Camera Document Restoration for OCR

SHIJIAN LU, CHEW LIM TAN, SENIOR MEMBER, IEEE
School of Computing, National University of Singapore
Singapore 117574

Abstract

As camera resolution increases, high-speed non-contact text capture through a digital camera is opening up a new channel for text capture and understanding. Unfortunately, the skew, perspective, and geometric distortions coupled within the captured images make it hard to recognize the document text using the generic OCR systems. In this paper, we propose a document restoration technique, which is capable of removing the three types of distortions, and reconstructing the fronto-parallel view of the document text using a single document image captured through a digital camera. Different from the reported techniques, the proposed restoration technique is carried out based on the vertical stroke boundary and the top line and base line of text lines. Experimental results show the proposed technique is fast, accurate, and robust.

1. Introduction

As sensor resolution increases in recent years, high-speed non-contact text capture through a digital camera is becoming an alternative choice. Unfortunately, the document images captured through a digital camera are often coupled with the distortions including rotation-induced skew, perspective, and geometric distortions. These three types of distortions must be removed before the camera documents are fed to the generic OCR system.

As Figure 1(a) shows, the rotation-induced skew normally occurs as the image plane *R* of the digital camera is parallel to the document plane *D*. While the camera image plane *R* is not parallel to the document plane *D*, the perspective distortion as illustrated in Figure 1(b) is inevitably introduced. In addition, as most of scene documents such as the hand-held newspaper, the paper sheets pasted on cylindrical containers and even the pages bound within the thick books generally lie on a smoothly curved instead of planar surface, the camera documents are often coupled with the geometric distortion as illustrated in Figure 1(c) as well.

A large number of document restoration methods [1-6] have been reported in the literature. Traditionally, document distortion generally refers to the rotation-induced skew and the main problems of the reported skew detection methods lie with the restriction on the detectable skew angle range [1] and the heavy computation load [2]. In recent years, a few perspective restoration techniques have been reported, but most of the reported techniques rely heavily on the image features such as the high contrasted document boundary (HDB) [3] and the paragraph formatting (PF) information such as paragraph margins [4]. A few geometric restoration techniques have been reported as well in recent years. Most of reported methods [5-6] approach the restoration problem through the 3D reconstruction, but auxiliary hardware is normally required for 3D measurements.

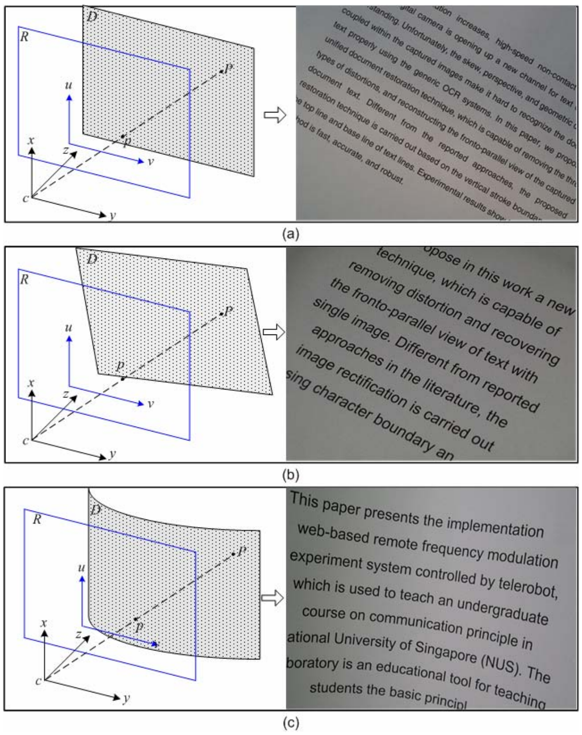

Figure 1: (a) camera document with skew; (b) camera document with perspective distortion; (c) camera document with geometric distortion

In [7], we propose to remove the perspective distortion through the detection of the vertical stroke boundary (VSB) and the top line and base line of text lines as labeled with (1) and (2) in Figure 2. VSBs are firstly identified based on three fuzzy sets that characterizes the size, linearity, and orientation of the extracted stroke boundaries. The top line and base line of text lines are then fitted using character tip points that are classified based on the structure of the typeset document text. For the sample document given in Figure 1(b), Figure 3 shows the identified VSB where text is printed in a light gray color to highlight its relative position to the identified VSBs.

