Script-Independent Handwritten Textlines Segmentation using Active Contours

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

Handwritten document images contain textlines with multi orientations, touching and overlapping characters within consecutive textlines, and small inter-line spacing making textline segmentation a difficult task. In this paper we propose a novel, script-independent textline segmentation approach for handwritten documents, which is robust against above mentioned problems. We model textline extraction as a general image segmentation task. We compute the central line of parts of textlines using ridges over the smoothed image. Then we adapt the state-of-the-art active contours (snakes) over ridges, which results in textline segmentation. Unlike the "Level Set" and "Mumford-Shah model" based handwritten textline segmentation methods, our method use matched filter bank approach for smoothing and does not require heuristic postprocessing steps for merging or splitting segmented textlines. Experimental results prove the effectiveness of the proposed algorithm. We evaluated our algorithm on ICDAR 2007 handwritten segmentation contest dataset and obtained an accuracy of 96.3%.

1 Introduction

Handwritten textline segmentation is the problem of textline extraction. It is an important step for handwritten document recognition. Handwritten document image contains, (i) multi-oriented textlines (Figure 3a), (ii) high degrees of curl (Figure 3b), (iii) touching and overlapping components (Figure 3c), (iv) small interline spacing (Figure 3c), (v) noisy components, like signature, ornamentation (Figure 3a), seal, check mark, etc., and (vi) complex historical writing style, like 'cross-hatching' (Figure 3d). As a result, textline segmentation of handwritten document image is a difficult task. Various approaches to textlines detection have been reported. These approaches can be classified into two main categories: (i) connected component based [10, 15, 9, 11] and (ii) deformable model-based [8, 3].

The method described by Nicolas, 2004 [10] is based on a rule based system. The approach described by Yin, 2008 [15] groups connected components using minimal spanning tree MST with distance measure. The textline segmentation method proposed by Loulouids, 2008 [9] is based on Hough transform on subset of connected components. A multi-oriented textline segmentation approach is introduced by Ouwayed, 2008 [11]. In this method, zone-segmentation is done by using projection and Wigner-Ville distribution WVD. Within each zone, textlines are segmented using proximity of connected components and zone orientation. The textline segmentation approach described by Li, 2008 [8] introduces density estimation for textlines; the image is smoothed by using an anisotropic Gaussian filtering and then boundary based level set method is adapted for textline segmentation. The textline segmentation approach described by Du, 2008 [3] follows the same smoothing step as proposed by Li, 2008 [8] and then region based Mumford-Shah energy function is used for textline segmentation. This approach [3] uses morphological postprocessing for splitting and joining segmented textlines.

In case of multi-oriented textlines handwritten document image, the above approaches depend on skew-correction and zone segmentation before textline segmentation. These approaches are sensitive to large number of touching and overlapping components and use heuristic postprocessing rules for splitting and joining segmented textlines to handle these cases. Furthermore, connected component based approaches [10, 15, 9, 11] are more sensitive to small interline spacing, topological changes and noisy components as compared to deformable model-based approaches [8, 3]. However, deformable model-based approaches [8, 3] are sensitive to the heuristic selection of σ_x and σ_y for an anisotropic Gaussian mask under high degrees of skew and curl. Furthermore, level set deformable model-based approach [8] is sensitive to the number of level set evolution iterations.

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.206

In this paper we describe a novel script-independent textline segmentation approach for handwritten document image based on active contours (snakes) [7]. Our method is related to curled textline segmentation approach described by Bukhari, 2008 [1], where active contours (snakes) were applied to camera-captured printed documents, referred as “baby-snakes” model. In baby-snakes model, slope-aligned straight line snakes are initialized over each smeared connected components. Snakes are deformed only in vertical direction, because of considering the horizontal nature of textlines. For deformation, external energy of smeared image is used to force each baby snake towards its neighboring components (words). After few deforming iterations, neighboring snakes join together and result in curled textline segmentation of camera-captured document image. Our approach of handwritten textline segmentation is different from the active contours based curled textline segmentation approach [1]. Our method is started by image smoothing using a multi-oriented anisotropic Gaussian filter bank approach, similar to matched filter bank approach [2]. Then we estimate ridges [6] over the smoothed image and then adapt an active contours (snakes) [7] over ridges to perform textlines segmentation.

