Dewarping of Document Images using Coupled-Snakes

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

Traditional OCR systems are designed for planar (dewarped) images and the accuracy is reduced when applied on warped images. Therefore, developing new OCR techniques for warped images or developing dewarping techniques are the possible solutions for improving OCR accuracy camera-captured documents. Among different types of dewarping techniques, curled textlines information based dewarping techniques are the most popular ones, but are sensitive to high degrees of curl and variable line spacing. In this paper we build a novel dewarping approach based on curled textlines information, which has been extracted using ridges based modified active contour model (coupled-snakes). Our dewarping approach is less sensitive different direction of curl and variable line spacing. Experimental results show that OCR error rate, from warped to dewarped documents, has been reduced from 5.15% to 1.92% on the dataset of CBDAR 2007 document image dewarping contest. We also report the performance of our method in comparison with other state-of-the-art methods.

1 Introduction

For document analysis and recognition, flat-bed scanners are traditionally and widely used in document capturing, that produce planar images with high resolution. From decades many novel approaches have been proposed for planar document image segmentation [1] and OCR [2]. Nowadays high performance cameras are available at low cost, that offer fast, easy, flexible and non-contact imaging. These advantages make camera a potential substitute of scanner for document capturing and at the same time open doors for many new applications, like mobile OCR, digitizing thick books, digitizing fragile historical documents, finding text-in-images, etc. But camera-captured document images suffer from various distortions, like non-planar (warped) shape, uneven light shading, motion blur, perspective distortion, under- and over-exposure. Therefore, current OCR systems, which are designed for planar document images, do not have capability to deal with these distortions and give poor performance when applied directly to warped camera-captured document images. There could be two possible solutions for improving the OCR performance of warped document images: (i) design new camera-based document analysis and recognition techniques, like specialized binarization, curled textlines detection which can help in layout analysis and character segmentation, blurred and low resolution character recognition, etc. or (ii) design dewarping techniques for flattening document images such that current OCR systems can be directly applied to them.

So far, much attention has not been given to developing new OCR techniques for camera-captured document images. But over last decade, different approaches have been proposed for document image dewarping [3, 4]. These approaches can be divided into two main categories based on the document capturing methodology: (i) approaches in which specialized hardware arrangement, like stereo-camera, is required for 3D shape reconstruction of warped document [5, 6, 7] and (ii) approaches in which dewarping method is designed for image which is captured using single hand-held camera in an uncontrolled environment [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. Hand-held camera-based dewarping approaches can be further classified into two groups: (i) approaches based on document geometry [8, 9, 10, 11] and (ii) approaches based on curl textlines information [12, 13, 14, 15, 16, 17, 18, 19].

Our dewarping approach, presented here, falls under the category of single hand-held camera-based approaches using curled textlines information. Literature review on curled textline based dewarping approaches [12, 13, 14, 15, 16, 17, 18, 19] and curled textlines finding techniques [20, 21], [Ridges based Coupled-Snakes[1]] are given below.

[1]Detailed explanation of "Textline Information Extraction from Grayscale Camera-Captured Documents" will be published elsewhere.

Zhang and Tan [12] proposed dewarping for scanned images from thick bound volumes. They consider that the major portion of the image is straight. They estimate neighboring curved portions of each straight textline by clustering connected components and then move these connected components parallel to their straight line portion.

Ulges et al. [13] proposed dewarping technique based on priori layout information and local textline approximation using RAST [22]. After local textlines finding, they estimate quadrilateral cell for each letter and then map to a rectangle of appropriate size and position in the dewarped image.

Lu et al. [14] introduced a rectification technique for restoring documents with perspective distortions. Their algorithm is based on tip-points and vertical stroke boundaries estimation using morphological operations. They use top points and bottom points for estimating top and bottom textlines respectively. They estimate source quadrilaterals using vertical stroke boundaries and textlines pairs and then construct rectification homography using each pair of source and target quadrilaterals.

Lu and Tan [15, 16] proposed dewarping algorithms which are the extension of work presented in [14]. These algorithm can remove skew, perspective and geometric distortions.

Gatos et al. [17] introduced dewarping techniques using textlines information. They perform horizontal smoothing [23] to combine characters into words and then find lines by grouping neighboring words. They rotate each word of a line individually based on its slope and then align all words of a line with respect to the left most word.

