DOCUMENT IMAGE RESTORATION

-For Document Images Scanned from Bound Volumes-

By

Zheng Zhang

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
NATIONAL UNIVERSITY OF SINGAPORE
REPUBLIC OF SINGAPORE
AUGUST 2004

To My Parents

Table of Contents

List of Figures

List of Tables

List of Publications

1. Z. Zhang, C.L. Tan, and L.Y. Fan, “Restoration of Curved Document Images through 3D Shape Modeling”, *Computer Vision and Pattern Recognition (CVPR)*, Volume 1, pp. 10-16, 2004.
2. Z. Zhang, C.L. Tan, and L.Y. Fan, “Estimation of 3D Shape of Warped Document Surface for Image Restoration”, *International Conference on Pattern Recognition (ICPR)*, 2004.
3. Z. Zhang and C.L. Tan, “Correcting Document Image Warping Based on Regression of Curved Text Lines”, *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 589-593, 2003.
4. Z. Zhang and C.L. Tan, “Straightening Warped Text Lines Using Polynomial Regression”, *International Conference on Image Processing (ICIP)*, pp. 977-980, 2002.
5. Z. Zhang and C.L. Tan, “Recovery of Distorted Document Image from Bound Volumes”, *International Conference on Document Analysis and Recognition (ICDAR)*, pp.429-433, 2001.

6. Z. Zhang and C.L. Tan, “Restoration of Document Images Scanned from Thick Bound Document”, *International Conference on Image Processing (ICIP)*, pp. 1074-1077, 2001.

Acknowledgments

There are a number of people who guided and assisted me in one way or another to accomplish this research. First of all, I wish to thank my supervisor, Professor Chew Lim Tan, for his continuous guidance, insightful suggestions and enthusiastic inspiration. He advised me in various ways to improve my research acumen and shape my research capability. He makes my 4-year research work a most nourishing experience for me. I would also like to thank Dr. Zhi Yong Huang and Dr. Terence Sim for their reviews and guidance.

I am particularly grateful to Dr. Tao Xia and Mr. Li Ying Fan for their assistance. They provide me lots of valuable suggestions and help me accomplish the heavy workload experiments. Working with my buddies, Rui Ni Cao, Yue Lu, Pei Yi Shen, Ji He, Lin Lin, Florence, and all the other members in CHIME lab, colored my research life.

Finally, but not the least, I would like to thank my beloved parents, for their endless love, forever.

Abstract

When one scans a document page from a thick bound volume, perspective distortion is a common problem due to the curvature of the page to be scanned. This results in two kinds of distortion in the scanned document images:

- Photometric distortion: shade along the ‘spine’ of the book
- Geometric distortion: warping in the shade area.

The distortion in the document images introduces problems not only for fast and painless human reading, but also for document image analysis, understanding and recognition.

In this thesis, we first propose two novel restoration approaches to tackle the above distortion problems:

Approach 1: Document image restoration based on 2D document image processing

Approach 2: Document image restoration based on 3D document shape discovery

We then evaluate the restoration results by comparing the OCR performance on the original document image and the corresponding restored images by different methods respectively, and compare the two approaches by discussing two issues: effectiveness

and efficiency.

In approach 1, we first obtain the shade boundary knowledge by a run-length method. We next binarize the image by a modified Niblack's method to remove the shade. Connected components based on 8-neighbors are constructed and analyzed to help improve the noise reduction and graphical object removal. We divide the connected components into two areas by the shade boundary detected earlier, namely the shade area where the text lines are warped and the clean area where the text lines are not distorted and remain as a straight line. In the clean area, we adopt a top-down approach to separate connected components into partial straight text lines by analyzing the horizontal projection profile. We apply linear regression to generate a pair of top and bottom straight reference lines for each partial straight text line. In the shade area, we adopt a bottom-up approach to cluster connected components into words, and then cluster words into partial curved text lines. We use polynomial regression to compute a pair of top and bottom quadratic reference curves fitting the warped text lines. We next connect the partial straight and warped text lines to form a set of complete text lines. The warped text lines are restored by correcting the quadratic curves accordingly based on the corresponding straight reference lines. The experimental results showed the proposed method can mostly correct both photometric and geometric distortion. Our work in approach 1 has led to publications [106, 107, 108, 109]

In approach 2, with the scanner information (gain and bias, focal length, incident angle of the light source, and so on) estimated as a priori knowledge, we propose a

novel method to tackle the distortion problems based on the 3D document shape. We first build practical models (consisting of a 3D geometric model and a 3D optical model) for the practical scanning conditions. We then propose a novel method to reconstruct the 3D shape and the albedo distribution of book surface. We build a de-shading model based on the discovered albedo distribution to correct the photometric distortion, and a de-warping model based on the discovered book surface to correct the geometric distortion. This method is tolerable for document skew, and can successfully remove both photometric and geometric distortion. Our work in approach 2 has led to papers [110, 111, 112].

Finally, we evaluate the restoration results by comparing the OCR (Optical Character Recognition) performance on the original and restored document images. We use the precision and recall defined in [35] as the metrics for OCR performance. We present a discussion to compare the two approaches.

Chapter 1
Introduction

1.1 The Document Domain

The document incorporates all aspects of written communication. Examples include technical reports, government files, books, newspapers, journals, magazines, letters, bank checks, and so on. Documents have been the dominant information medium in human society. They contain information and provide a way of transferring information across time and space. Though traditionally documents are paper-based, now documents are often in electronic format thanks to advances in computing and networking.

The move from bookshelves and filing cabinets to the paperless world has been prompted by the many advantages to be gained from the electronic document environment, such as efficient archiving, retrieval and maintenance. In the past few decades, document have been increasingly generated, maintained and stored on the computer. However, there is no evidence yet of less paper on our desks. Paper

documents are still printed for reading and various transactions. Besides, in the libraries, we still have huge volumes of old paper documents. So, the cry of the early 1980s for the “paperless office” has now given way to a different objective: dealing with the flow of electronic and paper documents in an efficient and integrated way. The ultimate solution would be for computers to deal with paper documents automatically as they deal with other forms of computer media, such as magnetic and optical disks. [61]

A conceptual representation of a document’s life cycle is shown in Figure 1.1. It indicates how documents can be transformed from the electronic format to paper and vice versa. We usually create a *document model* on the computer, and then print it out on paper by rendering and reproducing. The printed document may be scanned and stored into the computer as *document image*. The document image can be further restored, analyzed and recognized, and converted into some editable models to facilitate manipulation on the computer.

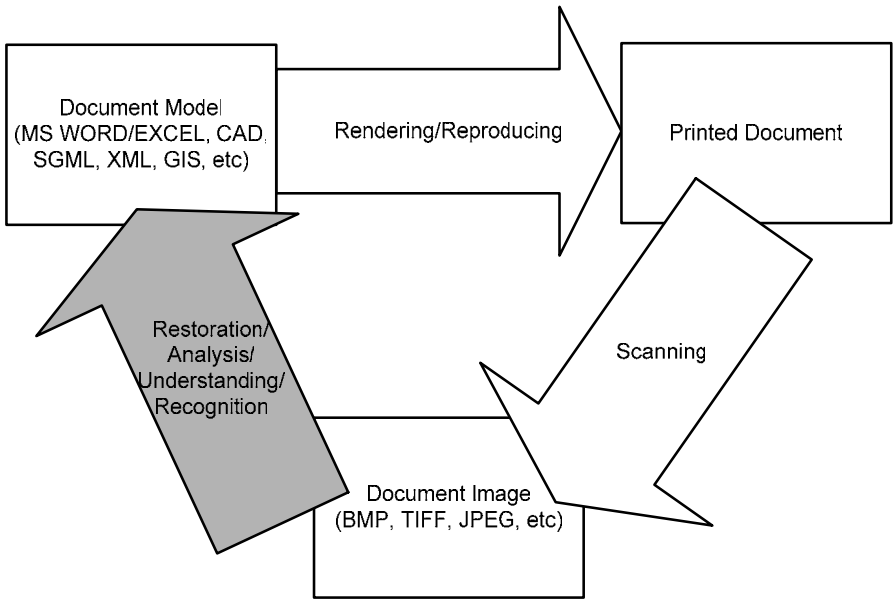


Figure 1.1: The conceptual representation of a document’s life cycle.

1.2 Document Image Restoration (DIR)

1.2.1 What is DIR?

In the cycle in Figure 1.1, while digitalizing the physical printed documents to images, the document images are almost inevitably degraded in the course of scanning, especially for the ones scanned from bound document volumes. This loss of quality – even when it appears negligible to human eyes – can cause problem for subsequent analysis, understanding, and recognition of the document images, for example, an abrupt decline in accuracy by the current generation of *Optical Character Recognition (OCR)* systems [8]. Thus various pre-processing methods that aim to suppress the document image degradation using knowledge of its nature have to be applied. This process is called *Document Image Restoration (DIR)*.

1.2.2 Problems of DIR for Document Images Scanned from Bound Volume

While scanning pages from a bound volume, the curving of the page facing the scanner glass causes both photometric and geometric distortion in the scanned grayscale document image as shown in Figure 1.2:

- **Photometric distortion**: shade along the ‘spine’ of the bound volume.
- **Geometric distortion**: warping of book surface in the shade area. Since the scanner picks up a 1D projection for each vertical column, the horizontal geometric distortion is due to orthogonal projection, and the vertical geometric distortion is due to perspective projection.

The distortion in the document images introduces problems not only for fast and painless human reading, but also for document image analysis, understanding and recognition [6, 7, 8, 57], such as:

- OCR for textual content
- Graphics recognition for engineer drawings, map conversion, music scores, schematic diagrams, organization charts, and so on
- Document layout analysis
- Script, language and font recognition
- Document image thresholding
- Document skew detection, and so on

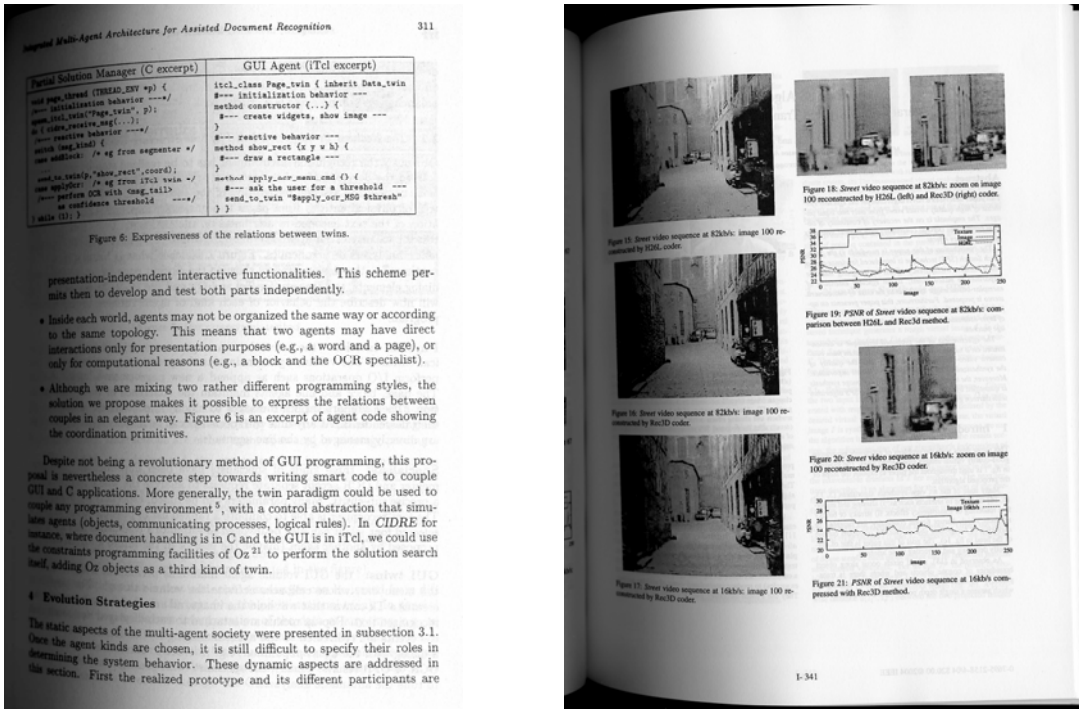


Figure 1.2: Two grayscale document images scanned from bound volumes.

1.3 The Objectives and Contributions

In this thesis, we present our solutions to address the issues of DIR for document images scanned from bound volumes. We discuss how to effectively and efficiently correct both photometric and geometric distortion using two different approaches as follows:

- **Approach 1 – DIR based on 2D document image processing:** We propose a novel binarization method to remove the photometric distortion, and a reference line/curve detection algorithm based on linear/quadratic regression to correct the geometric distortion.
- **Approach 2 – DIR based on 3D document shape discovery:** We introduce a practical model (consisting of a 3D geometric model and a 3D optical model) to reconstruct the book surface and recover the surface albedo distribution, and a restoration model (consisting of de-shading model and de-warping model) to correct both photometric and geometric distortion based on the discovered book surface and surface albedo distribution.

The evaluation of the restoration results is conducted to demonstrate the superiority of the proposed methods, and the comparison of the two restoration approaches is presented by discussing two issues: effectiveness and efficiency.

1.3.1 DIR based on 2D Document Image Processing

We remove the photometric distortion by binarizing the grayscale image. In the

literature, many binarization methods have been reported [66, 82, 103, 54, 73, 21, 32, 91, 22, 62, 67, 102, 53, 69, 70, 48, 27, 68, 4, 87, 55, 78, 71, 13, 46, 60, 44, 47, 93, 76, 98, 34, 104, 77, 58, 65, 23] since 1970's. Though extensive research has been done, as far as we know, there are no existing methods, which work efficiently and produce acceptable results for our problem. We thus propose a novel efficient local binarization method, which is modified from a well known binarization method – Niblack's method [60] (Experiments in [99, 88, 92] show that Niblack's method is most effective among eleven locally adaptive thresholding techniques). In our modification, each standard deviation is normalized by dividing its dynamic range. Furthermore, the local mean is utilized to multiply, instead of adding, the standard deviation terms. These modifications have the following effects:

- Amplifying the contribution of standard deviation, which leads to better binarization results with much less pepper noise than the ones binarized by Niblack's method.
- Reducing the sensitivity of control parameter, which makes the parameter to be constant for most of our testing document images.

This binarization method efficiently produces good binarization results for document images scanned from bound volumes, and thus tackles the photometric distortion.

We next propose a reference line/curve detection algorithm to correct the geometric distortion. For the binarized document image, noise is further removed using a connected component analysis. The connected components are divided into two classes: 1) connected components in the shade area, and 2) connected

components in the clean area. A top-down approach is applied to cluster connected components in the clean area into straight text lines, and the alignments of text are modeled by straight reference lines using linear regression. A bottom-up approach is applied to cluster the connected components in the shade area into warped text lines, and polynomial regression is used to model the warped text lines with quadratic reference curves. Corresponding warped text alignments and linear text alignments in both areas are then paired up. The warped text lines are restored by correcting the quadratic curves accordingly based on the corresponding straight text lines.

However, this method has the following disadvantages:

- The shapes of the characters are not changed. In the resulting images, while the orientation and location of the characters in the shade are restored, the shapes of these characters may still appear distorted and narrower than the ones in the other region.
- The graphical objects in the document image, such as diagrams, figures, charts, tables, and so on, cannot be restored.

Thus the geometric distortion is partially, but not all, corrected. With the scanner information as a priori knowledge, we propose another better restoration method based on discovering the 3D document surface, which can completely correct the photometric and geometric distortion. This is described in the next subsection.

1.3.2 DIR based on 3D Document Shape Discovery

We first build a 3D geometric model according to the geometric structure of the book

surface while scanning. With the scanner information (gain and bias, focal length, tilt angle of the light source, and so on, which are estimated as a priori knowledge), we next construct a 3D optical model for this *Shape-from-Shading (SFS)* [31] problem by considering the following four factors in real world environments:

- A proximal and a moving light source
- Lambertian reflection
- A non-uniform albedo distribution
- Document skew

We propose a method to reconstruct the book surface and recover the albedo distribution by adopting the 3D geometric model and 3D optical model. We build a de-shading model based on the discovered albedo distribution to correct the photometric distortion, and a de-warping model based on the discovered book surface to correct the geometric distortion by performing the following three corrections:

- Correcting the vertical geometric distortion caused by perspective projection
- Correcting the horizontal geometric distortion caused by orthogonal projection
- Correcting the document skew

This method is tolerable to document skew, and can successfully remove both photometric and geometric distortion. It works on the entire contents of the document page, irrespective of whether they are textual or graphic.

1.3.3 Experimental Evaluation & Comparison

Since one important purpose of our DIR is for subsequent document image analysis,

understanding, and, finally, recognition of the document images, and OCR played a fundamental role in document image recognition domain [57], we evaluate the restoration results by comparing the OCR performance on the original document image and the corresponding restored images by the two methods respectively. We use the precision and recall defined in [35] as the metrics for OCR performance. We compare the two methods by presenting a discussion on effectiveness and efficiency.

1.4 Organization of the Thesis

The organization of the rest of the thesis goes as follows:

In Chapter 2, we review an extensive related work in DIR literature: we classify the existing methods into two categories, and make a brief discussion on each category.

In Chapter 3, we propose a method to restore the document image based on several image processing techniques: we first binarize the document image, and then correct the text warping by the reference lines/curves of each text line.

In Chapter 4, we introduce a restoration method by knowing the scanner information as a priori knowledge: a practical model for the real scanning conditions, consisting of a 3D geometric model and a 3D optical model, is built, and an algorithm to reconstruct the book surface and recover the surface albedo distribution is introduced. Finally, the document image is restored by our de-shading model and de-warping model.

In Chapter 5, we present the evaluation of the restoration results and the comparison of the two restoration methods.

Chapter 6 concludes this thesis with some discussion on future work.

Chapter 2
Related Work

2.1 Introduction

To date, the literature sources of DIR research are quite rich. The *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, the journal of *Pattern Recognition*, and *Pattern Recognition Letters* cover many articles about DIR since 1970's. In 1998, Elsevier launched the *International Journal of Document Analysis and Recognition (IJDAR)*, which publishes papers in the whole area of document image restoration, analysis, understanding and recognition. The biennial *International Conference on Pattern Recognition (ICPR)*, *International Conference on Document Analysis and Recognition (ICDAR)*, and *IAPR International Workshop on Document Analysis System (DAS)* have been steady sources of ideas. Worthwhile contributions have appeared at the *International Conference on Image Processing (ICIP)*, *IAPR Workshop on Structural and Syntactic Pattern Recognition (SSPR)*, *SPIE Conference on Document Recognition and Retrieval (DR&R)*, and *ACM symposium on Document*

Engineering (ACM DocEng). There are evidences showing that, in recent years, an increasing number of papers about DIR [94, 110, 18, 15, 74, 97, 96] appeared in two reputable computer vision conferences – *Computer Vision and Pattern Recognition (CVPR)*, and *International Conference on Computer Vision (ICCV)*.

Though a number of DIR methods have been proposed in the literature, most of these methods are still far from providing a practical solution. As in Chapter 1, we classify the existing restoration methods, which can correct the photometric or geometric distortion over the document images, into two categories:

- **Category 1 – Approaches based on 2D document image processing:** The document images are restored by some document image processing techniques, such as binarization, connected component analysis, active contour, linear and nonlinear interpolation, and so on, without the document shape information.
- **Category 2 – Approaches based on 3D document shape discovery:** The document images are restored based on discovering the 3D document shape, which is estimated by various 3D models, such as applicable surface, depth map, cylindrical model, and so on.

In the rest of this chapter, we give a brief discussion on each category.

2.2 Approaches based on 2D Document Image Processing

Baird [12] discusses a single-stage parametric document image defect model of per-symbol and per-pixel defects based on physics of printing and imaging, and some

refinements of the model [9, 10, 11]. These models use ten parameters to approximate some aspects of printing and imaging of text, including symbol size, digitizing resolution, affine spatial deformations, jitter, blurring, etc. When the model is simulated, the ideal input image is first rotated, scaled, and translated; then the output resolution and per-pixel jitter determine the centers of each pixel sensor; for each pixel sensor the blurring kernel is applied, giving an analog intensity value to which per-pixel sensitivity noise is added; finally, each pixel's intensity is thresholded, giving the output image. Kanungo and Haralick further develop these models in [38, 39, 36, 37]. Kanungo and Zheng introduce a restoration algorithm with six parameters estimated in [42, 43, 105]. These works have opened up new challenges both in theoretical and practical aspects. For instance, the algorithm for estimating distributions on all of the model parameters to fit real image populations closely is still an open problem for further research.

Tang and Suen [85, 86] present an approach to address the nonlinear shape distortion problems for the document image. A number of image transformation models [52] based on the finite element theory are introduced. They adopt two-dimensional geometrical transformations to approximate the three-dimensional distortions. The general model of nonlinear shape distortion is described as follows: Let D^e be an element or a cell of the image. Let $\Phi_i(P)$ be a polynomial in element D^e, for $i = 1,2,..., p$, where P is a node in element D^e and p is the total number of nodes in element D^e. $\Phi_i(P)$ is the shape (distortion) function of element D^e, if it has a value 1 at node P but 0 at the other p-1 nodes. The method requires choosing a

certain number of reference points, depending on the model selected, to derive a transformation function on the image to be distorted. Note that $\Phi_i(P)$ describes the image distortion from a specified model. Once the above transformation has been established, its inverse transformation can be obtained by applying the method introduced in [64]. The inverse transformations of the geometrical models are able to remove nonlinear shape distortions [51]. They do not take account of photometric distortion. A uniform model of distortion on the entire image is assumed, and it is not applicable to images with different distortions at different portions of the images.

Lavialle et al [49, 5] introduce a new text line straightening method using an active contour [45] network based on an analytical model. They first propose a method based on Bezier curves [50], and then they present a model that requires cubic B-splines, which leads to more accurate results. Finally, they propose to automate the initialization by using an approach based on a particle system. However, while it provides good results, the initialization must be close to the desired result. This method only handles document images with pure texts, and do not tackle the photometric distortion. This method is computational expensive due to the computations of particles simulation.

O'Gorman [63] presents a method named the Document Spectrum for the detection of certain distorted text patterns. It is based on bottom-up and nearest-neighbor clustering of connected components, and yields an accurate measure of skew, within-line, and between-line spacing distortion for text lines and pages. It is able to process local regions of different text orientations for the same image.

However, the text lines found by the Document Spectrum analysis are always straight, and thus nonlinear distortion and its restoration are not discussed.

Weng and Zhu [97] propose a nonlinear shape restoration algorithm for document images using edge detection, thinning and linking [113]. This algorithm is based on a linear interpolation theory that is able to detect and restore nonlinear shape distortions in any irregular quadrilateral-shaped patterns. The main idea is the use of two-dimensional spline functions in bicubic, biquadratic, and/or bilinear models [84] to approximate the three-dimensional nonlinear distortion curves. The document images are restored by horizon restoration followed by vertical restoration. However, this method can only handle binarized document images, and many parameters have to be set manually to achieve good restoration results.

2.3 Approaches based on 3D Document Shape Discovery

Pilu [74, 75] presents a novel method based on the physical modeling of paper deformation with an applicable surface. The applicable surface [20] is represented by a polygonal mesh with suitable dimension and known distances between nodes. A relaxation algorithm [14] is then used to fit the applicable surface to noisy data so as to flatten it to produce the final undistorted image. Since this method represents the applicable surface as a polygonal mesh, the texts in the experimental results after correction are simply not legible even to human eyes. Moreover, this method does not tackle the photometric distortion, such as shading, in the testing images.

Brown et al [15] propose a general deskewing algorithm for arbitrary warped documents based on the 3D shape. This algorithm is to get the depth of each point in the image by some stereo vision method, hence to make a depth map, and then dewarp the image according to the depth map based on [89, 90] . Although it seems capable of dewarping any type of image distortions, how to map the points on the rough, noisy surface defined by the depth map to the points on the plane is still a problem. It also requires a special lighting setup [16, 17] and calibration [29] to do active lighting to obtain structural information. This algorithm cannot be applied to normal scanner or camera images, and it cannot correct the shading in the testing images.

Doncescu et al [26] propose a similar method, in which a laser projector is used to project a 2D light network on the surface of the document, and then two dimensional distortions of the surface are corrected with a two pass mesh warping proposed by [81]. This method shares the same problems with [15]， which requires a special lighting setup.

Yamashita et al [101] propose a shape reconstruction and image restoration method for paper document with curved surfaces or fold lines by using a stereo vision system. They first detect the corresponding points from the left and right images, and measure the 3D positions of these points through triangulation. They next reconstruct the 3D document surface shape by using NURBS (Non-Uniform Rational B-Spline) curve [72] representation. Finally, they transform the two original images of curved surfaces to those of flat surfaces by maintaining the distance between points, and

combine the clear regions of two images. This method requires a stereo vision system, and thus cannot be applied for normal scanner/camera images. Moreover, the runtime of their system takes several hours depending on the size of the image.

Tsoi and Brown [94] introduce an approach to restore the document images using the 3D boundaries of the imaged material modeled by [28]. They compute a corrective mapping based on boundary interpolation [24] to simultaneously undo common geometric distortions, such as skew, binder curl, and fold distortion. In addition, the same interpolation framework is used to estimate an intrinsic illumination image. This estimated illumination image together with the original image are used to remove the shading artifacts. However, this method requires that the document boundary must appear in the captured image, and a number of control points and function parameters have to be manually defined for each restoration.

Kanungo et al [40, 41] introduce a cylindrical model for perspective distortion, in which the optical distortion process is modeled morphologically by a combination of cylinder and plane. First a distance transform on the foreground is performed, followed by a random inversion of binary pixels where the probability of flip is a function of the distance of the pixel to the boundary of the foreground. Correlating the flipped pixels is modeled by a morphological closing operation [30, 31]. This method does not aim to estimate all parameters as in Baird's model [12].

Wada et al [95] develop a complicated model to reconstruct the book surface, incorporating interreflections (increased illumination on one part of the book caused by secondary reflections from another) with eight-parameters that are estimated a

priori by using calibrated images of white flat slopes with known slants. They compute the book surface shape by an iterative method, which has the same computation scheme as that proposed in [59]. This method assumes the book surface is cylindrical and requires that the book spine must be strictly parallel to the scanning light. However, usually this is not the case in real scanning conditions. Another limitation with the method is the high computational cost in dealing with interreflections even with the tessellation method they propose. In fact, the interreflections may be ignored, since they mainly affect the illumination on the space margin around the book spine and thus have only little effect on the estimated shape of book surface.

Cao et al [18, 19] introduce a general cylindrical surface model to rectify the warping of a bound document image captured by the camera. By the geometry of the camera image formation, the equations using the cue of directrixes to map the points on the surface in the 3-D scene to the points on image plane are achieved. Baselines of the horizontal text line are extracted based on [25] as projections of directrixes to estimate the bending extent of the surface, and then the images are rectified accordingly. This method cannot remove the photometric distortion like the shading along the book spine, and it would not work if there are no text lines or very few text lines.

Myers et al [56] propose a study of removing perspective distortion for camera images. They assume that cameras are placed such that vertical edges in scenes are still vertical and parallel in images. The vertical vanishing point where vertical edges

intersect is therefore at the infinity of the image plane, while the horizontal vanishing point is in the image plane. They detect and locate the texts in the image based on [80, 100]. They proceed by rotating each text line and observing the horizontal projection profile to find the top and base line, and observing the vertical projection profile to find the dominant vertical edge direction. From the three lines the foreshortening along horizontal axis and shearing along vertical axis are determined so that the original text line image is restored. Their work restores the text lines independently. However, they cannot tackle the photometric distortion and restore the graphical objects in the image.