We make the following contributions in this paper. We use a multi-oriented anisotropic Gaussian filter bank for density estimation and image smoothing. As compared to an anisotropic Gaussian filter, a multi-oriented anisotropic Gaussian filter bank enhances the textline structure well, especially under high degrees of multi-oriented skew and curl within a document image, as shown in Figure 1. Therefore, our approach is robust to multi-oriented documents and does not depend upon skew-correction and zone-segmentation. We introduce ridges detection for approximating central line parts of textlines, as shown in Figure 2b with red colors. These lines are linearly smooth as compared to irregular boundaries of textlines in the smoothed image. Therefore, we assume that initializing deformable model over ridges is more robust and accurate as compared to initializing deformable models over the smoothed image. We adapt an active contours (snakes) over ridges for performing segmentation, such that our deformable model is independent of number of deformation iterations. Our approach is free from heuristic parameters and postprocessing steps.

The rest of the paper is organized as follows: Section 2 deals with the technical and implementation details of proposed textline segmentation algorithm. Section 3 comprises the performance evaluation and experimental results. Section 4 discusses the results and conclusion.

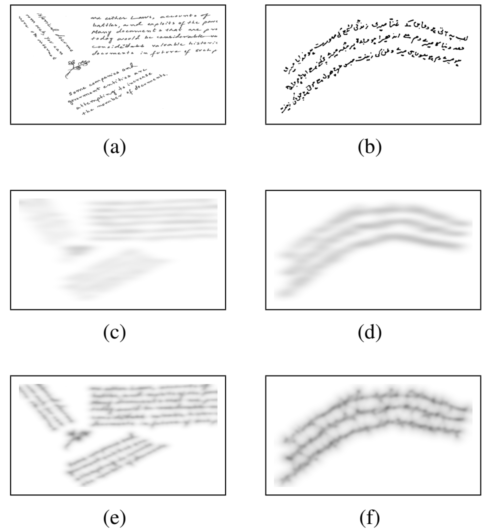

Figure 1: Textline enhancement: (a) Sample image of multi-oriented textlines of “English” script. (b) Sample image of high degrees of curl textlines of “Urdu” script. (c) and (d) Bad textline enhancement results due to anisotropic Gaussian smoothing, where $\sigma_x > \sigma_y$. (e) and (f) Proper textline enhancement results, using multi-oriented anisotropic Gaussian filter bank approach.

2 Handwritten Textline Segmentation

Our proposed approach for handwritten textline segmentation is capable enough to handle multi-oriented textlines, high degrees of curl, touching and overlapping components within textlines, large amount of noise, small interline space and noisy elements. In this section our proposed approach is discussed. In Section 2.1 a multi-oriented anisotropic Gaussian filter bank smoothing based on matched filter bank approach is described. In Section 2.2 detection of ridges over the smoothed image is discussed. In Section 2.3 adaptation of an active contours (snakes) over ridges is explained.

2.1 Multi-Oriented Textlines Smoothing: Matched Filter Bank

Several previous approaches [8, 3] to enhancing textline structure are based on anisotropic Gaussian filtering with heuristically fixed values for σ_x and σ_y, where $\sigma_x > \sigma_y$. The motivation for taking $\sigma_x > \sigma_y$ is the consideration of horizontal nature of textlines. But a handwritten document may contain skew, curl or multi-oriented textlines. In such cases, an anisotropic Gaussian smoothing, with $\sigma_x > \sigma_y$, is unable to enhance the textline structure, as shown in Figures 1c and 1d . The presence of skew, curl and multi-oriented textlines within handwritten document

