2017 14th IAPR International Conference on Document Analysis and Recognition

Robust Document Image Dewarping Method using Text-lines and Line Segments

Taeho Kil*, Wonkyo Seo*, Hyung Il Koo†, Nam Ik Cho*
*Department of Electrical and Computer Engineering and INMC, Seoul National University
†Department of Electrical and Computer Engineering, Ajou University

***Abstract*—Conventional text-line based document dewarping methods have problems when handling complex layout and/or very few text-lines. When there are few aligned text-lines in the image, this usually means that photos, graphics and/or tables take large portion of the input instead. Hence, for the robust document dewarping, we propose to use line segments in the image in addition to the aligned text-lines. Based on the assumption and observation that many of the line segments in the image are horizontally or vertically aligned in the well-rectified images, we encode this property into the cost function in addition to the text-line alignment cost. By minimizing the function, we can obtain transformation parameters for camera pose, page curve, etc., which are used for document rectification. Considering that there are many outliers in line segment directions and missed text-lines in some cases, the overall algorithm is designed in an iterative manner. At each step, we remove text components and line segments that are not well aligned, and then minimize the cost function with the updated information. Experimental results show that the proposed method is robust to the variety of page layouts.**

I. Introduction

Document image processing such as layout analysis and optical character recognition (OCR) is an important step to the document understanding, and numerous methods have been proposed for the scanned document image processing (printed documents are converted to digital images with flatbed scanners and document image processing algorithms are applied) [16], [17], [20], [23]. However, with the recent development of smartphones having high-resolution digital cameras, document image processing algorithms are required to handle camera-captured images as well as scanned documents [5], [9].

Compared with the scanned image processing, the processing of camera-captured images is considered a challenging task due to the geometric distortions caused by camera view and page curve. Although depth-measuring hardwares allow us to remove a range of geometric distortions [2], this approach is not applicable to common users. Therefore, the development of easy-to-use rectification methods has received lots of attentions for the decades. For instance, many text-line based methods were proposed for the single-image-based rectification. However, they focus on text-lines (and text-blocks) and have limitations in handling complex layout and/or very few text-lines. In this paper, in order to alleviate these limitations, we present a robust dewarping method that works for a range of documents (non conventional layouts and very few text-lines) by considering line segments as well as text-lines.

A. Dewarping Methods using Additional Information

For the document image rectification, many methods were developed by using depth measuring hardwares (e.g., structured light or laser scanners) [2], [15], [18], [28]. This approach is able to estimate the surfaces of curved pages very effectively, however, the requirements of special hardwares limit their application areas. In [10], [24], [26], curved pages were estimated from multiple images taken from different viewpoints. Although they could perform rectification without additional hardwares, taking multiple images are burdensome for common users and their computation complexity is also very high. In [4], [27], the shape-from-shading approach exploiting illumination conditions was proposed. Although these methods can be applied to a single document image, their assumptions on illumination may not hold in many situations.

B. Text-line based Dewarping Methods

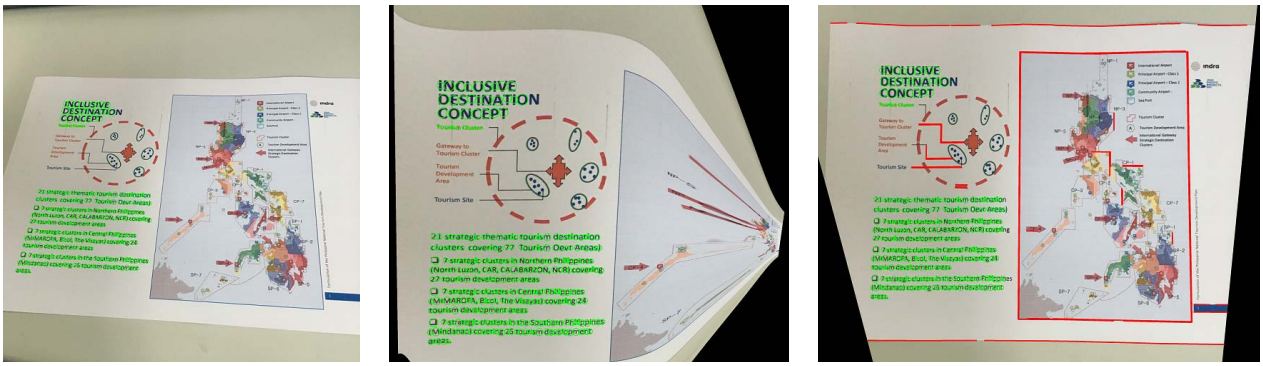
For the single document image rectification (without additional information), numerous methods using text-lines have been proposed. Since text-lines are common and show regular structures in document images, they are considered very useful features in the document rectification.

