Scale Space Binarization Using Edge Information Weighted by a Foreground Estimation

Florian Kleber, Markus Diem and Robert Sablatnig
Computer Vision Lab
Vienna University of Technology
1040 Vienna, Austria
Email: kleber@caa.tuwien.ac.at

***Abstract*—The proposed binarization algorithm uses a scale space to avoid the estimation of script size dependent parameters. Due to the continous smoothing from finer to coarse scales, noise such as background clutter is suppressed since coarse scales characterize homogeneous regions of the image. Thus, coarser scales of the scale space can be used as a foreground estimation to apply a weigthing scheme robust against noise present in, for instance carbon copies or ancient and degraded documents. Additionally the information of filled regions is propagated through the scales. The use of integral images for the calculation of the mean, standard deviation and morphological operations allow for an efficient implementation of the method presented. The binarization of each scale is based on changes of the local intensity as proposed by Su et al.**

***Keywords*-Document Binarization; degraded documents; scale space; foreground estimation;**

I. Introduction

For document binarization global or adaptive (local) thresholding techniques can be applied. Global binarization methods such as proposed by Otsu introduce errors if documents are degraded or uneven lighting conditions are present. Ancient manuscripts may be degraded or faded-out due to environmental effects such as mold or humidity (e.g. Missale Sinaiticum, a manuscript from the 11th century that has been exposed to water [1]), which leads to a high variation in the contrast of the image. In addition background clutter can produce errors if global methods are applied. Beside ancient documents, printed carbon copies can also contain noise (e.g. historic valuable records of the Stasi [2]). Based on this fact the binarization of historical documents is still a challenging research topic [3]. An evaluation of binarization algorithms is shown in the Document Image Binarization Contest (DIBCO) 2009 [4] and DIBCO 2010 [5]. The progress can be seen in the forthcoming DIBCO 2011. State-of-the-art algorithms (see Section II), such as Su et al. [3], [6] and Gatos [7], are local methods, which e.g. estimate the background (Su et al.) to define regions where a threshold depending on grayvalues is applied. An overview of state-of-the-art binarization algorithms is described in Section II.

An alternative to the methods mentioned is to use multispectral imaging and to exploit information in the non-visible wavelengths of the reflected and emitted light of historical documents. Lettner [8] shows the possibility to use the information within different wavelengths to enhance the binarization result. However, multispectral imaging techniques require customized equipment [9], [10]. As a result state-of-the-art binarization methods use traditional RGB/grayvalue images of manuscripts. Therefore the methods can be applied to images of digitized manuscripts and to images which have been captured with “low-cost” imaging devices.

The proposed binarization algorithm is an adaption of the algorithm published by Su et al. The drawback of the sensitivity of the contrast image to noise with a high frequency is reduced by weighting the image with a foreground estimation. Additionally, to avoid that pixels of a homogeneous region larger than the filtersize are not segmented a scale space is introduced, which makes the algorithm independent to the estimated stroke width and filtersize respectively.

This paper is organized as follows: Section II reviews the state-of-the-art of binarization algorithms. In Section III the proposed scale space binarization method is presented, while Section IV presents the results of the algorithm on the DIBCO dataset. Finally a conclusion is given in Section V.

II. Related Work

The objective of image segmentation is to group image pixels according to pre-defined rules. On document images this problem consists of two classes: foreground (text) and background. For the binarization of documents global and adaptive binarization methods exist. While a single threshold is applied on every pixel by global algorithms, adaptive methods define local regions in which individual threshold values are calculated. Global thresholds are suitable for images with a bimodal gray value distribution. Otsu’s thresholding method assumes a bimodal histogram and minimizes the intra-class variance, while maximizing the inter-class variance. Historical and degraded documents need adaptive algorithms due to the low contrast of faded-out text and the presence of background clutter/noise.

