Warped Image Restoration with Applications to Digital Libraries

Li Zhang, Chew Lim Tan
School of Computing, National University of Singapore
3 Science Drive 2, Singapore 117543
{zhangli, tancl}@comp.nus.edu.sg

Abstract

Recently, high resolution digital cameras have made the digitization process more flexible and convenient than traditional scanning technology. Therefore, document image analysis techniques need to be extended to handle images captured using ordinary hand-held cameras under an uncontrolled environment. In this paper, we propose an image restoration technique that corrects various warping distortions such as rolling curl, binding curl and fold distortions in camera images of geometrically distorted documents. Our goal is to restore these arbitrarily warped document images to a planar shape so as to facilitate OCR. Our method uses Spline interpolation technique based on a ruled surface model constructed from text lines in the 2D document image. Comparisons on OCR performance using Microsoft imaging software with two existing methods show an encouraging improvement in terms of recall and precision. The OCR precision on the corrected images using our method is improved up to 21.8% comparing to original images with a resolution of 3 mega pixels.

1. Introduction

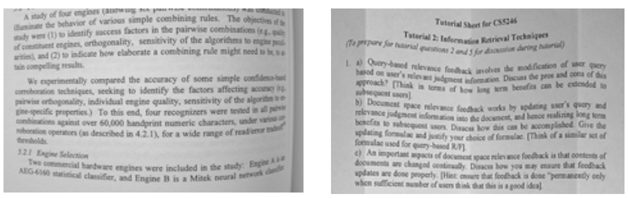
As we noticed that geometric distortions such as warping often exist when imaging non-planar documents using traditional scanning technology. This drastically degrades the recognition accuracy at the image processing stage. A typical example is when scanning thick bound documents. The binding area is often warped, shaded and blurred due to shape changes at the spine portion. This causes much problem both for human perception and machine recognition. To address this problem, Pilu [1] proposes a method of reconstructing the flat surface of a curled document by physically modeling the paper deformation with an applicable surface based on a set of 3D points and then unrolling it to a plane to produce the undistorted image. This method relies on a special 3D acquisition system to obtain a sparse set of 3D data and is prone to distortions if the calibration is not precise or if the polygon mesh for the applicable surface is coarse. Brown et al. [2] presents a general de-skewing algorithm for restoring arbitrary warped documents based on their 3D shape. It requires a special 3D digitization setup to acquire a 3D model using shape-to-texture registration under multiple lighting conditions. Therefore, it is not applicable to simple 2D images captured using normal scanners or digital cameras. Similarly, Zhang et al. [3] presented a novel method of restoring warping in scanned images by estimating the 3D surface shape of the document.

Nowadays with high resolution digital cameras widely available, digitization process becomes more flexible and convenient. However, it is observed that images captured using digital cameras often exhibit perspective distortions due to the camera optical system. In addition, geometric distortions such as rolling curl, binding curl, fold and skew also largely affect the recognition process. In recent years, many attempts have been tried to address these problems. Cao et al. [4] introduce a method to rectify warping distortions in camera images by constructing a general cylindrical model and using the skeleton of horizontal text lines in the image to help estimate the model parameters. Wu et al. [5] describe a method that infers the document's geometric distortion by constructing a source mesh based on the text lines extracted from the original 2D image and generating its corresponding non-uniform target mesh based on character density. The correction is then performed using a two-pass document image de-warping algorithm [6]. These approaches assume that the targeted distortion only appear once in the image. Tsoi et al. [7] present a novel approach that uses document boundary interpolation to correct various geometric distortions and shading artifacts present in images of art-like materials. This approach relies on the presence of boundary curves which may not be available when only a portion of the document is imaged or when the document color is mixed with the background.

