona, CVC (*R. Coll*)
33	Google, Inc., Mountain View, USA (*D. Bloomberg*)
34	Boise State University, USA Telecom ParisTech, France Math dept., UCLA, USA (*E. H. Barney Smith, L. Likforman and D. Jerome*)
35	Google, Inc., Mountain View, USA (*R. Romano*)

3.1 Definitions

– F-measure

$$F\text{-measure} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (1)$$

where recall $= \frac{TP}{TP+FN}$, precision $= \frac{TP}{TP+FP}$

TP, FP and FN denote the true-positive, false-positive and false-negative values, respectively.

– PSNR

$$\text{PSNR} = 10 \log \left(\frac{C^2}{MSE} \right) \quad (2)$$

where MSE $= \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x,y) - I\prime(x,y))^2}{MN}$

PSNR is a measure of how close is an image to another. Therefore, the higher the value of PSNR, the higher the similarity of the two images. We consider that the difference between foreground and background equals to 255 ($C = 255$).

– Negative rate metric (NRM)

Fig. 1 **a** Representative machine-printed image; **b–f** Binarization results from the first to the fifth best binarization algorithm, i.e. algorithms No. 26, 14, 24, 10 and 9a, respectively

The negative rate metric (NRM) is based on the pixel-wise mismatches between the GT and prediction. It combines the false-negative rate NR_{FN} and the false-positive rate NR_{FP}. It is denoted as follows:

$$NRM = \frac{NR_{FN} + NR_{FP}}{2} \tag{3}$$

where $NR_{FN} = \frac{N_{FN}}{N_{FN}+N_{TP}}$, $NR_{FP} = \frac{N_{FP}}{N_{FP}+N_{TN}}$

N_{TP}, N_{FP}, N_{TN} and N_{FN} denote the number of true positives, false positives, true negatives and false negatives, respectively.

In contrast to F-measure and PSNR, the binarization quality is better for lower NRM.

– Misclassification penalty metric (MPM)

The misclassification penalty metric (MPM) evaluates the prediction against the ground truth (GT) on an object-by-object basis. Misclassification pixels are penalized by their distance from the ground truth object's border.

$$\mathrm{MPM} = \frac{MP_{FN} + MP_{FP}}{2} \tag{4}$$

where $MP_{FN} = \frac{\sum_{i=1}^{N_{FN}} d_{FN}^{i}}{D}$, $MP_{FP} = \frac{\sum_{j=1}^{N_{FP}} d_{FP}^{j}}{D}$

d_{FN}^{i} and d_{FP}^{j} denote, respectively, the distance of the ith false-negative and the jth false-positive pixel from the contour of the GT segmentation. The normalization factor D is the sum over all the pixel-to-contour distances of the GT object. A low MPM score denotes that the algorithm is good at identifying an object's boundary.

4 Experimental results

The DIBCO testing dataset consists of five machine-printed and five handwritten images resulting in a total of 10 images for which the associated ground truth was built for the evaluation. Representative example images of the dataset for printed and handwritten images are shown in Figs. 1a and 2a, respectively. The documents of this dataset originate from the collections of the following libraries: The Goettingen State and University Library, The Bavarian State Library, the British Library and the Library of Congress. The selection of the images in the dataset was made so that should contain representative degradations which appear frequently (e.g. variable background intensity, shadows, smear, smudge, low contrast, bleed-through or show-through).

The evaluation was based upon the four distinct measures presented in Sect. 3. Apart the overall ranking (Table 2), we present the ranking on the machine-printed and the handwritten test images, separately (Table 3). For each type, the ranking was calculated after accumulating the ranking value of each method for all measures. Let $R(i,j)$ be the rank of the i-th method using the j-th measure, where $i = 1 \ldots t$, t denotes the number of the binarization techniques used in the evaluation and $j = 1 \ldots m$, m denotes the number of