Gatos and Ntirogiannis [18] proposed dewarping approach based on the estimation of words and textline by using the modified "box hand" [24, 25] approach. Similar to [17], they rotate each word of a line and then align all words of a line with respect to the left most word.

Dewarping technique by Fu et al. [26] starts by estimating sub lines using characters combination method [27].They cluster mid points of sub lines based on a proximity criteria, which results in textlines. From these textlines, they estimate left and right borders and top and bottom curves. They stretch cylinder surface area into planar surface area based on the model presented in [5].

Stamatopoulos et al. [19] introduced a two step dewarping algorithm. They estimate textlines using method introduced in [17]. In coarse dewarping step, they estimate left and right borders and top and bottom curves using textlines information and then transform curve area into 2D rectangular area. In fine dewarping step, they perform dewarping algorithm, presented in [17], over coarse dewarped image.

We have already described versatile active contours (snakes) based curled textlines detection techniques for binary and grayscale camera-captured document images [20, 21, Ridges based Coupled-Snakes[1]]. our curled textlines detection techniques are less sensitive to high degrees of curl, variable direction of curl, different line spacing and font sizes, as compared to above mentioned curled textlines approaches (as part of dewarping algorithms).

In this paper we introduce document image dewarping approach based on curled textline information. The method starts by estimating x-line:baseline pairs using ridges based coupled-snakes model[1]. The starting position of each straight textline in dewarped image is calculated using neighboring curled textlines in warped image and then geometric distortion (curl) is removed by mapping characters over each curled x-line:baseline pair to its corresponding straight x-line:baseline pair. Finally perspective distortion is removed by using four point homography algorithm.

The rest of the paper is organized as follows: Section 2 describes the technical and implementation details of dewarping algorithm. Section 3 comprises the performance evaluation and experimental results. Section 4 discusses the results and conclusion.

2 Dewarping Algorithm

Our dewarping algorithm comprises of three steps: (1) curled textline information extraction from binarized document image using ridges based coupled-snakes model, (2) geometric distortion handling using curled textline information and (3) perspective distortion handling using four point homography algorithm. All these steps are described below.

2.1 Curled Textline Information Extraction

We have already described curled textline information extraction from grayscale document images using ridges based coupled-snakes model[1]. The dewarping algorithm presented here use this technique for finding textlines information from binarized document images. Here is the brief overview of ridges based coupled-snakes model for curled textlines information extraction.

The process starts by enhancing curled textline structure using multi-oriented multi-scale anisotropic Gaussian smoothing, based on match filter bank approach [28, 29]. A single range is selected for both σ_x and σ_y, which is the function of the height of the document image (H), that is aH to bH with $a < b$. The range for θ is set from -45 to 45 degrees. From these ranges, a set of filters is generated for all possible combinations of σ_x, σ_y and θ. This set of filters is applied to each pixel of binary image and the maximum value among them is selected. Figures 1(a) and 1(b) show the input and smoothed images respectively.

Horn-Riley [30, 31] based ridges detection approach is used for finding the central lines structure for textlines. This approach is based on the information of local direction of gradient and second derivatives as the measure of curvature. From this information, given 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 1(b) are shown in Figure 1(c). It is clearly visible in the Figure 1(c) that each ridge covers the complete central line structure of a textline, which results in textlines detection.

After detecting textlines, the task is to determine the x-line:baseline pairs information for textlines. For this purpose, the gradient of binary image is computed by using Sobel filter. Then gradient image is divided into two images: one contains positive magnitudes (top-image) and another one contains absolute values of negative magnitudes (bottom-image). Top-image is dominated by the top parts of curled textlines and similarly bottom-image is dominated by the bottom parts of curled textlines. Then gradient vector flow (GVF) [32] of both images are calculated. Active contour (snake) [33] is adapted over ridges. The duplicated ridges are used as initial open-curve snakes pairs for curled textlines. For each pair, one snake is deformed with respect to the vertical components of GVF of top-image and another one with respect to the vertical components of GVF of bottom-image. Large percentage of GVF of bottom gradient image and small percentage of GVF of top gradient image are used during coupling, because of the assumption that more characters lie on baseline than on x-line. After each deformation iteration, the distances between each pair of snakes are adjusted to make them equal to average distance.

