

Recovery of Distorted Document Images from Bound Volumes

Zheng Zhang, Chew Lim Tan

School of Computing, National University of Singapore

3 Science Drive 2, Singapore 117543

Email: {zhangz, tancl}@comp.nus.edu.sg

Abstract

Recovery of document images scanned from thick bound volumes is necessary for the purpose of human reading and text retrieval. The main problem with scanning of bound volumes is that there always occurs perspective distortion. Such distortion causes two sources of degradation for the scanned images – 1) shadow at the book spine area, and 2) warping of the words in the shadow. In this paper, we have developed a restoration system to solve these two problems. First, the boundary between the shadow and the clean area is detected. Then the system applies a modified Niblack's method to remove the shadow. The system uses a connected component analysis to help improve the noise reduction and adjust the location and orientation of the warped word in the shadow area, i.e. the words within the boundary detected earlier. The implementation results for each step are presented. Our system will be used in the text retrieval projects for National Archives of Singapore and NUS Digital Library.

1. Introduction

There are two projects from National Archives of Singapore and NUS Digital Library respectively, which motivate the present work. The former involves thick volumes of old handwritten documents, while the second deals with bound copies of past students' theses and old books that are of historical value. Both of them require digitalizing physical pages to images. The main problem with scanning of bound volumes is that the page to be scanned cannot be laid flat on the document glass of the scanner, resulting in image distortion. There are two sources of degradation in such distorted images -- 1) shadow at the book spine area, and 2) warping of the words in the shadow. The degradation introduces problems in our project in image restoration of archive documents and in word spotting for purpose of document retrieval from the digital library.

To deal with the above two type of image degradation, we did a literature search and found related techniques but no direct solution. Baird discusses a 6-parameter degradation model [1][2]. But the algorithm for estimating distributions on all of the model parameters to fit real image populations closely is still an open problem. Kanungo introduces a model for perspective distortion [3]. However, in his model there are a number of assumptions, and some parameters may not be known in reality. Thus in our system we do not adopt any degradation model. We use a modified Niblack's binarization method to remove the shadow, and next construct connected components to do noise filtration and to adjust the warped words in the shadow.

Our method consists of four steps: 1) detecting shadow boundary; 2) binarizing the image; 3) constructing and analyzing connected components; 4) correcting warped word images. We shall discuss these four steps in sections 2 to 5.

2. Detecting shadow boundary

Figure 1 shows an example of a grayscale image scanned from a thick bound book. We assume that each image contains only one physical page. Note that only the words in the shadow are warped. Thus correction of word images only needs to be made within the shadow region. This step thus aims to detect the boundary between the shadow and the clean area in preparation for the distorted word image adjustment to be described in Section 5.

A vertical projection profile is obtained first to decide whether the shadow lies on the left side or right side of the image, based on the location of the peak in the profile. We next scan the text line by line horizontally starting from the side near the peak location. A break point, b , is determined for each scan line. The boundary between the shadow and the clean area is defined as a set, B , of all the break points, as follows:

$$B = \{(b, i) \mid 0 \leq i < n, b = F(i, T)\} \quad (1)$$

where n is the number of horizontal pixel lines in the image, T is a predefined threshold, and F is a function returning the length of the first run of pixels in i th line

whose intensity value is greater or equal to T . In our experiment, we use $T = Vmax / 4$, where $Vmax$ is the maximum intensity value in the grayscale image.

3. Binarizing the image

A modified version of Niblack's algorithm [4] is used to remove the shadow. Niblack's method works by varying the threshold over an image, based on the local mean, m , and local standard deviation, s , computed in a small neighborhood (normally a window size 15×15 is used) of each pixel. A threshold for each pixel, $p(x,y)$, is computed from $T(x,y) = m(x,y) + k \cdot s(x,y)$, where $m(x,y)$ and $s(x,y)$ are the local mean and local standard deviation calculated in a window centered at (x,y) , and k is a user defined parameter and is negative in value.

