H-DIBCO 2010 – Handwritten Document Image Binarization Competition

Ioannis Pratikakis[1,2], Basilis Gatos[1] and Konstantinos Ntirogiannis[1]

[1]Computational Intelligence Laboratory, Institute of Informatics and Telecommunications
National Center for Scientific Research "Demokritos"
GR-15310 Agia Paraskevi, Athens, Greece
e-mail: {ipratika, bgat, kntir}@iit.demokritos.gr

[2]Department of Electrical and Computer Engineering
Democritus University of Thrace
GR-67100 Xanthi, Greece
e-mail: ipratika@ee.duth.gr

***Abstract* — H-DIBCO 2010 is the International Document Image Binarization Contest which is dedicated to handwritten document images organized in conjunction with ICFHR 2010 conference. The general objective of the contest is to identify current advances in handwritten document image binarization using meaningful evaluation performance measures. This paper reports on the contest details including the evaluation measures used as well as the performance of the 17 submitted methods along with a short description of each method.**

Keywords - handwritten document image, binarization, performance evaluation

I. INTRODUCTION

Handwritten document image binarization contributes significantly in the success of the handwritten document image recognition challenging task. Motivated by this, it is imperative to create a framework for benchmarking purposes, i.e. a benchmarking dataset along with an objective evaluation methodology in order to capture the efficiency of current image binarization practices for handwritten document images. To this end, following the success of DIBCO 2009 [1] organized in conjunction with ICDAR'09, the follow-up of this contest has been organized, namely the Handwritten Document Image Binarization Contest (H-DIBCO 2010) in the context of ICFHR 2010 conference. In this contest, we focused on the evaluation of document image binarization methods using a variety of scanned handwritten documents for which the corresponding binary ground truth image has been created. The authors of submitted methods had initially registered in the competition and downloaded representative document image 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 (10 handwritten images with the associated ground truth) along with the evaluation software has been released as publicly available in the following link:

(http://www.iit.demokritos.gr/~bgat/H-DIBCO2010/benchmark).

The remainder of the paper is structured as follows: Each of the methods submitted to the competition is briefly described in Section II. The evaluation measures are detailed in Section III. Experimental results are shown in Section IV while in Section V conclusions are drawn.

II. METHODS AND PARTICIPANTS

Sixteen (16) distinct research groups have participated in the competition with seventeen (17) different algorithms Either certain participants have submitted more than one algorithm or representatives from more than one research groups joined efforts in a single submission. Brief descriptions of the methods are given in the following (The order of appearance is based upon the order of submission of the algorithm).

1) National University of Singapore & Institute for Infocomm Research, Singapore (B. Su, S. Lu, C.L. Tan): The proposed method consists of four main steps. First, image contrast which is evaluated by local maximum and minimum is used to select the high contrast points. Second, the stroke edges which are extracted using Canny's method are combined with those high contrast points to produce a better edge map. Third, the document image is binarized by a local threshold which is decided based on the constructed edge map. Finally, some post-processing work is applied to improve the final result.

2) Ben-Gurion University, Computer Science department, Israel (I. Bar-Yosef, K. Kedem, I. Dinstein): The proposed approach is composed by several steps, mainly adaptive binarization, removal of false objects and accurate local region-based active contour. The initial binarization step is based on normalized image gradients

978-0-7695-4221-8/10 $26.00 © 2010 IEEE
DOI 10.1109/ICFHR.2010.118

and local intensity averaging. At the second step, we analyze each connected component whether it should be omitted or processed in the last phase. Finally, a fast and accurate active contour method is applied based on local region statistics.

3) South University of Toulon-Var, LSIS, UMR CNRS 6168, France (T. Lelore and F. Bouchara): The algorithm is composed of three different steps. First, text position is roughly estimated thanks to an edge detection approach. Next to the previously estimated text location, a clustering algorithm is applied in order to produce a three valued image (Text, Background, Unknown). Finally, a post-processing step assigns a class to 'Unknown' pixels thanks to heuristic rules.