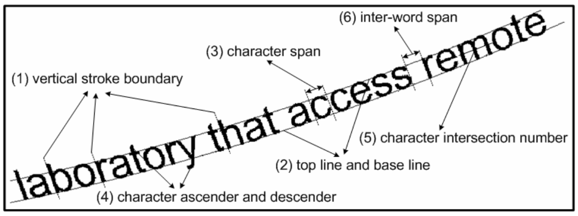

Figure 2: Text line definition

In this paper, we extend our work reported in [7] to restore the camera documents with three types of document distortions. The proposed technique has multiple advantages. Firstly, it is able to estimate the skew angle ranging from 0° to 360° and the estimation speed is totally independent of the skew angle. Secondly, it is able to rectify the perspectively distorted camera documents that have no HDB or PF features and may contain only one text line or even just a few words. Thirdly, it is able to restore the camera documents with geometric distortion with just a single document image captured through a digital camera. Lastly, the proposed technique needs no camera calibration and it requires only a camera image of document with good resolution.

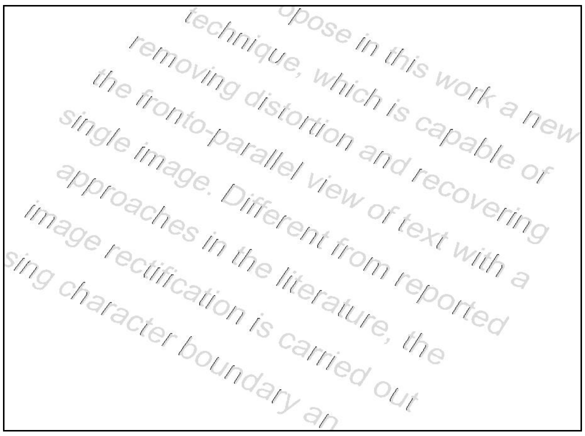

Figure 3: Identified VSB from document image given in Figure 1(b)

In our proposed approach, the skew angle can be simply estimated based on the orientation of the top line and base line. For the camera documents with perspective distortion, the restoration is carried out with the homography estimated based on the top line and base line of text lines and the identified VSB. For camera documents with geometric distortion, we propose to remove the distortion through the image segmentation, which partitions the camera documents into multiple small image patches where text can be approximated to lie on a planar surface. The global geometric distortion is then removed through the local rectification of the partitioned image patches one by one.

2. Proposed Restoration Technique

We present in this section the outline of the proposed document restoration technique. In particular, we will divide this section into a few subsections, which deal with the identification of document distortions and the restoration of camera documents with rotation-induced skew, perspective, and geometric distortions respectively.

2.1. Distortion Identification

For document images with rotation-induced skew, the restoration can be simply implemented through an image rotation operation. But for document images with geometric distortion, the restoration process is much more complex because it involves the VSB identification, the top line and base line fitting, and the image segmentation. Therefore, it is better to identify the distortion type first before the actual restoration operation. We propose to identify the distortion type based on the pattern of the classified character centroids, which normally fit well to a set of parallel straight lines, unparallel straight lines and smooth curves respectively for document images with the three types of distortions.

The document images with skew or perspective distortions can be firstly differentiated from the ones with geometric distortion based on the linear fitting error, which can be evaluated using the distance:

$$D = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m} Dist(C_j, L_i) \tag{1}$$

where parameters n and m refer to the number of the fitted middle lines and the number of characters centroids within the *ith* classified character centroid category. Parameter L_i refers to the straight middle line fitted with the character centroids within the *ith* category. Function $Dist(C_j, L_i)$ calculates the distance between the *jth* character centroid C_j within the *ith* character centroid category and the *ith* fitted straight middle line L_i.