Niblack's method could not be adopted directly for two reasons. One is that Niblack's method is sensitive to the value of k for our images. It is quite difficult to find a single k value that works for most of our test images. The other is the resultant large amount of pepper noise in the shadow area, even if a proper k value is chosen.

In our modification, each standard deviation, $s(x,y)$, is normalized by dividing it by the dynamic range of standard deviation, R . Furthermore, the local mean, $m(x,y)$, is utilized to multiply, instead of adding, the standard deviation terms. These have the effect of amplifying the contribution of standard deviation, which produces results with much less pepper noise than the original one. These modifications also reduce the sensitivity of parameter k . Equation (2) presents the revised formula.

$$T(x,y) = m(x,y) \cdot [1 + k \cdot (1 - \frac{s(x,y)}{R})] \quad (2)$$

where $m(x,y)$ and $s(x,y)$ are as in Niblack's formula. We use $R = 100$ and $k = 0.1$ for grayscale images. Figure 2 shows the binarization result for Figure 1.

4. Constructing and analyzing connected components

After binarization, sizeable noise may still remain as shown in Figure 2. To improve the binarization result, connected components are constructed based on 8-neighbor connectivity to realize noise filtration. Furthermore, analysis of the connected components also permits formation of bounding boxes of words for use in warped word adjustment in section 5.

4.1 Noise Filtration

The filter consists of two parts, namely shape-filter and size-filter. The shape-filter rejects long and irregular-shape objects that are usually not text characters, and the size-filter simply rejects large objects whose sizes are greater than a threshold value. The connected component area that is most common then becomes the shape-filter. This is reasonable since the text characters are almost similarly sized. Connected components whose sizes lie outside the permissible range of the shape-filter are then removed. Mathematically, the shape-filter can be expressed as Equation (3).

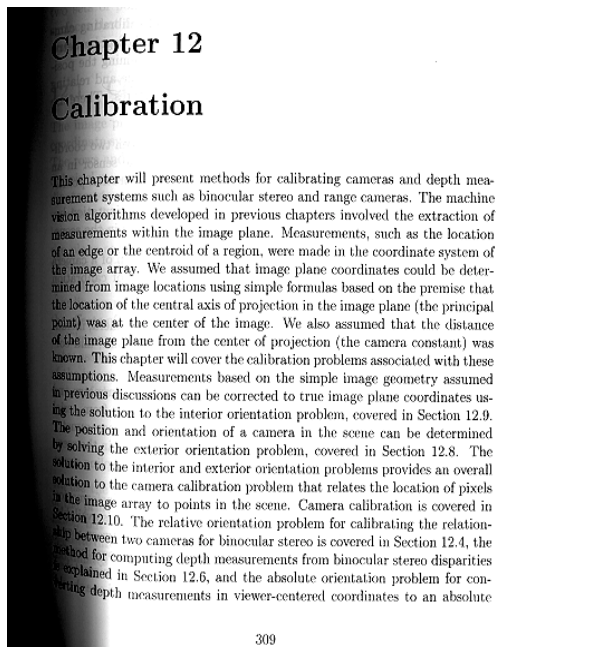


Figure 1. A grayscale image scanned from a thick bound volume.

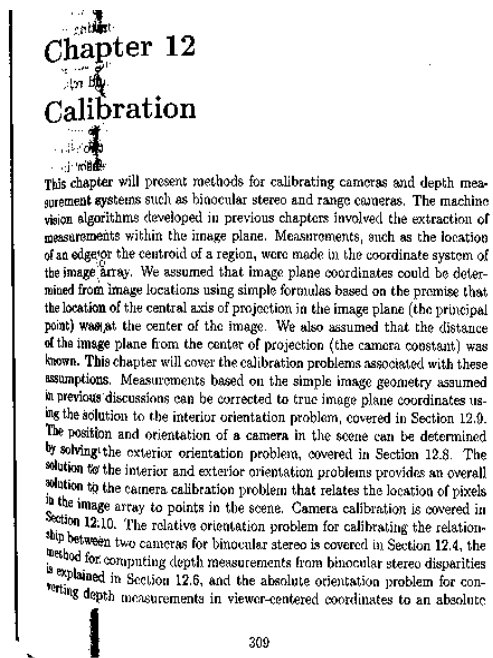


Figure 2. Binarization result for Figure 1.

