

Document Image Binarization

Konstantinos Ntirogiannis*

¹ National and Kapodistrian University of Athens
Department of Informatics and Telecommunications
`kntir@di.uoa.gr`

² National Centre for Scientific Research “Demokritos”
Institute of Informatics and Telecommunications
Computational Intelligence Laboratory
`kntir@iit.demokritos.gr`

Abstract. Principal stage of the document image analysis procedure is the binarization, according to which the pixels are classified into text and background. It is a crucial stage that can affect further stages including the final character recognition stage. This thesis is focused on document image binarization, including both binarization techniques and evaluation methodologies. Specifically, according to the developed performance evaluation methodologies, the pixel-level ground-truth image is constructed using a semi-automatic procedure based on the edges and the skeleton of the characters. The new measures use (a) weights that start from the ground truth contour and (b) the local stroke width to limit the weights close to the character areas and to properly normalize those weights. Experimental results prove the validity and effectiveness of the new measures for document images, while other measures concern the image or signal processing area in general. Concerning binarization techniques, some improvements were initially proposed for the well-known technique of Yang&Yan. To further enhance the quality of binarization and be more robust against different types of degradations (e.g. faint characters, bleed-through and non-uniform background), a new binarization technique was developed that was based on background estimation and on the combination of selected global and local binarization techniques. Additionally, a binarization technique was developed for the binarization of the text areas captured from video content. This technique is also based on the Yang&Yan binarization technique and sets low and high values in its global parameter for the inside and outside area of the text. Initially, the definition of the text areas is based on the baselines of the text and at the final stage the text areas are better defined by the convex hulls of neighbouring textual components. Furthermore, through the document image binarization contests that we organized, a publicly available benchmark has been created that aids in the development of document image binarization techniques and evaluation methodologies.

Keywords: pre-processing, binarization, evaluation metrics, ground-truth image, historical document image processing

* Dissertation Advisors:¹Sergios Theodoridis, Professor -²Basilis Gatos, Researcher

1 Introduction

Document image binarization (or thresholding) is the process that segments the grayscale or color document image into text and background by removing any existing degradations (such as bleed-through, large ink stains, non-uniform illumination and faint characters). It is an important pre-processing step of the document image processing and analysis pipeline that affects further stages as well as the final Optical Character Recognition (OCR) stage. This thesis is focused on document image binarization, including both binarization techniques and evaluation methodologies. Our core motivation for the binarization was to develop an easy to tune method that could be effective against characters of various sizes [1], as well as against many different degradation types [2] (e.g. faint characters and bleed-through). Apart from the binarization of document images, we developed a method for the binarization of textual content from video frames [3].

As far as the developed evaluation methodologies are concerned [4, 5], we were motivated by the fact that existing pixel-based evaluation measures concern the image or signal processing area in general, while for document image processing those measures do not always provide reliable results. Last but not least, using the ground-truth construction procedure of our methodology [5], we successfully organized Document Image Binarization Competitions (DIBCO) from 2009 to 2012 [6–10] and made publicly available the competition datasets. Therefore, we have created a benchmark which is widely used for the development of document image binarization techniques and evaluation methodologies.

In the following, in Section 2 we present the related works concerning the binarization methods along with the binarization methods developed through this thesis, in Section 3 we present the related works concerning the evaluation methodologies along with the evaluation methods developed through this thesis. In Section 4 we present the experimental results and finally, in Section 5, the conclusions are drawn.

2 Binarization Methods

2.1 Related Work

Many document image binarization methods have been proposed which are usually classified in two main categories, namely global and local. Reference points in binarization are considered the global thresholding method of Otsu [11] and the local adaptive methods of Niblack [12] and Sauvola et. al. [13] which are widely incorporated in binarization methods that followed, e.g. Kim et. al. [14], Gatos et. al. [15], Lu et. al. [16]. Certain document image binarization methods have incorporated background estimation and normalization steps, e.g. Gatos et. al. [15], Lu et. al. [16], as well as local contrast computations to provide improved binarization results, e.g. Su et. al. [17], Howe [18]. Other binarization

methods, aiming in an increased binarization performance, proposed combination methodologies of binarization methods, e.g. Gatos et. al. [19], Su et. al. [20].