Based on the distance defined in Equation (1), the distance threshold can be determined as:

$$TD = k_d \cdot VSB_{avg} \quad (2)$$

Parameter VSB_{avg} is the average size of the identified VSBs, which normally indicates the size of the captured document text. Parameter k_d [0.1 0.5] is designed to adjust the distance threshold and we set it as 0.3 in the implemented system. Therefore, geometric distortion is detected if the distance D determined using Equation (1) is bigger than the distance threshold TD given in Equation (2). Otherwise, document images are determined to contain rotation-induced skew or perspective distortion.

Skew and perspective distortions can be further differentiated from each other based on relative orientation of the fitted middle lines. For document images with skew or perspective distortion, the fitted middle lines normally correspond to a set of parallel or unparallel straight lines respectively. The relative orientation of the fitted middle lines can thus be evaluated as:

$$RO = \frac{1}{n}\sum_{i=1}^{n}\left(\varphi_i - \frac{1}{n}\sum_{i=1}^{n}\varphi_i\right)^2 \quad (3)$$

where parameter n refers to the number of the fitted middle lines. Parameter φ_i refers to the orientation angle of the *ith* fitted middle line. For the document images with rotation-induced skew, the relative orientation RO determined in Equation (3) is quite close to zero. But for the document images with perspective distortion, the relative orientation RO is normally much bigger. Skew and perspective distortions can thus be differentiated based on the relative text line orientation RO given in Equation (3)

Table 1 Distortion identification performance

	NO of sample images	No of correctly identified images	Identification rate
Camera documents with skew distortion	30	31	96.67%
Camera documents with perspective distortion	30	29	96.67%
Camera documents with geometric distortion	30	30	100%

The proposed distortion identification technique is able to differentiate the three types of document distortions in most cases. We test the identification performance using 90 distorted camera images of documents as given in Table 1 where the distortion types of the 88 images are correctly identified. The identification error normally occurs while the two related distortions are quite close to each other. For example, perspective distortion may be falsely identified as skew distortion as the angle between the optical axis of digital camera and document plane is close to 90° and the captured text lines are roughly parallel to each other.

2.2. Document Restoration

This section presents the restoration of the camera documents with the three types of distortions including rotation-induced skew, perspective and geometric distortions respectively.

2.2.1. Skew Detection and Correction

For the camera documents with rotation-induced skew, the top line and base line of text lines are actually a set of approximately parallel straight lines. The skew angle can thus be simply estimated based on the orientation of the top line and base line. The skew angle can be determined as:

$$\tan(\beta) = \frac{1}{n}\sum_{i=1}^{n} slp_i \quad (4)$$

where parameter n is the number of fitted top line and base line and slp_i refers to the slope of the *ith* top line or base line.

Based on the skew angle determined using the Equation (4), the camera documents with skew distortion can be restored through a simple image rotation operation. For the skewed camera document given in Figure 1(a), Figure 4 shows the restored document image.

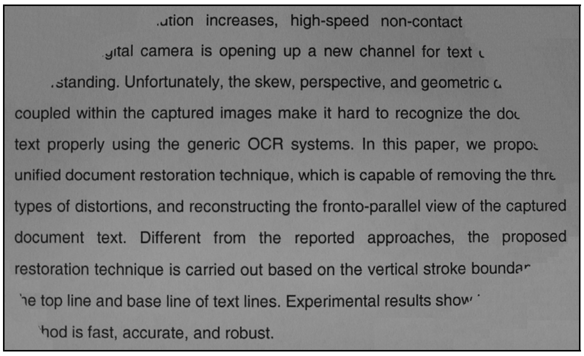
.ution increases, high-speed non-contact
gital camera is opening up a new channel for text
.standing. Unfortunately, the skew, perspective, and geometric
coupled within the captured images make it hard to recognize the do
text properly using the generic OCR systems. In this paper, we propo
unified document restoration technique, which is capable of removing the thre
types of distortions, and reconstructing the fronto-parallel view of the captured
document text. Different from the reported approaches, the proposed
restoration technique is carried out based on the vertical stroke bounda
he top line and base line of text lines. Experimental results show
hod is fast, accurate, and robust.

