The Restoration of Camera Documents Through Image Segmentation

Shijian Lu and Chew Lim Tan

School of Computing, National University of Singapore, 117543, Singapore
{lusj, tancl}@comp.nus.edu.sg
http://www.comp.nus.edu.sg/labs/chime/

Abstract. This paper presents a document restoration technique that is able to flatten curled document images captured through a digital camera. The proposed method corrects camera images of documents through image partition, which divides distorted text lines into multiple small patches based on the identified vertical stroke boundary (VSB) and the fitted x-line and baseline of text lines. Target rectangles are then constructed through the exploitation of the characters enclosed within the partitioned image patches. With the constructed target rectangles and the partitioned image patches, global geometric distortion is finally removed through the local rectification of partitioned image patches one by one. Experimental results show that the proposed technique is fast, accurate, and easy for implementation.

1 Introduction

As camera resolution increases in recent years, high-speed non-contact document capture through a digital camera is opening up a new channel for document capturing and processing. Unfortunately, most documents such as the newspaper held in human hands and the book pages bound in a thick volume lie on a curled instead of planar surface. At the same time, the generic optical character recognition (OCR) systems cannot handle the camera text lying on a non-flat surface. Similar to the compensation of rotation induced skew introduced during the scanning process, geometric distortion resulting from the non-flat document surfaces must be removed before captured documents are fed to the OCR systems.

A number of document restoration techniques have been reported in the literature. The proposed techniques can be classified into two categories, namely, "*hard*" approaches [1, 2, 3] and "*soft*" approaches [4, 5, 6], respectively. The difference between the two approaches is whether auxiliary hardware is required or not. In [1], Pilu proposes to remove the geometric distortion using an applicable surface. This method needs the special laser devices to acquire 3D data before the restoration. Moreover, the correction process is quite slow, as the relaxation requires a large number of iteration to converge. In [2], Brown proposes to restore arbitrarily warped documents using the reconstructed 3D model. This method requires the structured lighting system and the complicated calibration process

to establish the depth map. In addition, the mapping of points on the non-flat surfaces to the ones on the planar images is still a problem. In [3], Yamashita et al propose to flatten the distorted documents through 3D modeling where a stereo vision system is set up for 3D measurements.

Instead of setting up some auxiliary devices, Agam and Wu [4] proposed a new technique that removes the geometric distortion through 3D mesh manipulation. The 3D mesh is constructed based on stereo disparity, which is built with multiple images captured from different viewpoints. Similarly, Cao et al propose to model the non-flat document using a cylindrical surface in [5]. The restoration equation is constructed through the exploitation of the camera imaging geometry and the cylinder directrix. Apart from the cylindrical modeling restriction, this method requires that generatrix of the cylinder model parallel to the image plane. In another work [6] by Liang, general curved documents can be flattened with just a single image captured through an uncalibrated digital camera. But the process is quite slow because the restoration involves the texture flow computation and the developable surface estimation.

In [7], we propose a rectification technique that restores camera documents with perspective distortion to a fronto-parallel view. The rectification is accomplished through the exploitation of the vertical stroke boundary (VSB), x-line and baseline as labeled with (1) and (2) in Figure 1, which are determined using a few fuzzy sets and morphological operators. In this paper, we extend our earlier work to restore the camera documents where texts lie on a non-flat instead of planar surface. We focus on the documents where texts lie on a smoothly curved surface and text line orientations can be modeled using a cubic polynomial. The proposed restoration technique needs no camera calibration, no auxiliary hardware [1, 2, 3] or 3D reconstruction from multiple images [4]. The only thing required is a single document image captured through a common digital camera. In addition, the method assumes no specific 3D surface model as did in [5]. It is able to handle the camera documents captured from different viewpoints provided that the angle between document normal and camera optical axis lies within a reasonable range.
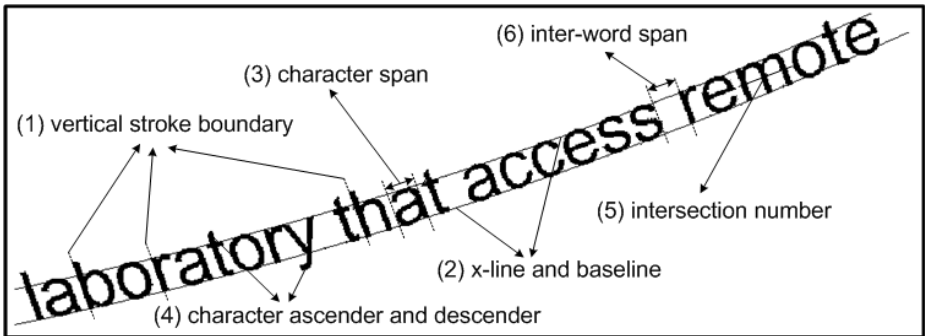