In this paper, we propose a document-boundary independent approach to correct arbitrarily warped document images taken using ordinary digital cameras. Our method is based on a Gordon surface model [8] constructed from a set of text lines extracted from the 2D image. The text lines are represented using Natural Cubic Splines interpolating a set of points extracted from connected component analysis. In addition, we compare our method with two existing methods in terms of their OCR results on restored images. One is Tsoi's boundary interpolation method and the other is Zhang's 3D surface shape estimation method. The results show that our method achieves a higher precision than Tsoi's method and meanwhile removes its boundary constraints in the input image. Our method also outperforms Zhang's method in terms of OCR precision. In addition, Zhang's method is particularly for cylindrical-like shapes near book spines and thus not applicable to arbitrary warped images such as internal warpings, folds, etc.

2. Ruled Surface Modeling

Due to geometric distortions such as those in folded paper or thick bound books, and perspective distortions of the camera, the captured images often end up with non-uniform curvatures along the text lines as shown in Figure 1. This results in poor OCR accuracy during recognition. Our goal is to rectify these various warping distortions and produce a flattened image for better OCR.



(a) binding curl (b) folds

Figure 1. Grayscale distorted images taken using normal digital camera

As we notice in Figure 1, the images do not have explicit boundary curves for boundary interpolation. However, we can construct a ruled surface model based on the text lines present in the document. This can be visualized from Figure 2. Given a set of text lines $(C_1(u), C_2(u), \cdots, C_n(u))$ in the 3D document and two vertical text boundaries, we can construct a surface $\boldsymbol{g}$ that interpolates all these isoparametric curves $g(u, v_i)$ $i = 0, \cdots, m$ and $g(u_j, v)$ $j = 0, 1$ in terms of surface $\boldsymbol{g}$. Suppose the surface $\boldsymbol{g_1}$ interpolates to one family of isoparametric curves $g(u, v_i)$ $i = 0, \cdots, m$, and then $g_1(u, v)$ is defined as:

$$g_1(u, v) = \sum_{i=0}^{m} g(u, v_i) L_i^m(v) \tag{1}$$

where L_i^m are linear Lagrange polynomials defined as:

$$L_i^m(v) = \frac{\prod_{j=0, j \neq i}^{m} v - v_j}{\prod_{j=0, j \neq i}^{m} v_i - v_j} \tag{2}$$

Similarly, the lofted surface $g_2(u, v)$ is obtained as:

$$g_2(u, v) = \sum_{j=0}^{n} g(u_j, v) L_j^n(u) \tag{3}$$

Next, we define the interpolating tensor product surface:

$$g_{12}(u, v) = \sum_{i=0}^{m} \sum_{j=0}^{n} g(u_j, v_i) L_j^n(u) L_i^m(v) \tag{4}$$

Finally, the Gordon surface [8] can be represented as:

$$g(u, v) = g_1(u, v) + g_2(u, v) - g_{12}(u, v) \tag{5}$$

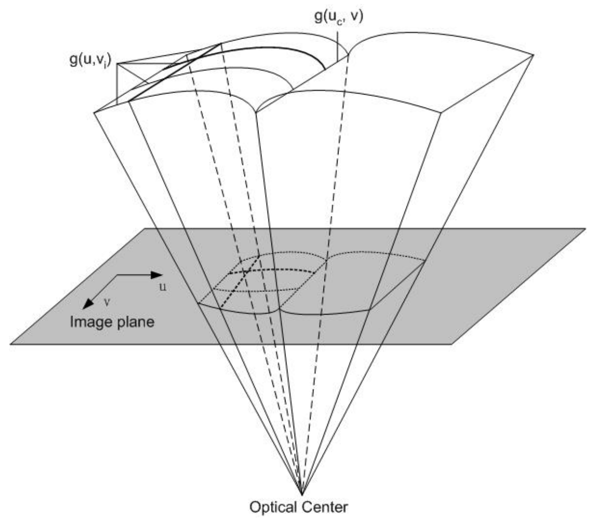


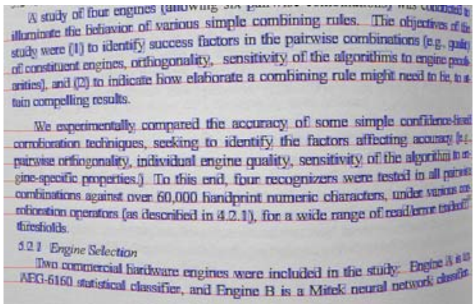
Figure 2. Gordon surface model for the distorted book surface. Its projection image can also be parameterized using the projected 2D curves