Gatos et al. [7] propose an adaptive binarization suitable for degraded documents. They use Sauvolas approach [11]

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.238

for a rough estimation of the foreground to estimate the background surface. This is done by interpolating the pixel values of the detected foreground regions by the surrounding background. The resulting background surface is combined with the image for the final thresholding.

Bolan Su et al. [3] present an algorithm for degraded historical documents which uses a contrast image to define threshold region candidates. The contrast image is defined by the normalized gradient image which is calculated using the local maximum and minimum. Su et al. states that the normalization "compensates for the effect of the image contrast/brightness variation" [3]. The contrast image is thresholded using Otsu's method to define candidates for the foreground region. The window size for the adaptive thresholding is estimated by the stroke width and the threshold is defined by $m + s/2$ where m is the mean value of the foreground candidates and s is the standard deviation within the local window. In addition, the amount of the estimated foreground pixels has to be higher than the estimated stroke width. An adaption of the algorithm uses a polynomial smoothing procedure to estimate the document background. The varying contrast is compensated using the estimated background. Edges in the compensated document image are maxima in the vertical and horizontal L1-norm image. For a detailed description see Shijian Lu et al. [6].

Wolf et al. [12] use a binarization algorithm suitable for multimedia documents and video frames. On the DIBCO 2009 dataset this method achieved rank 5. It is an adaption of Sauvola [11] by normalizing the contrast and the mean gray level of the image. Tabbone and Wendling [13] have published a general binarization algorithm using a multi-scale approach, at which the image is continously smoothed to test the homogeneity of regions and to decide if a region belongs to the background. A different morphological multiscale approach applied to document images is proposed by Dorini and Leite [14]. They introduce a morphological operator with scale-space properties, which identifies or delimit regions (Binary Self-dual MultiScale Morphological Toggle, see [14]).

Figure 1 shows an image of the DIBCO 2009 dataset with a variable background, the grayvalue distribution and the binarized image using Otsu and a manually determined threshold. The results show that degraded documents with a variable background can not be binarized using a global method, since the assumption of a bimodal grayvalue distribution is wrong. The manually determined global threshold shows, that parts of the characters are not segmented. Thus adaptive binarization methods are needed.

III. Proposed Method

The results of the binarization contest in 2009 and 2010 show that algorithms using (text) stroke edge regions as candidate pixels performed best for ancient manuscripts.

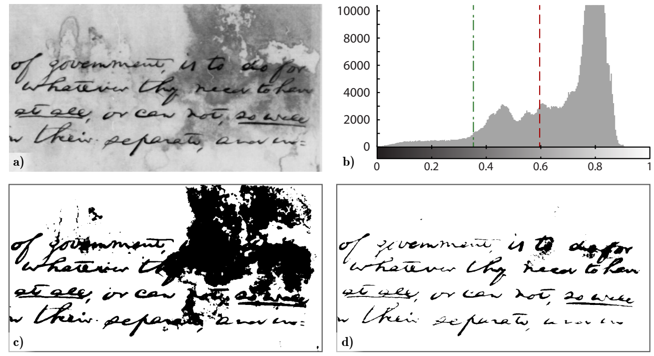

Figure 1. (a) Image of the DIBCO 2009 dataset (b) histogram with Otsu threshold (dashed) and manual threshold (dashed-dotted) (c) Otsu threshold image (d) manually thresholded image

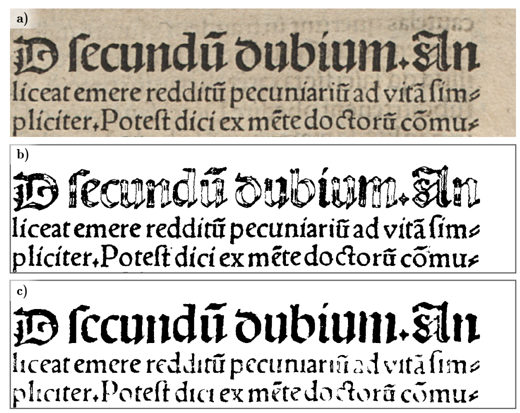

Figure 2. (a) Image of the DIBCO 2009 dataset (b) Su et al.'s approach (c) proposed approach