Fig. 1. Text line definition

We propose to restore the camera documents with perspective and geometric distortion through image partition, which divides distorted text lines to multiple small image patches where text can be approximated to lie on a planar surface. The partition is accomplished through the exploitation of the x-line and baseline and the VSBs. For each partitioned image patch, a target rectangle is constructed based on the number and the aspect ratios of enclosed characters. The character aspect ratios are determined based on character span, character ascender and descender, and character intersection numbers as labeled with (3), (4), and (5) in Figure 1. With the constructed quadrilateral correspondences, the global geometric distortion is finally removed through the local rectification of the partitioned image patches one by one.

2   The Proposed Approach

This section presents the proposed document restoration technique. In particular, we divide the description into three subsections, which deal with the camera document partition, the target rectangle construction, and the final camera document restoration respectively.

2.1   Camera Document Partition

The document partition process can be divided into two steps. The first step partitions the distorted document images to multiple text lines based on the x-line and baseline. Combined with the identified VSBs, the second step further divides the partitioned text lines into multiple smaller patches where text can be approximated to lie on a planar surface.

Text lines can be partitioned through the exploitation of the character tip points as proposed in [7]. For text lying over a smoothly curved surface, the orientation of text lines can be modeled well with a cubic polynomial in most cases. We therefore choose a polynomial instead of straight line to fit the extracted character tip points. For distorted word image given in Figure 2(a), Figure 2(b) shows the extracted character tip points near the x-line and baseline positions. Figure 2(c) shows the fitted x-line and baseline that are fitted using the classified character tip points.

Partitioned text lines can be further divided into multiple smaller patches based on the VSBs proposed in [7] where VSBs are identified with a few fuzzy sets and aggregation operators. For sample word image "laboratory" given in Figure 2(a), Figure 2(d) shows the VSBs identified from the left side of character strokes. Before the text line division, the identified VSBs must be processed further to facilitate the document partition and later restoration process. Firstly, to restrict the number of characters within each partitioned image patch, some VSBs must be deleted if they are too close to their left adjacent neighbor within the same text line. In our proposed technique, the distance threshold is determined as:

$$D_{thre} = k_d \cdot VSB_{avg} \tag{1}$$

where parameter VBS_{avg} represents the average length of identified VSBs, which normally reflects the size of the captured characters. Parameter k_d is designed to adjust the width of the partitioned image patches and we set it at 3 so that each image patches enclose as least three characters.

Secondly, for text lines that have no VSBs identified at their left or right end positions, a VSB must be estimated there to enclose all texts that belong to the processed text line. The orientation of the VSB at the text line end position can be determined through the linear interpolation:

$$slp = slp'' + \frac{(x - x'') \cdot (slp' - slp'')}{x' - x''} \qquad (2)$$

where x is x coordinates of the leftmost or rightmost text pixel within the studied text line. x', x'' are x coordinates of the centroids of the two VSBs nearest to x. Parameters slp' and slp'' denote the slopes of the straight lines fitted based on that two nearest VSBs. Therefore, the VSBs at the text line end positions can be estimated as the straight line segments that pass through x with orientation determined using Equation (2).