This Gordon surface model fits the definition of deployable surface and can be mapped to a planar surface without distortion. This model assumes that the book spines are along iso-parametric lines, e.g. $g(u_c, v)$ in terms of surface $\boldsymbol{g}$, is the straight line parallel to v-axis as shown in Figure 2. Consequently, this 3D Gordon surface model can be applied to the 2D projection image since straight lines are preserved under projection. The iso-parametric lines passing through $C_1(u_i), C_2(u_i), \cdots, C_n(u_i)$ must also pass through the corresponding 2D projection points. Therefore, the projection of this Gordon surface model can also be parameterized using the projected text lines in the 2D image.

3. Warping Correction

3.1 Text Line Extraction

Given an image without document boundary curves as shown in Figure 1, boundary interpolation techniques are thus not applicable. One solution is then to identify the curvature of the text lines by performing connected component analysis. First, a set of boundary points are extracted based on the bottom boundaries of the connected components. Then, an approximate Natural Cubic Spline can be constructed for each text line in the document. An example of the text lines detected for Figure 1(a) is shown in Figure 3.


A study of four engines (allowing ...) was conducted to illuminate the behavior of various simple combining rules. The objectives of the study were (1) to identify success factors in the pairwise combinations (e.g. quality of constituent engines, orthogonality, sensitivity of the algorithms to engine peculiarities), and (2) to indicate how elaborate a combining rule might need to be, to attain compelling results.

We experimentally compared the accuracy of some simple confidence-level corroboration techniques, seeking to identify the factors affecting accuracy (e.g. pairwise orthogonality, individual engine quality, sensitivity of the algorithm to engine-specific properties.) To this end, four recognizers were tested in all pairwise combinations against over 60,000 handprint numeric characters, under various corroboration operators (as described in 4.2.1), for a wide range of read/error tradeoff thresholds.

5.2.1 *Engine Selection*

Two commercial hardware engines were included in the study. Engine A is an AEG-6160 statistical classifier, and Engine B is a Mitek neural network classifier

Figure 3. Text lines extracted

3.2 Text Line Representation

Given a set of boundary points extracted for each text line, the curve can be represented using a Natural Cubic Spline that defines continuous image coordinates along the curve. Spline is used because it is advantageous in representing various warping curvatures. Moreover, due to the oscillatory nature of high-degree polynomials and the property of fluctuation over small intervals, Spline is clearly better off than polynomial. Each Natural Cubic Spline can be written as $s_i(t) = (x(t), y(t))$, where ***t*** is the knot value ranging from 0 to 1. Each knot value is associated with one 2D point on the curve. In other word, if there are n+1 2D point on curve ***i***, there will be n+1 corresponding knot values. Therefore, given (n+1) 2D points $(p_0, p_1, \cdots, p_n)$ and n+1 knot values $(t_0, t_1, \cdots, t_n)$ on a curve ***i***, a set of piece-wise functions are uniquely defined as follows:

$$s_i(t) = \begin{cases} s_{i,0}(t) & t \in [t_0, t_1] \\ s_{i,1}(t) & t \in [t_1, t_2] \\ \vdots & \\ s_{i,n-1}(t) & t \in [t_{n-1}, t_n] \end{cases} \tag{6}$$

where $s_{i,j}(t) = (x_{i,j}(t), y_{i,j}(t))$ $j = 0, \cdots, n-1$ and $x_{i,j}(t)$ and $y_{i,j}(t)$ are cubic functions. These piece-wise cubic functions can be derived easily if the n+1 2D points and the corresponding n+1 knot values are available.