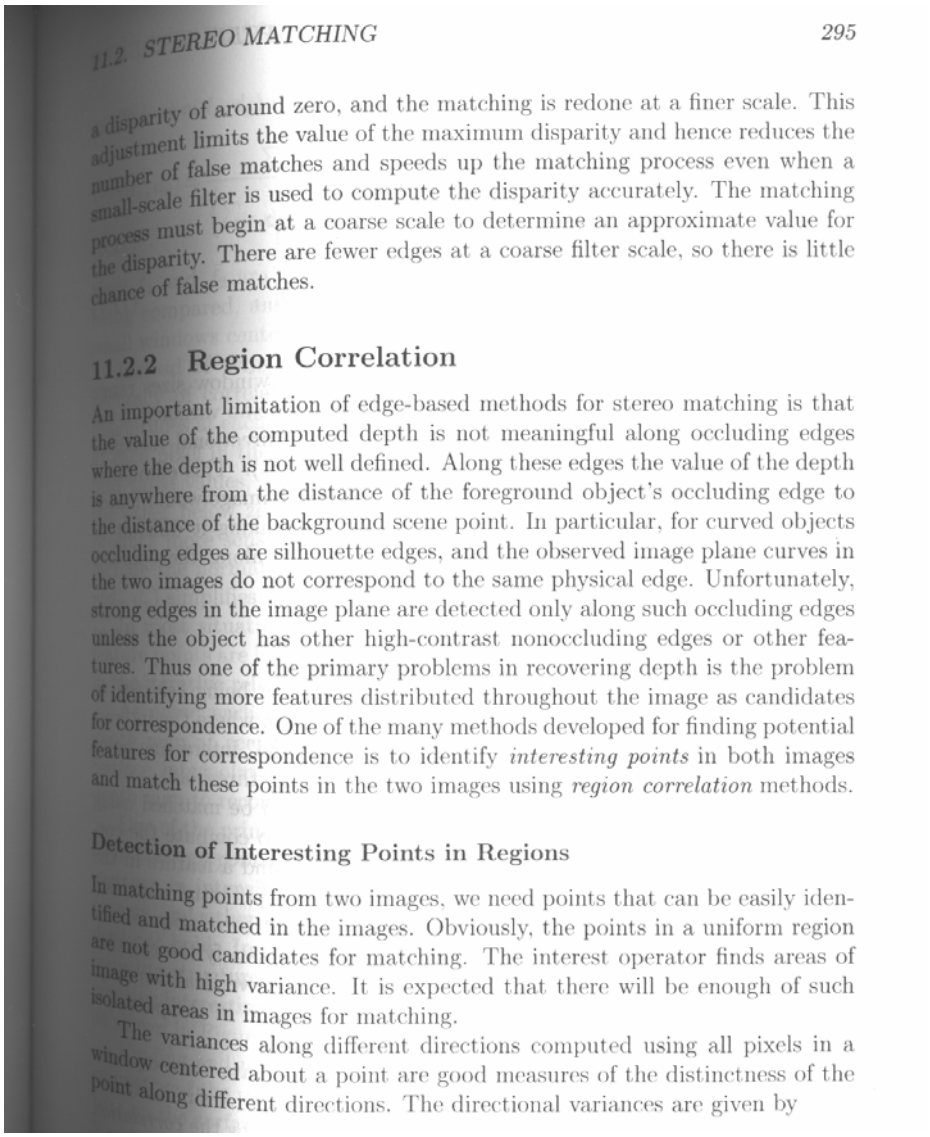
Chapter 3
DIR based on 2D Document Image Processing

3.1 Introduction

Our document images are scanned horizontally or skewed with a slight angle at different resolution level from bound volumes. Figure 3.1 shows an example. In this chapter, we propose a novel resolution-free restoration system that adopts a number of image processing techniques to correct both photometric and geometric distortion without the document shape information.

We first detect the shade position by two sums of pixel intensity values for the right-hand-side and the left-hand-side respectively, and obtain the shade boundary knowledge by a run-length method. We next binarize the image by a modified Niblack's method to remove the photometric distortion. Connected components are constructed and analyzed to help improve the noise reduction and graphical object removal. We divide the connected components into two areas by the shade boundary

detected earlier, namely the shade area where the text lines are warped and the clean area where the text lines are not distorted and remain as a straight line. In the clean area, we adopt a top-down approach to separate connected components into partial straight text lines by analyzing the horizontal projection profile. We apply linear regression to generate a pair of top and bottom straight reference lines for each partial straight text line. In the shade area, we adopt a bottom-up approach to cluster the



11.2. STEREO MATCHING 295

a disparity of around zero, and the matching is redone at a finer scale. This adjustment limits the value of the maximum disparity and hence reduces the number of false matches and speeds up the matching process even when a small-scale filter is used to compute the disparity accurately. The matching process must begin at a coarse scale to determine an approximate value for the disparity. There are fewer edges at a coarse filter scale, so there is little chance of false matches.

11.2.2 Region Correlation

An important limitation of edge-based methods for stereo matching is that the value of the computed depth is not meaningful along occluding edges where the depth is not well defined. Along these edges the value of the depth is anywhere from the distance of the foreground object's occluding edge to the distance of the background scene point. In particular, for curved objects occluding edges are silhouette edges, and the observed image plane curves in the two images do not correspond to the same physical edge. Unfortunately, strong edges in the image plane are detected only along such occluding edges unless the object has other high-contrast nonoccluding edges or other features. Thus one of the primary problems in recovering depth is the problem of identifying more features distributed throughout the image as candidates for correspondence. One of the many methods developed for finding potential features for correspondence is to identify *interesting points* in both images and match these points in the two images using *region correlation* methods.

Detection of Interesting Points in Regions

In matching points from two images, we need points that can be easily identified and matched in the images. Obviously, the points in a uniform region are not good candidates for matching. The interest operator finds areas of image with high variance. It is expected that there will be enough of such isolated areas in images for matching.

The variances along different directions computed using all pixels in a window centered about a point are good measures of the distinctness of the point along different directions. The directional variances are given by

Figure 3.1: A typical grayscale document image scanned from a bound volume.

connected components into words, and then cluster words into partial curved text lines. We use polynomial regression to compute a pair of top and bottom quadratic reference curves fitting the warped text lines. We next connect the partial straight and warped text lines to form a set of complete text lines. The warped text lines are restored by correcting the quadratic curves accordingly based on the corresponding straight reference lines. The experimental results showed that the proposed method can mostly correct both photometric and geometric distortion.

The rest of this chapter is organized as follows. In Section 3.2, we present a run-length method to detect the shade boundary. In Section 3.3, we present our binarization method. In Section 3.4, we construct the connected components based on 8-neighbors algorithm. In Section 3.5, we present two noise filters based on the size and shape of the connected components respectively. In Section 3.6, we present a novel restoration method to straighten warped text lines based on reference lines/curves. Finally, we summarize this chapter in Section 3.7.

3.2 Detecting Shade Boundary

Note that only the words in the shade are warped and need to be restored, which means we require the knowledge of the boundary between the shade and the clean area. The shade boundary detected in this section will be subsequently used for adjusting the warped text lines to be described in Section 3.6.

We first detect which side of the document image the shade lies on. Two sums of

pixel intensity values, S_{left} and S_{right}, of N vertical pixel columns for left-hand-side and right-hand-side of the image respectively are computed as follows:

$$S_{left} = \sum_{j_1=0}^{N-1} \sum_{i=0}^{I-1} P(i, j_1) \qquad (0 \le j_1 < N, \;\; 0 \le i < I) \tag{3.1}$$

$$S_{right} = \sum_{j_2=J-N}^{J-1} \sum_{i=0}^{I-1} P(i, j_2) \qquad (J - N \le j_2 < J, \;\; 0 \le i < I) \tag{3.2}$$

where

- I, J: The document image height (vertical length) and width (horizontal length) respectively.
- N: The number of vertical columns taking into account. In our system, we set $N = \frac{J}{10}$.
- $P(i, j)$: The pixel intensity value at image indices (i, j).

The shade lies on the side with the smaller sum.

We next scan the image row by row horizontally, starting from the shade side. For each row, a break point, *b*, is found, and the boundary between the shade and the clean area is defined as a set, *B*, of all the break points. Mathematically, *B* can be expressed as:

$$B = \{(i, b) \mid 0 \le i < I, b = L(i, T)\} \tag{3.3}$$

where

- I: The height of the document image.
- $L(i, T)$: A function returning the length of the first run of pixels, whose intensity value is less or equal to *T*, for *ith* horizontal pixel row.
- T: A predefined threshold parameter. In our experiment, we use $T = 0.75 \cdot P_{\max}$,

where $P_{\max}$ is the maximum intensity value over the grayscale image, to obtain a good result for *B*.

Figure 3.2 shows the shade boundary detected for the document image in Figure 3.1.

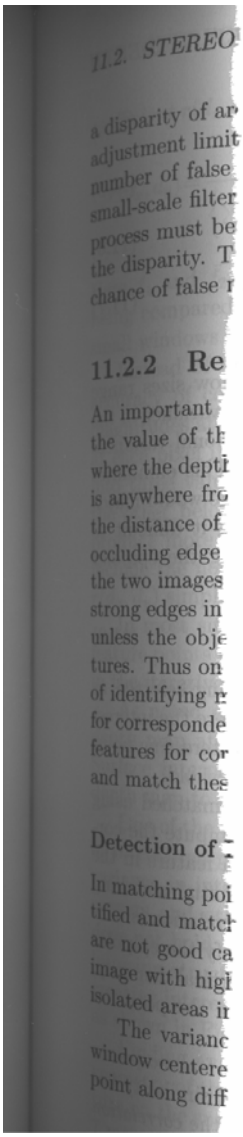


Figure 3.2: The shade boundary detected for the document image in Figure 3.1.

3.3 Binarizing the Document Image

A modified version of Niblack's algorithm is used to remove the shade. Niblack's

method [60] 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 is computed from $T(i, j) = m(i, j) + k \cdot s(i, j)$, where $m(i, j)$ and $s(i, j)$ are the local mean and local standard deviation computed in a window centered at pixel (i, j), and *k* is a user defined parameter and is negative in value.

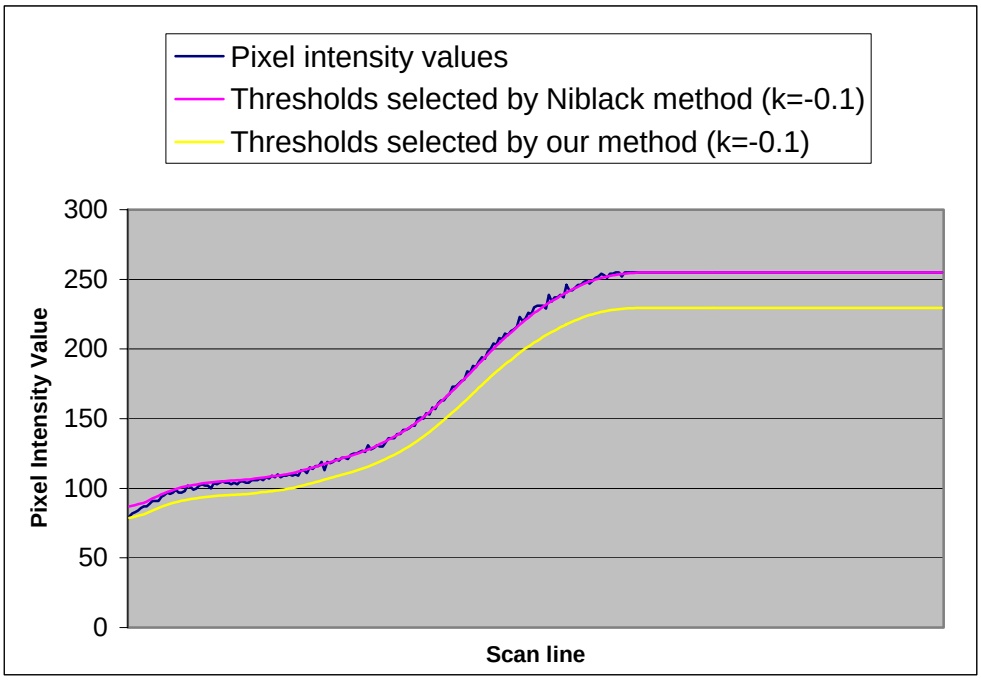
Niblack's method could not be adopted directly for two reasons. One is that Niblack's method is sensitive to the value of *k* for our images. It is quite difficult to find a single *k* value that works for most of our test images. The other is the resultant large amount of pepper noise in the shade area, even if a proper *k* value is chosen. Figure 3.3(a) shows this problem.

In our modification, the local mean, $m(i, j)$, is utilized to multiply, instead of adding, the standard deviation terms. Moreover, each standard deviation, $s(i, j)$, is normalized by dividing it by the dynamic range of standard deviation, *R*. These have the effect of amplifying the contribution of standard deviation, which produces results with much less pepper noise than the original Niblack's method. These modifications also reduce the sensitivity of parameter *k*. Equation (3.4) presents the revised formula.

$$T(i, j) = m(i, j) \cdot [1 + k \cdot (1 - \frac{s(i, j)}{R})] \qquad (3.4)$$

where $m(i, j)$ and $s(i, j)$ are as in Niblack's formula, *R* is computed as the maximum value over all $s(i, j)$ values. We use *k* = -0.1 for grayscale images. Figure 3.3(b) shows the comparison of threshold selection between Niblack's method and the modified method.

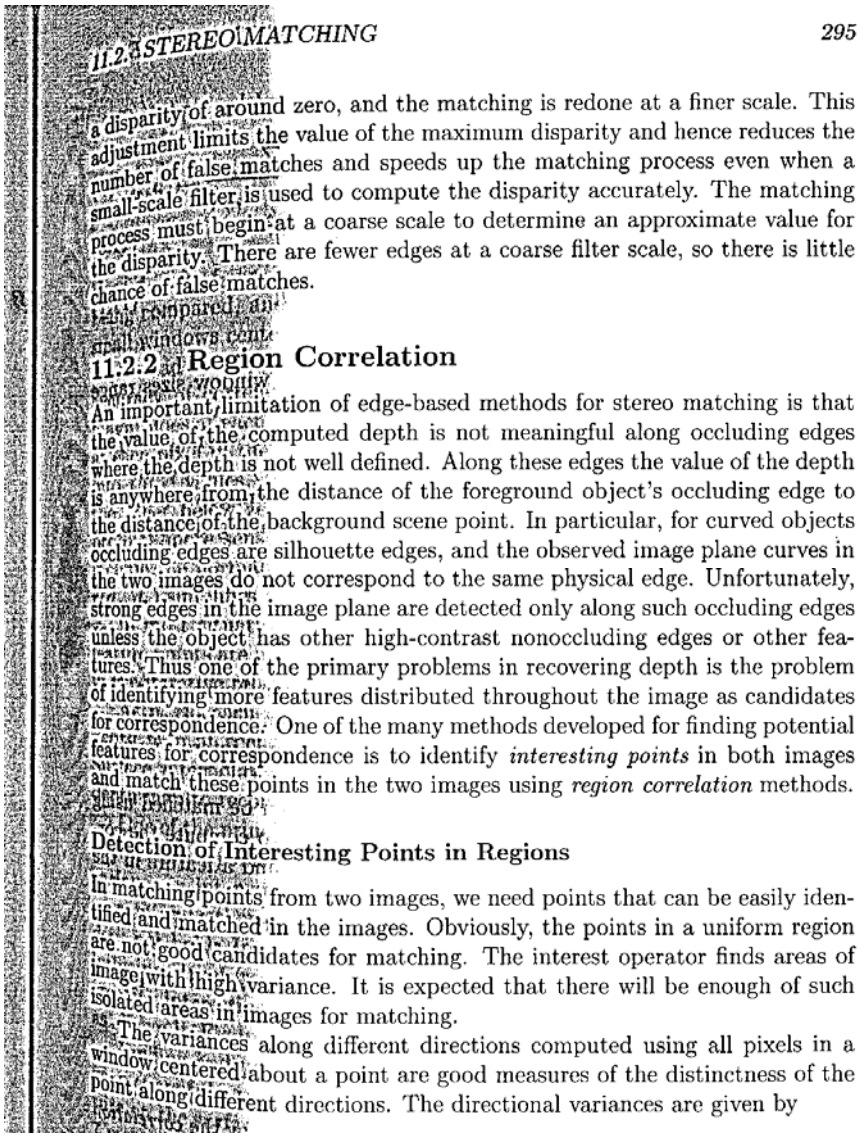
(a)



(b)

Figure 3.3: Comparison of threshold selection: (a) A strip from top margin of the image in Figure 3.1, and the binarization result using Niblack's method; (b) Comparison of threshold candidate selections of an example scanline in (a).

Figures 3.4 and 3.5 show the binarization results by Niblack's method and our method respectively for the document image in Figure 3.1. We observe that there are no blurring, broken, and loss of symbols and text in the binarized image. Furthermore, there is no pepper noise as generated by Niblack's method in Figure 3.4. The photometric distortion for the image in Figure 3.1 is completely removed.



11.2 STEREO MATCHING 295

a disparity of around zero, and the matching is redone at a finer scale. This adjustment limits the value of the maximum disparity and hence reduces the number of false matches and speeds up the matching process even when a small-scale filter is used to compute the disparity accurately. The matching process must begin at a coarse scale to determine an approximate value for the disparity. There are fewer edges at a coarse filter scale, so there is little chance of false matches.

11.2.2 Region Correlation

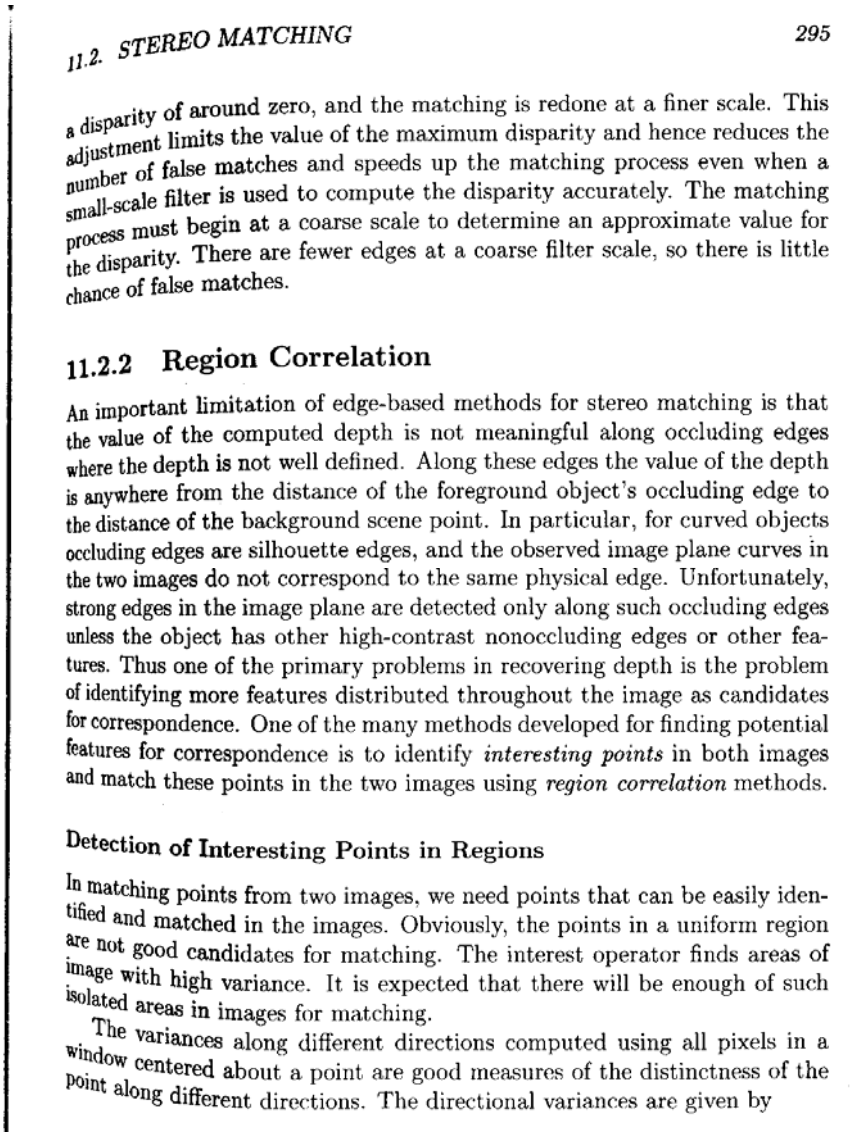
An important limitation of edge-based methods for stereo matching is that the value of the computed depth is not meaningful along occluding edges where the depth is not well defined. Along these edges the value of the depth is anywhere from the distance of the foreground object's occluding edge to the distance of the background scene point. In particular, for curved objects occluding edges are silhouette edges, and the observed image plane curves in the two images do not correspond to the same physical edge. Unfortunately, strong edges in the image plane are detected only along such occluding edges unless the object has other high-contrast nonoccluding edges or other features. Thus one of the primary problems in recovering depth is the problem of identifying more features distributed throughout the image as candidates for correspondence. One of the many methods developed for finding potential features for correspondence is to identify *interesting points* in both images and match these points in the two images using *region correlation* methods.

Detection of Interesting Points in Regions

In matching points from two images, we need points that can be easily identified and matched in the images. Obviously, the points in a uniform region are not good candidates for matching. The interest operator finds areas of image with high variance. It is expected that there will be enough of such isolated areas in images for matching.

The variances along different directions computed using all pixels in a window centered about a point are good measures of the distinctness of the point along different directions. The directional variances are given by

Figure 3.4: The binarization result using Niblack's method for the document image in Figure 3.1.



11.2. STEREO MATCHING 295

a disparity of around zero, and the matching is redone at a finer scale. This adjustment limits the value of the maximum disparity and hence reduces the number of false matches and speeds up the matching process even when a small-scale filter is used to compute the disparity accurately. The matching process must begin at a coarse scale to determine an approximate value for the disparity. There are fewer edges at a coarse filter scale, so there is little chance of false matches.

11.2.2 Region Correlation

An important limitation of edge-based methods for stereo matching is that the value of the computed depth is not meaningful along occluding edges where the depth is not well defined. Along these edges the value of the depth is anywhere from the distance of the foreground object's occluding edge to the distance of the background scene point. In particular, for curved objects occluding edges are silhouette edges, and the observed image plane curves in the two images do not correspond to the same physical edge. Unfortunately, strong edges in the image plane are detected only along such occluding edges unless the object has other high-contrast nonoccluding edges or other features. Thus one of the primary problems in recovering depth is the problem of identifying more features distributed throughout the image as candidates for correspondence. One of the many methods developed for finding potential features for correspondence is to identify *interesting points* in both images and match these points in the two images using *region correlation* methods.

Detection of Interesting Points in Regions

In matching points from two images, we need points that can be easily identified and matched in the images. Obviously, the points in a uniform region are not good candidates for matching. The interest operator finds areas of image with high variance. It is expected that there will be enough of such isolated areas in images for matching.

The variances along different directions computed using all pixels in a window centered about a point are good measures of the distinctness of the point along different directions. The directional variances are given by

Figure 3.5: The binarization result using our method for the document image in Figure 3.1.

3.4 Constructing Connected Components

After binarization, sizable noise may still remain as shown in Figure 3.5 (the book spine in this case). 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 straightening the warped text lines in Section 3.6.

We define the connected components for binary image as follows:

- **8-Neighbors:** For a pixel at (i, j), its *8-neighbors* are pixels at $(i-1, j-1)$, $(i-1, j)$, $(i-1, j+1)$, $(i, j-1)$, $(i, j+1)$, $(i+1, j-1)$, $(i+1, j)$, and $(i+1, j+1)$.
- **8-Path:** An *8-path* from the pixel at (i_0, j_0) to the pixel at (i_n, j_n) is a sequence of pixels (i_0, j_0), (i_1, j_1), (i_2, j_2), …, (i_n, j_n) such that the pixel at (i_k, j_k) is an 8-neighbor of the pixel at (i_{k+1}, j_{k+1}) for all k with $0 \le k < n$.
- **Foreground:** The set of all black pixels in an image is called the *foreground* and is denoted by *S*.
- **Connectivity:** A pixel $p \in S$ is said to be *8-connected* to $q \in S$ if there is an 8-path from *p* to *q* consisting entirely of pixels of *S*.
- **Connected Component:** A set of black pixels in which each pixel is 8-connected to all other pixels is called a *connected component*.

3.5 Noise Filtration

Filters utilizing the property of connected components are applied to remove the noise. Note that most of graphical objects are treated as noise and removed due to the irregular shape. Only text objects remain after noise filtration.

The filter consists of two parts, namely size-filter and shape-filter. The size-filter simply rejects 1) small connected components whose sizes are less than a threshold

value T_1, and 2) big connect components whose sizes are greater than a threshold value T_2. Let *c* denote a connected component, and $S(c)$ is a function returning the number of black pixels in *c*. Mathematically, the size filter F_{size} is represented as follows:

$$F_{size}(c) = \begin{cases} true & if \quad T_1 \le S(c) \le T_2 \\ false & otherwise \end{cases} \tag{3.5}$$

The shape-filter rejects long and irregular-shape connected components that are usually not text characters. The shape-filter "draws" a circle centered at the gravity center of a connected component, and the radius of the circle is decided by the number of black pixels in the connected component and a predefined parameter *s*. The connected components that have black pixel lies outside the permissible range of the shape-filter are then removed. Mathematically, the shape-filter can be expressed as Function (3).

$$F_{shape}(c) = \begin{cases} true & if \quad \sqrt{(i_1 - i_0)^2 + (j_1 - j_0)^2} \le \sqrt{s \cdot S(c)}, \quad \forall (i_1, j_1) \in c \\ false & otherwise \end{cases} \tag{3.6}$$

where

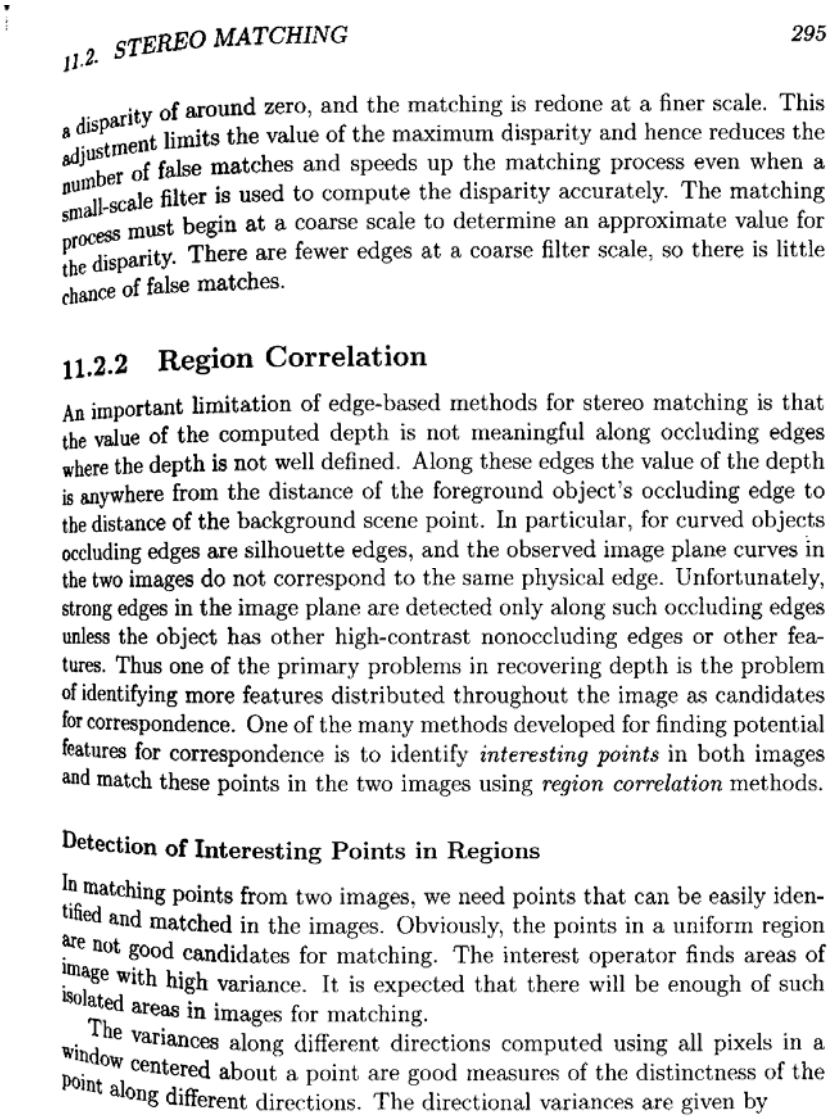
- (i_1, j_1): The image indices for an arbitrary pixel of *c*.
- (i_0, j_0): The image indices for the gravity center of *c*, which are computed as follows:

$$\begin{aligned} i_0 &= \frac{\sum i_1}{S(c)} \\ j_0 &= \frac{\sum j_1}{S(c)} \end{aligned} \qquad for \quad \forall (i_1, j_1) \in c \tag{3.7}$$

- *s*: A predefined parameter. We use *s* = 5 for our test images.

The binarization result after applying the two filters is shown in Figure 3.6. Note that

the book spine and some isolated pixels are successfully removed.



11.2. STEREO MATCHING 295

a disparity of around zero, and the matching is redone at a finer scale. This adjustment limits the value of the maximum disparity and hence reduces the number of false matches and speeds up the matching process even when a small-scale filter is used to compute the disparity accurately. The matching process must begin at a coarse scale to determine an approximate value for the disparity. There are fewer edges at a coarse filter scale, so there is little chance of false matches.

11.2.2 Region Correlation

An important limitation of edge-based methods for stereo matching is that the value of the computed depth is not meaningful along occluding edges where the depth is not well defined. Along these edges the value of the depth is anywhere from the distance of the foreground object's occluding edge to the distance of the background scene point. In particular, for curved objects occluding edges are silhouette edges, and the observed image plane curves in the two images do not correspond to the same physical edge. Unfortunately, strong edges in the image plane are detected only along such occluding edges unless the object has other high-contrast nonoccluding edges or other features. Thus one of the primary problems in recovering depth is the problem of identifying more features distributed throughout the image as candidates for correspondence. One of the many methods developed for finding potential features for correspondence is to identify *interesting points* in both images and match these points in the two images using *region correlation* methods.

Detection of Interesting Points in Regions

In matching points from two images, we need points that can be easily identified and matched in the images. Obviously, the points in a uniform region are not good candidates for matching. The interest operator finds areas of image with high variance. It is expected that there will be enough of such isolated areas in images for matching.

The variances along different directions computed using all pixels in a window centered about a point are good measures of the distinctness of the point along different directions. The directional variances are given by

Figure 3.6: Noise-removed binarization result for the document image in Figure 3.1.