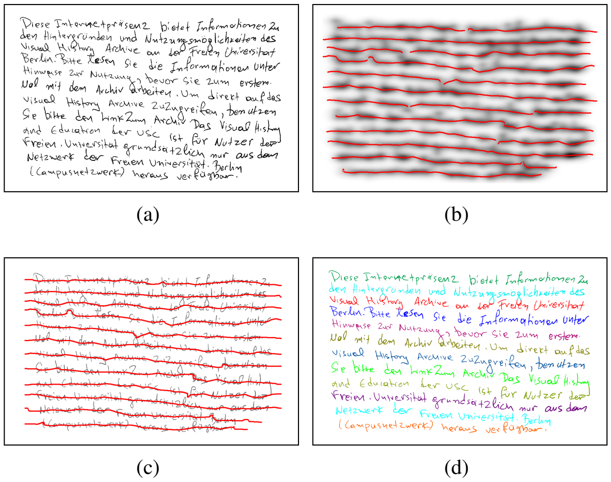

Figure 2: (a) Image portion from ICDAR 2007 contest dataset. (b) Ridges over smoothed image, shown in red colors. (d) Deformed ridges, shown in red colors. (f) Segmented textlines.

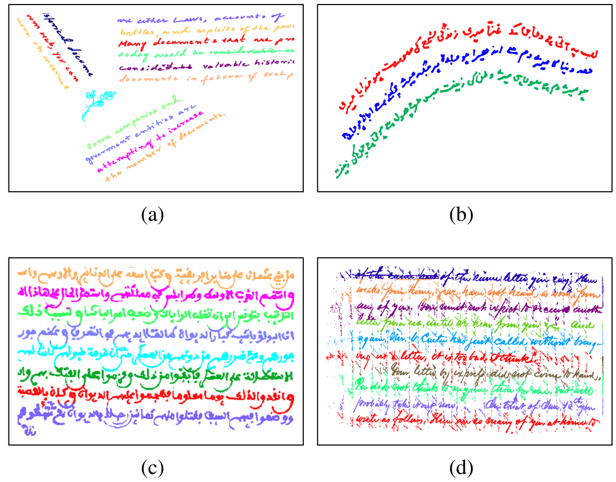

Figure 3: Segmentation results: (a) Multi-oriented "English" script textlines. (b) High degrees of curl textlines of "Urdu" script. (c) Touching and overlapping "Arabic" textlines with small interline space. (d) Ancient "Cross-Hatching" English writing style: Fill the page and then to turn the paper ninety degrees and continue writing over previous text.

image motivate us to use smoothing similar to matched filter bank approach [2] for enhancing textline structure. For this purpose, we are using a multi-oriented anisotropic Gaussian filter bank approach based on matched filter bank approach [2]. We generate a set of filters with different values for σ_x, σ_y and θ, by making all possible combinations among the ranges of σ_x, σ_y and θ. The range of θ is from -45° to 45°. σ_x and σ_y share the same range. Instead of selecting heuristic range for σ, we use average height of connected components. Let H be the average height of connected components, the range for σ is aH to bH, where $a > b$. In this way, our filter bank covers all the orientation with verity of σ_x and σ_y. For each image pixel value, we apply the set of filter bank and select the maximum value among them. Results of multi-oriented anisotropic Gaussian filter bank smoothing of Figures 1a and 1b are shown in Figures 1e and 1f respectively. As compared to anisotropic Gaussian smoothing, multi-oriented anisotropic Gaussian filter bank smoothing enhances the textlines structure reasonably well under skew, curl and multi-oriented textlines in handwritten document image.

2.2 Approximation of Central Line of Parts of Textline: Ridges-Detection

Several researches [6, 4] have been done for defining the edge- and region-based structure for objects of grayscale images by analyzing the intensity values of image and then identify ridges over it. In previous section, we have seen that a multi-oriented anisotropic Gaussian filter bank smoothing enhances the textline structure of handwritten documents well. Therefore, detection of ridges over the smoothed image can give good approximation of central line of parts of textline. We use Horn-Riley [6] based ridges detection approach. In this approach, ridges are constructed by finding zero crossing of the appropriate directional derivatives of smoothed image. Central line of parts of textlines, that are ridges, are shown in Figure 2b with red colors. In next section we model active contours (snakes) over the ridges, which results in textlines segmentation.