As far as the video frame binarization is concerned, there exist several techniques that perform binarization on the textual content in video frames aiming in an improved OCR performance. Many techniques [21–23] incorporate modifications of well-known binarization techniques, such as the Logical Level Technique of [24], Otsu [11] and Sauvola et. al. [13]. Other techniques [25, 26] are based on training using mainly SVM (Support Vector Machine) classifier or convolutional neural network. In the most recent related work [27], the Canny edge detector [28] was used to specify the text boundaries on the image. Then, a flood fill algorithm was used to fill the edge contour and form the characters. However, the Canny edges can be very confusing since they also depict non-text objects. Especially, in videos with high background complexity the edges of text may connect with background edges and hence deform the actual contour of the characters.

2.2 Improvement of Yang&Yan Method

The method of Yang&Yan [29] assumes a single stroke width for the document image. The value of the stroke width determines the size of the windows that are used to calculate the threshold at each point. However, characters of various sizes may exist within a document (e.g. a newspaper with big titles). To adaptively define the stroke width and consequently the size of the windows, we rely on the binarization output of [15]. Then, we detect the contour points and the skeleton using skeletonization method [30]. Afterwards, the local stroke width is assigned to each skeleton point by measuring the distance of that skeleton point from the nearest contour point. Then, each remaining point inherits the value of the nearest skeleton point found. However, for machine-printed documents that may suffer from internal holes at their strokes, the maximum of the local stroke widths is considered. All the aforementioned stages are shown in Fig. 1. Another improvement is the modification of the local threshold by a factor β ($T' = \beta \cdot T$). According to [1], this factor enhances the overall performance, especially for machine-printed documents. Representative results are shown in Section 4.

2.3 A Combined Binarization Approach

In degraded historical images, faint characters and bleed-through have quite similar characteristics. Thus, current methods are usually robust against one of the aforementioned degradation. In [2], we introduced a binarization method capable of achieving high performance in many different noise types. The main idea is to initially erase all the noisy components (false alarms) even if faint character parts are also removed. Then, perform binarization of high *Recall* such as Niblack [12] and perform combination at connected component level. In this way, noise is erased, the faint characters are completely detected, while the noise

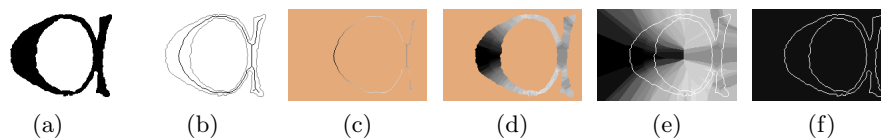


Fig. 1. The stages of the adaptive stroke width detection: (a) initial binary image; (b) contour points along with the skeleton; (c) local stroke width is assigned to each skeleton point; (d) the character stroke width image; (e) the final stroke width map (used for handwritten documents); (f) the final stroke width map (used for printed documents).

levels are very low. All the aforementioned stages are detailed below and are shown in Fig. 2:

1. Niblack binarization ($w=60 \times 60$, $k=-0.2$) and one iteration of dilation (3×3 element),
2. estimate the background, follow proposed inpainting [2] using the above Niblack result as inpainting mask,
3. normalize original image with the above estimated background (keep the range of the original image),
4. Otsu binarization and remove connected components of very small height,
5. calculate: (a) the stroke width map using the above binary image and (b) the global contrast,
6. Niblack binarization with window size and parameter k based on the stroke width map and the global contrast, respectively,
7. combination at connected component level. Large Niblack components that correspond to only a few foreground pixels of Otsu are not considered,
8. enhance the final result using binary image of step (4) (before the components removal).