3.6 Straightening the Warped Text Lines

Based on the shade boundary detected in Equation (3.3), we divide the connected components in the binarized image in Figure 3.6 into two classes, namely C_{shade} for the connected components in the shade area and C_{clean} for the rest in the clean area.

In the rest of this section, we model each warped partial text line, formed by the C_{shade} connected components, by a pair of top/bottom quadratic reference curves, and each straight partial text lines, formed by the C_{clean} connected components, by a pair of top/bottom straight reference lines. The warped partial text lines modeled by the quadratic curves are then corrected accordingly based on the corresponding straight reference lines.

3.6.1 Processing the C_{clean} Connected Components

We define the *bounding box* of a connected component *c* as a rectangle, which satisfies:

- The minimum-area rectangle that all the pixels of *c* fall in.
- The four edges of the rectangle are parallel to the four edges of the image boundary respectively.

The bounding box of a connected component *c* can be expressed as one pair of its diagonal corner points: $(i_1(c), j_1(c))$ and $(i_2(c), j_2(c))$, where $i_1(c) < i_2(c)$ and $j_1(c) < j_2(c)$. We next compute the horizontal projection profile *H* by projecting the bounding boxes of the C_{clean} connected components. Mathematically, *H* is a set represented as:

$$H = \{(i,h) \mid 0 \le i < I, h = \sum (j_2(c) - j_1(c)), \forall c \in C_{clean} \quad and \quad i \in [i_1(c), i_2(c)]\} \qquad (3.8)$$

where *I* is the image height. By the 0-runs in *H*, we cluster the C_{clean} connected components into straight partial text lines denoted as $L = \{l_1, l_2, ..., l_N\}$. Figure 3.7(a)

shows the straight partial text lines indicated by the bounding boxes.

For each partial straight text line $l_n \in L \ (1 \le n \le N)$, we classify the connected components $c \in l_n$ into two classes as follows:

$$Class(c) = \begin{cases} Character & if \quad S(c) \ge 0.5 \cdot \overline{S} \\ Symbol & otherwise \end{cases} \tag{3.9}$$

where $\overline{S}$ denotes the average size of the connected components in l_n, and $S(c)$ the number of black pixels in *c*. Let *M* denote the number of connected components belonging to "Character" class in l_n. We then apply linear regression twice to generate a pair of top and bottom straight reference lines to model the alignment of the text. The linear regression takes the $(i_m, j_m) \ (1 \le m \le M)$ indices of the centers of the top/bottom edge of the bounding boxes of "Character" class connected components in l_n as the inputs, and finds the equation $j = k \cdot i + b$ of a top/bottom straight reference line that best fits these top/bottom centers, where *k* and *b* are computed using Equations (3.10) and (3.11) respectively (linear regression formular):

$$k = \frac{(M\sum_{m=1}^{M} i_m j_m - (\sum_{m=1}^{M} i_m)(\sum_{m=1}^{M} j_m))}{(M\sum_{m=1}^{M} i_m^2 - (\sum_{m=1}^{M} i_m)^2)}$$

(3.10)

$$b = \frac{(\sum_{m=1}^{M} j_m - k\sum_{m=1}^{M} i_m)}{M}$$

(3.11)

Figure 3.7(b) shows the two straight reference lines of the last text line in second paragraph extracted from Figure 3.7(a).

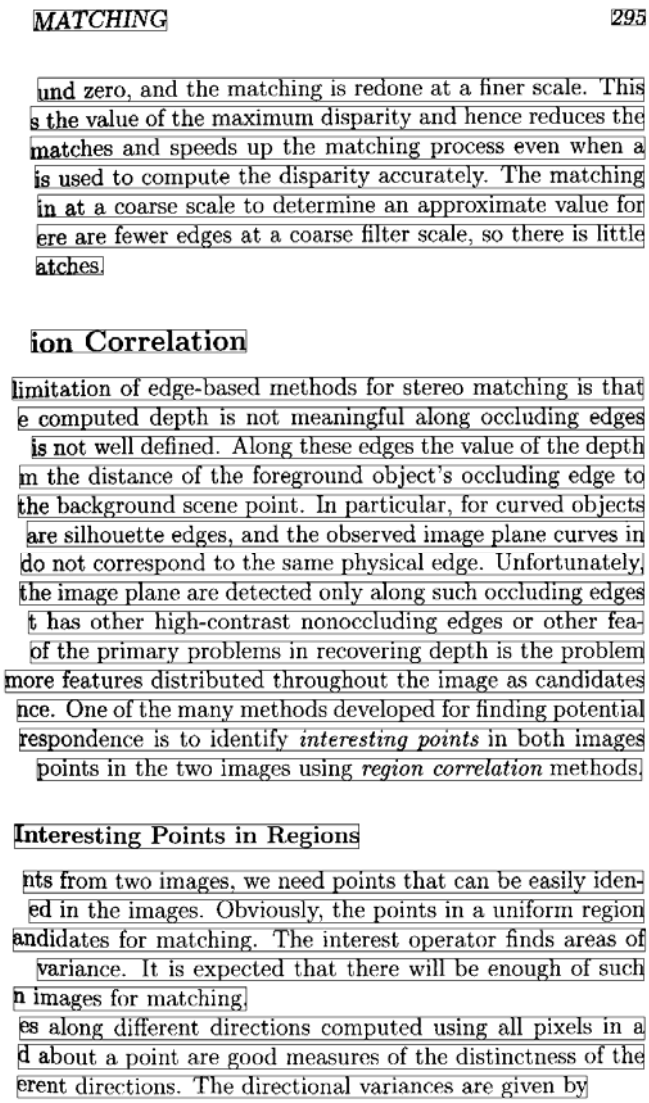


Figure 3.7: Partial straight text lines: (a) The straight partial text lines indicated by the bounding boxes; (b) The two straight reference lines of the last text line in the second paragraph in (a).

In this process, we will compute a threshold distance d^*, which is expected to satisfy the following inequality: inter-character spacing within word < d^* < inter-word spacing. d^* will be used subsequently in Section 3.6.2. We first compute the vertical projection profiles (projecting on the connected component bounding box) for each straight partial text line $l_n \in L$. d^* is determined by analyzing the distribution of the run-length of 0's on the projection profiles. In general, such a run-length distribution is bi-modal. One mode corresponds to the inter-character spacings within word, and the other to the inter-word spacings. d^* can be chosen in the valley between the two dominant histogram modes. Thus we determine the value of d^* based on Otsu's algorithm [65] as follows:

Let *D* denote the maximum run-length of 0's in the profile, and $f_1, f_2, ..., f_D$ the frequencies of run-length *1, 2,…, D* respectively. A threshold *d* ($1 \le d \le D$) partitions the run-length into two classes $C_0 = \{1,2,...,d\}$ and $C_1 = \{d+1, d+2, ..., D\}$. Let $\delta_B^2(d)$ and δ_T^2 be the between-class variance with respect to *d* and the total variance respectively, which are computed as follows:

$$\delta_B^2(d) = w_0 \cdot w_1 \cdot (\mu_1 \cdot \mu_0)^2 \tag{3.12}$$

$$\delta_T^2 = \sum_{i=1}^{D} (i - \mu_T)^2 \cdot p_i \tag{3.13}$$

where

$$w_0 = \sum_{i=1}^{d} (p_i), \qquad w_1 = 1 - w_0, \qquad p_i = \frac{f_i}{\sum_{j=1}^{D} f_j},$$

$$\mu_1 = \frac{\mu_T - \mu_d}{w_1}, \qquad \mu_0 = \frac{\mu_d}{w_0}, \qquad \mu_d = \sum_{i=1}^{d} (i \cdot p_i), \qquad \mu_T = \sum_{i=1}^{D} i \cdot p_i .$$

We define a criterion function with respect to *d* as follows:

$$\eta(d) = \frac{\delta_B^2(d)}{\delta_T^2} \tag{3.14}$$

Then $\eta(d)$ is computed for all possible values of *d*, and the *d* that gives the smallest η is the optimal threshold. Thus

$$d^* = \arg \min_d \eta(d) \tag{3.15}$$

3.6.2 Processing the C_{shade} Connected Components

We first cluster the C_{shade} connected components into words using d^*, detected in the last section, as the threshold distance. Two connected components are in one word if the Euclidean distance between them is less than d^*. We group the words into text lines by using a modified "box-hands" method. The reason we could not adopt the original "box-hands" method [83] is that it works when the text line is a straight horizontal line, which is obviously not the case for the partial curved text line in the shade. As shown in Figure 3.8(a), we extend the bounding box by adding to their left and right sides two equal parallel-quadrilateral extensions, called the "box hands", defined as follows:

- The width and height of the hands are equal to the height and half of the height of the word bounding box respectively.
- The orientation of the hands is determined by the orientation of the word. We approximate the word orientation by Equation (3.10) and (3.11), using the image indices of the center of "Character" connected components (classified by Equation (3.9) by setting $\overline{S}$ to be the average size of the connected components

in current word) in the word.

- The two box-hands are positioned at the mid-point of the touching side of the word bound box.

A partial curved text line is detected by clustering all the words whose hands of bounding boxes touch each other. Figure 3.8(b) shows the partial curved text line clustered. We denote the partial text lines as $L' = \{l_1', l_2', ..., l_{N'}'\}$.

For each curved text line $l_{n'}' \in L'$ $(1 \le n' \le N')$, we approximate the alignment of the C_{shade} connected components by two quadratic reference curves. Let M' denote the number of connected components belonging to "Character" class (classified by Equation (3.9)) in $l_{n'}'$. Polynomial regression using least mean square is applied twice with the $(i_{m'}, j_{m'})$ $(1 \le m' \le M')$ image indices of the centers of the top and bottom edges of bounding boxes of "Character" connected components in $l_{n'}'$ as the inputs respectively.

Suppose that the quadratic polynomial equation is $j = a_2 i^2 + a_1 i + a_0$, then $j_{m'}' = a_2 i_{m'}^2 + a_1 i_{m'} + a_0$ is the estimated value given by the quadratic function. The least mean square error *E* between the estimated value $j_{m'}'$ and the true value $j_{m'}$ is represented as:

$$E = \sum_{m'=1}^{M'} (j_{m'} - j_{m'}')^2 = \sum_{m'=1}^{M'} (j_{m'} - a_2 i_{m'}^2 - a_1 i_{m'} - a_0)^2 \qquad (3.16)$$

Hence:

$$\frac{\partial E}{\partial a_2} = \sum_{m'=1}^{M'} (i_{m'}^4 a_2 + i_{m'}^3 a_1 + i_{m'}^2 a_0 - i_{m'}^2 j_{m'}) \qquad (3.17)$$

$$\frac{\partial E}{\partial a_1} = \sum_{m'=1}^{M'} (i_{m'}^3 a_2 + i_{m'}^2 a_1 + i_{m'} a_0 - i_{m'} j_{m'}) \tag{3.18}$$

$$\frac{\partial E}{\partial a_0} = \sum_{m'=1}^{M'} (i_{m'}^2 a_2 + i_{m'} a_1 + a_0 - j_{m'}) \tag{3.19}$$

Equating these partial derivatives to zero and solve the three simultaneous equations. We obtain the values of coefficients a_2, a_1 and a_0:

$$a_2 = \frac{t_3 t_4 - t_2 t_5}{t_1 t_3 - t_2^2} \tag{3.20}$$

$$a_1 = \frac{t_2 t_4 - t_1 t_5}{t_2^2 - t_1 t_3} \tag{3.21}$$

$$a_0 = \frac{\sum_{m'=1}^{M'} j_{m'} - a_2 \sum_{m'=1}^{M'} i_{m'}^2 - a_1 \sum_{m'=1}^{M'} i_{m'}}{M'} \tag{3.22}$$

where

$$t_1 = M' \sum_{m'=1}^{M'} i_{m'}^4 - (\sum_{m'=1}^{M'} i_{m'}^2)^2 \tag{3.23}$$

$$t_2 = M' \sum_{m'=1}^{M'} i_{m'}^3 - \sum_{m'=1}^{M'} i_{m'}^2 \sum_{m'=1}^{M'} i_{m'} \tag{3.24}$$

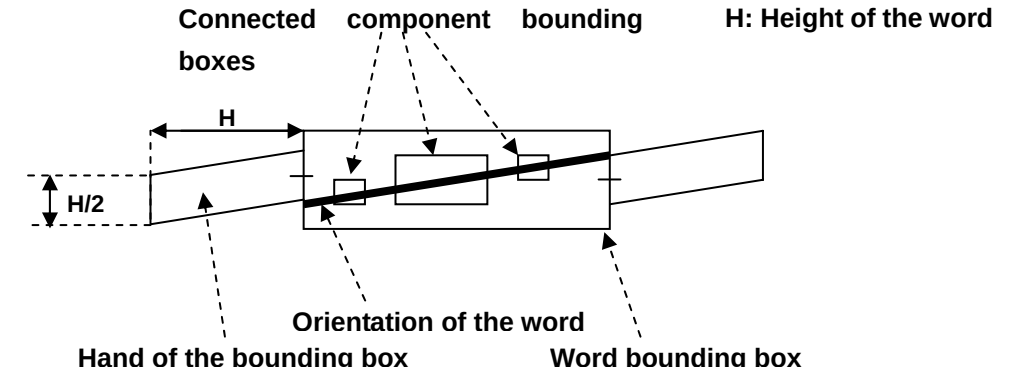
$$t_3 = M' \sum_{m'=1}^{M'} i_{m'}^2 - (\sum_{m'=1}^{M'} i_{m'})^2 \tag{3.25}$$

$$t_4 = M' \sum_{m'=1}^{M'} (i_{m'}^2 j_{m'}) - \sum_{m'=1}^{M'} i_{m'}^2 \sum_{m'=1}^{M'} j_{m'} \tag{3.26}$$

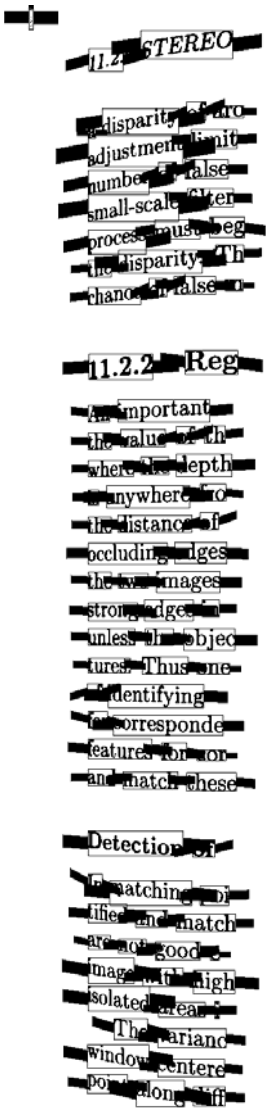
$$t_5 = M' \sum_{m'=1}^{M'} (i_{m'} j_{m'}) - \sum_{m'=1}^{M'} i_{m'} \sum_{m'=1}^{M'} j_{m'} \tag{3.27}$$

Figure 3.8(c) shows two quadratic reference curves for the last text line extracted from the second paragraph in Figure 3.8(b). Note that for each quadratic curve we

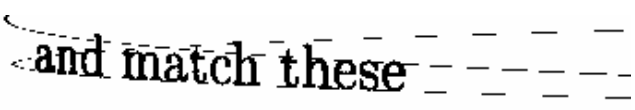
only use the half on which the input points lie.



(a)



(b)



(c)

Figure 3.8: Box-hands approach and partial curved text lines: (a) "Box-hands"— extension of word bounding box; (b) Partial curved text lines; (c) Two quadratic reference curves for the last text line in the second paragraph in (b).

3.6.3 Straightening the Warped Text Lines

The partial curved and straight text lines, $L' = \{l_1', l_2', ..., l_{N'}'\}$ and $L = \{l_1, l_2, ..., l_N\}$, are naturally combined in to a set of complete text lines by adding two horizontal hands of the bounding box of $\forall l_n \in L$ as shown in Figure 3.9. For each complete text line, we restore $\forall l_{n'} \in L'$ by adjusting the location and orientation of each C_{shade} connected component in $l_{n'}$.

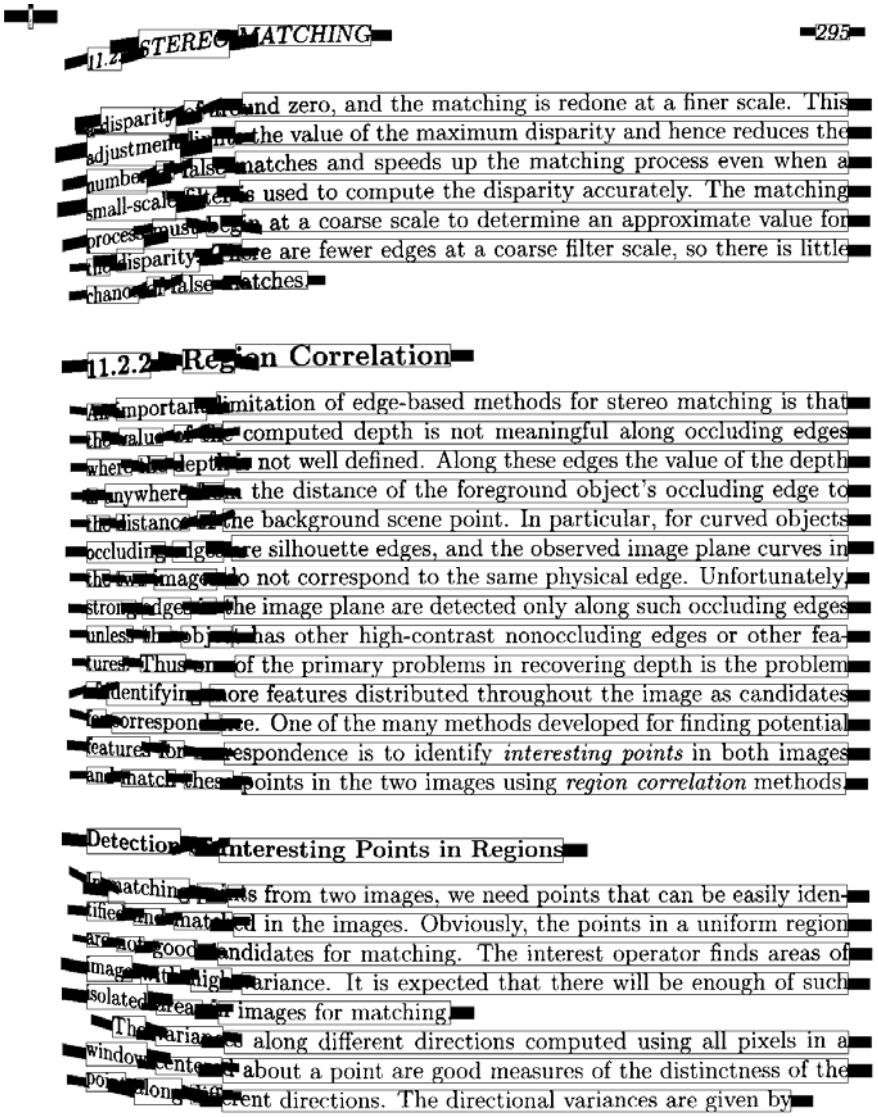
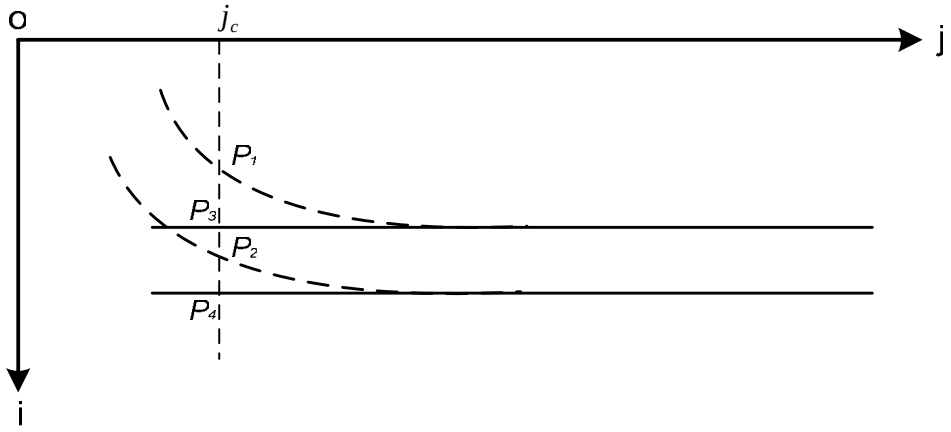


Figure 3.9: The complete text lines.

The location adjustment is done by calculating the relative position of this connected component with the two quadratic reference curves, and then vertically moving it to the same relative position with the two straight reference lines. Figure 3.10(a) shows a complete text line formed by the partial straight and curved text lines in Figures 3.7(b) and 3.8(c), and its four reference lines/curves. Figure 3.10(b) shows that suppose a vertical line $j = j_c$, which crosses the center (i_c, j_c) of the bounding box of a C_{shade} connected component *c*, intersects the two quadratic curves (dash line) at P_1 and P_2, and the two straight lines (solid line) at P_3 and P_4 respectively. The image indices of P_1, P_2, P_3, and P_4 can be computed using (i_c, j_c) by knowing the functions of the quadratic curves and the straight lines. The connected

(a)



(b)

Figure 3.10: Straightening the text lines:
(a) A complete text line formed by Figures 3.7(b) and 3.8(c);
(b) The image indices system for the text line.

component c is moved along $j = j_c$, such that the new indices $(i_c^{'}, j_c)$ of the bounding box center satisfy the following equation, i.e. we keep the relative position of *c* unchanged:

$$\frac{i_c - i_{P_1}}{i_{P_2} - i_{P_1}} = \frac{i_c^{'} - i_{P_3}}{i_{P_4} - i_{P_3}} \tag{3.28}$$

Thus

$$i_c^{'} = \frac{i_c - i_{P_1}}{i_{P_2} - i_{P_1}} \cdot (i_{P_4} - i_{P_3}) + i_{P_3} \tag{3.29}$$

Let $j = a_2 i^2 + a_1 i + a_0$ and $j = b_2 i^2 + b_1 i + b_0$ denote the functions for the top and bottom quadratic curves respectively. By the first order differentiation, the slopes at P_1 and P_2 are $s_{P_1} = 2a_2 i_{P_1} + a_1$ and $s_{P_2} = 2b_2 i_{P_2} + b_1$ respectively. The skew of *c* is estimated by the average of s_{p_1} and s_{p_2}:

$$s_c = \frac{s_{P_1} + s_{P_2}}{2} = a_2 i_{P_1} + b_2 i_{P_2} + \frac{a_1 + b_1}{2} \tag{3.30}$$

The orientation adjustment is done by rotating the C_{shade} connected component *c* by $-\arctan s_c$ degree.

The final restoration result using our system is shown in Figure 3.11. Both photometric and geometric distortion is generally corrected.

a disparity of around zero, and the matching is redone at a finer scale. This adjustment limits the value of the maximum disparity and hence reduces the number of false matches and speeds up the matching process even when a small-scale filter is used to compute the disparity accurately. The matching process must begin at a coarse scale to determine an approximate value for the disparity. There are fewer edges at a coarse filter scale, so there is little chance of false matches.

11.2.2 Region Correlation

An important limitation of edge-based methods for stereo matching is that the value of the computed depth is not meaningful along occluding edges where the depth is not well defined. Along these edges the value of the depth is anywhere from the distance of the foreground object's occluding edge to the distance of the background scene point. In particular, for curved objects occluding edges are silhouette edges, and the observed image plane curves in the two images do not correspond to the same physical edge. Unfortunately, strong edges in the image plane are detected only along such occluding edges unless the object has other high-contrast nonoccluding edges or other features. Thus one of the primary problems in recovering depth is the problem of identifying more features distributed throughout the image as candidates for correspondence. One of the many methods developed for finding potential features for correspondence is to identify *interesting points* in both images and match these points in the two images using *region correlation* methods.

Detection of Interesting Points in Regions

In matching points from two images, we need points that can be easily identified and matched in the images. Obviously, the points in a uniform region are not good candidates for matching. The interest operator finds areas of image with high variance. It is expected that there will be enough of such isolated areas in images for matching.

The variances along different directions computed using all pixels in a window centered about a point are good measures of the distinctness of the point along different directions. The directional variances are given by

Figure 3.11: The final restoration result for the image in Figure3.1.

3.6.4 Discussion

The proposed horizontal projection profile method to segment the C_{clean} connected components into partial straight text lines in Section 3.6.1 will be replaced by the modified box-hands method as proposed in Section 3.6.2 under the following two conditions:

- Large document skew: The document skew is significant, such that the horizontal projection of C_{clean} connected components belonging to different text lines overlap in the horizontal projection profile.
- Multi-column document: If the document image contains more than one column of text, and the text lines in different columns are not in the same horizontal line.

Figure 3.12 shows how the box-hands method clusters both C_{clean} and C_{shade} connected components into complete text lines for a double column document image with large document skew.

In general case, we do not apply the box-hands method to cluster the C_{clean} connected components into text lines, due to the following reasons:

- The number of C_{clean} connected components is much greater than the number of C_{shade} connected components.
- The computational cost for box-hands method is very expensive, since we have to 1) cluster the connected components into words, 2) apply linear regression to approximate the word orientation and construct the hands for each word, and 3) check whether one hand of a word touches any hand from all the other words.

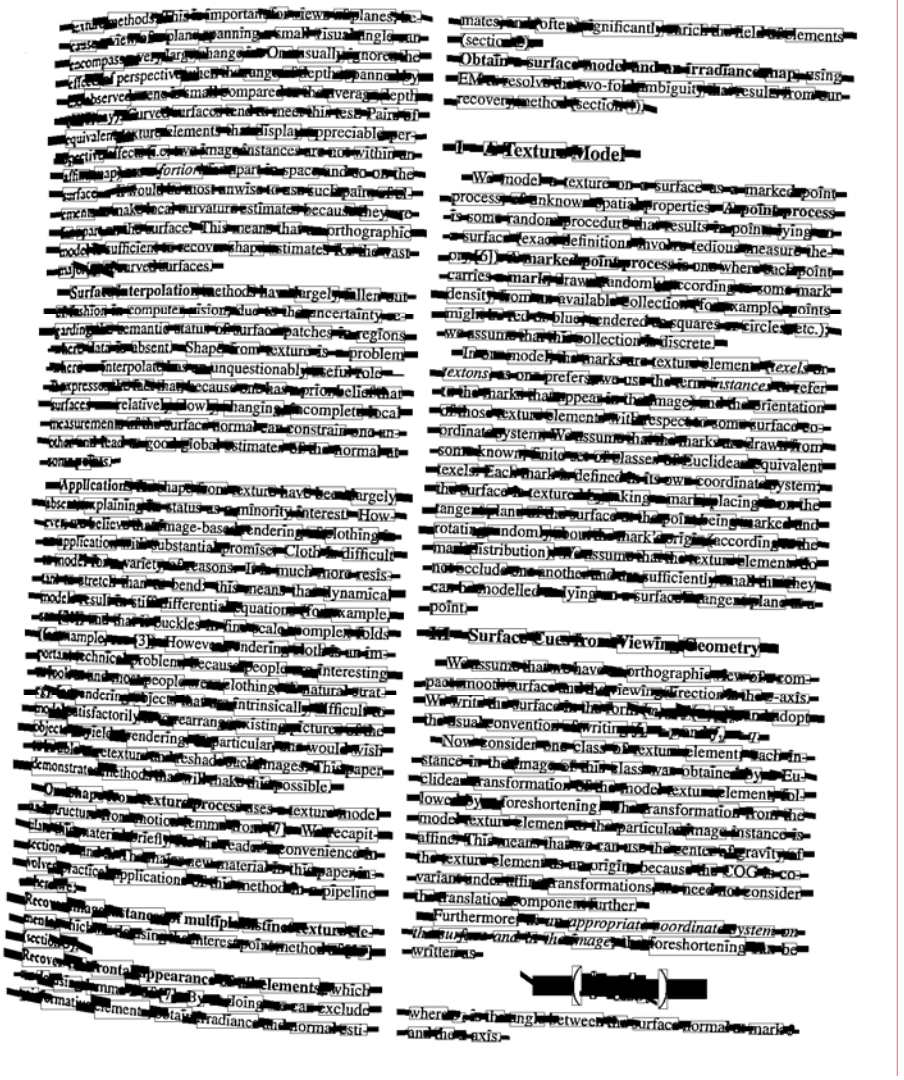


Figure 3.12: The complete text lines clustered by box-hands method for a double column document image with large document skew.

3.7 Summary

In this chapter, we have proposed a novel restoration method for the document images scanned from bound volumes. We first introduce a run-length method to detect the shade boundary in the grayscale image. We remove the photometric distortion by

binarizing the image using a modified Niblack method. Connected component based on 8-neighbor algorithm are constructed for noise filtration and subsequent restoration. We propose a novel method for word clustering and a modified box-hands method for text line clustering. We correct the geometric distortion by straightening the warped text lines based on reference line/curve detection.

As mentioned in Chapter 1, this method has the following two disadvantages:

- The shapes of the characters are not changed. In the resulting images, while the orientation and location of the characters in the shade are restored, the shapes of these characters may still appear distorted and narrower than the ones in the other region.
- The graphical objects in the binarized document image, such as diagrams, figures, charts, tables, and so on, are removed by the size and shape filters, cannot be restored by the proposed method.

