DEWARPING AND DESKEWING OF A DOCUMENT USING AFFINE TRANSFORMATION

Honey Kansal, Sudip Sanyal and Deepali Gupta
IIIT-Allahabad, Allahabad, Uttarpradesh, India

Keywords: Dewarping, Deskewing, Deshadowing, Noise in scanned documents, Affine transformations.

Abstract: An approach based on affine transformations is applied to solve the problem of dewarping of scanned text images. The technique is script independent and does not make any assumptions about the nature of the text image or the nature of warping. The attendant problems of deskewing and deshadowing are also dealt with using a vertical projection technique and filtering technique respectively. Experiments were performed on scanned text images with varying font sizes, shapes and from various scripts with varying degrees of warp, skew and shadow. The proposed method was found to give good results on all the text images, thus demonstrating the effect of the approach.

1 INTRODUCTION

While scanning a thick volume, a major problem often faced is warping along the spine of the volume which hampers the readability of the scanned text image. Also, if we want to process the text image consisting of such problems with an OCR, then the performance of the process degrades considerably. The problems discussed here can also be seen while photocopying a thick volume. This paper proposes an effective method to solve this problem in a manner that is independent of the script and also does not make any assumptions about the nature of the distortions.

Figure 1: Document image with warp, skew and shadow along the spines.

Warping refers to non-uniform curvature generated near the spine and it produces a distorted text image. The problem is compounded by shadowing and skewness. The cause of warping can be understood if we examine the optical model of a scanner. It captures the image of the document by sending the light rays over the whole page. Now, if the page is perfectly flat then the intensity of light is uniform across the whole area. However, when we consider a thick volume, the intensity of incident light is not uniform and consistent. The cause of the non-uniformity is the curvedness of the page at the 'spines'. This leads to two effects. The first is that some dark region is generated at the spines, which is termed as a shadow which is a photometric distortion. The second effect is "geometric distortion" which leads to bending or curving of the text lines in the scanned image. These lead to the problems of skewing and warping respectively.

In Figure 1 we show an example where all three problems are present. In the general case both warp and skew distortions exist together. However, most approaches applied till now discussed removal of warping assuming the absence of skewing or vice versa. There are mainly two classes of approach applied for dewarping. The first is 2D image processing technique as described in (Zhang, 2003), (Kakamanu, 2006), (Cao, 2003), (Ezaki, 2005), (Gatos, 2007) and the other is restoration of 3D document shapes as described in (Zhang, 2004), (Zhang, 2005), (Chew, 2006). The latter approach requires hardware like special camera setup which entails a high cost. In this work we follow the first approach. Zheng et al (Zhang, 2003) proposed a technique based on connected component analysis and regression of curved text lines. It was an

Kansal H., Sanyal S. and Gupta D. (2009).
DEWARPING AND DESKEWING OF A DOCUMENT USING AFFINE TRANSFORMATION.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 73-78
DOI: 10.5220/0001798100730078

extension of their previous work (Zhang, 2001) and used a new resolution free restoration system with a simpler connected component analysis. The technique was free from various resolution parameters although it makes a significant assumption of negligible skewing. Estimation of robust text lines was done in (Kakamanu, 2006) using the cue that the text lies on the surface of the page and are straight. Their approach assumed simple background which can be easily separated without considering shadowing at spines. Cao et al (Cao, 2003) assumed a cylindrical model that can be fitted over the warped document. They estimated the bending extent of the surface by extracting the horizontal baselines and then fitted a curve over the warped text. This assumption is not valid for all warped document, especially those with skew. A similar technique was proposed without any consideration of cylindrical model by Ezaki et al (Ezaki, 2005) by fitting a model in order to estimate the warp of each text line by fitting an elastic curve to the text line. Another approach was based on segmentation (Gatos, 2005) which detects all the words using image smoothing. Then the lines were identified using connected component labelling.

In general most of the techniques described above neglect skewing which is often present in the warped document and may be present independent of any warping. Also, most of the earlier approaches ignore the presence of shadowing. Thus, a number of techniques proposed by earlier workers will not work properly in the presence of skewing. On the other hand, a number of approaches have been proposed considering skewing as an individual problem without taking into account the problem of warping. Dhandra and his colleagues in their work (Dhandra, 2006) removed the skewing using image dilation and region labelling techniques. They calculated the average of all the angles at which each labelled region was tilted and then found the resultant skew angle. Another approach, based on least square method and saw tooth algorithm, was used to calculate the skew angle by Yu et al (Yu, 1995). Another method proposed by Ballard et al (Ballard, 1982) uses 2D Fourier Transform to estimate the skew angle. It was computationally easy and appropriate for uniformly distributed text. However, in the presence of warping the degree of alignment changes continuously as one moves over the warped line because of which the FFT method fails.This non-uniformity present across the text forced us to find a technique that should work on local distribution of the text. So we have opted for affine transformations as our approach for the removal of warping.