2.4 Thresholding of Video Text Areas

For the binarization of video frames, we assume that the text detection step has already been performed and we focus on the binarization step of the detected text boxes. We introduced in [3] a binarization technique that aims in improving the text/background separation. The main idea is to specify the main body of the text (Fig. 3a-3c) in order to extract valuable information concerning the textual content. The main body of the text is defined as the area which is limited by the upper and lower baselines. Then, within the main body of the text we detect the stroke width (SW) of the characters which is used in consecutive adaptive binarization steps that follow. At a next step, we perform adaptive binarization [29] with different valuation in parameters for the inside and outside area of the main body of the text (Fig. 3d). Hence, we remove most of the non-text information but in certain cases it results in the thinning and breaking of the textual

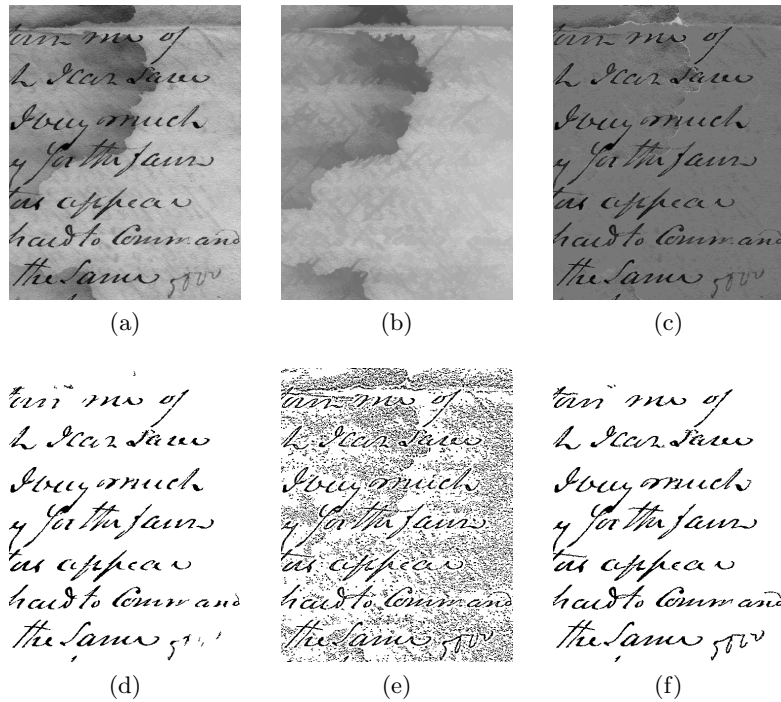


Fig. 2. The stages of the combined binarization approach: (a) original image; (b) estimated background; (c) normalized image; (d) Otsu binarization of (c) and small components removal; (e) Niblack binarization; (f) final combined result.

parts that are outside the main text body. Afterwards, we define the entire text body as the region inside the convex hulls of continuous connected components (Fig. 3e) and we perform the same adaptive binarization with different valuation in parameters for the inside and outside area of the entire text body (Fig. 3f).

3 Binarization Evaluation Methods

3.1 Related Work

Several efforts have been presented that strive towards evaluating the performance of document image binarization techniques. These efforts can be classified in three main categories (the human-oriented, the OCR-based and the pixel-based).

In the first category, evaluation is performed by the visual inspection of one or many human evaluators [31, 32]. For example, in [31], the amount of symbols that are broken or blurred, the loss of objects and the noise in background and foreground are used as visual evaluation criteria. In the second category, evaluation is addressed taking into account the OCR performance. The binarization

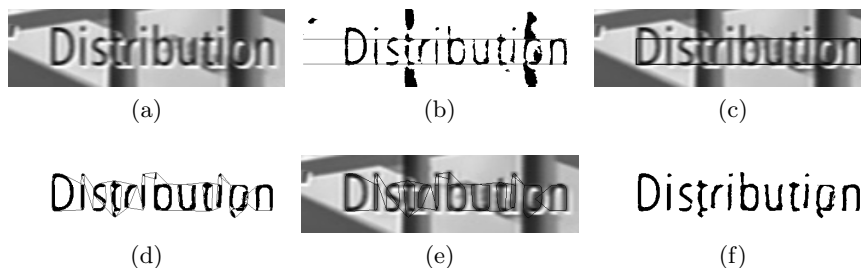


Fig. 3. The stages of the binarization method for video text areas: (a) original image; (b) binarization using [1] to detect the baselines; (c) main body defined by the baselines; (d) binarization [29] of (c) along with the convex hulls of neighbouring components; (e) main body defined by the convex hulls; (f) final binarization.