To avoid distortions arising from the variation of the background e.g. Gatos et al. [7] and Lu et al. [6] estimate the background for a contrast compensation. Both Algorithms estimate parameters such as the local window size based on e.g. the mean character height or the stroke width. Problems may arise if different text sizes and stroke widths occur on the same folio/page. Figure 2 shows a result of the algorithm of Su et al. if a wrong stroke width is applied and the result of the proposed scale space method. A scale space allows fixed parameters due to the different scales of the image. Propagating information from coarse to fine scales makes the algorithm independent to the text size. The continously gaussian smoothing of the image suppresses noise with a high variation. As a result, the coarse scales are used as a foreground estimation. The use of integral images allow an efficient implementation of the algorithm [15].The binarization of the image at each scale is done using the local maximum and minimum as published by Su et al. [3]. A Pseudo-Code of the method proposed is presented, see Algorithm 1.

The scale space theory has been studied by Lindeberg

[16]. A scale space $L(x, y, t)$ of an image $f(x, y)$ is gained by concolving $f(x, y)$ with Gaussians $G(x, y, t)$:

$$L(x, y, t) = G(x, y, t) * f(x, y) \tag{1}$$

where $*$ denotes the convolution and $t = \sigma^2$ the scale parameter. The Gaussian filter is defined by:

$$G(x, y, t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t} \tag{2}$$

Lindeberg [16] has shown that the Gaussian filter kernel is the only low-pass filter, which satisfies the following conditions:

- linear and shift/rotation invariance
- semigroup property
- continous signals ("no new local extrema or zero crossings are introduced with increasing scale parameter" [16])
- non-enhancement of local extrema (causality)

Thus, structures of a coarse scale represent simplified structures of finer scale levels, which allows to use a coarse scale (dependend on scale, see Section III-B) as a foreground estimation that supresses noise with a high frequency such as background clutter. Due to the fact that no new structures are introduced, no errors can be propagated from coarse to finer scales which allows to use the scale space for a binarization approach.

Algorithm 1 PseudoCode of the Scale-Space Binarization.

1: $L = scaleSpace(img);$
2: $i = size(L);$
3: $parent = binarizeSu(L(i-1));$
4: **for** $k = size(L) - 2 \rightarrow 1$ **do**
5: $\quad child = binarizeSu(L(k));$
6: $\quad$ **for all** (x, y) such that $parent(x, y) = 255$ **and** $child(x, y) = 0$ **do**
7: $\qquad thr = mean((parent$ **and** $BCChild) \cdot L(k), R);$
8: $\qquad child = L(k) < thr;$
9: $\quad$ **end for**
10: $\quad$ **for all** (x, y) such that $parent(x, y) = 0$ **and** $child(x, y) = 255$ **do**
11: $\qquad weightImg = L(k-2);$
12: $\qquad normalize(weightImg);$
13: $\qquad invert(weightImg);$
14: $\qquad thr = thr \cdot weightImg;$
15: $\qquad child = L(k) < thr(x, y);$
16: $\quad$ **end for**
17: $\quad parent = child$
18: **end for**

A. Scale Space Binarization

The scale space of a document image $f(x, y)$ is constructed as proposed by Lindeberg [16]. The information of the scale space is propagated from coarse to fine scales as follows:

Figure 3. (a) current scale (b) parent image (c) current segmentation (d) scale space binarization