Thus the geometric distortion is partially, but not all, corrected. In next chapter, with the scanner information as a priori knowledge, we propose another better restoration method based on discovering the 3D document shape, which can completely correct the photometric and geometric distortion.

The work reported in this chapter has led to the following publications:

[109] (ICIP 01): A restoration method based our binarization method and box-hands approach.

[108] (ICDAR 01): An improvement work of [109] with the word orientation approximated by linear regression.

[107] (ICIP 02): An improvement work of [108] with the text line straightening method based on the reference lines/curves.

[106] (ICDAR 03): An optimization of [107] with the horizontal projection profile for straight text line outside in the shade.

Chapter 4
DIR based on 3D Document Shape Discovery

4.1 Introduction

Figure 4.1 shows a typical grayscale document image containing graphical objects scanned from a skew bound volume. Note the shade along the book spine area and the warping in the shade. The restoration method proposed in Chapter 3 cannot correct the distortion for graphical objects, and has to adopt the computational expensive box-hands method for all the connected components due to the document skew.

In this chapter, with the scanner information (gain and bias, focal length, tilt angle of the light source, and so on) as a priori knowledge, we propose a restoration method to uncover the geometrical and optical information with respect to the book surface to enable us to transform the warped book surface image into its flattened rendition. This method is more efficient and effective, tolerable to document skew, and there is no need for image transformation from grayscale to binary.

From a technical point of view, this shape from shading (SFS) problem in real world environments is characterized by 1) proximal and moving light source, 2) Lambertian reflection, 3) nonuniform albedo distribution, and 4) document skew. Taking all these factors into account, we first build practical models (consisting of a 3D geometric model and a 3D optical model) for the practical scanning conditions. Based on the practical models, we propose a method to reconstruct the shape and

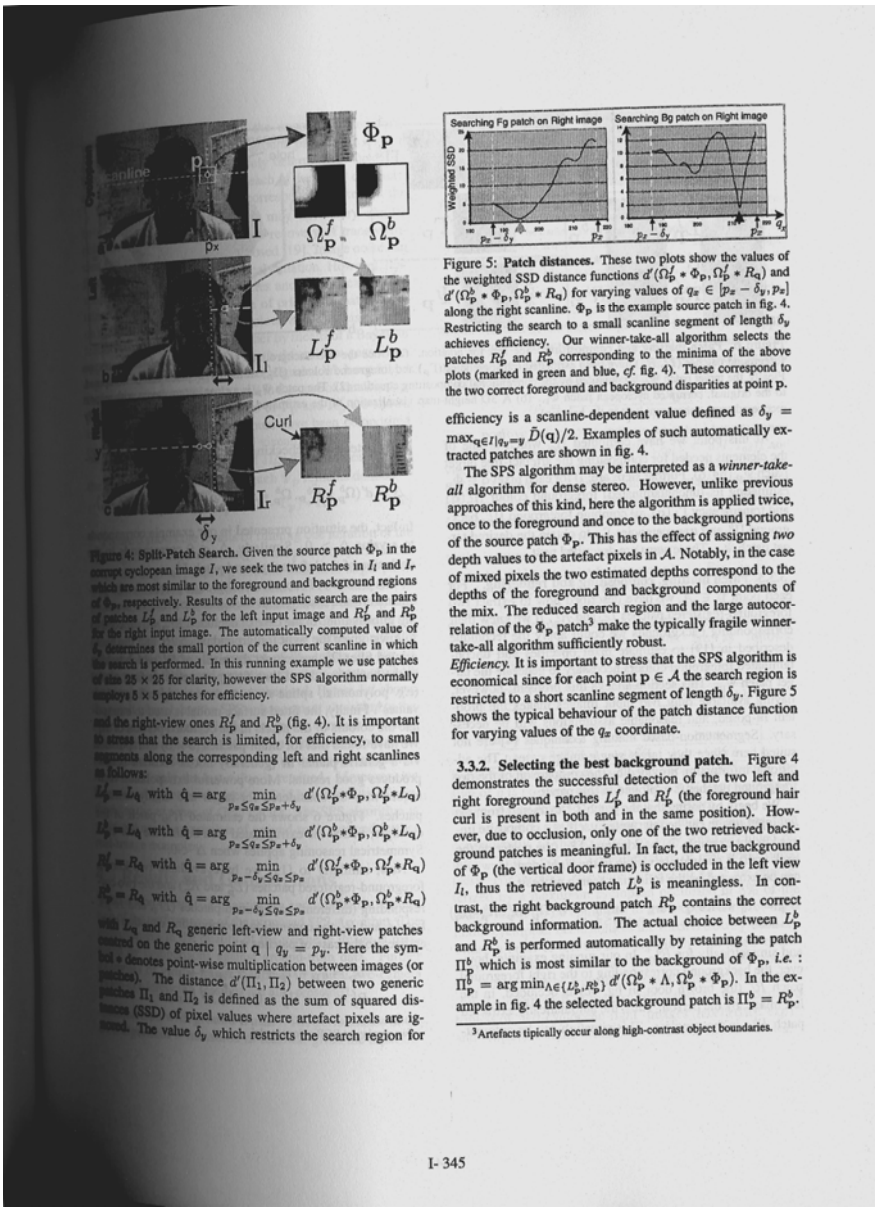


Figure 4: **Split-Patch Search.** Given the source patch Φ_p in the correct cyclopean image I, we seek the two patches in I_l and I_r which are most similar to the foreground and background regions of Φ_p, respectively. Results of the automatic search are the pairs of patches L_p^f and L_p^b for the left input image and R_p^f and R_p^b for the right input image. The automatically computed value of δ_y determines the small portion of the current scanline in which the search is performed. In this running example we use patches of size 25 × 25 for clarity, however the SPS algorithm normally employs 5 × 5 patches for efficiency.

and the right-view ones R_p^f and R_p^b (fig. 4). It is important to stress that the search is limited, for efficiency, to small segments along the corresponding left and right scanlines as follows:

$$L_p^f = L_q \text{ with } \hat{q} = \arg \min_{p_x \le q_x \le p_x + \delta_y} d'(\Omega_p^f * \Phi_p, \Omega_p^f * L_q)$$

$$L_p^b = L_q \text{ with } \hat{q} = \arg \min_{p_x \le q_x \le p_x + \delta_y} d'(\Omega_p^b * \Phi_p, \Omega_p^b * L_q)$$

$$R_p^f = R_q \text{ with } \hat{q} = \arg \min_{p_x - \delta_y \le q_x \le p_x} d'(\Omega_p^f * \Phi_p, \Omega_p^f * R_q)$$

$$R_p^b = R_q \text{ with } \hat{q} = \arg \min_{p_x - \delta_y \le q_x \le p_x} d'(\Omega_p^b * \Phi_p, \Omega_p^b * R_q)$$

with L_q and R_q generic left-view and right-view patches centred on the generic point $q \mid q_y = p_y$. Here the symbol $*$ denotes point-wise multiplication between images (or patches). The distance $d'(\Pi_1, \Pi_2)$ between two generic patches Π_1 and Π_2 is defined as the sum of squared distances (SSD) of pixel values where artefact pixels are ignored. The value δ_y which restricts the search region for

Figure 5: **Patch distances.** These two plots show the values of the weighted SSD distance functions $d'(\Omega_p^f * \Phi_p, \Omega_p^f * R_q)$ and $d'(\Omega_p^b * \Phi_p, \Omega_p^b * R_q)$ for varying values of $q_x \in [p_x - \delta_y, p_x]$ along the right scanline. Φ_p is the example source patch in fig. 4. Restricting the search to a small scanline segment of length δ_y achieves efficiency. Our winner-take-all algorithm selects the patches R_p^f and R_p^b corresponding to the minima of the above plots (marked in green and blue, *cf.* fig. 4). These correspond to the two correct foreground and background disparities at point p.

efficiency is a scanline-dependent value defined as $\delta_y = \max_{q \in I | q_y = y} \bar{D}(q)/2$. Examples of such automatically extracted patches are shown in fig. 4.

The SPS algorithm may be interpreted as a *winner-take-all* algorithm for dense stereo. However, unlike previous approaches of this kind, here the algorithm is applied twice, once to the foreground and once to the background portions of the source patch Φ_p. This has the effect of assigning *two* depth values to the artefact pixels in $\mathcal{A}$. Notably, in the case of mixed pixels the two estimated depths correspond to the depths of the foreground and background components of the mix. The reduced search region and the large autocorrelation of the Φ_p patch[3] make the typically fragile winner-take-all algorithm sufficiently robust.

Efficiency. It is important to stress that the SPS algorithm is economical since for each point $p \in \mathcal{A}$ the search region is restricted to a short scanline segment of length δ_y. Figure 5 shows the typical behaviour of the patch distance function for varying values of the q_x coordinate.

3.3.2. Selecting the best background patch. Figure 4 demonstrates the successful detection of the two left and right foreground patches L_p^f and R_p^f (the foreground hair curl is present in both and in the same position). However, due to occlusion, only one of the two retrieved background patches is meaningful. In fact, the true background of Φ_p (the vertical door frame) is occluded in the left view I_l, thus the retrieved patch L_p^b is meaningless. In contrast, the right background patch R_p^b contains the correct background information. The actual choice between L_p^b and R_p^b is performed automatically by retaining the patch Π_p^b which is most similar to the background of Φ_p, *i.e.* : $\Pi_p^b = \arg \min_{\Lambda \in \{L_p^b, R_p^b\}} d'(\Omega_p^b * \Lambda, \Omega_p^b * \Phi_p)$. In the example in fig. 4 the selected background patch is $\Pi_p^b = R_p^b$.

[3] Artefacts tipically occur along high-contrast object boundaries.

I- 345

Figure 4.1: A grayscale image containing graphical objects scanned from a skew bound document.

albedo distribution of the book surface. We next correct the photometric distortion using a de-shading model based on the albedo distribution. Finally, using a de-warping model based on the book surface shape, we tackle the geometric distortion by correcting: 1) the vertical geometric distortion caused by perspective projection, 2) the horizontal geometric distortion caused by orthogonal projection, and 3) the document skew.

The rest of this chapter is organized as follows. In Section 4.2, we present two practical modes for real-world scanning conditions. In Section 4.3, we reduce the 3D surface shape to a unique 2D cross section shape. In Section 4.4, we reconstruct the cross section shape and the albedo distribution of the book surface. In Section 4.5, we introduce two models, namely de-shading and de-warping models, to restore the document image based on the albedo distribution and cross section shape. We summarize this chapter in Section 4.6.

4.2 Practical Models

We present the practical scanning conditions in real-world in Figure 4.2. Figure 4.2(a) shows the structure of the image scanner and the coordinate system of the book surface. The image scanner consists of a light source L, a linear CCD sensor C, a mirror M and a lens E. The sensor C takes a 1D image $P(i)$ ($0 \le i < image\ \ length\ \ I$) along the scanning line S and moves with L, M, and E. The sequence $P(i)$ forms a 2D image $P(i, j)$ ($0 \le j < image\ \ width\ \ J$). Note

that in real scanning conditions, usually the book to be scanned may not be aligned strictly parallel to the scanning light, i.e. the scanning line *S* may not be parallel to the book spine. Let ε be the angle between the scanning line *S* and the projection of book spine on the scanning plane. As shown in Figure 4.2(a), we define the 3D coordinate system of the book surface as follows:

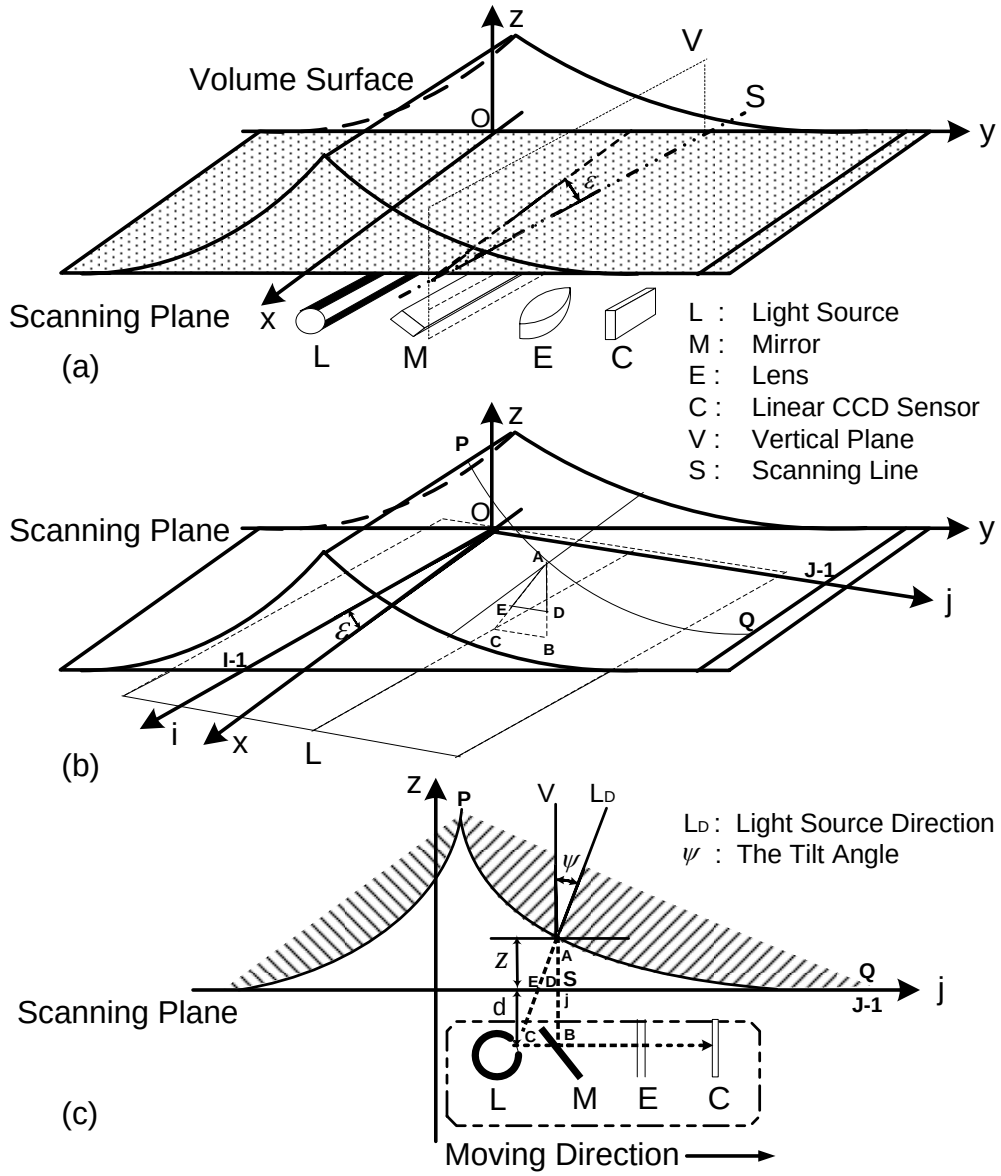


Figure 4.2: The practical scanning conditions.

- The origin O is located at the point on the scanning plane corresponding to $P(0,0)$ of the 2D image.
- The x-axis is located at the scanning plane and parallel to the book spine. It has positive values for the book spine.
- The y-axis is located at the scanning plane and parallel to the horizontal book boundary. It has positive values for the book surface captured by the scanner.
- The z-axis is perpendicular to the x-y plane, and has positive values for the book surface.

Thus, the image coordinate system and the book surface coordinate system share the same vertical axis, i.e. z axis. The transformation between the i-j image indices and x-y book surface coordinates is shown in Figure 4.3. As defined earlier, ε is the angle between *i-axis* and *x-axis*, and ε has positive values for counterclockwise document skew, and negative values for clockwise document skew. We then define the function $x(i, j)$ and $y(i, j)$, which return the corresponding x and y coordinates for $P(i, j)$ respectively, as follows:

$$x(i, j) = \sqrt{i^2 + j^2} \cdot \cos(\arctan\frac{j}{i} - \varepsilon) \tag{4.1}$$

$$y(i, j) = \sqrt{i^2 + j^2} \cdot \sin(\arctan\frac{j}{i} - \varepsilon) \tag{4.2}$$

where $0 \le i < I$ and $0 \le j < J$. Similarly, the functions $i(x, y)$ and $j(x, y)$, which return the corresponding image index (i, j) for given coordinates (x, y), are defined as:

$$i(x, y) = \sqrt{x^2 + y^2} \cdot \cos(\arctan\frac{y}{x} + \varepsilon) \tag{4.3}$$

$$j(x,y) = \sqrt{x^2 + y^2} \cdot \sin(\arctan\frac{y}{x} + \varepsilon) \qquad (4.4)$$

Note that the results of Equations 4.1-4.4 are rounded to integers

The document skew ε is detected by the following three steps:

Step 1. Detect the shade boundary using the run-length method in Chapter 3, Section 3.2. The shade boundary detected for the image in Figure 4.1 is shown in Figure 4.4.

Step 2. Extract the parts of the boundary that on the top and bottom margins, whose length is $I/10$ respectively, as shown in Figure 4.4, since we observe that most books have at least $I/10$ top and bottom unprinted margins. Note that even there are some graphical or textual contents in the margin we extracted, it will not affect the detection process too much, since we will adopt Hough

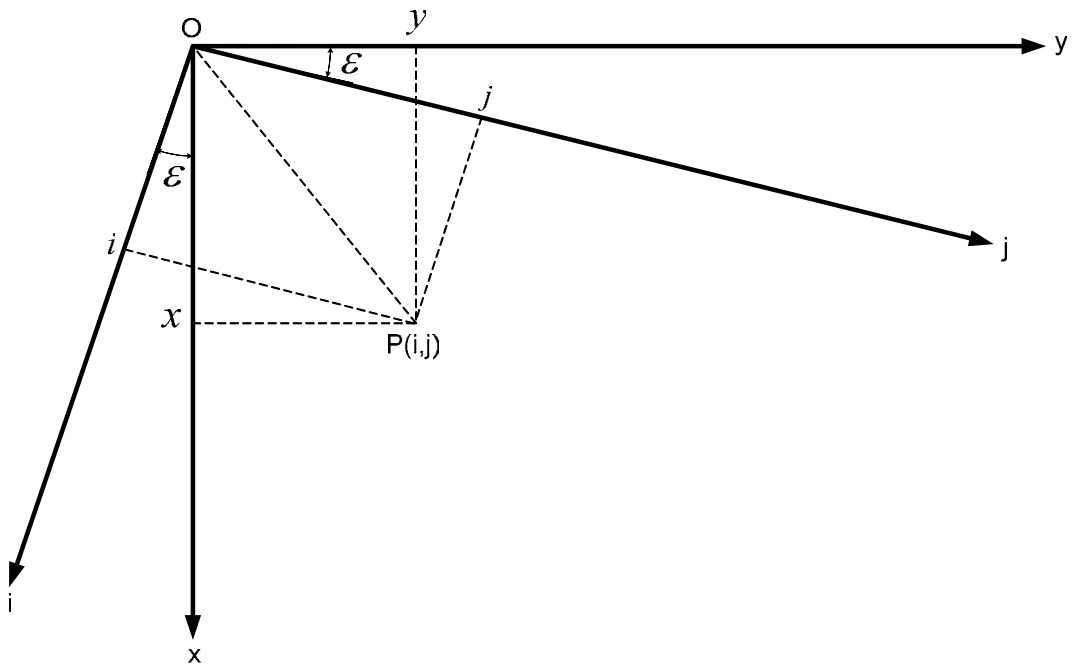


Figure 4.3: Transformation between the *I-w* image indices and the *x-y* coordinates.

Transform in the next step, which is robust to outliers, instead of Linear Regression, which is sensitive to outliers.

Step 3. Apply Hough Transform [79] on the extracted parts of the boundary to find the straight line L fitting maximum number of the shade boundary pixels. If the number of pixels on L over the number of pixels on the extracted partial boundary exceeds 50%, we set $\varepsilon =$ the angle between L and the vertical image boundary, otherwise, we set $\varepsilon = 0$.

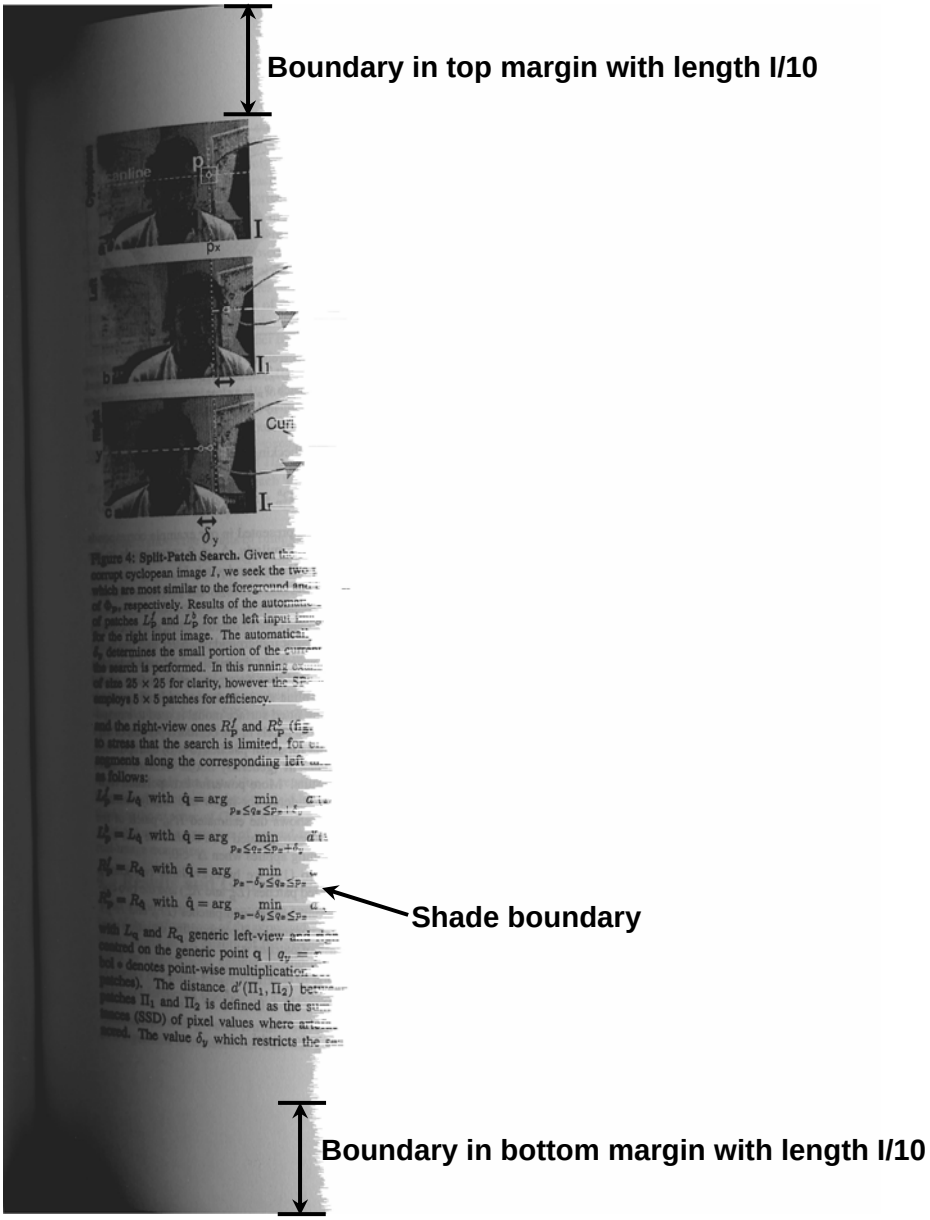


Figure 4.4: The shade boundary detected for Figure 4.1.

Figure 4.2(b) shows a condition where a point *C* on the directional linear light source *L* of the scanner casts a beam of light on a point *A* on the book surface. *E* is the intersection point of line *AC* and the scanning plane. *B* and *D* are the projection points of *A* on the scanning plane and the imaging plane (where the light source and CCD array lie on) respectively. Thus plane *ABC* is perpendicular to *L*, and curve *PAQ* is the intersection of plane *ABC* and the book surface. Note that curve *PAQ* is the cross section shape of the book surface on the *j*-*z* plane instead of the *y*-*z* plane, since the plane *ABC* is perpendicular to the light source *L* (also the *i*-*axis*) instead of *x*-*axis* (also the book spine). Figure 4.2(c) shows the cross section shape of the book surface in the *j*-*z* plane, where *V* is the vertical to the scanning plane, L_D the light source direction, ψ the tilt angle between the light source and the normal to the scanning plane, *z* the distance from *A* to the scanning plane, *d* the distance from scanning plane to light source.

We specify our problem by two practical models: a 3D geometric model and a 3D optical model. We introduce the following assumptions in our practical model:

Assumption 1: The cross section shape of the book surface is smooth on the *y*-*z* plane (except the points on the book spine) and constant along the *x*-axis. This is reasonable as the book pages are generally uncreased and the book is normally leveled because of its uniform weight.

Assumption 2: The book surface is Lambertian, i.e. no specular reflections and uniform brightness of reflected light in all directions. This is also reasonable as most papers are nearly Lambertian.

4.2.1 The 3D Geometric Model

Suppose the 3D point *A* in Figure 4.2(b) corresponds to the pixel $P(i, j)$ in the scanned document image. By Assumption 1, the *z* values are constant along *x*-axis. Thus the 3D coordinates of *A* in the *x*-*y*-*z* space can be represented as $(x(i, j), y(i, j), z(y(i, j)))$, where $x(i, j)$ and $y(i, j)$ are calculated from Equations (4.1) and (4.2) respectively. Figure 4.5(a) shows a cross section shape P'AQ' of the book surface at $x(i, j)$ in *y*-*z* plane. Note that plane *PAQ* in Figure 4.2(b) is perpendicular to the *i*-axis (also the light source), while plane *P'AQ'* here is perpendicular to the *x*-axis (also the book spine). *I* and *J* denote the image length and

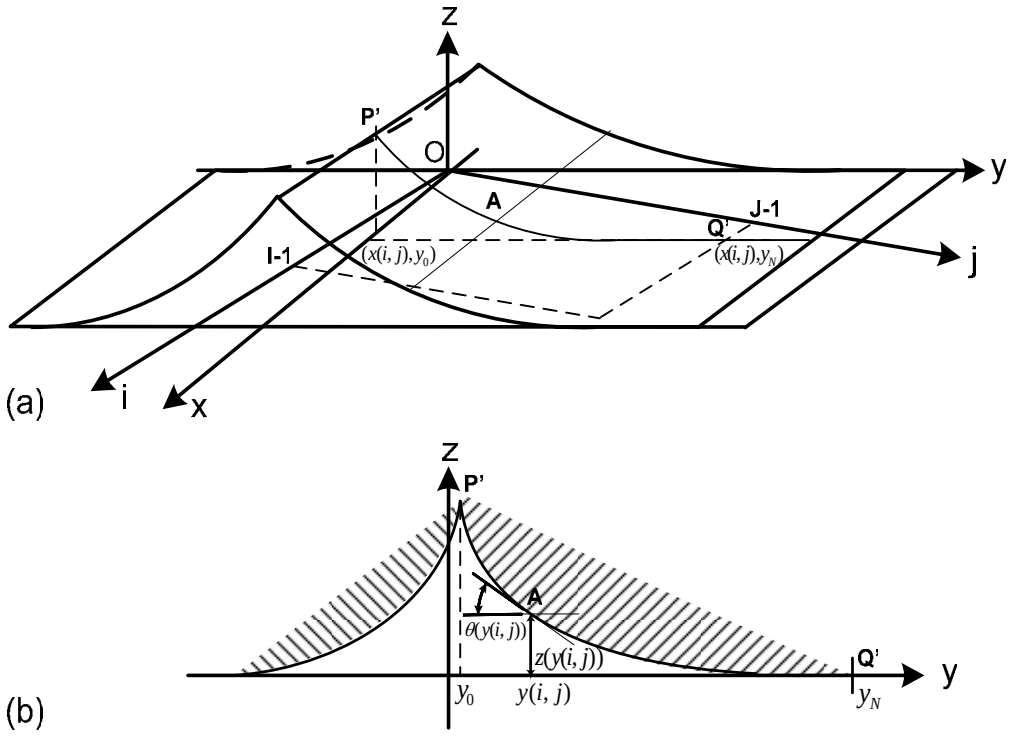


Figure 4.5: The cross section shape of the book surface in (a) *x*-*y*-*z* space and (b) *y*-*z* plane.

width respectively. y_0 and y_N denote the y value of the book spine and the maximum y value of the points on curve *P'AQ'* respectively.