In [22], two vanishing points were estimated by many horizontal (made by text-lines) and vertical lines (made by line feeds). This approach removes effectively perspective distortions, however is not suitable for geometric distortions by curved surfaces. In most of text-line based methods, curved surfaces are modeled with the generalized cylindrical surface (GCS) [24] and the shapes are estimated from the properties of text-lines. In [6], [21], curved page surfaces were estimated by fitting top and bottom text-lines to flat document regions. In [8], the properties of text-lines (in undistorted documents) were encoded into a cost function, and camera pose and curved page surfaces were estimated by minimizing the function. Although these text-line based methods are able to reduce geometric distortions without the additional information, they focus on text regions and sometimes yield severe distortions on non-text regions (e.g., photos, graphics or tables) as shown in Fig. 1. In summary, text-line based methods exploited regular structures of text-lines and text-blocks, and they basically work for text-abundant cases.

C. Our Approach

In order to alleviate the limitations of text-line based methods, we present a dewarping method that exploits the

2379-2140/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDAR.2017.146



(a) (b) (c)

Fig. 1: Comparison of the proposed method with a conventional text-line based method [8]: (a) Input image, (b) Result of [8]. Distortions on text regions are largely removed, however, new distortions are introduced in other regions. (c) Result of the proposed method. We exploit the properties of (red) line segments as well as (green) text-lines, and we can remove overall distortions.

properties of text and non-text regions: Our approach uses line segments as well as text-lines. Since non-text regions in documents usually have many line segments (e.g., tables and the boundaries of images) that are horizontally or vertically aligned in the well-rectified images, we encode this property into the proposed cost function, as well as the conventional properties of text-lines. The cost function is minimized via the Levenberg-Marquardt algorithm [11], [13] and we can obtain the rectification transformation that removes the distortions in both text and non-text regions as shown in Fig. 1.

Experimental results have shown that the proposed method yields the state-of-the-art rectification performance on text regions (evaluated in terms of character recognition rate) and produces visually pleasing rectification results on non-text regions. Also, we evaluated the rectification performance on non-text regions with several quantitative measures (e.g., the orthogonality).

II. Proposed Method

For the robust dewarping, we propose a new algorithm considering the alignment properties of text-lines and line segments. In this section, we first introduce a parametric model for the rectification transformation and present the proposed cost function reflecting the alignment properties. Then, we discuss the optimization method of the cost function. Although conventional methods were developed by assuming that there are no (significant) outliers in the text-line detection step [8], a few outliers can significantly deteriorate the performance and we deal with outliers in the proposed optimization step.

A. Parametric Model of Dewarping Process

For the parametric modeling of the dewarping process, we adopt the model in [8]. Given a document surface as shown in Fig. 2-(a), a point (α, β) on an image domain corresponds