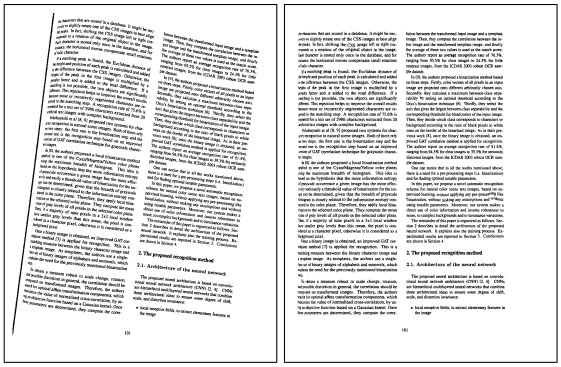
Figures 1(d) and 1(e) show the properly estimated pairs of x-line:baseline for curled textlines.

2.2 Handling of Geometric Distortions

An easy and efficient way of handling geometric (curled) distortion by using curled-line pairs is to estimate corresponding straight-line pairs and then map all pixel values from curled-line pairs to straight-line pairs. For each textline, the starting and ending x-coordinate values of straight-line pair is set similar to curled-line pair. Now the more critical task is the approximation of y-coordinate values for straight-line pairs. One way of calculating y-coordinates for each straight-line pair is to find the average y-coordinate values of top and bottom curled lines within a pair. But in document image some textlines are small and some are large and large textlines contain more curled information than smaller ones. Due to this fact, estimated y-coordinates of small textlines are not accurate and results in overlapping of textlines in dewarped image. Therefore we use neighboring textlines information for estimating y-coordinate values for straight-line pairs. For each textline, top and bottom curled lines from neighboring textlines are projected over the top curled line and bottom curled line of targeted textline respectively. Then, all these top curled lines are combined together and bottom curled lines are combined together, which results in an approximated curled-line pair for targeted textline. Approximated curled-textline pairs contain more curled information than actual curled-line pairs, especially for small textlines. We use these approximated pairs only for calculating y-coordinates for straight-line pairs. For each textline, the y-coordinates for straight-line pair are calculated from the approximated curled-textline pair through averaging y-coordinates of its top curled line and bottom curled line. After estimating straight-line pairs, all pixels over curled-line pairs are mapped to the corresponding straight-line pairs. Resulting dewarped image is shown in Figure 1(f), in which textlines are straight as compared to curled textlines of input image (Figure 1(a)).

2.3 Handling of Perspective Distortions

After handling geometric (curled) distortion gracefully, the next step is to remove perspective distortion in the image, as shown in Figure 1(f). For handling perspective distortion we use four point homography algorithm [34], in which homography matrix is calculated from source and target quadrilaterals. In our case, quadrilateral without perspective distortion is used as source and quadrilateral with perspective distortion is used as target. The process of removing perspective distortion starts by finding left and right vertical borders of warped image. Left and right border are calculated by applying RANSAC on staring and ending points of curled-line pairs, detected in section 2.1. Resulting borders are shown in Figure 1(d). For perspective distortion free rectangle, left border perpendicular to page width is calculated by finding minimum x-coordinate value from left border shown in Figure 1(d). Similarly right border perpendicular to page width, is calculated by finding maximum x-coordinate value of right border shown in Figure 1(d). Resulting source and target quadrilaterals are shown in Figures 1(f) and 1(g), in blue and red colors respectively. Rectifying homography matrix is calculated by using source and target quadrilateral. Then, x-and y- coordinates of source quadrilateral are transformed into target quadrilateral using homography matrix and bilinear-interpolation is applied for calculating intensity values for dewarped image. Final de-



(a) (b)

Figure 2. a) Two column document image. b) Dewarped documenrt image result.

warped image is represented in Figures 1(g) and 1(h). One can compare the good quality of final dewarped image (Figure 1(h)) with input warped image (Figure 1(a)). Our algorithm also gives good dewarping results for two column document images, shown in Figure 2.

3 Experiments and Results

To demonstrate the performance of our algorithm on real world documents, we evaluate it on the dataset of CBDAR document image dewarping contest [4]. The dataset consists of 102 documents, captured with hand-held camera. This dataset is freely available with ASCII text ground-truth. Three methods participated in this contest: SEG [17], SKEL [10] and CTM (un-cleaned results) and CTM2 (cleaned up results) [26]. We referred our dewarping method as "Ridges-Snakes". Together with our dewarped results, we also have dewarped results of all three participants of dewarping contest [4]. The results of all methods on some example documents from dataset are shown in Figure 3.

