Ridges based Curled Textline Region Detection from Grayscale Camera-Captured Document Images

Syed Saqib Bukhari[1], Faisal Shafait[2], and Thomas M. Breuel[1,2]

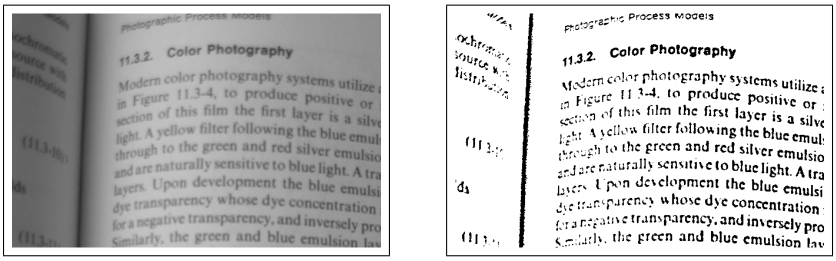
[1] Technical University of Kaiserslautern, Germany
[2] German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
bukhari@informatik.uni-kl.de
faisal@iupr.dfki.de
tmb@informatik.uni-kl.de
http://www.iupr.com

Abstract. As compared to scanners, cameras offer fast, flexible and non-contact document imaging, but with distortions like uneven shading and warped shape. Therefore, camera-captured document images need preprocessing steps like binarization and textline detection for dewarping so that traditional document image processing steps can be applied on them. Previous approaches of binarization and curled textline detection are sensitive to distortions and loose some crucial image information during each step, which badly affects dewarping and further processing. Here we introduce a novel algorithm for curled textline region detection directly from a grayscale camera-captured document image, in which matched filter bank approach is used for enhancing textline structure and then ridges detection is applied for finding central line of curled textlines. The resulting ridges can be potentially used for binarization, dewarping or designing new techniques for camera-captured document image processing. Our approach is robust against bad shading and high degrees of curl. We have achieved around 91% detection accuracy on the dataset of CBDAR 2007 document image dewarping contest.

Key words: Curled Textline Finding, Camera-Captured Document Images, Grayscale Document Images, Ridges, Anisotropic Gaussian Smoothing

1 Introduction

Since decades, scanners are being used for capturing document images. Nowadays, cameras are considered as a potential substitute of scanners due to their high production and low cost. Together with high production and low cost, cameras also offer long-ranged, non-contact and fast capturing as compared to scanners. These features speed up the traditional document image capturing and open doors for new versatile applications, like mobile OCR, digitizing thick books, digitizing fragile historical documents, etc.



(a) (b)

Fig. 1. a) Grayscale camera-captured document image portion from CDBAR 2007 dewarping contest dataset. b) Poor binarization (adaptive thresholding) result; which has lost most of its details.

Document imaging with hand-held camera is typically done in an uncontrolled environment, which induces several types of distortions in the captured image. Some of major distortions are: motion blur, low resolution, uneven light shading, under-exposure, over-exposure, perspective distortion and non-planar shape. Because of these distortions, traditional document image processing algorithms can not be applied directly to camera-captured document images.

Tremendous research is being devoted to make camera-captured document images suitable for traditional algorithms, by using dewarping techniques. Binarization and curled textline detection are the main steps of dewarping. Researchers use well known thresholding techniques [1] for binarization and have developed many new techniques for curled textline detection. Previous techniques of curled textline detection can be classified into two main categories: a) heuristic search [2–7] and b) active contours (snakes) [8, 9].

Heuristic search based approaches start from a single connected component and search other components of a textline in a growing neighborhood region. These approaches use complex and rule-based criteria for textline searching. Active contours (snakes) have been used in [8, 9] for curled textline segmentation. These approaches start with initializing open-curved snakes over connected components and result in textlines detection by deforming snakes in vertical directions only. In general, both heuristic search and active contours (snakes) approaches rely on adaptive thresholding for the binarization of camera-captured document image before textline detection. Binarization may give poor results, especially under the problems of uneven shading, low resolution, motion blur and under- or over-exposure, as shown in Figure 1. Poor binarization (Figure 1(b)) can negatively affect the textline detection results and later on text recognition results.