Let $L(x, y, t_{i+1})$ (parent image) denote a binarized image of the scale space with scale t_{i+1} and $L(x, y, t_i)$ (child image) denotes the binarized image of the next finer scale. The scale parameter is $t = \sigma^2$, and it is increased by σ between two successive scales. Both images are binarized using Su et al. 's approach with a constant stroke width and a constant size of the neighbourhood window (5 px). Each pixel (x, y) in $L(x, y, t_i)$ is compared with the associated pixel of the parent image $L(x, y, t_{i+1})$. If the pixel (x, y) is segmented in the parent image but not in the current image a new threshold $thr(x, y)$ is applied to the pixel (x, y) in $L(x, y, t_i)$. The new threshold is computed as follows: In contrast to the threshold candidates of Su et al. (local areas that contain a defined number of high contrast pixel proportional to the stroke width) the threshold candidates are redefined by segmented areas $R_1 \dots R_n$ of the binarized parent image $L(x, y, t_{i+1})$. The threshold thr of each area $R_i, i \in \{1 \dots n\}$ is defined by

$$thr_{R_i} = E_{mean} \tag{3}$$

where E_{mean} is the mean of all segmented binary high contrast pixel (BinaryContrastChild, BCChild, see Algorithm 1, line 7) in $L(x, y, t_i)$ with $(x, y) \in R_i$. Pixels that are segmented in $L(x, y, t_{i+1})$ and not in the finer scale belong to homogeneous regions, which are not segmented in the current scale if a wrong stroke width is estimated. Due to properties of the scale space (see Section III) noise can not be introduced in coarse scales. Hence, noise is not propagated through the scale space. Even areas with a dark background are not segmented due to the threshold which is calculated by the pixels defined in the binary high contrast image of $L(x, y, t)$. Figure 3 shows an image and the binarized images of two subsequent scales with the effect described.

Figure 4. weight image (foreground estimation)

B. Foreground Estimation

Pixels that are segmented in the finer scale and not in the coarse scale belong to finer structures, e.g. noise. In order to compute a foreground estimation that is robust with respect to noise the image at scale $L(x, y, t_{i-2})$ is used (see Algorithm 1, line 11-14). If a pixel is segmented in the current scale and not in the parent image the current threshold as defined by Su et al. is weighted with the foreground image. Figure 4 shows the foreground estimation of the image at the last scale. It can be seen that background noise is supressed (background is a homogeneous area).

IV. Results

The proposed algorithm has been tested on 3 sets of images: the dataset used at DIBCO 2009 Contest which consists of 5 machine printed and 5 handwritten images, the dataset of DIBCO 2010 and finally 3 synthetic images with noise and varying text sizes. The GT of the synthetic images is defined manually. For the evaluation of the binarization 3 measures, namely the F-Measure, the Peak Signal to Noise Ratio (PSNR) and the Negative Rate Metric (NRM) are used. These measures have also been applied at the Document Image Binarization Contest (DIBCO 2009) within the ICDAR Conference [4]. Another possibility for the evaluation of binarization methods is to apply an OCR and to evaluate its errors. A common state of the art OCR system is the FineReader [17]. An example of an evaluation using an OCR system can be seen in [7]. Figure 5 shows an image of a carbon copy, a synthetic image with background noise, the results of the binarization of the proposed approach and the result of Su et al. It can be seen that the background noise is suppressed due to weighting with the foreground estimation. Table I shows the results on the DIBCO dataset.

Method	F-Measure	PSNR	NRM ($\times 10^{-2}$)
proposed approach DIBCO 2009	86.5624	16.8254	10.0634
proposed approach DIBCO 2010	78.6865	16.4226	17.0166