The dewarped documents of all methods are processed through a commercial OCR system **ABBYY Fine Reader 9.0**. After obtaining text from the OCR software, the block edit distance[1] with the ASCII ground-truth has been used as the error measure. Table 1 shows the comparative results of all methods with respect to mean edit distance, median edit distance and the number of documents for each algorithm on which it has the lowest edit distance (in case of tie, all algorithms having the lowest edit distance are scored for that document).

Together with comparative performance evaluation, we have also compared mean edit distance error rate before and after dewarping. Before dewarping, mean edit distance is 5.153%. After dewarping, using our described method, mean edit distance is 1.917%, as mentioned in Table 1. This demonstrates that, after dewarping average edit distance error rate is reduced by 3.24%.

[1] http://sites.google.com/site/ocropus/release-notes

4 Discussion

We have presented a new approach for document image dewarping using curled textlines information. Unlike some other dewarping approaches, our dewarping approach does not use any type of postprocessing step for cleaning up resulting dewarped documents. We calculated edit distance error rate using our raw dewarped results. After applying our reported dewarping method on the dataset of CBDAR 2007 document image dewarping contest [4], OCR error rate is reduced by 3.24%. Additionally, a fair comparison of our algorithm with other three participants of dewarping contest [4] has been done. The winning method [26] of that contest had submitted two different results: (i) dewarped results without postprocessing (CTM) and (ii) dewarped results with postprocessing for removing graphics and images (CTM2). According to the statistics presented in the Table 1, our raw (un-cleaned) dewarped results are nearly similar to the cleaned dewarped results of winning method of dewarping contest, i.e. CTM2, but our raw (uncleaned) dewarped results are better then the un-cleaned dewarped results of winning method of dewarping contest, i.e. CTM. Furthermore as shown in Figure 3, our dewarped results look more planar than other three methods and our dewarping method performs better than other approaches in the presence of margin-notes, two-column documents and high degrees of perspective distortions.

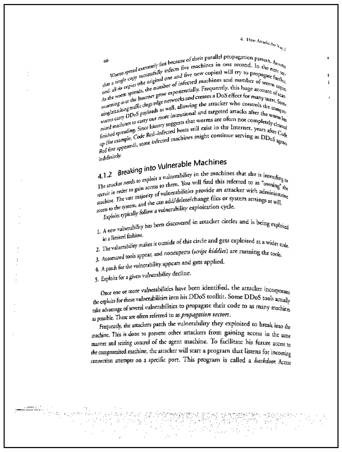
References

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

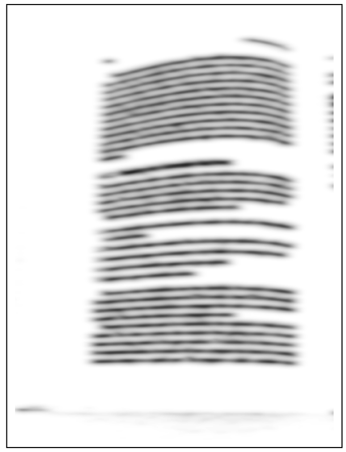
[2] S. Mori, C.Y. Suen, and K. Yamamoto. Historical review of OCR research and development. *Proceedings of the IEEE*, 80(7):1029–1058, 1992.

[3] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey. *International Journal of Document Analysis and Recognition*, 7(2-3):84–104, 2005.

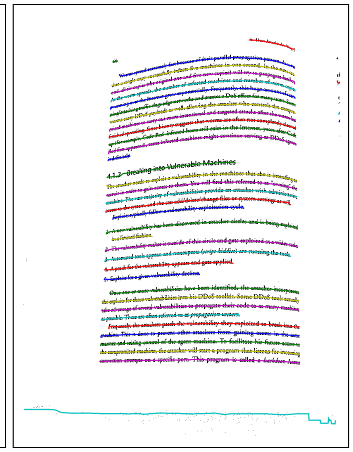
[4] F. Shafait and T. M. Breuel. Document image dewarping contest. In *Proceedings 2nd International*



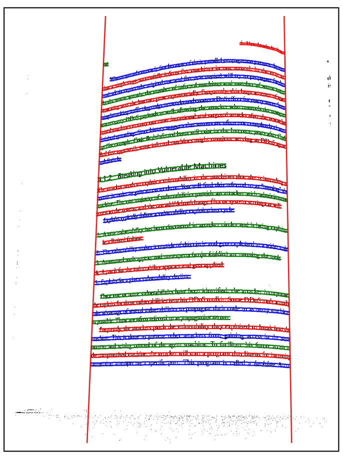
(a) Input Image.



