Page Curling Correction for Scanned Books Using Local Distortion Information

Vladimir Kluzner, Asaf Tzadok

Document Processing and Management Group

IBM Research - Haifa

Haifa, Israel

{*kvladi, asaf*}*@il.ibm.com*

Abstract—The correction of page curling in scanned document images has attracted a lot of attention in recent years. Fixing page curling is essential because of the resulting damage in the visual perception of the scanned text and the ensuing reduction in OCR performance on the distorted image. It has been generally concluded that correcting the distortion due to page curling will serve as a solid basis for increased OCR accuracy.

We present a novel approach for the efficient correction of page curling in the images of scanned book pages. The approach is based on the fact that approximately 70% of the words in any book are recurring terms. Thus, for many distorted words, a distinct and clear reference word can be found. Our work computes a global, polynomial transformation-based correction for the page distortion. This correction is based on the estimation of various local distortions in the given page, which are characterized by located words.

Experiments on the scanned page images of an 18th century book printed in Old Gothic font have demonstrated the effectiveness of the proposed technique.

Keywords-page curling, de-warping, polynomial transformation, distortion correction, distortion compensation.

I. INTRODUCTION

The non-linear warping of book page images is often observed when they are captured by flatbed scanning. This phenomenon may result from the scanner setup, the condition of the scanned material, environmental conditions such as dust or humidity that cause page shrinking, or trivial page curling at the page corners or near the book binding. Page curling can cause the heavy distortion of text in these images and consequently influence the performance of optical character recognition (OCR) systems, which cannot handle distorted text. It is now recognized that correcting the curling of scanned book pages can serve as a solid basis for increasing OCR accuracy.

In this paper, we present a novel approach for efficiently correcting page curling in scanned book page images. The approach is based on the fact that there is a recurrence of approximately 70% of the words in the book. Thus, for many distorted words, a high quality reference word can be identified. We compute the correction for the global, polynomial transformation-based page distortion, based on

previously estimated local distortions in the given page. These estimations are characterized by distorted words located on the page and the existence of their high-quality reference words.

The structure of this paper is as follows. Section II reviews various correction techniques for distorted document images. Section III briefly describes the pre-processing architecture of the suggested approach and its individual components, including word indexing, determining high-quality reference words, and computing local distortions. Our polynomial global approach for page curling correction using local distortion information is presented in Section IV. Section V considers the alternating minimization (AM) algorithm for the above approach and presents the curled and corrected pages of scanned 18th century book printed in Old Gothic font. Conclusions and future directions are discussed in Section VI.

II. RELATED WORK

Many different approaches for correcting distortion in document images were reviewed by Liang et. al. [1]. One of the popular approaches is 3D page shape reconstruction; this technique requires some kind of specialized hardware, such as laser or stereo-cameras, or prior knowledge concerning the distortion type. On the other hand, the approach for 2D document image processing uses only the page image from the scanner. Our work is related to this 2D category where all information comes from the scanner.

Lavialle et al. [2] introduced a method to straighten text lines, using an active contour network based on an analytical model containing cubic B-splines. However, the initialization must be very close to the desired solution. Wu and Agam [3] use the texture of a document image to infer the document structure distortion. A two-pass image warping algorithm is then used to correct the images. The approach developed by Zhang and Tan [4] uses polynomial regression to model the warped text lines with quadratic reference curves. It can be applied only to grey-level images with shaded curl regions. Tsoi and Brown [5] rely on boundaries to correct the geometric distortion in images of printed materials. This approach requires the use of a physical pattern to guide the uniform parametrization of distorted image. The model fitting technique [6] estimates the warp of each text line

