A Unified Framework for Document Image Restoration	Li Zhang	2008

A Unified Framework for Document Image Restoration

By

Li Zhang

A Thesis Submitted

For The Degree Of Doctor of Philosophy

at

Department of Computer Science

School of Computing

National University of Singapore

Computing 1, Law Link, Singapore 117590

August, 2008

Name: Li Zhang
Degree: Doctor of Philosophy
Department: Department of Computer Science
Thesis Title: **A Unified Framework for Document Image Restoration**

Abstract:

Document image processing and analysis has been an active research topic in recent years, which includes text detection and extraction, normalization, enhancement, recognition and their related applications. The work described in this thesis focuses on the normalization of various types of document images that display all sorts of distortions including shadings, shadows, background noise, perspective and geometric distortions. In particular, a unified framework is developed which takes in an input image and rectifies all the distortions at one go to produce a final image that facilitates human perception and subsequent document image analysis tasks. The whole framework consists of three main components: photometric correction, surface shape reconstruction, and geometric correction. The first component is designed to address distortions including shadings, shadows and background noise through an inpainting-based procedure. The second component is meant to derive the 3D geometry of the document for the succeeding perspective and geometric correction tasks. It comprises of two Shape-from-Shading methods with different solving schemes for the image irradiance equation formulated under various illumination conditions. Finally, the last component is targeted at perspective and geometric distortions with three proposed methods handling different types of images by utilizing different sets of input information. Results on synthetic and real document images demonstrate that each type of the distortions can be effectively corrected using a full or sub set of the procedures in the whole framework. OCR results on the restored images of those text-dominant documents also show great improvements over the original distorted images.

Keywords: Document Image Restoration, Inpainting, RBF-based Smoothing, Shape-from-Shading, Surface Interpolation, Physically-based Modeling.

ACKNOWLEDGMENT

I would like to express my deep and sincere gratitude to all those people who have offered their ingenious ideas and invaluable support continuously throughout this research work. This thesis would not have been possible without their generous contributions in one way or another.

I am deeply grateful to my supervisor, Professor Chew Lim Tan in School of Computing, National University of Singapore, for his valuable supervision and guidance along the way from the topic selection to the completion of this thesis. His wide knowledge and constructive advice have inspired me with various ideas to tackle the difficulties and attempt new directions. He has also been very supportive in purchasing softwares, hardwares and experiment materials used in this research. His kind guidance and support have been of great value to me.

I wish to express my warm and sincere thanks to Dr. Andy Yip in the Department of Mathematics, National University of Singapore, who has kindly shared with me his great

expertise in solving partial differential equations and his knowledge in digital inpainting and surface fitting techniques. His detailed and constructive suggestions have helped me greatly in improving several papers towards their final publication.

I wish to thank Dr. Yu Zhang, for his insightful advice and comprehensive comments on the work about physically-based modeling. His expertise in computer graphics simulations and modeling has enlightened me several evaluation strategies to better demonstrate the effectiveness of the proposed method.

I wish to express my deep appreciation to Dr. Michael S. Brown currently with School of Computing, National University of Singapore, for his generous help in conducting comparative experiments for us using their existing work on the physically-based restoration model, and for his constructive suggestions and efforts in improving our paper on the unified restoration framework recently submitted to PAMI.

I also wish to thank Dr. Ping-Sing Tsai with the Department of Computer Science, The University of Texas - Pan American, for generously sharing with me his code and sample images on their existing Shape-from-Shading work to support our comparative experiments.

I owe my sincere gratitude to Professor Kankanhalli Mohan, Dr. Kok Lim Low and Dr. Zhiyong Huang in School of Computing, National University of Singapore, for their detailed reviews, constructive comments and suggestions to my graduate research paper and thesis proposal during the whole research program.

I wish to extend my warmest thanks to all those colleagues and friends who have helped me and encouraged me in one way or another during my research study in the

Center for Information Mining and Extraction (CHIME) lab of School of Computing, National University of Singapore.

Last but not least, I wish to express my special gratitude to my loving parents, for their continuous support and understanding throughout my undergraduate and postgraduate studies abroad for all these years.

TABLE OF CONTENTS

SUMMARY

Document imaging is a fundamental application of computer vision and image processing. The ability to image printed documents has contributed greatly to the creation of vast digital collections now available from libraries and publishers. While traditional document imaging has been performed using flatbed scanning devices, a trend towards more flexible camera-based imaging is also emerging especially when modern imaging devices such as point-and-shoot cameras, cell phones and PDAs are highly mobile, low priced and easy to use. The large volume of scanned and camera-based document images has called for highly effective image processing and analysis techniques to facilitate machine recognition and interpretation tasks. Document image processing and analysis has thus attracted many attentions in recent years, which include text detection and extraction, normalization, enhancement, recognition and their applications. The challenges of the complex layout and presentation style, uncontrolled noise and distortions, and diversified document content has kept this area an active research field aiming to provide more effective and practical solutions to current document image applications.

The work described in this thesis focuses on the normalization of the various types of document images that display all sorts of distortions including shadings, shadows, background noise, perspective and geometric distortions. In particular, a unified framework is developed which takes in an input image and rectifies all the distortions at one go to produce a final image that facilitates human perception and subsequent document image analysis tasks. The whole framework consists of three main components: photometric correction, surface shape reconstruction, and geometric correction. The first component is designed to address distortions related to photometric artifacts such as shadings, shadows and background noise. An inpainting-based procedure is developed to reconstruct the background layer image and then extract the foreground reflectance image from the original intensity image based on the notion of intrinsic images. In addition, if the image contains merely smooth shadings without other background noise or shadows, an RBF-based smoothing technique can be applied to extract a smooth shading image which can then be used as the input to the surface shape reconstruction component. This second component is meant to derive the 3D geometry of the document so as to obtain an accurate representation of the physical warping for subsequent restoration procedures when geometric or perspective distortions exist. Two Shape-from-Shading methods are proposed to compute the surface depth map by solving the image irradiance equation formulated under different illumination conditions. In particular, the first method solves the image irradiance equation using an iterative Fast Marching scheme with a time complexity of $O(N \log N)$, while the second method formulates the image irradiance equation as a Hamilton-Jacobi equation and solves it using a fast sweeping strategy with a time complexity of $O(N)$. Finally, the last component is targeted at geometric and perspective distortions that often appear in images of non-planar documents. Three methods are proposed, in which

the first method handles only smooth geometric warpings in camera-captured images based on purely 2D text line interpolation while the last two methods can deal with both geometric and perspective distortions given the 3D geometry of the document surface. Typically, the second method is based on a surface interpolation methodology while the last method employs a physically-based modeling technique. Results on synthetic and real document images demonstrate that each type of the distortions can be effectively corrected using a full or sub set of the procedures in the whole framework. OCR results on some restored images of text-dominant documents also show great improvements over the original distorted images.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

The fast advancing technologies in this digital age have resulted in more and more information being captured or generated in electronic forms which can be easily made available on-line through digital libraries or other web resources. However, there is still a large collection of documents that are originally printed or handwritten in undigitized physical media and thus relatively difficult to find and access [Bai03]. This is especially true for many historical documents that are either out of print, deteriorated, or sealed in archives for preservation. Current digital libraries are thus urged to integrate these resources into

the large on-line database that is searchable, browsable, readable or even editable by people around the world. As a result, document digitization is playing an important role in the advancement of current digital libraries. Generally speaking, a complete digitization cycle consists of three phases: imaging phase, recognition phase and content recovery phase [DWPL04]. In the first phase, physical document pages are converted into electronic images using scanners or digital cameras. At this phase, computers are hardly aware of the content of the document, which makes indexing a difficult task. The second phase is to recognize the content, which can be further divided into three steps. The first step consists of a set of pre-processing routines that are designed to remove noise and correct various distortions that may affect the subsequent document image analysis (DIA) tasks. Some of these routines are also known as normalization procedures. The second step is to detect and extract the text regions by analyzing the layout of the document in terms of segmented text lines or text blocks. The third step is to feed the extracted text regions into the OCR engine for recognition. Some image enhancement procedures may be applied prior to the recognition step to further improve the image quality when necessary. After the recognition phase, text information is available for reading and editing purposes. However, the plain text does not contain any style or structural information. Therefore, the last phase is to recover the full logical and physical structure of the document so that it can be easily republished in other formats such as PDF, HTML, XML, etc., for easy preservation and dissemination. Thus, the original physical document is completely converted to its electronic form, which can be easily accessed from the large on-line database by a wide range of audience. Nevertheless, a link between the reconstructed electronic document and its image form may still be retained for further verification purposes. Figure 1.1 shows a typical work flow of the document digitization process.

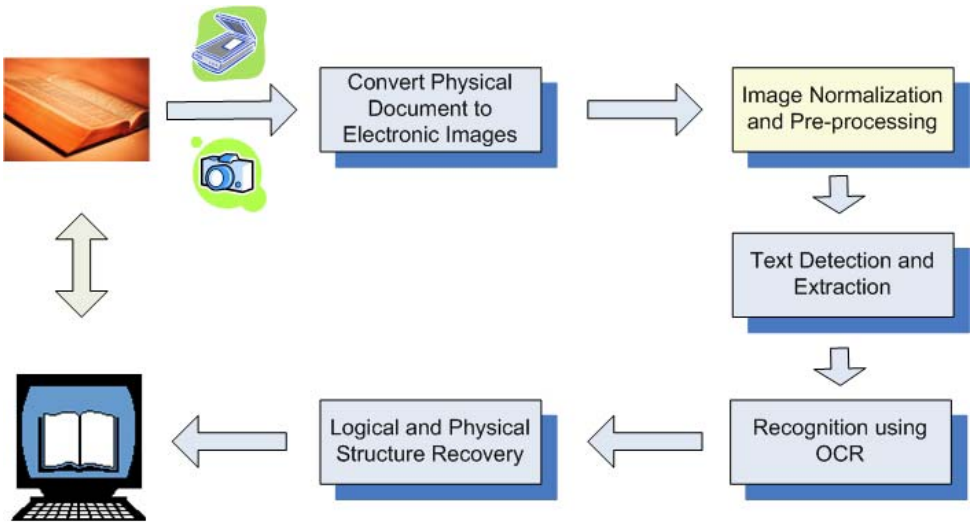


Figure 1.1: A typical work flow of the document digitization process.

The traditional way of converting physical documents to their electronic form is through flatbed scanners. A flatbed scanner is usually composed of a glass pane, an underneath bright light across the pane and a moving optical array being either a charge-coupled device (CCD) or contact image sensor (CIS). During scanning, documents are placed down-facing the scanning plane, and the sensor array and light source move through the pane to capture the entire area. However, due to the constraint of the imaging model, many non-planar materials such as rolled scripts, folded papers, etc, often need to be manually pushed to the glass pane in order to produce a quality output. Some fragile materials are thus easily damaged or destroyed in the scanning process and hence results in a permanent loss of information. Therefore, flatbed scanning is often undesirable in handling delicate historical materials that are easily broken under external forces. On the other hand, some thick bound documents can hardly be pressed down to the scanning plane especially near the spine region. This results in images with both geometric and shading distortions. The geometric distortion here refers to the distortion caused by the

non-planar geometric shape of the document being scanned. The shading distortion refers to the shading artifacts mainly at the spine region caused by a non-uniform illumination due to the shape of the spine. Other distortions such as smear, stains or background noise are also common problems in images of historical and ancient document collections. A large-scale test of current commercial OCR systems [RNN99] has demonstrated that the accuracy of current OCR devices falls abruptly when the image contains certain defects such as heavy and light print, stray marks, skew or curved baselines, shaded background, etc. In order to achieve a good OCR performance, pre-processing steps are thus needed to correct the distortions prior to the recognition phase. In response to this, many restoration methods have been developed over the years to address the distortions in scanned document images, which include various binarization methods, skew detection and correction methods, warping and shading correction methods, etc.

On the other hand, the increasing popularity and the high resolution property of current digital cameras have made camera imaging a new trend for digitizing paper documents especially those deteriorated ancient manuscripts. For example, there are collections of historical documents in our National Archive that are difficult to handle using flatbed scanners because the paper is so fragile that it will easily crack at the spine region when being pushed to the scanning pane. This is undesirable because such historical documents are usually invaluable and need to be well preserved to maintain their integrity for future studies. Current digital cameras realize this objective effectively with their non-contact imaging property, which captures the image from a distance without physically touching the manuscript. This is very document-friendly and is highly appreciated by the archivists across the world. Meanwhile, the quality of the image is retained with the high resolution property of the cameras, which allows details of the content being captured with great fi-

delity. Despite the several advantages in using digital cameras for document digitization, some common distortions still remain as in the scanned images such as geometric and shading distortions, background noise, etc. Some of these distortions may even become more complex if the imaging environment is not carefully controlled. Further investigations are thus needed to improve existing restoration techniques on scanned images and explore new methods that are specially tailored to camera-based document images.

Apart from digitizing historical documents, camera imaging also provides an easy way of recording daily information with its simple yet powerful snapshot functionality. This is especially true with the fast emerging hand-held digital imaging devices such as cell phones, PDAs, point-and-shoot cameras, etc. The high portability, versatile functionality and low pricing properties of such devices have promoted their use as personal photocopiers that can be carried everywhere with ease [DLL03]. It is now easy and highly tempting to take a picture of a presenter's slide during a conference than to jot down bits and pieces of the content using a pen or through a computer keyboard. People are also using their pocket cameras to capture interesting advertisements or newly-bought books to share with their friends. Another interesting example is the camera-based business card reader in some recent cell phone models. The idea is to capture an image of the business card using the phone camera and then recognize the content through an underlying OCR engine. This enables a direct transfer of the card information into the phone's address book without much manual effort. However, all these applications are useful only when the captured images are of good quality. Any distortions in the image can possibly cause inaccurate recognitions or even result in a system failure. For example, the business card reader will not allow the user to take a skew or perspectively distorted image and will often fail to recognize certain content when the image contains shadows due to some

occlusions of the light source. Generally speaking, images with less distortions are more pleasing to the human eyes and are also more acceptable by the machines. Most of the state-of-the-art document analysis techniques or softwares can only produce good results on images with high resolution, high quality and simple structure [DLL03]. To this extent, restoration procedures are desired to correct or remove the distortions and produce a qualitative frontal-parallel view of the document page for easy human perception and better machine recognition.

1.2 Contributions

In view of the large amount of scanned images in current digital libraries and the increasing number of camera-based document images produced in daily imaging activities, there is a great need to make these images more easily machine readable and accessible regardless of the distortions that may affect the traditional DIA tasks. Our objective is therefore to design and develop new restoration methods that can effectively correct various distortions in a wide range of document images and produce a flat rendering of the document to improve human illegibility and facilitate subsequent analysis and recognition tasks such as segmentation, classification and OCR.

Distortions in document images can be divided into two main categories: 1) distortions related to the imaging environment and the document property, such as background noise due to stains or material degradations, shadings and shadows due to variable illumination conditions, geometric distortions due to non-planar surface shapes, etc; 2) distortions related to the imaging device, such as sensor noise, lens distortions, quantization errors, etc. In this thesis, we mainly focus on those distortions in the first category. As mentioned

earlier, traditional flatbed scanners have usually constrained the imaging environment with their fixed hardware layout. In particular, the light source is fixed and the distance from the sensor to the imaging plane is also fixed. Therefore, the distortions related to these settings are relatively easier to model compared to those in camera-based images. For example, in a camera-based imaging environment, a document could be arbitrarily warped instead of being constrained to adhere to the scanning plane. This brings in more variations to the geometric distortions which are unable to be simply modeled as special shapes such as cylindrical surfaces as in the scanner-based case. On the other hand, the uncontrolled lighting environment also makes the shading distortions more complex unlike the uniform illumination in most flatbed scanners. Moreover, due to the perspective projection property of most cameras, perspective distortions may also appear in camera-based images which are uncommon in scanned images. Nevertheless, almost all the distortions are common to both types of images except perspective distortions. To summarize, we look at the following three types of distortions in general:

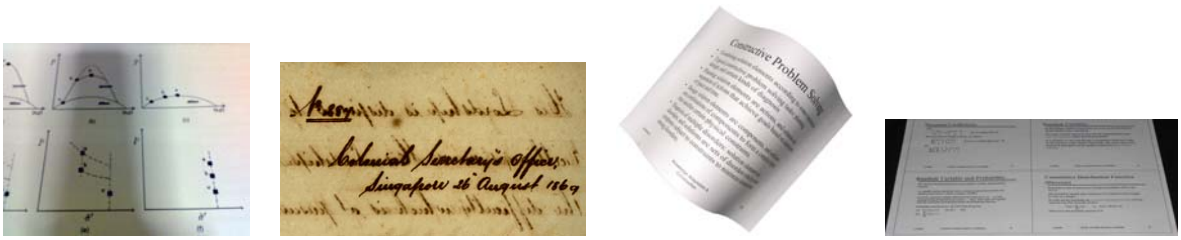
- **Shading distortions**. Shadings are often caused by changes of the surface normal with respect to the lighting direction. In scanned images, they are therefore mainly caused by the non-planar shape of the document since the light source is fixed. However, cameras usually have less control on the lighting conditions than scanners. A typical example is when imaging a document with an on-camera flash, the image will appear bright near the center of the view and gradually fade away toward the corners. Moreover, casting shadows may also occur when the light source is occluded by other objects and therefore results in large intensity variations at certain portions of the image. Such distortions will cause significant errors in the OCR process. Generally speaking, shadings are affected by four factors including illumination,

surface reflection property, surface shape and image projection model. A common assumption for shadings present in the document image domain is that they are produced under a point light source on a Lambertion surface. The image projection model is often taken as perspective projection. On the other hand, a shadow's sharpness depends on the shape of the light source, the reflection properties of the surface, the geometry and opacity of the occluder and the interreflections among the objects in the scene. In particular, a shadow can have infinitely sharp edges when it is produced by a point light source and extremely soft edges when it is generated by large area light sources. Here we only consider soft shadows that are possibly blended with other shadings.

- **Background noise**. Background noise in document images usually refers to defects in the original documents which are transfered to the images during digitization - scanning or camera imaging. Here we focus on two types of defects. One type refers to those in historical documents caused by long time preservation under bad environmental conditions such as ink bleed-through, water stains or smudges. The other type refers to the background noise caused by show-through effects in daily document images where the back-side content of a page is visible in the front-side image. Such degradations cause substantial noise in the digitized images, which present great challenges for machine segmentation and recognition tasks or even for human inspection.

- **Geometric and perspective distortions**. When scanning non-planar documents such as thick bound books, it is sometimes hard to press the whole document down to the glass pane and therefore some parts of the documents especially near the spine region may be at different distances from the imaging sensor. This appears

more frequently and more severely in camera-based images because there is no more confinement from the glass pane. Such situations cause various degrees of warping distortions in the images perceived as wavy text lines, distorted characters, etc. The geometric distortions here do not refer to lens distortions such as radial distortions, barrel distortions, etc. Instead, they refer to the distortions caused by the non-planar geometric shape of the document being imaged. Moreover, they are often accompanied by shading distortions since the surface normal certainly varies with respect to the illumination direction. On the other hand, perspective distortions can also be treated as one special kind of geometric distortions in which the document surface is not parallel but slanted with respect to the image plane. All these distortions not only impede the legibility of the image, but also affect many document analysis applications which are designed to obtain high-level semantics from frontal-parallel images. However, unlike the simple linear shape distortions such as translation, scaling and rotation, these non-linear shape distortions cannot be easily described using general mathematical models. The restoration of such distortions is therefore more difficult and remains as one of the many challenging problems in the computer vision domain.



(a) (b) (c) (d)

Figure 1.2: Document images with different types of distortions: (a) shadows; (b) background noise; (c) geometric and shading distortions; (d) perspective and shading distortions.

Besides the large varieties of distortions, there are different types of document images that need to be considered in terms of their content. In general, two categories of documents can be defined: textual documents and graphical documents. Textual documents contain mainly paragraphs of text with regular layout, which may include figures, tables or equations but those should only occupy a small portion of the whole document. Graphical documents, on the other hand, contain mainly figures or pictures, which can be line drawings, work flow diagrams, tables, etc. The reason that we classify the documents based on their content is that sometimes a method works on one type of documents but not the other, while some methods deal with various types of documents independently of their content. Nevertheless, methods that are tailored to a specific document type may be able to utilize certain properties of the document that are highly correlated to the distortions being studied to produce good restoration results efficiently. For example, if we know a given document is a textual document with regular text lines and layout, we can use the text lines to estimate the warping distortions and rectify the warpings accordingly. This method may be effective for textual documents, but it is not applicable to graphical documents in which no text lines can be found.

Due to the wide range of distortions and the diversified document content, it is difficult, if not impossible, to develop a single restoration technique that can deal with all kinds of distortions and all types of document content effectively. Therefore, in this thesis, we have explored and designed various methods and techniques, which are targeted at different types of distortions and document content. Some methods are general enough to handle several types of distortions with certain common properties. Some are able to deal with various types of documents regardless of their content. On top of all these uncorrelated methods and techniques, we propose a unified framework that integrates all the methods

together and groups them into three interconnected components with each component addressing certain distortions or deriving certain intermediate information for subsequent processes. Several methods and techniques may belong to the same component and each one may require different types of input data depending on the content of the document. The ultimate goal of this framework is to take in a distorted image with various distortions and output a final restored image with all the distortions rectified. Figure 1.3 gives an overview of the entire restoration framework.

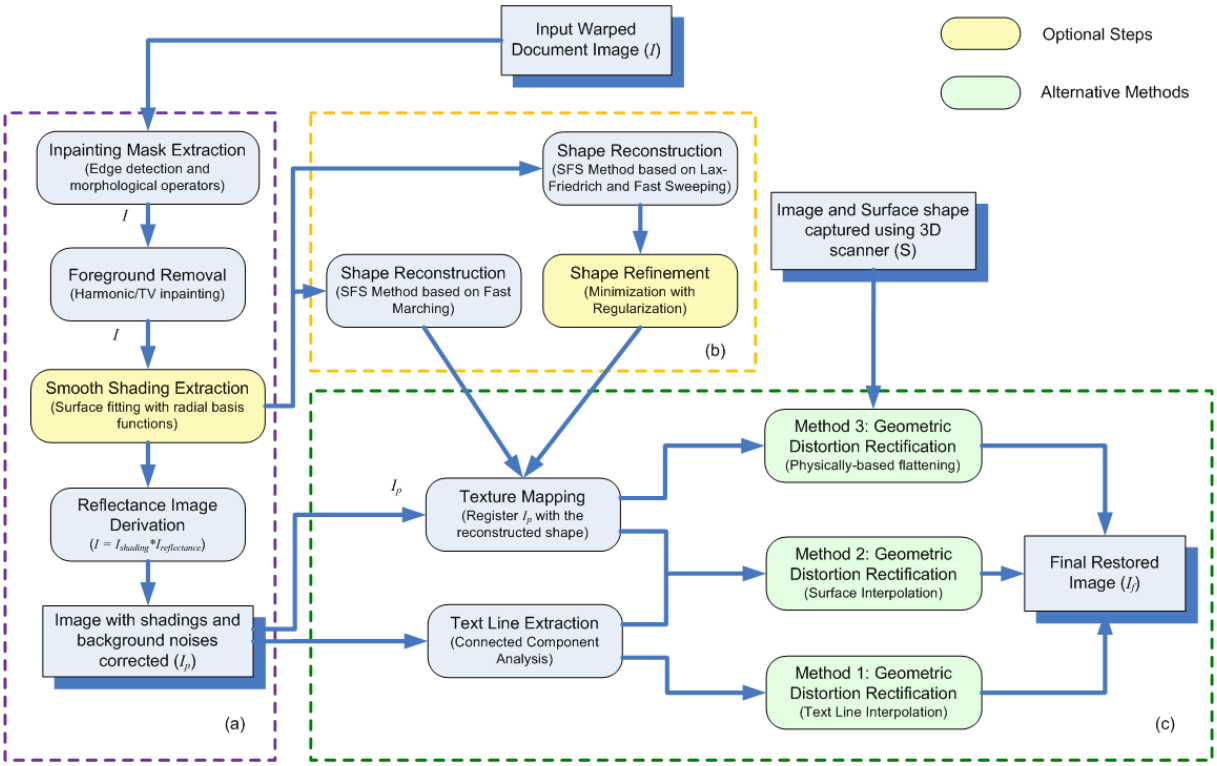


Figure 1.3: An overview of the restoration framework: (a) Photometric correction; (b) Surface shape reconstruction; (c) Geometric and perspective correction.

1.3 Structure of the Thesis

The remaining chapters of this thesis are organized to capture the three components of the entire framework including the photometric correction component, surface shape reconstruction component, geometric correction component. In particular, each of the component may consist of several methods either targeting at different sets of document images or employing different techniques and strategies. Finally, a separate chapter is dedicated to illustrate the functionality of the whole framework by building a typical assembly with a set of selected methods one from each component and evaluating their performance on real document images with different types of distortions. Below is a road map of the remaining chapters of this thesis.

Chapter 2 provides a review of the existing techniques and methods for document image restoration tasks including those for scanned images and camera-based images. In particular, we investigate the existing methods and approaches proposed to address each type of the distortions mentioned in Section 1.2. First, we review the two types of commonly-used approaches for correcting shading distortions in document images, namely binarization methods based on thresholding and restoration methods based on the surface shape, respectively. Second, we look at various methods designed to deal with background degradations in historical documents and analyze their performance and applicability in real archive digitization tasks. Last but not least, we study the large varieties of geometric correction methods in the document image domain and discuss the advantages and disadvantages of each type of methods as well as their relationships with image processing, computer graphics and computer vision.

Chapter 3 describes the photometric correction component of the framework, which mainly focuses on a generic method designed to handle several types of distortions in both scanned and camera-based document images including shadings, shadows and background noise as discussed in Section 1.2. Enlightened by the common properties of the three types of distortions, we have designed a series of procedures that first reconstruct the background layer image using an inpainting technique and then extract the foreground reflectance image based on the notion of intrinsic images. The idea is to treat all the unwanted pixels as background noise and remove them from the original image. In addition, an optional step can be applied to construct a smooth shading image based on a smoothing technique if the given image only contains smooth shading distortions. This shading image can be further utilized in the subsequent geometric and perspective correction phase when needed. Comprehensive experiments have been conducted to demonstrate how well the proposed method is able to correct all three types of distortions and in which cases the method fails to perform.

Chapter 4 presents two Shape-from-Shading methods designed to reconstruct the surface shape of the document when the document being scanned or captured has a non-planar surface. This is essentially what the surface shape reconstruction component is for, which is to obtain the surface shape of the non-planar document in preparation for the geometric and perspective correction phase because many geometric correction methods require the knowledge of the surface shape in order to produce a good restoration result. One way to get the surface shape is to capture it using special setups such as structured lighting or range scanners, which is simple and accurate but the cost is usually high. Another way as discussed here is to reconstruct the shape using shape recovery methods based on one or more 2D images. In this thesis, we propose two shape recovery methods

that are both based on the Shape-from-Shading methodology, which aims to recover the shape of the object based on the shading variations in a single 2D image. The difference of the two methods is mainly on the scheme used to solve the image irradiance equation formulated under different illumination conditions and with different projection models. Experiments on synthetic data are conducted by first generating synthetic shading images from functional geometric surfaces and then reconstructing the shape based on the synthetic shading image by solving the image irradiance equation using numerical methods. The results have demonstrated the great performance of the two proposed methods by comparing them with the ground truth shape and those from many existing approaches.

Chapter 5 focuses on the geometric correction component, which consists of three different methods designed to address either geometric distortions or both geometric and perspective distortions depending on the type of the document and the information that is obtainable from the input image such as text lines, shadings, etc. Typically, the first method is a purely 2D-based method that is tailored to camera-based textual document images with simple and regular warping distortions. This method uses existing text lines or document boundaries to estimate the physical warpings and rectify them through an interpolation procedure. The advantage of this method is its simplicity and efficiency. However, due to the inaccurate estimation of the 3D warpings using solely 2D information, the characters may not be fully rectified to their normal size at the warping regions. An improved version of this method is thus proposed to further address this problem by incorporating 3D information obtained from the surface shape reconstruction phase into the 2D text lines extracted from the input image. A subsequent surface interpolation technique is then applied to derive the final restored image. Finally, the third method is a content-independent method that works on both textual documents and graphical

documents which can be either scanned or camera-captured. The only input data required is a known surface shape of the document and its 2D image. Once the surface shape is available, we can model it as a particle system and flatten it to a plane through numerical simulation techniques. Since this component is usually executed after the photometric correction component, the shadings in the original image should have been corrected at a previous stage. By using the photometrically-corrected image as the texture to the surface, we obtain the final restored image after the surface is flattened. This physically-based restoration method is able to handle both geometric and perspective distortions because perspective distortions can essentially be treated as special geometric distortions with a slanted planar surface. Compared to the first two methods which are more specifically designed for textual documents with regular warping distortions, the last method is more general and suitable for a large set of images with different types of content. Experimental results are shown for each of the methods and their performances are evaluated based on the OCR accuracy on the restored images.