outcome is subject to OCR and the corresponding result is evaluated with respect to character and word accuracy [15, 27]. In the third category, pixel-based evaluation is used by taking into account the pixel-to-pixel correspondence between the ground truth and the binarized image. In this category, the evaluation is based either on synthetic images [33, 34] or on real images [35]. Ground truth images from real degraded images which correspond to real “challenging” cases for document image binarization were not publicly available. The Document Image Binarization (DIBCO) contests that were organized by us [6–10] made the datasets publicly available after each corresponding contest.

Concerning pixel-based evaluation, several measures have been used for the evaluation of document image binarization techniques, such as the F-Measure (Recall and Precision), the PSNR, the Negative Rate Metric (NRM) and the Misclassification Penalty Metric (MPM) [6], the chi-square metric [36], the geometric-mean accuracy [34], the normalized cross-correlation metric [35] and the DRD (Distance Reciprocal Distortion) [37]. Some researchers have stated the need for an improved pixel-based evaluation measure for document image binarization. For instance, in [35], wherein the ground truth generation from several users was studied, it was stated that there is a need for a weighted measure in relation to the ground truth borders in order to compensate the subjectivity of the ground truth.

3.2 Skeleton based Methodology

This method was presented in [4]. It consists of a semi-automatic procedure for the ground-truth construction and it also introduces the use of the skeleton of the characters for the evaluation of binarization output in terms of “Recall”. However, the ground-truth construction procedure has certain issues which were resolved in the latest evaluation methodology presented in [5]. Thus, in this section we will focus on the evaluation stage and not at the ground-truth construction procedure.



Fig. 5. Problematic cases concerning the skeletonized ground-truth: (a) original image; (b) ground-truth image along with the skeletonized ground-truth; (c) binarization output wherein the skeletonized ground truth is fully detected.

few factors that should be considered when penalization weights are based on the distance from the ground-truth contour. These factors are listed below:

- a breaking at a small/thin character part would have much less penalty than a bigger/thicker character part;
- noise inserted among the characters would have much less penalty than attached to a single character;
- noise attached to a big/thick character is less important than attached to a smaller/thinner character;
- noise far from the ground-truth that do not interfere with the textual content would be much grater penalized than noise among the characters that destroys the useful textual content.

In [5], we proposed proper weighting to minimize/diminish the effects of the aforementioned factors. In particular, to measure the amount of loss, the pseudo-Recall was introduced by which the distance-based weights are normalized according to local stroke width. In this way, each character breaking has the same importance regardless of the local thickness. Additionally, to measure the amount of the inserted noise, the pseudo-Precision was introduced according to which the weights are constrained within an area that extends to the background by the corresponding stroke width of each character. In this area the weights take values from 1 to 2, while outside this area the weights equal one. In this way, noise that is located among the characters has higher significance, while noise far from the ground-truth does not get exaggerating penalty. Furthermore, the distance between the characters is also considered to handle the cases of noise among the characters that result in merging.

The metrics of Recall and Precision are combined into F-Measure. Hence, the proposed pseudo-Recall/Precision are combined into pseudo-FMeasure F_{ps} . After many test cases examined in [5], the proposed pseudo-FMeasure offers more reliable results and it also has greater consistency to the OCR results. Representative results are given through Fig. 6 and Table 1.

4 Experimental Results

In this section the experimental results for the binarization of document images are shown. In the following, Fig. 7 shows representative results of the developed

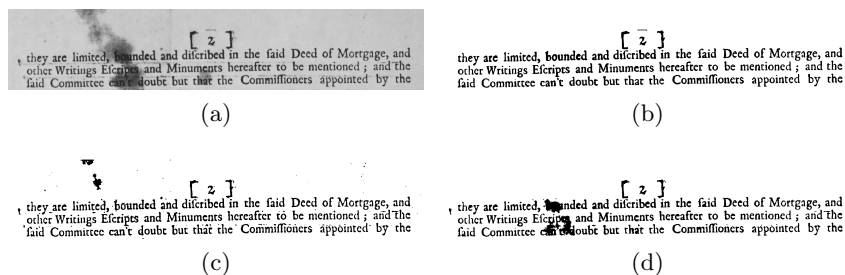


Fig. 6. (a) Original image; (b) ground-truth; (c) binarization where text is preserved but with background noise; (d) binarization with stains among the text.