2.3 Textlines Segmentation: Adaptation of Active Contours (Snakes) over Ridges

An active contour (snake) [7] is a curve, which moves over the image in a way to minimize the energy function:

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

where S represents a curve, with $S(s) = [x(s), y(s)]$ and $s \in [0, 1]$, E_{int} is internal snake energy that tries to keep the curve points together and E_{ext} is external energy calculated from the image data, which tries to move the curve points towards the targeted object. There are three main types of external energies: (i) Gaussian, (ii) gradient of Gaussian and (iii) gradient vector flow (GVF) [14]. Among them, GVF has large capturing range than others [14].

The main motivation of using active contours (snakes) [7] model in our handwritten textline segmentation is the presence of central line of parts of textlines, in the form of ridges. By considering ridges as open-curve snakes, we can join neighboring ridges without using any heuristic neighboring criteria, but by adapting snakes over ridges. We will refer this snake model as "ridge-snakes" model, which is different from our already proposed baby-snakes model [1]. The details of ridge-snakes model and textline segmentation are given below.

The ridges of the handwritten document image of Figure 2a are shown in Figure 2b. Here we use ridges as initial open-curve snakes. We increase the length of each open-curve snake from both ends. The size of the additional length is a factor of average width of connected components of handwritten document image and the slope of additional length is similar to the slope of open-curve snake. The motivation of initializing this type of snakes is to join the neighboring snakes together, using external forces of the image. We compute the horizontal gradient vector flow (GVF) image and vertical GVF image from ridges image; the image which contains only ridges. The GVF vectors in both GVF images point towards the ridges. We deform each snake either using horizontal GVF image or using vertical GVF image, based on the approximated slope of snake. If the slope of snake is in between -45° to 45° then we use vertical GVF image and update only vertical components of snake points. And if the slope of snake is in between 45° to 135° then we use horizontal GVF image and update only horizontal components of snake points. Once the snake reaches near a ridge, it does not change its position regardless of further deformation steps. After few deformation steps, neighboring snakes stick together. Each set of connected snakes, referred as chain, covers the complete range of textline, as shown in Figure 2c. A label is assigned to each connected component of handwritten document image based on the overlapping chain.

If more than one chain overlap over a connected component, then the connected component can be cut horizontally or vertically depending on the slope of underlying ridges into equal number of parts.

Each unlabeled connected component shares the nearest segmented textline label. Final textline segmentation result of handwritten document of Figure 2a is shown in Figure 2d. Sample textline segmentation results of complex handwritten document images are shown in Figure 3.

3 Experiments and Performance Evaluation

To test the performance of our algorithm on real-world documents, we evaluate it on the ICDAR 2007 Handwritten Segmentation Contest dataset [5]. This dataset consist of 80 unconstrained handwritten document images. The dataset is composed of handwritten historical archives, documents from several writers and scanned handwritten documents sample from web. The ground truth is generated by manual annotation. The total number of textlines in the dataset is 1770. Although comprehesive performance evaluation methods like [12, 13] could be used, we present results using ICDAR 2007 handwriting context evaluation methodology [5] for comparison purposes. This methodology considers a match only if the matching score is equal to or above a specified acceptance threshold, where matching score is defined as:

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

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, detection rate (DR), recognition accuracy (RA) and combine performance metric FM are calculated as follows:

$$DR = w_1 \frac{o2o}{N} + w_2 \frac{g\_o2m}{N} + w_3 \frac{g\_m2o}{N} \tag{3}$$

$$RA = w_4 \frac{o2o}{M} + w_5 \frac{d\_o2m}{M} + w_6 \frac{d\_m2o}{M} \tag{4}$$

$$FM = (2 * DR * RA)/(DR + RA) \tag{5}$$

where N and M are the total number of ground truth and results elements respectively, $w_1, w_2, w_3, w_4, w_5, w_6$ 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. The weights chosen to compute DR and RA were set to the same values as in [5].