In this paper we introduce a novel approach for curled textlines regions detection directly from grayscale camera-captured document images. Therefore, our method does not use binarization before textline detection. The method starts by enhancing the curled textline structure of a grayscale document image using multi-oriented multi-scale anisotropic Gaussian smoothing based on matched filter bank approach as in [10]. Then ridges detection [11, 12] technique is applied on the smoothed image. Resulting ridges represent the complete unbroken central line regions of curled textlines.

We make the following contributions in this paper. The method presented here works directly on grayscale camera-captured document image and is therefore independent of binarization and its errors under the problems of uneven shading, low resolution, motion blur and under- or over-exposure. Together with the independence of binarization, our method is robust against the problems of low resolution, motion blur and under- or over-exposure and detects textlines accurately under these problems. Furthermore, unlike previous approaches of curled textline detection, our method is also robust against high degrees of curl, variable directions of curl, different line spacing and font sizes problems.

2 Curled Textline Detection

As a first step, we apply matched filter bank approach from [10] for curl textlines enhancement of grayscale camera-captured document images. The main idea is to use oriented anisotropic Gaussian smoothing filter (Equation 2) to generate the set of Gaussian smoothing windows from the ranges of horizontal and vertical standard deviations (scales) and orientation.

$$g(x, y; \sigma_x, \sigma_y, \theta) = \frac{1}{2\pi\sigma_x\sigma_y} exp\{-\frac{1}{2}(\frac{(xcos\theta + ysin\theta)^2}{{\sigma_x}^2} + \frac{(-xsin\theta + ycos\theta)^2}{{\sigma_y}^2})\}$$

The reason of considering ranges for scales and orientation is because of variable font sizes and high degrees of multi-oriented curls within an image, respectively. To achieve this, we define an automatic way of selecting the ranges for σ_x, σ_y and θ. The same range is selected for both σ_x and σ_y, which is a function of the height of the document image (H), that is aH to bH with $a < b$. The suitable range for θ is from -45 to 45 degrees. The set of filters is defined by selecting all possible combinations of σ_x, σ_y and θ, from their ranges. It covers all typical font sizes and curl conditions in an image. The set of filters is applied on each pixel of grayscale document image and then the maximum resulting value among all the resulting values is selected for the final smoothed image. Multi-oriented multi-scale anisotropic Gaussian smoothing enhances the curled textlines structure well, which is clearly visible in Figure 2(b).

The next step is to detect textlines from the smoothed image. Some researchers [13, 14] use adaptive thresholding of the smoothed image to detect the textlines or the portions of textlines and then perform heuristic postprocessing steps to join neighboring portions. Our method breaks this tradition and uses ridges detection for finding textlines form smoothed image. Firstly, adaptive thresholding of smoothed grayscale image gives poor results and misses textlines or portions of textlines, as shown in Figure 2(c). Secondly, we do not want to use any type of heuristic postprocessing steps.