by fitting it with an elastic curve model. The authors note the need for a more accurate line-warping model. Ulges et. al. [7] presented a complete algorithm to remove distortion due to perspective and page curl from images of smooth yet non-planar book pages. Their work assumes that the original document contains parallel text lines of fixed line spacing; however, this assumption is not generic. Liang et al. [8] model the page surface using a developable surface and exploit the properties of the printed textual content on the page to recover the surface shape. This process is quite slow because of the computation time needed. Zhang and Tan [9] use a spline interpolation technique, which is based on a ruled surface model constructed from text lines in the 2D document image, in order to restore arbitrarily warped document images to a planar shape. However, this method relies on the presence of well-structured text lines that are long enough to reflect the warping curvature. The method, proposed by Lu and Tan [10], corrects camera images of documents through image partition, which divides distorted text lines into multiple small patches based on the identified vertical stroke boundary and the fitted x-line and baseline of text lines. This approach fails when the distortion angle between the document normal and camera optical axis is too big. Masalovitch and Mestetskiy [11] propose a technique for approximating the deformation of text in document images based on the continuous skeletal representation of an image; however, the approximation of deformation of the text blocks' vertical borders is not highly accurate. The segmentation based method developed by Gatos et. al. [12] tends to detect text lines for the efficient restoration of arbitrarily warped document images, but more efficient word baseline detection algorithm is desired. Wu et. al. [13] propose a coordinate transform model to rectify both the perspective and surface curving distortions of a book image that is captured by a 2D digital camera; the method proposed suffers from run-time issues. Rosman et. al. [14] present a de-warping method for documents, including the structured forms created by sheet-fed scanners. This method assumes the existence of a straight template form. The two-step approach developed by Stamatopoulos et. al. [15] includes the efficient de-warping of camera document images, which suffer from warping due to distortion and surface deformation, but takes approximately 10 seconds on average to process one page. Eventually, Bukhari et. al. [16] presented a new approach for document image de-warping using curled text line information that has been extracted using a ridges-based modified active contour model. Unlike some other de-warping approaches, this one does not use any type of post-processing step for cleaning up resulting de-warped documents.

III. Overview of Book Pre-Processing

The approach proposed here is relevant for work on large bodies of homogenous material, and is meant for application to text with a large percentage of recurring items. Thus, we found that scanned ancient books are an ideal application for our algorithm.
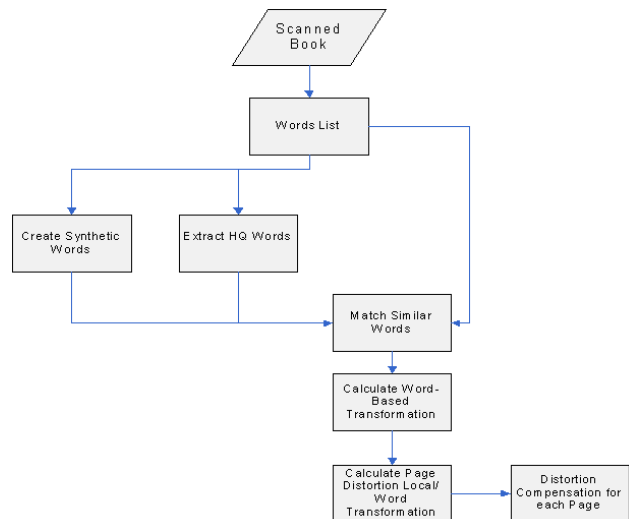

Figure 1. Page curling correction algorithm

The flowchart of the process is presented in Figure 1. As mentioned previously, page curling correction should serve as a basis for better book text recognition, hence, the process starts with image enhancement and layout analysis. Next, the scanned book pages are segmented into the individual word images. Our experiments use word segmentation provided by the ABBYY FineReader engine.

Once individual word images are created, the system proceeds to determine equivalence classes such that each class contains images presumed to show different instances of the same word (see, for example [17]). Then, for every class, the high-quality word (non-distorted one) is chosen. The choice of the high quality reference word can be based on the existence of high-confidence characters in the candidate word. In the worst case scenario, the possibility of synthetic word creation may be discussed.

Naturally, the efficiency of our approach depends on having a large number of word repetitions in the given text. Figure 2 depicts the ratio between the number of distinct words and the total number of words in the data set used. As expected, this ratio decreases exponentially, reaching its limit of 1/7 for texts above 40,000 words. That is, for a standard ancient book with approximately 200 pages (200 words per page), we have recurring items for more than 80% of the page words.

However, since the stages described in this section go beyond the scope of this paper, we do not describe them in detail.
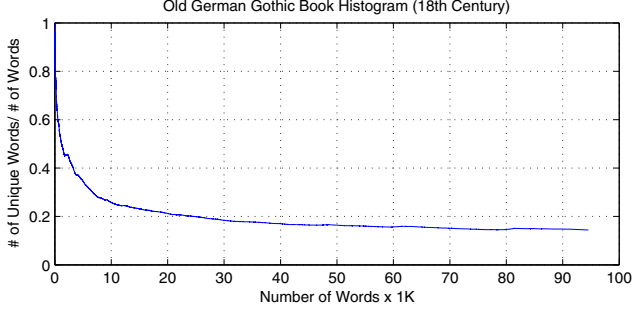
Figure 2. Ratio of unique words as a function of the amount of text

IV. POLYNOMIAL GLOBAL APPROACH FOR CORRECTION OF PAGE DISTORTION

This section presents the core element of the proposed algorithm. Given the curled page I. Let $D_1, \ldots, D_n \subset I$ denote the number of words in the page, with the number of high-quality reference words represented by $S_1, \ldots, S_n \subset \mathbb{R}^2$. The process of locating the reference words is straight-forward, taking into account the existence of word equivalence classes presented in previous section, and is beyond the scope of this paper.