Chapter 6 demonstrates the functionality and performance of the entire framework by selecting a sub set of methods and techniques from each component with a focus on daily snapshots of documents with shadings, perspective and geometric distortions. Several examples with different combinations of distortions have been used to demonstrate the intermediate result of each restoration step. The final restored images are also evaluated against the OCR metric which shows significant improvements compared to the original distorted images. Finally, we look at the framework as a whole and discuss its several advantages in relation to each component.

Chapter 7 summarizes the contributions of this thesis and outlines some possible directions for future research and investigation.

CHAPTER 2

RELATED WORK

As mentioned in Chapter 1, there are several types of distortions in scanned and camera-based document images which are related to the imaging environment and the property of the document. In particular, we look at three types of distortions that are commonly seen in images of ancient manuscripts and modern documents, which include shading distortions, background noise, geometric and perspective distortions. Over the years, many approaches have been proposed to address each type of the distortions to a certain extent with different focuses or using different methodologies. In this Chapter, we give a brief review of the existing work in each of the three distortions domain including methods for either scanned images or camera-based images.

This chapter is organized as follows. Section 2.1 reviews the history of shading correction in both scanned and camera-based document images. Section 2.2 reports the several categories of methods for restoring historical documents with background degradations such as ink bleed-through, smears or water stains. Finally, Section 2.3 analyzes a series of geometric and perspective correction methods from 2D-based geometric correction to 3D-based geometric correction to pure perspective correction and depicts the relationship between geometric correction and various image processing related domains.

2.1 Shading Correction

When scanning thick bound documents with deep spines, we often observe dark shadings near the spine region due to its larger distance from the light source than the rest. Such shadings result in a low contrast between the text and the background and thus makes OCR difficult to score. On the other hand, camera-based images are more susceptible to shading distortions due to the intricate and uncontrolled light sources in the natural imaging environment. In the real world, we have daylight and various electric lights. In computer graphics world, we have as many as Distant Light, Point Light, Spotlight, Area Light and Linear Light. Sometimes more than one light source can appear in the imaging process. Apart from the many types of light sources, the location of the light source is also a crucial factor in the shading formation process. In addition, the property of the surface material is what determines how the incident light is reflected and therefore also affects the shading. All these factors make the shading distortions rather complicated in camera-based images. Besides shadings caused by uneven illumination which is also referred to as self-shadows, cast-shadow is another type of shading distortions which is

usually caused by the occlusion of the light source. When such shadows are casted on the documents, they will cause large shading variations in the captured image and therefore affect the recognition of the content.

Shading distortions in document images are mostly addressed by various binarization methods, which aim to depict the object and the background separately using a bi-level representation in order to reduce storage cost, remove noise and simplify content analysis tasks. Another small stream is to correct shading distortions based on the shape information. This set of approaches mainly focus on shadings caused by the non-planar geometric shape of the document such as shadings in scanned images of a thick bound book.

2.1.1 Binarization-Based Methods

Most of the binarization methods are based on thresholding techniques, which try to classify the image pixels into the object class and the background class respectively based on one or more threshold values. Based on whether the threshold value is obtained globally or locally, these methods can be categorized as global thresholding method, local adaptive thresholding method or hybrid method. In particular, local adaptive thresholding methods have great advantages in dealing with shading distortions caused by uneven illumination as reported in many literatures. Most of these methods calculate the threshold at each pixel based on the local statistics information in the pixel's neighborhood such as mean, variance, etc.

One typical example is Niblack's local adaptive thresholding algorithm [Nib86], which estimates the local threshold for each pixel based on the pixel mean $m(x, y)$ and standard

deviation $s(x, y)$ within a small neighborhood using the formula:

$$T(x, y) = m(x, y) + ks(x, y) \tag{2.1}$$

where $k \in (-1, 0)$ is the variance gains which can be adjusted to control the performance of the algorithm. Meanwhile, the size of the neighborhood window should be small enough to maintain local details, and also large enough to suppress noise. Niblack's algorithm has been rated as one of the best-performed adaptive thresholding algorithms [TT95] notwithstanding the efficiency issue due to the computation of mean and variance for each image pixel and the sensitivity of the algorithm in relation to the manually adjusted parameters w for window size and k for variance gains. As an improved version of Niblack's method, Sauvola and Pietaksinen [SP00] propose to modify Niblack's formula as in Eq. 2.1 to:

$$T(x, y) = m(x, y) \left[1 + k \left(\frac{s(x, y)}{R} - 1\right)\right] \tag{2.2}$$

where R refers to the dynamic range of the standard deviation and $k \in (0, 1)$ is a user-defined parameter. This new formula is designed to suppress background noise and works well on badly illuminated documents or historical manuscripts with stains and ink bleed-through problems. However, it has an additional parameter R that needs to be tuned and requires the knowledge of the image contrast in order to choose an appropriate value.

White and Rohrer [WR83] have proposed another local thresholding method specially for document images with background noise. This method uses a dynamic threshold to classify each pixel p into either foreground or background. The dynamic threshold is computed as the average gray value of the pixels within a neighborhood of each pixel p. The size of the neighborhood is chosen to be approximately equal to the size of the characters in the document. In addition, the gray value of each pixel p is adjusted

with a biasing function before comparing with the neighborhood average, which helps to suppress noise caused by unwanted patterns smaller than the character size. In a similar dynamic method proposed by Bernsen [Ber86], the threshold is defined as the mean of the minimum and maximum values of the pixels in a neighborhood. Both methods determine the threshold based on the local contrast of a neighborhood window and the performance is therefore subject to the window size chosen for each document image.

Another method that has been shown to perform equally well [TT95] as Niblack's method is the one proposed by Yanowitz and Bruckstein [YB89], which constructs a threshold surface by interpolation over a set of object boundaries using a successive over-relaxation algorithm. The object boundaries are identified as pixels with high image gradient which usually indicate object edges. However, due to the iterative interpolation scheme, this method is not runtime efficient. A more efficient method has recently been proposed to compute the adaptive threshold surface by Blayvas *et al.* [IB06], which derives a smooth and continuous approximation of the threshold surface by summing up a set of scaled and shifted functions of the interpolated surface based on extracted object boundaries. However, the experiments have shown that this method does not work well on textual documents with texts in various font sizes.

Most recently, a global thresholding technique [LT07] has been proposed to binarize badly illuminated document images with smooth shading variations. The method estimates the global shading variation as a least square polynomial surface using a two-dimensional Savitzky-Golay filter. This surface is then used to produce a compensated image with roughly uniform illumination. Finally, a global threshold can be easily chosen to binarize the compensated image. This method is specially designed for images with smooth shadings which can be modeled using polynomials and therefore does not work

well on images with irregular non-smooth shadings.

2.1.2 Shape-Based Methods

As mentioned in Chapter 1, shading distortions often accompany geometric distortions and appear in images of non-planar shaped documents. If the shadings are purely caused by the variations of the surface normal with respect to the illumination direction, intuitively they should be able to be addressed together with the geometric distortions based on the available shape information. Some attempts have been made in this direction to correct shadings in scanned images of thick bound documents or warped camera-captured images.

Kanungo *et al.* [KHP93] is the first that proposes to remove the dark shadings near the book spine regions in scanned images of thick bound documents using a distance-based method. Due to the fixed light beam of the flatbed scanners, it is reasonable to assume that the illumination at a point P on the document surface is inversely proportional to the distance between the point P and the light source L. In addition, with the assumption of a diffuse lighting model, only the curved portion needs to be restored based on the estimated cylindrical shape of the spine.

Subsequently, Zhang *et al.* [ZZTX05, TZZX06] has presented a method that uses the image irradiance equation on Lambertian surfaces to associate the shading variations to the surface shape. The method first assumes a constant albedo across the whole document to reconstruct the 3D surface geometry. Then, it re-computes the correct albedo based on the estimated shape with the assumption that each column gets a uniform illumination. Iketani *et al.* [ISI06] has also used a similar method to correct the shadings in the

final image mosaicked from a series of video-captured document images after the surface shape is recovered. This method assumes a parallel lighting during the imaging process. Therefore, it is similar to the uniform light beam in normal scanners.

For camera-based document images, Tsoi and Brown [TB04, BT06] have introduced a shading correction method based on boundary interpolation. In particular, the document being imaged is assumed to have a white margin near the boundaries. This guarantees that a perfect boundary curve can be identified to approximate the shape of the document. During the restoration process, the internal pixel intensities are then computed by interpolating the boundary pixel intensities using Coons patch interpolation. However, this implies that the shading is uniform along the columns, which may not be true in certain situations. For example, with a close point light source like a camera's flash, the image will appear much brighter in the center than the corners.

Sun *et al.* [SYY+05] have also proposed a method that corrects non-smooth shadings in camera-based document images by classifying the intensity changes into two classes - illumination change and reflectance change. The method assumes that any noticeable visual discontinuity such as an edge is caused by the reflectance or illumination and the two types of edges do not overlap. First, the illumination edges are identified based on the captured 3D geometry of the document surface. A restoration procedure is then applied to correct the shadings by integrating the gradient field. This is followed by the removal of the overlapping reflectance edges to produce the final result. This method focuses on images with obvious edges such as folds in any directions. The content independence property makes it widely applicable to any types of distorted images.

2.1.3 Discussion

For binarization-based methods, global thresholding based methods are usually efficient although they do not work well on irregular shadings when the shadings do not fit into regular distributions as shown in [LT07]. A more effective technique for such irregular shadings are local adaptive thresholding which derives thresholds based on local information. However, one disadvantage of this technique is the high computational cost due to the iterative local computations especially when the window size is small. Some methods also have certain limitations such as manually adjusted window sizes and variance gains or known contrast levels, etc.

Shape-based methods are relatively new and less-widely studied. The main idea is to associate the shadings with the surface shape and restore it either based on shape-from-shading, interpolation or notion of intrinsic images. If the surface shape can be accurately captured or derived, the shading correction part is usually straightforward and shown to be effective. However, it is often hard to obtain an accurate representation of the surface shape especially for daily imaging applications where only a single input image is available. The shape dependence property of this set of methods thus becomes the main concern for its application to daily image restoration applications.

Generally speaking, binarization-based methods are targeted at a general set of document images with shading distortions, while the shape-based methods are more focusing on shadings in physically warped document images which require the presence of a warping in order to obtain the necessary shape information for estimating the shading. Therefore, binarization-based methods are in a sense more widely applicable than shape-based methods.

2.2 Background Noise Removal

As mentioned in Chapter 1, image degradations in historical documents appear in various forms, ink bleed-through due to ink's seeping through the document pages in a humid environment, water stains or other existing damages, etc. When such degradations are transfered to the digitized image, they will certainly produce significant background noise and affect further document analysis procedures. Moreover, show-through effects in modern printed document images of double-sided pages also present similar problems as ink bleed-through although they are normally less severe. Researchers in the document image analysis community have tried various ways to tackle this problem along different directions.

2.2.1 Thresholding-Based Methods

One stream uses binarization methods based on thresholding techniques. For example, Dubois and Pathak [ED01] design a method to solve the ink bleed-through problem by first registering two side images of a page and then identifying the ink bleed-through pixels using a thresholding technique to replace it with the background intensity.

Sharma [Sha01] proposes a method to deal with the show-through effects in scanned images of duplex printed document pages. First, a linear model is derived to capture the reflectance properties of both the front-side and the back-side images of a page with certain assumptions of the scanning process. Next, by using an adaptive filter to estimate the Point Spread Function (PSF) of the show-through effects, the corrected pixel intensities can be computed based on the front-side scan density and the back-side scan absorptance

adaptively. This approach requires the front-side and the back-side images of the duplex printed page to be well-aligned in order to produce a good result.

Nishida and Suzuki [NS02] describe another approach to address the show-through problem in scanned color document images. In particular, it first separates the foreground components from the background components including the show-through using a local adaptive binarization method for each color component followed by edge thresholding. Then, it estimates the background colors through a local color thresholding technique to obtain an initial restored image. Finally, this restored image is compared with the original image in terms of the edge distributions and corrected adaptively through a multi-scale analysis approach.

Leedham *et al.* have investigated three global thresholding algorithms and a multi-stage thresholding algorithm to separate text from background in degraded document images. They have concluded that the given global algorithms do not work well with difficult documents due to over-thresholding, while the multi-stage algorithm can do a better job by incrementally removing the noise [LVPG02].

More recently, Bar-Yosef *et al.* propose a multi-stage binarization method to restore and recognize ancient Hebrew calligraphy documents [BYBKD07]. The method consists of a global binarization step, which gives an initial binary image mainly for noise free characters, and a subsequent refinement step, which starts with a part of a character as seed and expands the character to its final form through a growing process. The method has demonstrated good classification performance using the restored images, but yet to be applied to a large dataset.

2.2.2 Other Methods

Besides thresholding-based methods, another set of methods is to separate the image into different layers using classification or clustering algorithms. For example, Drira proposes to solve the bleed-through problem by classifying the pixels of an image to background, original text and interfering text, and then replacing the last class with an average background value [Dri06].

Boussellaa *et al.* use a hybrid method to separate foreground from background in Arabic historical documents [BZA07]. In particular, they first use a normalization algorithm to roughly separate the foreground and the background images. Then, after applying gamma correction and contrast adjustment to the foreground image, they obtain an enhanced foreground image which is finally segmented to foreground and background based on k-means clustering with a maximum likelihood estimation. Their results on a set of Arabic historical manuscripts with bleed-through effects, uneven background color and ink spots have shown effective improvements after restoration.

Along another direction, Tan *et al.* [TCS02a] also propose a novel approach to tackle the bleed-through problem using a wavelet-based technique. More specifically, they use a wavelet reconstruction process to iteratively enhance the foreground strokes and dim the bleed-through strokes so that a subsequent edge detection process can distinguish the foreground and the bleed-through strokes more accurately and remove the unwanted strokes. However, this method does require two well-registered images from front and back of the page in order to achieve good results.

Another method [TCS+02b] they propose is to combine an edge detection algorithm with a stroke orientation filter to separate foreground edges from bleed-through edges.

The main problem of this method is on how to produce a good edge detection result when the interfering strokes are strong.

Tonazzini *et al.* [TBS04] also propose to model the bleed-through problem as one type of the blind source separation (BSS) problems, in which the foreground texts and the background texture are unknown and need to be recovered. Similar to [Sha01], a linear model is adopted to represent each color component image as a linear combination of the sources known as background, overwriting and underwriting, respectively. A fast independent component analysis (ICA) algorithm is then employed to solve the linear equations and produce the final corrected image. This method provides promising results only when there is no cross correlation between the sources, which is the assumption of the ICA algorithm.

2.2.3 Discussion

Background noise in images of degraded historical documents has been the attention of many researchers in the document image analysis community especially when it is related to many digitization projects in modern digital libraries or national archives. The methods therefore often focus on certain types of distortions depending on the given documents in the library. For example, in archives of tropical countries such as Singapore, many historical documents have the ink bleed-through problem and some can be disastrous. Several methods above have been specially designed to deal with this type of distortions and some have achieved promising results. However, many are not readily applicable to real restoration systems due to their high computational cost or the large number of parameters to be tuned which makes it difficult to use by librarians or archivists.

Although there have been a variety of methods along different directions, more applicable methods are desirable which will balance the accuracy and efficiency for a wider range of distortions.

2.3 Geometric and Perspective Correction

2.3.1 Image Warping and De-warping

Digital image warping is a growing branch in the image processing domain that deals with geometric transformations of digital images [Wol90]. A geometric transformation is an operation that redefines the spatial relationship between points in an image. The warping may range from something as simple as a translation, rotation, or scale, to something as complicated as a convoluted transformation. The problem of image warping has been the subject of considerable attention in various fields including remote sensing, medical imaging, computer vision and computer graphics for over thirty years. The earliest work can be traced back to its application in correcting camera nonlinearity and geometric distortions of aerial photographs in remote sensing field. In all these work, distortion models are estimated as bivariate polynomials whose coefficients are obtained by minimizing an error function over some reference points. Similar methods can be directly applied to medical imaging and computer vision fields as well, such as image registration and rotation for digital radiology. Other than these, people in the area of graphics design also make use of image warping to create interesting visual effects such as the transformation from ostrich to tiger to woman in Willow. However, in the field of computer graphics, the goal is more on geometric distortion rather than geometric

correction. One such application is known as texture mapping, a technique to map 2D images onto 3D surfaces, and then project them back onto a 2D viewing scene. The net effect is a warp from a 2D texture image to a 2D screen image which is usually like what we see in camera-captured images. Suppose we have an image of a flat document page (2D texture image) which is mapped to a multi-folded surface, now when we take a picture of the folded document page, we will have a 2D camera image that looks warped as illustrated in Figure 2.1. Reversely, if we have the warped 2D camera image, how can we recover the original flat texture image? This is known as the digital image de-warping problem, which tries to correct the warping distortions in digital images caused by a non-planar geometric shape of the object being imaged and return their original flat view for subsequent image processing tasks.

In relation to the distortions observed in camera-based document images, a warped 2D image is considered to have geometric distortions due to the non-planar geometric shape of the document being imaged. For example, when imaging an opened thick bound book, the warped surface shape causes the originally straight text lines to appear warped especially near the spine region. More complex warpings also exist in daily camera imaging applications such as crumpled pages and multi-folded materials. In particular, perspective distortions can also be considered as one special type of geometric distortions because perspective distortions are essentially produced due to the image plane not parallel to the document surface. The process of removing such geometric and perspective distortions can thus be regarded as a digital image de-warping problem. To address this problem, many approaches have been proposed to estimate the distortion parameters and restore a flat rendering of the document page. In the following Sections, we give a brief survey on the existing 2D-based and 3D-based geometric restoration approaches and the commonly-

used perspective restoration methods for document images. In addition, we discuss their pros and cons in terms of the applicability and the performance.

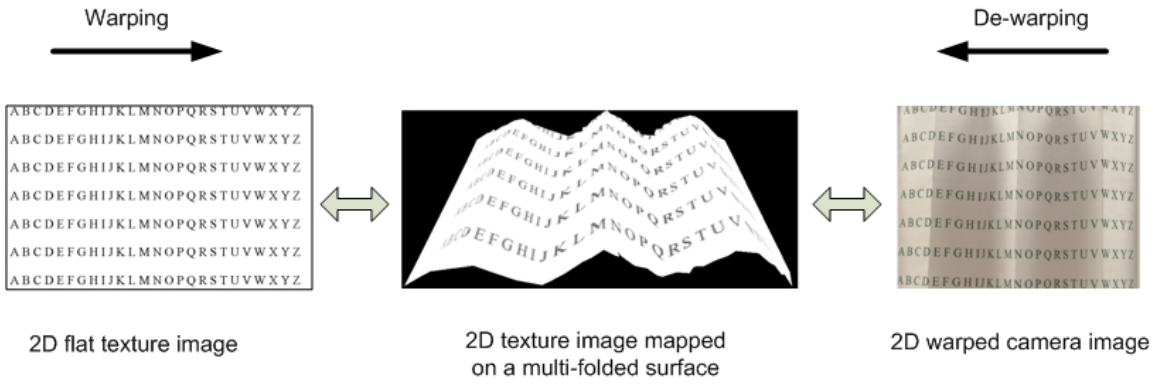


Figure 2.1: Digital image warping and de-warping.

2.3.2 2D-Based Geometric Correction Methods

Most of the earlier de-warping approaches make use of 2D image processing techniques to deal with geometric transformations between the input warped image and the output restored image. Each point in the output image is mapped to a corresponding point in the input image through a spatial transformation function. This function can take different forms depending on the application, such as affine, perspective, bilinear and polynomial transformations. In all cases, the whole process is to take an input image and apply a geometric transformation to produce an output image.

Baird discusses the use of a descriptive document image defect model to approximate document printing and imaging parameters including digitizing resolution, font size, affine spatial transformation, speckle, jitter, blurring, etc [BBY92, Bai93, Bai00]. Further developments on such defect/degradation models are made by Kanungo *et al.*, which handle

both global (perspective and non-linear illumination) and local (speckle, blur, jitter and threshold) defects [KHP93, KHB95, KHB+00]. Such degradation models can help to characterize the output performance of OCR, explicitly design OCR systems, evaluate various image processing algorithms for a continuum of degradation levels, and also develop algorithms for restoration purposes.

Tang and Suen try to approximate the 3D warping distortions by finding various 2D geometric transformations in bilinear, quadratic, biquadratic, cubic and bicubic models based on a set of matching reference points [TS93]. Coons and Harmonic models are also introduced to handle more complicated situations. The inverse transformation is then used to restore the nonlinear shape distortions with convincing results.

Wen and Zhu develop a multi-step algorithm that uses 2D spline functions in bilinear, biquadratic and bicubic models to estimate upper and lower text lines for linear interpolation [WZ96]. This algorithm handles text patterns in any irregular quadrilateral shapes of non-linear distortions.

Lavialle *et al.* introduce a text line straightening method based on an active contour model with the use of Bezier curves to represent the text lines [LMAB01]. More accurate results are obtained by using cubic B-splines. The restoration is done by minimizing an internal and external energy function based on the existing text lines. Automatic initializations are performed using a particle system.

Zhang and Tan design a polynomial regression model for straightening curved text lines [ZT01, ZT02, ZT03]. The proposed model is specifically designed for scanned images of thick bound document, in which text lines can be separated into a straight portion and a curved portion near the spine. The restoration is done by straightening the curved

portion in reference to the corresponding straight portion. Shading distortions are also removed using a binarization technique.

Tsoi and Brown recently propose a novel approach that uses document boundary interpolation to correct geometric and shading distortions present in images of art-like materials [TB04, BT06]. The method produces good results for a wide range of geometric warpings but restricts to iso-parametric folding lines. This method also handles shading distortions using a similar boundary interpolation algorithm. However, an obstructed white border needs to be enforced for the interpolation to work. Since they do not require the 3D shape of the warped surface, the uniform parametrization needs to be guided by a checkerboard pattern placed beneath the document. Furthermore, image boundaries must be present to reflect the warping curvature.

Most of the 2D-based approaches derive the distortion parameters based on the content of the document such as text lines, known reference points, or document boundaries. This limits their applications only to those text-dominant documents. On the other hand, if a document contains multiple columns like in common conference proceedings, newspapers, etc, these textline-based methods will not be applicable.

2.3.3 3D-Based Geometric Correction Methods

In view of the obstacles faced by the 2D-based shape-free approaches, many attempts have been made to utilize the shape information to design more general restoration approaches. Some of these approaches require special setups such as 3D scanners or structured lighting systems to capture the shape information, although their restoration process is less content dependent. Some try to reconstruct the shape using specific modeling techniques, but are

limited to certain types of surfaces. More general methods lead to the classic shape recovery problem, which is often challenging due to the limited input information. The following two sub-sections give a brief review of the approaches in this category.

Shape From Special Setups

To accommodate the fragile nature of some historical documents in their digitization process, some researchers exploited the power of those 3D range scanners to obtain the structure information of the document surface in aid of their restoration. Some efforts are also made to reconstruct the document surface shape based on special setups such as structured lighting systems.

Pilu presents a novel method that models the paper deformations of a curled document using an applicable surface, which is then unrolled to a plane using a relaxation algorithm to obtain the restored image [Pil01b]. In this method, the 3D data of the document shape is obtained through a special system using structured light. A correspondence between the 3D data and an initial flat mesh is established by translating each node vertically to its closest data point. This also determines how the texture mapping will be flattened even if the warped surface is not flattened. The geometric flattening is done using a gradient descent method, which gradually adjusts the position of the mesh node until a satisfactory low energy is achieved.

Brown *et al.* propose a restoration approach that captures the 3D geometry of the document using a structured light acquisition system and then flattens it numerically through a particle-based system with a mass-spring model [BS01b, BS04]. The 3D acquisition system is specially designed for the digitization of library materials [BS00], which

uses a mirror-deflected LCD projector and the British Library's Kontron camera. The projector and the camera are calibrated and their coordinates are registered through a structured lighting process. The restoration is done by modeling the captured 3D mesh as a particle system with springs connecting any two adjacent particles and then numerically pressing down the particles controlled by an external gravitational force and the internal spring forces. This method effectively restores arbitrary geometric distortions and is successfully applied to the digitization of some damaged manuscripts in current digital libraries [BS01a]. A recent application is to the restoration of arbitrarily distorted printed documents both geometrically and photometrically [SYY+05]. Its photometric correction technique is discussed in Section 2.1.2.

Other than flattening the surface using iterative-relaxation techniques, Brown and Pisula also propose the use of conformal mapping to parametrize a document's 3D shape to a 2D plane [BP05]. Strictly speaking, a conformal map is a 2D parametrization of a 3D surface such that angles are preserved. For general surfaces, a minimization of the angle distortions can be used. In the paper, the conformal mapping is found by solving a large linear system, which does not suffer the instability problems in previous simulation-based approaches.

Doncescu *et al.* introduce a similar method, in which a laser projector is used to project a 2D light network on the document surface to generate a set of 3D points, and then 2D distortions of the surface are corrected with a 2-pass mesh de-warping algorithm [DBQ97]. The 3D points are defined as the intersections of the 2D light network, which are extracted through a wavelet-based segmentation and filtering process. The restoration is done by fitting the 3D points to a continuous bivariate function and then mapping it to its flat counterpart.

Recently, Seales and Lin present a method to reveal invisible texts buried in damaged books or scrolls [SL04, LS05]. Unlike the previous approaches which mainly focus on opened document, this method considers impenetrable objects that often cause problems in the digitization process. In order to capture the shape of such objects, Computed Tomography (CT) scans are used to get the volumetric data. The property of CT scans makes it possible to distinguish different materials, such as pigment or paper or papyrus and therefore enables the creation of a surface from the 3D voxel set to reveal the intensity variations related to ink. Subsequently, similar physically-based simulation procedures are applied to unroll the scripts and disclose the embedded content.

Shape From Reconstruction

Apart from capturing the 3D surface shape using special equipment or setups, attempts are also made to reconstruct the shape from one or more 2D images through shape recovery techniques.

Kanungo *et al.* first introduce a global degradation model to simulate perspective distortions of scanned document images [KHP93]. This model is in reference to the flatbed scanning system. It assumes that the warped surface is circular cylindrical near the spine region and the lighting direction is vertical. The geometric distortion near the spine region is rectified by mapping the curved portion to a flat plane based on the cylindrical model.

Wada *et al.* develop a complicated model to reconstruct the 3D surface shape of thick bound books based on a single scanned image [WUM97]. The model integrates the image irradiance equation with the typical scanning system and also incorporates the

inter-reflections between the adjoining surfaces at the spine regions. This produces a set of equations with eight parameters that need to be estimated a priori by using a white paper that forms a known linear slope. The proposed method requires the book spine to be strictly parallel to the scanning light beam and the warping surface is approximately cylindrical.

Zhang *et al.* propose a more general model that also deals with scanned images of thick bound documents [ZTF04]. In comparison with Wada's work, this model handles more general cylindrical surfaces which include cross section shapes of circular, elliptic, hyperbolic and parabolic forms. This method claims to be more efficient than Wada's method although the inter-reflections are ignored. On the other hand, an improved version shows that the method is able to tackle skewed images in which the book spine is misaligned with the lighting beam [ZZTX05, TZZX06].

Similarly, Cao *et al.* build a general cylindrical model to restore geometric distortions of camera-based images [CDL03a, CDL03b]. Following the geometry of the camera image formation process, points on the 3D surface are mapped to points on the 2D image plane using the cue of directrixes. More specifically, the projections of directrixes are used to extract the baselines of the horizontal text lines as an estimation of the warping curvature. The restoration is then performed by eliminating the distance differences through x-distance normalization and y-distance normalization sequentially.

Moreover, Liang *et al.* propose to model the a warped page surface by a developable surface and exploit the properties of the printed textual content on the page to recover the surface shape [LDD05]. The developable surface basically models a smoothly rolled document page which can be unrolled to a plane without tearing or stretching. This

method makes use of the parallelism and equal spacing properties of text lines to detect projected rulings and find their vanishing points for restoration. The limitation of this method is the dependence on the text lines and the smoothness of the surface.

Instead of working on a single 2D image, Yamashita *et al.* propose a restoration method that reconstructs the document's surface using a stereo vision system and then restores the curved lines on the Non-Uniform Rational B-Spline (NURBS) surface to straight lines [YKKM04]. First, the stereo vision system captures two images from left and right two views. A set of 3D points are then detected through a triangulation process. Next, the shape of the document surface is reconstructed using NURBS curve representation. Finally, the original two images are transformed to their flat views and the clear regions are combined.

Recently, Iketani *et al.* describe a novel video mosaicking method based on 3D reconstruction of curved document using structure from motion techniques [ISI06]. In view of the low resolution problem when capturing a single large document image, this method is able to capture multiple frames of partial images and stitch them together to form one large, high resolution image. More specifically, the proposed method first tries to track a set of feature points and estimate their 3D positions together with the extrinsic parameters of the camera in each frame. The surface shape is reconstructed by fitting a 1D polynomial to a point cloud after feature selection and refinement steps. The unwarped image is obtained by mapping each 2D pixel to its 3D coordinate in the fitted surface. Typically, the surface is fitted to a 1D polynomial using the projection of all the selected feature points. This essentially implies that surfaces can only be horizontally distorted instead of any arbitrary distortions.

