Coupled Snakelet Model for Curled Textline Segmentation of Camera-Captured Document Images

Syed Saqib Bukhari[1], Faisal Shafait[2], 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

Abstract

Detection of curled textline is important for dewarping of hand-held camera-captured document images. Then baselines and the lines following the top of x-height of characters (x-lines) are estimated for dewarping. Existing curled textline segmentation approaches are sensitive to outlier points and perspective distortions. Furthermore these approaches use regression over top and bottom points of a segmented textline to estimate its x-line and baseline separately, which may results in inaccurate estimation. Here we propose a novel curled textline segmentation approach based on active contours (snakes) in which we perform segmentation by estimating the pairs of x-line and baseline; solving both problems together. Starting form a connected component we jointly trace a pair of x-line and baseline using coupled snakes and external energies of neighboring top-bottom points. We grow neighborhood region iteratively during tracing, which results in robustness to perspective distortions, and maintain a natural property of similar distance within the pair of x-line and baseline pair, which results in robustness to outlier points. We achieved 90.76% of one-to-one match-score recognition accuracy of curled textline segmentation on CBDAR 2007 Document Image Dewarping Contest dataset, with good estimation of pairs of x-line and baseline.

1 Introduction

Curled textline segmentation is the problem of textline detection. It is an important step for dewarping of hand-held camera-captured document images [19, 11, 10, 4, 6, 16, 17, 2]. For dewarping, these approaches rely on accurate detection of curled textlines with proper estimation of pairs of x-line and baseline. Previous curled textline segmentation approaches can be divided into two sub categories:

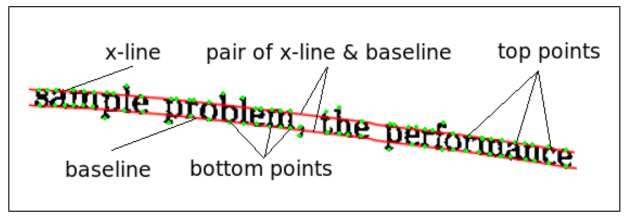


Figure 1: Curled textline definition

(a) heuristic search [19, 11, 10, 4, 6, 16, 17] and (b) active contours (snakes) [2]. In this paper our focus is on curled textline segmentation. Labeling of curled textline is shown in Figure 1, which we will use throughout the paper.

Generally, heuristic search based approaches start from a single component and search other components of a textline in growing neighborhood region. These techniques use complex and rule-based criteria for textline searching. Another general observation is that, they estimate the pair of x-line and baseline for each curled textline after performing segmentation, by fitting regression over top and bottom points separately. These approaches are sensitive to outlier points, high degrees of curls, variable directions of curls, multiple line spacings and different font sizes.

Active contours (snakes) have described in [2] for curled textline detection of camera-captured documents. That work introduced the use of multiple small snakes, referred as "baby-snakes" model. Small slope-aligned straight line snakes are initialized over each smeared connected component (word) in the baby-snakes model. Each snake has some additional length on both sides, which is the function of average width of the smeared connected components. External energy of smeared image is used to deformed each baby snake. After few iterations, neighboring snakes join together and results in curled textline segmentation. This approach considers textline extraction as a fundamental image segmentation problem. Furthermore, this approach can han-

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.204

dle several complex cases in which heuristic search based approaches usually fail, like high degrees of curls, variable directions of curls, different line spacings and font sizes, text-note aligned left to paragraph [2]. That approach [2] dose not deal with the estimation of the pairs of x-line and baseline.

An alternative approach for handwritten textline segmentation using level sets has recently been presented in [9]. We discuss this approach here because it also uses general image segmentation framework for textline segmentation like [2]. In this approach, Gaussian filtering is used to estimate probability density function PDF of pixel values. Level sets are initialized on high PDF values. Growing and merging of level sets is then performed iteratively. This approach has made some modification in the growing and merging criteria of level set, keeping straight horizontal assumption of textlines. Although this assumption generally holds for handwritten text, it breaks down on camera-captured documents due to the high degree of curl present in these documents. Hence this approach is not suitable for segmenting curled textlines.

In this paper, we describe a novel curled textline segmentation approach referred as snakelets model, based on active contours (snakes) [8] using growing neighborhood criteria. This approach differs from previous curled textline segmentation work of baby-snakes model [2] and heuristic searched based approaches [19, 11, 10, 4, 6, 16, 17], through following contributions.