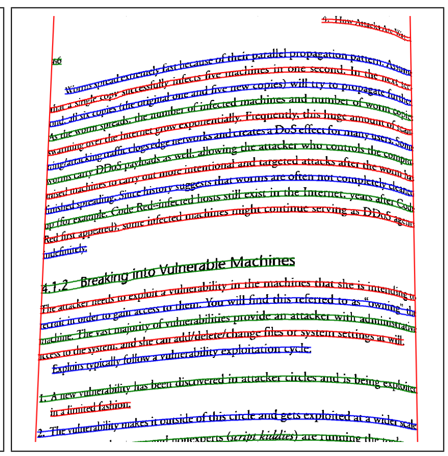
(b) Smoothed Image generated by using mutli-oriented anisotropic Gaussian smoothing.



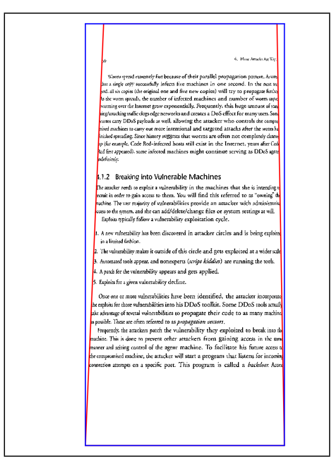
(c) Ridges have been detected from smoothed image using Horn-Riley method [30, 31].



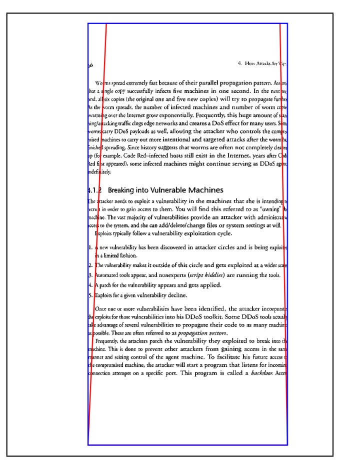
(d) Snake-pairs have been estimated using ridges based coupled-snakes model. Left and right vertical borders have been calculated using RANSAC on starting and ending points of each pair.



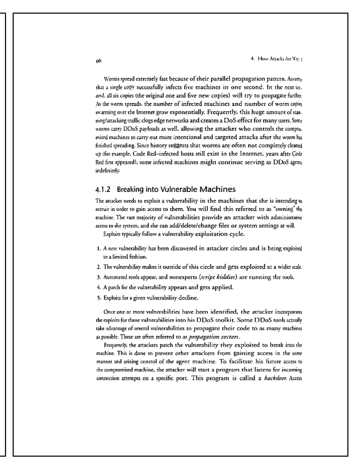
(e) Closeup portion of 1(d).



(f) Geometric distortion has been handled by straightening curled textlines snake-pairs, shown in Figure 1(d). Red borders show that document image still have perspective distortion. Blue rectangle has been calculated using red borders.