Figure 2(e) shows the processed VSBs where the second VSB in Figure 2(d) is deleted to ensure the width of the partitioned image patches. The rightmost VSB given in Figure 2(e) is constructed to enclose all text within the processed
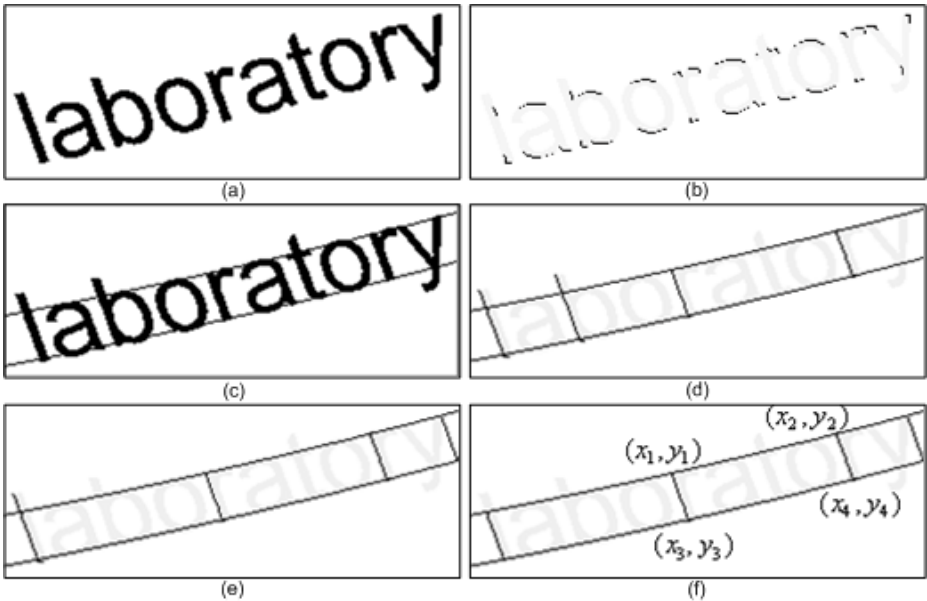
Fig. 2. Document image partition: (a) distorted document text; (b) extracted character tip points; (c) fitted x-line and baseline of text line; (d) identified VSBs; (c) processed VSBs; (d) partition result

text line. Some vertical lines can thus be fitted based on the processed VSBs. Combining the fitted x-line and baseline of the text lines, the distorted word image given in Figure 2(a) can thus be partitioned into three small image patches as shown in Figure 2(f).

2.2   Target Rectangle Construction

For each partitioned image patch, a target rectangle correspondence must be constructed within the target image to rectify that partitioned image patch. We propose to classify characters to six categories with six different aspect ratios. Characters are classified based on the features including character span, character ascender and descender, and character intersection numbers as shown in Figure 1.

Character span is defined as the distance between two parallel straight lines tangent to the left and right sides of character with the orientation same as that of the nearest VSB. Character ascender and descender can be determined based on the distance between the topmost and lowermost character pixels and the x-line and baseline. The last feature, character intersection number, defines the number of intersection between character strokes and the straight line that passes through the character centroid with orientation orthogonal to that of the nearest VSB.

The character classification algorithm can be formalized as follows:

Inputs: Character spans $CSpan$; Character ascender and descender
              $ADInfo$; Intersection numbers $Inter$
Procedure: CC($CSpan$, $ADInfo$, $Inter$)
  1) Initialize i = 1
  2) Calculate average character span $CSpan_{avg}$ based on the
     determined $CSpan$.
  3) If $Inter$(i)≥ 3 and $ADInfor$=1(with ascender), character is
     classified as $M, W, @$.
  4) Else if $Inter$(i) ≥ 3 and $ADInfor$=0(no ascender),
     character is classified as m, w.
  5) Else if $ADInfor$=1(with ascender) and $CSpan$(i) $\geq k_u \cdot CSpan_{avg}$,
     character is classified as $A - H, J - L, N - V, or X - Z$.
  6) Else if $CSpan$(i) $\geq k_l \cdot CSpan_{avg}$ and $CSpan$(i) $\leq k_l \cdot CSpan_{avg}$,
     character is classified as, $a - e, g - h, n - q, u - v, x - z, 2 - 9, \dot{e}, \acute{u} \ldots$
  7) Else if $CSpan$(i) $\leq k_s \cdot CSpan_{avg}$, character is classified as $i, l, I, (, ! \ldots,$
     or j.
  8) Else, character is classified as $t, f, or r \ldots$.
  9) i = i + 1

To classify characters into six categories, the average character span $CSpan_{avg}$ in Step 2) is firstly calculated. Parameter k_u, k_l, and k_s in Steps 5), 6), and 7) are three key parameters for character categorization and they are determined as 1.2, 0.7, and 0.3 based on the observation of character aspect ratios in different categories. We have tested 30 camera documents and the categorization rate reaches over 95%. We note that small classification errors will not lead to

Table 1. Restoration and recognition results (IS: image size; CON: number of characters; ET: execution time; RRB: recognition rate before restoration; RRA: recognition rate after restoration)

Classified character categories	Character aspect ratios (R)
M, W, @$\cdots$	1.6 : 1
m, w	1.4 : 1
A-H, J-L, N-V, X-Z $\cdots$	1.2 : 1
a-e, g-h, k, n-q, s, u-v, x-z, 0, 2-9, #, â, ë $\cdots$	0.8 : 1
t, f, r, - $\cdots$	0.5 : 1
i, j, l, I, 1, (,), !, [, —, { $\cdots$	0.2 : 1

the obvious restoration and later recognition errors because the partitioned document patches normally contain no less than three characters as determined by Equation (1).