We have formulated our approach using as few assumptions as possible. Moreover, our technique handles possible shadowing, skewing and warping of the text image. We follow the approach of vertical projection in order to remove skewing after dilation of the shadow free document. Deskewing is followed by dewarping using local affine transformation of the segmented words after finding the angle at which each word can be considered to be warped. In the following section we describe our approach. In the third section we present the results of the experiments that were performed in order to test our approach. Some directions for future improvements are presented in the concluding section.

2 SEQUENTIAL RESTORATION OF DISTORTED DOCUMENT

The scanned document of a thick volume consists of warping along with skewing and shadowing at the spines. The steps can be listed as:

a) Deshadowing in order to get the text region which we call as foreground and background separately.
b) Deskewing of the binary document after dilating the image uniformly.
c) Removal of warping from the deskewed document using local affine transformation.

The significant factor of our work is the consideration of a binary image that contains less information as compared to grey scale or colour. Without loss of generality, we will take white as background and black as foreground. In the following we describe each part of the process.

2.1 Deshadowing

Shadowing is a photometric effect which is independent of other two distortions. This paper proposes a filtering technique for deshadowing which will eliminate the noise spread that is present mainly near the spine as shown in Figure 2a. An examination of Figure 2a shows that pixel density can be used as a feature for distinguishing foreground from background. In all cases that we examined, the pixel density of the text area is significantly larger than all other regions, including the region under the shadow.

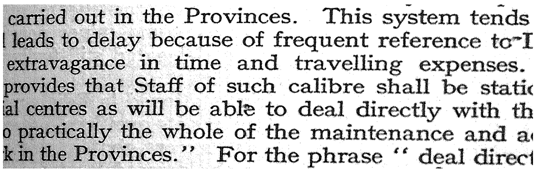
carried out in the Provinces. This system tends
l leads to delay because of frequent reference to I
extravagance in time and travelling expenses.
provides that Staff of such calibre shall be static
al centres as will be able to deal directly with th
o practically the whole of the maintenance and a
k in the Provinces." For the phrase " deal direct

Figure 2(a): A sample showing the shadow effect at the spines.

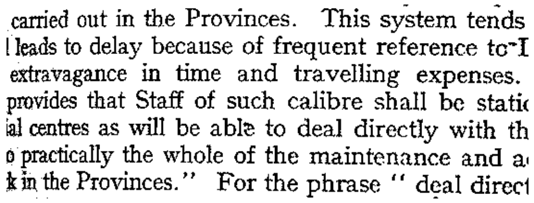
carried out in the Provinces. This system tends
l leads to delay because of frequent reference to I
extravagance in time and travelling expenses.
provides that Staff of such calibre shall be static
al centres as will be able to deal directly with th
o practically the whole of the maintenance and a
k in the Provinces." For the phrase " deal direct

Figure 2(b): Effect of deshadowing.

To estimate the pixel density we take a square window of size less than stroke width of the text and move it horizontally and vertically over the whole image. The pixel density in the window is calculated using eqn. (1) which is used to determine the true pixel value for the central pixel of the window. If the density of pixels in a given window is greater than some threshold then we mark the corresponding central pixel of the window as black.

$$\text{Pixel density} = \frac{\text{No of pixels of specified colour}}{\text{Total No of pixels}} \quad (1)$$

The threshold is determined empirically after observing a large number of documents. We find that a pixel density of > 0.9 signifies that the central pixel is black. In Figure 2b we show the effect of using the above technique. As can be observed, the shadow is removed quite effectively with the proposed technique.

2.2 Deskewing

After deshadowing we approach the first category of geometric distortion i.e. skewing. Our motive is to find out the skew angle at which whole text is tilted and rotate the text by this angle. We have examined two approaches for determining the skew angle. The first was based on 2D Fourier transforms (Ballard, 1982) and another using a histogram technique of local angles obtained using vertical projections. The latter approach is slower than the former but gives higher accuracy.

2.2.1 Problem with FFT Approach

One of the main objectives of our work was to make the technique script independent. Scripts can be divided into mainly two categories, either consisting of upper and lower modifier as in Devanagari and other Indian scripts where the vowels are represented as upper or lower modifiers or scripts free from upper and lower modifier like English.