Figure 4.5(b) shows the cross section shape in *y-z* plane. Our Assumption 1 reduces the 3D book surface in the *x-y-z* space to the 2D cross section shape in *y-z* plane. Mathematically, we have:

$$z(y_N) - z(y(i,j)) = \int_{y(i,j)}^{y_N} z'(y)dy \tag{4.5}$$

Since $z(y_N) = 0$, in our discrete *y-z* coordinate system, the depth map $z(y(i,j))$ is represented as:

$$\begin{aligned} & z(y(i,j)) \\ &= \sum_{y_k = y_N}^{y(i,j)+1} \int_{y_k - 1}^{y_k} - z'(y) \\ &= \sum_{y_k = y_N}^{y(i,j)+1} \tan\theta(y_k) \qquad (y_0 \le y(i,j) \le y_N, \theta \in [0, \frac{\pi}{2}]) \end{aligned} \tag{4.6}$$

where $y_N = y(l, J-1)$ for $x(l, J-1) = x(i,j)$ and $0 \le l < I$, y_0 is the y coordinate of the point separating the book pages in the cross section shape, $\theta(y_k)$ the slant angle at y_k. Note that y_N is not a constant and varies with $x(i,j)$. The image indices corresponding to y_k can be calculated by substituting $(x(i,j), y_k)$ to Equations (4.3) and (4.4) respectively.

4.2.2 The 3D Optical Model

In computer vision field, our problem can be classified to the Shape-from-Shading (SFS) area. However, the standard SFS requires: distance and fixed light source, uniform surface albedo, and so on, which are obvious not the case in our problem.

Thus for the real world scanning conditions shown in Figure 4.2, we formulate our SFS problem by considering the following four factors:

Factor 1: Proximal and moving light source: The linear directional light source of the scanner moves during the scanning process and is located very close to the book surface. This implies that the illuminant intensity varies with respect to the location on the book surface.

Factor 2: Lambertian reflection: The book surface is Lambertian, i.e. the reflected light intensity observed by the CCD sensor is proportional only to the cosine of angle of incidence.

Factor 3: Nonuniform albedo distribution: The albedo distribution over a printed book is not uniform.

Factor 4: Document skew ε: The book to be scanned is not necessarily aligned parallel to the linear directional light source of the scanner.

Let us first consider an ideal shape-from-shading problem, which satisfies

1) Distant and fixed light source: This implies that the illuminant intensity and the light source direction are constant all over the object surface.
2) Lambertian reflection.
3) Uniform albedo distribution: The albedo is constant all over the book surface.
4) No document skew: The book spine is parallel to the linear directional light source of the scanner, i.e. $\varepsilon = 0$.

The problem under these ideal conditions can be formulated as:

$$I_o(p) = I_s \cdot k \cdot \cos\varphi(p) \tag{4.7}$$

where p denotes a 2D point in the scanned image, $I_o(p)$ the reflected light intensity observed at p, I_s the illuminant intensity, k the albedo on the surface, and $\varphi(p)$ the angle between the light source direction and the surface normal at the 3D point on the book surface corresponding to p.

Proximal and moving light source (factor 1)

With a proximal and moving light source, the illuminant intensity is no longer constant over the object surface. The illuminant intensity on one point p is now a function of the location of p and that of the light source corresponding to p. We can formulate the problem as follows:

$$I_o(p) = I_s(s(p), l(p)) \cdot k \cdot \cos\varphi(p) \tag{4.8}$$

where $s(p)$ and $l(p)$ denote the 3D point on the book surface and light source location corresponding to p respectively.

Lambertian reflection and nonuniform albedo distribution (factor 2 and 3)

Next we can formulate our problem by incorporating Lambertian reflection and nonuniform albedo distribution characteristics into Equation (4.8):

$$I_o(p) = I_s(s(p), l(p)) \cdot k(s(p)) \cdot \cos\varphi(p) \tag{4.9}$$

where $k(s(p))$ denotes the albedo at $s(p)$.

Document skew ε (factor 4)

Finally, as shown in Figure 4.2(a) and 4.2(b), we take the document skew ε into account. By the coordinate system in Figure 4.2 and Equation (4.9), the relationship between the image intensity $P(i, j)$ (pixel value) and the reflected light intensity is represented as follows:

$$\begin{aligned} & P(i, j) \\ & = \alpha \cdot I_o(i, j) + \beta \\ & = \alpha \cdot I_s(z(y(i, j))) \cdot k(x(i, j), y(i, j)) \cdot \cos\varphi(y(i, j)) + \beta \end{aligned} \tag{4.10}$$

where

- $P(i, j)$: The image intensity at (i, j) in the observed image, i.e. the scanned document image.
- α, β: The gain and bias of the photoelectric transformation in the image scanner respectively.
- $x(i, j)$, $y(i, j)$: The *x* and *y* coordinates of the 3D point corresponding to $P(i, j)$, which are calculated from Equations (4.1) and (4.2) respectively.
- $z(y(i, j))$: The *z* coordinate of the 3D point on the book surface corresponding to $P(i, j)$ in the observed image. By Assumption 1, the *z* values are constant along *x*-axis. $z(y(i, j))$ is the distance between the scanning plane and the book surface, which is represented as Equation (4.6).
- $I_s(z(y(i, j)))$: The illuminant intensity distribution on the *y*-*z* plane when taking the 1D image at $(x(i, j), y(i, j))$, i.e. $I_s(z(y(i, j)))$ is the practical representation of $I_s(s(p), l(p))$. Due to the constant tilt angle ψ of the directional linear light source of the scanner, as shown in Figure 4.2(b) and (c), the 3D points with same z value have the same distance $(AD + BD)/\cos\psi$, i.e. $(z(y(i, j)) + d)/\cos\psi$,

from the light source. Note that the document skew ε has no effect on this distance, and the book surface depth $z(y(i, j))$ is the only effective factor. Based on the directional linear light source model, $I_s(z(y(i, j)))$ can be represented as follows:

$$I_s(z(y(i, j))) = \frac{I_D(\psi)}{((z(y(i, j)) + d)/\cos\psi)} = \frac{I_D(\psi) \cdot \cos\psi}{z(y(i, j)) + d} \tag{4.11}$$

where ψ is the tilt angle between the light source direction and the normal to scanning plane, $I_D(\psi)$ the directional distribution of the illuminant intensity.

- $k(x(i, j), y(i, j))$: The albedo of the 3D point corresponding to $P(i, j)$.
- $\cos\varphi(y(i, j))$: The cosine of the angle φ between the light source direction and the surface normal at the 3D point $(x(i, j), y(i, j), z(y(i, j)))$. By Assumption 2, the book surface is Lambertian. The reflected light intensity is proportional only to $\cos\varphi$. Since the directional linear light source of the scanner has a constant tilt angle ψ and Assumption 1, both the light source direction and the surface normal are constant along *x*-axis. We can represent φ as $\varphi(y(i, j))$. Note that $\varphi(y(i, j))$ is neither in the plane *PAQ* (Figure 4.2) nor the plane *P'AQ'* (Figure 4.5). We shall discuss the relation between $\varphi(y(i, j))$ and $\theta(y(i, j))$ (the slant angle in the cross section shape *P'AQ'*) in the next section.

In the experiments, parameters α, β, $I_D(\psi)$ are estimated a priori using calibration image attached with the scanner, parameters d, ψ are estimated by measurement of the geometry of the projected scanner light.

4.3 Reducing the 3D Shape Reconstruction Problem to a 2D Cross Section Shape

Reconstruction Problem

After we built the geometric model and optical model, we are facing the following two problems:

Problem 1: In the geometric model, the 2D cross section shape in the *y*-*z* plane of the 3D book surface is not unique, since y_N varies with $x(i, j)$ in Equation (4.6). That means the number of $z(y(i, j))$ and $\theta(y(i, j))$ pairs is different for different $x(i, j)$.

Problem 2: Both the slant angle $\theta(y(i, j))$ in the geometric model (Equation (4.6)) and the incident angle $\varphi(y(i, j))$ in the optical model (Equation (4.10)) are unknown, which leads to a greater number of unknown variables than the number of equations in our practical models.

The solution of the depth map $z(y(i, j))$ (2D cross section shape on *y*-*z* plane) cannot be found under these conditions. In this section, we describe how to reduce the 3D book surface reconstruction problem to the 2D cross section shape reconstruction problem by solving Problems 1 and 2.

4.3.1 The Processing Area of the Document Image (Solving Problem 1)

In order to reduce the 3D book surface to a unique 2D cross section shape, we define four lines in *x*-*y* plane: $x = x_0$, $x = x_M$, $y = y_0$, and $y = y_N$, where

$$x_0 = \max(0, x(0, J-1)) \tag{4.12}$$

$$x_M = \min(x(I-1,0), x(I-1, J-1)) \tag{4.13}$$

$$y_0 = \max(0, y(I-1,0)) \tag{4.14}$$

$$y_N = \min(y(0, J-1), y(I-1, J-1)) \tag{4.15}$$

I and J are the image length and width respectively, $x(0, J-1)$, $x(I-1,0)$, $x(I-1, J-1)$, $y(I-1,0)$, $y(0, J-1)$, $y(I-1, J-1)$ are calculated from Equations (4.1) and (4.2) respectively.

We define the processing area of the document to be the rectangle bounded by the four lines defined by Equations (4.12) – (4.15). Figure 4.6 shows the Processing Area

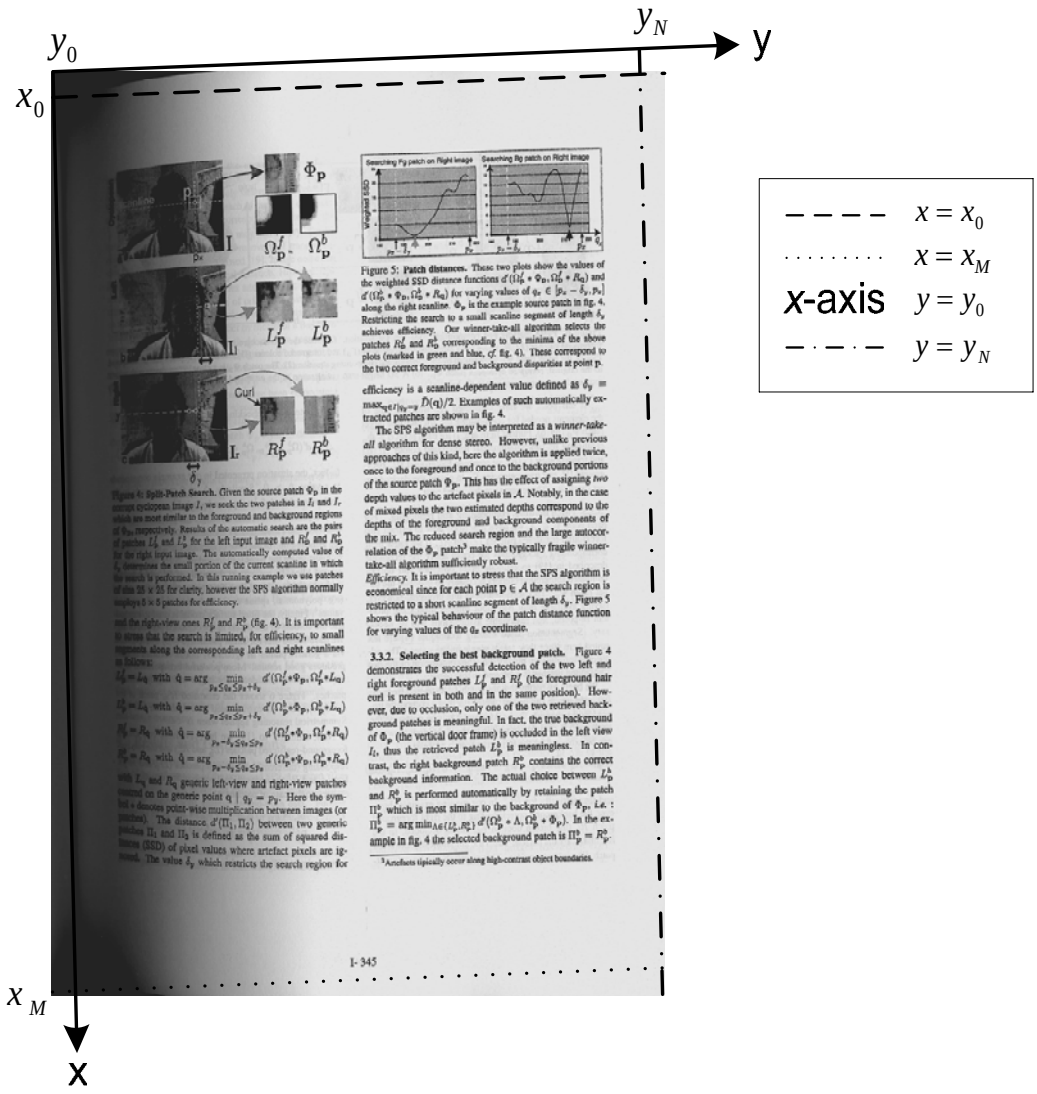


Figure 4.6: The processing area of the document image in Figure 4.1.

of Figure 4.1. The 3D book surface is now reduced to a unique 2D cross section shape between y_0 and y_N on *y*-*z* plane.

4.3.2 The Relation between $\theta(y(i, j))$ and $\varphi(y(i, j))$ (Solving Problem 2)

By knowing the document skew ε, we may derive the relation between $\theta(y(i, j))$ and $\varphi(y(i, j))$. Thus we can reduce the number of unknown variables by one. Figure 4.7(a) shows the book surface and its two coordinate systems. The plane ABC and ABC' are perpendicular to the light source (also the *i*-axis) and the book spine (also the *x*-axis) respectively, and intersect the book surface at curve PAQ (the cross section shape of the book surface in *j*-*z* plane) as shown in Figure 4.2 (b) and curve P'AQ' (the cross section shape of the book surface in *y*-*z* plane) as shown in Figure 4.5(a) respectively.

Figure 4.7(b) shows the geometric structure of $\theta(y(i, j))$ and $\varphi(y(i, j))$ at $y(i, j)$ ($y_0 \le y(i, j) \le y_N$, where y_0 and y_N are defined by Equations (4.14) and (4.15) respectively). $\theta(y(i, j))$ denotes the slant angle of the 2D cross section shape P'AQ' at $y(i, j)$, $\varphi(y(i, j))$ the angle between the light source direction L_D and the surface normal $N(y(i, j))$ at $y(i, j)$ (due to Assumption 1, the surface normal of the points along *x*-axis are constant, which can be expressed as $N(y(i, j))$), ψ the constant tilt angle of the scanner light source, and V the vertical to the scanning plane.

In Figure 4.7(c), we define a local coordinate system for point *A* on the book surface, whose origin is offset from the origin *O* in Figure 4.7(a) to the point *A* itself,

and the orientations of all the axes remain unchanged. We rename the axes as *x'*, *y'*, *z'*, *i'*, and *j'*, corresponding to the *x*, *y*, *z*, *i*, and *j* axes in Figure 4.7(a) respectively. Thus

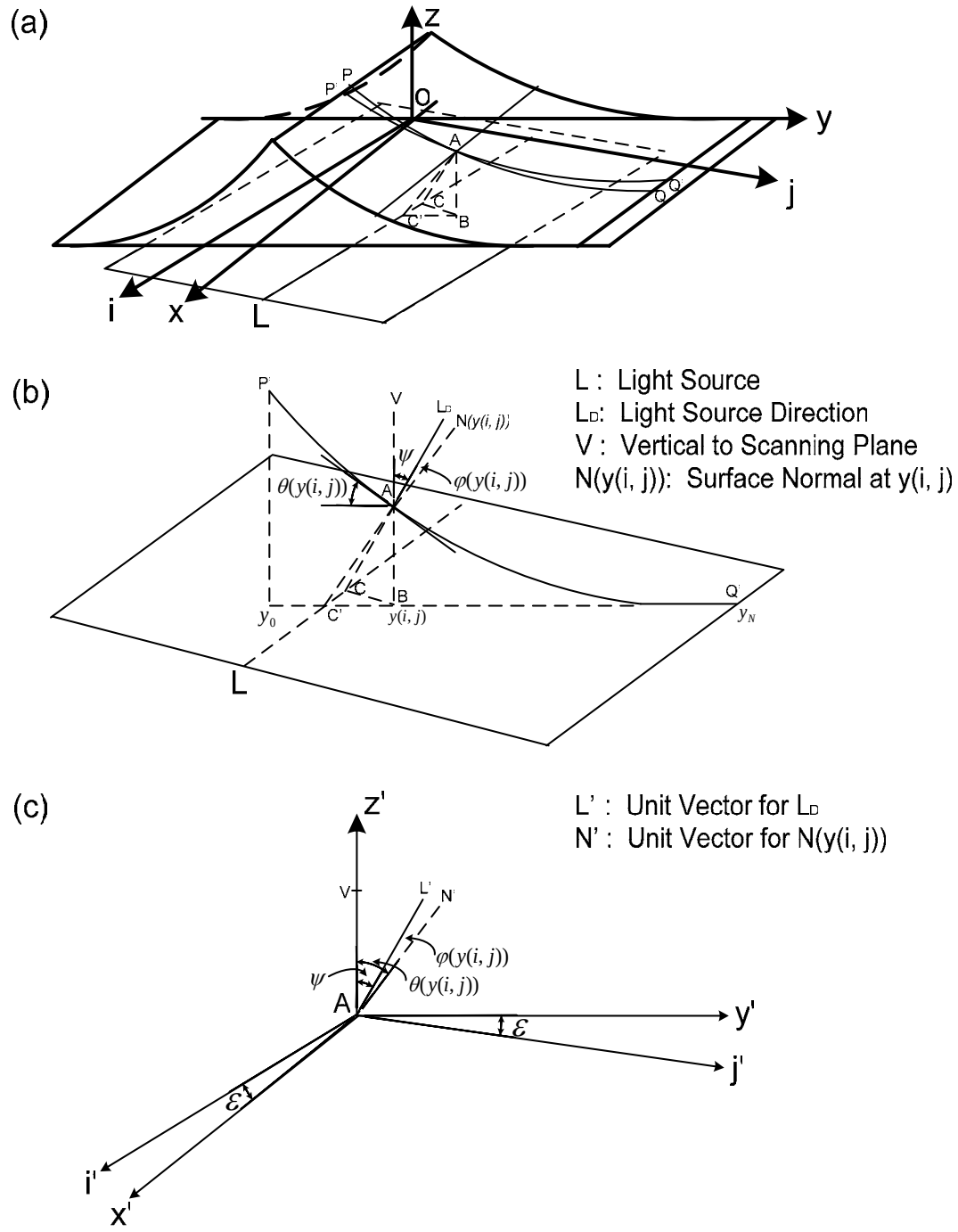


Figure 4.7: The schematic drawing of the relation between $\theta(y(i,j))$ and $\varphi(y(i,j))$.

the angle between *x'*-axis and *i'*-axis is equal to ε. We define *L'* and *N'* to be the unit vector for L_D and $N(y(i, j))$ in Figure 4.7(b) respectively, i.e. $\| L'\| = \| N'\| = 1$. In Figure 4.2(b), since $\theta(y(i, j))$, $N(y(i, j))$, and *V* are coplane, and $N(y(i, j))$ and *V* are perpendicular to the two edges of $\theta(y(i, j))$, the angle between $N(y(i, j))$ and *V* is equal to $\theta(y(i, j))$, i.e. the angle between unit vector $\| N'\|$ and $\| V'\|$ in Figure 4.7(c) is equal to $\theta(y(i, j))$.

Let $L'_{i'j'z'} = [i_{L'}, j_{L'}, z_{L'}]$ and $N'_{x'y'z'} = [x_{N'}, y_{N'}, z_{N'}]$ denote the unit vector representation for *L'* and *V'* in the *i'-j'-z'* and *x'-y'-z'* coordinate systems respectively. According to Figure 4.7(c), $N'_{x'y'z'} = [0, \sin\theta(y(i, j)), \cos\theta(y(i, j))]$ and $L'_{i'j'z'} = [0, \sin\psi, \cos\psi]$. Let $L'_{x'y'z'} = [x_{L'}, y_{L'}, z_{L'}]$ denote the vector representation for *L'* in the *x'-y'-z'* coordinate system. The relation between $L'_{i'j'z'} = [i_{L'}, j_{L'}, z_{L'}]$ and $L'_{x'y'z'} = [x_{L'}, y_{L'}, z_{L'}]$ are shown as follows:

$$x_{L'} = i_{L'} \cdot \cos\varepsilon + j_{L'} \cdot \sin\varepsilon \tag{4.16}$$

$$y_{L'} = -i_{L'} \cdot \sin\varepsilon + j_{L'} \cdot \cos\varepsilon \tag{4.17}$$

$$z_{L'} = z_{L'} \tag{4.18}$$

In matrix form, we have

$$\begin{bmatrix} x_{L'} \\ y_{L'} \\ z_{L'} \end{bmatrix} = \begin{bmatrix} \cos\varepsilon & \sin\varepsilon & 0 \\ -\sin\varepsilon & \cos\varepsilon & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} i_{L'} \\ j_{L'} \\ z_{L'} \end{bmatrix} \tag{4.19}$$

Substituting $L'_{x'y'z'} = [x_{L'}, y_{L'}, z_{L'}]$ and $L'_{i'j'z'} = [i_{L'}, j_{L'}, z_{L'}] = [0, \sin\psi, \cos\psi]$ to Equation (4.19), we represent $L'_{x'y'z'}$ as follows:

$$L'_{x'y'z'} = \begin{bmatrix} \cos\varepsilon & \sin\varepsilon & 0 \\ -\sin\varepsilon & \cos\varepsilon & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ \sin\psi \\ \cos\psi \end{bmatrix} = \begin{bmatrix} \sin\varepsilon \cdot \sin\psi \\ \cos\varepsilon \cdot \sin\psi \\ \cos\psi \end{bmatrix} \tag{4.20}$$

Since $\varphi(y(i, j))$ is the angle between *L'* and *N'*, we have:

$$\begin{aligned}&\cos\varphi(y(i, j))\\&=< L'_{x'y'z'}, N'_{x'y'z'} >\\&= \cos\varepsilon \cdot \sin\psi \cdot \sin\theta(y(i, j)) + \cos\psi \cdot \cos\theta(y(i, j))\end{aligned} \tag{4.21}$$

Note that both $\sin\theta(y(i, j))$ and $\cos\theta(y(i, j))$ appear in Equation (4.21). Although $\sin\theta(y(i, j))$ can be written as $\sqrt{1-\cos^2\theta(y(i, j))}$, we could not directly compute $\theta(y(i, j))$ by carrying inverse trigonometric functions. We thus need to transform Equation (4.21) to a monomial representation of $\theta(y(i, j))$.

By distilling $\cos\psi$ and let $a = \cos\varepsilon \cdot \tan\psi$, we have:

$$\begin{aligned}&\cos\varphi(y(i, j))\\&= \cos\psi \cdot [\cos\varepsilon \cdot \tan\psi \cdot \sin\theta(y(i, j)) + \cos\theta(y(i, j))]\\&= \cos\psi \cdot [a \cdot \sin\theta(y(i, j)) + \cos\theta(y(i, j))]\\&= \cos\psi \cdot \sqrt{a^2+1} \cdot \left[\frac{a}{\sqrt{a^2+1}} \cdot \sin\theta(y(i, j)) + \frac{1}{\sqrt{a^2+1}} \cdot \cos\theta(y(i, j))\right]\end{aligned} \tag{4.22}$$

Since $\left(\frac{a}{\sqrt{a^2+1}}\right)^2 + \left(\frac{1}{\sqrt{a^2+1}}\right)^2 = 1$, we introduce an angle γ, such that

$$\sin\gamma = \frac{a}{\sqrt{a^2+1}} \tag{4.23}$$

$$\cos\gamma = \frac{1}{\sqrt{a^2+1}} \tag{4.24}$$

$$\tan\gamma = \frac{\sin\gamma}{\cos\gamma} = a = \cos\varepsilon \cdot \tan\psi \tag{4.25}$$

By substituting Equations (4.23), (4.24) and $a = \cos\varepsilon \cdot \tan\psi$ to Equation (4.22), we have

$$\begin{aligned}&\cos\varphi(y(i, j))\\&= \cos\psi \cdot \sqrt{a^2+1} \cdot [\sin\gamma \cdot \sin\theta(y(i, j)) + \cos\gamma \cdot \cos\theta(y(i, j))]\\&= \sqrt{\cos^2\varepsilon \cdot \sin^2\psi + \cos^2\psi} \cdot [\sin\gamma \cdot \sin\theta(y(i, j)) + \cos\gamma \cdot \cos\theta(y(i, j))]\\&= \sqrt{\cos^2\varepsilon \cdot \sin^2\psi + \cos^2\psi} \cdot \cos(\theta(y(i, j)) - \gamma)\end{aligned} \tag{4.26}$$

From Equation (4.25), we derive that $\gamma = \arctan(\cos\varepsilon \cdot \tan\psi)$. Substituting γ to Equation (4.26), we have the relation between $\varphi(y(i, j))$ and $\theta(y(i, j))$ as follows

$$\begin{aligned}&\cos\varphi(y(i, j)) \\ &= \sqrt{\cos^2\varepsilon \cdot \sin^2\psi + \cos^2\psi} \cdot \cos(\theta(y(i, j)) - \arctan(\cos\varepsilon \cdot \tan\psi))\end{aligned} \tag{4.27}$$

Note that when the document skew $\varepsilon = 0$, i.e. *N'*, *L'*, and *V'* are coplane, Equation (4.27) is simplified to $\cos\varphi(y(i, j)) = \cos(\theta(y(i, j)) - \psi)$.

4.4 Reconstruction of Book Surface Shape and Albedo Distribution

Remember that our problem is how to restore the document image by removing the shade (de-shading) and correcting the warping (de-warping) in the shade. For de-shading we need the albedo distribution $k(x(i, j), y(i, j))$, and for de-warping we need the shape of the book surface shape $z(y(i, j))$. In this section, we will discuss how to solve $z(y(i, j))$ and $k(x(i, j), y(i, j))$ by adopting the two practical models.

4.4.1 Reconstruction of Book Surface Shape

Since most of book surfaces have a uniformly colored background (typically an unprinted white background), we can assume that pixel values $P_w(y_n)$ $(y_0 \le y_n \le y_N)$ corresponding to the background can be obtained as the maximum for each *y* value in the scanned document image, i.e. we scan the image pixels along the direction of *x*-axis (in the *x*-*y*-*z* 3D coordinate system) instead of *i*-axis (in the *i*-*j* 2D image index system), and set $P_w(y_n)$ as the maximum intensity value of each

scan:

$$P_w(y_n) = \max_{x_m} P(i(x_m, y_n), j(x_m, y_n)) \quad (x_0 \le x_m \le x_M, y_0 \le y_n \le y_N) \tag{4.28}$$

where x_0, x_M, y_0 and y_N are calculated from Equations (4.12), (4.13), (4.14) and (4.15) respectively, $(i(x_m, y_n), j(x_m, y_n))$ is the corresponding image index for (x_m, y_n), which are calculated from Equations (4.3) and (4.4).

We then extract the global maximum pixel value $P_w^{\max}$ over all y_n columns by:

$$P_w^{\max} = \max_{y_n} P_w(y_n) \tag{4.29}$$

By Equation (4.10), the constant albedo k_w of the white background is calculated as follows:

$$k_w = \frac{P_w^{\max} - \beta}{\alpha \cdot I_s(0) \cdot \cos\psi} \tag{4.30}$$

The optical model of the white background with the constant albedo k_w is represented as follows:

$$P_w(y_n) = \alpha \cdot k_w \cdot I_s(z(y_n)) \cdot \cos\varphi(y_n) + \beta \tag{4.31}$$

where $P_w(y_n)$ is from Equation (4.28), k_w from Equation (4.30).

By substituting Equation (4.11) into Equations (4.30) and (4.31), we derive the following representation of $z(y_n)$ from the optical model:

$$z(y_n) = \left[\frac{(P_w^{\max} - \beta) \cdot \cos\varphi(y_n)}{(P_w(y_n) - \beta) \cdot \cos\psi} - 1 \right] \cdot d \tag{4.32}$$

Replacing $\cos\varphi(y_n)$ by Equation (4.27), we have:

$$z(y_n) = \left[\rho \cdot \frac{\cos(\theta(y_n) - \arctan(\cos\varepsilon \cdot \tan\psi))}{P_w(y_n) - \beta} - 1 \right] \cdot d \tag{4.33}$$

where ρ is a constant represented as:

$$\rho = \frac{(P_w^{\max} - \beta) \cdot \sqrt{\cos^2\varepsilon \cdot \sin^2\psi + \cos^2\psi}}{\cos\psi} \tag{4.34}$$

By substituting y_n to Equation (4.6) (geometric model), we derive another representation of $z(y_n)$:

$$z(y_n) = \sum_{y_k = y_N}^{y_n + 1} \tan\theta(y_k) \qquad (y_0 \le y_n \le y_N, \theta \in [0, \frac{\pi}{2}]) \tag{4.35}$$

where y_0 and y_N come from Equations (4.14) and (4.15) respectively.

Note that now we have two representations of $z(y_n)$, i.e. Equation (4.33) and Equation (4.35) from the optical model and the geometric model respectively, and there are only two unknowns - $z(y_n)$ and $\theta(y_n)$. Furthermore, the initial values of $z(y_n)$ and $\theta(y_n)$ are known: $z(y_N) = 0$ and $\theta(y_N) = 0$. Thus in theory, the solution of the depth map $z(y_n)$ can be found.