Based on the proposed character classification technique, all characters are grouped to six categories. Table 1 shows six character categories together with the related aspect ratios.

Similar to character spans, the inter-word blanks as labeled with (6) in Figure 1 must be detected as well to estimate the aspect ratio of target rectangles. The inter-word blanks can be simply located based on the distance between the adjacent characters within, which is normally much less than two VBS_{avg} for adjacent characters within the same word. The contribution of the inter-word blanks to the width of target rectangle can thus be determined based on the relation between the size of inter-word blanks as shown in Figure 1 and the average character span $CSpan_{avg}$.

With VBS_{avg} as the height of target rectangles, the width of target rectangles can thus be determined as:

$$T_w = \sum_{i=1}^{n} R_i \cdot VSB_{avg} \tag{3}$$

where VBS_{avg} represents the average length of the identified VSBs and parameter n represents the number of characters and inter-word blanks within the partitioned image patch. Parameter R_i refers to the i_{th} aspect ratio of characters and inter-word blanks enclosed.
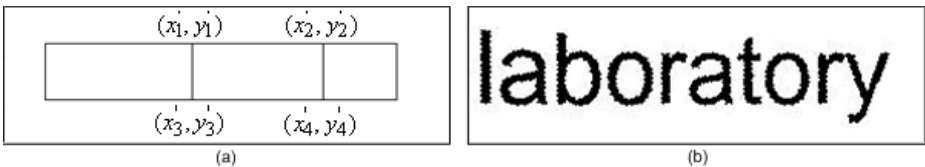
Fig. 3. Document restoration: (a) constructed target rectangles; (b) rectified text

Characters enclosed within the partitioned image patches can be easily located based on the relative position between character centroids and the fitted x-line, baseline, and two straight line segments fitted using the identified VSBs. Target rectangles can thus be restored with the pre-determined height (VBS_{avg}) and the restored width as given in Equation (3). For partitioned image patches given in Figure 2(d), Figure 3(a) shows the corresponding restored target rectangles, which lie on a horizontal line side by side.

2.3   Camera Document Restoration

With the established correspondences between partitioned image patches given in Figure 2(d) and the target rectangles given in Figure 3(a), the distorted camera documents can be restored patch by patch using the rectification homography estimated based on four point mapping algorithm. The rectification homography for a specific pair of quadrilaterals can be estimated as:

$$H = A^{-1} \cdot R \tag{4}$$

where H is the homography matrix and matrixes A, R are constructed using four point correspondences.

$$
A =
\begin{pmatrix}
-x_1 & -y_1 & -1 & 0 & 0 & 0 & x_1'x_1 & x_1'y_1 \\
0 & 0 & 0 & -x_1 & -y_1 & -1 & y_1'x_1 & y_1'y_1 \\
-x_2 & -y_2 & -1 & 0 & 0 & 0 & x_2'x_2 & x_2'y_2 \\
0 & 0 & 0 & -x_2 & -y_2 & -1 & y_2'x_2 & y_2'y_2 \\
-x_3 & -y_3 & -1 & 0 & 0 & 0 & x_3'x_3 & x_3'y_3 \\
0 & 0 & 0 & -x_3 & -y_3 & -1 & y_3'x_3 & y_3'y_3 \\
-x_4 & -y_4 & -1 & 0 & 0 & 0 & x_4'x_4 & x_4'y_4 \\
0 & 0 & 0 & -x_4 & -y_4 & -1 & y_4'x_4 & y_4'y_4
\end{pmatrix}
\quad
H =
\begin{pmatrix}
h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32}
\end{pmatrix}
\quad
R =
\begin{pmatrix}
x_1' \\ y_1' \\ x_2' \\ y_2' \\ x_3' \\ y_3' \\ x_4' \\ y_4'
\end{pmatrix}
$$

where the 3 homography matrix is expressed in a vector form and h33 is equal to 1 under homogeneous frame. Four point correspondences (x_i, y_i), (x_i', y_i'), i = 1...4, are taken as the four vertices of the partitioned document patch and the target rectangle as labeled in Figure 2(d) and Figure 3(a). The distorted sample word can thus be restored using the rectification homography and Figure 3(b) shows the restoration result.