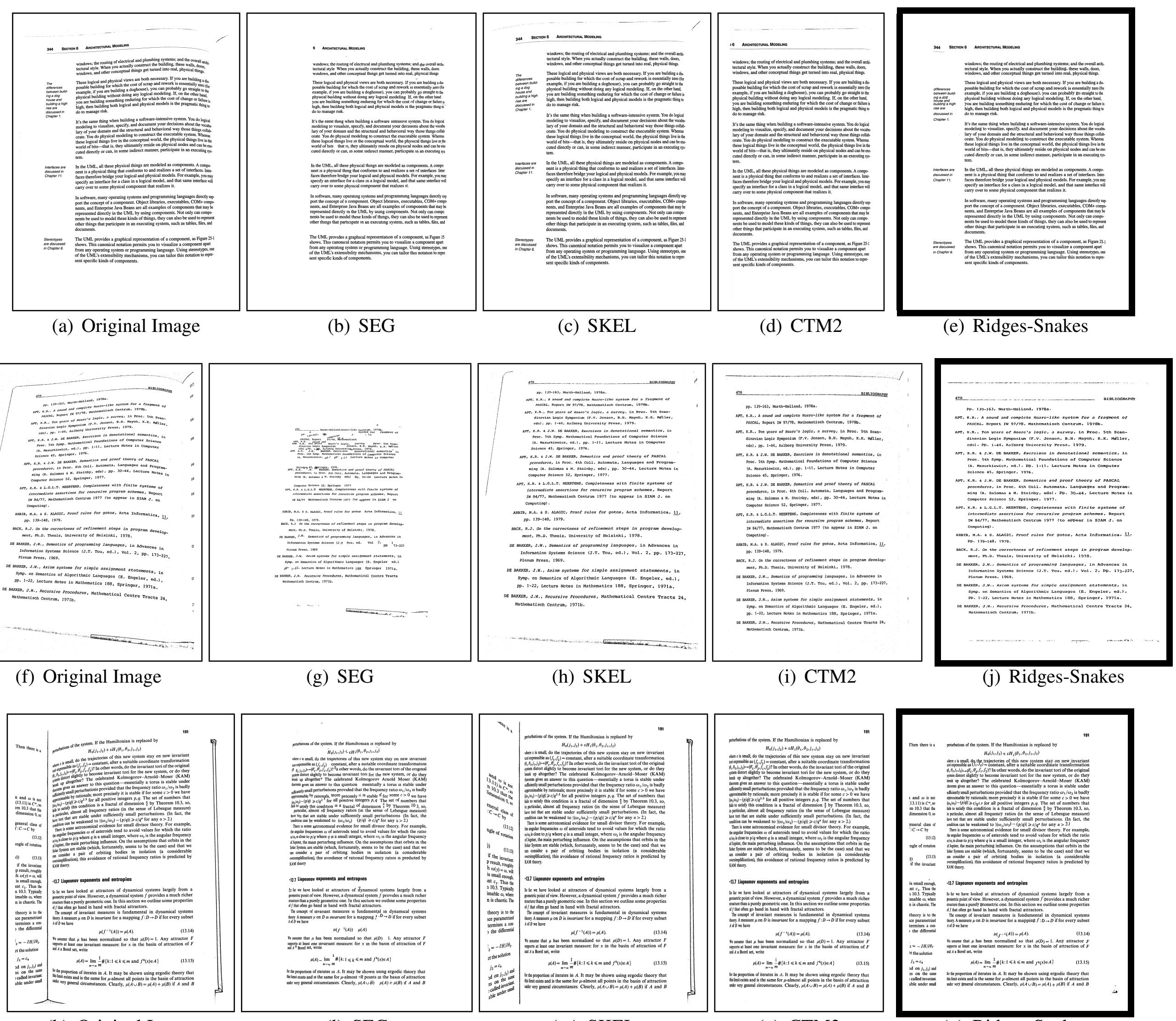


(g) Perspective distortion has been handled by using four point homography algorithm.



(h) Final dewarped Image.

Figure 1. Different stages of dewarping algorithm.



(a) Original Image (b) SEG (c) SKEL (d) CTM2 (e) Ridges-Snakes

(f) Original Image (g) SEG (h) SKEL (i) CTM2 (j) Ridges-Snakes

(k) Original Image (l) SEG (m) SKEL (n) CTM2 (o) Ridges-Snakes

Figure 3. Example results of different methods (SEG [17], SKEL [10], CTM [26] and Ridges-Snakes: For image 3(a) SEG and CTM2 methods removed text-note, among all SKEL and Ridges-Snakes have done proper dewarping. For image 3(f) SEG method failed to remove geometric and perspective distortions, SKEL method removed only geometric distortion; among all CTM2 and Ridges-Snakes have done proper dewarping by removing both geometric and perspective distortions.

Table 1. Comparative OCR error rate (block edit distance) results on the dataset of CBDAR 2007 Document Image Dewarping Contest based on ABBYY Fine Reader 9.0.

Algorithm	Mean Edit Distance %	Median Edit Distance %	Number of documents[a]
SEG [17]	4.088	2.122	02
SKEL [10]	2.162	0.972	29
CTM [26]	2.113	0.893	30
CTM2 [26]	**1.758**	0.827	38
Ridges-Snakes	1.917	**0.733**	**41**

[a]Number of documents for each algorithm on which it has the lowest edit distance.

Workshop on Camera Based Document Analysis and Recognition, pages 181–188, Curitiba, Brazil, 2007.