2.3.4 Pure Perspective Correction Methods

In traditional flatbed scanned images, perspective distortions are rare because the scanner's glass pane always ensures that the view angle between the imaging sensor and the document surface is orthogonal. However, in camera-based images, perspective distortions are pervasive because the view angle can hardly be controlled with the camera being placed at any position. One typical example is when imaging a straight road with the camera facing forward, the road in the image will tend to converge at its further end. Similarly, when imaging a planar document page from a non-orthographic viewpoint, characters that are further away from the camera will appear smaller and distorted in shape and therefore hardly recognizable by current OCR systems.

Various methods have been proposed in the document image domain to address such distortions based on the presence of orthogonal lines such as parallel text lines, page or paragraph boundary lines, etc. For example, Clark and Mirmehdi [CM03] try to use an extension of the 2D projection profile to locate horizontal vanishing point followed by the vertical vanishing point based on the change of line spacings. The vanishing points are then used to recover a frontal-parallel view of the document suitable for OCR. Pilu also proposes a method to extract horizontal and vertical features from the input image based on extracted text blocks [Pil01a]. On the other hand, Myers *et al.* [MBL+05] present a novel system that extracts text from real-world scenery images and performs perspective rectification and recognition accordingly. The algorithm assumes weak perspective projection in the vertical direction and tries to rectify foreshortening and shearing based on the top line, the base line and the vertical edge direction detected.

2.3.5 Discussion

As George Wolberg described in the book Digital Image Warping [Wol90], image warping is a subset of image processing. That is also why image warping is highly correlated with computer vision and computer graphics fields. More intuitively, Figure 2.2 shows the relationship between image warping and various image processing related areas. Looking at the existing document image de-warping approaches, all the three categories that we have summarized fit into this diagram nicely.

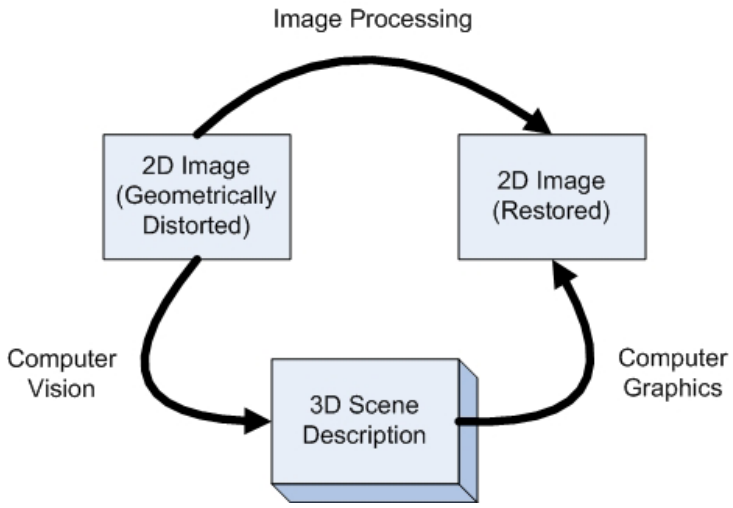


Figure 2.2: Relationship between image de-warping and image processing, computer graphics and computer vision.

As we described in Section 2.3.2, a wide range of methods have been proposed to restore geometric distortions based on solely 2D information. Generally speaking, they all use 2D image processing techniques to find geometric transformations ranging from simple affine transformation to more commonly used polynomial transformations. This can be represented by the "Image Processing" arrow in Figure 2.2.

The advantage of this set of approaches lies in its simplicity and directness. Given a warped image, the most straightforward way of restoring it will be to apply some spatial transformation function to get back its undistorted correspondence. The most common model of representing spatial distortion is low-order polynomials. This explains why many approaches are based on polynomial regression and interpolation techniques. However, to find the spatial transformation based on solely 2D information is difficult and sometimes inaccurate. One main reason is that the polynomials estimated based on text lines or document boundaries cannot precisely represent the geometric distortions present in the document due to the depth change under perspective distortions. Another reason is that the least-squares technique used to determine the coefficients of the polynomial averages a local geometric difference over the entire image independent of the position of the difference. Therefore, local geometric distortions cannot be handled accurately and often result in errors at distant locations. Weighted least-squares alleviated this problem by allowing the coefficients to be spatially varying subject to a user-specified parameter. On the other hand, to estimate the coefficients of the polynomial, we need to base on the presence of text line information or document boundary information. Approaches like this usually assume that the document contains single-column text or the document boundaries are easily identifiable from a distinguishable background. Some methods also assume that the document is clean so that the best fitting polynomial to the warping curvature can be found. Some of these assumptions may be hard to achieve in practical situations. Moreover, some arbitrary warping distortions such as wrinkles and folds are difficult to approximate globally using polynomials. All these restrict the application of such 2D approaches to a wider range of documents.

On the other hand, with the current powerful 3D scanners, not only 2D image but

also 3D structural information of the warped document can be easily captured. This brings up another stream of image de-warping techniques, which utilizes the 3D shape information to restore the distorted 2D image. The arrow from "3D Scene Description" to "2D Image (Restored)" in Figure 2.2 basically describes this scenario. The "3D Scene Description" can be interpreted as the 3D shape of the physical document. One way to realize the transition from 3D Scene Description to 2D image is through computer graphics techniques. This is where those physically-based restoration techniques fit into the picture.

The advantage of this group of approaches lies in its wide applicability to various types of geometric warpings, ranging from standard cylindrical surfaces to general developable surfaces. Moreover, most of these approaches do not depend on the content of the document, which allows applications to not only textual documents but also graphical documents. The only information that is required is the surface shape of the physically warped document. Using special 3D setups such as laser or structured light or even CT scan can help to capture the shape information easily. Some 3D scanners can even automatically generate a 3D triangular mesh to represent the surface shape. With this, the restoration step can be done either by surface interpolation techniques or through computer graphics simulations. Surface interpolation can be considered as a piecewise interpolation technique to fit a composite surface to a set of scattered 3D data points. The mapping function can be defined as piecewise polynomials or global splines depending on the type of applications. Splines are more general forms than piecewise polynomials. Most of the current studies are focusing on splines for two reasons. First, partial derivatives at each point needed by piecewise polynomial are difficult to estimate. Second, splines allow us to impose constraints on the properties of the inferred mapping functions. Besides

surface interpolation, computer graphics simulations can also be used to perform the de-warping task. In this case, the document surface shape is often represented as deformable models as in the problem of cloth simulation. Many cloth simulation techniques such as the use of a mass spring model can be adapted to the document de-warping procedure with a proper modeling scheme.

The disadvantage of the above 3D-based de-warping approaches is the requirement of a special 3D setup, which may not be readily available for some daily uses. Without a special 3D setup, it is usually hard to obtain the 3D geometric shape accurately merely based on its 2D image.

Nevertheless, as discussed in Section 2.3.3, some approaches have been successfully developed to derive the 3D surface shape analytically based on geometric and optical modeling techniques. This group of approaches can be described by the two arrows "computer vision" and "computer graphics" combined in Figure 2.2. The arrow "computer vision" basically interprets the process of shape recovery while the arrow "computer graphics" indicates the restoration process which could potentially overlap with those described in the precious category. Some of these approaches try to work on a single 2D image. However, most of them restrict the surface shape to be either strictly cylindrical or at least generally cylindrical. Some relax the condition to smooth developable surfaces but may require the textual information as a guideline to derive the curvature. Recently, some attempts are also made to use two or more images for reconstruction purposes. These basically correspond to the Shape-from-Stereo and Structure-from-Motion problems in computer vision. All in all, a key problem here is on how to recover the surface shape of a geometric distorted document accurately and efficiently based on a single 2D image.

CHAPTER 3

PHOTOMETRIC CORRECTION

Strictly speaking, shadings are caused by changes of the surface normal with respect to the illumination direction. Because of the fixed light source in most flatbed scanners, shadings in scanned document images are usually caused by non-planar shape of the document being scanned. However, due to the intricate light sources in the camera-based imaging environment and the inconsistent properties of the surface material, shadings in real-world camera images are often too complicated to be defined by a uniform model. Shadows and background noise are even more irregular, which could be in any shape, color and intensity. This makes it difficult to build a universal model that fits in all sorts of situations. Nevertheless, common properties still exist among these distortions in a sense that they all can be treated as a layer different from the foreground strokes layer. Based

on this observation, we have designed a generic method that is able to deal with different types of background distortions from shadings to shadows to background noise in either scanned or camera-based document images as part of a unified restoration framework.

3.1 Background and Motivation

As discussed in Section 2.1, various binarization-based or shape-based methods have been proposed to address the shading problems in document images and achieved promising results to a certain extent. On the other hand, methods have also been designed to tackle similar issues in natural scene images by separating reflectance and illumination images based on the notion of intrinsic images [BT78]. In particular, an image is defined as being composed of a reflectance component and an illumination component. The illumination intrinsic image here includes both shadings and shadows. Recently, color information is exploited to separate reflectance from shading based on the observation that shading is almost exclusively defined by luminance while reflectance is defined by both luminance and color [OK04]. Funt *et al.* [FDB92] propose a method for recovering shading from color images by removing reflectance component on the basis of associated abrupt chromaticity changes. In other word, they use the fact that the change of reflectance is usually caused by a change in color. Similarly, Tappen *et al.* [TFA05] introduce another method to recover shading and reflectance intrinsic images using both color information and a classifier trained to recognize local gray-scale patterns to distinguish derivatives caused by reflectance changes from those caused by shading. The intrinsic images are recovered from its derivatives using the same method as introduced by Weiss [Wei01]. In both methods, diffuse surfaces are assumed and the thresholding process can potentially

flatten out discontinuous geometric features that may appear in the shading image. Toro *et al.* [TZAF04] describe an approach that addresses both diffuse and specular reflections with a known illumination direction. Despite all these efforts made in deriving the intrinsic images, there is no single exact solution because the decomposition of the intensity image into its two intrinsic components is theoretically not unique.

In view of the large set of document images which contain mainly textual and graphical information, we may treat all the foreground layer as the reflectance component and the background layer as the illumination component. Although some contents in the background layer are not caused by illumination changes such as the background noise, they can still be separated from the foreground layer image and discarded. This matches perfectly with our main objective which is to remove the background noise and derive a clean foreground layer image. Based on this idea, we have carefully designed the following procedures to separate the background layer from the foreground layer effectively. First, a background layer is reconstructed using an edge detection procedure followed by an inpainting process. Then, the background layer is separated from the foreground layer based on the notion of intrinsic images. Finally, the reconstructed background layer can be further refined using a RBF-based smoothing technique to produce a smooth shading image that resembles the actual shading as closely as possible when shadows and background noise are not present. This shading image can then be used to reconstruct the surface shape of the document for the subsequent geometric and perspective correction tasks if needed.

3.2 A Generic Photometric Correction Method

3.2.1 Inpainting Mask Generation

To reconstruct a background layer, we need to first identify pixels that do not belong to this layer. Here we assume that in most images foreground stroke pixels have relatively higher contrast to the background than shadings, soft shadows and other noise such as bleed-through and water stains. The task is similar to text localization except that we are interested in all the pixels that may indicate a reflectance change including text and graphics. Text localization has been a widely researched area either on document images or digital videos. The techniques can be broadly classified as component-based [CM02, PO01] or texture-based [ZZJ00, LDK00]. The component-based methods usually try to analyze the geometrical arrangements of edges or uniform colored components of the characters. The texture-based methods utilize the texture characteristics of text lines to extract the text. Here we use a component-based method as the first step to identify pixels that are of high contrast to the background. Next, morphological operations are applied to the edge-detected image to generate a mask of the foreground pixels. The detailed procedures are as follows:

1) Convert color images into gray-scale. This can be done by picking the luminance component of the YUV color model;

2) Detect edges using canny edge detector. Post-processings such as non-maximum suppression and streaking elimination are also applied for better results. Different thresholds can be chosen to better distinguish foreground strokes from unwanted noise such as shadows and bleed-through pixels;

3) Perform morphological dilation followed by closing. The purpose of this step is to mask all the foreground pixels as completely as possible to avoid noise after the inpainting process. The size of the structuring element can be tuned manually or adjusted automatically based on an estimated average character height when applicable. Even if the mask does not fully cover the foreground pixels, an iterative enhancement process can be carried out to adaptively improve the mask in conjunction with the inpainting step.

3.2.2 Harmonic/TV Inpainting

Digital inpainting was pioneered by Bertalmio *et al.* [BSBC00] and has since been applied to a variety of image processing applications. Here we use it to recover the background layer of the input image which can be further refined to produce a smooth shading image in certain situations. The idea is to fill up the masked foreground regions using the neighboring background pixels. For images with uniformly colored background, this essentially recovers the shading in the masked regions based on the assumption that the local shading variation is small. On the other hand, if our main focus is on recovering the foreground content, the quality of the background layer is therefore not so crucial. For example, we probably do not need to recover a fine detail of the bleed-through and water stain pixels. To this purpose, we look into the two non-texture inpainting models - harmonic inpainting and Total Variation (TV) inpainting [CS02].

Mathematically, inpainting can be considered as a local interpolation problem: Given an image I_0 with a hole Ω_H inside, we want to find an image I that matches I_0 outside the hole and has consistent information inside the hole. To do this by harmonic inpainting,

we try to find I that minimizes the following energy in a continuous domain Ω:

$$E(I) = \int_{\Omega} \chi(I - I_0)^2 d\mathbf{x} + \lambda \int_{\Omega} |\nabla I|^2 d\mathbf{x} \tag{3.1}$$

where $\lambda > 0$ is a smoothness parameter and χ denotes the characteristic function:

$$\chi(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in \Omega \backslash \Omega_H \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

To minimize the energy in Eq. 3.1, we solve the Euler-Lagrange equation for the functional $F(I, I_u, I_v) = \chi(I - I_0)^2 + \lambda |\nabla I|^2$ corresponding to Eq. 3.1:

$$\frac{\partial E}{\partial I} = \frac{\partial F}{\partial I} - \frac{\partial}{\partial u}\frac{\partial F}{\partial I_u} - \frac{\partial}{\partial v}\frac{\partial F}{\partial I_v} = 2[\chi(I - I_0) - \lambda(I_{uu} + I_{vv})] = 0 \tag{3.3}$$

By applying a gradient-descent method and a discretization using finite difference, we obtain the iterative update formula:

$$I_{i,j}^{n+1} = I_{i,j}^n + \Delta t \left(\frac{\lambda}{h^2}(I_{i+1,j}^n + I_{i-1,j}^n + I_{i,j+1}^n + I_{i,j-1}^n - 4I_{i,j}^n) - \chi_{i,j}(I_{i,j}^n - I_{0i,j}) \right) \tag{3.4}$$

where h is the grid size and λ is the smoothness parameter which is chosen through trial and error. The time step Δt can be any small constant that makes the iteration stable, which is chosen as 0.2 in our experiments.

We notice that the harmonic inpainting constructs a smooth solution which may cause problems when the foreground pixels at the image boundaries are completely masked or when the interior edges such as folds are occluded due to the overlaid stroke pixels. The edges at these places are often missing after being inpainted with a smoothly filled interior. This can be mitigated with the use of TV inpainting. Instead of using a penalty term

$\int |\nabla I|^2 d\mathbf{x}$ in Eq. 3.1, which is infinite for discontinuous functions, we use $\int |\nabla I| d\mathbf{x}$ which allows discontinuous functions as minimizers. The energy function now becomes:

$$E(I) = \int_\Omega \chi (I - I_0)^2 d\mathbf{x} + \lambda \int_\Omega |\nabla I| d\mathbf{x} \tag{3.5}$$

where λ is the smoothness parameter. A minimizer for this energy function can be computed using a similar scheme as for harmonic inpainting.

Note that both harmonic and TV inpainting are local models, in which the inpainting is mainly determined by the existing information I_0 in the vicinity of the inpainted domain Ω_H. We can also choose to retain the original background pixels and only inpaint the masked regions to minimize computational cost. Moreover, Eq. 3.1 has a built-in denoising capacity so that it is robust to noise. The main difference is that harmonic inpainting builds very smooth solutions and thus does not cope well with edges, while TV inpainting is able to restore narrow broken smooth edges which often exist in document images due to overlaid stroke pixels.

3.2.3 Smoothing with RBF

Although the inpainting process is able to remove all the masked foreground pixels and return an estimated background layer image, the result is not perfect due to the errors in the extracted mask. This is acceptable if our purpose is mainly to remove the background noise and extract the foreground strokes such as bleed-through and water stain removal. However, if we need a smooth shading image for the sake of a subsequent surface reconstruction process, additional refinement step is then necessary.

One way to remove the pepper noise in the inpainted image is to iteratively improve

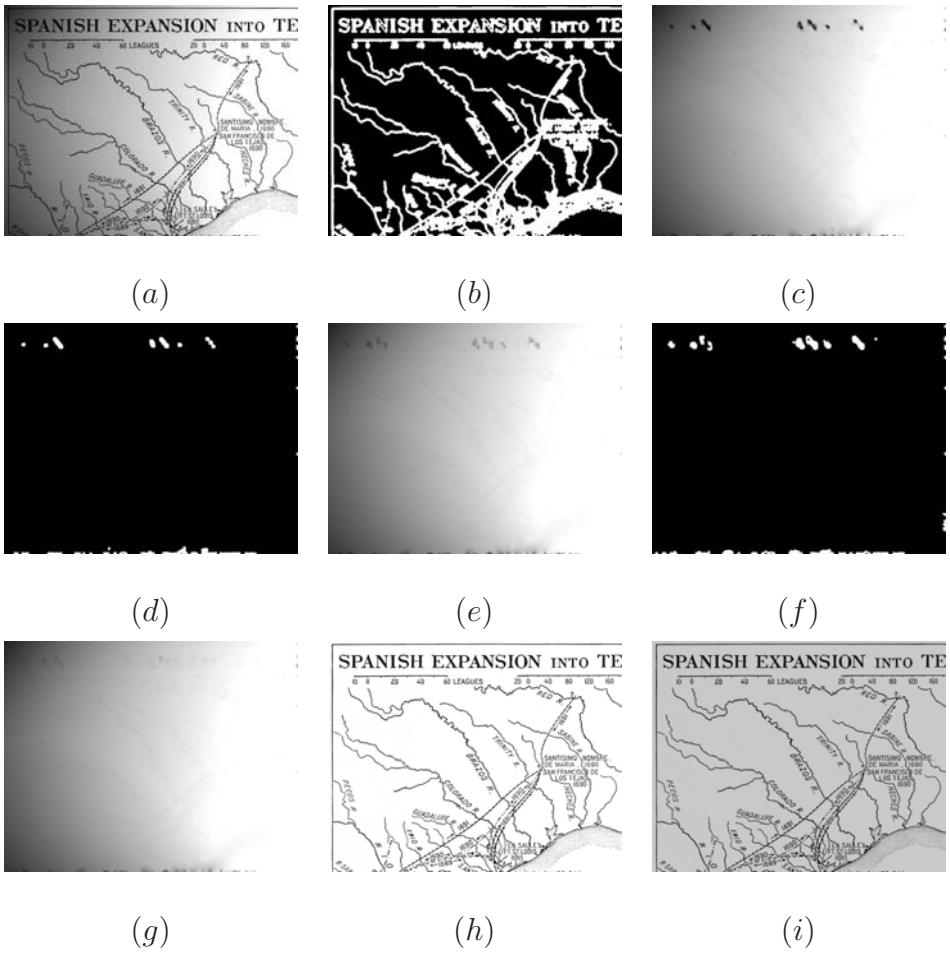
the mask and compute the inpainted image until all the foreground pixels are completely covered. More specifically, the first iteration extracts a mask from the original input image while the subsequent iterations extract masks from the inpainted images constructed in the previous step. Figure 3.1 shows an example of improving the inpainted image using an iterative approach. To separate reflectance edges from some illumination edges caused by creases or folds, the edge detector may need to be tuned specifically to avoid certain confusions. Nevertheless, even if some creases or folds are mis-identified as reflectance edges, the inpainting algorithm is still able to provide a close estimation of the original edge based on the neighboring pixel values to a certain extent.

Alternatively, we can smooth out the errors by using a smoothing algorithm with radial basis functions (RBF) [CBM+03]. This is especially useful when the background layer mainly comprises of shadings on smoothly warped surfaces. Typically, consider the pixels in the inpainted image as a set of noisy 3D points $\{(\mathbf{x}_i, I(\mathbf{x}_i)),\ i = 1, 2, \cdots, m\}$ where m is usually equal to width×height, our goal is to find an approximate fitting to these 3D points by using a selected subset of points as the collocation points that uniformly span the whole image horizontally and vertically denoted by $\{\mathbf{y}_j,\ j = 1, 2, \cdots, n\}$. The interval Δd between these collocation points can be adjusted to achieve different smoothness. In our experiments, we set it to 40 pixels. The task is therefore to find the coefficients α_j that minimize the least squares error defined as:

$$e = \min_{\alpha_1, \cdots, \alpha_n} \left\{ \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \alpha_j h(\mathbf{x}_i - \mathbf{y}_j) - f(\mathbf{x}_i) \right)^2 \right\} \tag{3.6}$$

with optional boundary conditions. The kernel function used here is Multiquadrics defined as:

$$h(\mathbf{x}) = \sqrt{||\mathbf{x}||^2 + c^2} \tag{3.7}$$



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 3.1: Background layer improvement using an iterative approach: (a) A map image with synthetic shadings; (b) Inpainting mask (Iteration 1); (c) Inpainted image (Iteration 1); (d) Inpainting mask (Iteration 2); (e) Inpainted image (Iteration 2); (f) Inpainting mask (Iteration 3); (g) Inpainted image (Iteration 3); (h) Restored reflectance image; (i) Final restored image.

where c is a constant with an empirical value of 10 for all our experiments. Other kernel functions can also be used such as Gaussian, thin-plate spline, etc.

The advantages of using RBF smoothing are:

1) It gives explicit formula for derivatives which are more accurate and less noisy than finite difference;

2) It is also easy to incorporate various types of boundary conditions;

3) Unlike polynomial fitting, RBF is more flexible and can be used to fit more complicated surfaces.

Figure 3.2 shows a typical example of how the RBF-based smoothing technique helps improve the shading image using a smoothly warped document page. In particular, the one-step inpainted image as shown in Figure 3.2 (c) contains obvious pepper noise due to unmasked ink pixels. To apply the smoothing technique, we first select 12×12 collocation points that uniformly sample the image plane as shown in Figure 3.2 (d). Next, we compute a smooth fitting to the original noisy data points based on the selected collocation points and the multiquadrics kernel function. A cross section view of the fitted surface at $v = 50$ is shown in Figure 3.2 (e). Finally, the smoothed shading image is shown in Figure 3.2 (f).

3.2.4 Background Layer Removal

Once the background layer is extracted, the foreground layer image can be derived based on the notion of intrinsic images [BT78]. For Lambertian surfaces, the intensity image is equal to the product of the shading image and the reflectance image. Therefore, the idea is to treat the foreground layer as the reflectance image and the background layer as the illumination image. Consider the luminance component of the YUV model, we have $I = I_s \cdot I_r$. Now given the background layer image I_s, the foreground layer image I_r can be computed as: $I_r = e^{log\, I - log\, I_s}$. The photometrically restored image can be computed as: $I_p = k \cdot I_r$, where $k \in [0, 1]$.

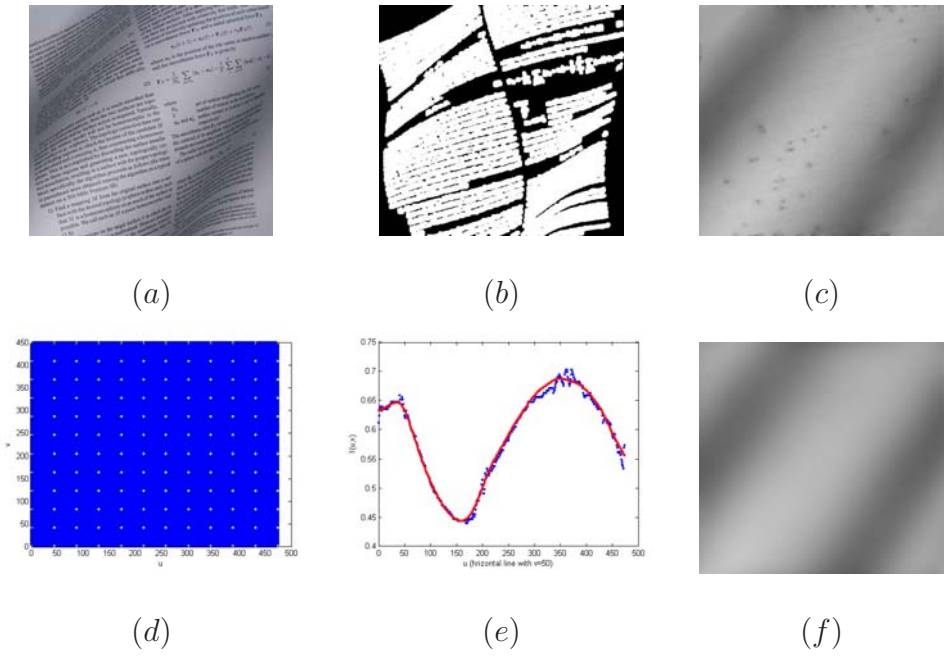


Figure 3.2: Results on RBF-based smoothing: (a) A warped document page with smooth shadings; (b) Inpainting mask; (c) Inpainted image; (d) Selected collocation points; (e) Cross section view of the fitted surface at $v = 50$; (f) Smoothed shading image.

3.3 Experimental Results

To demonstrate the performance of the above restoration procedures, we have conducted experiments on a data set including 18 synthetic images with smooth shading distortions and 35 real images with different types of distortions ranging from smooth shadings, non-smooth shadings to shadows and background degradations. The synthetic images were generated by adding controlled illumination such as an off-centered spot light source over an original clean document image. The real images were captured using either flatbed scanners, normal digital cameras or mobile phone cameras under unconstrained imaging conditions. In addition, the images include both modern documents and historical manuscripts which are either in color or gray-scale.

3.3.1 Results on Synthetic Document Images

Figure 3.3 demonstrates the performance of the proposed method using two synthetic examples picked up from the synthetic image set. The first example is a textual document containing mainly text information while the second example is a graphical document containing both texts and engineering drawings. The distorted images were obtained by simulating bad illumination conditions using an off-centered spot light source with large distance attenuations. The inpainted image shown in Figure 3.3 (d_1) was obtained after two iterations of the inpainting process using the iterative mask enhancement algorithm, while (d_2) was obtained after only one iteration. The final restored reflectance images are shown to be close resemblances of the original clean document images as given in Figure 3.3 (e_1)(e_2).

3.3.2 Results on Real Document Images

Figure 3.4 demonstrates the restoration results of two real images with smooth shading distortions due to a flash light source applied on the warped surface shape. Figure 3.4 (b_1)(b_2) show the inpainting masks and (c_1)(c_2) show the inpainted background layer images produced using the harmonic inpainting algorithm after one iteration. Figure 3.4 (d_1)(d_2) are the smooth shading images obtained after the refinement step using the RBF-based smoothing technique. Clearly, the refined images contain less noise than the inpainted image after one iteration. Finally, Figure 3.4 (e_1)(e_2) show the restored images after photometric correction, which are visually more legible notwithstanding the geometric distortions. We demonstrate how these refined shading images can be used to reconstruct the surface shape effectively for further geometric correction in Chapter 5.