Figure 4 illustrates the document restoration process where the distorted camera document given in Figure 4(a) is captured through a digital camera. Figure 4(b) and (c) show the identified VSBs and the fitted x-lines and baselines of text lines. With the VSBs and the x-line and baseline, distorted camera document is then partitioned into small patches as given in Figure 4(d). Target rectangles corresponding to the partitioned image patches are accordingly constructed based on the enclosed characters. As Figure 4(e) shows, target rectangles are arranged horizontally line by line where the bottom edge of all rectangles of the same text line lies on the same horizontal line and the adjacent rectangles share a common vertical edge. With partitioned image patches and the constructed target rectangles, distorted camera document
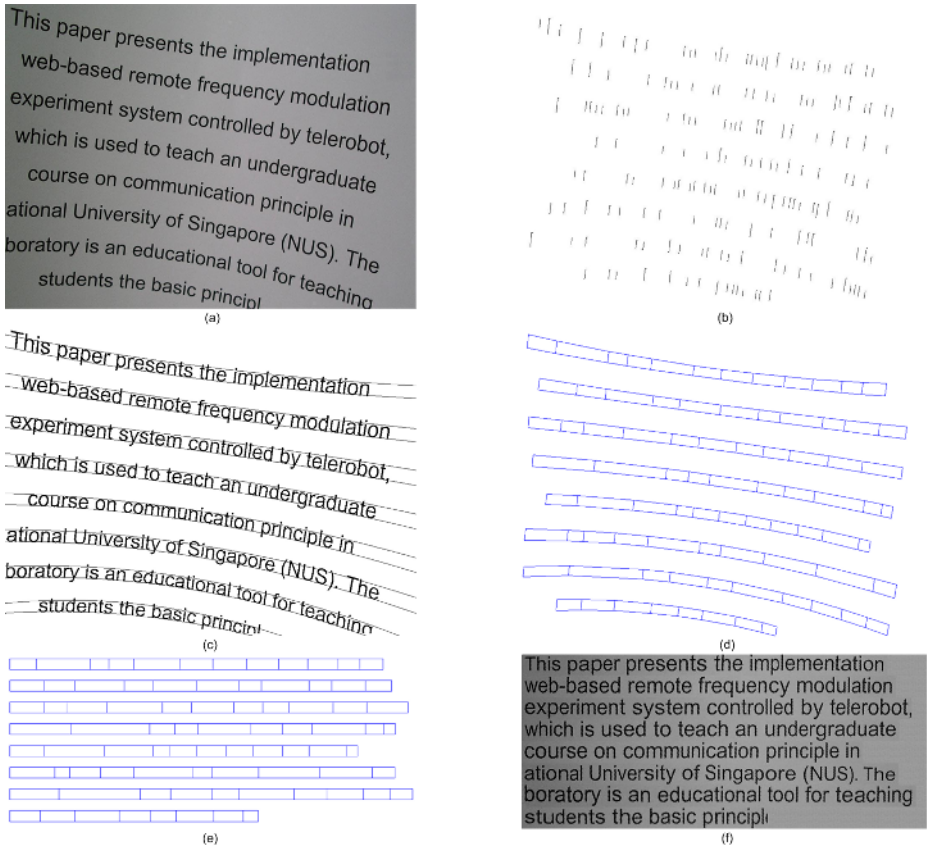
Fig. 4. Perspective and geometric distortion rectification: (a) distorted camera document; (b) identified VSBs; (c) fitted x-lines and baselines; (d) divided image patches; (e) constructed target rectangles; (f) restored document image

given in Figure 4(a) is finally restored patch by patch and Figure 4(f) shows the restored document image.

It should be clarified that character ascender and descender are actually above or below the divided image patches as shown in Figure 4(d). To rectify character ascender and descender correctly, the transformation must be extended above and below a bit beyond the constructed target rectangles as shown in Figure 4(e). The extension range is defined as:

$$E = k_e \cdot VSB_{avg} \tag{5}$$

where VBS_{avg} represents the average length of the identified VSBs. Parameter k_e is designed to adjust the extension range. We determined it as 0.5 in our implemented system so that the transformation is able to cover character ascender and descender and at the same time, it will not reach adjacent text lines.