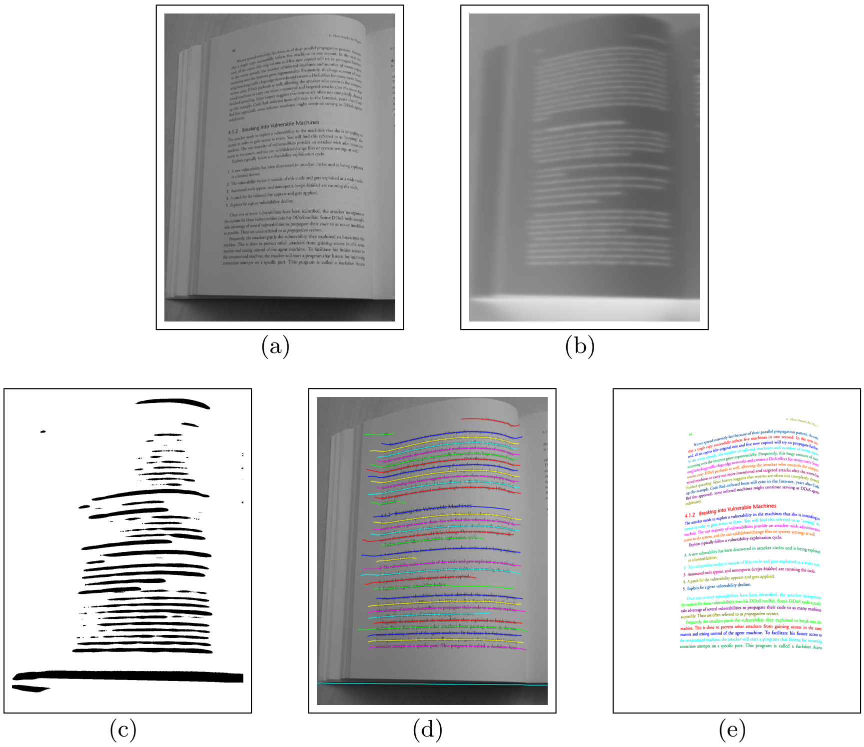


Fig. 2. a) Grayscale camera-captured document image from CDBAR 2007 dewarping contest dataset. b) Smoothed image : Enhanced curled textlines structure. c) Poor binarization (adaptive thresholding) result of smoothed image. d) Detected ridges from smooth image which are mapped over document image with different colors; represent curled textlines regions. e) Curled textlines detection result : Detected ridges are mapped over the binarized image (given in the dataset) and assign the corresponding ridges label to connected components.

Ridges are primary type of features, which provides important information about the existence of objects together with the symmetrical axes of these objects. Since decades, ridges detection has been used popularly for producing rich description of significant features from smoothed grayscale images [11] and speech-energy representation in time-frequency domain [12]. In this paper we are interested in detecting the central line regions of curl textlines. Ridges detection can do this efficiently by estimating the symmetrical axes from smoothed textlines structure. We have seen in the previous section that multi-oriented multi-scale anisotropic Gaussian smoothing generates well enhanced textlines structure. Therefore, ridges detection over smoothed image can produce unbroken central lines structure of textlines. Here, Horn-Riley [11, 12] based ridges detection approach is used. This approach is based on the informations of local direction of gradient and second derivatives as the measure of curvature. From this information, which is calculated by Hessian matrix, ridges are detected by finding the zero-crossing of the appropriate directional derivatives of smoothed image. Detected Ridges over the smoothed image of Figure 2(b) are shown in Figure 2(d). It is clearly visible in the Figure 2(d) that each ridge covers the complete central line region of a textline, which results in textlines detection.

3 Experiments and Performance Evaluation

To demonstrate the performance of presented approach, we evaluate it on the real-world hand-held camera-captured document images dataset used in CBDAR 2007 for document image dewarping contest [15]. This dataset consists of 102 grayscale and their binarized images, with ground truth of binarized images in color coded format. Previously, researchers have used only the binarized images and corresponding ground truth from this dataset, for the development and evaluation of their algorithms. We use grayscale images from this dataset. But, there is no direct evaluation method for grayscale textline detection results. Therefore, the ridges detected from grayscale image are mapped over its cleaned-up binarized image and the corresponding ridges label are assigned to connected components, as shown in Figure 2(e). Now we can compare this textlines detection result with ground truth. Here we use two different standard textline detection evaluation methods [16, 17]. Evaluation method presented in [17] is designed for handwritten textline segmentation without any background noise, therefore we are using 91 manually cleaned-up document images from the CBDAR 2007 dataset[3].

Descriptions of performance evaluation metrics for textline segmentation based on [18, 16] are as follows. Consider we have two segmented images, the ground truth G and hypothesized snake-segmentation H. We can compute a weighted bipartite graph called "pixel-correspondence graph" [19] between G and H for evaluating the quality of the segmentation algorithm. Each node in G or H represents a segmented component. An edge is constructed between two nodes such that the weight of the edge equals the number of foreground pixels in the intersection of the regions covered by the two segments represented by the nodes. The matching between G and H is perfect if there is only one edge incident on each component of G or H, otherwise it is not perfect, i.e. each node in G or H may have multiple edges. The edge incident on a node is significant if the value of (w_i/P) meets some thresholding criteria, where w_i is the edge-weight and P is the number of pixels corresponding to a node (segment).