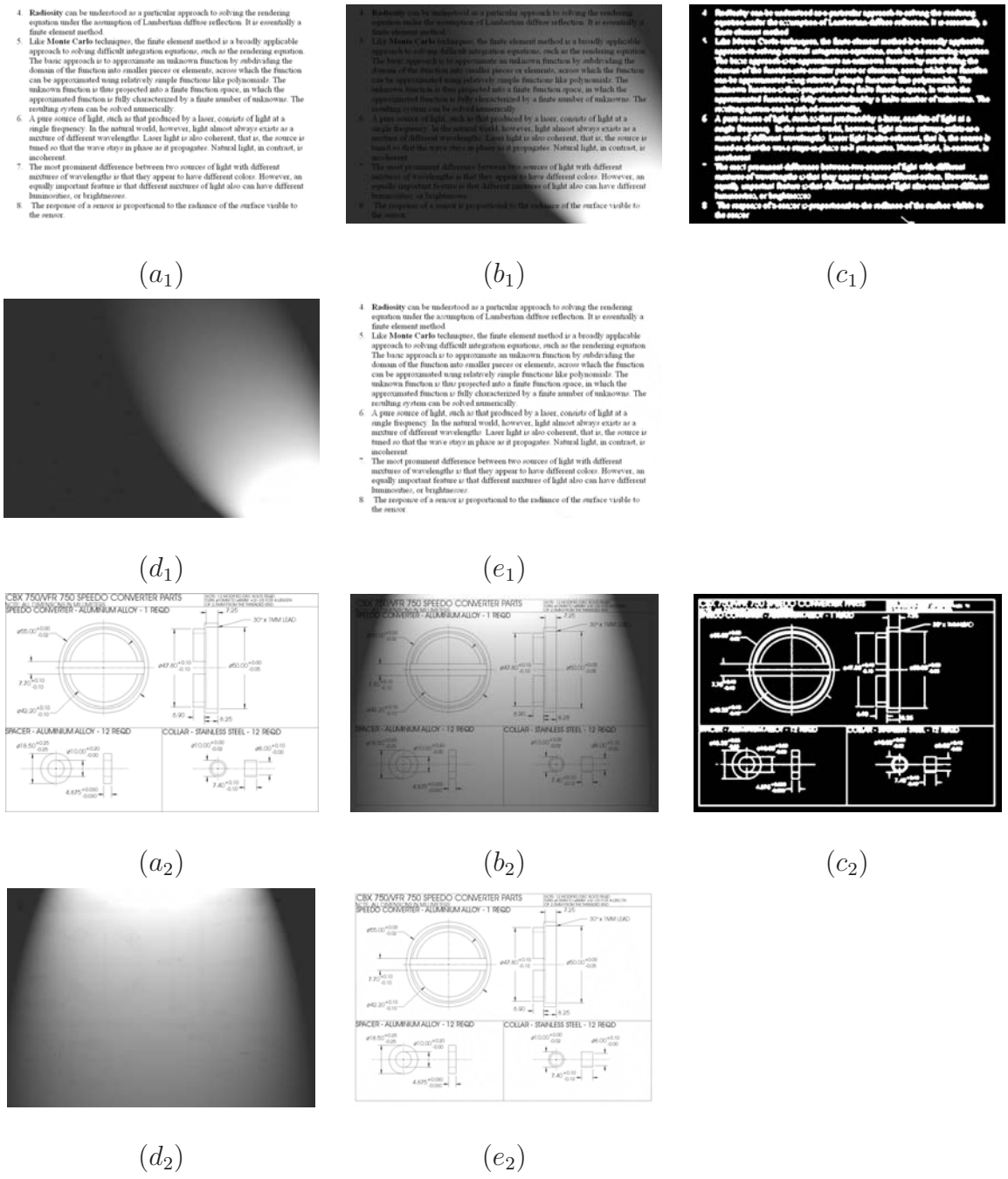
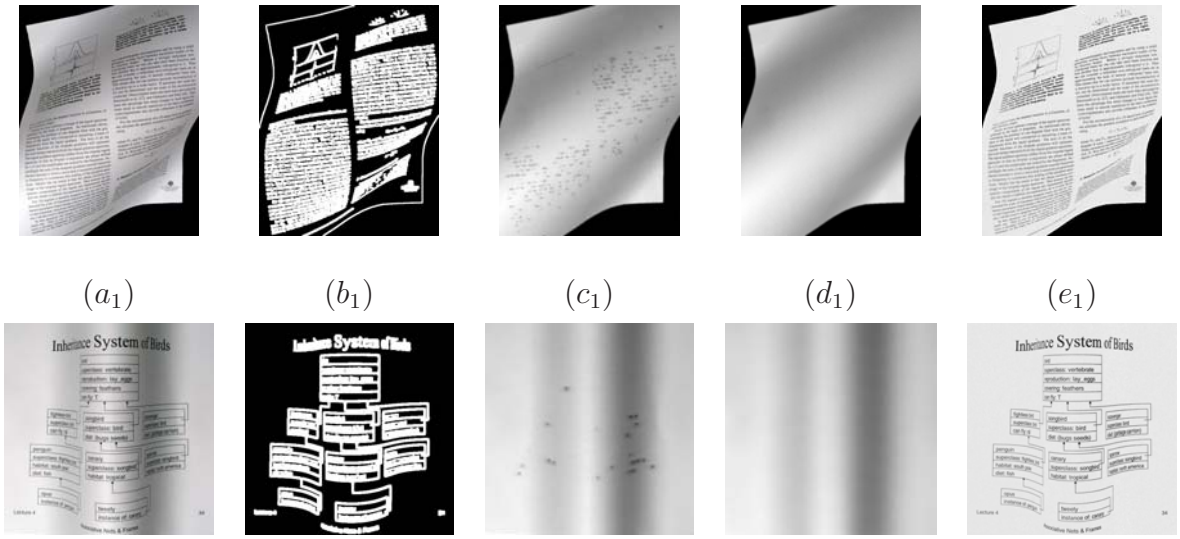


Figure 3.3: Restoration results of document images with synthetic shadings: $(a_1)(a_2)$ Original clean document image; $(b_1)(b_2)$ Distorted image with synthetic shadings; $(c_1)(c_2)$ Inpainting mask; $(d_1)(d_2)$ Inpainted image; $(e_1)(e_2)$ Photometrically restored image.



(a_1) (b_1) (c_1) (d_1) (e_1)

(a_2) (b_2) (c_2) (d_2) (e_2)

Figure 3.4: Restoration results of real document images with smooth shadings: $(a_1)(a_2)$ A real document image with shading distortions; $(b_1)(b_2)$ Inpainting mask; $(c_1)(c_2)$ One-pass inpainted image; $(d_1)(d_2)$ Smooth shading image extracted after RBF smoothing; $(e_1)(e_2)$ Photometrically restored image with $k = 0.9$.

Next, Figure 3.5 shows the results of four images with irregular shadings or different degrees of soft shadows. Figure 3.5 (a_1) is an image taken under two fluorescent lights on top of a warped document. The shadings can still be considered as smooth even though there are some irregular patterns due to the interactions between the two lights. Figure 3.5 (d_1) shows that the shadings are effectively removed regardless of the geometric distortions. Moreover, Figure 3.5 (a_2) is a folded page image with a strong folding line in the middle of the page. The inpainting mask shows that the folding line is mis-detected as foreground pixels and thus its interior is re-computed in the inpainting step. As a result, those parts that overlap with the textual content are not fully recovered because the large horizontal span of the text eliminated the valuable shading information near the folding line. Nevertheless, the recovered shading image does not deviate from the actual

shading too much because of the correct shading cue on the top and the bottom of the text lines. Therefore, this does not bring serious influences on the final restored image. Furthermore, Figure 3.5 (a_3-a_5) show three images taken using mobile phone cameras with normal indoor lightings. It is clear that shadows of the mobile phone or neighboring articles are cast on the document and result in large shading variations across the image. However, our method is able to separate the shadows effectively from the original image and produce a set of much more machine-processable images as shown in Figure 3.5 (d_3-d_5).

Moreover, Figure 3.6 shows the results of three scanned historical document images provided by the National Archive of Singapore with distortions including ink bleed-through, water stains or smudges. As we can see, the original images are seriously distorted and hard to read due to the ink bleed-through problems and material degradations. After restoration, the restored images as shown in Figure 3.6 (d_1-d_3) are much more readable which makes them suitable for various scholarly studies and investigations. In particular, the reconstructed background layer contains all the bleed-through pixels, water stain noise and smudges, which is then extracted from the original image to obtain the foreground reflectance image.

In addition, Figure 3.7 demonstrates the restoration of different degrees of show-through effects in duplex printed document pages using three typical examples. As we can see from the original images shown in Figure 3.7 (a_1-a_3) that some pixels from the back-side of the page are so strong that they interfere with the recognition of the foreground pixels. On the other hand, they are still relatively of low intensity which makes them distinguishable from the foreground strokes. Figure 3.7 (b_1-b_3) and (c_1-c_3) show the extracted inpainting mask and the inpainted background layer image, respectively. It is

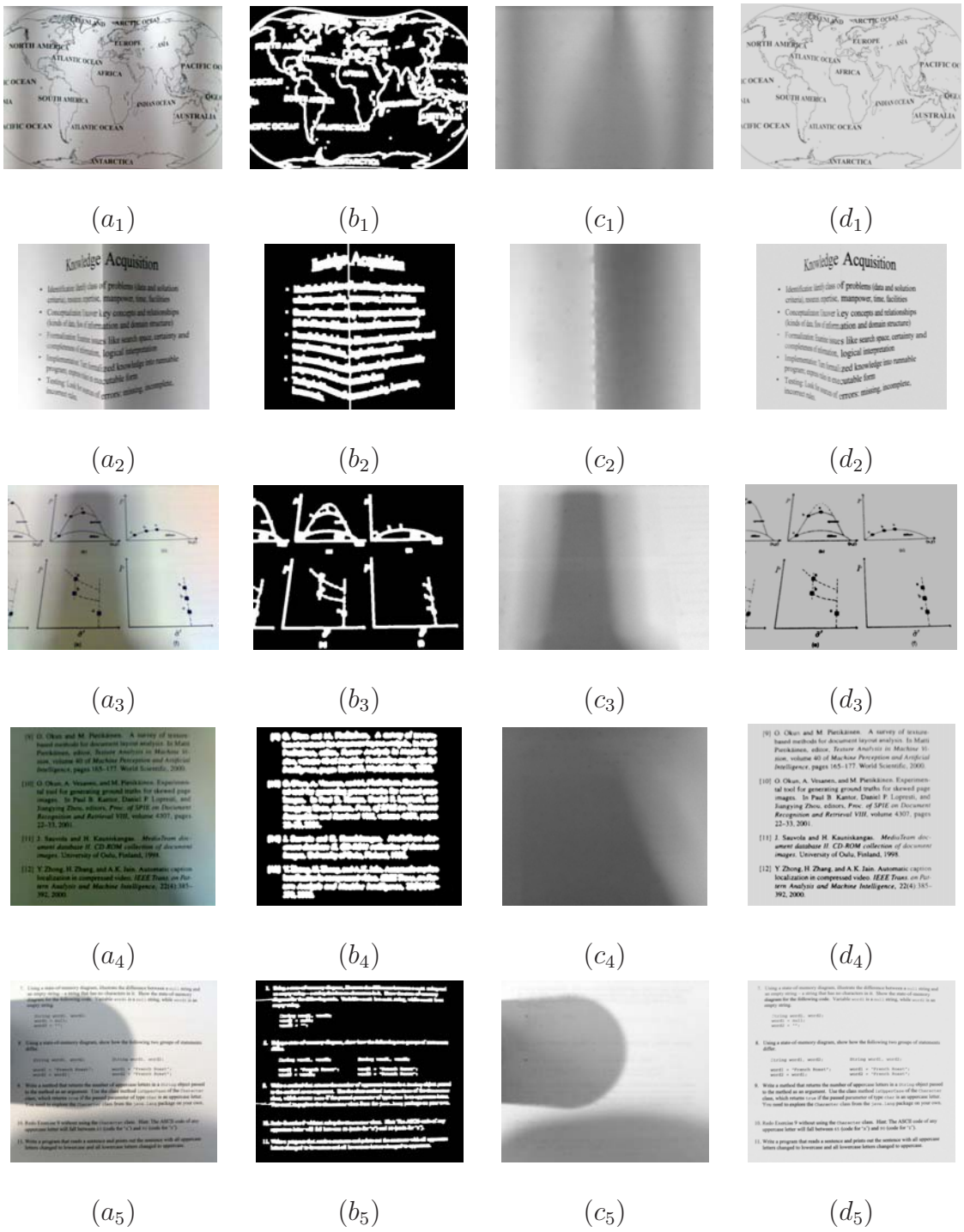


Figure 3.5: Restoration results of real document images with non-smooth shadings or shadows: (a_1-a_5) Original distorted image; (b_1-b_5) Extracted inpainting mask; (c_1-c_5) Reconstructed background layer image; (d_1-d_5) Restored image.

clear that the show-through effects are mostly captured in the background layer image which leaves a clean foreground image as shown in Figure 3.7 (d_1-d_3).

To demonstrate the computational complexity of the proposed restoration procedure, Table 3.1 tabulates the time taken for each of the restoration steps on the thirteen images shown in Figure 3.4, 3.5, 3.6 & 3.7. All the experiments were run on an Intel Pentium III 996MHz PC (512 MB RAM) with implementations using Matlab. The RBF-based smoothing step is only applicable to the two images with smooth shadings shown in Figure 3.4 as it is mainly for the purpose of extracting a smooth shading image. Therefore, the time of this step was only measured for the first two images. In particular, the time of the mask extraction step and the RBF-based smoothing step mainly depends on the size of the image. The time of the inpainting step depends on both the size of the image and the density of the masked regions.

3.3.3 Comparisons with Existing Methods

In this experiment, we have selected two images from the 18 synthetic image set with one from a textual document and the other from a graphical document to compare our results with those from two existing binarization methods. In particular, Niblack's method is a typical representative of the local adaptive thresholding methods and Lu & Tan's global thresholding method is specially designed to deal with badly illuminated document images which is similar to our focus here. For comparable reasons, we only chose images with smooth shadings since Lu & Tan's method only works on shadings that fit in a polynomial distribution. In particular, the variance gains k in Eq. 2.1 and the window size for Niblack's method is set to be -0.2 and 20, respectively. The parameters used in

Figure 3.6: Restoration results of real degraded historical document images with background noise: (a_1-a_3) Original distorted image; (b_1-b_3) Extracted inpainting mask; (c_1-c_3) Reconstructed background layer image; (d_1-d_3) Restored image.

Lu & Tan's method follow the suggested values as given in their paper [LT07]. Following this setup, Figure 3.8 $(d)(e)(f)$ show the restoration results of the distorted textual image obtained from our method, Niblack's method and Lu & Tan's method, respectively. It is observed that Niblack's method produces a large amount of background noise at regions with uniform pixel intensities such as the white margins. This is because when such white margins fully occupy a local window, the local threshold will be determined by the shading

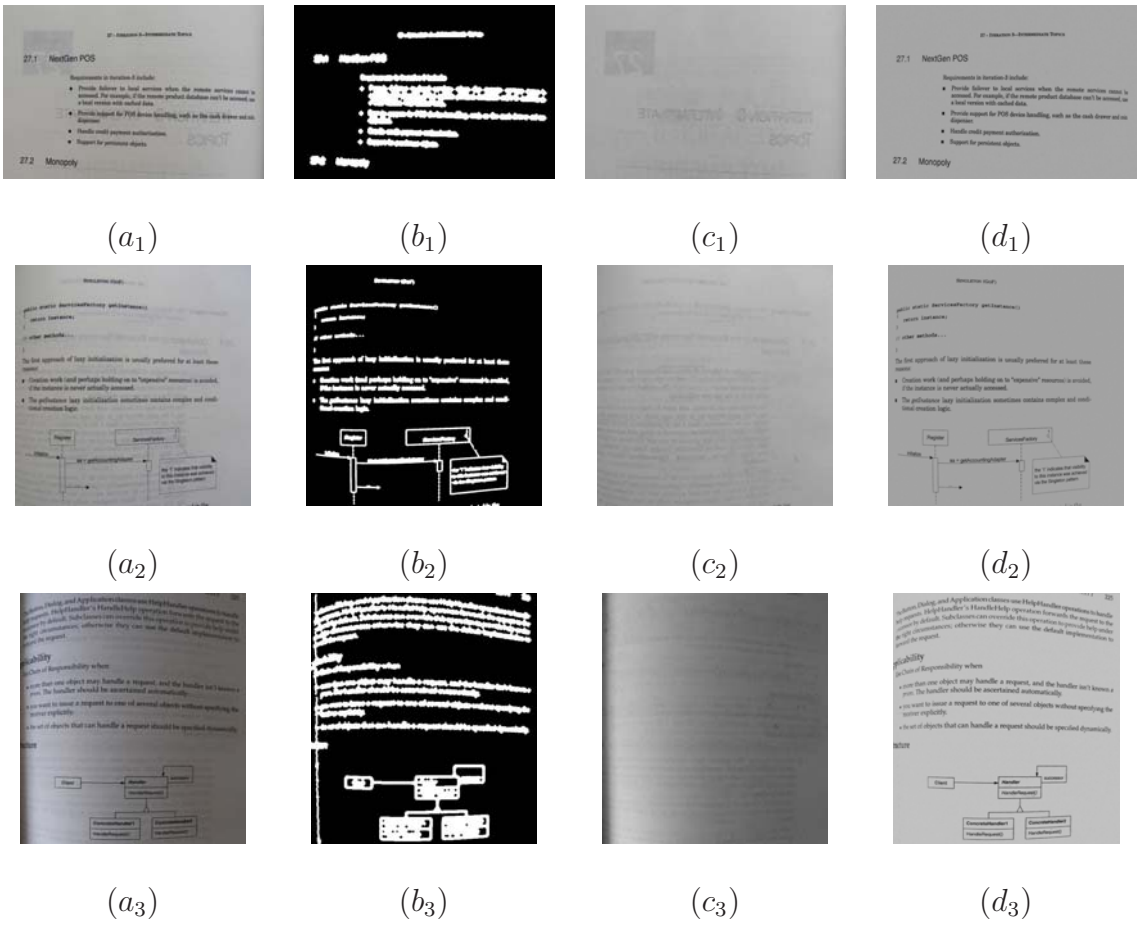


Figure 3.7: Restoration results of duplex printed document images with show-through effects: (a_1-a_3) Original distorted image; (b_1-b_3) Extracted inpainting mask; (c_1-c_3) Reconstructed background layer image; (d_1-d_3) Restored image.

variations instead of the uniform reflectance and thus results in an erroneous classification. Lu & Tan's method produces a better restored image but some text pixels in the upper right portion are eliminated due to over-thresholding. In contrast, our method achieves better result both visually and in term of OCR performance. Typically, the average word precision achieved for ten similar textual document images on their resorted images using the three methods are 77.5%, 90.4% and 96.8%, respectively. In addition, Figure 3.9 shows the results on a graphical document image that mainly consists of engineering drawings.

Table 3.1: Running time of each restoration step on the images shown in Figure 3.4, 3.5, 3.6 & 3.7.

Images	Size	Mask Gen. (s)	Inpainting (s)	RBF smoothing (s)	Total (s)
Figure 3.4 (a_1)	634×816	3.72	141.13	39.74	184.59
Figure 3.4 (a_2)	570×578	1.47	42.12	21.81	65.40
Figure 3.5 (a_1)	640×480	1.50	69.24	-	70.74
Figure 3.5 (a_2)	548×476	1.34	49.09	-	50.43
Figure 3.5 (a_3)	597×467	1.33	19.60	-	20.93
Figure 3.5 (a_4)	640×480	1.43	31.45	-	32.88
Figure 3.5 (a_5)	1056×928	3.64	102.59	-	106.23
Figure 3.6 (a_1)	640×1026	2.34	125.56	-	127.90
Figure 3.6 (a_2)	1104×597	2.35	97.54	-	99.89
Figure 3.6 (a_3)	640×1157	2.06	70.48	-	72.54
Figure 3.7 (a_1)	480×320	1.04	11.91	-	12.95
Figure 3.7 (a_2)	1112×1134	4.22	107.01	-	111.23
Figure 3.7 (a_3)	560×623	1.59	60.18	-	61.77

The performances are simialr to those on the textual document.

3.3.4 Method Evaluation

Despite the good performance on a general set of document images, the proposed procedure may also produce less satisfactory results in some extreme cases or on certain types of images. For example, in the case of bleed-through removal, if the degradations are too

(a) (b) (c)

(d) (e) (f)

Figure 3.8: Restoration result of a badly illuminated textual document image and its comparison with the results from existing methods: (*a*) Input image with shading distortions; (*b*) Extracted inpainting mask; (*c*) Reconstructed background layer image; (*d*) Restored foreground reflectance image; (*e*) Result from Niblack's method; (*f*) Result from Lu and Tan's global thresholding method.

severe such as when the bleed-through pixels are even more intense than the foreground pixels, confusions will arise. Figure 3.10 (a_1) gives an example of a degraded image with serious bleed-through distortions. After applying the inpainting method, the restored image as shown in Figure 3.10 (d_1) is still difficult to read because many bleed-through pixels are mis-detected as foreground pixels while many original foreground pixels are eliminated. Another situation is when the original image contains large color figures, some parts of the figures will not be correctly masked due to their large sizes and the

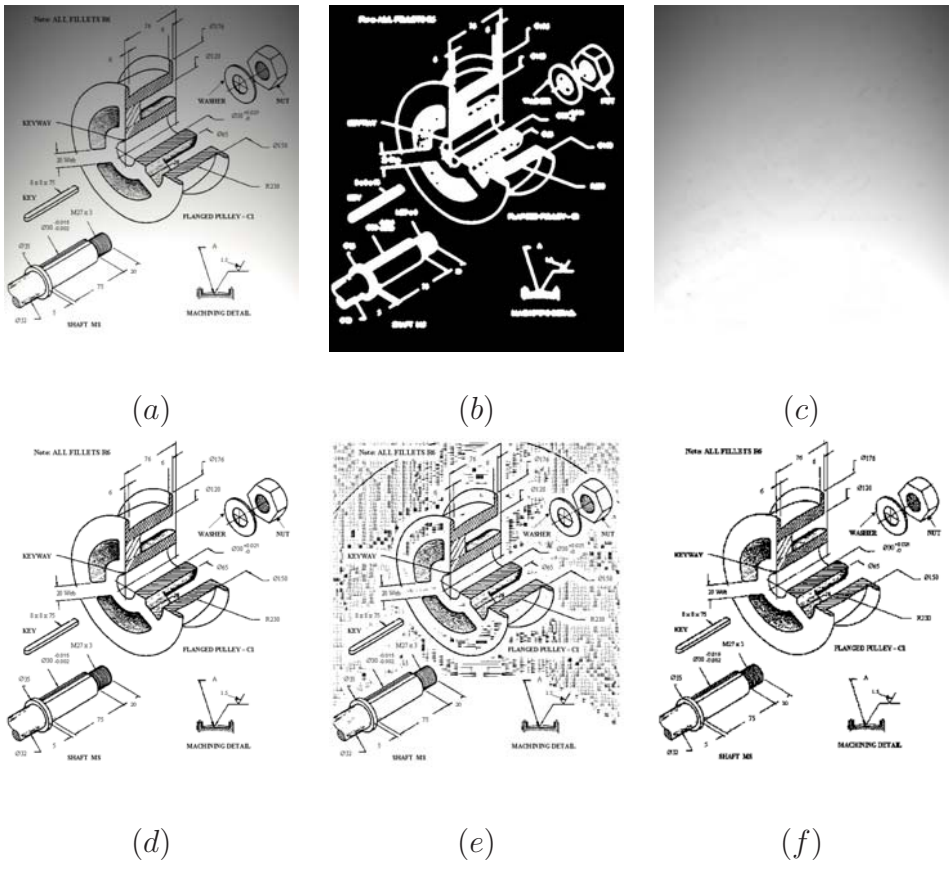


Figure 3.9: Restoration result of a badly illuminated graphical document image in comparison with the results from existing methods: (a) Input image with shading distortions; (b) Extracted inpainting mask; (c) Reconstructed background layer image; (d) Restored foreground reflectance image; (e) Result from Niblack's method; (f) Result from Lu and Tan's global thresholding method.

shadings around the figures will not be correctly approximated. This will result in errors in the final restored image at the figure region as shown in Figure 3.10 (d_2). One way to remedy this problem is to detect those large figure components beforehand and mask them as a whole. The inpainting method will then fill in the whole region accordingly. Although there are still some errors in the approximated shadings inside the filled region, the restored image is overall satisfactory if the primary concern is not on the fine-grained details of the figure itself as shown in Figure 3.10 (d_3).

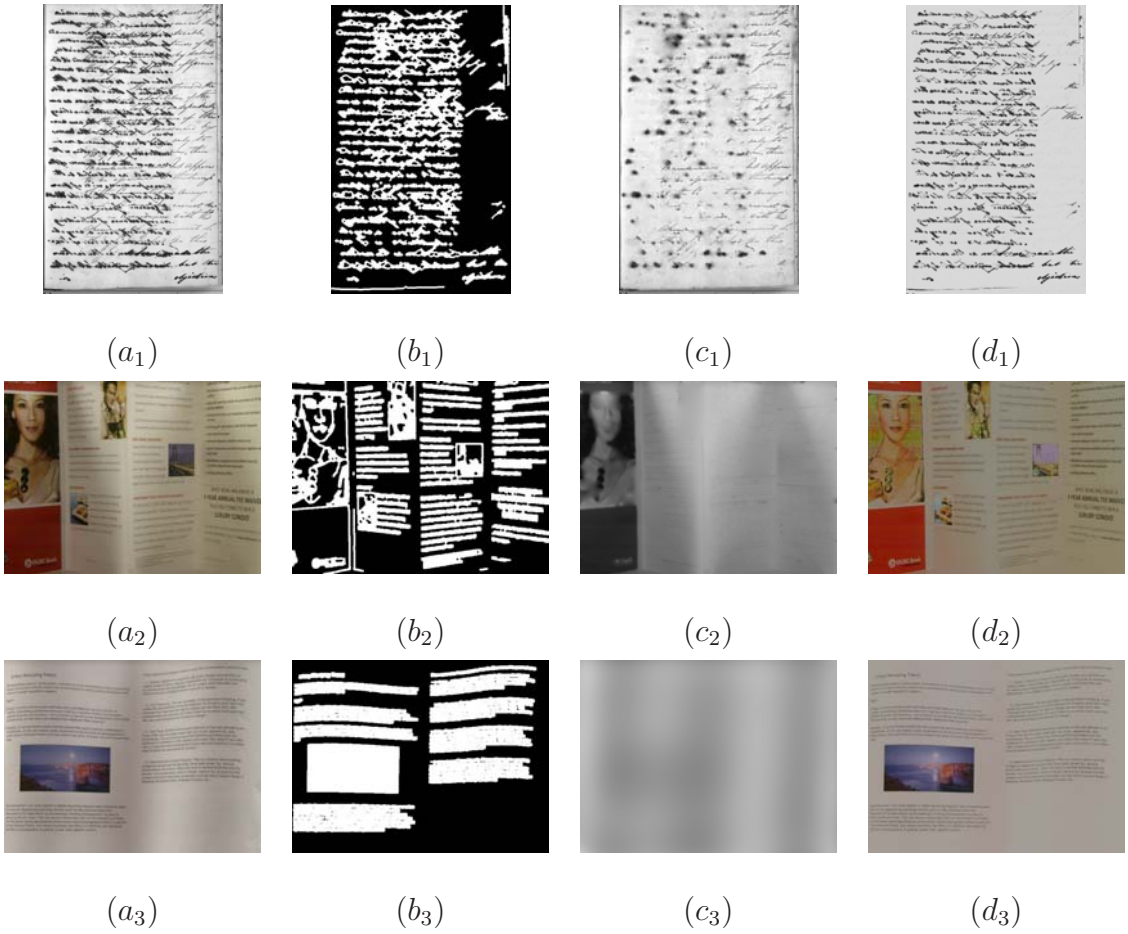


Figure 3.10: Restoration results of real document images with severe background noise or large embedded figures: (a_1-a_3) Original distorted image; (b_1-b_3) Extracted inpainting mask; (c_1-c_3) Reconstructed background layer image; (d_1-d_3) Restored image.

3.4 Discussion

In this chapter, we have described a generic photometric correction method which effectively removes shadings, shadows and background noise distortions in daily snapshots of modern documents or scanned images of degraded historical manuscripts. The idea is to construct a background layer image which approximates all the shadings, shadows or background noise and then separate it from the original intensity image to get the

foreground reflectance image. This method works well on document images with various types of irregular shadings due to bad illumination conditions, or soft shadows due to light source occlusions, or background noise in degraded historical documents such as bleed-through, water stains or smudges. In the case of smooth shadings or soft shadows, the foreground textual content can be easily identified using an edge detection procedure because the background pixels are usually in smooth transitions. Meanwhile, the inpainting process effectively fills in the foreground regions with the neighboring background information and thus produces a background layer image that closely approximates the distortions. When non-smooth shadings such as folds or sharp shadows are considered, there might be some loss of accuracy at such regions if the folds or sharp shadow edges are mis-identified as foreground edges. Nevertheless, the inpainting process is still able to recover part of the shadings at these edges and thus minimizes the effect on the restored images as shown in Figure 3.5.

Compared to the existing shading correction methods using adaptive thresholding techniques as discussed in Section 2.1.1, this method is able to handle more complex and more severe shadings with better performances as shown in several comparisons in Figure 3.8 and 3.9. On the other hand, compared to those shape-based methods as discussed in Section 2.1.2, this method does not require any shape information and is able to handle images with pure shading and shadow distortions instead of only those accompanied with geometric distortions.

The main assumption of this method is that the foreground pixels and the background pixels usually have a distinguishable contrast to the background or a different pattern in terms of continuity. When this condition is violated in certain cases such as the severe bleed-through problem as shown in Figure 3.10 (a_1), the bleed-through pixels essentially

have no difference from the foreground pixels and are therefore hard to be separated. On the other hand, when large figures are involved in the document image, the two inpainting algorithms discussed in our method here may have problems to fill in the correct shadings inside the figure box which consequently results in some loss of information in the restored image. However, this information loss is mainly on the figure itself, which may not affect the textual content analysis tasks significantly as shown in Figure 3.10 (a_3).