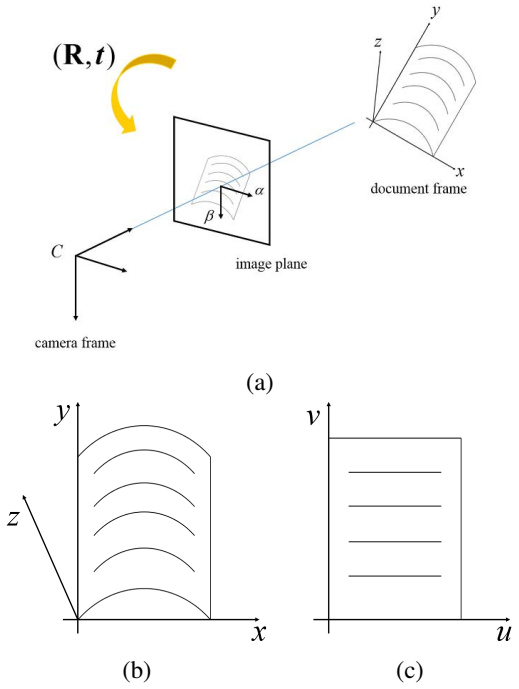


Fig. 2: Dewarping model: (a) A curved document viewed by a camera, (b) Curved document coordinate, (c) Flat document coordinate.

to a point (x, y, z) on the curved document surface with the relation

$$\begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \left(s\mathbf{R}^\top \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \boldsymbol{t} \right), \qquad (1)$$

where f is the focal length of a camera, (c_x, c_y) is an image center and $(s, \mathbf{R}, \boldsymbol{t})$ are scale, rotation and translation between

two frames, respectively. Since the parameters s and $\boldsymbol{t}$ are related to not rectification but image scale and resolution, we set $s = 1$ and $\boldsymbol{t} = [0, 0, f]^\top$ in rectification process.

For a GCS model as shown in Fig. 2-(b) and (c), a point on a curved surface can be transformed to a corresponding point (u, v) on the rectified document with

$$\begin{aligned} u &= \int_0^x \sqrt{1 + g'(t)^2}\, dt, \\ v &= y, \end{aligned} \tag{2}$$

where $g(x)$ is the document surface equation that is represented with polynomial

$$z = g(x) = \sum_{m=0}^{M} a_m x^m, \tag{3}$$

where M is a polynomial order. A small M could not describe the actual page curl well enough, and a large M is prone to over-fitting. Based on extensive experiments, we have found that fourth ($M = 4$) polynomial equation is a good choice to represent page curl.

By combining (1), (2), and (3), points on the image domain can be transformed to the corresponding points on rectified document images.

In summary, the geometric relation between the captured image domain and the rectified document domain can be parameterized with polynomial parameters $\{a_m\}_{m=0}^{M}$ and camera pose $\mathbf{R}$ (we assume that camera focal length f is known). Therefore, the document image rectification process can be formulated as an estimation problem of polynomial parameters and camera pose.

B. Proposed Cost Function

For the estimation of the dewarping parameters $\Theta = (\mathbf{R}, \{a_m\}_{m=0}^{M})$, we develop a cost function:

$$f_{cost}(\Theta) = f_{text}(\Theta) + \lambda f_{line}(\Theta), \tag{4}$$

where $f_{text}(\Theta)$ is a term reflecting the properties of text-lines in rectified images [8]. To be precise, $f_{text}(\Theta)$ becomes small when transformed text-lines are well-aligned: horizontally straight, line-spacings between two neighboring text-lines are regular, and text-blocks are either left-aligned, right-aligned, or justified. However, the optimization of $f_{text}(\Theta)$ sometimes yields severe distortions on non-text regions (as shown in Fig. 1-(b)), and we also exploit line segments in document images by introducing $f_{line}(\Theta)$.

C. Line Segment Alignment and Cost Function

For the design of $f_{line}(\Theta)$, we first extract line segments in given images by using Line Segment Detector (LSD) in [25]. Then, based on the observation that the majority of line segments are horizontally or vertically aligned in the rectified images, we define the term as

$$f_{line}(\Theta) = \sum_i \min\left(\cos^2\theta_i, \sin^2\theta_i\right), \tag{5}$$