Our snakelets model is a mixture of modified active contours (snakes) and growing-neighborhood-criteria. We perform curled textline segmentation by estimating the pairs of x-line and baseline. In this way we perform segmentation and estimation of pairs of x-line and baseline together, which is not present in previous heuristic search based and active contours (snakes) based approaches. Instead of estimating x-line and baseline of a curled textline individually using regression, we jointly estimate the pair of x-line and baseline. During tracing of each pair, we maintain a natural property of similar distance within a pair, which make our segmentation technique more robust to outliers points as compared to previous approaches. We jointly trace a pair using coupled snakes and external energies of neighboring top-bottom points, which results in proper estimation of x-line and baseline as compared to regression used by previous approaches. Our approach is insensitive to different directions of curls, small and variable line gaps, different font sizes and outliers.

The rest of the paper is organized as follows: Section 2 explains the technical and implementation details of snakelets model for curled textline segmentation. Section 3 deals with the experimental results. Section 4 describe

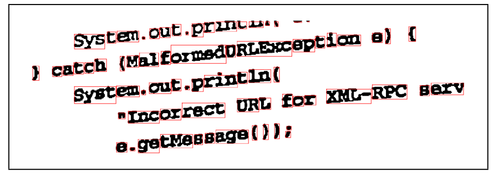


Figure 2: Bounding boxes are drawn around connected components of binarized document image.

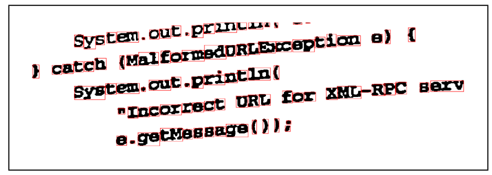


Figure 3: Result of curved-cut segmentation algorithm [1] on binarized document image. Bounding boxes around resulting connected components are drawn.

conclusions.

2 Snakelets

In this Section we discuss our curled textline segmentation approach in detail. In Section 2.1 necessary preprocessing steps are described. In Section 2.2 snakelets model is explained.

2.1 Pre-processing

This method starts by extracting connected components. The bounding box of each connected component is computed and the top and bottom points of connected component which touch the bounding box are calculated. Because of the uneven shading in camera-captured document images, binarization may results in joining of characters as shown in Figure 2. Top and bottom points of these bounding box are dispersed apart and do not accurately represent the information of x-lines and baselines of curled textlines. To overcome this problem, we use the curved-cut segmentation algorithm [1] to cut merged characters in binarized camera-captured document images. Figure 3 shows the result of curved-cut segmentation algorithm. For each connected component find all top and bottom points. From these top and bottom points, we generate two images one containing only top points and other one containing only bottom points, referred as top-points image and bottom-points image respectively. We assume a cleaned up document image as input. If marginal noise is present, it can be removed using page frame detection [15].

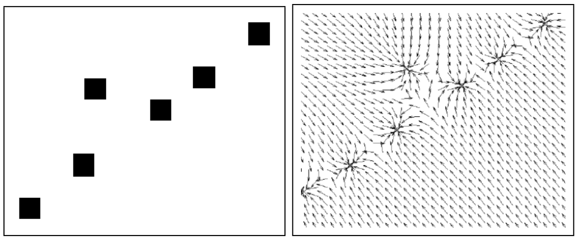


Figure 4: Discrete set of neighboring points (left) and their Gradient Vector Flow (right)

2.2 Curled Textline Segmentation

Our curled textline segmentation model is based on active contours (snakes) [8]. Traditionally, snakes are used for detecting edge boundary of an object in a image. It is a closed curve of points $S(s) = [x(s), y(s)]$, where $s \in [0, 1]$, that moves through the spatial domain of an image to minimize the energy function:

$$E = \int_0^1 [E_{int}\{S(s)\} + E_{ext}\{(S(s)\}ds \quad (1)$$

Internal forces try to keep the curve points close to each other and external forces try to move the curve points towards the boundary of the object. There are different types of external forces: (i) gradient, (ii) gradient of Gaussian and (iii) gradient vector flow (GVF) [18]. Among them GVF has larger capturing range than others.