Given n pairs of words $(S_i, D_i)_{i=1}^n$. For every pair i one may easily estimate the desired elastic registration function $T_i(x, y) : \mathbb{R}^2 \to \mathbb{R}^2$, mapping S_i to D_i (see, for example, [17]). That is, for a set of m_i source points $\{s_{ij}^{loc}\}_{j=1}^{m_i} \in S_i$ and a set of m_i destination points $\{d_{ij}^{loc}\}_{j=1}^{m_i} \in D_i$, s.t. $T_i(s_{ij}^{loc}) = d_{ij}^{loc}$ for every $j = 1, \ldots, m_i$, T_i is a solution for the following minimization problem:

$$T_i = \arg\min_{Tr} \sum_{j=1}^{m_i} \|D_i(Tr(s_{ij}^{loc})) - S_i(s_{ij}^{loc})\|^2.$$

Here s_{ij}^{loc} and d_{ij}^{loc}, $j = 1, \ldots, m_i$ denote the local point coordinates with respect to word images S_i and D_i, $i = 1, \ldots, n$. Also, one can consider the mapping T_i as the characteristics of local distortion of page I in the region of the word's D_i location.

At this stage, we have the information concerning the local distortions in various regions of the given page. In order to consider the global page distortion compensation, let us denote by $UL_i, \ldots, UL_n$ the upper-left corner coordinates of word images $D_1, \ldots, D_n$, respectively. Denote also by $\{s_{ij}\}$ and $\{d_{ij}\}$, $j = 1, \ldots, m_i$, $i = 1, \ldots, n$ the global (page) coordinates of points $\{s_{ij}^{loc}\}$ and $\{d_{ij}\}^{loc}$, $j = 1, \ldots, m_i$, $i = 1, \ldots, n$, respectively. Then:

$$\begin{cases} s_{ij} = UL_i + s_{ij}^{loc} \\ d_{ij} = UL_i + d_{ij}^{loc} \end{cases}, \quad j = 1, \ldots, m_i, \ i = 1, \ldots, n.$$

Our next step is to find the unified (global) transformation $T(x, y) : \mathbb{R}^2 \to I$, which should be a solution for the following minimization problem:

$$T = \arg\min_{Tr} \sum_{i=1}^n \sum_{j=1}^{m_i} \|Tr(s_{ij}) - d_{ij}\|^2. \tag{1}$$

That is, if the above minimization function is equal to zero, the inverse of T should transfer distorted words $D_1, \ldots, D_n$ to the reference words $S_1, \ldots, S_n$, respectively.

The polynomial interpolation type was chosen to represent the transformation T due to ability of polynomial functions to approximate every function up to a desired level of accuracy. Polynomial interpolation has long been used for image registration (see [18]). Hence, we tend to find the coefficients a_{uvx}, a_{uvy} of the polynomial functions $T_x(x, y)$, $T_y(x, y)$

$$T(x, y) = (T_x(x, y), T_y(x, y)) =$$
$$= \left(\sum_{k=0}^N \sum_{u+v=k} a_{uvx} x^u y^v, \sum_{k=0}^N \sum_{u+v=k} a_{uvy} x^u y^v \right),$$

where N is the polynomial degree, such that (1) takes place. Thus, denoting $s_{ij} = (s_{ijx}, s_{ijy})$, $d_{ij} = (d_{ijx}, d_{ijy})$ for every $j = 1, \ldots, m_i$, $i = 1, \ldots, n$, we define the following functional:

$$F(a_{uvx}, a_{uvy}) =$$
$$F(a_{00x}, \ldots, a_{N0x}, \ldots, a_{0Nx}, a_{00y}, \ldots, a_{N0y}, \ldots, a_{0Ny})$$
$$= \sum_{i=1}^n \sum_{j=1}^{m_i} \left(\left\| \sum_{k=0}^N \sum_{u+v=k} a_{uvx} s_{ijx}^u s_{ijy}^v - d_{ijx} \right\|^2 + \right.$$
$$\left. + \left\| \sum_{k=0}^N \sum_{u+v=k} a_{uvy} s_{ijx}^u s_{ijy}^v - d_{ijy} \right\|^2 \right).$$

The aforementioned function is convex, thus there exists the unique minimum, which is the desired solution for our problem.