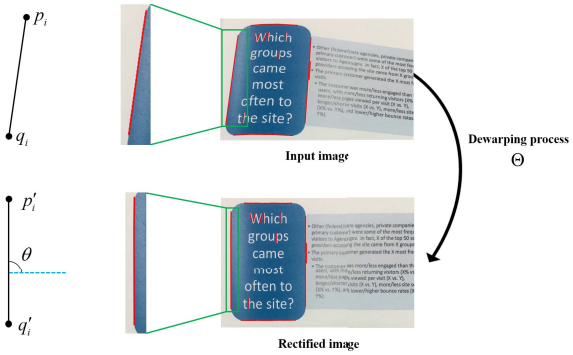
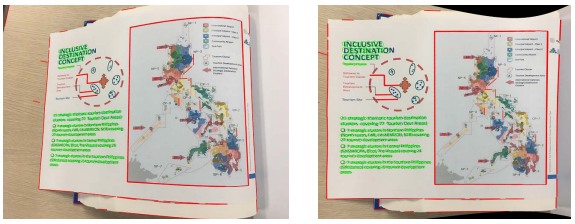


Fig. 3: Illustration of line segment and its angle.



(a) (b)

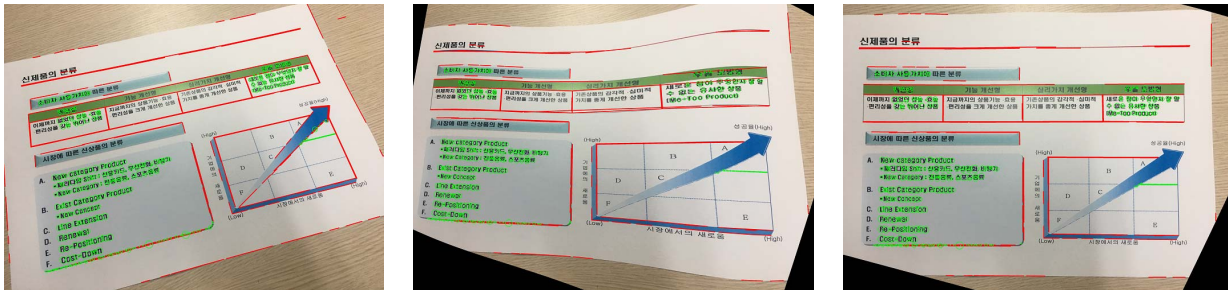
Fig. 4: Text-lines and line segments in camera-captured images and rectified images. Text components are represented as (green) ellipses and line segments are (red) lines. The line segment alignment terms (5) are 16.282 and 0.04, respectively.

where θ_i is the angle of the transformed i-th line segment (when rectified with the current parameters Θ) as illustrated in Fig. 3. Let us denote the end points of the i-th line segment (in the original image) as p_i and q_i, and their transformed points by the dewarping process (using Θ) as p'_i and q'_i. Then θ_i is defined as the orientation of a line segment connecting p'_i and q'_i.

As line segments are aligned in either vertical or horizontal directions, this term becomes small as shown in Fig. 4. Although there are outliers (line segments having arbitrary orientations), (5) minimizes these effects by using bounded penalty functions ($\cos^2\theta_i \leq 1$, $\sin^2\theta_i \leq 1$). Also, we develop an optimization step that alleviates the outlier problem.

D. Outlier Removal and Optimization

Although the optimization method used in [8] assumes that there are no outliers (or their effects are not critical), the direct optimization of $f_{cost}(\Theta)$ may yield poorly rectified results as shown in Fig. 5-(b), due to outliers. We treat two outlier types that are missed text-lines and line segments having arbitrary direction (non horizontal/vertical). For the outlier removal, we design an iterative method. At each step, we refine the features (text components and line segments) by removing outliers and minimize the cost function with updated inliers. To be precise, an updated inlier set of line segments at the $(j+1)$-th iteration



(a) (b) (c)

Fig. 5: Our iterative scheme: (a) An input image, (b) Result after the first iteration, (c) Result after the second iteration.