4) EPITA Research and Development Laboratory (LRDE), France (T. Geraud, G. Lazzara): The method is based on a multi-scale implementation of Sauvola's binarization [2]. A post-processing is applied to remove small connected components and fill holes inside characters.

5) Synchromedia Laboratory, École de technologie supérieure, Montréal, Québec, Canada (R.F. Moghaddam and M. Cheriet): A generalized multi-scale adaptive binarization method [3] is implemented. The method uses the multi-level classifiers [4,5,6]. The key classifier is the estimated background. Thanks to automatic estimation of the parameters, the method is able to provide binarized document images without any need for human interaction. The estimated parameters are the average stroke width and the average line height. In order to obtain better results, post-processing steps are also performed.

6) EPITA Research and Development Laboratory (LRDE) & MINES ParisTech, Centre de morphologie mathématique, Mathématiques et Systèmes (CMM), Fontainebleau, France (J. Fabrizio, B. Marcotegui): The algorithm used is the same as the one used in the DIBCO contest in the framework of ICDAR 09 [1][7] and is based on the toggle mapping operator [8]. The image I is mapped 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'. To reduce noise in homogeneous regions, pixels whose erosion and dilation are too close are excluded from the analysis. In other words, every pixel p with the difference between the dilation and the erosion is under a threshold t is considered as included in an homogeneous region and it is excluded from the analysis. Pixels are then classified into three classes: 'foreground', 'background' and 'homogeneous'. Finally, 'homogeneous' regions are assigned to 'foreground' or 'background' according to the class of their boundaries. Quality of results highly depends on the choice of t and this value is difficult to choose. A hysteresis threshold is used in order to reduce the critical effect of the threshold parameter. Since the version used in DIBCO contest, two main improvements are to be noticed.

The first improvement is that color images are now segmented several times. Each channel and the luminance are segmented. The final result is the union of results from each channel. Segmenting only the luminance leads to missing some regions.

The second improvement is that 'background' is estimated and removed. The choice of t is difficult to set and the value is the same for the whole document but document may not be homogeneous (especially background variation due to stains). The background is coarsely detected by a large opening and then removed from the image. The resulting document is more homogeneous and the use of a unique t for the whole document is less problematic.

7) Institute of Space Technology, Pakistan (K. Khurshid): In the proposed algorithm NICK, where the thresholding formula has been derived from the basic Niblack algorithm [9], binarization threshold is found out for each pixel by taking into account its neighbouring pixels in a sliding window using the formula in the following:

$$T = m + k\sqrt{\frac{(\sum_{i=1}^{NP} p_i^2 - m^2)}{NP}} \tag{1}$$

where ,

k is the NICK factor ranging between –0.2 and –0.1

p_i = pixel value of gray scale image

NP = number of pixels in the window

m = mean gray value of the NP pixels

During experiments it has been observed that one major advantage of NICK over Niblack is that it considerably improves binarization for 'white' and 'light' page images by shifting down the binarization threshold to ensure that no 'non-text' areas are taken erroneously as 'text'. The value of NICK factor k can vary from -0.1 to -0.2 depending upon the application requirement. Value of k close to -0.2 makes sure that noise is all but eliminated but characters can break a little bit, while with values close to -0.1, some noise pixels can be left but the text will be extracted crisply and unbroken.

8) CMM, Mines Paristech, France (J. Hernandez): This method is based on a morphological operator named ultimate opening (UO). The method consists of three steps: Firstly, ultimate openings of height and width attributes are carried out in order to extract the most contrasted structures in both horizontal and vertical directions. UO provides two pieces of information, contrast R(I) and size q(I). Then, a classical Otsu's binarization [10] is performed from contrast output. Finally, we apply a post-processing step: eliminate small and isolated structures, and remove the connections to the background.