$$F(c) = \begin{cases} \text{true} & \text{if } \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \leq \sqrt{j \cdot \text{size_of}(c)}, \nabla(x_1, y_1) \in c \\ \text{false} & \text{otherwise} \end{cases} \quad (3)$$

where c denotes a connected component, (x_0, y_0) and (x_1, y_1) are coordinates for the center of gravity and an arbitrary point of c respectively, size_of is a function returning the number of black pixels in c , and j is a predefined parameter. We use $j = 2.5$ for our grayscale images. The result after applying filters is shown in Figure 3.

4.2. Connected component basis clustering

Some of the connected components may consist of one or several characters, while others are part of some characters, such as the top dot and the body of a character like ‘i’ or ‘j’. If we adjust the warped words by moving and rotating connected components directly, the characters in the same word may not lie on the same straight line and may have different orientations, resulting in poor restoration. So the connected components are first clustered into words by applying a nearest neighbor algorithm. The algorithm initializes all the connected components to be unmarked and then perform the following iterative steps:

Step 1: Generate an empty word, i.e. the word contains zero connected components. Push the first unmarked connected component into the word, and make it marked.

Step 2: While there are still unmarked connected components, for each unmarked connected component,

This chapter will present methods for calibrating cameras and depth measurement systems such as binocular stereo and range cameras. The machine vision algorithms developed in previous chapters involved the extraction of measurements within the image plane. Measurements, such as the location of an edge or the centroid of a region, were made in the coordinate system of the image array. We assumed that image plane coordinates could be determined from image locations using simple formulas based on the premise that the location of the central axis of projection in the image plane (the principal point) was at the center of the image. We also assumed that the distance of the image plane from the center of projection (the camera constant) was known. This chapter will cover the calibration problems associated with these assumptions. Measurements based on the simple image geometry assumed in previous discussions can be corrected to true image plane coordinates using the solution to the interior orientation problem, covered in Section 12.9. The position and orientation of a camera in the scene can be determined by solving the exterior orientation problem, covered in Section 12.8. The solution to the interior and exterior orientation problems provides an overall solution to the camera calibration problem that relates the location of pixels in the image array to points in the scene. Camera calibration is covered in Section 12.10. The relative orientation problem for calibrating the relationship between two cameras for binocular stereo is covered in Section 12.4, the method for computing depth measurements from binocular stereo disparities is explained in Section 12.6, and the absolute orientation problem for converting depth measurements in viewer-centered coordinates to an absolute

Figure 3. Enhanced binarization result for Figure 1.

$C1$, and each component, $C2$, in the current word, if there exists a pixel $p1$ in $C1$ and a pixel $p2$ in $C2$ such that the distance between $p1$ and $p2$ is less than a threshold distance, D , $C1$ is marked and pushed into current word. If no connected component is added in, the current iteration of Step 2 will stop and the algorithm will check whether all the connected components are marked, if yes, the algorithm terminates, else go back to Step 1 and start a new iteration.

The result of applying this algorithm is shown in Figure 4. Each word is indicated by its bounding box.

4.3. Word basis clustering

Merely clustering connected components into words is not enough, since for a given word we do not know which text line it belongs to, i.e. there is no way to move the word to the correct location. Our system groups the words into text lines by using a modified “box-hands” method [5]. The original “box-hands” method assumes that the text line is a straight horizontal line, which is obviously not the case in our test images.