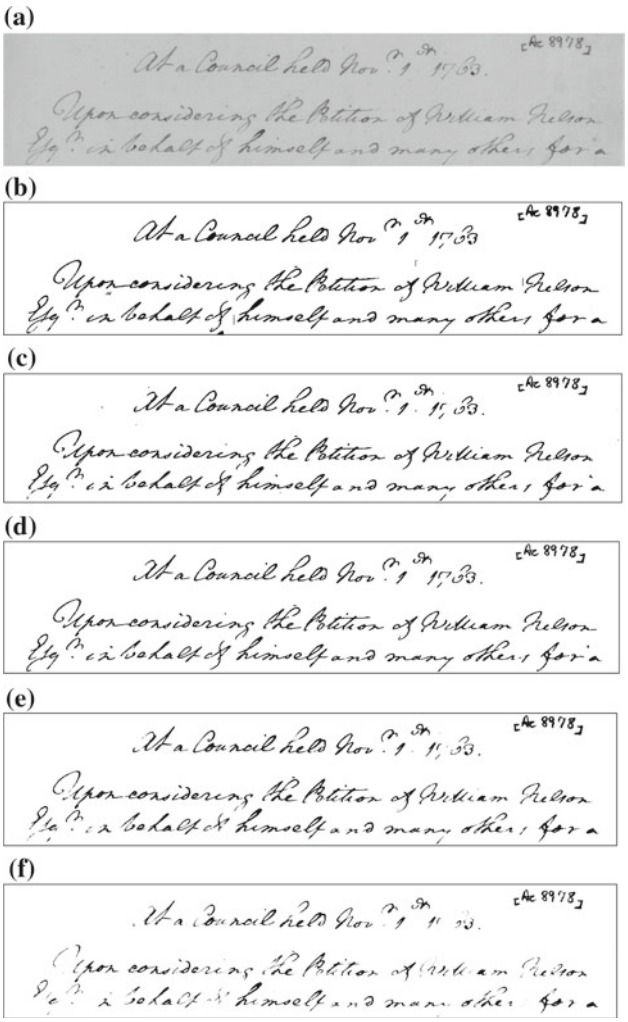

Fig. 2 **a** Representative handwritten image; **b–f** Binarization results from the first to the fifth best binarization algorithm, i.e. algorithms No. 26, 14, 24, 10 and 9a, respectively

the evaluation measures. For each binarization method, the final ranking is achieved by the summation of the four rankings $S_i = \sum_{j=1}^{4} R(i, j)$. The smaller value of S_i the better performance is achieved by the corresponding binarization method i. We further provide graphs that show the performance of the binarization algorithms in terms of F-measure and NRM (Fig. 3). Overall, the best performance is achieved by *Algorithm 26* which has been submitted by S. Lu and C.L. Tan of the *Institute for Infocomm Research* in Singapore. To further provide a comparison with representative state-of-the-art binarization algorithms, Otsu [3] and Sauvola et al. [4] algorithms were applied at the DIBCO 2009 dataset using the DIBCO 2009 measures. Results are shown in Tables 2, 3. Example binarization results of the top five algorithms for machine-printed and handwritten images are shown in Figs. 1b–f and 2b–f.

5 Top five best performing algorithms

5.1 Algorithm No. 26: Institute for Infocomm Research, Singapore (S. Lu, C. L. Tan)

The algorithm includes four parts, which deal with document background extraction, stroke edge detection, local thresholding and post-processing, respectively. The local threshold is estimated by averaging the detected edge pixels within a local neighborhood window. A detailed description follows.

The background is estimated in a two-round strategy. In the first round, the row-scanned background surface is generated by fitting a polynomial for each row of the document image. In the second round, we smooth the first-round background surface by applying the polynomial smooth on that surface column by column. In each round of the smoothing, the data is 1-D image signal that is sampled from one row/column of the document image under study. Equation 5 below specifies the sampled data.

$$X_{sp}(i) = \frac{\sum_{j=i*\text{step}}^{(i+1)*\text{step}} X_{sm}(j)}{\text{windowsize}} \tag{5}$$

where $X_{sm}(j)$ refers to the origin pixel value, $X_{sp}(i)$ refers to the sampled pixel value, and *step* denotes the sampling step. This polynomial smooth procedure is employed iteratively for each row/column. After each round of smoothing, the sampled data that is farthest from the fitted smoothing polynomial is removed. The smoothing proceeds iteratively until the maximum difference between the sampled data and the fitted smoothing polynomial is smaller than a pre-defined threshold.

When the background surface is estimated, the gradient information (GI) of each pixel is calculated as follows: First, the pixel differences between one pixel and its eight neighbors are calculated in four directions as specified in Eq. 6. Then, the GI value of one pixel is defined as sum of the magnitude of the four direction differences. In addition, the GI (7) is further normalized by mdn/bg to deal with the uneven illumination.