Our snaklets model is based on active contours (snakes), which track pairs of x-line and baseline of curled textlines in the segmentation step. Therefore, snakelets model produced here can be seen as an extension of the baby-snakes model [2]. The features of snakelet model are explained below:

1. **External forces calculation from discrete points:** For deformation, external forces are calculated from an edge map in image segmentation. We need to estimate the proper x-line and baseline through our open-curve snakelets model, therefore we use discrete set of neighboring points for GVF calculation. Discrete set of points and their GVF are represented in Figure 4. Like the baby-snakes model, we deform only vertical components of open-curve snake points with respect to the vertical components of GVF.

2. **Evolving snake:** Traditionally, snakes update their position and number of points using the same external force throughout the deformation life cycle. We have observed that the large amount of noise/unfocused-objects within image causes poor segmentation due to fixing the external force area. Therefore here we introduce the concept of evolving snake with growing external force area. We grow the snake length and GVF calculation area before every next deformation step, with respect to the starting position. In this way we can accurately perform segmentation, especially the territory of x-lines and baselines under high degrees and different directions of curls.

3. **Weighted-coupled snakes pair:** In medical image segmentation, researchers [7, 3] have introduced coupled snakes to perform segmentation of similarly shaped objects. In this paper we use this idea to estimate pair of x-line and baseline, as the x-line and baseline deform in a similar way for any given textline. In ”weighted-coupled snakes” we initialize a pair of open-curve snakes on top and bottom points of a connected component. During coupling, we give high weights to baseline because in general more components lie on the baseline than on the x-line due to the higher ratio of ascender than descenders. We use small percentage of GVF of top neighboring points to deform top open-curve snake and high percentage of GVF of bottom neighboring points to deform bottom open-curve snake. After deformation we adjust the distances, at each common x-coordinate point, between them and make them equal to the average distance. In this way we can accurately estimate the pair of x-line and baseline.

Based on the snakelet model, curled textline segmentation is performed as follow. Top-points and bottom-points images are calculated using curved-cut connected component image in pre-processing step. Let L and H be the average length and average height of connected components. Consider a single connected component as the source component. A pair of open-curve snakes of length L are initialized over the top- and bottom-points of that connected component, referred as x-line- and baseline-snake respectively. Regions, equivalent to area $2L \times H$, are selected from top- and bottom-points images, taking the source connected component as center. These regions are referred as top- and bottom-neighborhood. Vertical components of GVFs of these regions are calculated. X-line-snake is deformed using the 50% weights of vertical components of top-neighborhood GVF and baseline-snake is deformed using the 100% weights of vertical components of bottom-neighborhood GVF. After deformation, distances between each common x-coordinate of x-line- and baseline-snake are calculated. Then x-line- and baseline-snake points are adjusted to make the distances equivalent to average distance. For the same connected component, similar proce-

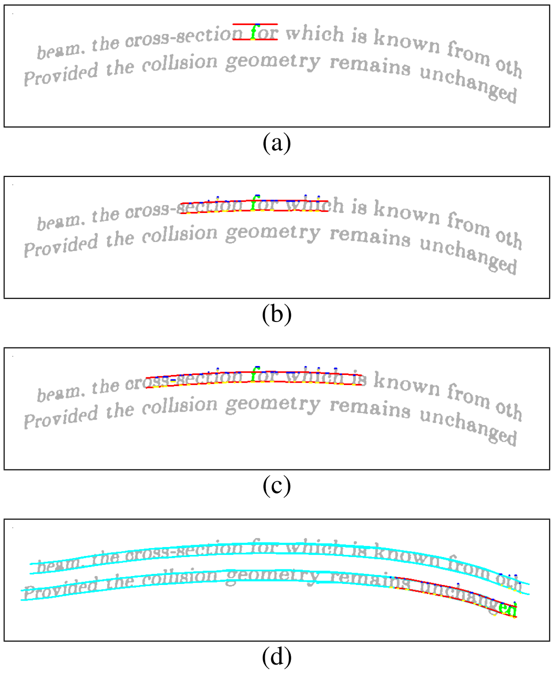


Figure 5: Flow of Snakelets algorithm for segmenting curled textlines by estimating the pairs of x-line and baseline. (a) Initial pair of x-line and baseline snakes (in red color) over a connected component is represented. (b) Deformed snakes pair after 1st iteration. (c) Deformed snakes pair after last iteration. (d) Final result: Already deformed pairs are represented with blue color and last deformed pair of last connected component is represented with red color.

dure is repeated for 3-4 times, every time with growing x-line- and baseline-snake by units L and growing neighborhood region by units $2L \times H$, with respect to previous one. We do the same for all connected components within document image. All overlapping pairs of x-line and baseline snakes are parts of single curled textline. The flow of algorithm steps is shown in Figure 5.