We created a collection of 10 scanned documents of different scripts. An artificial skew of range 2°-15° was introduced in them. While testing the approach over the document for scripts like English, the difference of calculated skew angle and exact skew angle came to be within the range of 0.02°-0.3°. On the other hand, when the same technique is applied to Indian scripts that have upper and lower vowel modifiers, we found that this method has an error in the range of 0.5-3°.

While the error does not appear alarming at first glance, it is sufficient to cause line segmentation errors for OCR. Thus we had to device a new approach for the calculation of skew angle.

2.2.2 Vertical Projection Histogram Technique

The first step is to dilate the whole image to the extent such that different lines become a single object. For this purpose, we will choose a rectangular window for dilating and make the whole window black instead of central pixel. The effect of dilation is shown in Figure 3 below. Since the skew angle is a property of the entire page, therefore only the topmost line is considered for its determination. The next step is to find the first foreground pixel while moving in the vertical direction. This is performed at fixed horizontal gaps. The grey vertical lines in Figure 3 show these vertical lines. The lengths of the grey lines are the starting points of the text area. This is similar to drawing 'vertical projection' at some fixed interval and finding the topmost point of intersection of projection and the dilated image. Let the points of intersection be labelled as (x_i, y_i). We can get a local estimate of the skew angle as

$$\theta_i = \tan^{-1} ((y_{i+1} - y_i) / (x_{i+1} - x_i)) \quad (2)$$

In practice, there will be some variations due to the non-uniform heights of the original characters in the text as well as due to warping. Therefore, we compute the histogram of the local skew angles. The histogram has a sharp and clearly identifiable peak corresponding to the skew angle. Thus, the skew angle for the entire page can be obtained as the median of the local skew angles. The advantage of this technique is that it gives a control over local

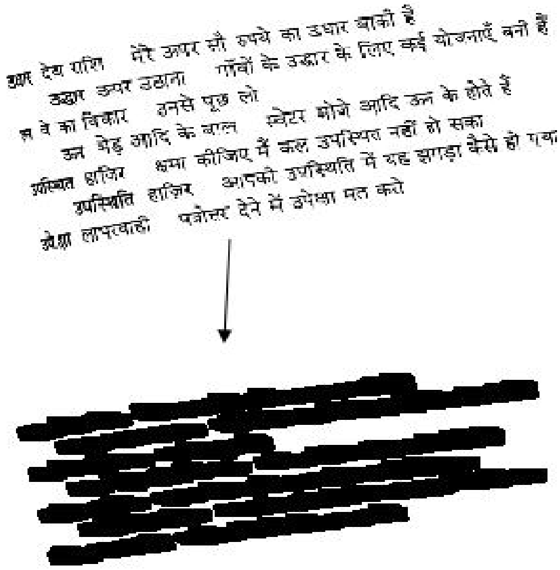

Figure 3: Document before and after dilation.

variation in the skewing. Moreover, the process works well even in the presence of warping. The horizontal gap between vertical projections should be chosen with extreme care. For a normal image of size 5000 * 7000 pixels, we tested with 50 pixels as this parameter and got good results even for a complex script like Devanagari which has lower and upper modifiers along with normal characters. The final output of deskewing is shown in Figure 4.

2.3 Dewarping

A warped text can be considered as a set of words that have been rotated and translated from their original position. Thus, our basic approach to dewarping is to estimate the extent of rotation and translation of the original word and perform the corresponding inverse transformations.

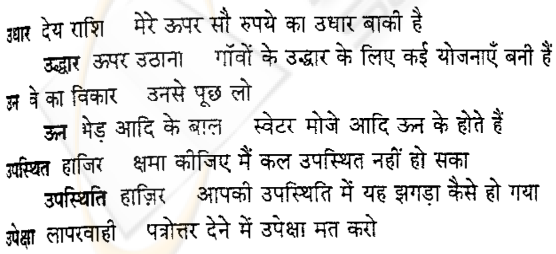

Figure 4: Output after de-skewing.

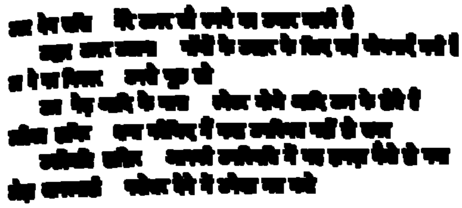

Figure 5: Dilated deskewed output.

Let us consider a left warped document (as in Figure 5) in order to make the approach clear.

2.3.1 Dilation

The first step in this process is to estimate the extent of rotation and translation of each word. Thus, we first have to identify each word by dilating the document in a manner similar to that adopted for deskewing but to different extent such that such that each word of the text document becomes an object as shown in Figure 5.