9) NifiSoft, Saint-Etienne, France (A. Hassaïne): The proposed method classifies each pixel as 'foreground' or 'background' taking into account its global k-means segmentation, the otsu segmentation, the values of its neighbors in these two segmented images, the values of the basic morphological operations with several sizes and the values of gaussian filters with several sizes. Moreover, if the color information is present, the hue and saturation gradient are also taken into account. All these descriptors are combined using a logistic regression to perform the classification.

10) Université de Strasbourg, Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection - Équipe Modèles, Images et Vision (MIV), France (B. Perret): The proposed algorithm is based on two steps: (i) background removal and (ii) adaptive thresholding. Foreground objects are identified with an analysis of a modified version of the connected component tree (taking advantage of the hyper-connection theory to allow non-flat nodes). For each leaf of the tree, the background level is defined by a criterion on the evolution of the "area/gray level" curve [11]. Then, the edges of the flattened image are detected using a Sobel operator and a global Otsu threshold. The local threshold value is then determined by the pixel values of the edges.

11) Brigham Young University, USA (O. Nina): This method combines a recursive version of the Otsu algorithm after a preprocess step of background normalization and smoothing using a bilateral filter and a final despeckle step.

12) Synchromedia Laboratory, École de technologie supérieure, Montréal, Québec, Canada (D. Rivest-Henault, R.F. Moghaddam and M. Cheriet): The proposed method can automatically process both color and gray-level images. In the first step, the input image is converted to a gray-level image. Then, in several steps, the binarized version of the input is created. The core of the method is based on the multi-level classifiers. These classifiers are capable to extract and identify information on different levels from local to global. On each level, a set of parameters is used as the a priori information of the document image. In an automatic way, these parameters are estimated by analysis of the input image. In this method, the most important level is the content level. Among many classifiers on this level, stroke map classifier is the key one. Stroke map tries to capture pixels on the document image which may belong to the text strokes. This is achieved by analysis of image structures around the target pixel. This kernel-based classifier depends on an a priori parameter, i.e. the estimated stroke width. This parameter is computed automatically at the beginning of the process. In order to obtain better results, removal of the noise pixels is also performed.

13) SMCC, Jadavpur University, Kolkata, India (A.F. Mollah): In the proposed method first, the given image is pre-processed and then, edges are extracted. Some edges are filtered out and the remaining is used to extract text regions from the original image. These text regions are partitioned and then a global binarization technique (i.e. Otsu) is applied on each such partition.

14) Smith College, MA, USA (N.R. Howe): This algorithm is built upon recent work in figure/ground segmentation for video. It uses the Laplacian operator to assess the local likelihood of foreground and background labels, Canny edge detection to identify likely discontinuities, and a graph cut implementation to efficiently find the minimum energy solution of an objective function combining these concepts.

15) MCKV Institute of Engineering, Dept. of CSE, Howrah, India & Jadavpur University, CSE Department, Kolkata, India (S. Saha, S. Basu, M. Nasipuri, D.K. Basu): In the proposed technique, histogram equalization is applied over the whole image as well as at different levels of localization or partition independently over the original grey values of the pixels in those localized areas. Each of these partitions is again sub-divided into four partitions. At any level, whenever the image is histogram equalized, each pixel gets a new grey value increasing contrast with its neighbours. This new grey value divided by 255 provides a membership of greyness for each pixel. Membership value is an indication of the inclination of the pixel towards black or white. Each pixel gets a set of membership values for global operation as well as for different levels or depth of local operations. Ultimately, all the membership values of each pixel that are obtained for different levels are combined to get the net membership value for each pixel. Net membership value is the weighted average of all the membership values for each pixel with respect to its grey value. During calculation, membership values obtained from local histogram equalization are given more weight than membership values obtained from global histogram equalization. Each pixel is then binarized depending on whether the net membership value of that pixel crosses 0.5 or not. If it crosses 0.5 then it is decided as 'white' otherwise, it is decided as 'black'.