Figure 4: Skew restoration result

2.2.2 Perspective distortion detection and correction

For camera documents with perspective distortion, the top line and base line generally correspond to a set of unparallel straight lines. In this paper, we propose to remove the perspective distortion through the exploitation of the quadrilateral correspondences where the source quadrilaterals are constructed using the identified VSB and the top line and base line of text lines. For each determined source quadrilateral, a target quadrilateral is constructed based on the number of the characters enclosed within the source quadrilateral and the specific character width height ratios. With the constructed source and target quadrilateral correspondences, optimal rectification homography is estimated and perspective distortion is finally removed.

For each straight line fitted based on an identified VSB, there will be multiple intersections between it and the top line and base line pairs. The identified VSB must be classified to the text line from which they are extracted to construct the desired source quadrilaterals and this can be achieved based on the distances between the centroids of the identified VSB and the fitted top line and base line pairs. Thus, source quadrilaterals can be constructed using the identified VSB and the related top line and base line pairs.

As the camera capturing process impairs the geometric relation between the straight lines, we propose to construct the target quadrilateral based on the number of the characters enclosed within the source quadrilaterals and the approximation that the width height ratio of characters is 1:1. It should be clarified that the 1:1 width height ratio is only an average approximation, as the width height ratios of different characters such as “m” and “i” may differentiate quite a lot. To make the approximation more close to the fact, the constructed source quadrilaterals must be wide enough to enclose more characters. In our proposed technique, the distance threshold between two VSB is defined based on the average length of the captured text lines:

$$L = k_r \cdot \frac{1}{n}\sum_{i=1}^{n} leg_i \tag{5}$$

where n is the number of detected text lines. Parameter k_r is used to adjust distance threshold and we take it as 0.4 in our system. Symbol leg_i represent the length of *ith* text line, which is calculated as the distance between the leftmost and rightmost pixels of characters that belong to the same text line.

Characters within the constructed source quadrilaterals can thus be determined based on relative position between the character centroids and the four source quadrilateral edges. The inter-word blank can be detected as well based on the distance between the centroids of the adjacent characters and it takes the width same to a character. With the approximated character height width ratio, the relation between the width and height of target quadrilaterals can be restored as

$$l_q = n \cdot h_q \tag{6}$$

where parameters l_q and h_q are the length and height of the target quadrilaterals. Parameter n is the number of character enclosed within the source quadrilateral, including the detected inter-word blanks. The height of the target quadrilateral h_q can be commonly determined as the average size of the identified VSB.

With multiple pairs of source and target quadrilaterals, multiple rectification homographies can be determined using the four point algorithm [8]. The homography between the distorted and restored document images can be estimated as:

$$H = A^{-1} \cdot R \tag{7}$$

where H is the homography matrix and matrixes A, R are constructed using four point correspondences. The three matrixes take the following form:

$$H = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix}, \quad A = \begin{bmatrix} -x_1 & -y_1 & -1 & 0 & 0 & 0 & x'_1 \cdot x_1 & x'_1 \cdot y_1 \\ 0 & 0 & 0 & -x_1 & -y_1 & -1 & y'_1 \cdot x_1 & y'_1 \cdot y_1 \\ -x_2 & -y_2 & -1 & 0 & 0 & 0 & x'_2 \cdot x_2 & x'_2 \cdot y_2 \\ 0 & 0 & 0 & -x_2 & -y_2 & -1 & y'_2 \cdot x_2 & y'_2 \cdot y_2 \\ -x_3 & -y_3 & -1 & 0 & 0 & 0 & x'_3 \cdot x_3 & x'_3 \cdot y_3 \\ 0 & 0 & 0 & -x_3 & -y_3 & -1 & y'_3 \cdot x_3 & y'_3 \cdot y_3 \\ -x_4 & -y_4 & -1 & 0 & 0 & 0 & x'_4 \cdot x_4 & x'_4 \cdot y_4 \\ 0 & 0 & 0 & -x_4 & -y_4 & -1 & y'_4 \cdot x_4 & y'_4 \cdot y_4 \end{bmatrix}, \quad R = \begin{bmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \\ x'_4 \\ y'_4 \end{bmatrix} \tag{8}$$

where the 3×3 homography matrix is expressed in vector form and h_{33} is equal to 1 under homogeneous frame. Four point correspondences $\langle (x_i, y_i), (x'_i, y'_i) \rangle$, $i = 1, \ldots, 4$, are taken as the four vertices of the constructed source and target quadrilateral pairs.