3   Experimental Results

We have implemented the proposed restoration technique described in Section 2. Experimental results show that the proposed method can restore camera documents with perspective and geometric distortions efficiently. The programs are written in C++ and run on a personal computer equipped with Window XP and Pentium 4 CPU. The system was evaluated with an image database that contains 30 distorted camera documents captured from different distances and viewpoints. At the present stage, the average rectification process takes around 2-3 seconds for document images with size 640×480.

As most researchers rely on 3D measurement devices or multiple images for 3D reconstruction, it is difficult to compare our method with theirs. We therefore evaluate the performance of the proposed method based on the recognition rate of restored document images. In our experiments, 30 sample documents from different sources including newspaper, journal articles, and books are utilized for OCR testing. Documents are firstly captured and digitalized to the electronic images through a digital camera. The camera images are then restored using our proposed restoration technique. Lastly, the restored document images are fed to the OCR software for recognition. We test the recognition performance using Omnipage Pro [8], one of the best OCR software available in the market.

Distorted document images are firstly tested using Omnipage and the average recognition rate is below 10%. This result can be expected, as the generic OCR systems cannot handle the camera documents with perspective and geometric distortions well. Figure 5 shows the recognition results of the 30 restored document images. As the figure shows, the recognition performance of the restored
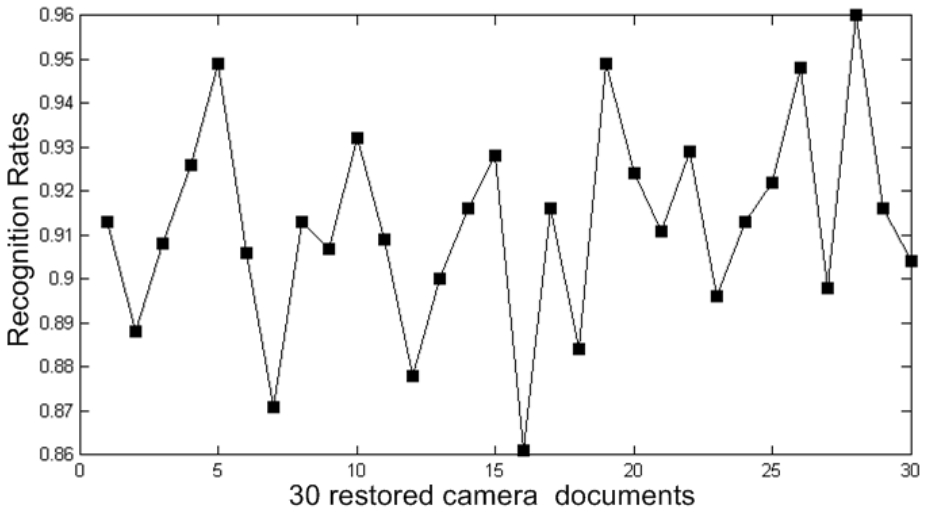

Fig. 5. Recognition rates of the restored document images

image is improved greatly and the average recognition rate reaches over 90%. Even for the worst cases where document may not be partitioned and restored nicely, the recognition rate still reaches over 80%. This is mainly because that the OCR systems are normally tolerant of some minor distortion such as the slight skew and perspective distortions.

The proposed restoration technique is also able to handle the camera documents with only perspective distortion where text lies on a planar instead of curved document surface. Unlike those reported perspective rectification techniques that rely heavily on the document boundary or specific paragraph formatting information [9, 10], our method requires only the VSBs and the x-line and baseline that can be extracted from character strokes directly. Perspective distortion can be removed based on the partitioned image patches and the constructed target rectangles as described in Section 2. For the camera document with perspective distortion given in Figure 6(a), Figure 6(b) shows the restored document image where the bounding box labels the falsely classified characters.
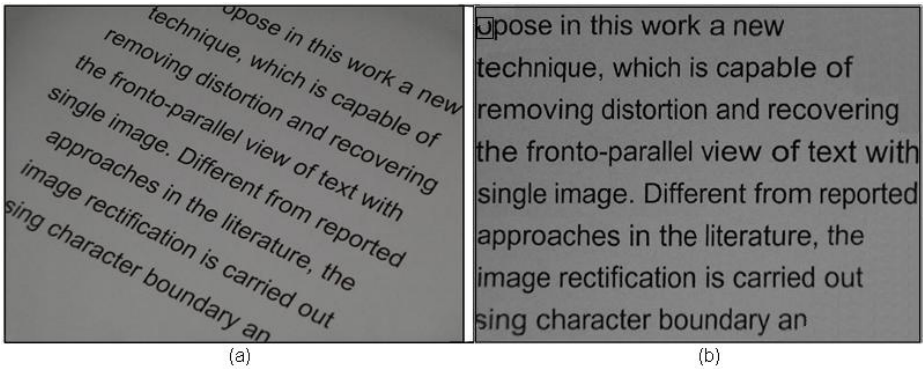
Fig. 6. Experiment results: (a) distorted document image; (b) rectified document image