On the basis of the above description the performance evaluation metrics are:

- **Correct segmentation (N_c):** the total number of G and H components' pairs which have only one significant edge in between.
- **Total oversegmentations (N_{tos}):** the total number of significant edges that ground truth lines have, minus the number of ground truth lines.

[3] The cleaned-up data set and ground truth are available by contacting the authors

Table 1. Performance evaluation results based on [16]

Number of ground truth lines (N_g)	2713
Number of segmented lines (N_s)	2704
Number of correct segmentation (N_c)	2455
Total oversegmentations (N_{tos})	105
Total undersegmentations (N_{tus})	99
Oversegmented components (N_{oc})	99
Undersegmented components (N_{us})	93
Missed components (N_{mc})	6
Correct segmentation accuracy ($100 * N_c/N_g$)	90.50%

- **Total undersegmentations (N_{tus}):** the total number of significant edges that segmented lines have, minus the number of segmented lines.
- **Oversegmented components (N_{oc}):** the number of ground truth lines having more than one significant edge.
- **Undersegmented components (N_{uc}):** the number of segmented lines having more than one significant edge.
- **Missed components (N_{mc}):** the number of ground truth components that matched the background in the hypothesized segmentation.

Performance evaluation results of our textline segmentation algorithm, based on the above metrics, are given in a Table 1.

According to the methodology [17], the matching score is equal to or above a specified acceptance threshold (i.e. 95%), where matching score is defined as:

$$MatchScore(i, j) = T(G_j \cap R_i \cap I)/T(G_j \cup R_i \cup I) \tag{1}$$

where I is the set of all image pixels, G_j and R_i are the sets of all pixels covering the j^{th} ground truth region and i^{th} result region respectively. Based on the matching scores, detection rate (DR), recognition accuracy (RA) and combine performance metric FM are calculated as follows:

$$DR = w_1 \frac{o2o}{N} + w_2 \frac{g\_o2m}{N} + w_3 \frac{g\_m2o}{N} \tag{2}$$

$$RA = w_4 \frac{o2o}{M} + w_5 \frac{d\_o2m}{M} + w_6 \frac{d\_m2o}{M} \tag{3}$$

where N and M are the total number of ground truth and result elements respectively, $w_1(1), w_2(0.25), w_3(0.25), w_4(1), w_5(0.25), w_6(0.25)$ are pre-determined weights, $o2o$ is one to one match, g_o2m is one ground truth to many detected, g_m2o is many ground truth to one detected, d_o2m one detected to many ground truth and d_m2o is many detected to one ground truth. Performance evaluation results based on equations (1), (2) and (3) are given in Table 2.

Table 2. Performance evaluation results based on [17]

Ground truth elements (N)	2749
Detected elements (M)	2717
One to one match ($o2o$)	2503
One ground truth to many detected (g_o2m)	64
Many ground truth to one detected (g_m2o)	127
One detected to many ground truth (d_{o2m})	62
Many detected to one ground truth (d_{m2o})	130
Detection rate (DR)	92.79%
Recognition accuracy (RA)	93.89%
$FM = (2 * DR * RA)/(DR + RA)$	93.34%
Correct segmentation accuracy ($100 * o2o/N$)	91.05%

4 Discussion

We have proposed a novel approach for curled textlines regions detection from grayscale camera-captured document images. We introduced a combination of multi-scale multi-oriented Gaussian smoothing and ridges detection for finding curled textlines directly from grayscale intensity values of camera-captured document images. Therefore, our approach does not use binarization and is independent of binarization errors. We have achieved around 91% of one-to-one correct segmentation textline detection accuracy on the dataset of CDBAR 2007 document image dewarping contest, which proves the effectiveness of our method. The 9% of errors are mainly because of oversegmentation, that is more than one textline detected for a single ground truth textline. These oversegmentation errors can be easily overcome by grouping ridges in horizontal neighborhood region, as a post-processing step. Our approach is robust against uneven shading, low resolution, motion blur, under- or over-exposure, high degrees of curl, variable directions of curl, different line spacing and font sizes problems and detects textlines under these problems. Therefore, our method can be integrated with vareity of camera devices (from low to high image quality) and can be used under vareity of image capturing environment (from rough and uncontrolled to highly controlled environment). The textline detection results of presented method can be used for image dewarping, developing efficient binarization techniques and introducing new grayscale OCR algorithms.