The vertex position of the constructed source quadrilateral normally contains errors. As a small error in source quadrilateral vertices may introduce a big error to the restored document images, a criterion must be set to choose the homography that optimize the restoration performance. Based on the facts that the top line and base line should be restored to multiple horizontal lines and the identified VSB should lie on multiple vertical lines within the restored document image, we define the objection function as:

$$J = \frac{1}{m}\sum_{i=1}^{m} abs\left(\frac{S_{l_i}}{S_{avg}}\right) + \frac{1}{n}\sum_{j=1}^{n} abs\left(\frac{ptx_j - pbx_j}{Dist_{avg}}\right) \tag{9}$$

where m is the number of detected text lines and n is the number of the identified VSB. S_{li} is the orientation of *i-th* restored text line and S_{avg} is the orientation average. ptx_j and pbx_j represent two horizontal coordinates of vertices of *j-th* restored VSB and the component $abs((ptx_j - pbx_j)/ Dist_{avg})$ is the normalized distance

$abs((ptx_j - pbx_j)/ Dist_{avg})$ is the normalized distance in horizontal direction between the vertices of that vertical stroke boundary. The first part on the right side of Equation (9) represents the sum of normalized orientation of the restored text lines, which should be zero ideally, and the second part refers to the sum of normalized vertex distance of the restored VSB in horizontal direction, which ideally should be zero as well. The optimal homography can accordingly be determined as the one that minimizes the objection function defined in Equation (9).

The camera document with perspective distortion as given in 1(b) can be finally restored based on the estimated optimal homography. Figure 5 shows the restored document image.

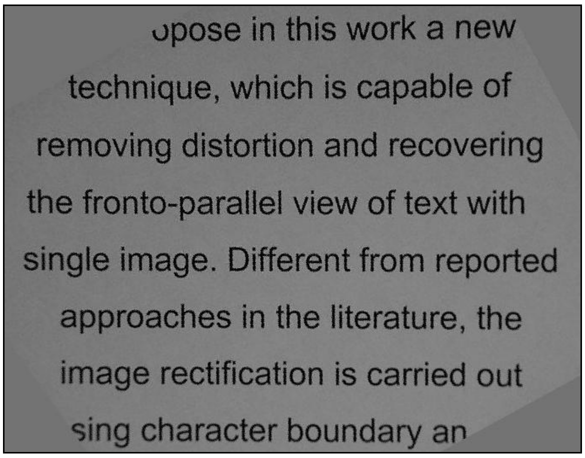

Figure 5 Perspective restoration result

2.2.3. Geometric distortion detection and correction

Based on the top line and base line pairs and the identified VSB, we propose to remove the geometric distortion through the image segmentation, which partitions the distorted camera documents into multiple small image patches where text can be approximated to lie on a planar surface. For each partitioned image patch, a target rectangle is then constructed based on the characters within that partitioned image patch and the specific character width height ratios. Lastly, the global geometric distortion is corrected through the local rectification of the partitioned image patches one by one.

Before camera document segmentation, the identified VSB must be processed further to facilitate the later restoration and make sure that the partitioned image patches enclose all captured text. Firstly, some VSB must be deleted if they are too close to their left adjacent neighbor. VSB deletion operation is designed to control the size of partitioned image patches, but the identified VSB cannot be deleted arbitrarily. In our proposed technique, we delete the VSB based on its distance to the left adjacent VSB and the distance threshold is determined as:

$$D_{thre} = k_d \cdot VBS_{avg} \tag{10}$$

where parameter VBS_{avg} represents the average size of the identified VSB. Parameter k_d is designed to adjust the distance threshold and it is determined as a number between 3 and 5 so that each partitioned image patch enclose 3-5 characters.

In addition, for the text lines that have no VSB identified at their left or right end, a VSB must be constructed there so that the partitioned image patches are able to enclose all characters that belong to the studied text line. The orientation of the VSB at the text line end positions can be estimated through linear interpolation:

$$slp = slp'' + \frac{(x - x'') \cdot (slp' - slp'')}{(x' - x'')} \tag{11}$$

where x is x coordinates of the leftmost or rightmost text pixel and x', x'' are x coordinates of centroids of two VSB that are nearest to the related leftmost or rightmost text pixel. Parameters slp' and slp'' are slopes of the straight lines fitted based on the two nearest VSB neighbors.