is defined as

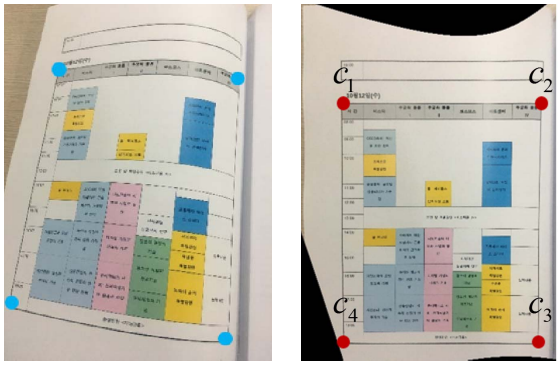
$$L_{j+1} \quad = \quad \{l | l \in L_j, f_{line}(l) < \tau_j\} \tag{6}$$

where L_j is an inlier set at the j-th iteration. In the case of that the transformed line segments (at each step) are not vertical/horizontal aligned, we determine these line segments as outliers. The outlier removal of text is similarly defined. In the case of that the transformed text-lines are not aligned (not horizontal lines), we determine these text components that are not on the horizontal line as outliers. Since the cost terms reflect the alignments of text-lines and line segments, we detect the outliers by computing the cost terms. This iteration is repeated until the number of inliers becomes stable. After extensive experiments, we set the maximum iteration number to 3. The cost function consists of the square terms and Levenberg-Marquardt (LM) algorithm is used for the cost function optimization [11], [13].

III. Experimental Results

For the evaluation, we first conducted experiments on text-abundant images, CBDAR2007 dewarping contest dataset [19]. The CBDAR2007 dewarping contest dataset consists of 102 binarized document images as shown in the first row in Fig. 7. Since the focal length information is not available in this dataset, we test our algorithm on a range of focal length values and select the focal length that minimizes the cost. Since there are no publicly available dataset (to the best of our knowledge) consisting of non conventional document images (i.e., not text-abundant cases), we collected 78 images having various layouts (e.g., three column documents, documents containing large tables and/or figures, presentation slides, and so on) as shown in the second and third rows in Fig. 7.

In the experiments, we set the weight λ in (4) so that it is proportional to $\frac{N_{text}}{N_{line}}$, where N_{text} and N_{line} are the numbers of text components and line segments, respectively. Also, we set the thresholds for outlier removal τ_1, τ_2 and τ_3 to 0.01, 0.005, and 0.001, respectively. We implemented our method with C++ and our implementation takes $3 \sim 8$ (s) for the rectification of an image (4000×3000) in our tested dataset on Intel(R) i5(TM) CPU(3.40GHz). In detail, many of the processing time is spent on the text-line extraction ($2 \sim 5$s) and rendering process ($1 \sim 2$s), the proposed optimization process takes $1 \sim 3$s.



(a) (b)

Fig. 6: Extraction of four corner points for evaluation: (a) The distorted images and manually annotated corner (blue) points, (b) The transformed images and corner (red) points.

We first evaluate our method on text-abundant cases [19] and compare the performance with the conventional methods [3], [6]–[8], [14] in terms of OCR accuracy. To be precise, the accuracy is defined as

$$\text{accuracy}(R, G) = 1 - \frac{L(R, G)}{\max(\#R, \#G)}, \tag{7}$$

where R is a recognition result, G is the ground truth, $\#(\cdot)$ is the number of characters in the string, and $L(x, y)$ means the Levenshtein distance between two strings [12]. The distance is defined as the minimum number of character edits (insertion, deletion, and substitution) to transform one string to the other. For the OCR, we use the Google tesseract OCR engine [20]. Experimental results are summarized in Table. I. Since samples in CBDAR2007 dewarping contest dataset are text-abundant images (having single columns), the conventional text-line based method [8] showed good performance. However, as can be seen, the proposed method can also handle all these cases and shows improved accuracy (probably due to the our optimization method considering outliers).

(a) (b) (c) (d) (e) (f)

Fig. 7: Experimental results on CBDAR2007 dewarping contest and our datasets: (a), (d) Distorted input images and text components (green ellipses), (b), (e) Rectified images by the text-line based method [8], (c), (f) Rectified images by the proposed method and (red) vertical/horizontal aligned line segments.