For proper binarization, selecting the number of level or depth becomes an issue. It depends on two factors:

A. Size of a partition: During the process of division, the height and width of each partition is divided by 2 provided each having a value of greater than 2. If one of either height or width reaches that limit then its value is 'frozen' and the other's value is divided by 2 for subsequent level of partitioning. When both height and width reaches the limit, further division of sub-images is stopped.

B. Standard deviation of grey values of pixels in each partition: For any partition at any level, the standard deviation of the grey values of the pixels in that partition is calculated. If it is very low compared to the 0^{th} level of standard deviation then the corresponding partition is not histogram equalized. Rather, if the grey value of a pixel in that partition is less than the mean grey value of that partition then its membership value is given as 0.4 (i.e.

forcing it towards black) otherwise, its membership value is given as 0.8 (i.e. forcing it towards white).

16) Jean Monnet University St. Etienne, FRANCE (S. Karaoglu): This methodology presents an innovative methodology for binarization which does not require any parameter tuning by the user; and can deal with degradations which occur due to shadows, non-uniform illumination, low-contrast, large signal-dependent noise, smear and strain. A pre-processing procedure based connected opening for image enhancement has been applied to suppress the dark structures in the image. Difference of gamma functions in approximation with Generalized Extreme Value Distribution has been used to realize the proper threshold.

17) Technological Educational Institute (TEI) of Athens, Greece (A. Nikolaou): This method is a naive/rough exploration of the idea to use Integral Images of Histograms for binarization. In this implementation we assumed that most problems arise from dark background pixels due to stains etc. Using the Integral Image of Histograms, we can obtain the histogram of any rectangular region in constant complexity. For each pixel, the otsu threshold of a square window with the size of the minimum dimension of the grayscale image is attributed. As long as it is classified as 'foreground' we reduce the window by a factor of 11/20 up to 5 times.

III. EVALUATION MEASURES

For the evaluation, the measures used comprise an ensemble of measures that have been widely used for evaluation purposes. These measures consist of (a) F-Measure; (b) pseudo F-Measure; (c) PSNR; (d) Negative Rate Metric and (e) Misclassification Penalty Metric.

A. F-Measure

$$FM = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{2}$$

where $\text{Recall} = \frac{TP}{TP+FN}$, $\text{Precision} = \frac{TP}{TP+FP}$

TP, FP, FN denote the True positive, False positive and False Negative values, respectively.

B. pseudo F-Measure

This measure has been introduced in [12]. It was motivated by the fact that each character has a unique silhouette which can be represented by its skeleton. In this respect, we assume that a perfect recall can be achieved in the case that each skeleton constituent of the ground truth has been detected. Compared with the typical F-Measure as presented in III.A, there exist a difference which concerns an alternate measure for recall, namely *pseudo-Recall (p-Recall)* which is based on the *skeletonized ground truth image*.

The skeletonized ground truth image is defined by the following equations:

$$SG(x,y) = \begin{cases} 0, & \text{background} \\ 1, & \text{text} \end{cases} \tag{3}$$

Taking into account the skeletonized ground truth image, we are able to automatically measure the performance of any binarization algorithm in terms of recall.

p-Recall is defined as the percentage of the skeletonized ground truth image ***SG*** that is detected in the resulting MxN binary image ***B***. p-Recall is given by the following equation:

$$\text{p-Recall} = \frac{\sum_{x=1,y=1}^{x=M,y=N} SG(x,y)\cdot B(x,y)}{\sum_{x=1,y=1}^{x=M,y=N} SG(x,y)} \; 100\ \% \tag{4}$$

$$p-FM = \frac{2 \times \text{p-Recall} \times \text{Precision}}{\text{p-Recall} + \text{Precision}} \tag{5}$$

C. PSNR

$$PSNR = 10\log\left(\frac{C^2}{MSE}\right) \tag{6}$$

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

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 *M*x*N* images is. We consider that the difference between foreground and background equals to *C*.