The bounding boxes constructed in section 4.2 are extended to give chains of connected boxes. As shown in Figure 5, we extend each bounding box by adding to its left and right sides two equal parallel-quadrilateral extensions, called the “box hands”. The length and height of the hands are equal to the height and half of the height of the word bounding box respectively. The two box-hands are positioned in such a way that the following three points lie on the same horizontal line: the mid-point

This chapter will present methods for calibrating cameras and depth measurement systems such as binocular stereo and range cameras. The machine vision algorithms developed in previous chapters involved the extraction of measurements within the image plane. Measurements, such as the location of an edge or the centroid of a region, were made in the coordinate system of the image array. We assumed that image plane coordinates could be determined from image locations using simple formulas based on the premise that the location of the central axis of projection in the image plane (the principal point) was at the center of the image. We also assumed that the distance of the image plane from the center of projection (the camera constant) was known. This chapter will cover the calibration problems associated with these assumptions. Measurements based on the simple image geometry assumed in previous discussions can be corrected to true image plane coordinates using the solution to the interior orientation problem, covered in Section 12.9. The position and orientation of a camera in the scene can be determined by solving the exterior orientation problem, covered in Section 12.8. The solution to the interior and exterior orientation problems provides an overall solution to the camera calibration problem that relates the location of pixels in the image array to points in the scene. Camera calibration is covered in Section 12.10. The relative orientation problem for calibrating the relationship between two cameras for binocular stereo is covered in Section 12.4, the method for computing depth measurements from binocular stereo disparities is explained in Section 12.6, and the absolute orientation problem for converting depth measurements in viewer-centered coordinates to an absolute

Figure 4. Words indicated by their bounding boxes.

of the right side of the left box-hand, the center of the word bounding box, and the mid-point of the left side of the right box-hand, as shown in figure 5. The orientation of the hands is decided by the orientation of the word. Our system detects the word orientation by adopting a Linear Regression [6]. It takes the coordinates, x_i and y_i , of the centers of all the connected components that belong to a word as inputs of the Linear Regression, and finds the equation $y = m \cdot x + c$ of a line that best fits all these centers, where m and c are computed as follows:

$$m = \frac{(n \sum x_i y_i - (\sum x_i)(\sum y_i))}{(n \sum x_i^2 - (\sum x_i)^2)} \quad (4)$$

$$c = \frac{(\sum y_i - m \sum x_i)}{n} \quad (5)$$

A text line is found by clustering all the words whose hands of bounding boxes touch each other.

Comparing with [7], we do not calculate the two intersection points of the word orientation line and the edges of the bounding box. Also, instead of Hough transform we adopt Linear Regression, which has much less computation than Hough transform. These simplifications improved the system performance greatly without affecting the quality of the resultant image.

5. Warped word adjustment

The warped word adjustment consists of two steps – location adjustment and orientation adjustment. For a text

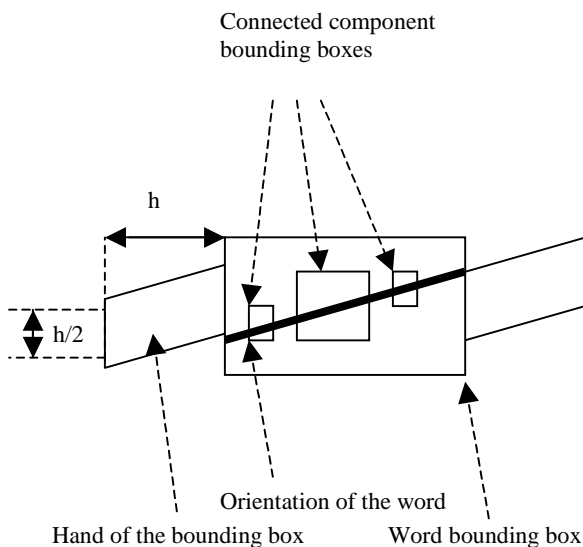


Figure 5. Extension of the word bounding box.

line, L , found in section 4.3., we perform a Hough transform on the centers of all the connected components in L , and return a straight line, S , for L . Then for all the warped words in L , i.e. the words whose centers of bounding boxes lie within the shadow boundary (detected in section 2.), the location adjustment is done by moving the warped word such that the center of the word bounding box lies on its vertical projection point on S . For each warped word, the angle, A , between the word orientation (detected in section 4.3) and S can be calculated, since the functions of the two lines are known. The orientation adjustment is done by rotating the word around its center of bounding box by $-A$ degrees.