2.3.2 Object Labeling

The next task is to label the words and sort them in such a way that we can identify the line to which different words belong and obtain the sequence of words in a line. A very straightforward approach is employed in the present work for object labelling (Fang, 1987). Moving in a direction from right to left (for left warped documents) we locate the first foreground pixel and mark it with a label. The foreground pixels that are connected to the marked pixel are then assigned the same label. This process continues recursively till all connected components of the object get the same label. The label for the next object found is incremented by one and the process is repeated till all objects get labelled. The coordinates of the bounding rectangle of each object is also noted. The problem with this labelling algorithm is that it does not label the objects in a sorted order. The labelling algorithm can assign labels to objects with numbers that may have large differences between adjacent objects. For further processing, each line should have its objects labelled in increasing order from right to left and consecutive objects marked with consecutive integers. This post labelling work is summarized next.

1) Sort objects in decreasing order according to their rightmost x-coordinate and store them in a list S.

2) Pickup the topmost object from S which is the rightmost word of the line that is processed first. Mark it as <i, L> where L is line number and i=1,2, ….n where n is number of words in this line L.

3) For next word of this line, traverse S from top to bottom and search for all objects (not marked) having at least 85% overlap in Y direction with the currently marked word.The first word of S, while traversing from top to bottom, that belongs to the list obtained from step 3 is the next word in the line.
5) This word is marked and steps 3 and 4 are repeated till all words are marked.

2.3.3 Local Affine Transformation

The word "affine transformation" here means all the transformations done to the pixels are based on some kind for 'affinity' or 'relation' of them with their neighbouring pixels. After correctly labelling each word, we can now find the extent of translation and rotation that each word has been subjected to, due to warping. In order to achieve this we estimate the transformation required to dewarp each word.
1) Obtain the horizontal profile of each word W_i of the line,traverse vertically downward in the word W_i and find out the Y-level Y_i having horizontal profile value as at least 90% of the width of word.
2) Note the X-coordinate, X_i, of leftmost pixel of this Y-level Y_i. Do steps 1-3 for all words.
3) Completion of step 3 yields the coordinates (X_1, Y_1), (X_2, Y_2) … (X_n, Y_n) where n is number of words in this line and (X_1, Y_1) is the coordinate corresponding to the rightmost word W_1.
4) Now for each word W_i (i=2,3,….,n) calculate the local angle Θ_i with which it should be rotated by calculating the slope of straight line joining the points (X_i,Y_i) and (X_{i-1},Y_{i-1}) using (2).
5) Rotate each word W_i of the original image (not dilated) with Θ_i keeping W_1 fixed and write them in a new temporary image. This step effectively removes the rotation effect of warping. The translation effect is removed in the following steps.
6) Repeat dilation and object labelling, for this new image. Then perform step 1 and 2 to get Y_i (i=1,2,….,n) for all words.
7) Translate each word W_i in vertical upward direction by $|Y_i-Y_1|$ again keeping W_1 fixed.
8) Repeat steps 1-8 for all lines to get the dewarped image as shown in Figure 6.

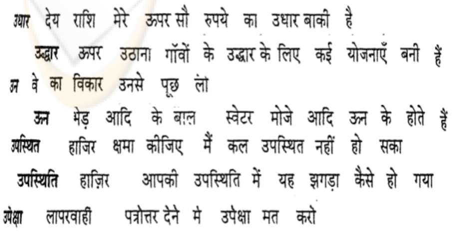
उधार देय राशि मेरे ऊपर सौ रुपये का उधार बाकी है
उद्धार ऊपर उठाना गॉवों के उद्धार के लिए कई योजनाएँ बनी हैं
ऊन वे का विकार उनसे पूछ लो
ऊन भेड़ आदि के बाल स्वेटर मोजे आदि ऊन के होते हैं
उपस्थित हाजिर क्षमा कीजिए मैं कल उपस्थित नहीं हो सका
उपस्थिति हाज़िर आपकी उपस्थिति में यह झगड़ा कैसे हो गया
उपेक्षा लापरवाही पत्रोत्तर देने मे उपेक्षा मत करो

Figure 6: Final output free from skewing and warping.

The results of our approach are summarized in the next section.

3 RESULTS

In order to show the result of the approaches there is no proper method prescribed other than visualizing the output. Generally the accuracy of skewing and warping is tested over the performance of OCR (Zhang, 2003; Kakamanu, 2006; Cao, 2003) since the words of a warped document will not be correctly recognized by an OCR. The technique we discussed is mainly based on local transformation of words after calculating their degree of rotation and translation. Therefore we have used a word based accuracy measure. This measure can be described using the Figures 5 and 6. The document, Figure 5, consists of 70 words warped at different angles. On reviewing the output of Figure 6, we can analyze the error in calculating the degree of rotation and translation.