However, the source points s_{ij}, $j = 1, \ldots, m_i$, $i = 1, \ldots, n$, corresponding to the destination points d_{ij}, $j = 1, \ldots, m_i$, $i = 1, \ldots, n$ are not exactly in the right place. Their current location was calculated for a local mapping between the reference word images $S_1, \ldots, S_n$ and distorted word images $D_1, \ldots, D_n$. Thus, it should be assumed, that for reference word images $S_1, \ldots, S_n$ there exist the unknown displacement vectors $\delta_1, \ldots, \delta_n$ such that denoting $\delta_i = (\delta_{ix}, \delta_{iy})$, $i = 1, \ldots, n$, the proper minimization functional for the page curling problem is:

$$F(a_{uvx}, a_{uvy}, \delta_{1x}, \ldots, \delta_{nx}, \delta_{1y}, \ldots, \delta_{ny}) = \sum_{i=1}^n \sum_{j=1}^{m_i} \tag{2}$$
$$\left(\left\| \sum_{k=0}^N \sum_{u+v=k} a_{uvx} (s_{ijx} + \delta_{ix})^u (s_{ijy} + \delta_{iy})^v - d_{ijx} \right\|^2 + \right.$$
$$\left. + \left\| \sum_{k=0}^N \sum_{u+v=k} a_{uvy} (s_{ijx} + \delta_{ix})^u (s_{ijy} + \delta_{iy})^v - d_{ijy} \right\|^2 \right).$$

Hence, the number of minimization parameters depends on the polynomial degree N and the number of distorted words n and equals to $(N+1)(N+2)+2n$.

V. IMPLEMENTATION CONSIDERATIONS AND RESULTS

It is useful to note, that for given $\delta_1, \ldots, \delta_n$ the functional (2) is convex with respect to a_{uvx}, a_{uvy}, $u+v=k$, $k = 0, \ldots, N$. Also, for a given a_{uvx}, a_{uvy} the functional (2) is definitely convex with respect to $\delta_1, \ldots, \delta_n$ only for $N=1$. However, we remark that (2) is not jointly convex in general. Therefore, with an initial guess $(0,0)$ for $\delta_1, \ldots, \delta_n$, we can minimize (2) with respect to a_{uvx}, a_{uvy} and then, taking the solution as an initial guess for a_{uvx}, a_{uvy}, $u+v=k$, $k = 0, \ldots, N$, minimize (2) with respect to $\delta_1, \ldots, \delta_n$, and so on. Hence, we develop an alternating minimization (AM) algorithm in which the function value always decreases as the iteration number increases. More precisely, the algorithm is stated as follows:

- Start with $\delta_i^{(0)} = (0,0)$, $i = 1, \ldots, n$. Assume we have $\delta_i^{(m)}$, $i = 1, \ldots, n$.
- Solve
$$\left(a_{uvx}^{(m)}, a_{uvy}^{(m)}\right) = \arg\min_{a_{uvx}, a_{uvy}} F(a_{uvx}, a_{uvy}, \delta_1^{(m)}, \ldots, \delta_n^{(m)}).$$
- Solve
$$\left\{\delta_i^{(m+1)}\right\}_{i=1}^n = \arg\min_{\delta_1, \ldots, \delta_n} F(a_{uvx}^{(m)}, a_{uvy}^{(m)}, \delta_1, \ldots, \delta_n).$$
- Iterate on m until convergence criterion takes place.

The first minimization problem eventually converges to the solution of the linear equations system $\mathbf{Vc} = \mathbf{f}$, where $\mathbf{c}$ is the vector of polynomial coefficients we tend to find, $\mathbf{f}$ is vector of destination points d_i, and $\mathbf{V}$ is the Vandermonde matrix with rows $(\mathbf{V})_i$, consisting of the elements $x_i^u y_i^v$, $u+v = k$, $k = 1, \ldots, N$, where x_i, y_i denote the x and y coordinates of the source point s_i and N is the degree of the polynomial.