Table I
BINARIZATION RESULTS OF THE PROPOSED APPROACH - DIBCO

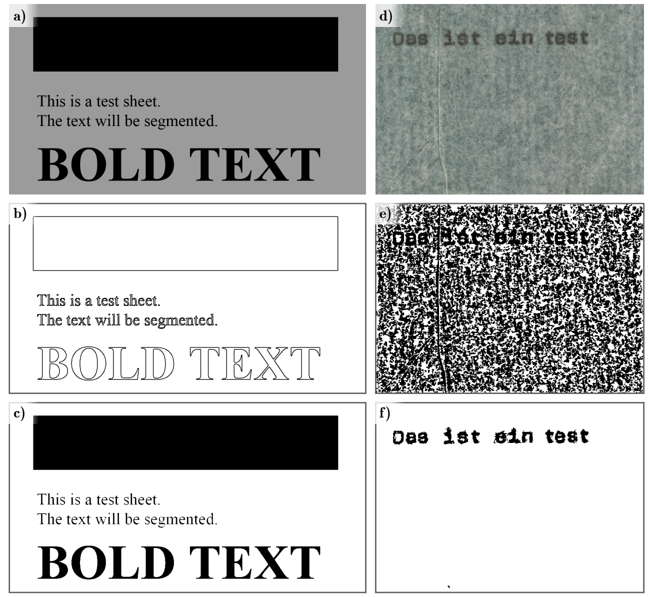

Figure 5. (a) synthetic test image (b) Su et al. (c) proposed method (d) carbon copy (e) Su et al. (f) proposed method

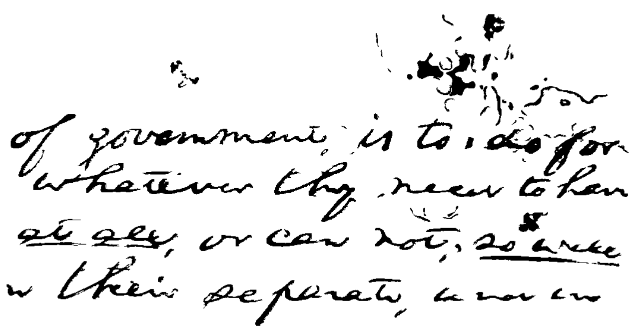

Figure 6. result of the proposed approach of an image of the DIBCO 2009 dataset

The final result is the average of the measures of each single image. Although Su et al. perform best, the proposed method performs best on the dibco dataset combined with the dataset containing a carbon copy with background noise and synthetic images with background noise (see Table II). Figure 5 shows the results of the images, whereas Figure 6 shows the errors of the proposed approach. The wrong segmentation arise from the edge of the background variation, which is detected in the binary high contrast pixel image. Table III shows the result of the proposed approach on synthetic images. It can be seen that the approach can handle noise with a high frequency. The performance on the single DIBCO dataset arise from the foreground estimation. The weigthing with the foreground let thin strokes disappear in the last scale.

Method	F-Measure	PSNR	NRM ($\times 10^{-2}$)
Otsu	82.8263	17.5926	6.0132
Sauvola and Pietikainen	59.6585	8.4239	12.2673
Su et al.	83.1956	16.4907	11.324
proposed approach	88.3845	19.3848	8.465

Table II
BINARIZATION RESULTS OF DIBCO 2009 AND SYNTHETIC IMAGE DATABASE

Method	F-Measure	PSNR	NRM ($\times 10^{-2}$)
Otsu	83.0921	23.2997	6.8747
Sauvola and Pietikainen	72.2764	10.8566	15.3871
Su et al.	58.3462	9.1918	29.4971
proposed approach	92.6757	25.7834	4.4691

Table III
BINARIZATION RESULTS OF SYNTHETIC IMAGE DATABASE

V. CONCLUSION

A scale space binarization algorithm with a foreground estimation has been proposed. Although the images at each scale are binarized with the winner of the DIBCO 2009 Su et al. each binarization can be used for the scale space. However, tests have shown, that Su et al. performs best as the "base" algorithm. The main advantage is the indepence to scale dependent parameters. As a result the information is propagated through the scale. Additionally a foreground is estimated to suppress background noise. As future work the foreground estimation will be improved to avoid the segmentation of low frequency noise as shown in the Section IV.

ACKNOWLEDGMENT

The authors would like to thank the Fraunhofer-Institute for Production Systems and Design Technology (IPK), Berlin for supporting the work.