We propose the following algorithm to compute the numerical solutions of $z(y_n)$ and $\theta(y_n)$:

Step 1. Rewrite Equation (4.33) in the following form:

$$\cos(\theta(y_n) - \arctan(\cos\varepsilon \cdot \tan\psi)) = \left[\frac{P_w(y_n) - \beta}{\rho} \right] \cdot \left[\frac{z(y_n) + d}{d} \right] \tag{4.36}$$

Thus if $\theta(y_n) < \arctan(\cos\varepsilon \cdot \tan\psi)$

$$\theta(y_n) = \arctan(\cos\varepsilon \cdot \tan\psi) - \arccos\left[\frac{P_w(y_n) - \beta}{\rho} \cdot \frac{z(y_n) + d}{d} \right] \tag{4.37}$$

else

$$\theta(y_n) = \arctan(\cos\varepsilon \cdot \tan\psi) + \arccos\left[\frac{P_w(y_n) - \beta}{\rho} \cdot \frac{z(y_n) + d}{d}\right] \qquad (4.38)$$

Step 2. Set $z(y_N) = 0$, $\theta(y_N) = 0$, and Exceed=0 (since $\theta(y_n)$ is monotonously increasing starting with value 0 from y_N to y_0, this control flag "Exceed" indicates whether the θ value has exceeded $\arctan(\cos\varepsilon \cdot \tan\psi)$).

Step 3. For $(y_n = y_N - 1, y_n \geq y_0, y_n --)$ {

3.1 Compute $z(y_n)$ by Equation (4.35).

3.2 If $\theta(y_n + 1) \geq \arctan(\cos\varepsilon \cdot \tan\psi)$, set Exceed =1.

3.3 If Exceed =0, compute $\theta(y_n)$ by substituting $z(y_n)$ to Equation (4.37), else compute $\theta(y_n)$ by substituting $z(y_n)$ to Equation (4.38).

}

End of the Algorithm

Figure 4.8 shows book surface shape discovered by our algorithm.

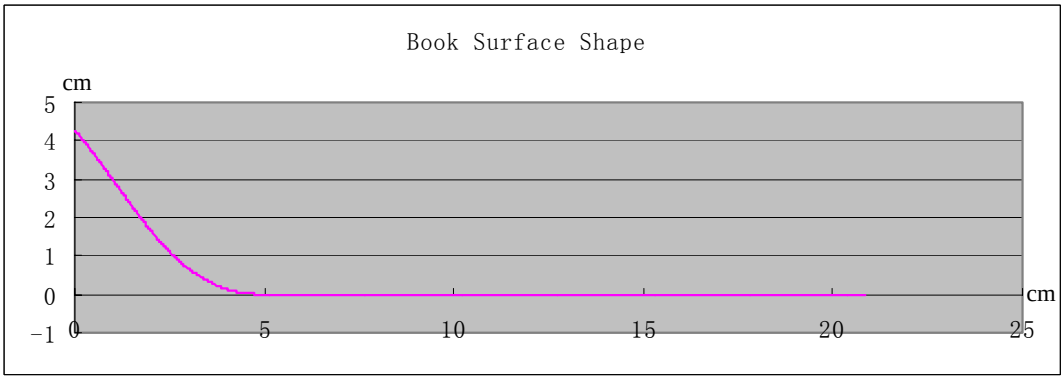


Figure 4.8: Cross section shape on *y-z* plane of the book surface in Figure 4.1.

4.4.2 Reconstruction of Albedo Distribution

By Equation (4.10), the albedo distribution is represented as:

$$k(x(i,j),y(i,j)) = \frac{P(i,j)-\beta}{\alpha \cdot I_s(z(y(i,j))) \cdot \cos\varphi(y(i,j))} \tag{4.39}$$

Substituting Equations (4.27) into (4.39), we have:

$$k(x(i,j),y(i,j)) = \frac{P(i,j)-\beta}{\alpha \cdot I_s(z(y(i,j))) \cdot \sqrt{\cos^2\varepsilon \cdot \sin^2\psi + \cos^2\psi} \cdot \cos(\theta(y(i,j)) - \arctan(\cos\varepsilon \cdot \tan\psi))} \tag{4.40}$$

where $z(y(i,j))$ and $\theta(y(i,j))$ are looked up from the depth map $z(y_n)$ and $\theta(y_n)$ respectively, $I_s(z(y(i,j)))$ is calculated from Equation (4.11), $P(i,j)$ is the pixel value in the observed image (scanned image), the document skew ε is detected in Section 4.2, α, β, and ψ are known constants.

4.5 Restoration of Document Image

We propose two models, namely de-shading and de-warping models to restore the document image by using the book surface shape and albedo distribution obtained in last section.

4.5.1 De-shading Model

If the physical document page is flat while scanning, the z values are all zero, i.e. $z(y(i,j)) = 0$, and the page surface normal is the vertical, i.e. $\varphi(y(i,j)) = \psi$. By

Equation (4.10), the optimal image intensity $P^*(i, j)$ for point $(x(i, j), y(i, j),0)$ would be

$$P^*(i, j) = \alpha \cdot k(x(i, j), y(i, j)) \cdot I_s(0) \cdot \cos\psi + \beta \tag{4.41}$$

where $k(x(i, j), y(i, j))$ is calculated by Equation (4.40), $I_s(0)$ is calculated by Equation (4.11), and α, β, ψ are known constants.

Therefore, our restoration system recalculates the image intensity for each pixel

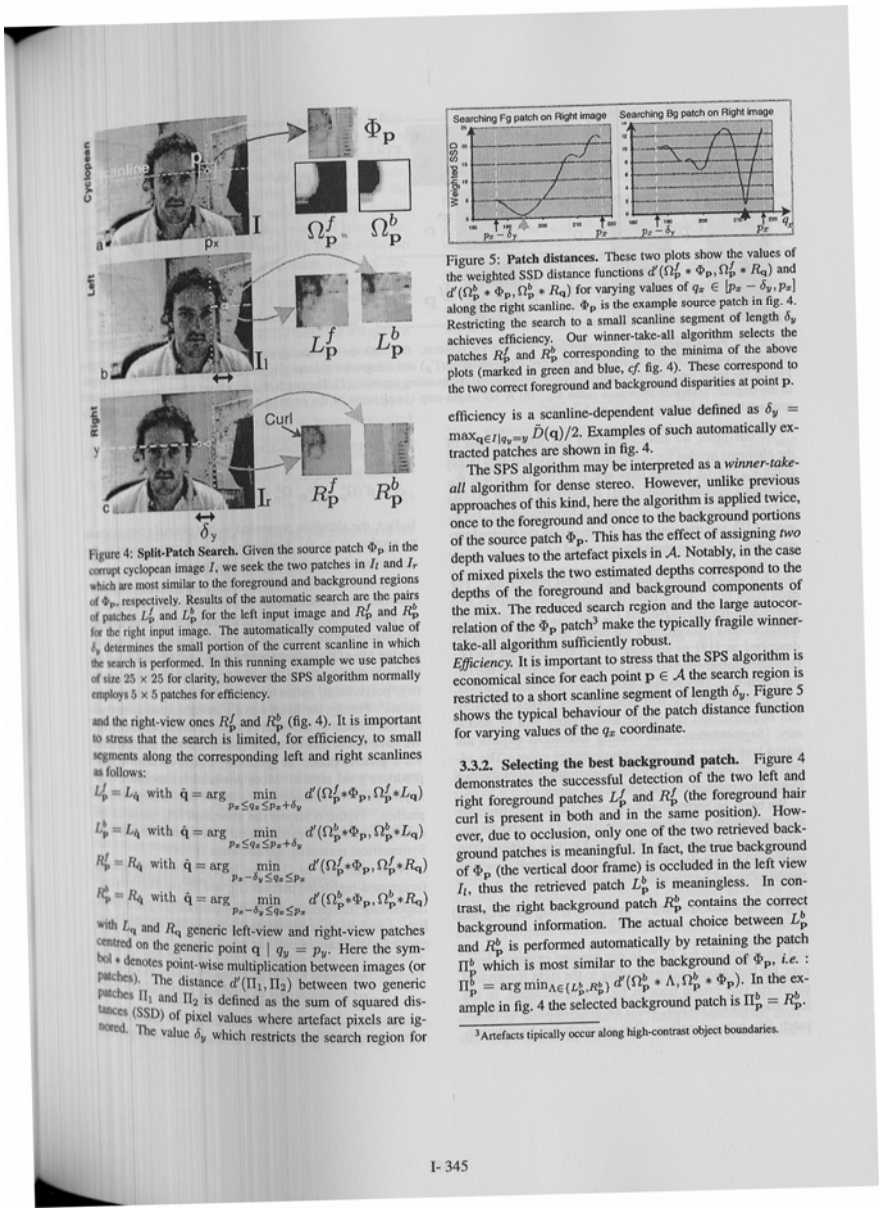


Figure 4: **Split-Patch Search.** Given the source patch $\Phi_{\mathbf{p}}$ in the corrupt cyclopean image I, we seek the two patches in I_l and I_r which are most similar to the foreground and background regions of $\Phi_{\mathbf{p}}$, respectively. Results of the automatic search are the pairs of patches $L^f_{\mathbf{p}}$ and $L^b_{\mathbf{p}}$ for the left input image and $R^f_{\mathbf{p}}$ and $R^b_{\mathbf{p}}$ for the right input image. The automatically computed value of δ_y determines the small portion of the current scanline in which the search is performed. In this running example we use patches of size 25×25 for clarity, however the SPS algorithm normally employs 5×5 patches for efficiency.

and the right-view ones $R^f_{\mathbf{p}}$ and $R^b_{\mathbf{p}}$ (fig. 4). It is important to stress that the search is limited, for efficiency, to small segments along the corresponding left and right scanlines as follows:

$$L^f_{\mathbf{p}} = L_{\hat{\mathbf{q}}} \text{ with } \hat{\mathbf{q}} = \arg \min_{p_x \le q_x \le p_x + \delta_y} d'(\Omega^f_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^f_{\mathbf{p}} * L_{\mathbf{q}})$$

$$L^b_{\mathbf{p}} = L_{\hat{\mathbf{q}}} \text{ with } \hat{\mathbf{q}} = \arg \min_{p_x \le q_x \le p_x + \delta_y} d'(\Omega^b_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^b_{\mathbf{p}} * L_{\mathbf{q}})$$

$$R^f_{\mathbf{p}} = R_{\hat{\mathbf{q}}} \text{ with } \hat{\mathbf{q}} = \arg \min_{p_x - \delta_y \le q_x \le p_x} d'(\Omega^f_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^f_{\mathbf{p}} * R_{\mathbf{q}})$$

$$R^b_{\mathbf{p}} = R_{\hat{\mathbf{q}}} \text{ with } \hat{\mathbf{q}} = \arg \min_{p_x - \delta_y \le q_x \le p_x} d'(\Omega^b_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^b_{\mathbf{p}} * R_{\mathbf{q}})$$

with $L_{\mathbf{q}}$ and $R_{\mathbf{q}}$ generic left-view and right-view patches centred on the generic point $\mathbf{q} \mid q_y = p_y$. Here the symbol $*$ denotes point-wise multiplication between images (or patches). The distance $d'(\Pi_1, \Pi_2)$ between two generic patches Π_1 and Π_2 is defined as the sum of squared distances (SSD) of pixel values where artefact pixels are ignored. The value δ_y which restricts the search region for

Figure 5: **Patch distances.** These two plots show the values of the weighted SSD distance functions $d'(\Omega^f_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^f_{\mathbf{p}} * R_{\mathbf{q}})$ and $d'(\Omega^b_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^b_{\mathbf{p}} * R_{\mathbf{q}})$ for varying values of $q_x \in [p_x - \delta_y, p_x]$ along the right scanline. $\Phi_{\mathbf{p}}$ is the example source patch in fig. 4. Restricting the search to a small scanline segment of length δ_y achieves efficiency. Our winner-take-all algorithm selects the patches $R^f_{\mathbf{p}}$ and $R^b_{\mathbf{p}}$ corresponding to the minima of the above plots (marked in green and blue, *cf.* fig. 4). These correspond to the two correct foreground and background disparities at point $\mathbf{p}$.

efficiency is a scanline-dependent value defined as $\delta_y = \max_{\mathbf{q} \in I \mid q_y = y} \tilde{D}(\mathbf{q})/2$. Examples of such automatically extracted patches are shown in fig. 4.

The SPS algorithm may be interpreted as a *winner-take-all* algorithm for dense stereo. However, unlike previous approaches of this kind, here the algorithm is applied twice, once to the foreground and once to the background portions of the source patch $\Phi_{\mathbf{p}}$. This has the effect of assigning *two* depth values to the artefact pixels in $\mathcal{A}$. Notably, in the case of mixed pixels the two estimated depths correspond to the depths of the foreground and background components of the mix. The reduced search region and the large autocorrelation of the $\Phi_{\mathbf{p}}$ patch[3] make the typically fragile winner-take-all algorithm sufficiently robust.

Efficiency. It is important to stress that the SPS algorithm is economical since for each point $\mathbf{p} \in \mathcal{A}$ the search region is restricted to a short scanline segment of length δ_y. Figure 5 shows the typical behaviour of the patch distance function for varying values of the q_x coordinate.

3.3.2. Selecting the best background patch. Figure 4 demonstrates the successful detection of the two left and right foreground patches $L^f_{\mathbf{p}}$ and $R^f_{\mathbf{p}}$ (the foreground hair curl is present in both and in the same position). However, due to occlusion, only one of the two retrieved background patches is meaningful. In fact, the true background of $\Phi_{\mathbf{p}}$ (the vertical door frame) is occluded in the left view I_l, thus the retrieved patch $L^b_{\mathbf{p}}$ is meaningless. In contrast, the right background patch $R^b_{\mathbf{p}}$ contains the correct background information. The actual choice between $L^b_{\mathbf{p}}$ and $R^b_{\mathbf{p}}$ is performed automatically by retaining the patch $\Pi^b_{\mathbf{p}}$ which is most similar to the background of $\Phi_{\mathbf{p}}$, *i.e.* : $\Pi^b_{\mathbf{p}} = \arg \min_{\Lambda \in \{L^b_{\mathbf{p}}, R^b_{\mathbf{p}}\}} d'(\Omega^b_{\mathbf{p}} * \Lambda, \Omega^b_{\mathbf{p}} * \Phi_{\mathbf{p}})$. In the example in fig. 4 the selected background patch is $\Pi^b_{\mathbf{p}} = R^b_{\mathbf{p}}$.

[3] Artefacts tipically occur along high-contrast object boundaries.

I- 345

Figure 4.9: Image generated by de-shading model for the Processing Area defined in Figure 4.6.

by Equation (4.41). The de-shading result is shown in Figure 4.9.

4.5.2 De-warping Model

Taking the image generated by de-shading model as input, we next build our de-warping model. The removal of distortion is done based on the cross section shape $z(y_n)$ of the book surface. Note that distortion occurs along y-axis only due to orthogonal projection, while perspective distortion occurs along x-axis due to perspective projection. Thus the de-warping model consists of three parts: 1) Restoration along *x*-axis; 2) Restoration along *y*-axis; 3) Correction of document skew ε.

4.5.2.1 Restoration along *x*-axis

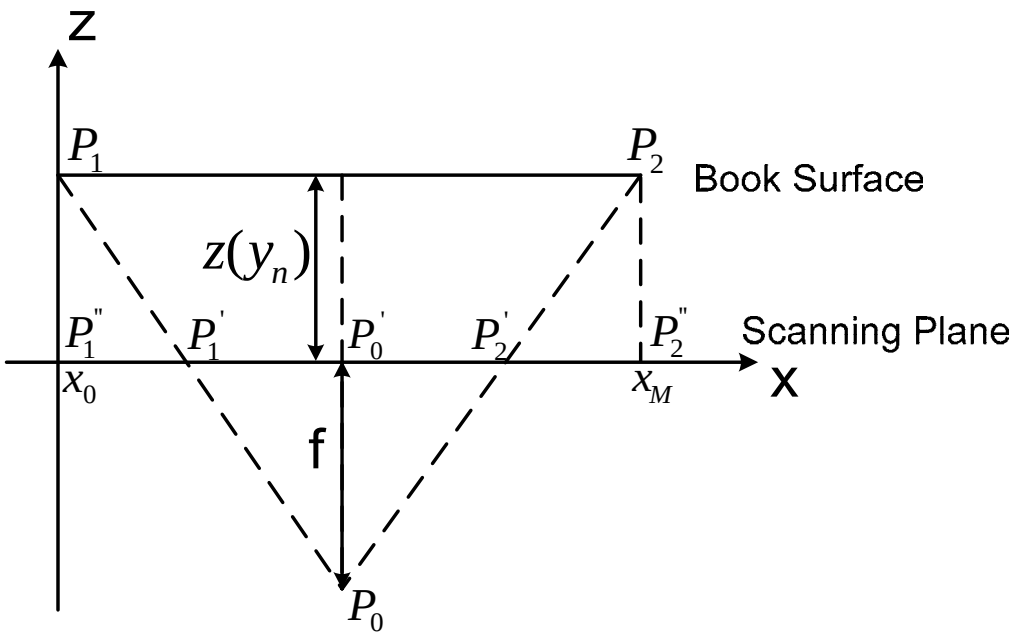


Figure 4.10: Perspective projection on a slice of the *x-z* plane at y_n.

Figure 4.10 shows a slice of x-z plane at y_n ($y_0 \le y_n \le y_N$), where P_1P_2 denotes a line parallel to x-axis on the book surface, $P_1'P_2'$ the line on the scanning plane corresponding to P_1P_2 viewed by the CCD sensor (i.e. $P_1'P_2'$ is the observed P_1P_2 in the scanned image), $P_1P_1'' = z(y_n)$, $P_1''P_2'' = 2 \times P_1''P_0' = x_M - x_0$ (x_0 and x_M are constants from Equations (4.12) and (4.13)), the focal length $f = P_0 P_0'$.

We restore the *x*-axis distortion by regenerating the image intensity $P^{**}(i, j)$ according to Equation (4.42) for each pixel on y_n (stretching $P_1'P_2'$ to the length of P_1P_2, from perspective equations, this stretching proves to be uniform for each y_n and proportional to the $z(y_n)$ value).

$$P^{**}(i, j) = \begin{cases} P^*(i(x_r, y(i, j)), j(x_r, y(i, j))) & \text{if } x_r \text{ is an integer} \\ P^*(i(floor(x_r), y(i, j)), j(floor(x_r), y(i, j))) \cdot (ceil(x_r) - x_r) & \\ + P^*(i(ceil(x_r), y(i, j)), j(ceil(x_r), y(i, j))) \cdot (x_r - floor(x_r)) & \text{otherwise} \end{cases} \tag{4.42}$$

where

- $P^*(i(x_r, y(i, j)), j(x_r, y(i, j)))$: Pixel value regenerated from Equation (4.41), which is corresponding to the 3D point $(x_r, y(i, j), z(y(i, j)))$.
- x_r: The corresponding *x* value on P_1P_2 for $x(i, j)$ on $P_1'P_2'$. x_r is calculated as follows:

$$
\begin{aligned}
x_r &= x(i,j)\cdot\frac{P_1'P_2'}{P_1P_2}+P_1'P_1'' \\
&= x(i,j)\cdot\frac{f}{z(y(i,j))+f}+\frac{z(y(i,j))\cdot\frac{1}{2}(x_M-x_0)}{z(y(i,j))+f} \\
&= \frac{2\cdot x(i,j)\cdot f+z(y(i,j))\cdot(x_M-x_0)}{2\cdot(z(y(i,j))+f)}
\end{aligned}
\tag{4.43}
$$

where $x(i,j)$ and $y(i,j)$ are calculated according to Equations (4.1) and (4.2), f a known constant.

- $floor(x_r)$: Returns the largest integer that is less or equal to x_r.
- $ceil(x_r)$: Returns the smallest integer that is greater or equal to x_r.

4.5.2.2 Restoration along *y*-axis

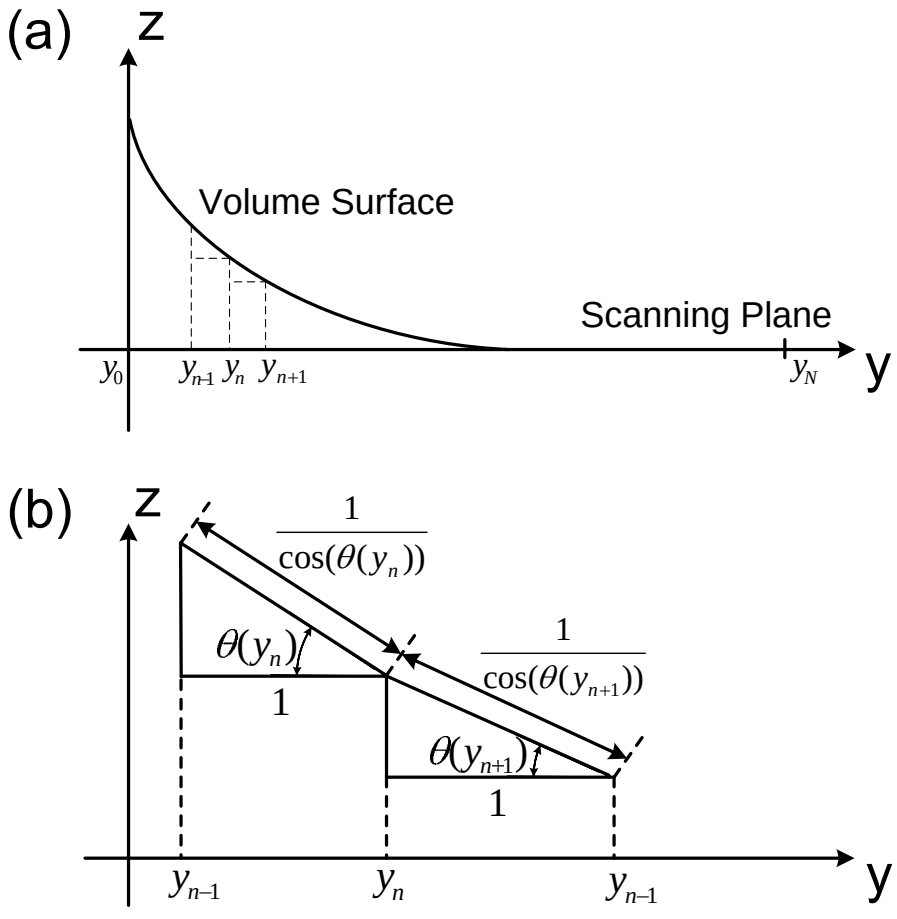


Figure 4.11: Orthogonal projection on a slice of the *y-z* plane.

Distortion along *y*-axis is only due to orthogonal projection as shown in Figure 4.11(a). Figure 4.11(b) shows that the true width between two adjacent *y* value y_{n-1} and y_n is equal to $\frac{1}{\cos\theta(y_n)}$. Thus the correction along *y*-axis is done by the following two steps: First we add the estimated true width w calculated from Equation (4.44), and then we stretch the observed image to its true width along *y*-axis

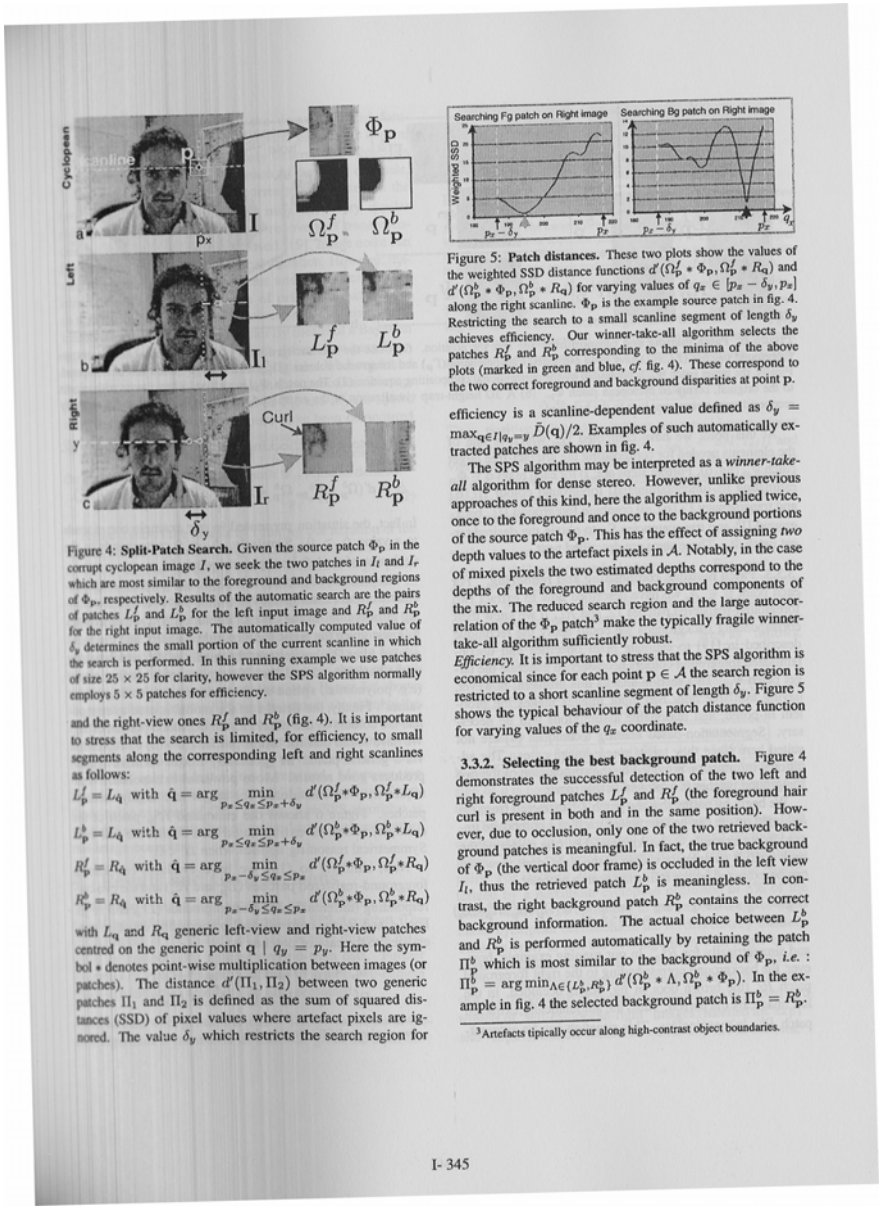


Figure 4: **Split-Patch Search.** Given the source patch $\Phi_{\mathbf{p}}$ in the corrupt cyclopean image I, we seek the two patches in I_l and I_r which are most similar to the foreground and background regions of $\Phi_{\mathbf{p}}$, respectively. Results of the automatic search are the pairs of patches $L^f_{\mathbf{p}}$ and $L^b_{\mathbf{p}}$ for the left input image and $R^f_{\mathbf{p}}$ and $R^b_{\mathbf{p}}$ for the right input image. The automatically computed value of δ_y determines the small portion of the current scanline in which the search is performed. In this running example we use patches of size 25 × 25 for clarity, however the SPS algorithm normally employs 5 × 5 patches for efficiency.

and the right-view ones $R^f_{\mathbf{p}}$ and $R^b_{\mathbf{p}}$ (fig. 4). It is important to stress that the search is limited, for efficiency, to small segments along the corresponding left and right scanlines as follows:

$$L^f_{\mathbf{p}} = L_{\hat{\mathbf{q}}} \text{ with } \hat{\mathbf{q}} = \arg\min_{p_x \le q_x \le p_x+\delta_y} d'(\Omega^f_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^f_{\mathbf{p}} * L_{\mathbf{q}})$$

$$L^b_{\mathbf{p}} = L_{\hat{\mathbf{q}}} \text{ with } \hat{\mathbf{q}} = \arg\min_{p_x \le q_x \le p_x+\delta_y} d'(\Omega^b_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^b_{\mathbf{p}} * L_{\mathbf{q}})$$

$$R^f_{\mathbf{p}} = R_{\hat{\mathbf{q}}} \text{ with } \hat{\mathbf{q}} = \arg\min_{p_x-\delta_y \le q_x \le p_x} d'(\Omega^f_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^f_{\mathbf{p}} * R_{\mathbf{q}})$$

$$R^b_{\mathbf{p}} = R_{\hat{\mathbf{q}}} \text{ with } \hat{\mathbf{q}} = \arg\min_{p_x-\delta_y \le q_x \le p_x} d'(\Omega^b_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^b_{\mathbf{p}} * R_{\mathbf{q}})$$

with $L_{\mathbf{q}}$ and $R_{\mathbf{q}}$ generic left-view and right-view patches centred on the generic point $\mathbf{q} \mid q_y = p_y$. Here the symbol * denotes point-wise multiplication between images (or patches). The distance $d'(\Pi_1, \Pi_2)$ between two generic patches Π_1 and Π_2 is defined as the sum of squared distances (SSD) of pixel values where artefact pixels are ignored. The value δ_y which restricts the search region for

Figure 5: **Patch distances.** These two plots show the values of the weighted SSD distance functions $d'(\Omega^f_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^f_{\mathbf{p}} * R_{\mathbf{q}})$ and $d'(\Omega^b_{\mathbf{p}} * \Phi_{\mathbf{p}}, \Omega^b_{\mathbf{p}} * R_{\mathbf{q}})$ for varying values of $q_x \in [p_x - \delta_y, p_x]$ along the right scanline. $\Phi_{\mathbf{p}}$ is the example source patch in fig. 4. Restricting the search to a small scanline segment of length δ_y achieves efficiency. Our winner-take-all algorithm selects the patches $R^f_{\mathbf{p}}$ and $R^b_{\mathbf{p}}$ corresponding to the minima of the above plots (marked in green and blue, *cf.* fig. 4). These correspond to the two correct foreground and background disparities at point **p**.

efficiency is a scanline-dependent value defined as $\delta_y = \max_{\mathbf{q} \in I | q_y = y} \bar{D}(\mathbf{q})/2$. Examples of such automatically extracted patches are shown in fig. 4.