3 Experiments and Results

To test the performance of our curled textline segmentation algorithm on real-world documents, we evaluated it on the data set used in the CBDAR 2007 document image dewarping contest [12]. This data set consists of documents images captured with a hand held camera in an uncontrolled environment. Although comprehesive performance evaluation methods like [13, 14] could be used, we present results using ICDAR 2007 handwriting context evaluation methodology [5] for comparison purposes. In our experiments, we used the cleaned up version of dataset and selected 92 documents containing textlines only. The evaluation methodology is based on match score computation:

Table 1: Performance evaluation results

Detected elements (M)	2790
One to one match ($o2o$)	2465
One detected to many ground truth (d_{o2m})	79
Many detected to one ground truth (d_{m2o})	190
Recognition accuracy (RA)	90.76%

$$MatchScore(i, j) = \frac{T(G_j \cap R_i \cap I)}{T(G_j \cup R_i \cup I)} \quad (2)$$

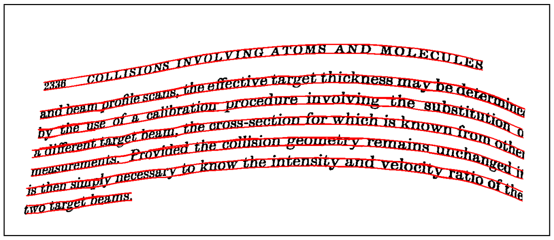
where I is the set of all image pixels, G_j and R_i are the sets of all pixels covering the j^{th} ground truth region and i^{th} result region respectively. Based on the matching scores recognition accuracy (RA) is calculated as follows:

$$RA = w_1 \frac{o2o}{M} + w_2 \frac{d\_o2m}{M} + w_3 \frac{d\_m2o}{M} \quad (3)$$

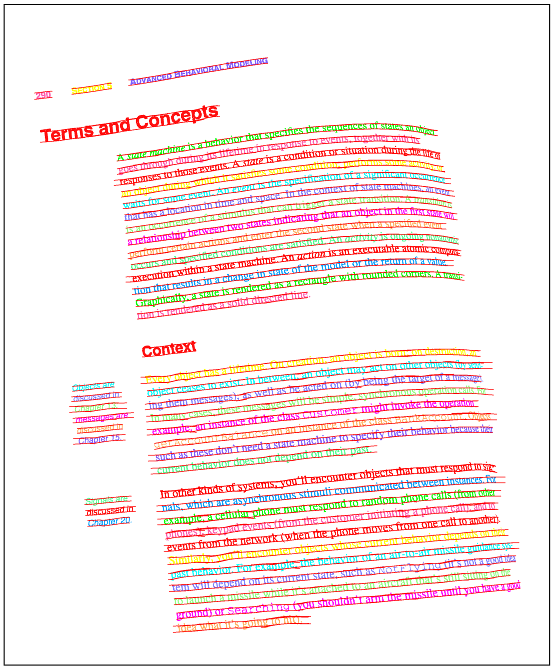
where, M is the number of detected components, $o2o$ is one to one match with ground truth components, d_o2m is one detected to many ground truths, d_m2o is many detected to one ground truth, and $w_1 = 1$, $w_2 = 0.25$ and $w_3 = 0.25$ are pre-determined weights. The weights were set to the same values as in [5]. Performance evaluation results based on equations (2) and (3) are given in Table 1.

4 Conclusion

The paper describes a novel approach to finding curled textlines of camera-captured document images using active contour models. Our snakelets model is simpler then the baby-snake model [2], but yields more information as shown in Figure 6. We achieved 90.76% one-to-one match-score recognition accuracy on the data set used in the CB-DAR 2007 Document Image Dewarping Contest. Most of the errors are due of oversegmentation, that is more than one textlines detected for a single groundtruth textline. We can improve our recognition accuracy by performing neighborhood proximity criteria as a postprocessing step. Without doing any postprocessing, our initial recognition accuracy of 90.76% implies good segmentation results. Unlike other approaches, our approach is robust against high degrees of curls, variable directions of curls, different line spacings and font sizes, text-note aligned left to paragraph, as shown in Figure 6. Furthermore, our approach estimates the pairs of x-line and baseline using GVF forces which gives more accurate estimation, shown in Figure 6, as compared to regression used by other approaches.