Table 1. Comparison between existing and the proposed (F_{ps}) pixel-based measure. The OCR accuracy is also shown. Notice that for metrics MPM and DRD lower values denote higher performance.

	FM	PSNR	MPM	DRD	F_{ps}	OCR accuracy
Fig. 4b	94.37	22.89	0.85	1.72	95.02	-
Fig. 4c	93.69	22.56	0.07	1.53	98.24	-
Fig. 6c	90.30	16.89	7.33	3.85	93.38	92.86
Fig. 6d	91.48	17.37	0.66	3.39	92.37	87.50

methods [1] and [2]. In Table 2 the detailed evaluation results are shown using the winning method of each DIBCO competition [6–10] as well as results from the current state-of-the-art methods [17, 18] that used the same DIBCO datasets. From Table 2, it is shown that the latest method presented in [2] achieves the highest performance for the majority of the evaluation metrics.

5 Conclusions

Though this thesis, we have thoroughly studied the research area of document image binarization by focusing not only at the development of novel binarization techniques but also at the corresponding evaluation methods and metrics. An initial binarization method was developed that is more robust for machine-printed documents, while it has poor performance in handwritten images. The latest binarization method achieves high performance in documents with many different degradations types and it also achieves higher performance than state-of-the-art methods or methods from the DIBCO contests. Furthermore, the idea of using the baselines and the convex hulls for binarization purposes seems promising for the video processing area. Additionally, the initial evaluation methodology revealed some benefits of using a skeletonized ground-truth for evaluation purposes but it also revealed some drawbacks. The latest evaluation methodology was developed on the premise that the effect of flipped pixels on the image should be considered, and not just the fact that pixels had been flipped, which

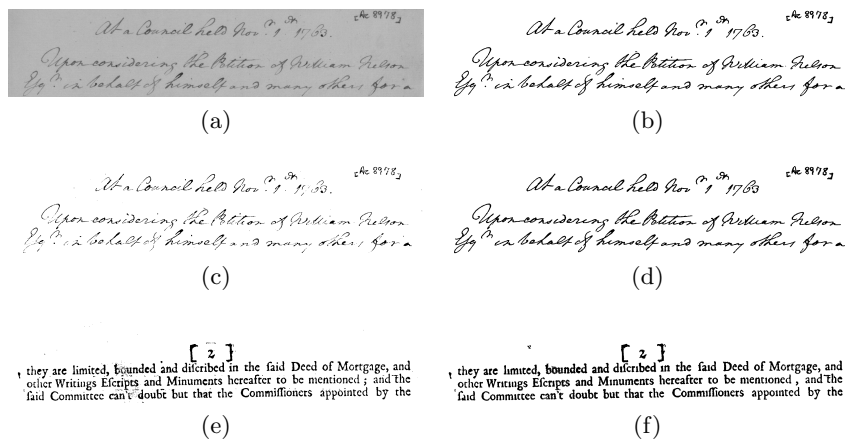


Fig. 7. (a) Original image; (b) ground-truth; (c)-(d) Ntirogiannis et. al. [1] and [2], respectively; (e)-(f) Ntirogiannis et. al. [1] and [2], respectively for Fig. 6.

leads to more reliable document-oriented evaluation. Last but not least, using the ground-truth construction procedure of the latest evaluation methodology, we made ground-truth from real degraded images and organized international document image binarization competitions. The datasets were made publicly available after each competition and have been widely used ever since.