The SPS algorithm may be interpreted as a *winner-take-all* algorithm for dense stereo. However, unlike previous approaches of this kind, here the algorithm is applied twice, once to the foreground and once to the background portions of the source patch $\Phi_{\mathbf{p}}$. This has the effect of assigning *two* depth values to the artefact pixels in $\mathcal{A}$. Notably, in the case of mixed pixels the two estimated depths correspond to the depths of the foreground and background components of the mix. The reduced search region and the large autocorrelation of the $\Phi_{\mathbf{p}}$ patch[3] make the typically fragile winner-take-all algorithm sufficiently robust.

Efficiency. It is important to stress that the SPS algorithm is economical since for each point $\mathbf{p} \in \mathcal{A}$ the search region is restricted to a short scanline segment of length δ_y. Figure 5 shows the typical behaviour of the patch distance function for varying values of the q_x coordinate.

3.3.2. Selecting the best background patch. Figure 4 demonstrates the successful detection of the two left and right foreground patches $L^f_{\mathbf{p}}$ and $R^f_{\mathbf{p}}$ (the foreground hair curl is present in both and in the same position). However, due to occlusion, only one of the two retrieved background patches is meaningful. In fact, the true background of $\Phi_{\mathbf{p}}$ (the vertical door frame) is occluded in the left view I_l, thus the retrieved patch $L^b_{\mathbf{p}}$ is meaningless. In contrast, the right background patch $R^b_{\mathbf{p}}$ contains the correct background information. The actual choice between $L^b_{\mathbf{p}}$ and $R^b_{\mathbf{p}}$ is performed automatically by retaining the patch $\Pi^b_{\mathbf{p}}$ which is most similar to the background of $\Phi_{\mathbf{p}}$, *i.e.* : $\Pi^b_{\mathbf{p}} = \arg\min_{\Lambda \in \{L^b_{\mathbf{p}}, R^b_{\mathbf{p}}\}} d'(\Omega^b_{\mathbf{p}} * \Lambda, \Omega^b_{\mathbf{p}} * \Phi_{\mathbf{p}})$. In the example in fig. 4 the selected background patch is $\Pi^b_{\mathbf{p}} = R^b_{\mathbf{p}}$.

[3] Artefacts tipically occur along high-contrast object boundaries.

Figure 4.12: Image generated by de-warping model for the Processing Area defined in Figure 4.6.

using similar method discussed in the last section.

$$w = \sum_{y_n = y_0}^{y_N} \left(\frac{1}{\cos(\theta(y_n))} - 1 \right) \tag{4.44}$$

The de-warping result is presented in Figure 4.12.

4.5.2.3 Correction of document skew ε

As shown in Figure 4.12, the document has a skew angle ε with the image boundary. We simply rotate the image at the center by ε degree, and the final restored image is shown in Figure 4.13. The experiment shows that the geometric and photometric distortions are mostly removed. This result demonstrates that the shape is accurately computed enough for the image restoration task.

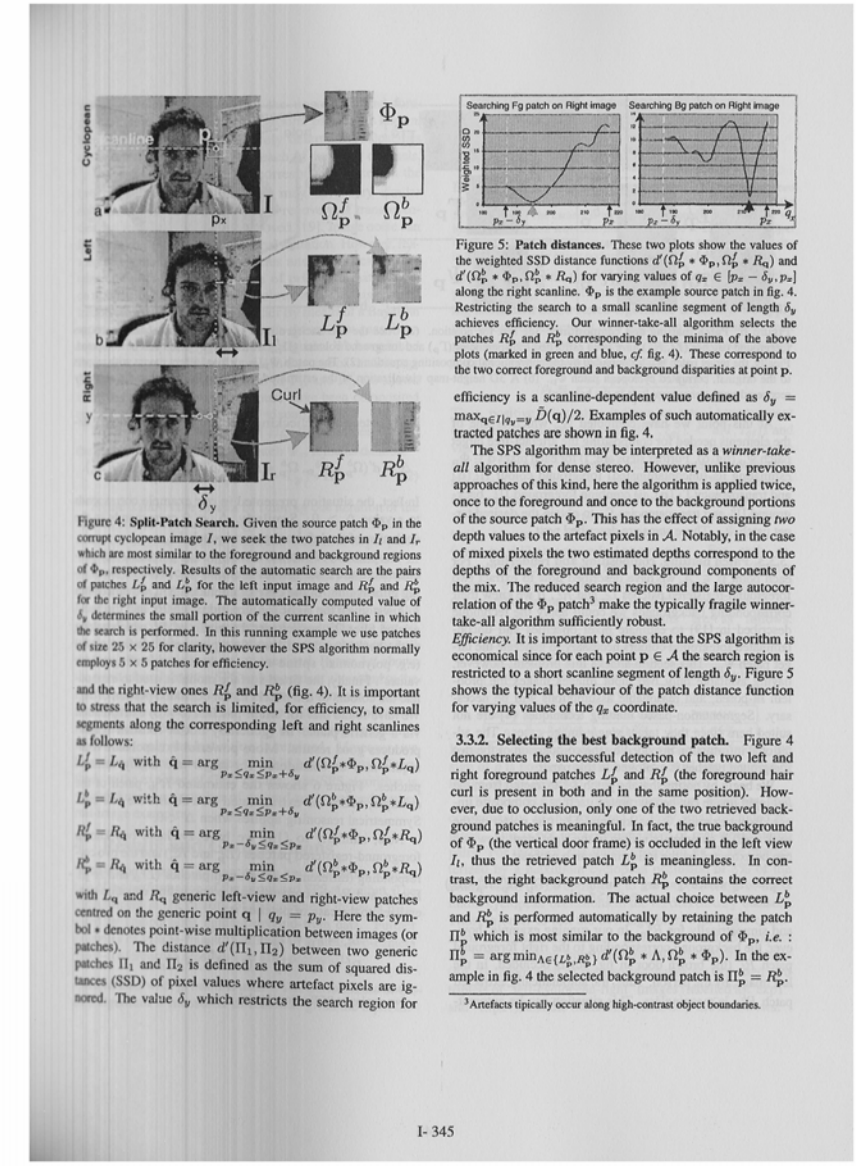


Figure 4: **Split-Patch Search.** Given the source patch Φ_p in the corrupt cyclopean image I, we seek the two patches in I_l and I_r which are most similar to the foreground and background regions of Φ_p, respectively. Results of the automatic search are the pairs of patches L^f_p and L^b_p for the left input image and R^f_p and R^b_p for the right input image. The automatically computed value of δ_y determines the small portion of the current scanline in which the search is performed. In this running example we use patches of size 25×25 for clarity, however the SPS algorithm normally employs 5×5 patches for efficiency.

and the right-view ones R^f_p and R^b_p (fig. 4). It is important to stress that the search is limited, for efficiency, to small segments along the corresponding left and right scanlines as follows:

$$L^f_p = L_{\hat{q}} \text{ with } \hat{q} = \arg \min_{p_x \le q_x \le p_x + \delta_y} d'(\Omega^f_p * \Phi_p, \Omega^f_p * L_q)$$

$$L^b_p = L_{\hat{q}} \text{ with } \hat{q} = \arg \min_{p_x \le q_x \le p_x + \delta_y} d'(\Omega^b_p * \Phi_p, \Omega^b_p * L_q)$$

$$R^f_p = R_{\hat{q}} \text{ with } \hat{q} = \arg \min_{p_x - \delta_y \le q_x \le p_x} d'(\Omega^f_p * \Phi_p, \Omega^f_p * R_q)$$

$$R^b_p = R_{\hat{q}} \text{ with } \hat{q} = \arg \min_{p_x - \delta_y \le q_x \le p_x} d'(\Omega^b_p * \Phi_p, \Omega^b_p * R_q)$$

with L_q and R_q generic left-view and right-view patches centred on the generic point $\mathbf{q} \mid q_y = p_y$. Here the symbol $*$ denotes point-wise multiplication between images (or patches). The distance $d'(\Pi_1, \Pi_2)$ between two generic patches Π_1 and Π_2 is defined as the sum of squared distances (SSD) of pixel values where artefact pixels are ignored. The value δ_y which restricts the search region for

Figure 5: **Patch distances.** These two plots show the values of the weighted SSD distance functions $d'(\Omega^f_p * \Phi_p, \Omega^f_p * R_q)$ and $d'(\Omega^b_p * \Phi_p, \Omega^b_p * R_q)$ for varying values of $q_x \in [p_x - \delta_y, p_x]$ along the right scanline. Φ_p is the example source patch in fig. 4. Restricting the search to a small scanline segment of length δ_y achieves efficiency. Our winner-take-all algorithm selects the patches R^f_p and R^b_p corresponding to the minima of the above plots (marked in green and blue, *cf.* fig. 4). These correspond to the two correct foreground and background disparities at point p.

efficiency is a scanline-dependent value defined as $\delta_y = \max_{\mathbf{q} \in I \mid q_y = y} \bar{D}(\mathbf{q})/2$. Examples of such automatically extracted patches are shown in fig. 4.

The SPS algorithm may be interpreted as a *winner-take-all* algorithm for dense stereo. However, unlike previous approaches of this kind, here the algorithm is applied twice, once to the foreground and once to the background portions of the source patch Φ_p. This has the effect of assigning *two* depth values to the artefact pixels in $\mathcal{A}$. Notably, in the case of mixed pixels the two estimated depths correspond to the depths of the foreground and background components of the mix. The reduced search region and the large autocorrelation of the Φ_p patch[3] make the typically fragile winner-take-all algorithm sufficiently robust.

Efficiency. It is important to stress that the SPS algorithm is economical since for each point $\mathbf{p} \in \mathcal{A}$ the search region is restricted to a short scanline segment of length δ_y. Figure 5 shows the typical behaviour of the patch distance function for varying values of the q_x coordinate.

3.3.2. Selecting the best background patch. Figure 4 demonstrates the successful detection of the two left and right foreground patches L^f_p and R^f_p (the foreground hair curl is present in both and in the same position). However, due to occlusion, only one of the two retrieved background patches is meaningful. In fact, the true background of Φ_p (the vertical door frame) is occluded in the left view I_l, thus the retrieved patch L^b_p is meaningless. In contrast, the right background patch R^b_p contains the correct background information. The actual choice between L^b_p and R^b_p is performed automatically by retaining the patch Π^b_p which is most similar to the background of Φ_p, *i.e.* : $\Pi^b_p = \arg\min_{\Lambda \in \{L^b_p, R^b_p\}} d'(\Omega^b_p * \Lambda, \Omega^b_p * \Phi_p)$. In the example in fig. 4 the selected background patch is $\Pi^b_p = R^b_p$.

[3] Artefacts tipically occur along high-contrast object boundaries.

I- 345

Figure 4.13: The final restored document image for Figure 4.1.

4.6 Summary

In this chapter, we have addressed the problems of distorted images while scanning bound documents, and discussed the real world shape-from-shading problems recovering the 3D shape and the albedo distribution of the book surface from a scanner image. We have presented a restoration method that solves these problems

using the shape and albedo distribution obtained based on the 3D geometric model and 3D optical model. This method has successfully removed the photometric distortion by de-shading the image based on our de-shading model, corrected the geometric distortion by adjusting the warped book surface to its flat rendition, and de-skewed the document images based on our de-warping model.

The method proposed in this chapter is efficient and effective. We will present the evaluation of the restoration results and comparison of the proposed methods in Chapter 3 and Chapter 4 in the next chapter.

The work described in this chapter has been variously reported in the following papers:

[111] (ICPR 04): A restoration method based on the 3D geometric model and 3D optical model, which restore the document images by de-shading model and de-warping model.

[110] (CVPR 04): An improvement of [111] by using an iterative method to reconstruct the book surface shape.

[112] (To be submitted to PAMI): An improvement of [110] by taking account of the document skew and proposing a new algorithm for reconstruction of the book surface shape.

Chapter 5
Experimental Evaluation & Comparison

5.1 Introduction

In this chapter, we present the evaluation of the restoration results and the comparison of the two restoration methods introduced in Chapter 3 and 4 respectively. Since one important purpose of our DIR is for subsequent document image analysis, understanding, and, finally, recognition of the document images, and OCR played a fundamental role in document image recognition domain [57], we evaluate the restoration results by presenting an experiment to show the OCR improvement. We compare the two methods on the effectiveness and efficiency, and present a discussion on the advantages and disadvantages.

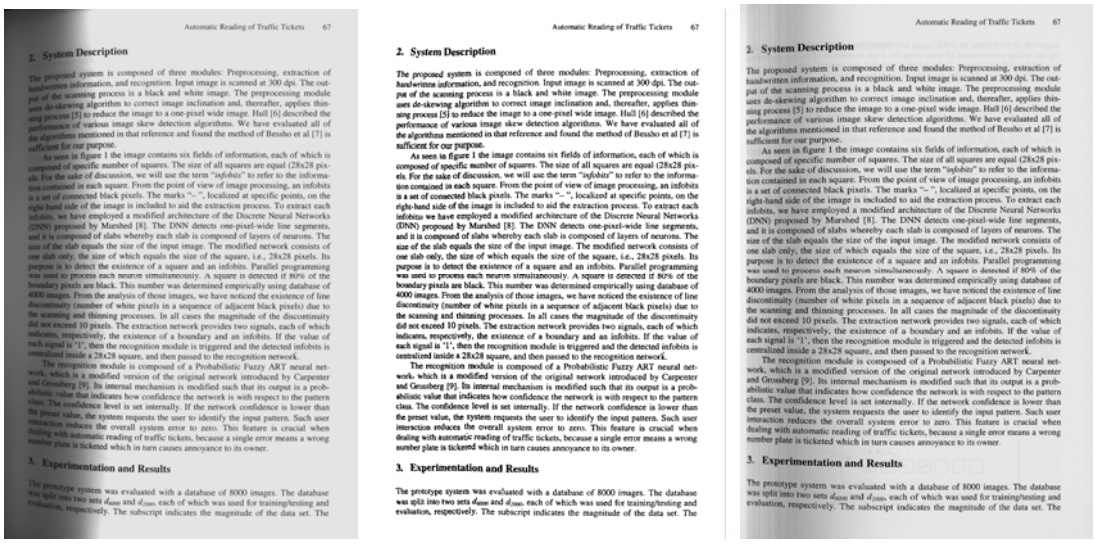
The rest of this chapter is organized as follows: The OCR evaluation is presented in Section 5.2. We compare the two methods on effectiveness and efficiency, and present a discussion in Section 5.3. Finally, Section 5.4 summarizes this chapter.

5.2 Experimental Evaluation

As a measure of success, we evaluate the restoration results by carrying out the OCR test on the original document image and the corresponding restored images by the two methods respectively. We perform the OCR test using three well-known commercial OCR products (latest versions):

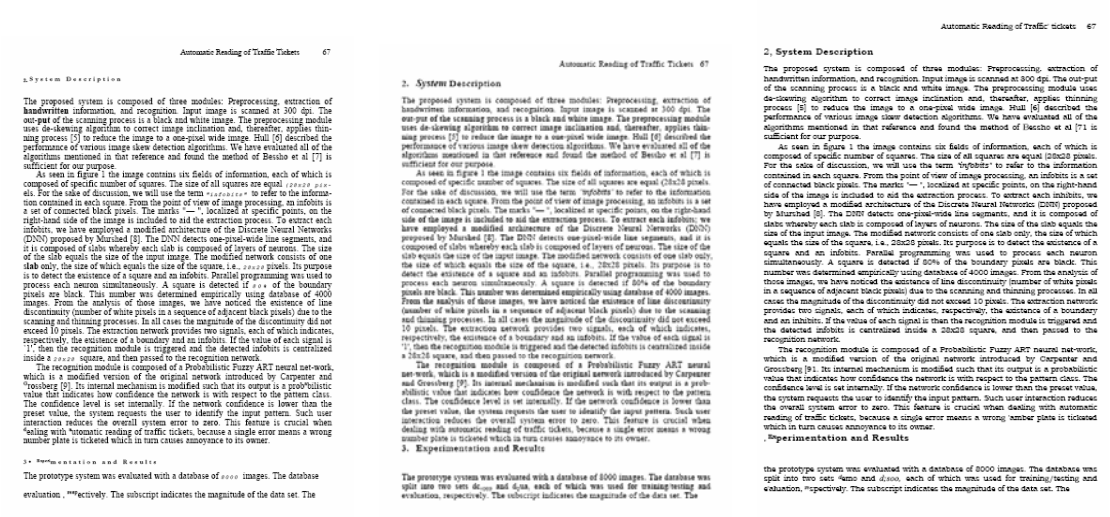
- Scansoft OmniPage Pro 14.0 [3]
- IRIS Readiris Pro 9.0 [2]
- ABBYY FineReader Pro 7.0 [1]

Figure 5.1 (a-c) shows a grayscale scanner document image and the corresponding restored images by the methods in Chapter 3 and 4 respectively. Figures 5.2 (a-c), 5.3 (a-c), and 5.4(a-c) show the OCR results by OmniPage, Readiris, and FineReader respectively on Figure 5.1 (a-c) correspondingly.



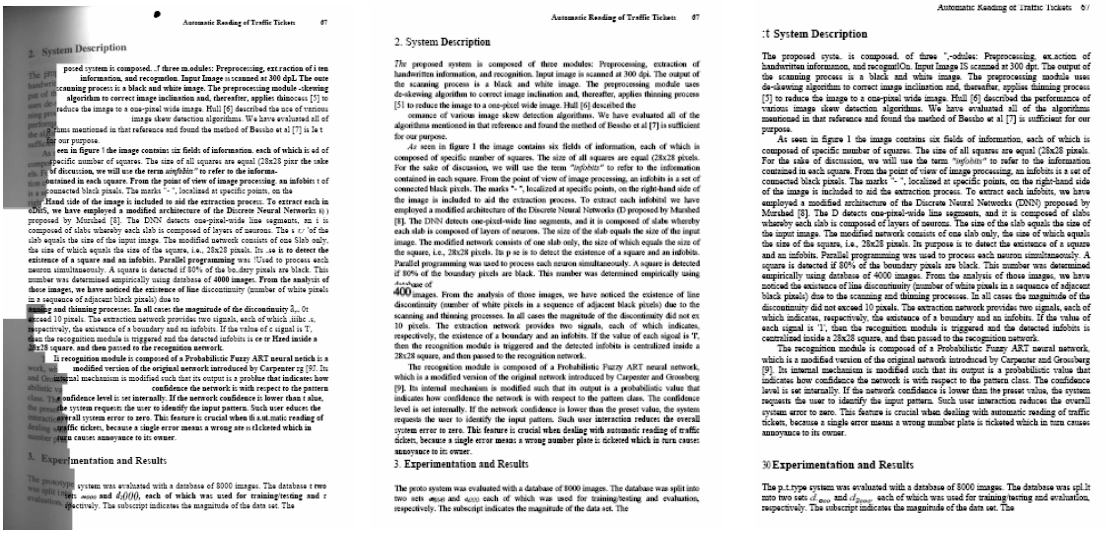
(a) **(b)** **(c)**

Figure 5.1: Distorted image and restored images: (a) A grayscale image scanned from a book; (b) The restored image by the method introduced in Chapter 3; (c) The restored image by the method introduced in Chapter 4.



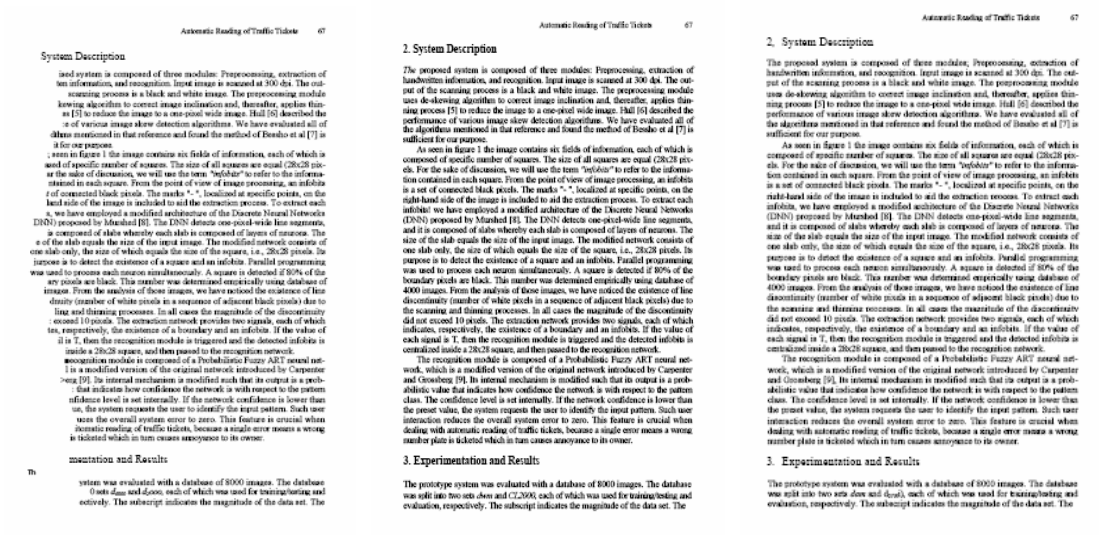
(a) (b) (c)

Figure 5.2: OmniPage OCR results for Figure 5.1(a), (b) and (c) respectively.



(a) (b) (c)

Figure 5.3: Readiris OCR results for Figure 5.1(a), (b) and (c) respectively.



(a) **(b)** **(c)**

Figure 5.4: FineReader OCR results for Figure 5.1(a), (b) and (c) respectively.

Note that all the three OCR products perform not well for the original scanned image in Figure 5.1(a), due to the serious shading and warping. We observe that most of the characters and words in Figures 5.2 (b-c), 5.3(b-c) and 5.4(b-c) are recognized correctly.

Precision and recall [35] on characters and words defined below are used as the evaluation metrics for the OCR results:

$$\text{Character Precision} = \frac{\text{Number of characters correctly detected by OCR}}{\text{Total number of characters detected by OCR}}$$

$$\text{Character Recall} = \frac{\text{Number of characters correctly detected by OCR}}{\text{Total number of characters in the document}}$$

$$\text{Word Precision} = \frac{\text{Number of words correctly detected by OCR}}{\text{Total number of words detected by OCR}}$$

$$\text{Word Recall} = \frac{\text{Number of words correctly detected by OCR}}{\text{Total number of words in the document}}$$

We scanned 100 document pages with resolution 200, 300 and 400 ppi respectively. In Tables 5.1-5.6, we show the average character precision/recall and word precision/recall respectively for the original scanner document images, the images restored by the method introduced in Chapter 3, and the images restored by the method introduced in Chapter 4 respectively. We summarize the improvements on character precision/recall and word precision/recall by the two methods in Tables 5.7 and 5.8 respectively. Our two methods significantly improve the character precision (ranging from 10.43% to 13.47% and from 10.55% to 13.82% respectively), character recall (ranging from 7.62% to 17.33% and from 7.47% to 17.43% respectively), word precision (ranging from 11.15% to 15.65% and from 11.48% to 15.28% respectively), and word recall (ranging from 10.04% to 18.42% and from 9.93% to 17.94% respectively) for the three tested OCR products. These improvements demonstrate the success of the two restoration methods proposed in Chapters 3 and 4.

Table 5.1: Average character precision and recall for the original scanner document images.

	Average Character Precision (%)			Average Character Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	83.96	80.34	82.93	90.49	79.13	79.01
300 ppi	84.14	83.31	84.13	90.94	80.92	80.16
400 ppi	84.96	85.14	85.01	91.43	81.22	80.92

Table 5.2: Average word precision and recall for the original scanner document images.

	Average Word Precision (%)			Average Word Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	81.87	78.39	80.48	87.11	76.13	77.64
300 ppi	82.54	82.15	81.11	87.93	78.74	78.60
400 ppi	82.59	84.19	82.60	88.01	78.89	79.29

Table 5.3: Average character precision and recall for the images restored by the method proposed in Chapter 3.

	Average Character Precision (%)			Average Character Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	96.82	93.11	96.24	98.56	94.17	96.34
300 ppi	97.61	94.45	96.88	98.85	94.96	96.95
400 ppi	98.12	95.57	97.37	99.05	95.73	97.15

Table 5.4: Average word precision and recall for the images restored by the method proposed in Chapter 3.

	Average Word Precision (%)			Average Word Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	96.34	93.04	95.82	97.66	94.12	96.06
300 ppi	96.39	94.18	96.76	97.97	94.54	96.55
400 ppi	96.87	95.34	97.12	98.17	94.99	96.86

Table 5.5: Average character precision and recall for the images restored by the method proposed in Chapter 4.

	Average Character Precision (%)			Average Character Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	96.80	93.45	96.75	98.53	94.10	96.44
300 ppi	97.54	94.74	97.71	98.73	94.87	96.98
400 ppi	97.91	95.69	98.46	98.90	95.30	97.27

Table 5.6: Average word precision and recall for the images restored by the method proposed in Chapter 4.

	Average Word Precision (%)			Average Word Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	96.02	93.27	95.75	97.54	93.22	95.43
300 ppi	96.41	94.63	96.39	97.89	94.10	96.54
400 ppi	96.43	95.67	96.76	97.94	94.85	96.61

Table 5.7: Improvement on average precision and recall by the method proposed in Chapter 3.

	Improvement on Average Character Precision (%)			Improvement on Average Character Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	12.86	12.77	13.31	8.07	15.04	17.33
300 ppi	13.47	11.14	12.75	7.91	14.04	16.79
400 ppi	13.16	10.43	12.36	7.62	14.51	16.23

	Improvement on Average Word Precision (%)			Improvement on Average Word Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	14.47	14.65	15.34	10.55	17.99	18.42
300 ppi	13.85	12.03	15.65	10.04	15.80	17.95
400 ppi	14.28	11.15	14.52	10.16	16.10	17.57

Table 5.8: Improvement on average precision and recall by the method proposed in Chapter 4.

	Improvement on Average Character Precision (%)			Improvement on Average Character Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	12.84	13.11	13.82	8.04	14.79	17.43
300 ppi	13.40	11.43	13.58	7.79	13.95	16.82
400 ppi	12.95	10.55	13.45	7.47	14.08	16.35
	Improvement on Average Word Precision (%)			Improvement on Average Word Recall (%)		
Resolution	OmniPage	Readiris	FineReader	OmniPage	Readiris	FineReader
200 ppi	14.15	14.88	15.27	10.43	17.09	17.78
300 ppi	13.87	12.48	15.28	9.96	15.36	17.94
400 ppi	13.84	11.48	14.16	9.93	15.96	17.32

5.3 Comparison

We compare the two restoration methods proposed in Chapters 3 and 4 based on two issues: effectiveness and efficiency. Finally, we present a discussion on the two

methods.

5.3.1 Effectiveness

The Chapter 3 method removes the shading by binarizing the document image, and corrects the text warping by projecting and rotating the characters based on a reference line/curve approach. The graphical objects are removed by the filters, and no restoration is made on them. The method only works on textual contents. Although the characters are moved and rotated to the proper location and orientation, the shapes of the characters in the shade still appear distorted. For example, the characters in the shade appear narrower than the ones in the other area. Some parameters of this method have to vary with the resolution of the document image to be restored. For example, we set the window size of our binarization method to be 7×7, 11×11, and 15×15 for resolution 200, 300, and 400 ppi images respectively. Some parameters may require manually setting for each testing image. For example, the parameters of the noise filters may be manually set to remove the small noise and large graphical objects for each image.

The Chapter 4 method restores the document images by the de-shading model and de-warping model using the albedo distribution and book surface shape respectively, which are discovered based on the 3D geometric model and 3D optical model. The restoration is pixel-based, instead of connected-component-based as in the Chapter 3 method. Thus the photometric distortion is removed without downgrading the image from grayscale to binary, and the geometric distortion is corrected not only

for the textual contents but also for the graphical contents. Furthermore, the shapes of both the textual contents and graphical contents are fully restored, instead of partially restored by the Chapter 3 method. Since the scanner's parameters are fixed for one type of scanner, even for one series of scanners, the document images scanned at any resolutions by the same type of scanner may be restored without changing of method parameters. We summarize the comparison on the effectiveness of the two restoration methods in Table 5.9.

Table 5.9: Comparison on effectiveness between the methods proposed in Chapters 3 and 4 respectively.