(a)



(b)

Figure 6: Result of Snakelets model.

5 Acknowledgement

We would like to thanks Basilis Gatos of Computational Intelligence Laboratory, National Center for Scientific Research "Demokritos", Greece, for providing us the textlines segmentation evaluation software.

References

[1] T. M. Breuel. Segmentation of handprinted letter strings using a dynamic programming algorithm. In *6th Int. Conference on Document Analysis and Recognition*, pages 821–826, 2001.

[2] S. S. Bukhari, F. Shafait, and T. M. Breuel. Segmentation of curled textlines using active contours. In *8th IAPR Workshop on Document Analysis Systems*, pages 270–277, 2008.

[3] R. Chandrashekara, R. H. mohiaddin, and D. Rueckert. Analysis of myocardial motion in tagged MR images using nonrigid image registration. *Progress in biomedical optics and imaging*, 3(2):1168–1179, 2002.

[4] B. Fu, M. Wu, R. Li, W. Li, and Z. Xu. A model-based dewarping method using text line detection. In *2nd Int. Workshop on Camera Based Document Analysis and Recognition*, September 2007.

[5] B. Gatos, A. Antonacopoulos, and N. Stamatopoulos. ICDAR2007 handwriting segmenentation contest. In *9th Int. Conference on Document Analysis and Recognition*, pages 1284–1288, 2007.

[6] B. Gatos, I. Pratikakis, and K. Ntirogiannis. Segmentation based recovery of arbitrarily warped document images. In *Proceedings of Int. Conference on Document Analysis and Recognition*, 2007.

[7] B. Hohnhaeuser and G. Hommel. 3d pose estimation using coupled snakes. *J. WSCG*, 12(1-3):1213–6972, Feb 2003.

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. Journal of Computer Vision*, 1(4):1162–1173, 1988.

[9] Y. Li, Y. Zheng, D. Doermann, and S. Jaeger. Script-independent text line segmentation in freestyle handwritten documents. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(8):1313–1329, 2008.

[10] S. J. Lu, B. M. Chen, and C. C. Ko. Perspective rectification of document images using fuzzy set and morphological operations. *Image and Vision Computing*, 23:541–553, 2005.

[11] S. J. Lu and C. L. Tan. The restoration of camera documents through image segmentation. In *7th IAPR workshop on Document Analysis Systems*, pages 484–495, 2006.

[12] F. Shafait and T. M. Breuel. Document image dewarping contest. In *2nd Int. Workshop on Camera-Based Document Analysis and Recognition*, Curitiba, Brazil, Sep 2007.

[13] F. Shafait, D. Keysers, and T. M. Breuel. Pixel-accurate representation and evaluation of page segmentation in document images. In *Int. Conference on Pattern Recognition*, pages 872–875, Hong Kong, China, Aug 2006.

[14] F. Shafait, D. Keysers, and T. M. Breuel. Performance evaluation and benchmarking of six page segmentation algorithms. *IEEE Tans. on Pattern Analysis and Machine Intelligence*, 30(6):941–954, Jun 2008.

[15] F. Shafait, J. van Beusekom, D. Keysers, and T. M. Breuel. Document cleanup using page frame detection. *Int. Jour. on Document Analysis and Recognition*, 11(2):81–96, 2008.

[16] N. Stamatopoulos, B. Gatos, I. Pratikakis, and S. J. Perantonis. A two-step dewarping of camera document images. In *8th IAPR Workshop on Document Analysis Systems*, pages 209–216, 2008.

[17] A. Ulges, C. H. Lampert, and T. M. Breuel. Document image dewarping using robust estimation of curled text lines. In *Int. Conference on Document Analysis and Recognition*, pages 1001–1005, 2005.

[18] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. In *IEEE Transaction of Image Processing*, pages 359–369, 1998.

[19] Z. Zhang and C. L. Tan. Correcting document image warping based on regression of curved text lines. In *7th Int. Conference on Document Analysis and Recognition*, pages 589–593, 2003.