As the proposed technique relies on the VSBs, the x-line and baseline for document restoration, it is able to handle the document text printed in most frequently used fonts. For italic texts, VSBs can still be identified though they do not give the vertical direction. The proposed method may fail when the size of the captured characters is too small. We test some document images with small characters and experiments show that recognition results deteriorate quickly while the average VSBs size becomes smaller than 12 pixels. This can be explained by the fact that VSBs may not be identified properly while characters become too small. Furthermore, the vertical direction cannot be estimated accurately even if the VSBs are correctly identified from characters with small size. This problem can be remedied through the image interpolation, which enlarges the image when the captured characters are too small.

Furthermore, the restoration may fail when the distortion angle between the document normal and camera optical axis is too big. For the 30 document samples we tested, the distortion angle is all within 45 degrees. Experiments show

the restoration may fail when the distortion angle becomes bigger. In fact, even human eyes cannot read the text correctly while the distortion angle is bigger than 70 degrees. With the similar reason, the proposed method cannot restore the camera documents with arbitrary geometric distortion such as the image of the creased documents. We will work on the restoration of arbitrarily distorted documents next.

4   Conclusion

In this paper, a computationally efficient technique is proposed to restore the document images with perspective and geometric distortion captured through a digital camera. The restoration is carried out through the image partition, which is implemented based on the identified VSBs and the x-line and baseline of text lines. Different from reported rectification methods that depend heavily on some auxiliary hardware or complicated 3D reconstruction process with multiple images captured from different viewpoints, the proposed rectification technique needs only a single document image captured by a digital camera. Experimental results show that rectification process is fast and easy for implementation.

With a digital camera, the proposed document restoration technique may open a new channel for document capture and understanding. Furthermore, with a little adaptation, it may be applied to some other portable devices such as the digital camera, the mobile phone and the personal digital assistant (PDA) for document capture and management. As a result, these devices embedded with camera sensor need only to store and transmit recognized ASCII text instead of the huge document images.

References

1. M. Pilu, Undoing Paper Curl Distortion Using Applicable Surfaces, International Conference on Computer Vision and Pattern Recognition, Kauai, USA, page 67–72, 2001.
2. M. S. Brown, W. B. Seales, Document restoration using 3D shape: a general deskewing algorithm for arbitrarily warped documents, International Conference on Computer Vision, vol. 2, July 2001, Vancouver, Canada, pp. 367–374.
3. A.Yamashita, A. Kawarago, T. Kaneko, K. T. Miura, Shape Reconstruction and Image Resto-ration for Non-Flat Surfaces of Documents with a Stereo Vision System, International Con-ference on Pattern Recognition, vol. 1, August 2004, Cambridge, UK, page 482–485.
4. G. Agam and C. H. Wu Structural rectification of non-planar document images: application to graphics recognition, Fourth International Workshop on Graphics Recognition Algorithms and Applications, Kingston, Ontario, Canada, pp. 289–298, 2001.
5. H. Cao, X. Ding, C. Liu, A Cylindrical Surface Model to Rectify the Bound Document Image, Ninth IEEE International Conference on Computer Vision, vol. 1, 2003, Nice, France, pp. 228–233.

6. J. Liang, D. DeMenthon, and D. Doermann, Flattening curved documents in images, Interna-tional Conference on Computer Vision and Pattern Recognition, June, 2005, San Diego, USA, pp. 338–345.
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. http://www.scansoft.com/omnipage/.
9. C.R. Dance, Perspective estimation for document images, Proceedings of the SPIE Confer-ence on Document Recognition and Retrieval IX, 2002, pp. 244–254.
10. P. Clark, M. Mirmhedi, Rectifying perspective views of text in 3Dscenes using vanishing points, Pattern Recognition, vol. 36, pp. 2673–2686, 2003.