$$\left\{\begin{array}{l} A = I(i, j+1) + I(i, j) - 2^*I(i, j-1) \\ B = I(i+1, j) + I(i, j) - 2^*I(i-1, j) \\ C = I(i+1, j+1) + I(i, j) - 2^*I(i-1, j-1) \\ D = I(i-1, j+1) + I(i, j) - 2^*I(i+1, j-1) \end{array}\right\} \tag{6}$$

$$GI(i, j) = \frac{mdn}{bg}(|A| + |B| + |C| + |D|) \tag{7}$$

where I refers to the intensities of document image, A, B, C and D denote the pixel intensity difference of four directions, mdn is median value of BG, and bg refers to the intensity of that point in the estimated document background surface. The GI of document images usually has a bimodal pattern

Table 2 Detailed evaluation results of all methods participating to DIBCO 2009

Rank	Method	F-measure (%)	PSNR	NRM ($\times 10^{-2}$)	MPM ($\times 10^{-3}$)
1	**26**	91.24	18.66	4.31	0.55
2	**14**	90.06	18.23	4.75	0.89
3	**24**	89.34	17.79	5.32	1.90
4	**10**	87.03	17.21	7.03	0.57
5	**9a**	87.89	17.12	7.73	0.97
6	**8**	87.71	16.86	5.99	2.19
7	**33c**	86.35	16.66	6.03	1.45
8	**9b**	87.16	17.08	8.5	0.74
9	**4**	86.53	16.47	5.41	1.76
10	**34a**	87.49	16.83	7.76	1.57
11	**33b**	85.66	17.01	11.37	0.52
12	**6**	86.93	16.61	7.29	2.58
13	**11**	85.72	16.44	8.94	1.12
14	**34b**	85.99	16.37	8.28	1.46
15	**35**	85.11	15.75	5.38	2.22
16	**33a**	84.59	16.66	11.48	0.61
17	**1**	85.06	16.36	6.49	3.78
18	**34c**	84.78	16.02	8.73	1.50
19	**25**	83.99	15.58	4.18	4.60
20	**3**	85.30	15.68	7.59	4.18
21	**7c**	85.17	16.04	9.93	1.93
22	**17**	83.98	15.81	4.51	5.48
23	**34d**	84.03	15.86	8.78	1.40
24	**29**	84.69	16.33	7.96	3.83
25	**18**	83.74	15.22	4.62	3.86
26	**23**	82.50	15.11	4.47	3.62
27	**12**	83.53	15.59	4.91	5.34
28	**22**	83.54	15.53	7.62	3.54
29	**7a**	84.57	15.67	7.81	5.84
30	**28**	84.25	16.42	9.13	7.46
31	**30**	83.62	15.57	7.67	5.53
32	**2**	83.10	14.74	5.18	7.11
33	**19**	79.71	16.62	9.93	4.55
34	**7b**	80.74	14.86	5.98	9.60
35	**5**	80.90	14.64	8.17	4.22
36	**15**	74.12	15.05	18.07	2.57
37	**16**	82.27	14.96	8.04	41.30
38	**21**	75.86	13.34	15.45	2.51
39	**20**	80.43	14.37	8.21	7.70
40	**27**	82.74	14.78	10.12	56.22
41	**13**	35.28	12.44	36.60	2.68
42	**31**	61.48	9.22	14.69	86.03
43	**32**	58.77	9.27	18.77	118.02
Otsu [3]		82.17	15.06	5.63	13.86
Sauvola et al. [4]		87.26	16.69	6.61	3.39

Bold face numerals are shown for "rank" and "method" categories in order to differentiate them from evaluation measures namely "F-measure", "PSNR", "NRM" and "MPM"

Table 3 Detailed evaluation results for machine-printed and handwritten document images