From section 3.1, a set of boundary points can be extracted for each text line. These constitute the (n+1) 2D points. Intuitively, its corresponding set of knot values can be defined based on the Euclidean distance between these 2D points, i.e.:

$$t_i = \begin{cases} 0 & if\ i = 0 \\ \frac{1}{|d|} \sum_{0<j<n+1} |p_j - p_{j-1}|_2 & if\ i > 0 \end{cases} \tag{7}$$

where $|d| = \sum_{0<j<n+1} |p_j - p_{j-1}|_2$

This parameterization assigns larger t values for segments with more obvious curvature and thus allows more pixels to be mapped to the original image during restoration so that the characters in the warping area will not appear too squeezed after restoration.

3.3 Text Line Interpolation

After obtaining a set of Natural Cubic Splines that represent all well identified text lines, the restored image can be generated through a 2D mapping function derived from Gordon surface model in section 2. The restored image is defined over the space u and v, where $u \in [0,1]$ and $v \in [0,1]$. For each point (u_i, v_j) in the restored image, its corresponding point (x, y) can be calculated by substituting the values of u_i and v_j to the function $g(u, v)$ to compute the component $x_1(u_i,v_j)$, $y_1(u_i,v_j)$ and $x_2(u_i,v_j)$, $y_2(u_i,v_j)$ respectively:

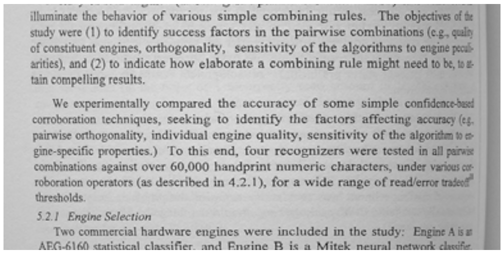
$$\begin{aligned} x(u_i, v_j) &= x_1(u_i, v_j) + x_2(u_i, v_j) - x_{12}(u_i, v_j) \\ y(u_i, v_j) &= y_1(u_i, v_j) + y_2(u_i, v_j) - y_{12}(u_i, v_j) \end{aligned} \tag{8}$$

For simplicity, we can also select only the top and bottom two text lines as the boundary curves for interpolation. A special case of Gordon surface model known as coons patch [9] can thus be applied to this four boundary interpolation problem as described in Tsoi's paper [7].

4. Experimental Results

We tested our method on a set of 20 document images captured using ordinary digital cameras with resolutions of 2 mega pixels and 3 mega pixels respectively. The document images are of different font sizes and taken with different binding curvatures such as bindings on the left or right, folds and rolling curves. The images can be taken from any angle and the corresponding transformation will be reflected in the restored image. The restored image for Figure 1(a) is shown in Figure 4. As we can see that the warped text lines are rectified to be straight lines. Due to the shading problem in some images, the restored image may still look distorted. This can be easily resolved by a de-shading procedure as described by Tsoi [7] using a boundary interpolation method.

In addition, both the distorted image and the restored image are fed into Microsoft OCR engine for text recognition. Precision and recall defined in terms of the number of words correctly detected are used as the evaluation metrics for the OCR results.

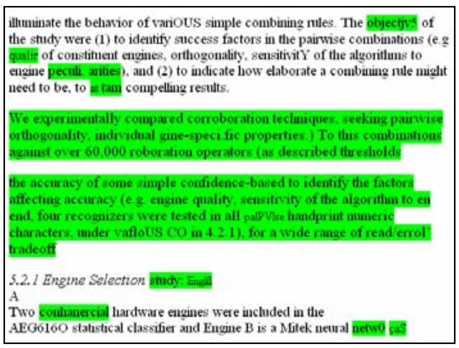

illuminate the behavior of various simple combining rules. The objectives of the study were (1) to identify success factors in the pairwise combinations (e.g., quality of constituent engines, orthogonality, sensitivity of the algorithms to engine peculiarities), and (2) to indicate how elaborate a combining rule might need to be, to attain compelling results.