	Chapter 3 Method	Chapter 4 Method
1. Correct the shading?	Yes (but downgrade the image from grayscale to binary)	Yes
2. Correct the warping?	Yes (but the distorted shapes of the characters remain unchanged)	Yes
3. Restore textual contents?	Yes	Yes
4. Restore graphical contents?	No (graphical contents are removed by the noise filters)	Yes
5. Fixed parameter setting for all experimental images?	No (generally, the parameters vary with the scanning resolutions, and some parameters may require manual setting for each image)	Yes (for the images captured by the same scanner at any resolution)

5.3.2 Efficiency

The Chapter 3 method adopts three computational expensive algorithms:

- The binarization algorithm (Section 3.3): this algorithm is a local thresholding technique, which requires calculating the mean and standard deviation for a number of the neighbors of each pixel in the document image.
- The Otsu algorithm (Section 3.6.1): The criterion function must be computed for each possible values of run-length in the profile.
- The box-hands method (Section 3.6.2): the connected components are first clustered into words; the linear regression is applied to approximate the word orientation; the two box-hands are generated for each word, according to the word orientation and height; finally, we have to check whether the two hands of each word touch any hand of all the other words.

The Chapter 4 method basically consists of the processes for document skew detection, the book surface reconstruction, de-shading, de-warping along *x*-axis and *y*-axis, and de-skew of the document image respectively. The runtime of this method is of the order of few seconds. This method is very efficient even comparing with other methods based on 3D surface shape. For example, we simulate Wada's method [95] on the same machine, and the runtime is beyond 1 minute even for 200 ppi images. This is due to several computational expensive processes, such as piecewise polynomial fitting, iterative objective function minimization, interreflection computing by the 3D tessellation of the book surface.

In Table 5.10, we summarize the efficiency by presenting the average runtime on a

Pentium IV 2.6G PC of the two restoration methods with the default parameter settings.

Table 5.10: Comparison on efficiency between the methods proposed in Chapters 3 and 4 respectively.

	Average Runtime on a PIV 2.6G PC (sec)	
Resolution	Chapter 3 Method	Chapter 4 Method
200 ppi	31.67	2.12
300 ppi	40.15	4.71
400 ppi	58.61	8.38

5.3.3 Discussion

The Chapter 3 method can be controlled step by step, and the restored images may achieve higher OCR precision/recall, if the parameters are set properly. Some techniques introduced in this method may contribute the research in Document Image Analysis/Understanding/Recognition. For example, the modified Niblack's method for image thresholding, and the box-hands approach for document image segmentation and layout analysis.

From the comparison on effectiveness and efficiency in the last two sections, generally the Chapter 4 method is more practical than the Chapter 3 method. Table 5.11 shows the scanning time for an Epson 1640XL scanner (the scanner for our experimental images). Tables 5.9-5.11 prove that Chapter 4 method may be incorporated in a scanner software to correct the photometric and geometric distortion

for practical use. The above study of the OCR performance will provide opportunities for future analysis of the causes of OCR errors so as to further enhance the effectiveness of our methods.

Table 5.11: Scanning time of Epson 1640XL scanner for different size documents.

	Scanning Time (sec)				
Resolution	Document Size				
	B5	Letter	A4	B4	A3
200 ppi	13	14	15	17	20
300 ppi	19	20	21	23	26
400 ppi	26	27	28	35	37

Furthermore, comparing the Chapter 4 method with other methods based on the 3D document shape, which are introduced in Chapter 2, it has the following five advantages:

- It does not require textual contents or document boundary in the image to be restored as in Cao's method [18, 19] and Tsoi's method [94] respectively.
- It does not require special setups as in Brown's method [15], Doncescu's method [26] and Yamashita's method [101].
- It does not require aligning the book to be scanned strictly parallel to the scanning direction, which is a strong assumption in Kanungo's model [40, 41] and Wada's method [95].

- It does not assume the book surface is cylindrical, which may not be true in real scanning conditions, as in Cao's method [18, 19], Kanungo's model [40, 41] and Wada's method [95].
- It is much more efficient than Brown's method [15], Pilu's method [74, 75], Wada's method [95], and Yamashita's method [101]. For instance, through communication with Brown and Yamashita, it is known that their methods take hours to complete. We also re-implemented Wada's method and found that its processing time is in the order of minutes.

5.4 Summary

In this Chapter, we have presented the OCR results from the original image and the ones restored by the two methods in Chapters 3 and 4 respectively. The significant improvements prove the success of the two methods. We have compared the two methods on the effectiveness and efficiency, and presented a discussion.

Chapter 6
Conclusions

6.1 Summary

In this thesis, we address the problems of DIR for the document images scanned from bound volumes. We propose two restoration approaches based on 2D document image processing and 3D document shape discovery respectively. These methods have been shown the ability to correct both photometric and geometric distortion in the scanner images. The improvement on OCR precision/recall of three well-known commercial OCR products proves the success of the two approaches. We compare the two approaches based on two issues: effectiveness and efficiency, and present a discussion on the advantages of each approach.

6.2 Contributions

In this thesis, we have proposed two restoration methods that are reported in Chapters

3 and 4, respectively. The two methods are unique in their own right, in comparison with existing works.

First, comparing with the other methods based on 2D document image processing, the Chapter 3 method adopts a novel binarization method to remove the photometric distortion, which is not considered in Lavialle's method [49, 5], Tang's method [85, 86] and Weng's method [97]. This method corrects the nonuniform geometric distortion of the textual contents, which cannot be tackled by O'Gorman's method [63] and Tang's method [85, 86]. This is done through an oriented box-hands method and a reference line/curve approach. The Chapter 3 method is a preliminary solution to our DIR problem which provides insight and motivation for us to venture into the next method which is reported in Chapter 4. Essentially, we wanted to find a totally different approach which, unlike the method in Chapter 3, will dispense with the need for textual contents in the document page.

Chapter 4's method is a new attempt that has proven superior to other methods based on the 3D document shape discovery. Comparing with the existing methods, our method requires neither textual contents and document boundary in the document image as in Cao's method [18, 19] and Tsoi's method [94], nor special camera setups as in Brown's method [15], Doncescu's method [26] and Yamashita's method [101]. Our method neither assumes the document to be scanned aligned strictly parallel to the scanning direction, which is a strong assumption in Kanungo's model [40, 41] and Wada's method [95], nor does it assume the document surface to be cylindrical as in Cao's method [18, 19], Kanungo's model [40, 41] and Wada's method [95]. Moreover,

it is much more efficient than Brown's method [15], Pilu's method [74, 75], Wada's method [95] and Yamashita's method [101].

We are pleased to note that the research has resulted in a number of technical papers in the literature [106, 107, 108, 109, 110, 111]. The contributions are summarized as follows:

1. A DIR approach based on 2D document image processing, which mainly consists of:

- A run-length method to detect the shade boundary: This run-length method can successfully detect the shade boundary, which classifies the connected components into two classes – connected components in the shade area and connected components in the clean area. This method also helps the document skew detection based on Hough Transform in the restoration approach based on 3D document surface discovery.
- A modified Niblack's method: This method amplifies the contribution of standard deviation, which led to produce results with much less pepper noise than the original Nibalck's method. This method also reduces the sensitivity of parameter *k*, which allows us to find a single *k* value for most of testing document images. As shown in Chapter 3, the binarization results using our method are much better than the ones using Niblack's method.
- Two noise filters based on the size and shape of connected components: The filters can successfully remove the noise and graphical objects by utilizing the irregular size and shape of them, which are quite different with the ones of the

textual objects.

- A clustering method based on box-hands with orientation: This box-hands method decides the hands orientation and length by the word orientation approximated by linear regression and the word height respectively. It can effectively cluster the words into text lines, irrespective of they are in warped or straight text lines.
- A text line straightening method based on the reference lines/curves detected by linear/quadratic regression: This method models the warped part of the text lines by two quadratic reference curves, and the straight part of the text lines by two straight reference lines. It straightens the warped text line by moving and rotating the characters to the proper location and orientation respectively, based on the reference lines/curves.

2. A DIR approach based on 3D document shape discovery, which mainly consists of:

- Two practical models tolerable to document skew – the 3D geometric model and 3D optical model: The 3D geometric model describes the geometric property of the book surface shape. The 3D optical model describes the optical process of generating the scanner images by a special Shape-from-Shading approach considering four factors. The two practical models adequately simulate the real world scanning conditions for our purpose.
- A method to reduce the 3D surface shape to a unique 2D cross section shape: This method reduces the 3D surface shape to a unique 2D cross section shape by

defining a processing area of the document image and discovering the relation between the incident angle of the light source and the corresponding slant angle of the book surface shape.

- A surface shape reconstruction and albedo distribution discovery method based on the two practical models: This method reconstruct the book surface shape with the help of the background with constant albedo, based on the geometric model and optical model.
- A de-shading model based on the surface albedo distribution: This model can correct the photometric distortion by recalculating the optimal pixel intensity with zero *z* values based on the surface albedo distribution and the 3D optical model.
- A de-warping model based on the surface cross section shape: This model tackles the geometric distortion by correcting the distortion along *x*-axis and *y*-axis, and de-skewing the document, based on the surface cross section shape and the document skew detected by Hough Transform.

6.3 Future Work

For the DIR approach based 2D image processing, we propose the following future works:

- To restore the graphical objects, which are removed by the noise filters, we may explore a way by edge detection with help of our reference lines/curves from textual objects.

- To correct the shape of the distorted characters, we may apply some shear techniques, after projecting and rotating the characters.
- To improve the runtime, we may further optimize the implementation.

For the DIR approach based on 3D document shape discovery, we suggest the future works in the following ways:

- Extend the system to handle color document images, instead of only grayscale ones.
- Though most of papers are Lambertian, some documents may contain plates, which have specular reflection property. Some specular models may be further studied to find the best one fitting this situation.

Bibliography

[1] http://www.abbyy.com/finereader/

[2] http://www.irisusa.com/products/readiris/pc/

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

[4] A. S. Abutaleb, “Automatic Thresholding of Gray-level Pictures using Two-dimensional entropy”, *Computer Vision, Graphics, and Image Processing*, Volume 47, pp. 22-32, 1989.

[5] F. Angella, O. Lavialle, and P. Baylou. “A Deformable and Expansible Tree for Structure Recovery”, *International Conference on Image Processing (ICIP)*, Volume 1, pp 241-245, 1998.

[6] H. Baird. “Difficult and Urgent Open Problems in Document Analysis for Libraries”, *International Workshop on Document Image Analysis for Libraries (DIAL)*, pp. 25-32, 2004.

[7] H. Baird. “Digital Libraries and Document Image Analysis”, *International Conference on Document Analysis and Recognition (ICDAR)*, Volume 1, pp. 2-14, 2003.

[8] H. Baird. “The State of the Art of Document Image Degradation Modeling”, *IAPR*

International Workshop on Document Analysis System (DAS), pp. 1-16, 2000.

[9] H. Baird. “Document Image Quality: Making Fine Discriminations”, *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 459-462, 1999.

[10] H. Baird. “Document Image Defect Models and Their Uses”, *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 730-734, 1993.

[11] H. Baird. “Calibration of Document Image Defect Models”, *Annual Symposium on Document Analysis and Information Retrieval (SDAIR)*, pp. 1-16, 1993.

[12] H. Baird. “Document Image Defect Models”, *International Workshop on Syntactic and Structural Pattern Recognition (SSPR)*, pp. 38-46, 1990.

[13] J. Bernsen, “Dynamic Thresholding of Grey-level Images”, *International Conference on Pattern Recognition (ICPR)*, pp.1251-1255, 1986.

[14] A. Blake and A. Zisserman, Visual Reconstruction, MIT Press, Cambridge, MA, 1987.

[15] 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 (ICCV)*, Volume 2, pp. 367-374, 2001.

[16] M. S. Brown and W. B. Seales, “Digital Atheneum: New Approach for Preserving, Restoring and Analyzing Damaged Manuscripts”, *IEEE/ACM Joint Conference on Digital Library*, pp. 437-443, 2001.

[17] M. S. Brown and W. B. Seales, “Beyond 2D Images: Effective 3D Imaging for Library Materials”, *ACM Conference on Digital Library (ACM DL)*, pp. 27-36, 2000.

[18] H. Cao, X. Ding, and C. Liu. “A Cylindrical Model to Rectify the Bound

Document Image", *International Conference on Computer Vision (ICCV)*, Volume 2, pp. 228-233, 2003.

[19] H. Cao, X. Ding, and C. Liu. "Rectifying the Bound Document Image Captured by the Camera: A Model Based Approach", *International Conference on Document Analysis and Pattern Recognition (ICDAR)*, Volume 1, pp. 71-74, 2003.

[20] D. Carmo, Differential Geometry of Curves and Surfaces, Prentice Hall, 1976.

[21] M. Chang, S. Kang, W. Rho, H. Kim, and D. Kim, "Improved Binarization Algorithm for Document Image by Histogram and Edge Detection", *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 636-643, 1995.

[22] W. Chen, C. Wen, and C. Yang, "A Fast Two-dimensional Entropic Thresholding Algorithm", *Pattern Recognition*, Volume 27(7), pp. 885-893, 1994.

[23] C. K. Chow and T. Kaneko, "Automatic Detection of the Left Ventricle from Cineangiograms", *Computers and Biomedical Research*, Volume 5, pp. 388-410, 1972.

[24] S. Coons, "Surface for Computer Aided Design", Technical Report, MIT, 1968.

[25] A. K. Das and B. Chanda, "A Fast Algorithm for Skew Detection of Document Images using Morphology", *International Journal on Document Analysis and Recognition (IJDAR)*, Volume 4, 109-114, 2001.

[26] A. Doncescu, A. Bouju, and V. Quillet. "Former Books Digital Processing: Image Warping", *International Workshop of Document Image Analysis (DIA)*, pp. 5-9, 1997.

[27] L. Eikvil, T. Taxt, K. Moen, "A Fast Adaptive Method for Binarization of Document Images", *International Conference on Document Analysis and Recognition*

(ICDAR), pp. 435-443, 1991.

[28] G. Farin. Cruves and Surfaces for Computer Aided Geometric Design, Academic Press San Diego, CA, 1990.

[29] O. Faugeras, Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, Cambridge, MA, 1993.

[30] R. M. Haralick and L. G. Shapiro, Machine Vision, Addison-Wesley Publishing Co., Inc., Reading, MA, 1992.

[31] B. K. P. Horn, Robot Vision, MIT Press, Cambridge, MA, 1986.

[32] L. K. Huang and M. J. Wang, "Image Thresholding by Minimizing the Measure of Fuzziness", *Pattern Recognition*, Volume 28, pp. 41-51, 1995.

[33] R. Jain, R. Kasturi, and B. G. Schunck, Machine Vision, MIT Press and McGraw-Hill, 1995.

[34] G. Johannsen and J. Bille, "A Threshold Selection Method using Information Measures", *International Conference on Pattern Recognition (ICPR)*, pp. 140-143, 1982.

[35] M. Junker, R. Hoch, A. Dengle. "On the Evaluation of Document Analysis Components by Recall, Precision and Accuracy", *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 713-716, 1999.

[36] T. Kanungo, R.M. Haralick, and H. Baird. "Validation and Estimation of Document Degradation Models", *Annual Symposium on Document Analysis and Information Retrieval (SDAIR)*, pp. 217-225, 1995.

[37] T. Kanungo, R.M. Haralick, and H. Baird. "Power Functions and Their Use in

Selecting Distance Functions for Document Degradation Model Validation", *International Conference on Document Analysis and Recognition (ICDAR)*, Volume 2, pp. 734-739, 1995.

[38] T. Kanungo, R.M. Haralick, H. Baird, W. Stuetzle, and D. Madigan. "A Statistical, Nonparametric Methodology for Document Degradation Model Validation", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, pp. 1209-1223, 2000.

[39] T. Kanungo, R.M. Haralick, and H. Baird, W. Stuetzle, and D. Madigan. "Document Degradation Models: Parameter Estimation and Model Validation", *International Workshop on Machine Vision Applications (MVA)*, pp. 11-15, 1994.

[40] T. Kanungo, R.M. Haralick and I. Phillips. "Nonlinear Global and Local Document Degradation Models", *International Journal of Imaging System and Technology (IJIST)*, pp. 220-230, 1994.

[41] T. Kanungo, R.M. Haralick and I. Phillips. "Global and Local Document Degradation Models", *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 730-734, 1993.

[42] T. Kanungo and Q. Zheng. "Estimating Degradation Model Parameters using Neighborhood Pattern Distributions: an Optimization Approach", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, pp. 520-524, 2004.

[43] T. Kanungo and Q. Zheng. "Estimation of Morphological Degradation Model Parameters", *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1961-1964, 2001.

[44] J. N. Kapur, P. K. Sahoo and A. K. Wong, “A New Method for Gray-level Picture Thresholding using the Entropy of the Histogram”, *Computer Vision Graphics and Image Processing (CVGIP)*, Volume 29, pp. 273-285, 1985.

[45] M. Kass, A. Witkin and D. Terzopoulos, “Snakes: Active Contour Models”, *International Journal of Computer Vision (IJCV)*, Volume 1, pp. 321-331, 1988.

[46] J. Kittler and J. Illingworth, “Minimum Error Thresholding”, *Pattern Recognition*, Volume 19, pp. 41-47, 1986.

[47] J. Kittler and J. Illingworth, “On Threshold Selection using Clustering Criteria”, *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 15, pp. 652-655, 1985.

[48] T. Kurita, N. Otsu and N. Abdelmalek, “Maximum Likelihood Thresolding based on Population Mixture Models”, *Pattern Recognition*, Volume 25(10), pp. 1231-1240, 1992.

[49] O. Lavaille, X. Molines, F. Angella, and P. Baylou. “Active Contours Network to Straighten Distorted Text Lines”, *International Conference on Image Processing (ICIP)*, pp. 1074-1077, 2001.

[50] T. M. Lehmann, C. Gonner and K. Spitzer, “Survey: Interpolation Methods in Medical Image Processing”, *IEEE Transactions on Medical Image*, Volume 18(11), pp.1049-1075, 1999.

[51] Z. C. Li, “Splitting Integrating Method for Normalizing Image by Inverse Transformation”, *Pattern Recognition*, pp. 678-686, 1992.

[52] Z. C. Li, Y. Y. Tang, T. D. Bui, and C. Y. Suen, “Shape Transformation Models and Their Applications in Pattern Recognition”, International Journal on Pattern

Recognition and Artificial Intelligence, Volume 4(1), pp.65-94, 1990.

[53] Y. Liu, R. Fenrich, and S. N. Srihari, “An Object Attribute Thresholding Algorithm for Document Image Binarization”, *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 278-281, 1993.

[54] Y. Liu and S. N. Srihari, “Document Image Binarization based on Texture Features”, *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 19(5), pp. 540-544, 1997.

[55] K. V. Mardia and T. J. Hainsworth, “A Spatial Thresholding Method for Image Segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 10(6), pp. 919-927, 1988.

[56] G. K. Myers, R. C. Bolles, Q. T. Luong, and J. A. Herson, “Recognition of Text in 3-D Scenes”, *Symposium on Document Image Understanding Technologies (SDIUT)*, pp.85-100, 2001.

[57] G. Nagy, “Twenty years of document image analysis in PAMI”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 22, No. 1, pp. 38-62, 2000.

[58] Y. Nakagawa and A. Rosenfeld, “Some Experiments on Variable Thresholding”, *Pattern Recognition*, Volume 11(3), pp. 191-204, 1979.

[59] S. K. Nayar, K. Ikeuchi and T. Kanade, “Shape from Interreflections”, *International Conference on Computer Vision (ICCV)*, pp. 2-11, 1990.

[60] W. Niblack, *An Introduction to Image Processing*, Prentice Hall, pp. 115-116, 1986.

[61] L. O'Gorman and R. Kasturi, "Document Image Analysis", *IEEE Computer Society Press*, 1995.

[62] L. O'Gorman, "Binarization and multithresholding of document images using connectivity", *Computer Vision Graphics and Image Processing (CVGIP): Graphical Models and Image Processing*, Volume 56(6), pp. 494-506, 1994.

[63] L. O'Gorman, "The Document Spectrum for Page Layout Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 15(11), pp. 1162-1173, 1993.

[64] J. Ortega and W. Rheinboldt, "Iterative Solution of Nonlinear Equations in Several Variables", *New York: Academic*, 1970.

[65] N. Otsu, "A Threshold Selection Method from Gray-level Histograms", *IEEE Transactions on System Man and Cybernetics*, Volume 9(1), pp. 62-69, 1979.

[66] N. Papamarkos, "A Technique for Fuzzy Document Binarization", *ACM Symposium on Document Engineering (ACM DocEng)*, pp. 152-156, 2001.

[67] N. Papamarkos and B. Gatos, "A New Approach for Multithreshold Selection", *Computer Vision Graphics and Image Processing (CVGIP): Graphical Models and Image Processing*, Volume 56(5), pp. 357-370, 1994.

[68] J. R. Parker, "Gray Level Thresholding on Badly Illuminated Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 13(8), pp. 813-819, 1991.

[69] J. R. Parker, C. Jennings, A. G. Salkauskas, "Thresholding using an Illumination Model", *International Conference on Document Analysis and Recognition (ICDAR)*,

pp. 270-273, 1993.

[70] T. Pavlidis, “Threshold Selection using Second Derivatives of the Grayscale Image”, *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 274-277, 1993.

[71] A. Perez and R. C. Gonzalez, “An Iterative Thresholding Algorithm for Image Segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 9(6), pp. 742-751, 1987.

[72] L. Piegl and W. Tiller, The NURBS Book (2nd Edition), Springer-Verlag, New York, 1997.

[73] A. Pikaz and A. Averbuch, “Digital Image Thresholding based on Topological Stable-state”, *Pattern Recognition*, Volume 29(5), pp. 829-843, 1996.

[74] M. Pilu, "Undoing Page Curl Distortion Using Applicable Surfaces", *Computer Vision and Pattern Recognition (CVPR)*, Volume 1, pp. 67-72, 2001.

[75] M. Pilu, "Undoing Page Curl Distortion Using Applicable Surfaces", *International Conference on Image Processing (ICIP)*, pp. 237-240, 2001.

[76] S. S. Reddi, S. F. Rudin and H. R. Keshavan, “An Optimal Multiple Threshold Scheme for Image Segmentation”, *IEEE Transactions on System Man and Cybernetics*, Volume 14(4), pp. 661-665, 1984.

[77] A. Rosenfeld, R. C. Smith, “Thresholding using Relaxation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 3(5), pp. 598-606, 1981.

[78] P. K. Sahoo, S. Soltani, and A. K. Wong, “A Survey of Thresholding Techniques”,

Computer Vision Graphics and Image Processing (CVGIP), Volume 41, pp.233-260, 1988.

[79] L. Shapiro and G. Stockman. *Computer Vision*, Prentice Hall, pp. 306-310, 2001.

[80] M. A. Smith and T. Kanade, "Video Skimming for Quick Browsing Based on Audio and Image Characterization", Technical Report CMU-CS-95-186, Carnegie Mellon University, 1995.

[81] D. B. Smythe, "A Two-Pass Mesh Warping Algorithm for Object Transformation and Image Interpolation", *ILM Technical Memo #1030, Computer Graphics Department, Lucasfilm Ltd*, 1990.

[82] C. Strouthopoulos and N. Papamarkos, "Multithresholding of Mixed Type Documents", *Engineering Application of Artificial Intelligence*, Volume 13(3), pp. 323-343, 2000.

[83] C. Strouthopoulos, N. Papamarkos and C. Chamzas, "Identification of text-only areas in mixed type documents", *Engineering Applications of Artificial Intelligence*, Volume 10(4), pp. 387-401, 1997.

[84] B. Su and D. Liu, "Computational Geometry Curve and Surface Modeling", *Academic Press, Shanghai Scientific and Technical Publishers*, 1989.

[85] Y. Y. Tang and C. Y. Suen. "Image Transformation Approach to Nonlinear Shape Restoration", *IEEE Transactions on Systems, Man, and Cybernetics*, Volume 23, No. 1, pp. 155-171, 1993.

[86] Y. Y. Tang and C. Y. Suen. "Nonlinear Shape Restoration by Transformation Models", *International Conference on Pattern Recognition (ICPR)*, Volume 2, pp.

14-19, 1990.

[87] T. Taxt, P. J. Flynn and A. K. Jain, “Segmentation of Document Images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 11(12), pp. 1322-1329, 1989.

[88] M. J. Taylor and C. R. Dance, “Enhancement of Document Images from Cameras”, *IS&T/SPIE Symposium on Electronic Imaging: Science and Technology*, pp. 230-241, 1998.

[89] D. Terzopoulos and K. Fleischer, “Modeling Inelastic Deformations Viscoelasticity, Plasticity, Fracture”, *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 269-278, 1988.

[90] D. Terzopoulos, J. C. Platt, and A. H. Barr, “Elastically Deformable Models”, *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 205-214, 1987.

[91] O. D. Trier and T. Taxt, “Improvement of Integrated Function Algorithm for Binarization of Document Images”, *Pattern Recognition Letters*, Volume 16(3), pp. 277-283, 1995.

[92] O. D. Trier and T. Taxt, “Evaluation of Binarization Method for Document Images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Volume 17(3), pp. 312-315, 1995.

[93] W. Tsai, “Moment-preserving Thresholding: a New Approach”, *Computer Vision, Graphics, and Image Processing (CVGIP)*, Volume 29, pp.377-393, 1985.

[94] Y. C. Tsoi and M. S. Brown, “Geometric and Shading Correction for Images of

Printed Materials – A Unified Approach Using Boundary", *Computer Vision and Pattern Recognition (CVPR)*, Volume 1, pp. 240-246, 2004.

[95] T. Wada, H. Ukida, and T. Matsuyama, "Shape from Shading with Interreflections Under a Proximal Light Source: Distortion-Free Copying of an unfolded Book", *International Journal of Computer Vision (IJCV)*, 24(2), pp. 125-135, 1997.

[96] T. Wada, H. Ukida, and T. Matsuyama, "Shape from Shading with Interreflections Under Proximal Light Source – 3D Shape Reconstruction of Unfolded Book Surface from a Scanner Image", *International Conference on Computer Vision (ICCV)*, pp. 66-71, 1995.

[97] Y. Weng and Q. Zhu, "Nonlinear Shape Restoration for Document Images", *Computer Vision and Pattern Recognition (CVPR)*, pp. 568-573, 1996.

[98] J. M. White and G. D. Rohrer, "Image Thresholding for Optical Character Recognition and Other Application Requiring Character Image Extraction", *IBM Journal on Research and Development*, Volume 27(4), pp. 400-411, 1983.

[99] C. Wolf, J. M. Jolion, and F. Chassaing, "Text Localization, Enhancement and Binarization in Multimedia Documents", *International Conference on Pattern Recognition (ICPR)*, Volume 4, pp. 1037-1040, 2002.

[100] V. Wu, R. Manmatha and E. Riseman, "Automatic Text Detection and Recognition", *In Proceedings of the DARPA Image Understanding Workshop*, pp. 707-712, 1997.

[101] A. Yamashita, A. Kawarago, T. Kaneko, and K. T. Miura, "Shape

Reconstruction and Image Restoration for Non-Flat Surface of Document with a Stereo Vision System", *International Conference on Pattern Recognition (ICPR)*, 2004.

[102] J. Yang, Y. Chen, and W. Hsu, "Adaptive Thresholding Algorithm and its Hardware Implementation", *Pattern Recognition Letters*, Volume 15(2), pp. 141-150, 1994.

[103] Y. Yang and H. Yan, "An Adaptive Logical Method for Binarization of Degraded Document Images", *Pattern Recognition*, Volume 33(5), pp.787-807, 2000.

[104] S. D. Yanowitz and A. M. Bruckstein, "A New Method for Image Segmentation", *Computer Vision Graphics and Image Processing (CVGIP)*, pp. 82-95, 1982.

[105] Q. Zheng and T. Kanungo, "Morphological Degradation Models and Their Use in Document Image Restoration", *International Conference on Image Processing (ICIP)*, pp. 193-196, 2001.

[106] Z. Zhang and C.L. Tan, "Correcting Document Image Warping Based on Regression of Curved Text Lines", *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 589-593, 2003.

[107] Z. Zhang and C.L. Tan, "Straightening Warped Text Lines Using Polynomial Regression", *International Conference on Image Processing (ICIP)*, pp. 977-980, 2002.

[108] Z. Zhang and C.L. Tan, "Recovery of Distorted Document Image from Bound Volumes", *International Conference on Document Analysis and Recognition (ICDAR)*,

pp.429-433, 2001.

[109] Z. Zhang and C.L. Tan, “Restoration of Document Images Scanned from Thick Bound Document”, *International Conference on Image Processing (ICIP)*, pp. 1074-1077, 2001.

[110] Z. Zhang, C.L. Tan, and L.Y. Fan, “Restoration of Curved Document Images through 3D Shape Modeling”, *Computer Vision and Pattern Recognition (CVPR)*, Volume 1, pp. 10-16, 2004.

[111] Z. Zhang, C.L. Tan, and L.Y. Fan, “Estimation of 3D Shape of Warped Document Surface for Image Restoration”, *International Conference on Pattern Recognition (ICPR)*, 2004.

[112] Z. Zhang, C. L. Tan, T. Xia, and L. Zhang, “Restoring Warped Document Images through 3D Shape Modeling”, to be submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, manuscript available at www.comp.nus.edu.sg/~zhangz/PAMI.pdf .

[113] Q. Zhu, M. Payne and V. Riordan, “Edge Linking by a Directional Potential Function (DPF), *Image and Vision Computing*, pp. 59-70, 1996.