D. Negative Rate Metric (NRM)

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:

$$NRM = \frac{NR_{FN} + NR_{FP}}{2} \tag{7}$$

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

N_{TP} denotes the number of true positives, N_{FP} denotes the number of false positives, N_{TN} denotes the number of true negatives, N_{FN} denotes the number of false negatives.
In contrast to F-Measure and PSNR, the binarization quality is better for lower NRM.

E. 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.

$$MPM = \frac{MP_{FN} + MP_{FP}}{2} \quad (8)$$

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 the distance of the i^{th} false negative and the j^{th} false positive pixel from the contour of the text in the GT image. 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.

IV. Experimental Results

The H-DIBCO testing dataset consists of 10 handwritten document images for which the associated ground truth was built for the evaluation following a semi-automatic procedure based on [12]. Representative examples of the dataset along with the associated ground truth images are shown in Fig. 1(a),(e) and Fig. 1(b),(f), respectively. The document images of this dataset originate from the collections of the Library of Congress [13]. 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).

TABLE I. Evaluation results wrt to the measures used for all methods submitted to H-DIBCO 2010

Rank	Method	FM (%)	p-FM (%)	PSNR	NRM ($x10^{-2}$)	MPM ($x10^{-3}$)
1	**1**	**91,50**	**93,58**	**19,78**	**5,981**	**0,492**
	2	**89,70**	**95,15**	**19,15**	**8,180**	**0,288**
2	3	91,78	94,43	19,67	4,771	1,334
3	14	89,73	90,11	18,90	5,776	0,412
4	10	87,98	90,83	18,26	7,677	0,377
5	13	86,85	92,43	18,19	9,989	0,231
6	8	86,13	88,8	17,62	8,686	0,378
7	17	85,71	91,68	17,63	10,42	1,188
8	16	83,22	91,24	17,19	13,15	0,507
9	12	85,06	89,63	17,56	10,48	3,807
10	9	83,51	86,88	17,24	13,02	0,949
	11	82,99	87,55	17,02	12,83	0,695
11	15	81,39	81,91	15,60	5,534	1,666
12	6	84,95	86,89	16,82	11,47	48,63
13	7	82,29	89,56	16,61	13,19	2,844
14	5	73,51	78,96	15,95	19,95	1,044
15	4	57,73	66,42	14,29	28,41	1,107

The evaluation was based upon the five distinct measures presented in Section III. At Table I, the detailed performance of each algorithm for each encountered measure is given. The final ranking as shown in Table I was calculated after sorting the accumulated ranking value for all measures. Specifically, 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 the evaluation measures. As denoted in (9), for each binarization method, the final ranking S_i is achieved by the five rankings summation.

$$S_i = \sum_{j=1}^{5} R(i,j) \quad (9)$$

Overall, the best performance is achieved by two algorithms that have been equally performed taking into account the final ranking over all measures. The top ranked algorithms are : ***Algorithm 1*** which has been submitted by B. Su, S. Lu and C.L. Tan as collaboration between the National University of Singapore and the Institute for Infocomm Research in Singapore and ***Algorithm 2*** which has been submitted by I. Bar-Yosef, K. Kedem, I. Dinstein from the Ben-Gurion University in Israel. Example binarization results of those two algorithms are shown in Fig. 1(c),(g) and Fig. 1(d),(h), respectively.

V. Conclusions

The H-DIBCO 2010 Handwritten Document Image Binarization Contest attracted 16 research groups that are currently active in document image analysis. The general objective of the contest is to identify current advances in handwritten document image binarization using meaningful evaluation performance measures. This objective is fulfilled by firstly, providing short descriptions of each submitted algorithm, thus, enabling the interested researchers to be aware of the highly performing algorithms and be able to push forward the state of the art by a new more advanced approach. Secondly, the public availability of the testing dataset and the evaluation software permits further benchmarking and comparison with H-DIBCO results. The authors hope that this effort will stimulate fruitful discussions which will provide substantial aid towards advancing the state of the art in handwritten document image binarization.