Rank	Machine-printed					Handwritten				
	Method	F-measure (%)	PSNR	NRM ($\times 10^{-2}$)	MPM ($\times 10^{-3}$)	Method	F-measure (%)	PSNR	NRM ($\times 10^{-2}$)	MPM ($\times 10^{-3}$)
1	**14**	94.09	17.9	3.12	0.83	**26**	88.65	19.42	5.11	0.34
2	**26**	93.81	17.89	3.5	0.75	**14**	86.02	18.57	6.39	0.95
3	**24**	93.18	17.44	4.73	0.47	**34a**	86.16	18.32	6.25	1.54
4	**33c**	92.73	17.09	2.84	1.47	**8**	84.49	17.76	5.54	2.03
5	**1**	92.82	17.26	4.3	1.39	**24**	85.29	18.14	5.9	3.33
6	**10**	92.17	17.02	5.93	0.4	**6**	84.75	17.82	6.86	1.67
7	**19**	92.76	17	4.69	1.17	**4**	81.9	16.96	4.5	1.88
8	**9b**	91.94	16.6	5.58	1.1	**34b**	83.36	17.53	7.36	1.43
9	**28**	91.59	16.77	5.17	1.17	**3**	82.93	17	3.73	5.65
10	**12**	91.65	16.47	3.33	2	**9a**	84.16	17.82	9.8	0.58
11	**18**	91.42	15.93	3.52	1.36	**33b**	82.17	17.82	13.8	0.24
12	**33a**	91.04	16.43	6.99	0.71	**9b**	82.35	17.56	11.4	0.38
13	**9a**	91.61	16.42	5.66	1.35	**10**	81.58	17.4	8.14	0.74
14	**7b**	91.75	16.89	3.55	3.41	**11**	82.41	17.45	9.7	0.74
15	**17**	91.55	16.73	3.62	2.48	**34c**	81.52	17.1	8.14	1.41
16	**25**	89.76	15.6	2.61	1.38	**16**	81.67	16.78	5.53	10.1
17	**27**	91.19	16.43	4.61	2.48	**2**	80.4	16.08	5.55	5.47
18	**29**	90.4	16.45	7.54	1.44	**35**	80.43	16.14	7.24	2
19	**20**	89.78	15.94	7.21	0.88	**7a**	81.59	16.83	7.43	6.02
20	**33b**	89.02	16.19	8.95	0.8	**7c**	81.81	16.96	10.6	2.19
21	**4**	90.36	15.98	6.33	1.65	**34d**	79.52	16.75	9.55	1.34
22	**23**	89.84	15.43	2.9	2.81	**22**	78.42	16.41	7.42	3.52
23	**8**	90.54	15.97	6.45	2.36	**33c**	79.97	16.22	9.21	1.43
24	**35**	89.79	15.35	3.51	2.44	**33a**	77.87	16.89	16	0.51
25	**15**	90.97	15.93	5.24	3.96	**25**	78.05	15.55	5.75	7.82
26	**5**	89.86	15.53	7.88	1.63	**23**	74.71	14.78	6.04	4.42
27	**30**	89.03	15.75	7.45	2.19	**17**	75.13	14.89	5.4	8.49
28	**11**	88.95	15.44	8.18	1.5	**29**	78.19	16.21	8.39	6.22
29	**34d**	88.32	14.96	8.01	1.45	**18**	74.53	14.51	5.72	6.36
30	**34b**	88.11	15.21	9.19	1.49	**1**	76.89	15.46	8.68	6.16
31	**34a**	88.34	15.33	9.27	1.59	**30**	77.09	15.38	7.89	8.88
32	**6**	88.93	15.4	7.71	3.49	**12**	74.44	14.7	6.5	8.69
33	**7c**	88.4	15.13	9.31	1.66	**28**	76.91	16.07	13.1	13.8
34	**34c**	87.54	14.93	9.31	1.59	**15**	52.54	14.16	30.9	1.17
35	**2**	85.8	13.4	4.82	8.75	**13**	17.92	14.2	41.3	0.81
36	**21**	78.56	12.81	17.2	1.17	**5**	68.77	13.75	8.45	6.8
37	**22**	88.04	14.65	7.83	3.57	**21**	70.8	13.87	13.7	3.86
38	**7a**	87.27	14.52	8.19	5.65	**19**	66.41	16.25	15.2	7.92
39	**3**	85.5	14.36	11.4	2.71	**7b**	67.72	12.82	8.42	15.8
40	**16**	82.87	13.14	10.5	72.5	**20**	68.07	12.8	9.22	14.5
41	**13**	50.92	10.69	31.9	4.54	**27**	72.68	13.12	15.6	110
42	**31**	74.93	10.56	12.1	94.5	**31**	46.1	7.876	17.3	77.6
43	**32**	72.05	10.32	18	118	**32**	43.12	8.216	19.5	118
Otsu [3]		90.75	16.19	3.86	3.49	**Otsu [3]**	71.92	13.93	7.41	24.23
Sauvola et al. [4]		91.93	16.47	5.29	3.28	**Sauvola et al. [4]**	82.48	16.91	7.93	3.49

Bold face numerals are shown for "rank" and "method" categories in order to differentiate them from evaluation measures namely "F-measure", "PSNR", "NRM" and "MPM"