CHAPTER 4

SURFACE SHAPE RECONSTRUCTION

Shape-based geometric correction methods have achieved rather promising results in the past decades. All these attribute to the critical shape acquisition step which provides an accurate 3D model to represent the actual physical warpings. As discussed in Chapter 2, the 3D-based geometric correction methods can be further divided into two sub-categories depending on how the surface shape is acquired, particularly whether they are acquired using special setups or reconstructed based on one or more 2D images. In this chapter, we present two surface reconstruction methods based on Shape-from-Shading (SFS) techniques using Fast Marching scheme or Lax-Friedrichs-based Fast Sweeping scheme, respectively. As mentioned in Section 3.2.3, the surface fitting step is able to construct a smooth shading image for the subsequent surface reconstruction process. This is where

the constructed shading image is utilized to perform surface reconstruction.

4.1 Background and Motivation

Shape recovery is a classic and fundamental problem in computer vision. Its goal is to derive a 3D scene description from one or more 2D images. Over the years, researchers have developed a variety of techniques to tackle this problem known as Shape-from-X where X can be shading, stereo, motion, texture, etc. In particular, Shape-from-Shading tries to make use of the shading variations in a single 2D image to reconstruct the original surface shape. The research in this field was pioneered by Horn who first formulated the SFS problem as that of finding the solution of a nonlinear first-order Partial Differential Equation (PDE) called the brightness equation [Hor75]. Following this, a series of variational methods [IH81, HB86, FC88] are developed which try to minimize an energy function that often comprises of an integral of the brightness error to find the solution. Later Oliensis and Dupuis propose to cast the SFS problem as an optimal control problem and directly find the depth map of the surfaces [OD91]. This brought out a new set of propagational approaches based on the theory of viscosity solutions to Hamilton-Jacobi equations [RT92, PFR02, PCF06].

According to the numerical schemes used to estimate the viscosity solutions, these methods can be further divided into two categories. The first class of methods are based on the monotonicity of the solution along the characteristic direction. Examples are the level set method [OS88, KB95] and the Fast Marching method [Set99] proposed by Sethian. The idea of the level set method is to track topology changes of a surface by considering it as the zero level set of a higher dimensional function called the level set function.

Associated with the initial value formulation of the surface motion, the level set method is essentially an entropy-satisfying scheme that solves a time-dependent Hamilton-Jacobi equation using an upwind finite difference approximations from hyperbolic conservation laws. Because of the time dependence property, the level set method usually runs slower with a complexity of $O(N^2)$, where N is the total number of grid points. This can be further improved to $O(kN)$ with the narrow band level set method [AS95], where k is the width of the narrow band. In the Fast Marching method, the solutions are found by using Dijkstra algorithm with a dynamic programming strategy. The time complexity of such a method is $O(N \log N)$. Various adaptations of the Fast Marching method have been developed to handle situations like oblique light sources [KS01] and perspective projection [TSY04, YTLC02]. However, most of them assume the Hamiltonian is convex and homogeneous. Recently, Prados and Soatto have extended the Fast Marching method to handle situations in which the solution is not systematically decreasing along the optimal trajectories [PS05] with results on some synthetic images. On the other hand, the second class of methods make use of iteration strategies. Rouy and Tourin have exploited an upwind and monotone scheme to solve the discretized Eikonal equation iteratively and the convergence property is shown [RT92]. Tsai *et al.* have combined the upwind monotone Godunov Hamiltonian with a Gauss-Seidel iteration method to reconstruct surfaces with good efficiency [TCOZ03]. Besides the variational methods and propagation methods, linear approaches and local approaches have also been developed. Linear approaches compute the solution by performing certain linearization to the reflectance map [TS94]. Local approaches derive the surface shape based on the assumption of certain surface type such as spherical surfaces [Pen84]. More comprehensive surveys can be found in [ZTCS99, DFS04].

Enlightened by the various attempts of using SFS methodologies to reconstruct shapes of objects, we try to adapt the SFS concept to perform surface reconstruction in our document image domain. The requirement of a single input shading image for the SFS-based method fits in well with the daily imaging applications where a single 2D image is often available. To ensure that unambiguous geometrical properties of a document surface can be inferred from the image irradiance, some assumptions need to be imposed on the SFS formulation: 1) the document surfaces follow Lambertian reflection; 2) there is no inter-reflections between document pages; 3) sensor noise is not considered; 4) the illumination direction is known. Based on these assumptions, we derive the image irradiance equation (IIE) under different projection models and with different types of light sources depending on the imaging environment. The main task is then to solve the IIE which is usually a non-linear PDE to obtain the surface depth map.

As discussed earlier, perspective distortions are caused by the perspective projection property of the pinhole camera model, which makes the faraway points appear smaller than those closer to the camera. On the other hand, when the document being imaged has a non-planar surface shape such as a thick bound book, the resultant image will often appear warped because the distance from the surface points to the camera varies. The surface shape can thus be represented as a depth map that records the distance from each surface point to the image plane. Similarly, we can treat pure perspective distortions on planar surfaces as geometric distortions too. The shape is now a slanting surface with respect to the image plane as illustrated in Figure 4.1. In the following sections, we present two methods to solve the image irradiance equations which are formulated in different forms depending on the projection model used and the light source imposed during the imaging process.

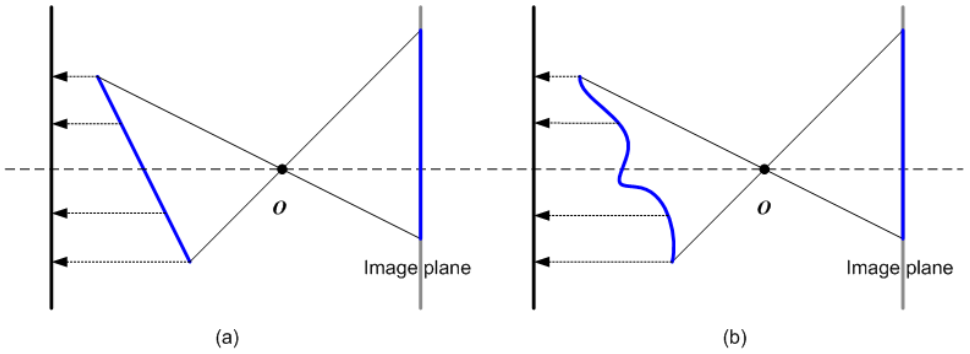


Figure 4.1: 2D illustration of (a) a pure perspective distortion and (b) a general geometric distortion with perspective distortion.

4.2 SFS Using Fast Marching Method

4.2.1 Problem Formulation

Following the document imaging system shown in Figure 4.2 with the camera facing up toward the document's surface and assuming an imaging environment with a pinhole camera and a close point light source at a known location, the reflectance map which specifies the radiance of the surface as a function of its orientation can be derived.

First, some of the notations are defined as follows. The surface shape is represented using the depth function $z(x, y)$ for a set of surface points $(x, y, z(x, y))$. The projection of the surface point $(x, y, z(x, y))$ is denoted by (u, v) in the image plane Ω. By definition, $z(x, y) \equiv z(u, v)$. The real-world coordinates of the point (u, v) in the 2D image is defined as (u', v', f), where f is the focal length of the camera and $u' = u - u_0, v' = v - v_0$. The focal length f and the principle point coordinates (u_0, v_0) are obtained through a simple camera calibration process [Zha99]. Let $\hat{z}(u, v) = \frac{z(u,v)}{f}$, the document surface can be

represented using a similar scheme as presented in [PF03]:

$$S = \{\hat{z}(u,v)\cdot(u',v',f),(u,v)\in\Omega\} \tag{4.1}$$

Assuming the illumination is a known close point light source located at (α,β,γ), the unit vector of the illumination direction at each surface point $\mathbf{x}$ can be written as:

$$L(\mathbf{x}) = (\alpha-\hat{z}(u,v)u',\beta-\hat{z}(u,v)v',\gamma-\hat{z}(u,v)f) \tag{4.2}$$

The downward surface normal at each point $\mathbf{x}$ can be derived by calculating the tangent vectors in both u and v directions and computing their cross product:

$$N(\mathbf{x}) = (fp',fq',-1-u'p'-v'q') \tag{4.3}$$

where $p' = \frac{\partial ln\hat{z}}{\partial u}$ and $q' = \frac{\partial ln\hat{z}}{\partial v}$.

Based on Eq. 4.2 and 4.3 with the assumption of Lambertian reflection, we derive the image irradiance equation based on Lambert's cosine law as:

$$I(u,v) = \frac{(\alpha f-\gamma u')p'+(\beta f-\gamma v')q'+\hat{z}f-\gamma}{ar^2\|N(\mathbf{x})\|\|L(\mathbf{x})\|} \tag{4.4}$$

where $r = \sqrt{(\alpha-\hat{z}u')^2+(\beta-\hat{z}v')^2+(\gamma-\hat{z}f)^2}$ is the distance from the point light source to the surface point and a is a constant that accounts for the distance attenuation of the close point light source.

4.2.2 Solving the IIE Using a Fast Marching Method

To solve the PDE in Eq. 4.4, we refer to the Fast Marching algorithm presented by Tankus *et al.* [TSY04], which uses an iterative scheme to solve the perspective SFS problem with

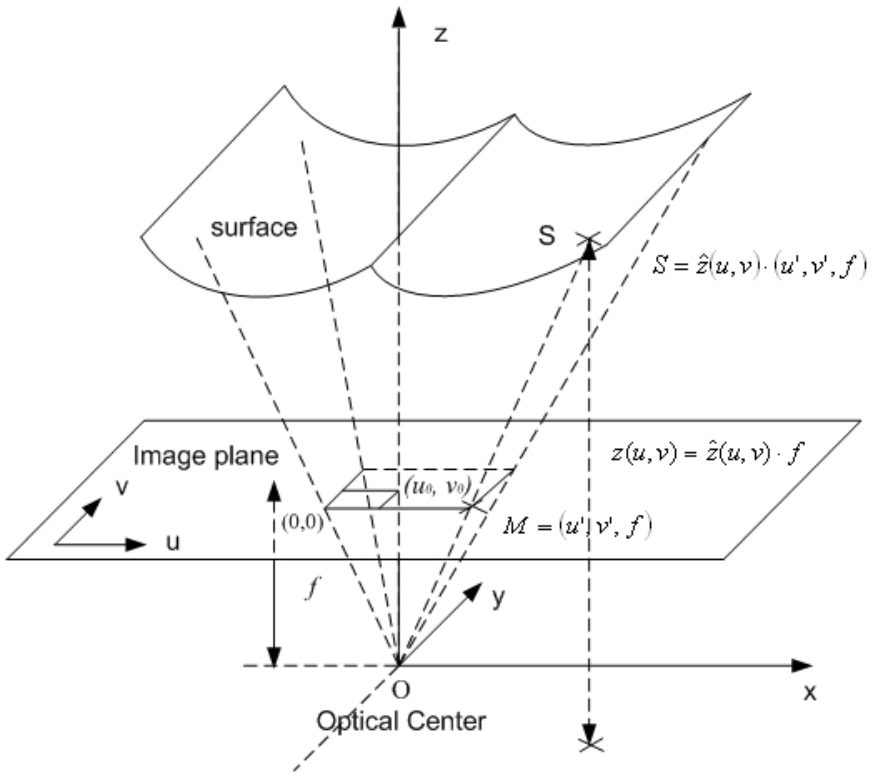


Figure 4.2: Camera imaging system and the SFS formulation with a close point light source.

a distant point light source. In contrast, our problem here requires a close point light source for which the distance attenuation needs to be considered.

Based on [KS01], the image irradiance equation for the orthographic case with a vertical illumination $L = (0, 0, -1)$ can be written as an Eikonal equation:

$$\sqrt{z_p^2 + z_q^2} = \sqrt{\frac{1}{I(u,v)^2} - 1} \tag{4.5}$$

where (z_p, z_q) is the surface gradient.

Similarly, our formulation of the image irradiance equation in Eq. 4.4 can be simplified into the form:

$$p'^2 A_1 + q'^2 B_1 = F \tag{4.6}$$

where A_1 and B_1 are non-negative and independent of p', q' and $\hat{z}$, while F depends on all p', q' and $\hat{z}$. More specifically, following [TSY04], Eq. 4.4 can be transformed into:

$$p'^2 A_1 + q'^2 B_1 = p'^2 A_2 + q'^2 B_2 - 2p'q'C - 2p'D - 2q'E - F_1 \tag{4.7}$$

where

$$\begin{aligned}
&A_1 = a^2 I^2 (f^2 + u'^2), B_1 = a^2 I^2 (f^2 + v'^2) \\
&A_2 = (\alpha f - \gamma u')^2 / r^4, B_2 = (\beta f - \gamma v')^2 / r^4 \\
&C = a^2 I^2 u'v' - (\alpha f - \gamma u')(\beta f - \gamma v') / r^4 \\
&D = a^2 I^2 u' - (\hat{z} f - \gamma)(\alpha f - \gamma u') / r^4 \\
&E = a^2 I^2 v' - (\hat{z} f - \gamma)(\beta f - \gamma v') / r^4 \\
&F_1 = a^2 I^2 - (\hat{z} f - \gamma)^2 / r^4
\end{aligned} \tag{4.8}$$

To solve Eq. 4.7, an iterative scheme can be used by first estimating the value of F in Eq. 4.6 using the values of p', q' and $\hat{z}$ in the $(n-1)^{th}$ iteration and then computing the new values of $\hat{z}$ for the n^{th} iteration using the numerical approximation scheme given in [TSY04]. The approximations for p' and q' are given as:

$$\begin{aligned}
p' &\equiv max \left\{ \frac{\hat{z'}(i,j) - \hat{z'}(i,j-1)}{\Delta u}, -\frac{\hat{z'}(i,j+1) - \hat{z'}(i,j)}{\Delta u}, 0 \right\} \\
q' &\equiv max \left\{ \frac{\hat{z'}(i,j) - \hat{z'}(i-1,j)}{\Delta v}, -\frac{\hat{z'}(i+1,j) - \hat{z'}(i,j)}{\Delta v}, 0 \right\}
\end{aligned} \tag{4.9}$$

where $\hat{z'}(i,j) = ln\hat{z}(i,j)$.

As shown by Rouy and Tourin [RT92], this numerical approximation gives the viscosity solution for the orthographic SFS problem. Similarly, the solution for Eq. 4.7 can be

derived as:

$$\hat{z'}^{n+1} = \begin{cases} \hat{z'}_1^n + \sqrt{\frac{F^n}{A_1}}, & \text{if } \hat{z'}_2^n - \hat{z'}_1^n > \sqrt{\frac{F^n}{A_1}} \\ \hat{z'}_2^n + \sqrt{\frac{F^n}{B_1}}, & \text{if } \hat{z'}_1^n - \hat{z'}_2^n > \sqrt{\frac{F^n}{B_1}} \\ \frac{\hat{z'}_1^n A_1 + \hat{z'}_2^n B_1 + \Delta F}{A_1 + B_1}, & \text{otherwise} \end{cases} \tag{4.10}$$

where $\hat{z'}^{n+1} = ln\hat{z}^{n+1}$, and $\hat{z}^{n+1}$ denotes the surface depth at the $(n+1)^{th}$ iteration,

$$\begin{aligned} \Delta F &= \sqrt{(A_1 + B_1)F^n - A_1 B_1 (\hat{z'}_1^n - \hat{z'}_2^n)^2} \\ \hat{z'}_1 &= min\{\hat{z'}_{i,j-1}, \hat{z'}_{i,j+1}\} \\ \hat{z'}_2 &= min\{\hat{z'}_{i-1,j}, \hat{z'}_{i+1,j}\} \end{aligned} \tag{4.11}$$

To initialize the iterative process, we use the depth value obtained through the orthographic SFS problem with an oblique light source. The adaptation from vertical light source to oblique light source can be considered as transforming the surface from the original coordinate system to the light source coordinate system with the light source located at $(0, 0, -1)$. Once the surface is reconstructed in the light source coordinate system, it is transformed back to the original coordinate system accordingly.

4.3 Experimental Results I

We have tested the proposed approach on both synthetic shading images and real document images. The experiments on synthetic shading images show that the modified SFS algorithm is able to reconstruct various warped surfaces effectively. In addition, we have tested on real images taken from pages in conference proceedings and compared the restored images with those obtained from range-scanned surface shapes. The results are encouraging.

4.3.1 Experiments on Synthetic Images

To generate a synthetic shading image, we first created a warped surface with the function: $\hat{z}(x, y) = 3x^3 + 100$. This gives us a surface as shown in Figure 4.3 (a).The shading image was generated using Eq. 4.4 with the light source given as $L = (1, 0, -1)$, the focal length set as $f = 200$ and $a = 1$. The camera was assumed to be on top of the document surface. Figure 4.3 (c) shows the reconstructed surface using the proposed method. Because of a fast convergence rate, here we show the result after 2 iterations. To show its similarity with the original ground truth shape, we contrast the two shapes in Figure 4.3 (d). Due to the inaccurate approximations of the initialization process, the reconstructed shape does not perfectly resemble the ground truth shape. However, this does provide a rough description of the surface structure.

We have also compared our results with those presented by Tankus *et al.* in [TSY03] on perspective reconstruction of the surface: $\hat{z}(x, y) = 2\cos\sqrt{x^2 + (y-2)^2} + 100$. Figure 4.4 (a) and (b) show the original ground truth surface and the corresponding shading image produced under perspective projection with a vertical light source $L = (0, 0, -1)$. Figure 4.4 (c) and (d) show the contrast of our reconstructed surface over the ground truth surface and that presented by Tankus *et al.*, respectively. It demonstrates that our method produces a better recovery of the original surface with less abrupt errors.

4.3.2 Experiments on Real Images

First, warped document images were taken using a normal digital camera Olympus C7070 ($f = 518$). These initial images were pre-processed to obtain the shading image using the

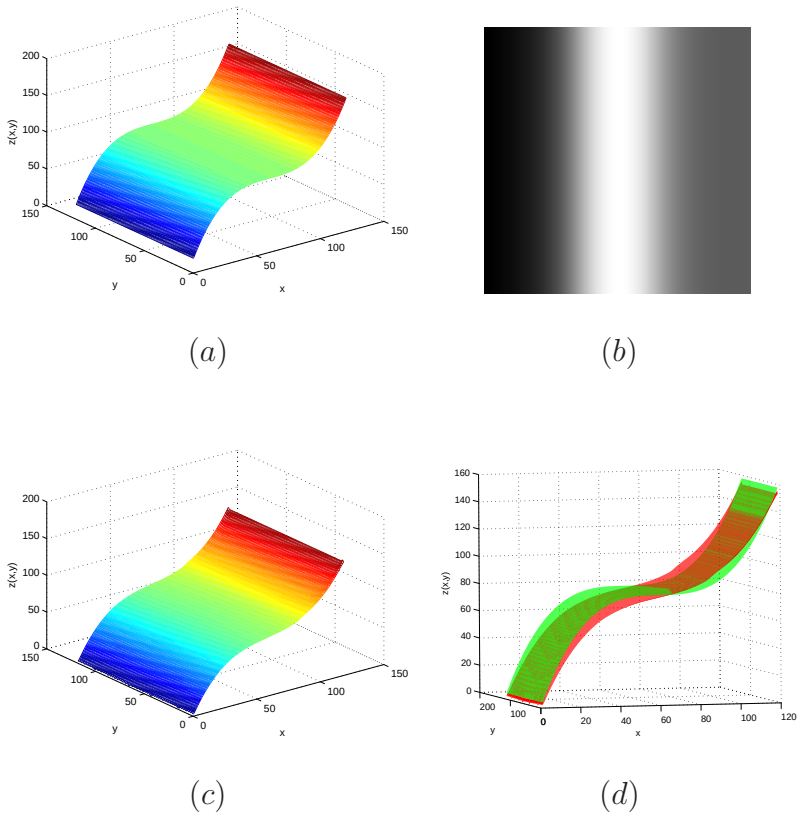


Figure 4.3: Surface reconstruction from a synthetic shading image using the Fast Marching method: (a) Synthetic surface: $\hat{z}(x,y) = 3x^3 + 100$; (b) Shading image under perspective projection with $L = (1, 0, -1)$; (c) Reconstructed surface from (b); (d) Real (green) vs. Reconstructed (red) surface.

method described in Chapter 3. Here we assume no shadows or interreflection effects, and the paper is thick enough to avoid the show-through effects from its backside contents. Proper gamma correction was also applied to the shading image when necessary. Next, the SFS technique was used to reconstruct the surface shape. As mentioned in Section 4.2, we first used an orthographic Fast Marching algorithm to provide an initial estimation of the surface depth. To start with, we initialized the singular point with a height of 40 cm which is the distance between the camera and the document surface. The subsequent iterative scheme would then compute the surface depth map accordingly. Figure 4.5 shows an example of the reconstructed surface of a document page in a thick bound conference

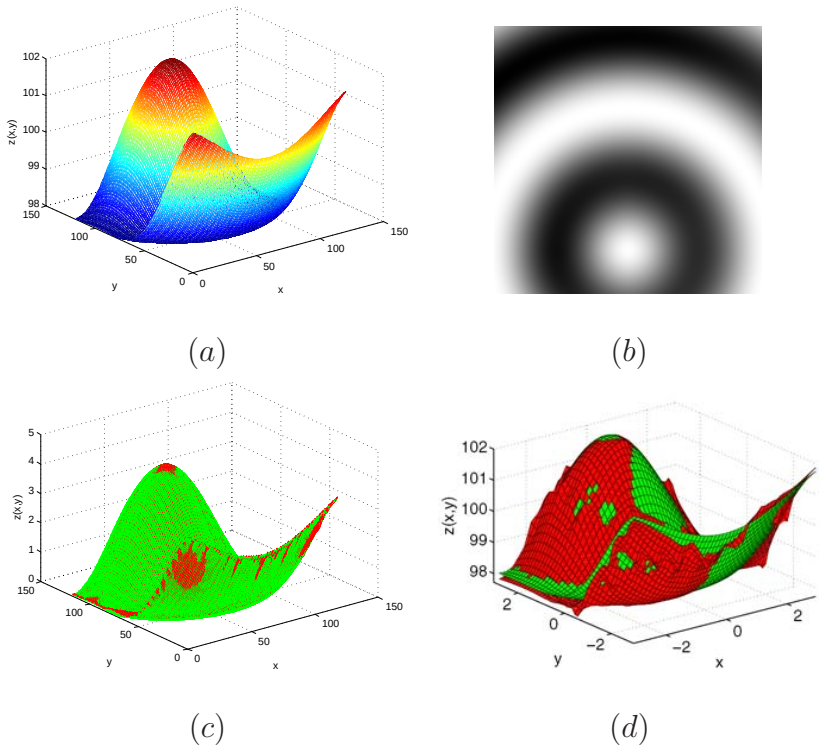


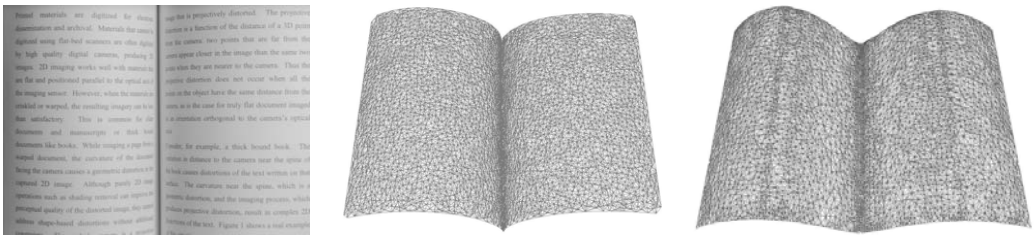
Figure 4.4: Surface reconstruction from a synthetic shading image using the Fast Marching method and its comparison with the result from an existing method: (a) Original surface; (b) Shading image under perspective projection with $L = (0, 0, -1)$; (c) Real (green) vs. Recovered (red) surface by our method; (d) Original (green) vs. Recovered (red) surface by Tankus *et al.*.

proceeding. The image was captured using a normal digital camera with a resolution of 1280×960.

To show how well the surface shape is recovered using the above SFS technique, we compare it with the real shape captured using a 3D scanner. The 2D image used to recover the shape was captured using the 3D scanner for comparison purposes, which has a relatively low resolution of 640×480 compared to normal digital images. In practice, normal digital cameras can capture images of much higher resolution. Therefore, the restored images will produce better OCR results. This is also one of the advantages of the surface reconstruction process, which can recover shapes from higher resolution images

(a) (b) (c)

Figure 4.5: Surface reconstruction result from a real document image using the Fast Marching method: (a) Initial warped page image; (b) Extracted shading image; (c) Reconstructed surface (triangular mesh).



(a) (b) (c)

Figure 4.6: Surface reconstruction result for a real document image using the Fast Marching method and its comparison with the actual 3D scanned surface: (a) Initial warped image; (b) Mesh captured using 3D scanner; (c) Mesh reconstructed using the proposed SFS technique.

without being limited by the low resolution of most current 3D scanners. Figure 4.6 (b) and (c) show a comparison of the range-scanned mesh and the reconstructed mesh for the image in Figure 4.6 (a). As we can see, the reconstructed mesh is quite close to the range-scanned mesh although it does not re-produce the actual depth in the spine area. This is because the dark shadings at the steep spine region provide too few variations to obtain an accurate shape reconstruction result. However, such areas usually do not contain textual information and thus do not bring significant effects to our restoration problem.

4.3.3 Method Evaluation

The main idea of the SFS method described in Section 4.2 is to use an iterative Fast Marching algorithm to solve the image irradiance equation formulated for the document imaging system under a close point light source and a perspective projection model. More intuitively, given a shading image which can be derived from the photometric correction routine, this SFS method will return the corresponding surface shape that produces that shading image under a known illumination condition. This provides one alternative way of obtaining the surface shape when special devices such as 3D scanners or structured lightings are not available. The shading image needs to be a good approximation of the real shading in order to arrive at a good surface reconstruction within a minimum number of iterations. Inaccurate shadings may result in a loss of depth information in the reconstructed surface as illustrated by the example shown in Figure 4.6. Although the reconstructed surface is not as accurate as that captured from the 3D range scanner, it does help to produce better restored images when geometric distortions are concerned.

Different from the one-pass Fast Marching method which propagates the surface one grid at a time in the normal direction with a speed that depends only on the position and is always positive, our image irradiance equation does not transform perfectly to the Eikonal equation that the Fast Marching method is used to solve. Nevertheless, we use an iterative strategy that updates the surface depth value based on the same upwind finite difference approximation scheme at each iteration and therefore improves the surface shape gradually till convergence. The iteration condition can be controlled by the time steps taken or the change of the shape between two consecutive iterations. The main difficulty of this method is to provide a good surface initialization since the

iterative step evolves the shape based on the result in the previous step. To relax this requirement, we propose another method that formulates the image irradiance equation as a static Hamilton-Jacobi equation and solves it using a Lax-Friedrichs based Fast Sweeping method as described in the next section.

4.4 SFS Using Fast Sweeping Method

4.4.1 SFS Formulation

Following the same imaging system as shown in Figure 4.2 with a pinhole camera facing upward, we use $z(x, y)$ to represent the depth value of a surface point $(x, y, z(x, y))$ whose projection is denoted by (u, v) in the image plane Ω. By definition, $z(x, y) \equiv z(u, v)$. Depending on the type of the illumination condition, the image irradiance equation can be formulated in different forms. To be more comprehensive, we discuss two image irradiance equations formulated under two different illumination conditions: a directional light source and a close point light source. In particular, most scanners' uniform light ray can be considered as a directional light source, while most cameras' flash light can be considered as a close point light source.

Directional light source

Given a Lambertion surface, the reflectance map under a directional light source is defined as:

$$I(u, v) = N \cdot L = \frac{(-p, -q, 1)}{\sqrt{p^2 + q^2 + 1}} \cdot \frac{(\alpha, \beta, \gamma)}{\sqrt{\alpha^2 + \beta^2 + \gamma^2}} \tag{4.12}$$

where $I(u, v)$ is the image irradiance at the image point (u, v) corresponding to the surface point (x, y), (α, β, γ) is the illumination direction and $(-p, -q, 1)$ is the surface normal at point (x, y) with $p = \frac{\partial z}{\partial u}$ and $q = \frac{\partial z}{\partial v}$. This is the general image irradiance equation under a directional light source. In particular, if the light source is right on top of the surface with $L = (0, 0, 1)$, the image irradiance equation becomes the Eikonal equation:

$$\sqrt{p^2 + q^2} = \sqrt{\frac{1}{I(u, v)^2} - 1} \tag{4.13}$$

Close point light source

Directional light sources are often difficult to obtain in real-life situations. In fact, it is easier and more practical to capture images using close point light sources such as the on-camera flash than under some specially-built lighting environment. The on-camera flash can be modeled as a close point light source with known position in the camera's coordinate system. More generally, this can be any point light source with a known location in reference to the optical center. Now, Let's represent the document's surface $S(u, v)$ with respect to the image domain Ω as:

$$\left\{ S(u, v) = \frac{z(u, v)}{f} \cdot \ (u', v', f) \quad (u, v) \in \Omega \right\} \tag{4.14}$$

where $z(u, v)$ is the distance from $x - y$ plane to the surface point (x, y), f is the camera's focal length, (u_0, v_0) is the principle component and $u' = u - u_0$, $v' = v - v_0$. This is similar to the surface representation as given in Eq. 4.1.