REFERENCES

[1] F. Kleber, M. Lettner, M. Diem, M. Vill, R. Sablatnig, H. Miklas, and M. Gau, "Multispectral Acquisition and Analysis of Ancient Documents," in *Proc. of the 14th International Conference on Virtual Systems and MultiMedia (VSMM 2008), Dedicated to Cultural Heritage - Project Papers*, M. Ioannides, A. Addison, A. Georgopoulos, and L. Kalisperis, Eds. Limassol, Cyprus: Archaeolingua, 2008, pp. 184–191.

[2] J. Schneider and B. Nickolay, "The Stasi puzzle," *Fraunhofer Magazine Special Issue*, pp. 32–33, 2008.

[3] B. Su, S. Lu, and C. L. Tan, "Binarization of historical document images using the local maximum and minimum," in *DAS '10: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. New York, NY, USA: ACM, 2010, pp. 159–166.

[4] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "ICDAR 2009 Document Image Binarization Contest (DIBCO 2009)," in *10th International Conference onDocument Analysis and Recognition, 2009. ICDAR '09.*, jul. 2009, pp. 1375 –1382.

[5] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "H-DIBCO 2010 - Handwritten Document Image Binarization Competition," *Frontiers in Handwriting Recognition, International Conference on*, pp. 727–732, 2010.

[6] S. Lu, B. Su, and C. L. Tan, "Document image binarization using background estimation and stroke edges," *Int. J. Doc. Anal. Recognit.*, vol. 13, pp. 303–314, December 2010. [Online]. Available: http://dx.doi.org/10.1007/s10032-010-0130-8

[7] B. Gatos, I. Pratikakis, and I. Perantonis, "Adaptive degraded document image binarization." *Pattern Recognition*, vol. 39, no. 3, pp. 317–327, 2006.

[8] M. Lettner and R. Sablatnig, "Higher Order MRF for Foreground-Background Separation in Multi-Spectral Images of Historical Manuscripts," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems DAS 2010*. Boston, USA: ACM, 2010, pp. 317–324.

[9] R. Easton, K. Knox, and W. Christens-Barry, "Multispectral Imaging of the Archimedes Palimpsest," in *32nd Applied Image Pattern Recognition Workshop, AIPR 2003*. Washington, DC: IEEE Computer Society, October 2003, pp. 111–118.

[10] M. Hain, J. Bartl, and V. Jacko, "Multispectral Analysis of Cultural Heritage Artefacts," in *Measurement Science Review*, vol. 3, 2003, pp. 9–12.

[11] J. Sauvola and M. Pietikainen, "Page segmentation and classification using fast feature extraction and connectivity analysis," in *Proceedings of the Third International Conference on Document Analysis and Recognition, 1995.*, vol. 2, aug. 1995, pp. 1127 –1131.

[12] C. Wolf, J.-M. Jolion, and F. Chassaing, "Text Localization, Enhancement and Binarization in Multimedia Documents," *Pattern Recognition, International Conference on*, vol. 2, pp. 1037–1040, 2002.

[13] S. Tabbone and L. Wendling, "Multi-scale binarization of images," *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 403 – 411, 2003.

[14] L. B. Dorini and N. J. Leite, "A Multiscale Morphological Binarization Algorithm," *Computer Vision, Imaging and Computer Graphics. Theory and Applications*, vol. 68, pp. 283–295, 2010.

[15] M. E. Hussein, F. Porikli, and L. S. Davis, "Kernel integral images: A framework for fast non-uniform filtering," in *International Conference on Computer Vision and Pattern Recognition, CVPR*, 2008, pp. 1–8.

[16] T. Lindeberg, "Scale-Space Theory: A Basic Tool for Analysing Structures at Different Scales," *Journal of Applied Statistics*, vol. 21(2), pp. 224–270, 1994.

[17] "ABBYY, www.finereader.com, accessed 09, 2010."