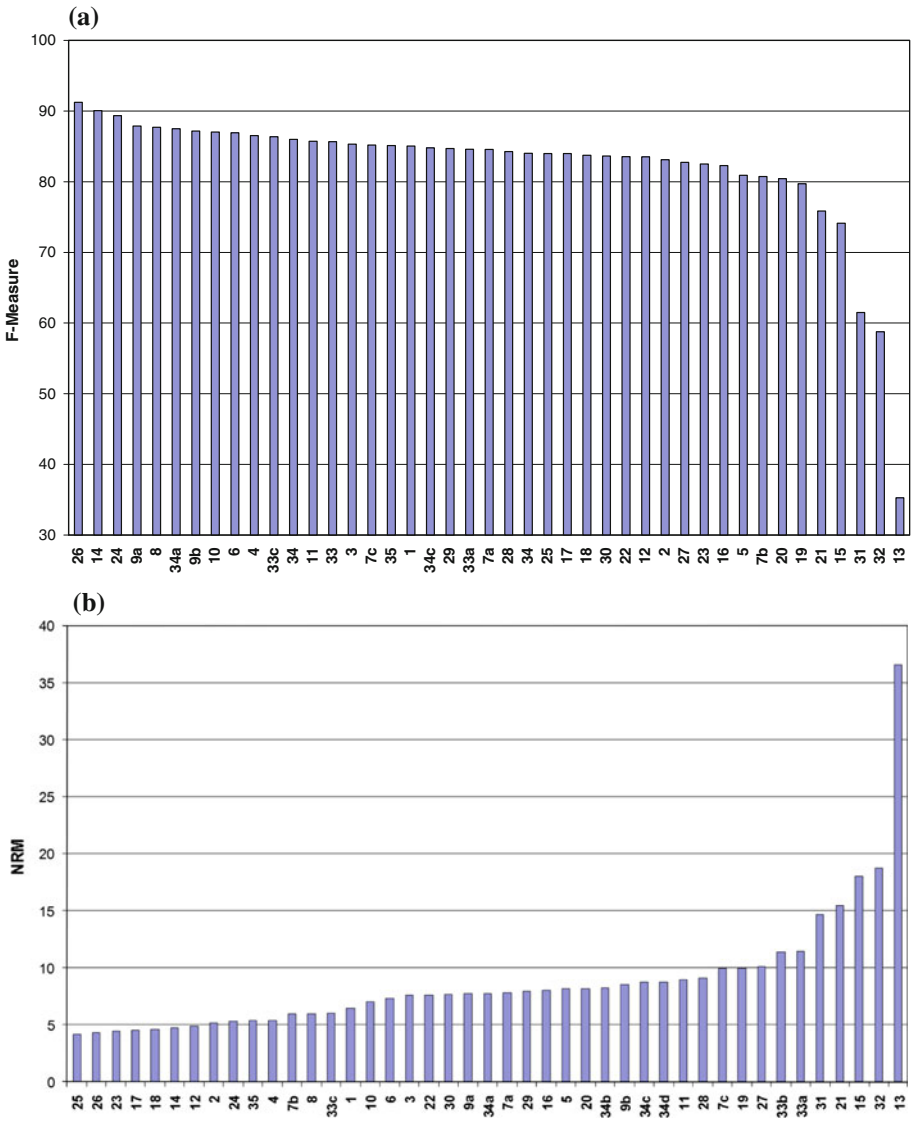

Fig. 3 Graphs that show the performance of the binarization algorithms submitted in DIBCO 2009 in terms of **a** F-measure and **b** NRM

because *GI* of stroke edge points is generally larger than that of other points. The stroke edges can therefore be extracted by Otsu's global thresholding method.

When the stroke edges are detected, document image pixels can be classified by Eq. 8 as follows:

$$R(i,j) = \begin{cases} 1 & N_e \geq N_{\min} \text{ AND } I(i,j) \leq E_{\text{mean}} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$E_{\text{mean}} = \frac{\sum_{\text{neighbor}} I(i,j) * E(i,j)}{N_e} \tag{9}$$

where I refers to the input document image, E refers to the binary edge image, *(i, j)* refer to the corresponding pixels in document image. The threshold $N_{\min}$ is defined by users, N_e refers to the number of edge points within the neighborhood window. E_{mean} (9) is the sum of the mean of the intensities of the edge points within the neighborhood window. So, if N_e is larger than $N_{\min}$ and *I(i, j)* is smaller than E_{mean}, *R(i, j)* is set at 1. Otherwise, *R(i, j)* is set at 0.

5.2 Algorithm No. 14: Université Pierre et Marie Curie & CMM, France (J. Fabrizio, B. Marcotegui)

The algorithm is based on the toggle mapping [5] morphological operator, and it is further developed to TMMS "Toggle Mapping Morphological Operation" for text localization in natural scenes [6]. According to Ref. [6], the algorithm outperforms standard thresholding criteria like Niblack's or Sauvola's. In the following, a detailed description of the algorithm is given.