TABLE I: OCR performance on CBDAR2007 dewarping contest dataset

	Original	SEG [7]	SKEL [14]	CTM [6]	Snakes [3]	Kim [8]	Proposed
OCR accuracy	62.47	89.47	93.17	96.22	96.47	97.42	**97.54**

TABLE II: Geometric measures of the proposed and conventional methods on our dataset

	Original	Kim [8]	Proposed
Orthogonality	21.4819	10.1254	**2.4284**
Diagonal ratio	0.0586	0.0375	**0.0096**
Vertical ratio	0.1427	0.0974	**0.0341**
Horizontal ratio	0.1245	0.0879	**0.0274**

For the evaluation on non conventional cases, we also conducted experiments on our dataset (78 images). Since we want to evaluate the rectification performance on non-text regions, we computed the geometric quantities of rectangles in rectified results. To be precise, we used orthogonality θ_o, diagonal ratio r_d, and length ratios for opposite sides r_h and r_v [1], which are defined as

$$\begin{aligned}
\theta_o &= \cos^{-1}\left(\frac{(\boldsymbol{c_1}-\boldsymbol{c_2})\cdot(\boldsymbol{c_1}-\boldsymbol{c_4})}{d(\boldsymbol{c_1},\boldsymbol{c_2})\times d(\boldsymbol{c_1},\boldsymbol{c_4})}\right),\\
r_d &= \max\left(\frac{d(\boldsymbol{c_1},\boldsymbol{c_3})}{d(\boldsymbol{c_2},\boldsymbol{c_4})},\frac{d(\boldsymbol{c_2},\boldsymbol{c_4})}{d(\boldsymbol{c_1},\boldsymbol{c_3})}\right),\\
r_h &= \max\left(\frac{d(\boldsymbol{c_1},\boldsymbol{c_4})}{d(\boldsymbol{c_2},\boldsymbol{c_3})},\frac{d(\boldsymbol{c_2},\boldsymbol{c_3})}{d(\boldsymbol{c_1},\boldsymbol{c_4})}\right),\\
r_v &= \max\left(\frac{d(\boldsymbol{c_1},\boldsymbol{c_2})}{d(\boldsymbol{c_3},\boldsymbol{c_4})},\frac{d(\boldsymbol{c_3},\boldsymbol{c_4})}{d(\boldsymbol{c_1},\boldsymbol{c_2})}\right)
\end{aligned}$$

where $\boldsymbol{c_i}$ $(i = 1, 2, 3, 4)$ is a corner point of a rectangle as shown in Fig. 6. Since $\theta_o = 90°$ and $r_d = r_h = r_v = 1$ in ideally rectified images, we measure the remaining geometric distortions with $|\theta_o - 90°|$, $|r_d - 1|$, $|r_h - 1|$ and $|r_v - 1|$. In the ideal case, the values of these four measures are 0. Experimental results are summarized in Table. II. Since

the executable of [8] is publicly available, we compare the proposed method with the authors' implementation. As shown, the proposed method shows improved geometric rectification performance in terms of all measures, showing the robustness of the proposed method for text and non-text regions.

Some experimental results are shown in Fig. 7. Unlike the proposed method, the conventional text-line based method has difficulties when there are few aligned text-lines or mis-detected text-lines (false positives). However, since the proposed method exploits line segments and removes outliers, the proposed method works robustly for a variety of inputs.

IV. Conclusion

In this paper, we have proposed a document dewarping method exploiting the properties of line segments and text-lines. We encoded the alignment property of line segments into the proposed cost function so that the method works on both text and non-text regions. Also, we developed an iterative optimization scheme in order to handle outliers. Experimental results showed that the proposed method yields the state of the art OCR performance on text regions and visually pleasing rectified results on non-text regions.

Acknowledgment

This research was supported by Hancom Inc.

References

[1] J. An, H. I. Koo, and N. I. Cho, "Rectification of planar targets using line segments," *Machine Vision and Applications*, vol. 28, no. 1, pp. 91–100, 2017.

[2] M. S. Brown, M. Sun, R. Yang, L. Yun, and W. B. Seales, "Restoring 2d content from distorted documents," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1904–1916, 2007.

[3] S. S. Bukhari, F. Shafait, and T. M. Breuel, "Dewarping of document images using coupled-snakes," in *International Workshop on Camera Based Document Analysis and Recognition*, 2009, pp. 34–41.

[4] F. Courteille, A. Crouzil, J.-D. Durou, and P. Gurdjos, "Shape from shading for the digitization of curved documents," *Machine Vision and Applications*, vol. 18, no. 5, pp. 301–316, 2007.