Five different methods participated in the ICDAR 2007 Handwritten Segmentation Contest [5]. Based on the above mentioned equations (2) to (5), the handwritten textlines segmentation results of all participated algorithms [5] together with our segmentation results are shown in Table 1.

4 Discussion

We have described a novel approach for script independent handwritten textlines segmentation based on active contours (snakes). We achieved 97.7% detection rate, 95.4% recognition accuracy and 96.3% FM, over ICDAR 2007 Handwritten Segmentation Contest dataset, which proves the effectiveness of our approach. We introduced

Table 1: Results on ICDAR 2007 handwriting segmentation contest dataset. For comparison, results of methods that participated in the contest are taken from [5] and have been included in the table.

Algorithm	M	o2o	g_o2m	g_m2o	d_o2m	d_m2o	DR	RA	FM
BESUS	1904	1494	9	151	72	21	86.6%	79.7%	83.0
DUTH-ARLSA	1894	1214	149	227	107	354	73.9%	70.2%	72.0%
ILSP-LWSeg	1773	1713	5	34	17	10	97.3%	97.0%	97.1%
PARC	1756	1604	40	76	34	85	92.2%	93.0%	92.6%
UoA-HT	1770	1674	14	54	27	28	95.5%	95.4%	95.4%
Ridges-Snakes	**1798**	**1704**	**32**	**41**	**19**	**25**	**97.3%**	**95.4%**	**96.3%**

a multi-oriented anisotropic Gaussian filter bank smoothing and used ridges detection for approximating the central line of parts of textlines. Ridges are uniform, linear and smooth as compared to irregular boundaries of textlines in smoothed images. In this way, we contribute in introducing the more uniform structure for the initialization of deformable models (level set, Mumford-Shah and snakes). We model active contours (snakes) over ridges, such that our segmentation results are insensitive to the number of iterations, unlike level set based textlines segmentation [8]. Because of the novel combination of multi-oriented anisotropic Gaussian filter bank smoothing, ridges and active contours (snakes), our handwritten textlines segmentation approach is robust against high degrees of skew, curls, multi-oriented textlines, small interline space and noise. Our method requires no postprocessing. Unlike other approaches, our approach does not require skew-correction or zone-segmentation.

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] 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.

[2] S. Chaudhuri, S. Chatterjee, N. K. abd M. Nelson, and M. Goldbaum. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Tans. on Medical Imaging*, 8(3):263–269, 1989.

[3] X. Du, W. Pan, and T. D. Bui. Text line segmentation in handwritten documents using mumford-shah model. In *11th Int. Conf. on Frontiers in Handwriting Recognition*, pages 1–6, 2008.

[4] D. Eberly, R. Gardner, B. Morse, S. Pizer, and C. Scharlach. Ridges for image analysis. *Journal of Mathematical Imaging and Vision*, 4(4):353–373, 1994.

[5] B. Gatos, A. Antonacopoulos, and N. Stamatopoulos. Handwriting segmentation contest. In *9th Int. Conf. on Document Analysis and Recognition*, pages 1284–1288, 2007.

[6] B. K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. *PhD Thesis, MIT*, 1970.

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

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

[9] G. Louloudis, B. Gatos, I. Pratikakis, and C. Halatsis. Line and word segmentation of handwritten documents. In *11th Int. Conf. on Frontiers in Handwriting Recognition*, 2008.

[10] S. Nicolas, T. Paquet, and L. Heutte. Text line segmentation in handwritten document using a production system. In *9th Int. Workshop on Frontiers in Handwriting Recognition*, pages 245–250, 2004.

[11] N. Ouwayed and A. Belad. Multi-oriented text line extraction from handwritten arabic documents. In *8th IAPR Int. Workshop on Document Analysis Systems*, pages 339–346, 2008.

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

[13] 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.

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

[15] F. Yin and C.-L. Liu. Handwritten text line segmentation by clustering with distance metric learning. In *11th Int. Conf. on Frontiers and Handwritten Recognition*, pages 229–234, 2008.