The *toggle mapping morphological operator* maps a function f on a set of n other functions g_i by replacing the

value in each point of the function by the value of the local nearest function. The result r of the mapping is defined by the following Eq. (10).

$$\forall x : r(x) = g_i(x); \min_i(|f(x) - g_i(x)|) \tag{10}$$

A common use of this operator is contrast enhancement. The background-to-foreground segmentation is based on the toggle mapping operator. They choose to map the image I on two functions: the morphological erosion E of the image and the morphological dilation D of the image. Then, for each pixel, if the given pixel value is closer to the erosion, it is marked as background and if the pixel is closer to the dilation it is marked as foreground.

The aforementioned strategy handles efficiently boundaries of patterns but generates salt and pepper noise on homogeneous regions. To avoid this issue, pixels whose erosion and dilation are too close are excluded from the analysis. In other words, every pixel with the difference between the dilation and the erosion is under a threshold $t_{\min}$ is considered as included in a homogeneous region and is excluded from the analysis. Pixels are then classified into three classes: foreground (F), background (B) and homogeneous (H). Finally, homogeneous regions are assigned to foreground or background according to the class of their boundaries. This special treatment of homogeneous regions avoids a lot of noise in the background but may lead to miss major regions. Quality of results depends on the choice of $t_{\min}$ for which the selection is not trivial since a high value may lead to open contours while a low value keeps noise in homogeneous regions. A hysteresis threshold is used in order to reduce the critical effect of the threshold parameter. This hysteresis thresholding comprises two thresholds instead of one: (a) A high threshold (for DIBCO 2009, a multiple of the distance of the two modes of the histogram was assigned) to select regions and (b) a lower threshold to define the boundaries of the selected regions. Foreground regions or background regions in the low thresholded image are kept if and only if they have a seed (marker) in the high thresholded result. Otherwise, if low thresholded regions do not contain at least one pixel on the high thresholded image, they are classified as homogeneous. In order to improve the quality of the output, a parameter p is added to manage the thickness of patterns.

$$\text{tmms(x)}= \begin{cases} H & if\,(D(x) - E(x)) < t_{\min} \\ B & if\,(D(x) - E(x)) \geq t_{\min} \\ & AND\ (f(x) - E(x)) < p * (D(x) - f(x)) \\ \text{F} & \text{otherwise} \end{cases} \tag{11}$$

Notice that in this definition, F and B can be interchanged whether the text is on a darker or a lighter background.

5.3 Algorithm No. 24: University of Quebec, Canada (D. Rivest-Hénault, R. F. Moghaddam and M. Cheriet)

The method takes advantage of local probabilistic models and the calculus of variation. The statistics of the input image are used for the automatic estimation of the stroke width. Based on this, very small regions with small confidence scores are removed. The produced stroke map is eroded using a curve evolution approach implemented in the level-set framework using an energy term which measures the fitness of the stroke pixels with respect to the stroke gray level map. A detailed description of the algorithm is given in the following.

At first, an initialization map is required to identify high-probable text pixels. The remaining text pixels which may be degraded will be recovered by the local-linear evolution of the level-set function. For this purpose, we use one of multilevel classifiers [7], the stroke map (SM). In multilevel classifiers, there are many classifiers which uses different features to locate text pixels. Although the information at the pixel-level is helpful, a major part of the document image information is hidden in the spatial correlations. The classifiers at the content level, such a the SM, the stroke profile (SP) [7] and the stroke cavity map (SCM) [8], seek for this information based on the stroke-based features. In other words, theses classifiers try to use the document-related nature of the images. In the case of the SM, likelihood of having a stroke around the pixel under question is examined based on the structure of the text pixels around it. In this analysis, the average stroke width ws [9] is used to determine the possibility of a stroke around the pixel. In a new kernel-based approach, on a neighborhood of size $2ws+1$, a score is calculated based on which a SM value is assigned to the pixel. The SM can operate on different operation regimes. For the purpose of this work, which is to avoid as much as false-positive pixels, the SM is set to internal high-confidence operating mode. As it is obvious, the SM itself needs an initialization map to have an pre-estimation of the text pixels. We use the grid-based Sauvola's method [9] to generate a fast and adequate initialization map for the SM.