We found that the estimate of rotation was correct in 65 words out of 70. Similarly, estimate of translation of only 4 words out of 70 had an error of more than 5 pixels. So the error coefficient can be given as max (5, 4) / 70 = 0.07 which results in an accuracy of 93% per document. Since the document was in Devanagari, so we could not find an OCR for performing the conventional test of accuracy. Similar experiments were performed with ten other documents with different levels of warp, skew and shadow. The average accuracy obtained was 94%. Also, with English documents, we performed the test using OCR and found that 95% of the words were correctly recognized. Thus, we can conclude that the word based accuracy measure and the OCR based measure yield similar performance. Moreover, the results of the word based measure are independent of font and script.

The parameters used in the technique are empirically found after working over different scripts. The local affine transformation technique used by the system overcomes the local variations over the distorted document.

4 CONCLUSIONS

Warping is a major problem which arises due to improper curvedness of thick volumes while scanning or photocopying. This defect affects the readability of text as well as its OCRability. It is a difficult task to invent a technique that could remove it perfectly considering the different components that could exist in a document like figures and various fonts. In the technique developed in the present work we were successful in removing the shadow, skew and warp, even if they co-existed in the same document. Moreover, our approach was script and font independent. One problem that we have noticed is the warping of individual characters. This can be noticed in the final output, Figure 6. This can be overcome by reducing the granularity of our approach. . In the present work we have taken the word as an individual unit and assume that the degree of rotation and translation will be a constant for a given word. However, the presence of warping of individual characters indicates that we should estimate the warping at a sub-word level. Similarly, in order to apply our technique to documents containing images, we will have to identify the text area first and apply the dewarping technique to the text regions only. For regions containing graphics or images, we can estimate the extent of warping using interpolation techniques. These are important open problems that require further investigations.

ACKNOWLEDGEMENTS

Authors gratefully acknowledge financial assistance from TDIL, MCIT.

REFERENCES

Zheng Zhang and Chew Lim Tan, Scotland, 2003.Correcting Document Image warping based on Regression of Curved Text Lines. *Proc. Seventh Int. Conf. Document Analysis and Recognition*, *vol.1*-589-593.

P.Kakamanu et al, 2006.Document Image Dewarping based on Line Estimation for Visually Impaired. *Proc. 18th IEEE Int. Conf. Tools with Artificial Intelligence*,p 625-631

Huaigu Cao et al.A, France, 2003.Cylindrical Surface Model to Rectify the Bound Document Image. *Proc. Ninth IEEE Int. Conf. Computer Vision*, *vol. 1*-pp 228-233.

Hironori Ezaki et al,Korea, 2005.Dewarping of document image by global optimization. *Proc. Eighth Int. Conf. Document Analysis and Recognition*, p 302-306.

B. Gatos et al, Brazil, 2007.Segmentation Based Recovery of Arbitrarily Warped Document Images. *Proc. Int. Conf. Document Analysis and Recognition, vol.2*-pp 989-993.

B. V. Dhandra et al, Hong Kong, 2006.Skew Detection in Binary Image Documents Based on Image Dilation and Region labelling Approach. *Proc. 18th Int. Conf. Pattern Recognition*, *vol.2*-pp 954-957.

Chiu L. Yu et al, Washington, 1995.Document Skew Detection Based on the Fractal and Least Squares Method. *Proc. 3rd Int. Conf. Document Analysis and Recognition*, vol.2- pp 1149-1152.

Zheng Zhang et al, Washington D.C, 2004.Restoration of Curved Document Images through 3D Shape Modeling. *Proc. 2004 IEEE Computer Society Conf. Computer Vision and Pattern Recognition, vol.1*- pp 1-10-1-15.

Li Zhang et al, C.A., 2005.3D Geometric and Optical Modeling of Warped Document Images from Scanners.*Proc. 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, *vol.1*-pp 337-342.

Chew Lim Tan et al, 2006. Restoring Warped Document Images through 3D Shape Modeling. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *Vol. 28*-pp 195-208

Zheng Zhang and Chew Lim Tan, Seattle, 2001. Recovery of Distorted Document Image from Bound Volumes. *Proc. 6th Int. Conf. Document Analysis and Recognition,* p 429-433.

Zhixi Fang and Xiaobo Li, 1987.A parallel processing approach to image object labeling problems.*ACM Annual Computer Science Conf., Proc. 15th Annual Conf. Computer Science*, 1987, p 423

D. Ballard and C. Brown, 1982.*Computer Vision* (pp 24 – 30). Prentice-Hall.