References

1. Ntirogiannis, K., Gatos, B., Pratikakis, I.: A modified adaptive logical level binarization technique for historical document images. In: Proc. Int. Conf. on Document Analysis and Recognition. (2009) 1171–1175
2. Ntirogiannis, K., Gatos, B., Pratikakis, I.: A combined approach for the binarization of handwritten document images. Pattern Recognition Letters (2012) DOI: 10.1016/j.patrec.2012.09.026.
3. Ntirogiannis, K., Gatos, B., Pratikakis, I.: Binarization of textual content in video frames. In: Proc. Int. Conf. on Document Analysis and Recognition. (2011) 673–677
4. Ntirogiannis, K., Gatos, B., Pratikakis, I.: An objective evaluation methodology for document image binarization techniques. In: Proc. Int. Workshop on Document Analysis Systems. (2008) 217–224
5. Ntirogiannis, K., Gatos, B., Pratikakis, I.: Performance evaluation methodology for historical document image binarization. IEEE Transactions on Image Processing **22**(2) (2013) 595–609
6. Gatos, B., Ntirogiannis, K., Pratikakis, I.: ICDAR 2009 Document Image Binarization Contest - DIBCO 2009). In: Proc. Int. Conf. on Document Analysis and Recognition. (2009) 1375–1382
7. Gatos, B., Ntirogiannis, K., Pratikakis, I.: DIBCO 2009: Document Image Binarization Contest. International Journal on Document Analysis and Recognition **14**(1) (2011) 35–44

Table 2. Comparison of the proposed methods [1] and [2] to the winning method of each DIBCO contest as well as to methods [17] and [18].

Method	FM	PSNR	NRM	MPM	DRD	FM _{skel}
Winner DIBCO09	91.24	18.66	4.31	0.55	-	-
Su [17]	93.50	19.65	3.74	0.43	-	-
Ntirogiannis [1]	84.71	16.33	11.17	1.17	-	-
Ntirogiannis [2]	94.09	20.40	2.68	0.70	-	-
Winner H-DIBCO10	91.50	19.78	5.98	0.49	-	93.58
Su [17]	92.03	20.12	6.14	0.25	-	94.85
Ntirogiannis [1]	70.64	15.22	22.16	1.44	-	84.22
Ntirogiannis [2]	94.49	21.72	3.18	0.30	-	94.32
Winner DIBCO11	80.86	16.14	-	64.42	104.48	-
Su [17]	87.80	17.56	-	5.17	4.84	-
Howe [18]	91.70	19.30	-	3.87	3.48	-
Ntirogiannis [1]	80.39	15.47	-	5.78	6.68	-
Ntirogiannis [2]	92.64	19.93	-	5.12	3.13	-
Winner H-DIBCO12	89.47	21.8	-	-	3.44	90.18
Ntirogiannis [1]	76.96	16.21	-	-	7.77	87.28
Ntirogiannis [2]	95.12	22.29	-	-	1.89	94.84

8. Pratikakis, I., Gatos, B., Ntirogiannis, K.: H-DIBCO 2010 - Handwritten Document Image Binarization Competition. In: Proc. Int. Conf. on Frontiers in Handwriting Recognition. (2010) 727–732
9. Pratikakis, I., Gatos, B., Ntirogiannis, K.: ICDAR 2011 Document Image Binarization Contest (DIBCO 2011). In: Proc. Int. Conf. on Document Analysis and Recognition. (2011) 1506–1510
10. Pratikakis, I., Gatos, B., Ntirogiannis, K.: H-DIBCO 2012 - Handwritten Document Image Binarization Competition. In: Proc. Int. Conf. on Frontiers in Handwriting Recognition. (2012) 813–818
11. Otsu, N.: A thresholding selection method from gray-level histogram. IEEE Transactions on Systems, Man and Cybernetics **9**(1) (1979) 62–66
12. Niblack, W. In: An Introduction to Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall (1986) 115–116
13. Sauvola, J., Pietikainen, M.: Adaptive document image binarization. Pattern Recognition **33**(2) (2000) 225–236
14. Kim, I.K., Jung, D.W., Park, R.H.: Document image binarization based on topographic analysis using a water flow model. Pattern Recognition **35**(1) (2002) 265–277
15. Gatos, B., Pratikakis, I., Perantonis, S.J.: Adaptive degraded document image binarization. Pattern Recognition **39**(3) (2006) 317–327
16. Lu, S., Su, B., Tan, C.L.: Document image binarization using background estimation and stroke edges. International Journal on Document Analysis and Recognition **13**(4) (2010) 303–314
17. Su, B., Lu, S., Tan, C.L.: A robust document image binarization technique for degraded document images. IEEE Transactions in Image Processing **22**(4) (2013) 1408–1417