In a next step, the stroke map is eroded in order to keep only the pixel that can be considered as stroke with an even higher confidence level and to remove salt and pepper noise. This process that helps in the subsequent creation of an accurate stroke gray level map is done by using a level-set based curve evolution approach. Within this scheme, we use two distinct local-linear models to represent the expected intensity of the strokes and that of the background, respectively. Also, a curvature-based term, which tends to smooth the contour, and a balloon force, which reduces the area of the stroke region, are used. The level-set methodology that is used is similar to that presented in Ref. [8] and [10] where a level-set function ϕ is evolved with respect to an artificial time variable t. The initial position of the contour, $\phi\ (t\ = 0)$, is given

by the SM obtained at step 1 and we set ϕ so that $\phi \geq 0$ indicates the strokes and $\phi < 0$ indicates the background.

At a final step, a dense stroke gray level map (SGL) is created to represent the expected intensity of stroke pixels at any spatial position. Let $\phi(t = t_{\text{Erosion}})$ corresponds to the level-set function at the end of the preceding step. Then, this map is computed by using only the information of the stroke pixel, i.e. where $\phi\left(t = t_{\text{Erosion}}\right) \geq 0$. After that, the binarization is refined by using a level-set function similar to that of the previous step. However, in this case, the initialization is given by $\phi\left(t = t_{\text{Erosion}}\right)$, the balloon force is canceled and a term that measure the fitness of the stroke pixels with respect to the stroke gray level map is added to the curve evolution scheme. The final segmentation is given by a thresholding of the resulting level-set function $\phi \geq 0$.

5.4 Algorithm No. 10: Tsinghua University, China, (X. Shen)

It is a three-step binarization algorithm. The first step locates the text area according to edge information, the second step binarizes the text area, and the third step modifies the binarized result from semantic perspective.

At the first step, we convert the original image I_{ori} to gray scale I_{gray}, then count gradient image I_{grd} from it. We binarize I_{grd} to get an edge pixel set $S_{\text{edg}} = \left\{p : I_{\text{grd}}(p) > T_1\right\}$ with threshold $T_1 = 120$. We extract connected component from S_{edg}. Area within bounding boxes of all the connected components are regarded as 'text area', denoted as S_{text}.

At the second step, we first find a global threshold V_g of text area using adaptive thresholding. For all the pixels $p \in S_{\text{text}}$, we take their mean gray value as initial threshold. The following iteration will find the optimal threshold.

(a) Divide all the pixel to two sets: $S_1 = \left\{p : p \in S_{\text{text}}, I_{\text{gray}}(p) < V_g\right\}$, $S_2 = \left\{p : p \in S_{\text{text}}, I_{\text{gray}}(p) \geq V_g\right\}$
(b) Count mean value m_1, m_2 of S_1, S_2
(c) Update threshold $V_g = (m_1 + m_2)/2$
(d) Go back to (a). Repeat until the new threshold matches the one before it.

The optimal threshold achieved by the aforementioned procedure is in fact the iterative threshold proposed by Trussell [11]. We binarize all the pixels in S_{text} using V_g and regard all the non-text-area pixels as background to get a rough binary result I_{bw}.

At the final step, connected component analysis is run on all the black pixels in I_{bw}, and count the average bounding box height L. L can be regarded as character size in the whole document. Image height H, image width W and L give us some semantic information about the document. We have three steps to modify I_{bw}.

(i) To handle non-even illumination, we add local threshold V_l to V_g. For each pixel $p \in S_{\text{text}}$, local threshold $V_l(p)$ is defined as the mean value of a window around p of size $2L \times 5L$. The final threshold for each pixel is $V(p) = \alpha \cdot V_g + (1 - \alpha) \cdot V_l(p)$ with $\alpha = 0.7$. A refined result I_{bw} is achieved from S_{text} with V.
(ii) Once again, we extract connected components from black pixels of I_{bw} and label them as $C_i, i = 1, 2 \ldots N$. For each C_i, if its width exceeds $0.8\,W$, or its height exceed $0.5\,H$, and the black pixel of it is less than 5% of its total pixel number, we erase it. Such connected components are often shadow on the fringe of the page.
(iii) Similar to (ii), if C_i's width and height are smaller than $0.1\,L$, erase it as noise.

5.5 Algorithm No. 9a: Université de Lyon, INSA, France (C. Wolf, J-M Jolion)

The method features an adaptive thresholding algorithm which has been designed to increase the local contrast in the text image.