Again, to find the surface normal at each point (u, v), we calculate the tangent vectors in both u and v directions, respectively and compute their cross product. The downward

normal can thus be derived as:

$$N(u,v) = \left(\frac{pz}{f}, \frac{qz}{f}, -\frac{z(u'p + v'q + z)}{f^2} \right) \tag{4.15}$$

where $p = \frac{\partial z}{\partial u}$ and $q = \frac{\partial z}{\partial v}$. Different from Eq. 4.3, here we work on the surface depth $z(x, y)$ and its gradient (p, q) directly instead of using $\ln z(x, y)$.

Suppose a close point light source is located at (α, β, γ), the illumination direction for each point (u, v) can be written as:

$$\begin{aligned} L(u,v) &= (l_1, l_2, l_3) = (\alpha - x, \beta - y, \gamma - z) \\ &= \left(\alpha - \frac{u'z}{f}, \beta - \frac{v'z}{f}, \gamma - z \right) \end{aligned} \tag{4.16}$$

Using Eq. 4.15 and 4.16 with the assumption of Lambertian reflection, the image irradiance equation can thus be derived from Lambert's cosine law as:

$$I(u,v) = \frac{N(u,v) \cdot L(u,v)}{\|N(u,v)\| \, \|L(u,v)\|} = \frac{n_1 l_1 + n_2 l_2 + n_3 l_3}{\sqrt{n_1^2 + n_2^2 + n_3^2}\sqrt{l_1^2 + l_2^2 + l_3^2}} \tag{4.17}$$

where $(n_1, n_2, n_3) = (pf, qf, -u'p - v'q - z)$.

4.4.2 Lax-Friedrichs-Based Viscosity Solution

To solve the image irradiance equation in Eq. 4.12 & 4.17, we first write it in the form of a static Hamilton-Jacobi equation:

$$\begin{cases} H(u, v, z, \nabla z) = R(u, v), & (u, v) \in \Omega \\ z(u, v) = b(u, v), & (u, v) \in \Gamma \subset \Omega \end{cases} \tag{4.18}$$

where Ω denotes the image plane, Γ denotes a set of points whose value $z(u, v)$ is known to be $b(u, v)$, although they may be located in the interior of Ω.

In the case of a directional light source as given by Eq. 4.12, we have:

$$\begin{cases} H(u,v,\nabla z) = I\sqrt{p^2+q^2+1} + p\bar{\alpha} + q\bar{\beta} - \bar{\gamma} \\ R(u,v) = 0 \end{cases} \tag{4.19}$$

where $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$ is the normalized illumination direction and $\nabla z = (p, q)$.

Similarly, for a close point light source as described by Eq. 4.17, we have:

$$\begin{cases} H(u,v,z,\nabla z) \quad = I\sqrt{n_1^2+n_2^2+n_3^2}\sqrt{l_1^2+l_2^2+l_3^2} + Ap + Bq + C \\ R(u,v) = 0 \end{cases} \tag{4.20}$$

where $A = f(\alpha f - \gamma u)$, $B = f(\beta f - \gamma v)$, $C = -fz(\gamma - z)$.

To solve Eq. 4.19 & 4.20, we use an iterative sweeping strategy [TCOZ03] to solve for $z(u, v)$ with an update formula based on Lax-Friedrichs Hamiltonian [KOQ04] given as:

$$\begin{aligned} z_{u,v}^{n+1} &= \frac{1}{\frac{\sigma_u}{\Delta u} + \frac{\sigma_v}{\Delta v}}(R(u,v) - H(u,v,z,p,q) + \sigma_u u_m + \sigma_v v_m) \\ p = \frac{z_{u+1,v} - z_{u-1,v}}{2\Delta u} \quad & q = \frac{z_{u,v+1} - z_{u,v-1}}{2\Delta v} \\ u_m = \frac{z_{u+1,v} + z_{u-1,v}}{2\Delta u} \quad & v_m = \frac{z_{u,v+1} + z_{u,v-1}}{2\Delta v} \end{aligned} \tag{4.21}$$

where $(\Delta u, \Delta v)$ is the grid size, σ_u and σ_v are artificial viscosities satisfying $\sigma_u \geq \max|\frac{\partial H}{\partial p}|$ and $\sigma_v \geq \max|\frac{\partial H}{\partial q}|$.

In particular, for Eq. 4.19, we let

$$\begin{aligned} \sigma_u &= \max_{u,v,p,q}\left|\frac{\partial H}{\partial p}\right| = \max_{u,v}\{\max\{|I+\bar{\alpha}|, |I-\bar{\alpha}|\}\} \\ \sigma_v &= \max_{u,v,p,q}\left|\frac{\partial H}{\partial q}\right| = \max_{u,v}\{\max\{|I+\bar{\beta}|, |I-\bar{\beta}|\}\} \end{aligned} \tag{4.22}$$

Similarly, for Eq. 4.20, we let

$$\begin{aligned}
\sigma_u &= \max_{u,v,z,p,q} \left| \frac{\partial H}{\partial p} \right| = \max_{u,v} \{\max\{|p_m I_p + A|, |p_m I_p - A|\}\} \\
\sigma_v &= \max_{u,v,z,p,q} \left| \frac{\partial H}{\partial q} \right| = \max_{u,v} \{\max\{|q_m I_p + B|, |q_m I_p - B|\}\} \\
p_m &= \max_{p,q} \left| \frac{\partial H}{\partial p} \right| = \sqrt{f^2 + u'^2} \\
q_m &= \max_{p,q} \left| \frac{\partial H}{\partial q} \right| = \sqrt{f^2 + v'^2}
\end{aligned} \tag{4.23}$$

where $I_p = I\sqrt{l_1^2 + l_2^2 + l_3^2}$.

An iterative sweeping strategy is applied to sweep through the image grid in four alternating directions to update the values of $z(u, v)$ which are initialized with the boundary values $b(u, v)$. After each sweep, height values are re-calculated at the four boundaries where the update formula fails to compute. The complexity of the Fast Sweeping algorithm is $O(N)$ where N is the number of the grid points. Compared to the iterative Fast Marching method, this method is clearly more efficient when the image size is large. On the other hand, we only need to initialize a number of boundary values instead of the whole surface, which makes it easier to apply to real image data.

4.4.3 Shape Refinement

The method presented in the previous section gives us an initial surface reconstruction based on the viscosity solution framework. The resulting surface can be used for the subsequent geometric restoration process and produce a visually satisfactory restored image. This is acceptable if the application is mainly to restore images with sparse graphical content for better visual appearance. However, if higher recognition accuracy is desired for those text-dominant images with less constraint on the computational cost, additional

refinement step can be applied. In this step, we try to apply a least squares method with a regularization term to further improve the shape obtained from the viscosity framework. Meanwhile, the current shape also provides a good initialization for the minimization-based refinement method and thus avoids the problem of being trapped in local minima.

The minimization method is based on the variational SFS formulation discussed in [HB86, CDD03], with the energy:

$$F_1(p,q) = \iint_\Omega \left[I(p,q) - E(u,v)\right]^2 dudv + \lambda_i \iint_\Omega \left[\frac{\partial p}{\partial v} - \frac{\partial q}{\partial u}\right]^2 dudv + \lambda_s \iint_\Omega \left[|\nabla p|^2 + |\nabla q|^2\right] dudv \tag{4.24}$$

where p and q are defined the same as before, $I(p,q)$ is the image irradiance equation defined in Section 4.4.1, $E(u,v)$ is the image intensity, λ_i and λ_s are the integrability and smoothing coefficient, respectively. Similarly, in order to derive the height z from p and q, we use the second energy:

$$F_2(z) = \iint_\Omega \left[\left(\frac{\partial z}{\partial u} - p\right)^2 + \left(\frac{\partial z}{\partial v} - q\right)^2\right] dudv \tag{4.25}$$

To numerically minimize the above two energy functions, we minimize their discrete counterparts. By using forward finite difference approximation to the partial derivatives of p and q, we have the first discrete energy ϵ_1 corresponding to the energy F_1 as follows:

$$\epsilon_1(p,q) = \sum_{(u,v)\in D_\Omega} \left[I(u,v) - E(u,v)\right]^2 + \lambda_i \sum_{(u,v)\in D_\Omega} \left[(p_{u,v+1} - p_{u,v}) - (q_{u+1,v} - q_{u,v})\right]^2 + \lambda_s \sum_{(u,v)\in D_\Omega} \left[(p_{u+1,v} - p_{u,v})^2 + (p_{u,v+1} - p_{u,v})^2 + (q_{u+1,v} - q_{u,v})^2 + (q_{u,v+1} - q_{u,v})^2\right] \tag{4.26}$$

where $I(u,v)$ corresponds to Eq. 4.12 or Eq. 4.17 under different illumination conditions, D_Ω is the discrete domain of Ω. Similarly, the discrete energy ϵ_2 associated with F_2 is:

$$\epsilon_2(z) = \sum_{(u,v)\in D_\Omega} [(z_{u+1,v} - z_{u,v} - p_{u,v})^2 + (z_{u,v+1} - z_{u,v} - q_{u,v})^2] \tag{4.27}$$

The energy ϵ_1 and ϵ_2 are minimized by the steepest descent method with a simple line search with Armijo condition [Fle87].

Note that in order to consider all the boundary points, we need to enforce a Neumann boundary condition on p and q. Suppose (m,n) is the size of D_Ω, we then define:

$$\begin{aligned} &p(0,\cdot) = p(1,\cdot), p(m,\cdot) = p(m+1,\cdot), p(\cdot,0) = p(\cdot,1), p(\cdot,n) = p(\cdot,n+1) \\ &q(0,\cdot) = q(1,\cdot), q(m,\cdot) = q(m+1,\cdot), q(\cdot,0) = q(\cdot,1), q(\cdot,n) = q(\cdot,n+1) \end{aligned} \tag{4.28}$$

Given an initialization of (p,q) from the viscosity solution denoted by $z_1(u,v)$, we first apply an iterative process to find a better configuration $(p,q)_m$ that minimizes ϵ_1 within a certain number of iterations. Next, this new configuration will be used in the evaluation of ϵ_2 in which z is initialized as $z_1(u,v)$. Similar to the previous procedure, we can obtain a new configuration z_m which is the final result.

4.5 Experimental Results II

To evaluate our SFS formulations and the performance of the viscosity framework, we first show several experiments on synthetic surfaces which are either parametric surfaces generated using mathematical functions or geometric surfaces captured from range scans. Synthetic shadings computed using our formula are compared with real shadings and show very similar patterns. The reconstructed shapes based on synthetic shadings are

also shown to be close resemblances of the ground truth shapes. Moreover, we also use several real document images to demonstrate how effectively our SFS method and digital flattening procedures can handle both geometric and perspective distortions in real situations.

4.5.1 Results on Synthetic Surfaces

First, to evaluate the accuracy of our shading extraction routine and the correctness of our image irradiance equation, we used an analytical cylindrical shape to generate a synthetic shading image using Eq. 4.17 and compared it with the real shading image extracted using our inpainting and RBF-based smoothing algorithm. Figure 4.7 (a) is a camera image captured from a real cylinder textured with a document page. The extracted shading image is shown in Figure 4.7 (b). On the other hand, Figure 4.7 (d) is an analytical cylindrical shape used to simulate the real cylinder. Figure 4.7 (e) shows the synthetic shading image generated using the same camera's focal length and the same light source location as the real imaging condition. The shading meshes shown in (c) and (f) demonstrate a high similarity between the reconstructed shading and the synthetic shading. This shows that our shading extraction procedure is fairly accurate and the general image irradiance equation in Eq. 4.17 can be used to model real imaging situations effectively. Finally, Figure 4.7 (g) and (h) show the reconstructed shape based on the real shading image in (b) and the corresponding restored image, respectively.

Next, we have also evaluated the viscosity framework by using three synthetic shading images generated from known parametric surfaces. First, the synthetic vase is generated using the formula provided in [ZTCS99] as shown in Figure 4.8 (a_1). The grid size is set

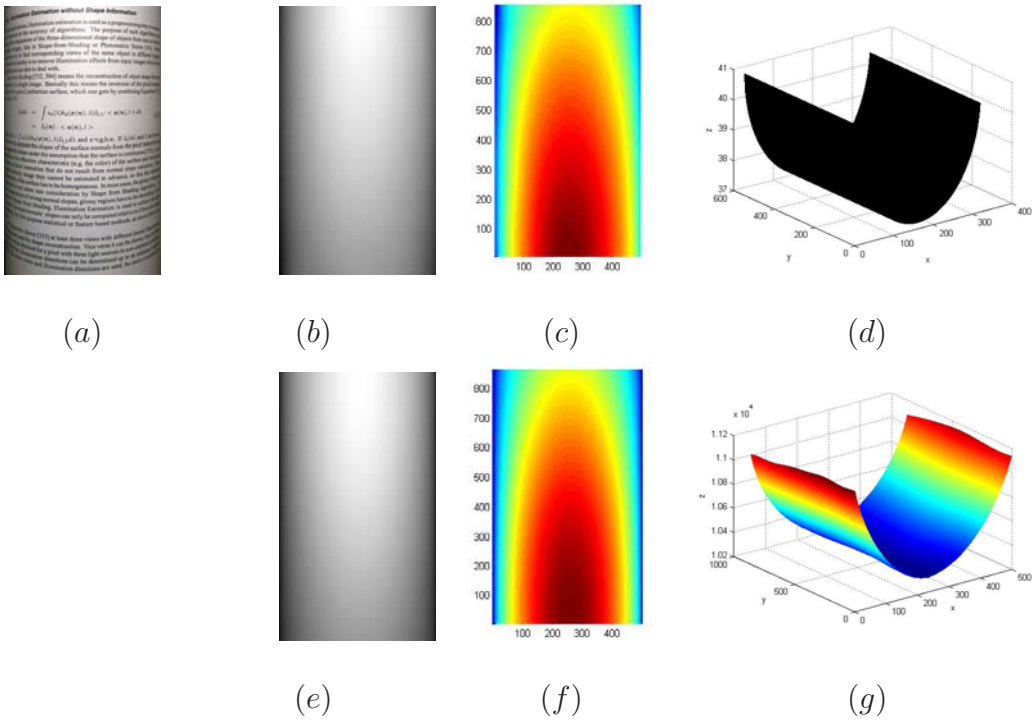


Figure 4.7: Comparisons between the synthetic shading image and the real shading image of a cylindrical shape: (a) Image of a real cylinder with a document texture; (b) Extracted shading image; (c) Color mapped shading mesh of (b); (d) Analytical cylinder; (e) Synthetic shading image produced under the same imaging condition as (a); (f) Color mapped shading mesh of (e); (g) Reconstructed surface shape from (b).

to be $\Delta x = \Delta y = 0.00625$ with an image of size 161×161. The second shape is given by Tankus [TSY03]: $z(x, y) = 2\cos\sqrt{x^2 + (y-2)^2} + 100$ as shown in Figure 4.8 (b_1). The third shape is obtained from [ZZQ04]: $f(x, y) = 2\pi\sqrt{[\cos(2\pi x)\sin(2\pi y)]^2 + [\sin(2\pi x)\cos(2\pi y)]^2}$ as shown in Figure 4.8 (c_1). The shading images are generated based on Eq. 4.13 and 4.12 under a distant frontal light source $L = (0, 0, 1)$ and an oblique directional light source $L = (1, 0, 1)$ as shown in the second and fourth column of Figure 4.8, respectively. In addition, p and q are discretized using the forward difference of the surface height z.

In the case of a distant frontal light source, we use the formulation described in Eq. 4.19

with $(\alpha, \beta, \gamma) = (0, 0, 1)$. By applying the iterative sweeping scheme based on Lax-Friedrichs Hamiltonian, we obtain the reconstructed surface as shown in the third column of Figure 4.8. In particular, the vase surface is initialized with $z = 0$ along the two vertical boundaries. The second shape is initialized with its four boundary values. The last shape is initialized with the five singular points at $(0.25, 0.25)$, $(0.75, 0.75)$, $(0.25, 0.75)$, $(0.75, 0.25)$ and $(0.5, 0.5)$. As we can see that the results are close resemblances of their original surfaces. In addition, we have also tried to apply a high order WENO scheme [ZZQ04] to solve our SFS formulation in the case of an oblique directional light source. This is in comparison with the low order Lax-Friedrich Hamiltonian. More accurate results are obtained as shown in the fifth column of Figure 4.8. The number of iterations and the total time taken to converge to the solution are given in Table. 4.1. The convergence criterion used in our experiments is $\max_{u,v} |z_{u,v}^{n+1} - z_{u,v}^{n}| \leq 0.01$.

Table 4.1: Efficiency evaluation on the three synthetic surfaces shown in Figure 4.8.

Surfaces	First order Lax-Friedrichs scheme				High order scheme	
	Frontal light		Oblique light		Oblique light	
	Iterations	Time (s)	Iterations	Time (s)	Iterations	Time (s)
Figure 4.7(a_1)	17	3.6007	77	19.7733	81	27.6636
Figure 4.7(b_1)	45	5.0929	93	16.3444	192	39.3084
Figure 4.7(c_1)	26	5.1727	177	50.9447	195	73.7268

4.5.2 Comparisons Using Mozart Bust

To compare our results with those of existing approaches, we have used the classic example of Mozart Bust provided by Tsai [ZTCS99]. The true depth map is captured using a

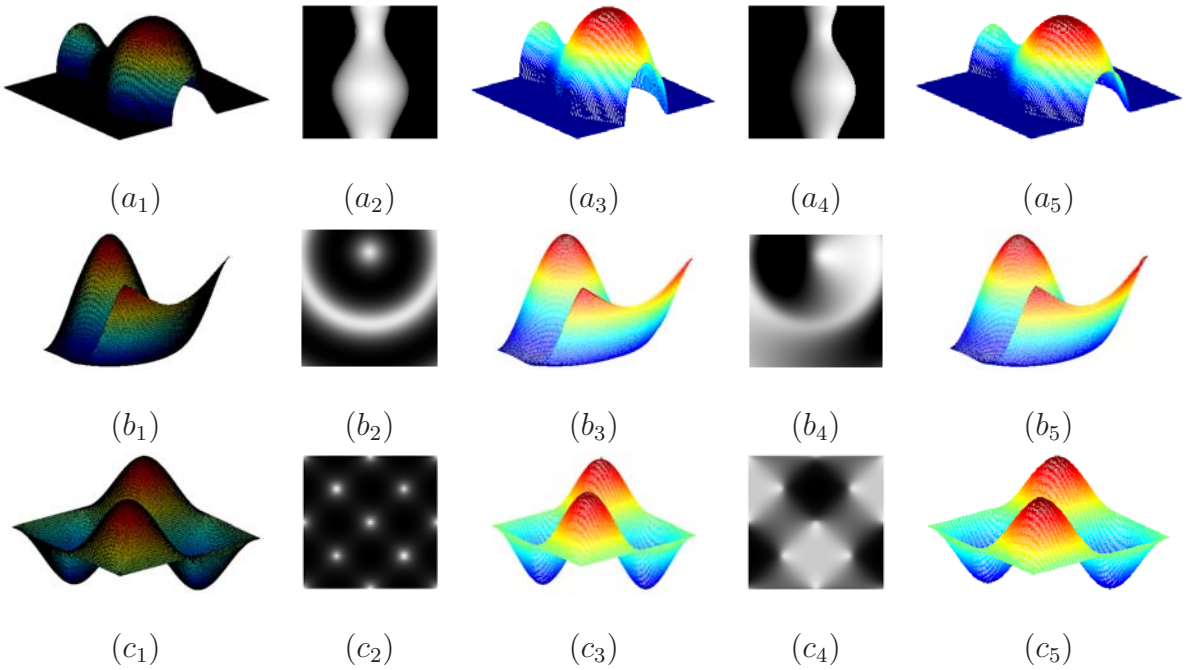


Figure 4.8: Surface reconstruction results on synthetic shading images: $(a_1)(b_1)(c_1)$ Original surface (ground truth); $(a_2)(b_2)(c_2)$ Shading image with a distant frontal light source $L = (0, 0, 1)$; $(a_3)(b_3)(c_3)$ Reconstructed surface based on Lax-Friedrichs Hamiltonian; $(a_4)(b_4)(c_4)$ Shading image with an oblique light source $L = (1, 0, 1)$; $(a_5)(b_5)(c_5)$ Reconstructed surface based on a high order WENO scheme.

range scanner as shown in Figure 4.9 (a). Using the same shading generation method described in Section 4.4.1, we obtain the shading image under an oblique directional light source $(L = (1, 0, 1))$ as shown in Figure 4.9 (b). Figure 4.9 (c) gives the reconstructed shape based on the Hamilton-Jacobi equation solver discussed in Section 4.4.2 with an initialization of the singular point on the nose tip. To evaluate the accuracy of the reconstructed shape, we have measured its absolute distance from the original true depth map. The result shows that most regions are well aligned with an average distance of 1.18 mm. In addition, we have compared our method with all the algorithms reported in [ZTCS99]. Figure 4.9 $(d)(e)(f)$ show the results on three selected approaches. We can see that the current method gives a relatively better reconstruction.

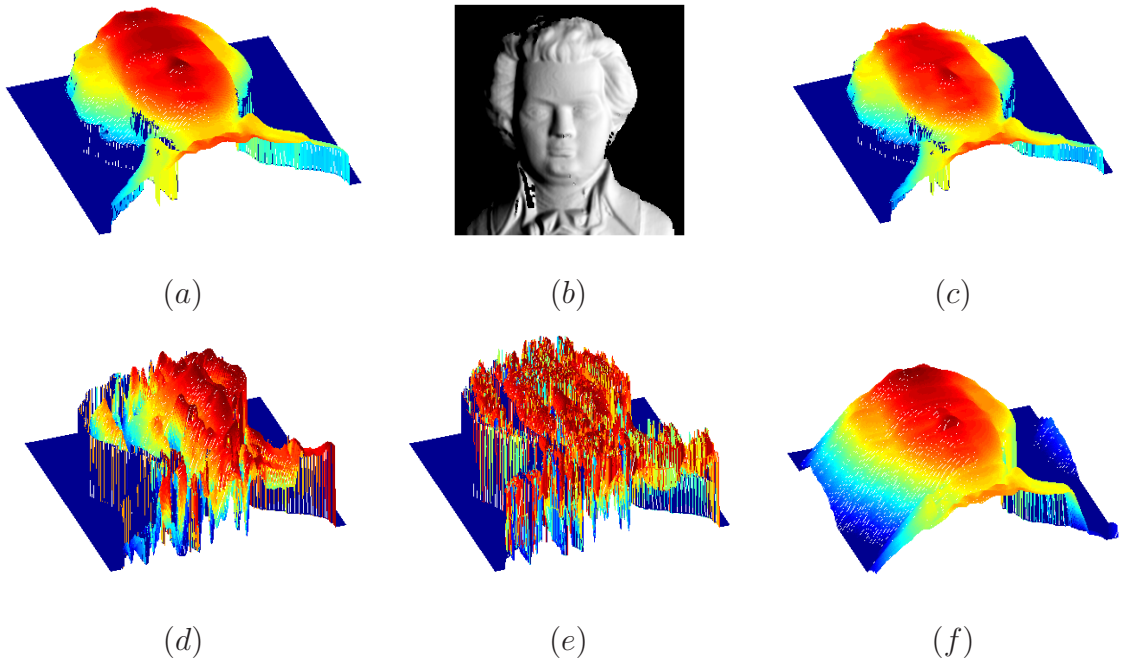


Figure 4.9: Surface reconstruction result on Mozart Bust image and its comparisons with the results from some existing methods: (a) Original depth map of Mozart Bust; (b) Shading image generated with $L = (1, 0, 1)$; (c) Shape reconstructed using the viscosity framework; (d) Shape reconstructed using Tsai's linear approximation method; (e) Shape reconstructed using Lee & Rosenfeld's method; (f) Shape reconstructed using Bichsel & Pentland's method.

4.5.3 Results on Real Document Images

One of the applications of the SFS technique is in the area of DIR [PCF06, CCDG04]. It can be used to reconstruct the surface shape of a warped document and thus provides a priori knowledge to the geometric and perspective restoration procedures. In this experiment, we used both scanned and camera-based real document images to demonstrate how our SFS method works in real situations. In particular, all the scanned images used were obtained from thick bound books using an Epson color scanner. The scanner's uniform light ray is modeled as an oblique directional light source with $L = (1, 0, 5.67)$. The lens follows perspective projection along the spine direction and the focal length is obtained

through a calibration process. On the other hand, all the camera images used were taken in a relatively dark environment with a small halogen light simulating a close point light source. In addition, the camera's focal length and principle components were obtained through a simple calibration procedure. Typically, for an image of size 1600×1200, we have $f = 1286$ and $(u_0, v_0) = (820, 585)$ in pixel size. The images were cropped to avoid lens distortions near the corners. Moreover, gamma correction was performed by applying an inverse power function $1/gamma$ to the RGB values, since most of cameras are calibrated to compensate for the display device with a gamma value.

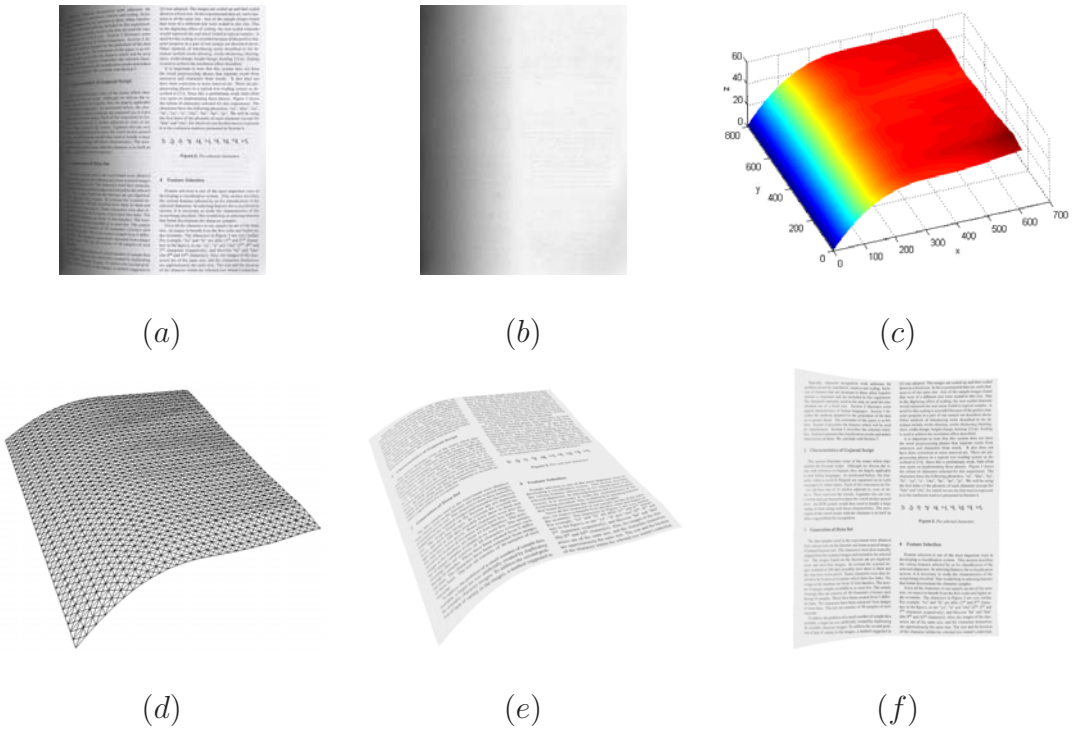


Figure 4.10: Surface reconstruction results on a real scanned document image: (a) Scanned thick bound book image; (b) Shading image obtained after one-step inpainting; (c) Shape reconstructed using the viscosity framework with an oblique light source; (d) 3D mesh of the reconstructed surface; (e) 3D mesh textured with the shading corrected image; (f) Final restored image.

Figure 4.10 shows an example of scanned thick bound document with geometric distor-

tions at the spine region. Because of the constraint imaging environment of the scanner, we have a fixed light source and a frontal imaging view. Images of thick bound books obtained from this typical setup therefore have similar properties. Figure 4.10 shows a typical example of the restoration process and some of the intermediate results. As we can see from Figure 4.10 (d), the reconstructed surface shape is a close estimation of the cylindrical warping of the book spine. Some minor uneven surface patches due to water creases are also detected near the page boundary in accordance to a slight change of shading. Finally, the restored image as shown in Figure 4.10 (f) demonstrates a great improvement over the original distorted image. Because of the constrained scanning environment, the warping distortions in scanned images of thick bound documents display similar characteristics as the typical example shown here. Therefore, this examples gives a representative result when applying our framework on common scanned thick bound document images.

Figure 4.11 shows three examples of geometrically distorted document images with different types of content captured under the halogen light source at a known position. In particular, Figure 4.11 (a_1) is an arbitrarily warped document page captured under a light source at $L = (-25, 0, -34)$. Figure 4.11 (b_1) gives the shading image extracted using the method given in Chapter 3. Finally, Figure 4.11 (c_1) is the reconstructed surface shape by initializing the singular points with an estimated distance from the surface to the camera set as $b_0 = 35.2$. Figure 4.11 $(a_2)(a_3)$ are two warped page images captured from a diagonal viewpoint with the same light source location. Figure 4.11 $(c_2)(c_3)$ show the reconstructed surface by initializing the singular points estimated based on the 2D image. These examples demonstrate that the Lax-Friedrichs-based Fast Sweeping method is able to produce a good approximation of the actual warped document surface, which

will be very useful for the subsequent geometric correction process. We will show the geometrically corrected images for these examples in Chapter 6.