We experimentally compared the accuracy of some simple confidence-based corroboration techniques, seeking to identify the factors affecting accuracy (e.g., pairwise orthogonality, individual engine quality, sensitivity of the algorithm to engine-specific properties.) To this end, four recognizers were tested in all pairwise combinations against over 60,000 handprint numeric characters, under various corroboration operators (as described in 4.2.1), for a wide range of read/error tradeoff thresholds.

5.2.1 Engine Selection

Two commercial hardware engines were included in the study: Engine A is an AEG-6160 statistical classifier, and Engine B is a Mitek neural network classifier

Figure 4. Restored image

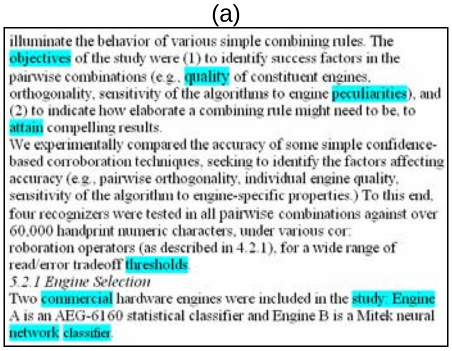

illuminate the behavior of variOUS simple combining rules. The objectjys of the study were (1) to identify success factors in the pairwise combinations (e.g qualir of constituent engines, orthogonality, sensitivitY of the algorithms to engine peculi. arities), and (2) to indicate how elaborate a combining rule might need to be, to at tam compelling results.

We experimentally compared corroboration techniques, seeking pairwise orthogonality, individual gine-speci.fic properties.) To this combinations against over 60,000 roboration operators (as described thresholds

the accuracy of some simple confidence-based to identify the factors affecting accuracy (e.g. engine quality, sensitivity of the algorithm to en end, four recognizers were tested in all palPVlse handprint numeric characters, under vafloUS CO in 4.2.1), for a wide range of read/errol' tradeoff

5.2.1 Engine Selection study: Engil
A
Two conhanercial hardware engines were included in the AEG6160 statistical classifier and Engine B is a Mitek neural netw0 ca$

(a)

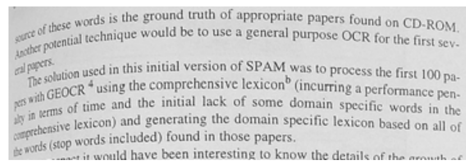

illuminate the behavior of various simple combining rules. The objectives of the study were (1) to identify success factors in the pairwise combinations (e.g., quality of constituent engines, orthogonality, sensitivity of the algorithms to engine peculiarities), and (2) to indicate how elaborate a combining rule might need to be, to attain compelling results.
We experimentally compared the accuracy of some simple confidence-based corroboration techniques, seeking to identify the factors affecting accuracy (e.g., pairwise orthogonality, individual engine quality, sensitivity of the algorithm to engine-specific properties.) To this end, four recognizers were tested in all pairwise combinations against over 60,000 handprint numeric characters, under various cor:
roboration operators (as described in 4.2.1), for a wide range of read/error tradeoff thresholds.
5.2.1 Engine Selection
Two commercial hardware engines were included in the study: Engine A is an AEG-6160 statistical classifier and Engine B is a Mitek neural network classifier.