Acknowledgment

This work has been partially funded by the European Community's Seventh Framework Programme under grant agreement n° 215064 (project IMPACT).

References

[1] B. Gatos, K. Ntirogiannis and I. Pratikakis (2009), "ICDAR 2009 Document Image Binarization Contest (DIBCO2009)", 10th International Conference on Document analysis and Recognition (ICDAR'09), July. 26-29, 2009, Barcelona, Spain, pp. 1375-1382.

[2] J. Sauvola and M.Pietikainen, “Adaptive Document Image Binarization”, Pattern Recognition, vol. 33, no. 2, pp. 225-236, 2000.

[3] R.F. Moghaddam and M. Cheriet, “A multi-scale framework for adaptive binarization of degraded document images”, Pattern Recognition, 43:2186-2198, June 2010.

[4] R.F. Moghaddam and M. Cheriet, “RSLDI: Restoration of single-sided low-quality document images”, Pattern Recognition, 42:3355-3364, 2009.

[5] R.F. Moghaddam and M. Cheriet, “EFDM: Restoration of Single-sided Low-quality Document Images”, ICFHR'08, 2008, 204-209.

[6] R.F. Moghaddam and M. Cheriet, “Application of Multi-level Classifiers and Clustering for Automatic Word-spotting in Historical Document Images”, ICDAR'09, 2009, 511-515.

[7] J. Fabrizio, B. Marcotegui and M. Cord, “Text segmentation in natural scenes using Toggle-Mapping”, IEEE International Conference on Image Processing (ICIP'09). Cairo, Egypt, November 7-10, 2009.

[8] J. Serra, “Toggle mappings”, From pixels to features. J.C. Simon (ed.), North-Holland, Elsevier (1998).

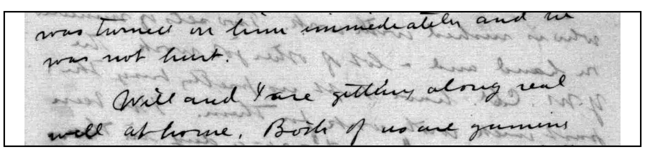
[9] W. Niblack, An Introduction to Digital Image Processing, Englewood Cliffs, NJ: Prentice-Hall, 1986, pp. 115–116

[10] N. Otsu, “A Thresholding Selection Method from Gray-level Histogram”, IEEE Trans. Systems, Man Cybern., vol. 9, pp. 62-66, Mar. 1979.

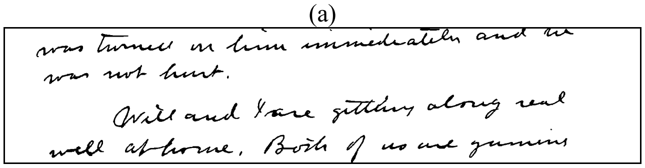
[11] B. Perret, S. Lefèvre, C. Collet, and E. Slezak. “From hyperconnections to hypercomponent tree : Application to document image binarization”, In Workshop on Applications of Digital Geometry and Mathematical Morphology, August 2010.

[12] K. Ntirogiannis, B. Gatos and I. Pratikakis, An Objective Evaluation Methodology for Document Image Binarization Techniques, The 8th IAPR International Workshop on Document Analysis Systems (DAS 2008), Nara Prefectural New Public Hall, Nara, Japan, September 17-19, 2008, pp. 217-224.

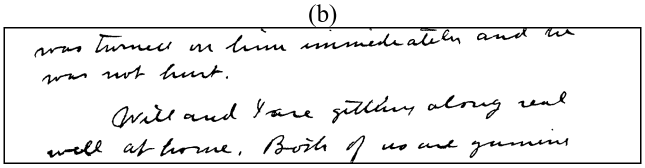
[13] Library of Congress (http://memory.loc.gov/ammem).