The final result for Figure 1 is shown in Figure 6. Comparing with Figure 3, the warped words have been moved to the correct location and rotated to the correct orientation. At the moment, no correction is made yet to the shape of the word. However, the general appearance and the readability of the original document image has been greatly improved and enhanced.

6. Conclusion and future works

In this paper, we first describe the problems of distorted document images from scanning of bound volumes. A restoration system is then proposed to solve these problems. The system basically first makes use of an adapted Niblack's thresholding method to binarize the image. Connected components are then constructed to help to further remove noise based on shapes and sizes of the connected components. Further analysis is done on the

Chapter 12 Calibration

This chapter will present methods for calibrating cameras and depth measurement systems such as binocular stereo and range cameras. The machine vision algorithms developed in previous chapters involved the extraction of measurements within the image plane. Measurements, such as the location of an edge or the centroid of a region, were made in the coordinate system of the image array. We assumed that image plane coordinates could be determined from image locations using simple formulas based on the premise that the location of the central axis of projection in the image plane (the principal point) was at the center of the image. We also assumed that the distance of the image plane from the center of projection (the camera constant) was known. This chapter will cover the calibration problems associated with these assumptions. Measurements based on the simple image geometry assumed in previous discussions can be corrected to true image plane coordinates using the solution to the interior orientation problem, covered in Section 12.9. The position and orientation of a camera in the scene can be determined by solving the exterior orientation problem, covered in Section 12.8. The solution to the interior and exterior orientation problems provides an overall solution to the camera calibration problem that relates the location of pixels in the image array to points in the scene. Camera calibration is covered in Section 12.10. The relative orientation problem for calibrating the relationship between two cameras for binocular stereo is covered in Section 12.4. The method for computing depth measurements from binocular stereo disparities is explained in Section 12.6, and the absolute orientation problem for converting depth measurements in viewer-centered coordinates to an absolute

Figure 6. Final result for Figure 1.

connected components to construct bounding boxes of the words in the text images to allow determination of text lines in order to correct warped text lines. As the text line correction basically entails adjustment of location and orientation, the shape of the word is not fully restored at the moment. One of our future research works is to replace the connected components analysis with an “optical flow” technique. We will first obtain surface distortion information through variation of gray levels and alignment of any textual components, and then the information obtained will be used to compute the surface orientation and curvature so that the system can also correct the shape of the warped words. More complex document images will also be examined in future, involving multiple pages and graphical contents.

Acknowledgement

This research is supported by a joint research grant R252-000-071-112/303 provided by National Science and Technology Board and Ministry of Education, Singapore.

References

- [1] H. Baird. “The State of the Art of Document Image Degradation Modeling”, In Proc. of 4th IAPR *International Workshop on Document Analysis Systems*, Rio de Janeiro, Brazil, pp.1-16, 2000, December 10-13, 2000.
- [2] H. Baird. “Document Image Quality: Making Fine Discriminations”, In Proc. of IAPR Int’l Conf. on Document Analysis and Recognition, Bangalore, India, pp. 459-462, September 20-22, 1999.
- [3] T. Kanungo, R.M. Haralick and I. Phillips. “Global and Local Document Degradation Models”, In Proc. of IEEE *International Conference on Document Analysis and Recognition*, Tsukuba Science City, Japan, pp. 730-734, October 1993.
- [4] J. Sauvola, and M.Pietikainen “Adaptive Document Image Binarization”, *Pattern Recognition 33(2000)*, pp. 225-236, 1999.
- [5] C. Strouthopoulos, N.Papamarkos, and C.Chamzas. “Identification of Text-Only Areas in Mixed-Type Documents”, *Engng Applic. Artif. Intell.*, Elsevier Science Ltd, Great Britain, Vol. 10, No. 4, pp. 387-401, 1997.
- [6] R. Jain, R.Kasturi, and B.G.Chamzas. “Machine Vision”, McGRAW-HILL International editions, pp. 506-507, 1995.
- [7] Z. Zhang, and C.L. Tan. “Restoration of Images Scanned from Thick Bound Documents”, accepted by International Conference on Image Processing 2001.