(b)

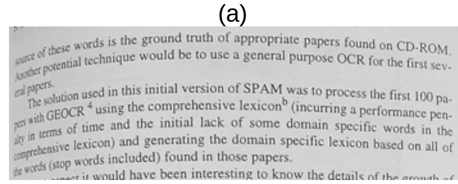
Figure 5. OCR output for: (a) original distorted image in Figure 1(a); (b) restored image

It is shown that after restoration the OCR precision is improved by 13.1% and 21.8% respectively for images of 2 and 3 mega pixels resolutions. Generally speaking, higher resolution images with larger font sizes produce better OCR outputs. This mainly attributes to the accuracy of connected component analysis. Examples of the OCR results of the original distorted image and the restored image of 2 mega pixels resolution are shown in Figure 5.(a) and (b) respectively. As we can see in (a), most words in the warping area cannot be correctly recognized such as “objectives”, “quality”, etc. Although some words are correctly recognized, they are placed in the wrong position due to the curvature in the text line. This results in a misinterpreted context as shown in the highlighted portion. On the contrary, the restored image is much better recognized with a higher precision as shown in (b).

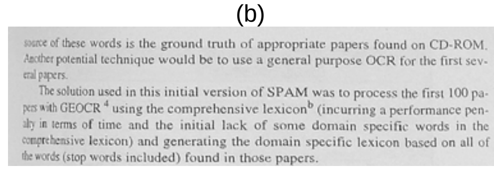
To compare our method with Tsoi’s boundary interpolation method and Zhang’s 3D surface modeling technique, we chose a set of images with boundary curves for testing. As an example, the restored images for a portion of a page image in Figure 6.(a) using each method are shown in Figure 6.(b)(c)(d). The resulting images of our method and Tsoi’s boundary interpolation method do not exhibit much difference through human perception. However, OCR testing using Microsoft OCR engine shows an average increase of 2.1% in OCR precision whereas Zhang’s method does not improve much on the original distorted image. The OCR results on the restored images in Figure 6 are shown in Figure 7.



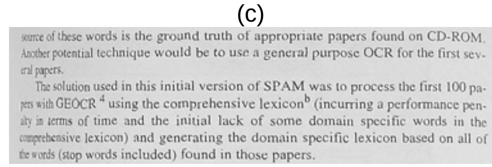
(a)



(b)



(c)


source of these words is the ground truth of appropriate papers found on CD-ROM. Another potential technique would be to use a general purpose OCR for the first several papers.

The solution used in this initial version of SPAM was to process the first 100 papers with GEOCR [4] using the comprehensive lexicon[b] (incurring a performance penalty in terms of time and the initial lack of some domain specific words in the comprehensive lexicon) and generating the domain specific lexicon based on all of the words (stop words included) found in those papers.

(d)

Figure 6. Part of the restored images for a 2D page image using three de-warping methods: (a) original warped image (b) Zhang’s method (c) Brown’s method (d) Our method

More specifically, a comprehensive comparison of OCR precision and recall among all three methods on images with or without boundaries is shown in Table 1. Generally speaking, Zhang’s method achieves a relatively lower precision but higher recall comparing to our method and Tsoi’s method. This is mainly because the cross section shape it models is particular for cylindrical-like book spines and is not well applicable to ordinary camera images. On the other hand, our method achieves a relatively higher precision than Tsoi’s method but with a slight lower recall due to the elimination of short text lines on the top or bottom of the page. Nevertheless, our method is a more general method than Tsoi's method in the sense that we accommodate the situation in which the boundary curves are not available or difficult to identify. Moreover, our method can be used to model Tsoi's method by selecting the topmost and bottommost text lines as the two boundary curves for interpolation.