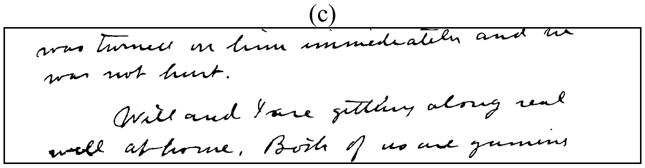
(a)



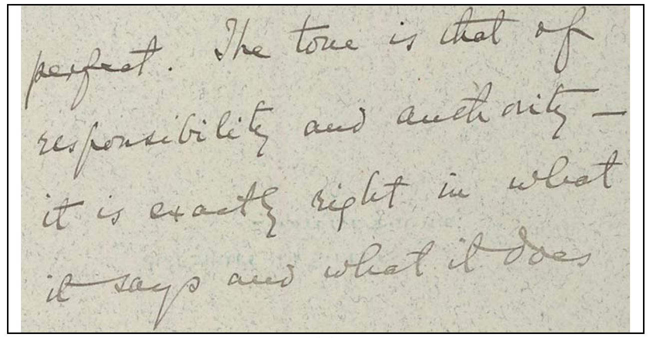
(b)



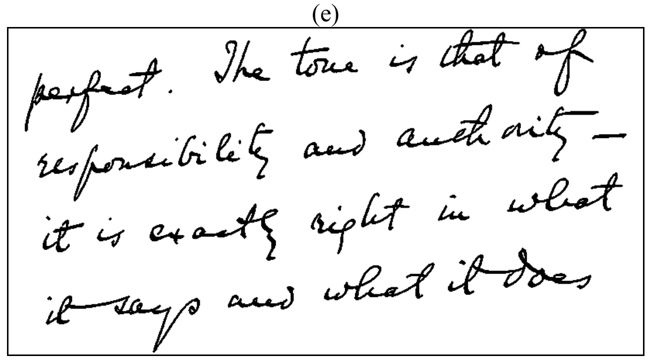
(c)



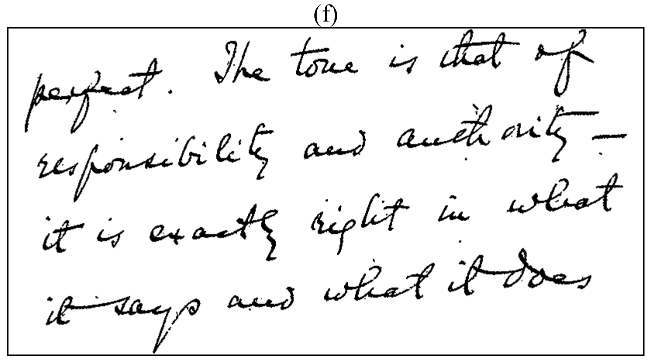
(d)



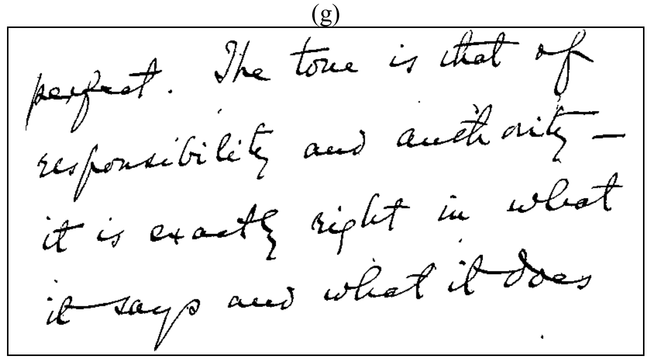
(e)



(f)



(g)



(h)

Figure 1. (a),(e) Representative original handwritten images included in the testing dataset; (b),(f) Ground truth image; (c),(g) Binarization results from ***Algorithm 1***; (d),(h) Binarization results from ***Algorithm 2***