Accordingly, the VSB at the leftmost or rightmost end can be estimated as a straight line that passes through the leftmost or rightmost character pixel with orientation same to the one estimated in Equation (11). For the distorted word given in Figure 6(a), Figure 6(b) shows the top line and base line and the identified VSB. Figure 6(c) gives the processed VSB after the deletion and addition operations where the second VSB from the left is deleted and the VSB at the rightmost end of text line is estimated. For each processed VSB, a straight line can thus be fitted using the least square method. With the top line and base line of text lines and the straight lines fitted based on the processed VSB, distorted text as given in Figure 6(a) is finally segmented into three small patches as given in Figure 6(d).

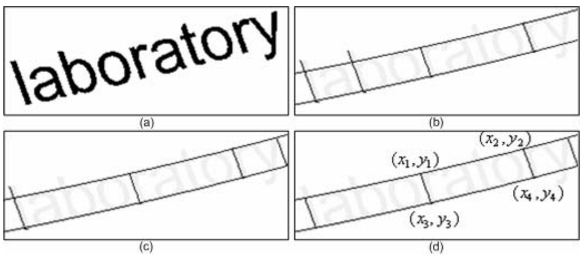

Figure 6: (a) Original sample word; (b) identified VSB and the top and base line; (c) VSB after the deletion and addition operation; (d) three segmented document patches

For each partitioned image patch, a target rectangle correspondence must be constructed within the target image to rectify that partitioned image patch. The height of target rectangles, which is same to the height of rectified characters with no ascender and descender, can be commonly determined as the average size of identified VSB. We thus propose to restore the width of target rectangles based on the characters enclosed within the partitioned image patches and specific character width height ratios.

While restoring the camera documents with only perspective distortion, we approximated the character width height ratio as 1:1. The approximated ratio works well because the width of the source quadrilateral determined using Equation (5) is large enough to contain a large number of characters. But for camera documents with geometric distortion, the partitioned image patches enclose only 3-5 characters and so the 1:1 approximation cannot be used here. We thus propose a rough character classification process to classify characters into different categories with different width height ratios.

We propose to classify characters into six categories with different character width height ratios. The classification is carried out based on multiple image features extracted from character strokes including character span, character ascender and descender, character intersection numbers, and inter-word spans as labeled with (3), (4), (5), and (6) in Figure 2. Character span is defined as the distance between two parallel straight lines tangent to the left and right sides of the studied character with the orientation same to that of the straight line fitted based on the nearest VSB. Inter-word span can be determined in the similar way as character span. The intersection numbers are equal to the number of intersection between character strokes and the straight lines that pass through the character centroid with orientation orthogonal to that of the straight line fitted based on the nearest VSB. Character ascender and descender can be determined based on the distance between the highest and lowest character pixel and the top line and base line.

With the determined text line features, the character classification algorithm is as follows:

Inputs: Binarized document image *BDI*; Calculated character spans *CSpan*; Ascender & descender information *ADInfo*; intersection numbers *Inter*

Procedure: *CC* (*BDI*, *CSpan, ADInfo, Inter*)

1) Initialize $i = 1$
2) Calculate the average character span *CSpan_avg* based on *CSpan.*
3) Repeat:
4) If *Inter(i)* ≥ 3 and *ADInfor(i)* = 1(with ascender), character is classified as ‘M’ or ‘W’.
5) Else if *Inter(i)* ≥ 3 and *ADInfor(i)* = 0(no ascender), character is classified as ‘m’ or ‘w’.
6) Else if *ADInfor(i)* = 1(with ascender) and $CSpan(i) > k_u{\cdot}CSpan\_avg$, character is classified as A-H, J-L, N-V, or X-Z.
7) Else if $CSpan(i) > k_l{\cdot}CSpan\_avg$ and $CSpan(i) < k_u{\cdot}CSpan\_avg$, character is classified as, a-e, g-h, k, n-q, s, or u-v, or x-z
8) Else if $CSpan(i) < k_s{\cdot}CSpan\_avg$, character is classified as ‘i’, ‘l’, ‘I’ or ‘j’.
9) Else, character is classified as t, f, or r.
10) $i = i + 1$
11) Until i is equal to the number of characters within *BDI*

Table 5.1 shows the proposed six categories and the related character width-height ratios.