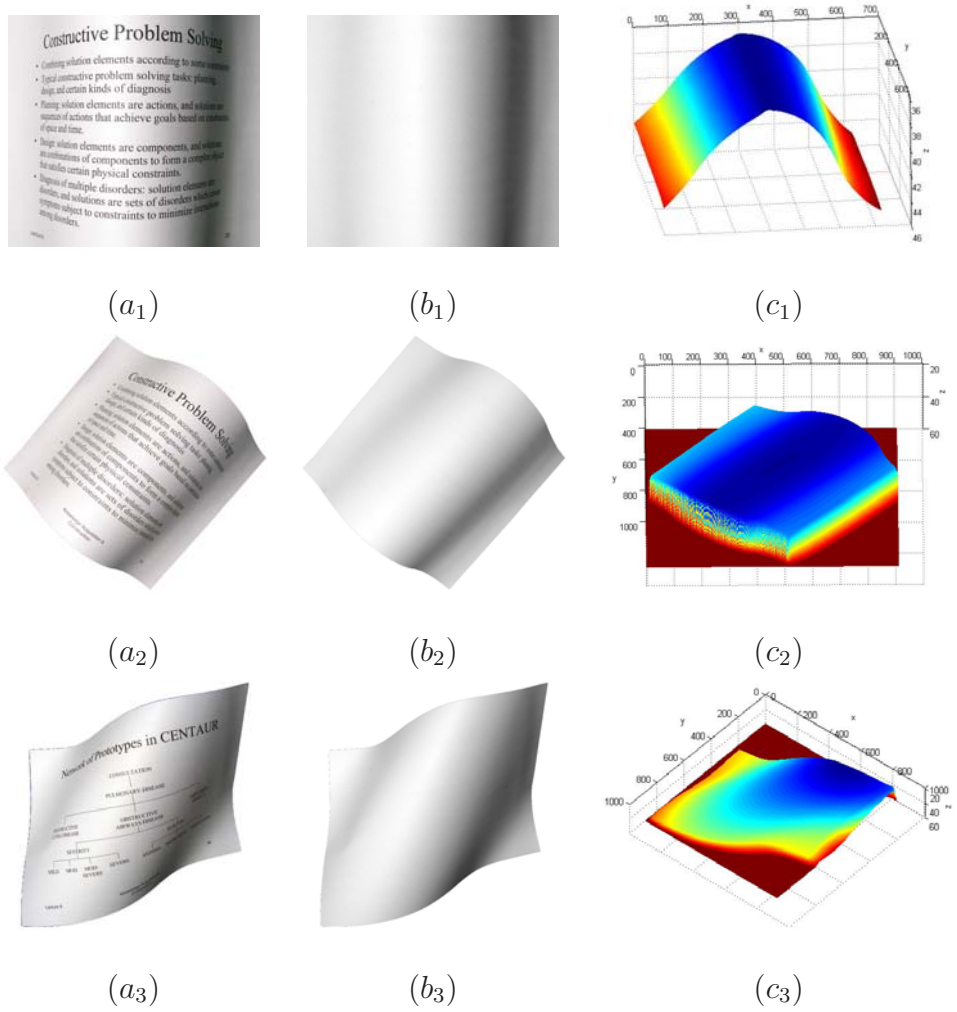
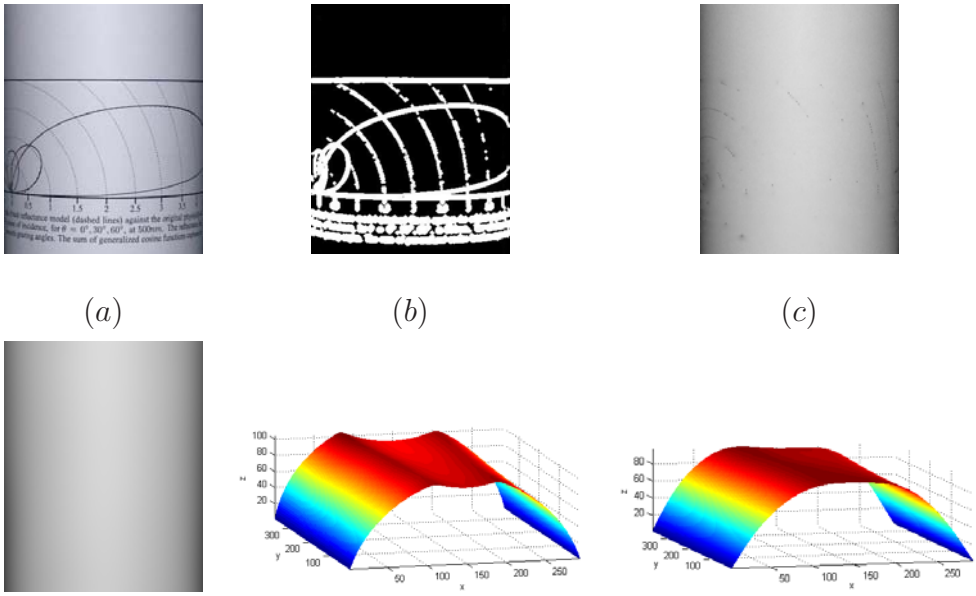


Figure 4.11: Surface reconstruction results on real camera-based document images: $(a_1)(a_2)(a_3)$ Distorted real document images; $(b_1)(b_2)(b_3)$ Extracted shading image; $(c_1)(c_2)(c_3)$ Reconstructed surface shape.

In addition, Figure 4.12 demonstrates how the shape refinement step improves the surface reconstruction result using a warped document page. Figure 4.12 (a) is a page image with a cylinder-like surface captured using a normal digital camera with its on-camera flash. Following the photometric correction steps discussed in Chapter 3, we obtain a smooth shading image as shown in Figure 4.12 (d). Now if we do not know

the exact location of the on-camera flash and we assume it is close enough to the optical center, then using the SFS method discussed in Section 4.4 we obtain the reconstructed surface as shown in Figure 4.12 (e). Due to the inaccurate approximation of the light source location, the reconstructed shape shows abrupt ridges at some regions. After going through the refinement step, the shape as shown in Figure 4.12 (f) becomes much smoother and better approximates the original shape.



(a) (b) (c)

(d) (e) (f)

Figure 4.12: Evaluation of the shape refinement step on real camera-based document images: (a) A real warped document page; (b) Inpainting mask; (c) Inpainted image; (d) Smooth shading image extracted after RBF fitting; (e) Reconstructed surface shape using the viscosity framework; (f) Refined surface shape.

4.5.4 Method Evaluation

Different from the iterative Fast Marching method described in Section 4.2, this method tries to formulate the image irradiance equation into the form of a Hamilton-Jacobi equa-

tion and then solve it using a Lax-Friedrichs-based Fast Sweeping scheme. Moreover, a shape refinement step is also proposed to produce an improved surface by using the reconstructed shape from the first step as the initialization. Our results on synthetic surfaces have shown that the sweeping method based on Lax-Friedrichs Hamiltonian can produce good results on perfect shading images with accurate initialization conditions. However, in real situations, shadings are often imperfect due to several hard-to-control factors such as lighting, surface material, lens distortion, etc. The use of the refinement step can thus further improve the first-step reconstruction result and produce a better surface for subsequent restoration processes. We have conducted experiments on some existing SFS approaches in [ZTCS99], but the results are not comparable since most of them are assuming distant light sources and thus not applicable to our real document imaging conditions. For those considering close point light sources, they are mainly targeting at endoscopic images [TSY04] which is also different from our document imaging environment.

4.6 Discussion

The successful application of the SFS technique to document image restoration (DIR) provides solutions to several problems that traditional 2D-based DIR methods cannot handle. In particular, for the image shown in Figure 4.11 (a_3), those textline-based DIR methods [ZT02, ZT03] will fail to function because of the lack of text lines. Moreover, the SFS-based approaches are not restricted to cylindrical surfaces as opposed to some existing 3D modeling methods [CDL03a, ZZTX05]. In addition, compared to methods that use stereo-based [YKKM04] or motion-based [ISI06] shape recovery techniques, the SFS-based

methods provide a good start in terms of working around a single image instead of multiple images. More importantly, the SFS-based methods are especially of low cost compared to those 3D range scanners or structured lighting setups and are thus more widely applicable to daily applications. However, the proposed two SFS methods currently only deal with smoothly curved documents but not folds as discussed in [SYY+05], so sharp edges are often smoothed out when ridges are expected. Further studies will be extended in this direction.

Comparing the two SFS methods we have presented in this Chapter, the main difference between them is in the formulation of the image irradiance equation and their corresponding solving schemes. The first method uses an iterative Fast Marching scheme. The standard Fast Marching algorithm has a time complexity of $O(N \log N)$ with the dynamic programming strategy, where N is the total number of grid points in the image. The number of iterations taken to get the final result depend on the initialization condition, the complexity of the shape and the accuracy of the estimated shading image. On the other hand, the second method uses a Fast Sweeping strategy which iterates through the image in four different directions to obtain the final depth map. The time complexity is therefore $O(N)$. A computational study of the Fast Marching method and the Fast Sweeping method used to solve the Eikonal equation can be found in [GK06]. In addition, the refinement step is a variational SFS method which tries to minimize a cost function of brightness errors. The computational complexity of this type of method is usually proportional to $O(N^2)$ as discussed in [LK93].

Despite the simplicity and effectiveness of the SFS method, there are still limitations. For example, it handles smooth warpings better than irregular folds or crumples although it is still able to recover the ridges to a certain smoothness level as shown in Figure 6.5

(c_2). Furthermore, due to the importance of the shadings in the reconstruction process, certain requirements on the imaging environment need to be followed such as a single point light source. Interreflections also need to be avoided because they may affect shadings significantly. Last but not least, if large text blocks or figures are masked, the inpainting method may not be able to estimate an accurate shading image and therefore cause errors in the reconstructed surface shape. Nevertheless, we can always substitute the SFS method with any other shape recovery methods such as methods based on Shape-from-Stereo, Structure-from-Motion or even use special setups to capture the surface shape if they are available. In all the cases, the rest components of the whole framework are not affected, which demonstrates the high flexibility and modularity of our framework.

CHAPTER 5

GEOMETRIC CORRECTION

In this chapter, we describe three geometric restoration methods that are proposed to deal with different types of document images with various geometric or perspective distortions. In reference to the categorization of the existing geometric correction methods described in Section 2.3, the first newly proposed method here belongs to the 2D-based shape-free approaches and the other two methods belong to the 3D-based shape-dependent approaches. We first give a brief description of the proposed methods and then move on to the detailed illustrations in the sub-sections.

- The first approach is specially designed to correct warping distortions in camera-based images with predominant textual content, which relies on merely top and

bottom two text lines to perform a 2D interpolation. The idea is based on the observation that the boundary text lines can reasonably reflect the warping curvature to a certain extent. In particular, the text lines are obtained as piecewise cubic splines by fitting a set of connected components boundaries. Splines for document boundaries can be more easily obtained by fitting the boundary pixels. The 2D interpolation methodology is based on the concept of ruled surface interpolation but applied to the projected 2D image. This simple 2D approach can effectively straighten the warped text lines although it may not restore the character spacings at the warping regions due to the lack of 3D information. Moreover, its high dependence on the single column text lines also restricts its application to only a limited set of document images.

- The second approach is a 3D-based method in a sense that it requires the 3D surface shape of the document. This 3D information helps to capture the physical warpings more accurately and therefore achieves a better restoration result especially in terms of the character spacings and sizes near the warping area. The main idea is to incorporate the 3D surface shape information into the first 2D interpolation methodology, which naturally leads to a surface interpolation algorithm to restore the warpings. This method returns better results compared to the first approach especially near the warping regions because the 3D curvature is obviously more accurate than its 2D estimation.

- The third approach is designed for a wide range of document images with its content independent property. Unlike most 2D-based approaches that rely on the existence of text lines or document boundaries, this approach requires nothing but a 2D image and the 3D surface shape of the document. As discussed in Chapter 4, one way to

get the surface shape is to capture it directly through current range scanners which allow us to capture both the 2D image and the 3D geometry easily at one go under any kind of environment. Alternatively, we can also derive the 3D geometry by reconstructing it from the given 2D image. Once the 3D geometry is available, we can then use a physically-based modeling technique to model the 3D document surface as a particle system and flatten it through a numerical simulation process. Meanwhile, the warped 2D image being the texture of the 3D surface is also restored to its planar form accordingly. This approach is therefore able to handle a wide range of document images with various types of content. In addition, our extensive experiments have also demonstrated its great stability and efficiency in handling different types of surface shapes.

5.1 Method 1: Geometric Correction Based on 2D Interpolation

In this Section, we present a document-boundary independent approach to correct arbitrary warping distortions in camera-based document images. This method is based on a Gordon surface model constructed from a set of text lines extracted from the 2D image. The text lines are represented using Natural Cubic Splines interpolating a set of points extracted from connected component analysis on the input document image. For simplicity, only the top and bottom two text lines are used for interpolation by using Coons patch algorithm [Far90]. We have compared our method with two existing methods in terms of the OCR performance on the restored images. One is Tsoi's boundary interpolation method [TB04] and the other is Z.Zhang's 3D modeling approach [ZTF04]. The

results show that our method achieves higher precision than Tsoi's method and meanwhile removes the boundary constraints they imposed on the input images. Our method also outperforms Z.Zhang's method in terms of OCR precision. In addition, Z.Zhang's method is particularly designed for scanned images with cylindrical-like shapes near the book spine and thus not applicable to arbitrary warped images with internal warpings, folds, etc.

5.1.1 Ruled Surface Modeling

Due to geometric distortions such as those in folded paper or thick bound books and the perspective projection model of the camera, many camera-based document images often display non-uniform curvatures along the text lines as shown in Figure 5.1. Current OCR systems often fail to recognize the characters in the warping area or produce text chunks in a wrong text flow sequence. The goal of this method is therefore to rectify the various geometric distortions in those textual document images by straightening the warped text lines and produce a flattened image for better OCR.

As shown in Figure 5.1 (c) and (d), the images do not have explicit document boundaries as those shown in (a) and (b). Nevertheless, we can still construct a ruled surface model based on the text lines present in the document. The surface model for a typical warped thick bound book is shown in Figure 5.2. Given a set of text lines in the warped document and two vertical text boundaries, we can construct a surface g that interpolates all the isoparametric curves $\{g(u, v_i), \quad i = 0, \cdots, m\}$ and $\{g(u_j, v), \quad j = 0, \cdots, n\}$:

$$g(u, v) = g_1(u, v) + g_2(u, v) - g_{12}(u, v) \tag{5.1}$$

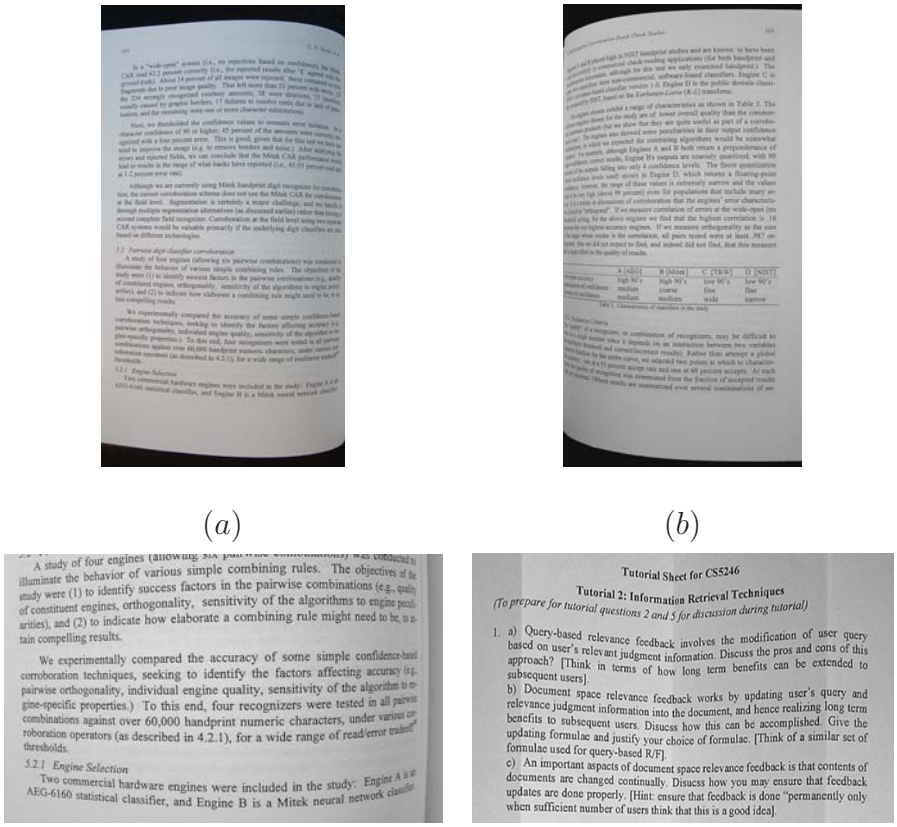


Figure 5.1: Examples of geometrically distorted document images with text-dominant features: $(a)(b)$ Binding curl with document boundaries; (c) Binding curl without document boundaries; (d) Multiple folds.

where

$$
\begin{aligned}
g_1(u,v) &= \sum_{i=0}^{m} g(u,v_i)L_i^m(v) \qquad g_2(u,v) = \sum_{j=1}^{n} g(u_j,v)L_j^n(u) \\
g_{12}(u,v) &= \sum_{i=0}^{m}\sum_{j=0}^{n} g(u_j,v_i)L_j^n(u)L_i^m(v) \\
L_i^m(v) &= \frac{\prod_{j=0,j\neq i}^{m} v - v_j}{\prod_{j=0,j\neq i}^{m} v_i - v_j} \qquad L_j^n(u) = \frac{\prod_{i=0,i\neq j}^{n} u - u_i}{\prod_{i=0,i\neq j}^{n} u_j - u_i}
\end{aligned}
\tag{5.2}
$$

This Gordon surface model fits the definition of developable surface and can be mapped to a planar surface without distortion. The projection of this Gordon surface model can

also be parameterized using the projected text lines in the 2D image.

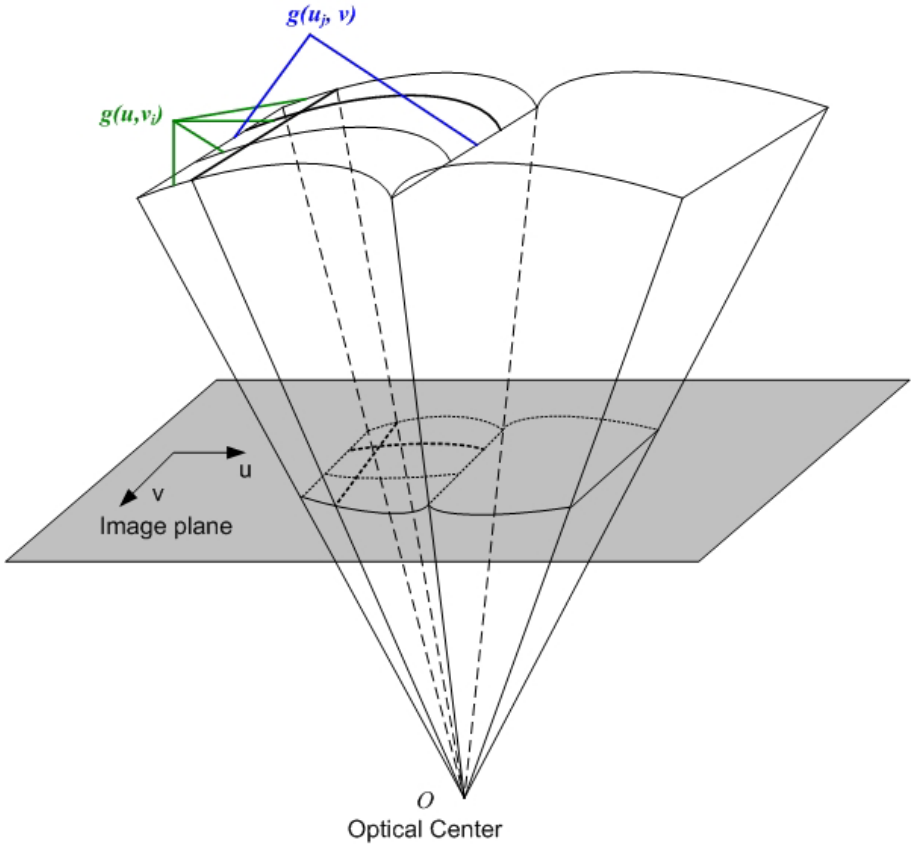


Figure 5.2: Gordon surface model for distorted book surfaces. Its projection image can be parametrized using the projected 2D curves.

5.1.2 Warping Correction

Given an image without explicit document boundaries as shown in Figure 5.1 (c) and (d), boundary interpolation techniques will not be applicable. One remedy to this problem is to make use of the existing text lines. To identify the text lines, we first extract a set of boundary points according to the bottom boundaries of the connected components. Next, for each text line, a Natural Cubic Spline is constructed which defines continuous image

coordinates along the curve. More specifically, given $n+1$ points $(P_0, P_1, \cdots, P_n)$ and $n+1$ knot values $(t_0, t_1, \cdots, t_n)$ on a curve C_i, the spline is defined as a set of piecewise functions:

$$s_i(t) = \begin{cases} s_{i,0}(t), & t \in [t_0, t_1] \\ s_{i,1}(t), & t \in [t_1, t_2] \\ \vdots & \\ s_{i,n-1}(t), & t \in [t_{n-1}, t_n] \end{cases} \tag{5.3}$$

While the $n+1$ points can be easily obtained from the boundary points of the connected components, the knot values need to be carefully chosen. In a simple way, they can be selected based on the 2D Euclidean distance between any two adjacent points, i.e.:

$$t_i = \begin{cases} 0, & \text{if } i = 0 \\ \frac{1}{|d|} \sum\limits_{0<j<i+1} |P_j(x,y) - P_{j-1}(x,y)|_2, & \text{if } i > 0 \end{cases} \tag{5.4}$$

where $|d| = \sum_{0<j<n+1} |P_j(x,y) - P_{j-1}(x,y)|_2$.

For example, Figure 5.3 shows the text lines detected for Figure 5.1 (c). After obtaining a set of splines that represent all the well-identified text lines, the restored image is produced through a 2D mapping function derived from the Gordon surface model. For each point (u_i, v_j) in the restored image, its corresponding point (x, y) in the warped image can be calculated by substituting the values of u_i and v_j to the function $g(u, v)$ to compute the component $x_1(u_i, v_j)$, $y_1(u_i, v_j)$ and $x_2(u_i, v_j)$, $y_2(u_i, v_j)$, respectively:

$$\begin{pmatrix} x(u_i, v_j) \\ y(u_i, v_j) \end{pmatrix} = \begin{pmatrix} x_1(u_i, v_j) \\ y_1(u_i, v_j) \end{pmatrix} + \begin{pmatrix} x_2(u_i, v_j) \\ y_2(u_i, v_j) \end{pmatrix} - \begin{pmatrix} x_{12}(u_i, v_j) \\ y_{12}(u_i, v_j) \end{pmatrix} \tag{5.5}$$

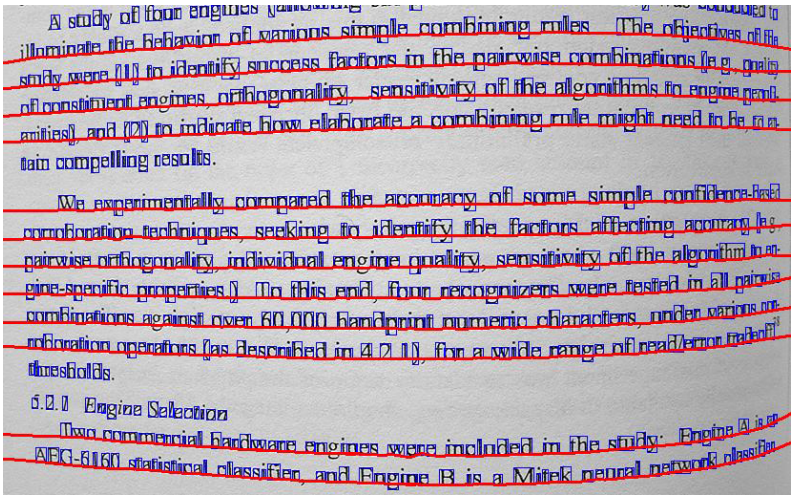


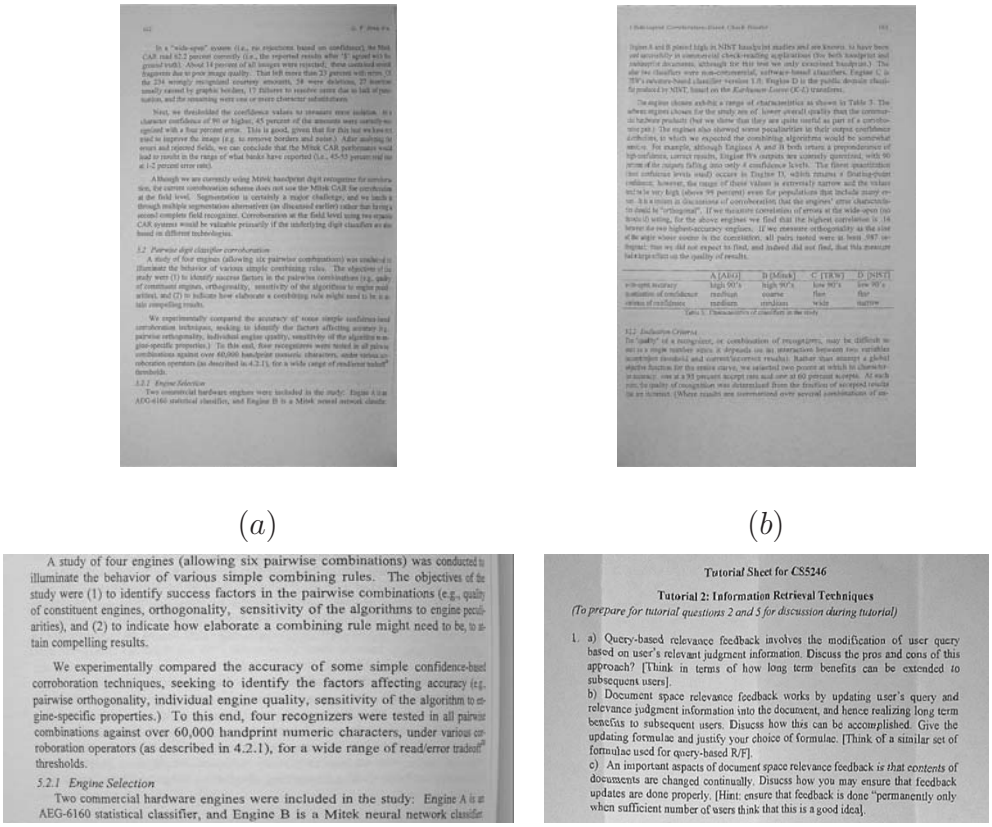
Figure 5.3: Text lines detected for the image shown in Figure 5.1 (*c*).

5.2 Experimental Results I

5.2.1 Results on Real Document Images

We have tested the proposed method on about 20 document images captured using ordinary digital cameras with resolutions of 1600×1200 and 2048×1536, respectively. All the images are of different font sizes and taken with various binding curvatures such as bindings on the left or the right of the image, folds and rolling curls, etc.

In particular, Figure 5.4 shows the restored images of the original distorted images given in Figure 5.1. As we can see, the warped text lines from either smoothly curved documents or multiple folded documents are all straightened out. Due to the shading distortions, the restored images may still look warped. This can be solved by incorporating the photometric correction method as discussed in Chapter 3.



(a) (b)

(c) (d)

Figure 5.4: Geometrically restored images corresponding to the input images shown in Figure 5.1.

5.2.2 Comparisons with Existing Methods on OCR Results

In addition, both the distorted images and the restored images were fed into the Microsoft OCR engine for text recognition. It is shown that after restoration the OCR precision is improved by 13.1% and 21.8% for images with resolutions of 1600×1200 and 2048×1536, respectively. Generally speaking, higher resolution images with larger font sizes produce better OCR outputs. This mainly attributes to the accuracy of the connected component analysis. To compare our method with Tsoi's boundary interpolation method and Z.Zhang's 3D surface modeling technique, we chose a set of images with boundary curves so that all the methods are applicable. A comprehensive comparison of OCR

precision and recall among all the three methods on images with or without boundaries is shown in Table 5.1. The results show that our method achieves much higher precision than Z.Zhang's method and has similar performance as Tsoi's method. However, Tsoi's method requires the presence of document boundaries and thus not applicable to partial images without boundary information as shown in Table 5.1.

Table 5.1: Comparisons of OCR results using three de-warping methods.