[5] L.-Y. Duan, R. Ji, Z. Chen, T. Huang, and W. Gao, "Towards mobile document image retrieval for digital library," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 346–359, 2014.

[6] B. Fu, M. Wu, R. Li, W. Li, Z. Xu, and C. Yang, "A model-based book dewarping method using text line detection," in *International Workshop on Camera Based Document Analysis and Recognition*, 2007, pp. 63–70.

[7] B. Gatos, I. Pratikakis, and K. Ntirogiannis, "Segmentation based recovery of arbitrarily warped document images," in *International Conference on Document Analysis and Recognition*, vol. 2, 2007, pp. 989–993.

[8] B. S. Kim, H. I. Koo, and N. I. Cho, "Document dewarping via text-line based optimization," *Pattern Recognition*, vol. 48, no. 11, pp. 3600–3614, 2015.

[9] H. I. Koo, "Text-line detection in camera-captured document images using the state estimation of connected components," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5358–5368, 2016.

[10] H. I. Koo, J. Kim, and N. I. Cho, "Composition of a dewarped and enhanced document image from two view images," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1551–1562, 2009.

[11] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[12] V. Levenstein, "Binary codes capable of correcting spurious insertions and deletions of ones," *Problems of Information Transmission*, vol. 1, no. 1, pp. 8–17, 1965.

[13] D. W. Marquardt, "An algorithm for least-squares estimation of non-linear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[14] A. Masalovitch and L. Mestetskiy, "Usage of continuous skeletal image representation for document images de-warping," in *International Workshop on Camera Based Document Analysis and Recognition*, 2007, pp. 45–53.

[15] G. Meng, Y. Wang, S. Qu, S. Xiang, and C. Pan, "Active flattening of curved document images via two structured beams," in *International Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3890–3897.

[16] G. Nagy, "Twenty years of document image analysis in pami," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 38–62, 2000.

[17] L. O'Gorman, "The document spectrum for page layout analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1162–1173, 1993.

[18] O. Samko, Y.-K. Lai, D. Marshall, and P. L. Rosin, "Virtual unrolling and information recovery from scanned scrolled historical documents," *Pattern Recognition*, vol. 47, no. 1, pp. 248–259, 2014.

[19] F. Shafait and T. M. Breuel, "Document image dewarping contest," in *International Workshop on Camera Based Document Analysis and Recognition*, 2007, pp. 181–188.

[20] R. Smith, "An overview of the tesseract ocr engine," in *International Conference on Document Analysis and Recognition*, vol. 2, 2007, pp. 629–633.

[21] N. Stamatopoulos, B. Gatos, I. Pratikakis, and S. J. Perantonis, "A two-step dewarping of camera document images," in *Internetional Workshop on Document Analysis Systems*, 2008, pp. 209–216.

[22] Y. Takezawa, M. Hasegawa, and S. Tabbone, "Camera-captured document image perspective distortion correction using vanishing point detection based on radon transform," in *International Conference on Pattern Recognition*, 2016, pp. 3968–3974.

[23] T. A. Tran, I. S. Na, and S. H. Kim, "Page segmentation using minimum homogeneity algorithm and adaptive mathematical morphology," *International Journal on Document Analysis and Recognition*, vol. 19, no. 3, pp. 191–209, 2016.

[24] Y.-C. Tsoi and M. S. Brown, "Multi-view document rectification using boundary," in *International Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[25] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD: a line segment detector," *Image Processing On Line*, vol. 2, pp. 35–55, 2012.

[26] S. You, Y. Matsushita, S. Sinha, Y. Bou, and K. Ikeuchi, "Multiview rectification of folded documents," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[27] L. Zhang, A. M. Yip, M. S. Brown, and C. L. Tan, "A unified framework for document restoration using inpainting and shape-from-shading," *Pattern Recognition*, vol. 42, no. 11, pp. 2961–2978, 2009.

[28] L. Zhang, Y. Zhang, and C. Tan, "An improved physically-based method for geometric restoration of distorted document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 728–734, 2008.