The second minimization problem implies several solutions in case the polynomial degree $N > 1$. In order to simplify the optimization process we calculated the word displacement values $\left\{\delta_i^{(m+1)}\right\}_{i=1}^n$ in the following way:

$$\delta_i = \frac{\sum_{j=1}^{m_i} (d_{ij} - T(s_{ij}))}{m_i},$$

which is identical to assuming that for this case the polynomial degree N equals to 1.

The proposed algorithm was tested on the scanned pages of a book printed in 18th century Old Gothic font (see its page on the Figure 3 on the left). According to the parabolic type of observed distortion, the polynomial degree N was chosen to be 2. However, we observed that it has appeared impossible to approximate the existing distortion through the x axis by one parabolic function. Therefore, in this particular case, we suggest dividing the entire page along its y axis in the so-called vertical division point c

(see red line on the Figure 3), such that it will be possible to find the ideal parabolic transformations, $T^L(x,y)$ and $T^R(x,y)$, approximating the left and right sides of the page, respectively:

$$T(x,y) = \left\{ \begin{array}{ll} T^L(x,y), & x < c \\ T^R(x,y), & x > c \end{array} \right. .$$

In order to guarantee a visual smoothness of this piecewise transformation, the continuity and the x axis smoothness of a first order for the transformation $T(x,y)$ are required:

$$\left\{ \begin{array}{l} \left(T_x^L(c,y), T_y^L(c,y)\right) = \left(T_x^R(c,y), T_y^R(c,y)\right) \\ \left(\dfrac{\partial T_x^L(c,y)}{\partial x}, \dfrac{\partial T_y^L(c,y)}{\partial x}\right) = \left(\dfrac{\partial T_x^R(c,y)}{\partial x}, \dfrac{\partial T_y^R(c,y)}{\partial x}\right) \end{array} \right. .$$

Assuming initially (without loss of generality), that the vertical division point c equals half the page's width W, we can calculate it for every iteration m as a function of the differences between the destination points d_i and source point transformations $T(s_i)$:

$$c^{(m+1)} = \frac{W * R}{(L + R)},$$

where L and R are the averages of the differences' absolute values $|d_i - T(s_i)|$ for the cases $s_{ix} < c^{(m)}$ and $s_{ix} > c^{(m)}$, respectively.

On Figure 3 one can see the curled page from the book mentioned above and its corrected one. There are 190 words on the page. We used 50 reference words instead of the 130 that could be found according to statistics. The polynomial degree was taken to be 2. It took 5 iterations to reach the presented result. The red vertical line passes through the vertical division point c.
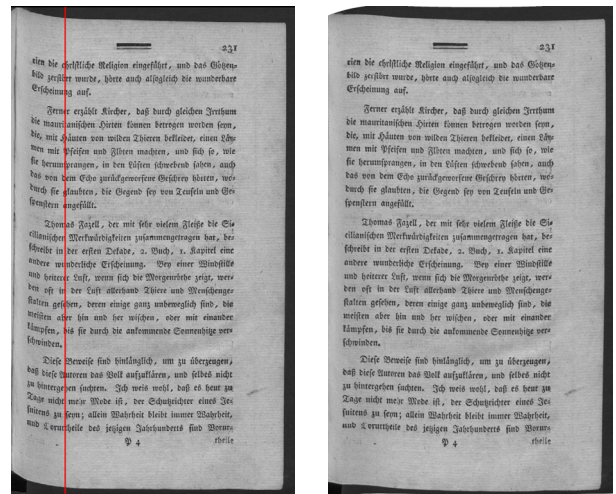

Figure 3. Original page (on the left) and corrected page (on the right)

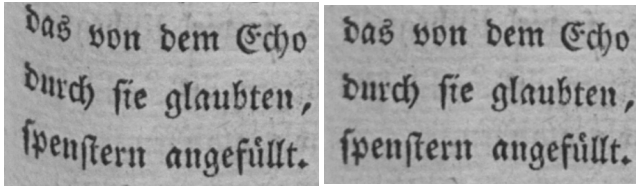Figure 4 shows a section of the distorted page and its dewarping, which were cut from their reference images above.

Figure 4. Curled part of page (on the left) and its de-warping (on the right)

VI. Conclusions and Future Work

Our work implements an algorithm for correcting page distortion due to page curling. We start by identifying a set of high quality (reference) words in undistorted regions of pages, which have content related to the content of the target page. We then generate a global polynomial-based transformation function for the distorted page image, such that its inverse can be applied to the target page image to transform every distorted word into its corresponding high quality word.