The authors propose to normalize the different elements used in the equation which calculates the threshold T at Niblack's algorithm [12], i.e. to formulate the binarization decision in terms of contrast instead of in terms of gray values, which is a natural way considering the motivation behind Niblack's technique. How can contrast be defined? In Niblack [12], the local contrast of the center pixel of a window of gray levels is defined as

$$C_L = \frac{|m - I|}{s} \tag{12}$$

where I is the gray value of the center pixel and, as mentioned earlier, m is the mean and s is the standard deviation of the gray values in the window. Since dark text is considered on bright background, points having a gray value which is higher than the local mean are not considered; therefore, the absolute value can be eliminated. Denoting by M the minimum value of the gray levels of the whole image, the maximum value of this local contrast is given by

$$C_{\max} = \frac{m - M}{s} \tag{13}$$

It is difficult to formulate a thresholding strategy defined only on the local contrast and its maximum value, since this does not take into account the variation of the window with respect to the rest of the image. This is the reason why the definition of a more global contrast is proposed; the contrast of the window centered on the given pixel is denoted as:

$$C_W = \frac{m - M}{R} \tag{14}$$

where R = max(s) is the maximum value of the standard deviations of all windows of the image. This contrast indicates whether the window is rather dark or bright with respect to the rest of the image (a high value implies the absence of text).

At a final step, a simple thresholding criterion is used to keep only pixels which have a high local contrast compared to its maximum value corrected by the contrast of the window centered on this pixel:

$$I : C_L > a(C_{\max} - C_W) \tag{15}$$

where a is a gain parameter. Developing this equation, the following threshold is obtained:

$$T = (1 - a)m + aM + a\frac{s}{R}(m - M) \tag{16}$$

In the case that the given pixel is the center of a window with maximum contrast, i.e. $s = R$, we get $T = m$, the algorithm is forced to keep the maximum number of points of the window. On the other hand, if the variation is low ($s \ll R$), then the probability that the window contains text is very low. Therefore, a pixel is only kept if its local contrast is very high. The threshold is denoted as $T \approx (1 - a)m + aM$. The gain parameter a allows to control the uncertainty around the mean value. A simple solution is to fix it at 0.5, which situates the threshold between m and M.

6 Conclusions

DIBCO 2009 attracted 35 research groups that are currently active in document image analysis. The increased interest in this competition is a twofold proof: first, it shows the importance of binarization as a step toward effective document image recognition and second, the need for pursuing a benchmark that will lead to a meaningful and objective evaluation.

References

1. Ntirogiannis, K., Gatos, B., Pratikakis, I.: An objective evaluation methodology for document image binarization techniques. Proceedings of 8th International Workshop on Document Analysis Systems (DAS'08), Nara, Japan, September, pp. 217–224 (2008)
2. Gatos, B., Ntirogiannis, K., Pratikakis, I.: ICDAR 2009 document image binarization contest (DIBCO 2009). Proceedings of 10th International Conference on Document Analysis and Recognition (ICDAR'09), Barcelona, Spain, pp. 1375–1382 (July 2009)
3. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**((1), 62–66 (1979)
4. Sauvola, J., Pietikainen, M.: Adaptive document image binarization. Pattern Recognit **33**, 225–236 (2000)
5. Jean, S.: Toggle mappings. In: Simon, J.C. (ed.) From Pixels to Features, North-Holland, Amsterdam (1989)
6. Fabrizio, J., Marcotegui, B.: Text Segmentation in Natural Scenes using Toggle-mapping. IEEE International Conference on Image Processing (2009), Cairo, Egypt, Nov. 2009
7. Farrahi Moghaddam, R., Cheriet, M.: RSLDI: restoration of single-sided low-quality document images. Pattern Recognit **42**, 3355–3364 (2009)
8. Farrahi Moghaddam, R., Rivest-Henault, D., Cheriet, M.: Restoration and Segmentation of Highly Degraded Characters Using a Shape-Independent Level Set Approach and Multi-level Classifiers. ICDAR'09, Barcelona, Spain, 26–29 July, pp. 828–832
9. Farrahi Moghaddam, M., Cheriet, R.: A multi-scale framework for adaptive binarization of degraded document images. Pattern Recognit **43**(6), 2186–2198 (2010)
10. Farrahi Moghaddam, R., Rivest-Henault, D., Cheriet, M.: A unified framework based on the level set approach for segmentation of unconstrained double-sided document images suffering from bleed-through. Proceedings of 10th International Conference on Document Analysis and Recognition (ICDAR'09), Barcelona, Spain, 26–29 July, pp. 441–445
11. Trussell, H.J.: Comments on "Picture thresholding using an iterative selection method". IEEE Trans. Syst. Man Cybern. SMC-9, No. 5, 311 (May 1979)
12. Niblack, W.: An Introduction to Digital Image Processing, pages 115–116. Englewood Cliffs, Prentice Hall, NJ (1986)