Table 1: Character classification and related width-height ratio

Classified characters	Character width height ratios (***R***)
M, W	1.6:1
m, w	1.4:1
A-H, J-L, N-V, X-Z	1.2:1
Inter-word span	1:1
a-e, g-h, k, n-q, s, u-v, x-z	0.8:1
t, f, r	0.5:1
i, j, l, I,	0.2:1

The average character span *CSpan_avg* in Step 2) is firstly determined before the classification. Parameter k_u, k_l, and k_s in Steps 6), 7) and 8) are three key parameters for character classification. In our implemented system, the three parameters are determined as 1.2, 0.7, and 0.3 respectively based on the relative width of characters in different categories.

We evaluate the proposed character categorization technique using the same 90 sample images as used for distortion identification. Experiment results show that the correct classification rate can reach over 96%. The small classification error will not affect the recognition performance of the restored document images seriously because each partitioned image patch normally contain 3-5 characters.

The width of target rectangles can thus be determined as:

$$T_w = \sum_{i=1}^{n} R_i \cdot VBS_{avg} \tag{12}$$

where VBS_{avg} represent the average size of identified VSB and parameter n represents the number of characters and inter-word blanks enclosed within the partitioned image patches. Parameter R_i refers to width height ratios of characters and inter-word blanks within the partitioned image patches as given in Table 1.

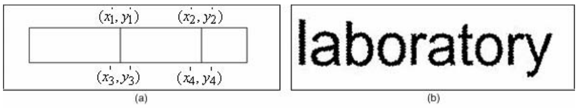

Figure 7 (a) constructed target quadrilateral; (b) restored document image

For the segmented image patches given in Figure 6(d), Figure 7(a) shows the constructed target rectangles. For each quadrilateral correspondence, a homography can be determined using Equation (8). Document text with geometric distortion given in Figure 6(a) can thus be restored through the rectification of three partitioned image patches. Figure 7(b) gives the restoration result.

Figure 8 illustrates the geometric restoration process where Figure 8(a) gives a camera document with geometric distortion. Based on the top line and base line and the identified VSB, the camera document is segmented into multiple image patches as shown in Figure 8(b). With the partitioned image patches, target rectangles are then constructed based on the enclosed characters and the specific character width height ratios as given in Figure 8(c). Finally, the camera document is restored based on the partitioned image patches and the constructed target rectangles. Figure 8(d) shows the restored image.

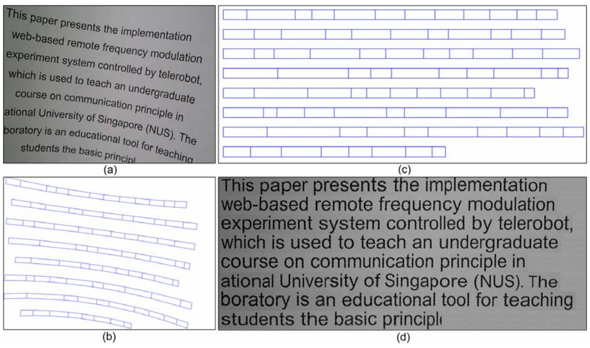

Figure 8(a) camera document with geometric distortion; (b) segmented image patches; (c) constructed target patches; (d) restored document image

3. Experimental Results

We implement the proposed technique based on the methods described above. The system is implemented in C++ and runs on a personal computer equipped with Window XP and Pentium 4 CPU. We evaluate the proposed technique with an image database that contains 90 camera documents with each 30 coupling with the skew, perspective and geometric distortions respectively. Experimental results show the proposed technique is able to restore the camera documents with three types of distortions efficiently.