[5] H. Cao, X. Ding, and C. Liu. Rectifying the bound document image captured by the camera: a model based approach. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, pages 71–75, Edinburgh, Scotland, 2003.

[6] M. S. Brown and W. B. Seales. Image restoration of arbitrarily warped documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1295–1306, 2004.

[7] C. L. Tan, L. Zhang, Z. Zhang, and T. Xia. Restoring warped document images through 3d shape modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):195–208, 2006.

[8] J. Liang, D. DeMenthon, and D. Doermann. Flattening curved documents in images. In *Proceedings 18th International Conference on Computer Vision and Pattern Recognition*, pages 338–345, San Diego, CA, USA, 2005.

[9] S. Lu and C.L. Tan. Document flattening through grid modeling and regularization. In *Proceedings 18th International Conference on Pattern Recognition*, pages 971–974, 2006.

[10] A. Masalovitch and L. Mestetskiy. Usage of continuous skeletal image representation for document images de-warping. In *Proceedings 2nd International Workshop on Camera-Based Document Analysis and Recognition*, pages 45–52, Curitiba, Brazil, 2007.

[11] P. Clark and M. Mirmehdi. Rectifying perspective views of text in 3d scenes using vanishing points. *Pattern Recognition*, 36(11):2673–2686, 2003.

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

[13] A. Ulges, C. H. Lampert, and T. M. Breuel. Document image dewarping using robust estimation of curled text lines. In *Proceedings 8th International Conference on Document Analysis and Recognition*, pages 1001–1005, Seoul, Korea, 2005.

[14] 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.

[15] S. J. Lu and C. L. Tan. Camera document restoration for ocr. In *Proceedings of First International Workshop on Camera-Based Document Analysis and Recognition*, pages 17–24, Seoul, Korea, 2005.

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

[17] B. Gatos, I. Pratikakis, and K. Ntirogiannis. Segmentation based recovery of arbitrarily warped document images. In *Proceedings 9th International Conference on Document Analysis and Recognition*, pages 989–993, Curitiba, Brazi, 2007.

[18] B. Gatos and K. Ntirogiannis. Restoration of arbitrarily warped document images based on text line and word detection. In *Proceedings 4th IASTED International Conference on Signal Processing, Pattern*

Recognition, and Applications, pages 203–208, Innsbruck, Austria, 2007.

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

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

[21] S. S. Bukhari, F. Shafait, and T. M. Breuel. Coupled snakelet model for curled textline segmentation of camera-captured document images. In *Proceedings 10th International Conference on Document Analysis and Recognition*, Barcelona, Spain, 2009.

[22] T. M. Breuel. Robust least square baseline finding using a branch and bound algorithm. In *Proceedings 9th International Conference on Document Recognition and Retrieval*, pages 20–27, San Jose, CA, USA, 2002.

[23] F. M. Wahl, K. Y. Wong, and R. G. Casey. Block segmentation and text extraction in mixed text/image documents. *Computer Graphics and Image Processing*, 20:375–390, 2006.

[24] Z. Zhang and C. L. Tan. Recovery of distorted document images from bound volumes. In *Proceedings nternational Conference on Document Analysis and Recognition*, pages 429–433, Seattle, WA, USA, 2001.

[25] C. Strouthopoulos, N. Papamarkos, and C. Chamzas. Identification of text-only areas in mixed-type documents. *Engineering Applications of Artificial Intelligence*, 10(4):387–401, 1997.

[26] B. Fu, M. Wu, R. Li, W. Li, and Z. Xu. A model-based book dewarping method using text line detection. In *Proceedings 2nd International Workshop on Camera Based Document Analysis and Recognition*, pages 63–70, Curitiba, Barazil, 2007.

[27] C. T. Hsieh, E. Lai, and Y. C. Wang. An effective algorithm for fingerprint image enhancement based on wavelet transform. *Pattern Recognition*, 36(2):302–312, 2003.

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

[29] L. O. Gorman. Matched filter design for fingerprint image enhancement. In *Proceedings International Conference on Acoustics, Speech, and Signal Processing*, pages 916–919, New York, NY, USA, 1988.

[30] 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.

[31] M. D. Riley. Time-frequency representation for speech signals. *PhD Thesis, MIT*, 1987.

[32] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. In *IEEE Transaction of Image Processing*, volume 7, pages 359–369, 1998.

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

[34] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.