In our experiments, we disregarded the issue of word indexing and locating the high quality words. These issues were widely explored previously (e.g., [17] for word indexing) and are beyond the scope of this paper.

This work focuses on the correction of page curling in ancient books. However, the developed algorithm does not imply any special properties of a scanned book. Thus, this approach should also work successfully for modern books.

Future research could concentrate on the evaluation of the algorithm's performance on a benchmark of multiple book pages, including the estimation of OCR accuracy before and after algorithm implementation.

References

[1] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," *Int.Journal on Document Analysis and Recognition*, vol. 7, no. 2-3, pp. 84–104, 2005.

[2] O. Lavaille, X. Molines, F. Angella, and P. Baylou, "Active contours network to straighten distorted text lines," in *Proc. of Int. Conf. on Image Processing*, Thessaloniki, Greece, 2001, pp. 1074–1077.

[3] C. Wu and G. Agam, "Document image dewarping for text/graphics recognition," in *Proc. of Joint IAPR Int. Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, Windsor, Canada, 2002, pp. 348–357.

[4] Z. Zhang and C. L. Tan, "Correcting document image warping based on regression of curved text lines," in *Proc. of 7th Int. Conf. on Document Analysis and Recognition*, Edinburgh, Scotland, 2003, pp. 589–593.

[5] Y. C. Tsoi and M. S. Brown, "Geometric and shading correction for images of printed materials - a unified approach using boundary," in *Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, Washington, DC, 2004, pp. 240–246.

[6] H. Ezaki, S. Uchida, A. Asano, and H. Sakoe, "Dewarping of document image by global optimization," in *Proc. of 8th Int. Conf. on Document Analysis and Recognition*, Seoul, Korea, 2005, pp. 500–506.

[7] A. Ulges, C. Lampert, and T. Breuele, "Document image dewarping using robust estimation of curled text lines," in *Proc. of 8th Int. Conf. on Document Analysis and Recognition*, Seoul, Korea, 2005, pp. 1001–1005.

[8] J. Liang, D. de Menthon, and D. Doermann, "Flattening curved documents in images," in *Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, San Diego, California, 2005, pp. 338–345.

[9] Z. Zhang and C. L. Tan, "Warped image restoration with applications to digital libraries," in *Proc. of 8th Int. Conf. on Document Analysis and Recognition*, Seoul, Korea, 2005, pp. 192–196.

[10] S. Lu and C. L. Tan, "The restoration of camera documents through image segmentation," in *Proc. of 8th Workshop on Document Analysis Systems*, Nelson, New Zealand, 2006, pp. 484–495.

[11] A. Masalovitch and L. Mestetskiy, "Usage of continuous skeletal image representation for document images dewarping," in *Proc. of 2nd Int. Workshop on Camera-Based Document Analysis and Recognition*, Curitiba, Brazil, 2007, pp. 45–53.

[12] B. Gatos, I. Pratikakis, and I. Ntirogiannis, "Segmentation based recovery of arbitrarily warped document images," in *Proc. of 9th Int. Conf. on Document Analysis and Recognition*, Curitiba, Brazil, 2007, pp. 989–993.

[13] M. Wu, R. Li, B. Fu, W. Li, and Z. Xu, "A page content independent book dewarping method to handle 2D images captured by a digital camera," in *Proc. of Int. Conf. on Image Analysis and Recognition*, Montreal, Canada, 2007, pp. 1242–1253.

[14] G. Rosman, A. Tzadok, and D. Tal, "A new physically motivated warping model for form drop-out," in *Proc. of 9th Int. Conf. on Document Analysis and Recognition*, Curitiba, Brazil, 2007, pp. 774–778.

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

[16] S. S. Bukhari, F. Shafait, and T. M. Breuel, "Dewarping of document images using coupled-snakes," in *Proc. of 3rd Int. Workshop on Camera-Based Document Analysis and Recognition*, Barcelona, Spain, July 2009, pp. 17–24.

[17] V. Kluzner, A. Tzadok, Y. Shimony, E. Walach, and A. Antonacopoulos, "Word-based adaptive OCR for historical books," in *Proc. of 10th Int. Conf. on Document Analysis and Recognition*, Barcelona, Spain, August 2009, pp. 501–505.

[18] R. Bernstein, "Digital image processing of earth observation sensor data," *IBM Journal of Research and Development*, vol. 20, no. 1, pp. 40–57, 1976.