18. Howe, N.R.: Document binarization with automatic parameter tuning. *International Journal on Document Analysis and Recognition* (2012) DOI: 10.1007/s10032-012-0192-x.
19. Gatos, B., Pratikakis, I., Perantonis, S.J.: Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information. In: *Proc. Int. Conf. on Pattern Recognition*. (2008) 1–4
20. Su, B., Lu, S., Tan, C.L.: Combination of document image binarization techniques. In: *Proc. Int. Conf. on Document Analysis and Recognition*. (2011) 22–26
21. Kwak, S., Chung, K., Choi, Y.: Video caption image enhancement for an efficient character recognition. In: *Proc. Int. Conf. on Pattern Recognition*. (2000) 606–609
22. Wolf, C., Jolion, J.M., Chassaing, F.: Text localization, enhancement and binarization in multimedia documents. In: *Proc. Int. Conf. on Pattern Recognition*. (2002) 1037–1040
23. Merler, M., Kender, J.R.: Semantic keyword extraction via adaptive text binarization of unstructured unsourced video. In: *Proc. Inter. Conf. on Image Processing*. (2009) 261–264
24. Kamel, M., Zhao, A.: Extraction of binary character-graphics images from grayscale document images. *CVGIP: Computer Vision Graphics and Image Processing* **55**(3) (1993) 203–217
25. Li, J., Tian, Y., Huang, T., Gao, W.: Multi-polarity text segmentation using graph theory. In: *Proc. Int. Conf. on Image Processing*. (2008) 3008–3011
26. Saidane, Z., Garcia, C.: Robust binarization for video text recognition. In: *Proc. Int. Conf. on Document Analysis and Recognition*. (2007) 874–879
27. Zhou, Z., Li, L., Tan, C.L.: Edge based binarization for video text images. In: *Proc. Int. Conf. on Pattern Recognition*. (2010) 133–136
28. Canny, J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(6) (1986) 679–698
29. Yang, Y., Yan, H.: An adaptive logical method for binarization of degraded document images. *Pattern Recognition* **33**(5) (2000) 787–807
30. Lee, H.J., Chen, B.: Recognition of handwritten chinese characters via short line segments. *Pattern Recognition* **25**(5) (1992) 543–552
31. Trier, D., Taxt, T.: Evaluation of binarization methods for document images. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(3) (1995) 312–315
32. Kavallieratou, E., Stathis, S.: Adaptive binarization of historical document images. In: *Int. Conf. on Pattern Recognition*. Volume 3. (2006) 742–745
33. Stathis, P., Kavallieratou, E., Papamarkos, N.: An evaluation survey of binarization algorithms on historical document images. In: *Proc. Int. Conf. on Pattern Recognition*. (2008) 1–4
34. Paredes, R., Kavallieratou, E., Lins, R.D.: ICFHR 2010 Contest: Quantitative evaluation of binarization algorithms. In: *Proc. Int. Conf. on Frontiers in Handwriting Recognition*. (2010) 733–736
35. Barney Smith, E.H.: An analysis of binarization ground truthing. In: *Proc. Int. Workshop on Document Analysis Systems*. (2010) 27–33
36. Badekas, E., Papamarkos, N.: Automatic evaluation of document binarization results. In: *Proc. Iberoamerican Congress on Pattern Recognition*. (2005) 1005–1014
37. Lu, H., Kot, A.C., Shi, Y.Q.: Distance-reciprocal distortion measure for binary document images. *IEEE Signal Process. Lett.* **11**(2) (2004) 228–231