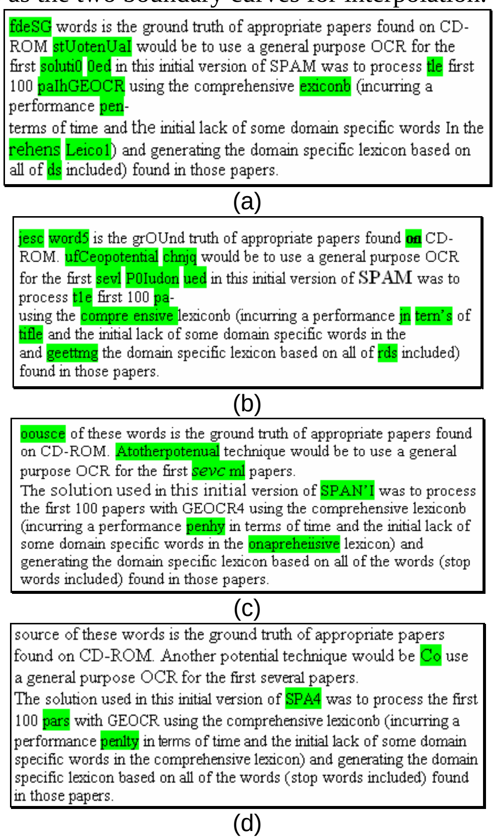


Figure 7. OCR results for the restored partial images in Figure 6: (a) original image (b) Zhang's method (c) Tsoi's method (d) Ours

Table 1. OCR result comparisons on restored images using three de-warping methods

OCR accuracy		Original	Zhang's	Brown's	Ours
Images w/ boundary	P	73.4%	75.3%	93.6%	95.7%
	R	90.1%	93.2%	92.7%	91.3%
Images w/o boundary	P	76.7%	78.2%	-	97.5%
	R	90.3%	93.3%	-	90.2%

5. Conclusion and future works

In this paper, we presented an image de-warping method that corrects arbitrary image warping distortions such as rolling curl, binding curl, and folds present in document images captured using ordinary digital camera. Our method directly works on distorted 2D images without the need of a special 3D acquisition system. A ruled surface model is applied to a set of Natural Cubic Splines representing the text lines extracted from the 2D document image. The method is a general version of Tsoi's boundary interpolation method in the sense that text lines are extracted to reflect warping curvatures and Gordon surface model is used for a more accurate interpolation. This eliminates the constraint of Tsoi's method that requires the original distorted image with clearly identifiable document boundary curves. The OCR results on restored images show an obvious improvement in precision over the original distorted images.

On the other hand, our method relies on the presence of well structured text lines that are long enough to reflect the warping curvature. Therefore, in the case where the top or bottom text lines are too short, they will be eliminated and result in a low OCR recall. A de-shading process is also desirable to generate a high quality binary image for text line detection. Our future work is thus to find ways to address these problems.

Acknowledgement

This research is supported in part by A*STAR under grant R252-000-206-305 and NUS URC under grant R252-000-202-112.

References

[1] Pilu, M. "Undoing page curl distortion using applicable surfaces". *Computer Vision and Pattern Recognition Conference*, Vol. 1, pp. 67-72, December 2001.

[2] Brown, M. S., Seales, W. B. "Document restoration using 3D shape: a general deskewing algorithm for arbitrarily warped documents". *International Conference on Computer Vision*, Vol 2, pp. 367-374, July 2001.

[3] Zhang Z., Tan, C. L., Fan, L. Y. "Estimation of 3D shape of warped document surface for image restoration". *International Conference on Pattern Recognition, ICPR 2004*, 23-26 Aug 2004, UK.

[4] Cao, H., Ding, X., Liu, C. "A cylindrical model to rectify the bound document image". *International Conference on Computer Vision*, Vol. 2, pp. 228-233, October 2003.

[5] Wu, C.H., Agam, G. "Document image de-warping for text/graphics recognition". *Proc. Of Joint IAPR2002 and SPR 2002*, Canada, Aug. 2002, pp. 348-357.

[6] Worlberg, G. *Digital image warping*. IEEE Computer Society Press, Los Alamitos, California, 1990.

[7] Tsoi, Y. C., Brown, M. S. "Geometric and shading correction for images of printed materials: a unified approach using boundary". *Computer Vision and Pattern Recognition conference*, pp. 240-246, 2004.

[8] Farin, G. *Curves and Surfaces for Computer Aided Geometric Design*. A practical guide, 4th edition.

[9] Coons, S. "Surfaces for computer aided design". Technical report, MIT, 1968.