We evaluate the proposed restoration technique based on the recognition rates of the document image after our proposed restoration operation. The OCR performance is tested using the software Omnipage Pro 14.0 [9]. The average recognition rates of 90 camera documents before the restoration operation are less than 10%. This result can be expected since the generic OCR systems can not deal with the perspective and geometric distortions well. At the same time, most of OCR systems perform poorly while the skew angle is bigger than 20˚. Figure 9 gives the experimental results where the recognition rates of three groups of images are illustrated with three types of curves labeled with pentagram, star, and diamond symbols respectively.

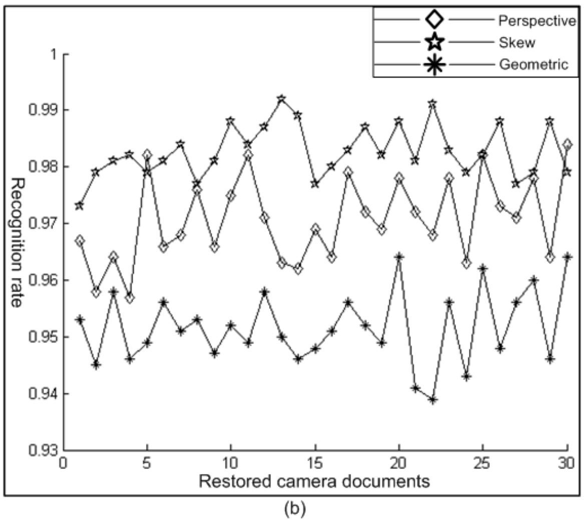

Figure 9: Recognition rate after the proposed restoration operation

As Figure 9 shows, the average recognition rate of the restored document images can reach over 95%. Such recognition rate shows that the proposed technique has the potential to be applied for the recognition of the distorted camera documents in practice. As the figure shows, the recognition rate of the camera document with skew and perspective distortion is normally a bit higher than that of the ones with geometric distortion. Such difference can be explained by the bigger errors introduced during the image segmentation process, which is required for the restoration of the geometric distortion.

Though the proposed technique is able to handle most of camera documents, some problems still exist.

Firstly, the proposed technique depends heavily on the resolution of the captured camera documents, as the required VSB component may not be identified properly from the camera documents with poor resolution. With the same reason, the proposed technique cannot handle the camera documents with arbitrary geometric distortion such as the crumpled paper sheets and the ones printed in handwritten text. Some new approaches will be investigated to solve these problems next.

4. Conclusion

In this paper, a unified document restoration technique is proposed to correct the document images with skew, perspective, and geometric distortions captured through a digital camera. The restoration of the camera documents is implemented through the exploitation of the vertical stroke boundary and the top line and base line of text lines. Different from the reported document restoration techniques that depend heavily on HDB, PF, and the auxiliary hardware equipments, the proposed technique needs only a single document image captured through a digital camera. Experimental results show that the proposed document restoration technique is fast, accurate, and robust.

References

[1] T. Akiyama and N. Hagita, "Automated Entry System for Printed Documents," *Pattern Recognition*, vol. 23, no.11, pp. 1141-1154, 1990

[2] D. S. Le, G. R. Thoma and H. Wechsler, "Automated Page Orientation and Skew angle Detection for Binary Document Image," *Pattern Recognition,* vol. 27, no. 10, pp. 1325-1344, 1994.

[3] P. Clark, M. Mirmehdi, "Recognizing text in real scenes," *International Journal of Document Analysis and Recognition*, vol. 4 no. 4, pp. 243–257, 2002

[4] C. R. Dance, "Perspective estimation for document images," *SPIE Conference on Document Recognition and Retrieval IX,* San Jose, USA, pp. 244–254, 2002.

[5] M. S. Brown and W. B. Seales, "Document restoration using 3D shape: a general deskewing algorithm for arbitrarily warped documents," *International Conference on Computer Vision,* Vancouver, Canada, pp. 117-124, 2001.

[6] A. Doncescu, A. Bouju and V. Quillet, "Former books digital processing: image warping," *Proceedings of Workshop on Document Image Analysis*, San Juan, Puerto Rico, pp. 5-9, 1997.

[7] 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,* vol. 23, pp. 541-553, 2005.

[8] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision,* Cambridge University Press, New York, 2000.

[9] http://www.scansoft.com/omnipage/