References

1. Shafait, F., Keysers, D., Breuel, T.M.: Efficient implementation of local adaptive thresholding techniques using integral images. In: Proc. Document Recognition and Retrieval XV. Volume 6815., San Jose, CA, USA (2008) 81510
2. Zhang, Z., Tan, C.L.: Correcting document image warping based on regression of curved text lines. In: Proc. 7th Int. Conf. on Document Analysis and Recognition, Edinburgh, Scotland (2003) 589–593

3. Lu, S.J., Tan, C.L.: The restoration of camera documents through image segmentation. In: Proc. 7th IAPR workshop on Document Analysis Systems, Nelson, New Zealand (2006) 484–495
4. Fu, B., Wu, M., Li, R., Li, W., Xu, Z.: A model-based book dewarping method using text line detection. In: Proc. 2nd Int. Workshop on Camera Based Document Analysis and Recognition, Curitiba, Barazil (2007) 63–70
5. Gatos, B., Pratikakis, I., Ntirogiannis, K.: Segmentation based recovery of arbitrarily warped document images. In: Proc. 9th Int. Conf. on Document Analysis and Recognition, Curitiba, Barazil (2007) 989–993
6. Stamatopoulos, N., Gatos, B., Pratikakis, I., Perantonis, S.J.: A two-step dewarping of camera document images. In: Proc. 8th IAPR Workshop on Document Analysis Systems, Nara, Japan (2008) 209–216
7. Ulges, A., Lampert, C.H., Breuel, T.M.: Document image dewarping using robust estimation of curled text lines. In: Proc. 8th Int. Conf. on Document Analysis and Recognition, Seoul, Korea (2005) 1001–1005
8. Bukhari, S.S., Shafait, F., Breuel, T.M.: Segmentation of curled textlines using active contours. In: Proc. 8th IAPR Workshop on Document Analysis Systems, Nara, Japan (2008) 270–277
9. Bukhari, S.S., Shafait, F., Breuel, T.M.: Coupled snakelet model for curled textline segmentation of camera-captured document images. In: Proc. 10th Int. Conf. on Document Analysis and Recognition, Barcelona, Spain (2009)
10. Bukhari, S.S., Shafait, F., Breuel, T.M.: Script-independent handwritten textlines segmentation using active contours. In: Proc. 10th Int. Conf. on Document Analysis and Recognition, Barcelona, Spain (2009)
11. Horn, B.K.P.: Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. PhD Thesis, MIT (1970)
12. Riley, M.D.: Time-frequency representation for speech signals. PhD Thesis, MIT (1987)
13. Li, Y., Zheng, Y., Doermann, D., Jaeger, S.: Script-independent text line segmentation in freestyle handwritten documents. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(8) (2008) 1313–1329
14. Du, X., Pan, W., Bui, T.D.: Text line segmentation in handwritten documents using Mumford-Shah model. In: Proc. Int. Conf. on Frontiers in Handwriting Recognition, Montreal, Canada (2008) 1–6
15. Shafait, F., Breuel, T.M.: Document image dewarping contest. In: Proc. 2nd Int. Workshop on Camera Based Document Analysis and Recognition, Curitiba, Brazil (2007) 181–188
16. Shafait, F., Keysers, D., Breuel, T.M.: Performance evaluation and benchmarking of six page segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(6) (2008) 941–954
17. Gatos, B., Antonacopoulos, A., Stamatopoulos, N.: ICDAR 2007 handwriting segmenentation contest. In: Proc. 9th Intelligence Conf. on Document Analysis and Recognition, Curitiba, Brazil (2007) 1284–1288
18. Shafait, F., Keysers, D., Breuel, T.M.: Pixel-accurate representation and evaluation of page segmentation in document images. In: Proc. Int. Conf. on Pattern Recognition, Hong Kong, China (Aug 2006) 872–875
19. Breuel, T.M.: Representations and metrics for off-line handwriting segmentation. In: Proc. 8th Int. Workshop on Frontiers in Handwriting Recognition, Ontario, Canada (2002) 428–433