OCR accuracy		Original Image	Restored Image		
			Z.Zhang's	Tsoi's	Ours
Images w/ boundary	Precision	73.4%	75.3%	93.6%	95.7%
	Recall	90.1%	93.2%	92.7%	91.3%
Images w/o boundary	Precision	76.7%	78.2%	-	97.5%
	Recall	90.3%	93.3%	-	90.2%

5.3 Method 2: Geometric Correction Based on Surface Interpolation

As discussed in Section 2.3, geometric restorations based on the approximated 2D curvatures do not produce accurate results due to the missing of the depth cue. 3D-based methods are more desirable because the surface shape provides more accurate information about the exact physical warping and thus helps its restoration process. Based on the text line interpolation scheme described in Section 5.1, we propose a hybrid method that incorporates the surface shape information derived from the SFS methods discussed in Chapter 4 into a 2D interpolation algorithm to correct various warping distortions in

textual document images. In our experiments, a set of more than 20 smoothly curved document images were taken using normal digital cameras under a known lighting environment. The restored images show clear improvements especially near the warping area compared to the 2D-based approach described in Section 5.1.

5.3.1 Warping Correction

Similar to the 2D interpolation method discussed in Section 5.1, we still model the document surface as a developable surface. Previously, we have only worked on the projections of the warped text lines in the 2D image without depth information. Now, we could directly manipulate the 3D text lines which provide a more accurate representation of the warping curvature. Following the same procedures, we can extract a set of text lines from the boundaries of the connected components. Next, the depth information is incorporated into the construction of the 3D text lines. Earlier on, the 2D text lines were represented using piecewise splines with the knot values defined on the Euclidean distances of the 2D image points. Here, we choose the knot values differently as the Euclidean distances between the corresponding 3D points on the surface. This is defined as:

$$t_i = \begin{cases} 0, & \text{if } i = 0 \\ t_{i-1} + \frac{1}{|d|}|P_i(x,y,z) - P_{i-1}(x,y,z)|_2, & \text{if } i > 0 \end{cases} \tag{5.6}$$

where $|d| = \sum_{i=1}^{n+1} |P_i(x,y,z) - P_{i-1}(x,y,z)|_2$. This parametrization assigns larger t values to those regions with more severe warpings such as the spine area. This allows more pixels to be allocated to those regions in the restored image and thus bring back the correct font size and character spacing in the warping area. Finally, the restored image can be obtained by mapping each point (u_j, v_i) to its corresponding point (x, y) in the original image based

on the function derived from the ruled surface modeling. For simplicity, we can select the top and bottom two text lines and the two vertical boundaries for interpolation using coons patch which is a special case of the Gordon surface model as discussed in Section 5.1.

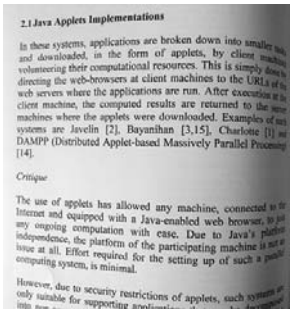
5.4 Experimental Results II

5.4.1 Results on Real Document Images

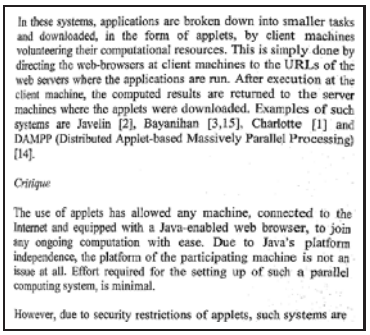
We have tested the proposed method on images captured using a normal digital camera with various resolutions and different warping distortions. Figure 5.5 shows two example images and their corresponding restored images with resolutions 1200×1200 and 748×1064, respectively. In addition, the restored images were also fed into the OmniPage OCR engine for recognition. Table 5.2 shows the OCR precisions on four input images with different levels of distortions as well as the results on their corresponding restored images. Clearly, the precisions on the restored images are much higher than those on the original images.

Table 5.2: Comparisons of the OCR results on both the original and the restored images.

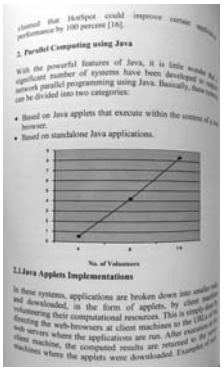
Images	Resolution	OCR word precision	
		Original Images	Restored Images
Image1	1200×1200	89.6%	99.4%
Image2	748×1064	69.4%	91.7%
Image3	1024×768	79.4%	94.5%
Image4	1280×960	84.7%	96.4%



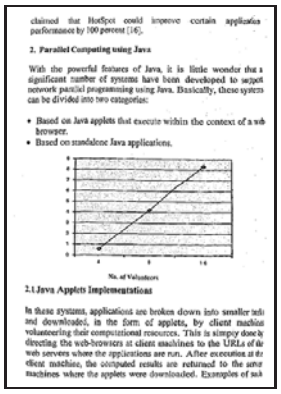
(a)



(b)



(c)

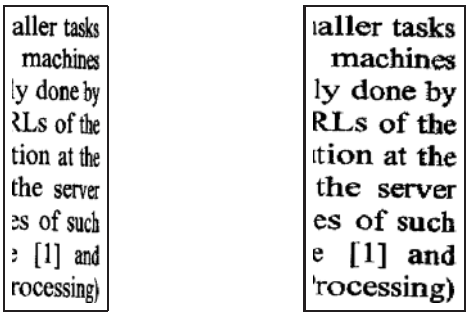


(d)

Figure 5.5: Examples of smoothly warped images with the binarized version of the restored images: $(a)(c)$ Original image; $(b)(d)$ Restored image.

5.4.2 Comparison with the 2D-Interpolation Approach

We have also compared the present method with the earlier attempt based on 2D interpolation. It is observed that the 2D method is unable to restore the character size and spacing near the warping area as shown in Figure 5.6 (a), while the current method does straighten out the text lines and restore the character font size and spacing as shown in (b). Moreover, OCR experiments also show a better performance on the restored images compared to those from the 2D method.



(a) (b)

Figure 5.6: Comparison between the 2D interpolation and 3D interpolation based geometric correction methods: (*a*) Restored image portion using the 2D interpolation method; (*b*) Restored image portion using the 3D interpolation method.

5.5 Method 3: Geometric Correction Using Physically-Based Modeling

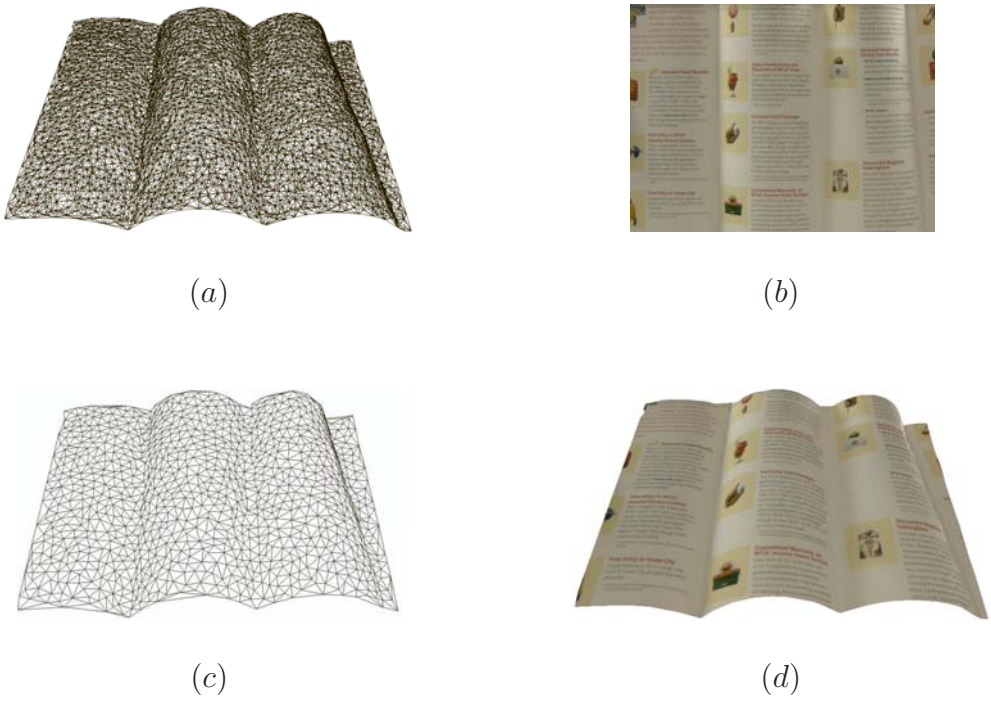
In this section, we present a geometric restoration method that uses the structural information of the physical warping to restore the warped 2D image through a physically-based flattening procedure. As a standalone geometric correction method, the structural information can be obtained either from special devices or through image-based shape recovery methods. To demonstrate the performance of the restoration procedure alone, we use surface shapes captured from a laser range scanner. The laser range scanner can effectively produce an irregular triangular mesh for any kind of shape under any imaging environment. Meanwhile, it also captures a warped 2D image as the corresponding texture mapping to the warped surface. Once the surface shape is available, we can then model the 3D surface mesh as a particle system and flatten it to a plane through numerical methods. Previous work [Pil01b, BS01b, BS04, BP05] has demonstrated that arbitrary warped documents can be successfully restored by flattening a 3D scan of the document.

These approaches used physically-based or relaxation-based techniques in their flattening process. While this has been demonstrated to be effective in rectifying the image content and improving OCR, these previous approaches have several limitations in terms of speed and stability. Here we propose a distance-based penalty metric to replace the mass-spring model and introduce additional bending resistance and drag forces to improve the efficiency of the existing approaches. The use of Verlet integration and special plane collision handling schemes also helps to achieve better stability without sacrificing efficiency. Experiments on various document images captured from books, brochures and historical documents with arbitrary warpings have demonstrated large improvements over the existing approaches in terms of stability and efficiency.

5.5.1 3D Acquisition

To get the 3D surface geometry, we used a dedicated laser range scanner - Minolta Vivid 900 [Viv] to capture the 3D structure of the warped document. The document was scanned by a plane of laser light coming from the Vivid's source aperture. The scanner captures range scan and color data simultaneously and has a sample pitch of 0.068 mm within an optimal measuring range from 0.6 m to 1.2 m. Therefore, it produces an accurate triangular mesh together with a texture image of size 640×480 as shown in Figure 5.7 (a) and (b), respectively. The laser range scanner is able to capture 307,000 points with an output image of size 640×480 in 2.5 seconds. However, working with all the 307,000 points is rather computationally expensive. Our experiments show that using about 1,000 points is enough to produce a satisfactory restoration result. The sub-sampling process was done using the Minolta Polygon Editing Tool, which took approximately 30 seconds to complete on a Pentium III 996 MHz PC. Figure 5.7 (c) and (d) show the sub-sampled

mesh and the textured appearance, respectively.



(a) (b)

(c) (d)

Figure 5.7: Data acquired through a 3D range scanner: (*a*) Original 3D mesh; (*b*) 2D texture image; (*c*) Sub-sampled sparse mesh with 1,010 points; (*d*) Textured 3D mesh.

5.5.2 Particle System Modeling

The restoration is done using a physically-based modeling technique, which incorporates classical Newtonian physics into the behavior of 3D structural models [Wit97, Bar97]. They are often used in computer graphics applications to produce realistic simulations by applying the laws of physics to deform the objects and allowing them to react and dynamically change their geometric structure based on forces in a simulated environment. One particular example is the simulation of the deformation of an initial flat sheet [TPB87]. In this experiment, a flat sheet is dropped and collides with various obstacles during its fall. As a result, gravity and obstacle collision produce forces that cause the sheet to deform.

Our restoration can be considered as the reverse problem. Starting with a deformed sheet which models the warped document surface, we need to reverse the deformation process and move back to its planar state. However, this is only a first-order approximation of how exactly the deformation takes place. It is possible that a complex warped document is produced by a series of irregular external physical events. Therefore, to establish an accurate model to simulate the complex deformation process may be impossible. Our goal is to manipulate the shape in a reasonably accurate and flexible way so that the restored image can have a better perceptual quality and also facilitate some document analysis tasks to a certain extent.

As discussed in Brown and Seales' paper [BS04], given the 3D geometry of the surface, a particle system is constructed with each point on the mesh as a particle. The whole system is governed by the classic second order Newtonian equation:

$$\mathbf{f} = \mathbf{ma} \tag{5.7}$$

where $\mathbf{f}$ is the force exerted on the particle, $\mathbf{m}$ is the mass of the particle (without loss of generality, we can let $m = 1$), and $\mathbf{a}$ is the acceleration. Each particle is associated with two variables, position $\mathbf{x}$ and velocity $\mathbf{v}$. During the simulation process, dynamic forces are exerted onto the particles over time to make them move to new positions. At the same time, distance constraint needs to be satisfied as the property of rigid shape deformation states: any change in shape must preserve the distance between the points on the surface.

To achieve the above constraint, various mass spring models have been proposed [Pro95, Oli01, BS01b]. In these models, particles are connected by virtual springs. The magnitude of the spring force is proportional to the difference between the actual spring

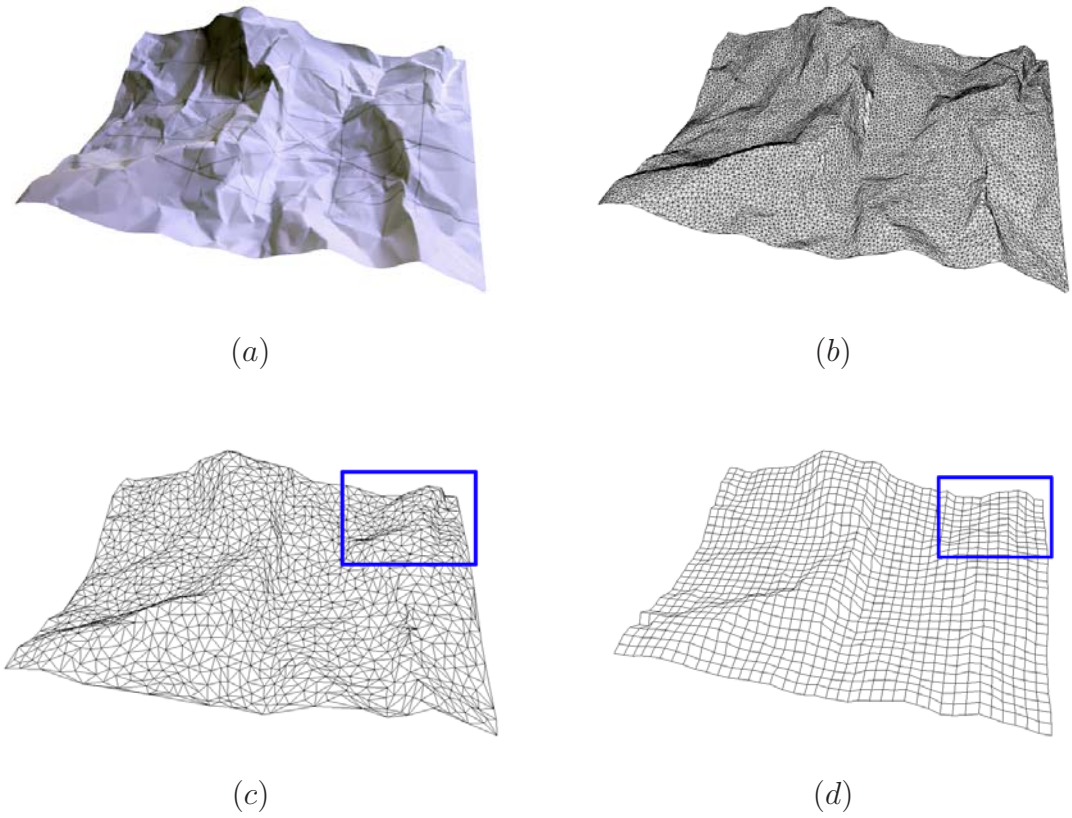
length and its rest length, which can be approximated using Hooke's Law:

$$f_a = -\left[k_s(|x_a - x_b| - r) + k_d \frac{(v_a - v_b)(x_a - x_b)}{|x_a - x_b|}\right] \frac{x_a - x_b}{|x_a - x_b|}$$
$$f_b = -f_a \tag{5.8}$$

where f_a and f_b are the forces exerted on particles a and b respectively, r is the rest length of the spring, x and v are the position and velocity at a particle's current state, k_s and k_d are the spring stiffness and damping constants, and $|\cdot|$ is the L_2 norm. At each time step, the spring forces are induced when external forces begin to move each individual particle from its rest state. These forces act together to keep the particles at a correct distance from each other and give the mesh its rigidity. However, it is often difficult to choose a suitable spring stiffness and damping constant. Strong springs lead to stiff meshes, while weak springs make the meshes too elastic. In our approach, instead of using springs, we use virtual sticks to connect the particles. These sticks can be considered as springs with infinite stiffness. In this case, we can exclude the spring force computation in favor of a point relaxation solver scheme [Jak03].

5.5.3 Mesh Refinement

In Brown and Seales's method [BS04], finite-element meshes are used to represent the document's 3D geometry with each basic finite-element being a quadrangle. This is similar to a quad mesh. In general, triangular meshes are more desirable because three vertices of a triangle solely define a plane but four vertices of a quadrilateral may not lie in the same plane. Therefore, triangular meshes usually provide a better and more accurate representation for the various surfaces that we are dealing with. This was also brought out by Choi and Ko [CK03] in their simulation of complex cloth. Moreover, as mentioned

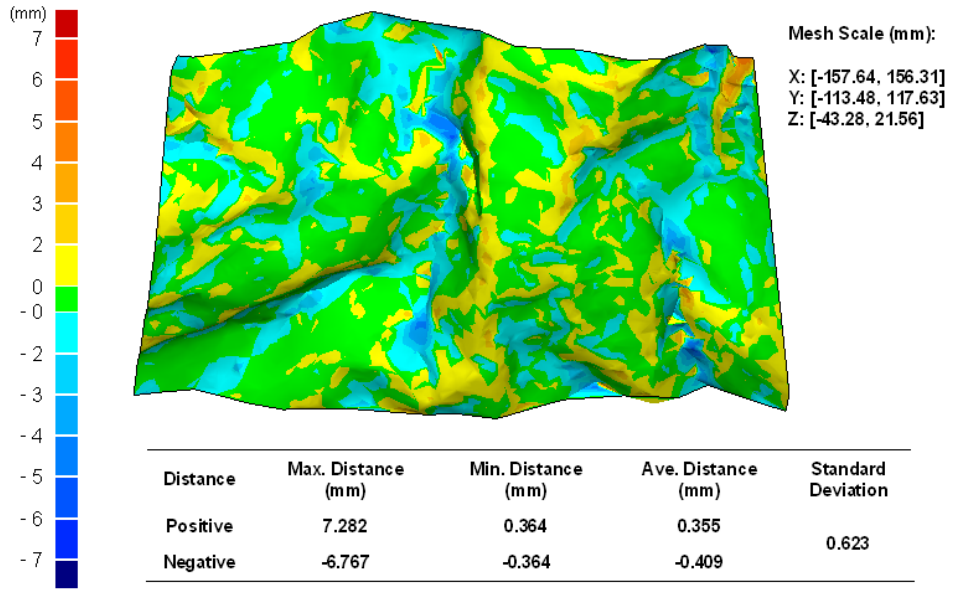


(a) (b)

(c) (d)

Figure 5.8: The shape of a crumpled paper represented using an irregular triangular mesh and a uniform quad mesh (both have 1,200 points): (*a*) Original shape with texture; (*b*) Original dense mesh; (*c*) Irregular triangular mesh; (*d*) Uniform quad mesh.

in Section 5.5.1, in order to improve computational efficiency, we propose to sub-sample the dense mesh into a sparse mesh beforehand. This again makes triangular mesh a better choice than quad mesh. Figure 5.8 shows a comparison of using a triangular mesh and a quad mesh to represent the shape of a crumpled paper. Both meshes have 1,200 points, which is much sparser than the original dense mesh captured by the range scanner. It can be seen from the marked partial image that the triangular mesh is clearly more delicate in representing some complex curvatures than the quad mesh. On the other hand, the sub-sampled sparse triangular mesh with 1,200 points does not degrade much from the original dense mesh. The measured distance between the sparse mesh and the dense mesh



Distance	Max. Distance (mm)	Min. Distance (mm)	Ave. Distance (mm)	Standard Deviation
Positive	7.282	0.364	0.355	0.623
Negative	-6.767	-0.364	-0.409	

Figure 5.9: Distance measure between the sub-sampled triangular mesh and the original dense mesh.

in terms of the depth Z for the crumpled page image as shown in Figure 5.9 is only about 0.38 mm on average and the standard deviation is 0.623 mm with a maximum depth of 64.84 mm.

During our experiments we found that irregular triangular mesh may cause instability problems during the flattening process. The mesh may bend along the common edge of two triangles and result in self-folding effect near some regions with large curvatures. To solve this problem, we propose to add a bending resistance between any pair of triangles that share a common edge as shown in Figure 5.10. This increases the rigidity of the mesh and avoids the self-folding phenomenon. For example, Figure 5.11 (a) shows the result of flattening an original mesh without adding bending resistance. We can see that the mesh near the spine region bends too much and results in self-folding after the flattening process. However, with the added bending resistance, the flattened mesh can hold its shape nicely as shown in Figure 5.11 (b).

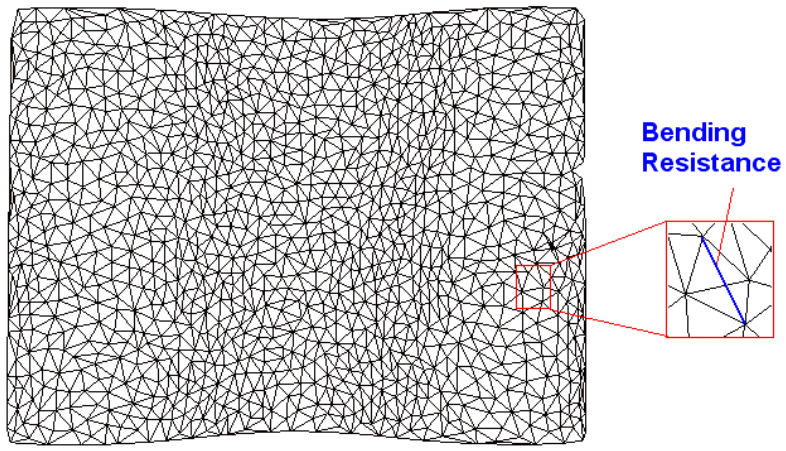


Figure 5.10: Bending resistance is added between any two adjacent triangles that share a common edge.

5.5.4 Constraints and External Forces

Stick Constraint. As mentioned earlier, instead of using springs to connect the particles, we use sticks with fixed lengths. This is to avoid the computation of spring forces which is often error-prune due to the difficulty of choosing a suitable spring stiffness and damping constant. To simulate the sticks between the particles, we use the constraint that enforces a fixed distance between any pair of particles connected by an edge. This distance is defined as the initial Euclidean distance between any two points on the original mesh. During the simulation, the particles are either pushed or pulled away in order to maintain the correct distance that is the stick's rest length as shown in Figure 5.12. Mathematically, suppose A and B are the two end points of a stick with their positions at the time step n denoted as $P_a^n = (x_a^n, y_a^n, z_a^n)$ and $P_b^n = (x_b^n, y_b^n, z_b^n)$. The stick's rest length is defined as:

$$d_{ab}^0 = \sqrt{(x_a^0 - x_b^0)^2 + (y_a^0 - y_b^0)^2 + (z_a^0 - z_b^0)^2} \tag{5.9}$$

where $P_a^0 = (x_a^0, y_a^0, z_a^0)$ and $P_b^0 = (x_b^0, y_b^0, z_b^0)$ are the initial positions of A and B. Due to

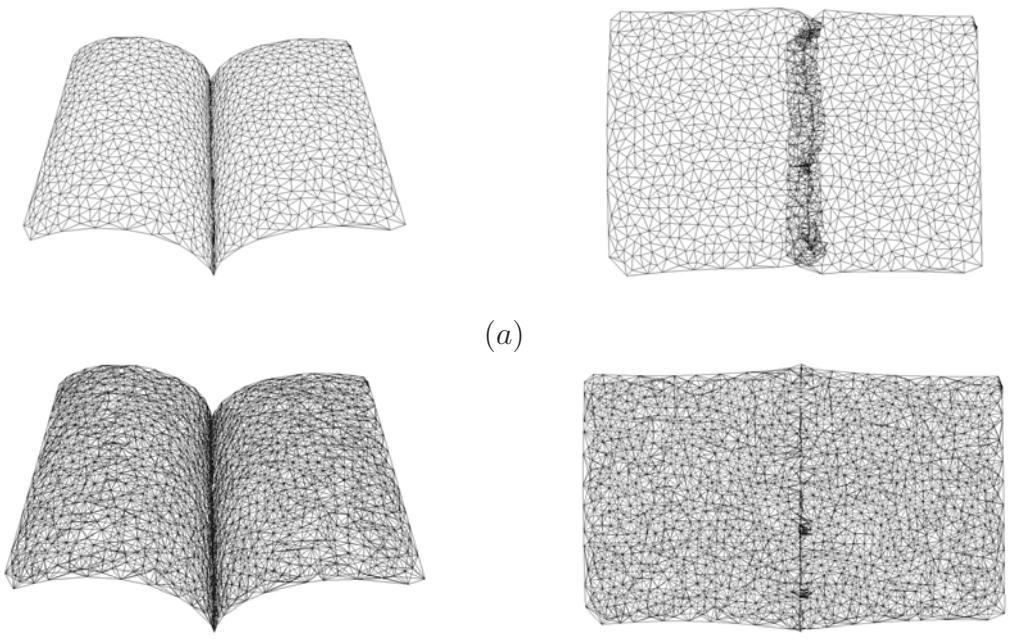


Figure 5.11: Flattening results with and without bending resistance: (a) Flattening without bending resistance results in self-folding; (b) Flattening with bending resistance spreads out the mesh nicely especially near the spine region.

the external forces, A and B will move to positions $P_a^{n'}$ and $P_b^{n'}$ defined as:

$$\begin{aligned} P_a^{n'} &= 2P_a^n - P_a^{n-1} + a\Delta t^2 \\ P_b^{n'} &= 2P_b^n - P_b^{n-1} + a\Delta t^2 \end{aligned} \tag{5.10}$$

To maintain the stick's rest length d_{ab}^0, the two end points A and B are thus moved away or towards each other by the same distance to reach their new positions P_a^{n+1} and P_b^{n+1} at the time step $n+1$:

$$\begin{aligned} P_a^{n+1} &= P_a^{n'} - \frac{0.5(P_a^{n'} - P_b^{n'})(d_{ab}^n - d_{ab}^0)}{d_{ab}^n} \\ P_b^{n+1} &= P_b^{n'} + \frac{0.5(P_a^{n'} - P_b^{n'})(d_{ab}^n - d_{ab}^0)}{d_{ab}^n} \end{aligned} \tag{5.11}$$

where $d_{ab}^n = \sqrt{(x_a^n - x_b^n)^2 + (y_a^n - y_b^n)^2 + (z_a^n - z_b^n)^2}$.

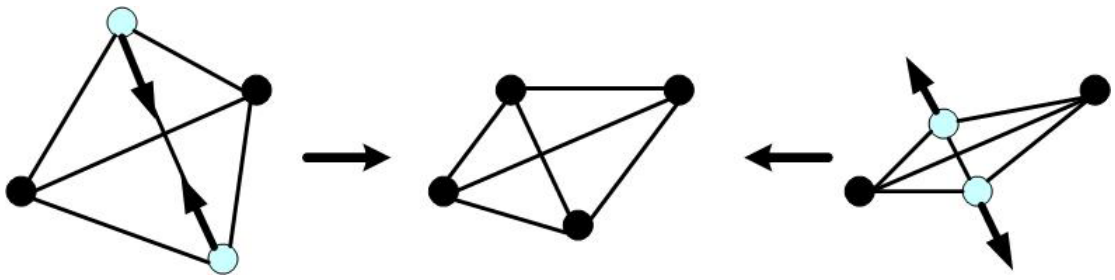


Figure 5.12: The end point particles are pushed or pulled along the stick direction to maintain the stick's rest length during the flattening process.

Plane Collision Constraint. The goal of the flattening process is to drive all the particles on the 3D mesh to the $z = 0$ plane. This plane defines where the warped document is placed during the shape acquisition phase. During the flattening process, some particles might collide with this plane or penetrate through it. In Brown and Seales' work [BS01b, BS04], this situation is handled by changing the velocity of the particle at the time of collision into the normal component and the tangential component. The normal component is then dampened with coefficient of restitution. However, this may cause the simulation to run very slowly when there are lots of collisions and penetrations.

In our method, we introduce a dynamic inverse constraint on the displacement of the particle [Pro95]. Particles that penetrated through the $z = 0$ plane will be projected out [Jak03], i.e. move the particles perpendicularly toward the $z = 0$ plane as defined by:

$$(x_i, y_i, z_i) = (x_i, y_i, \max(z_i, 0)) \tag{5.12}$$

where (x_i, y_i, z_i) is the position of the particle at the time of collision. This method of resolving collision allows a larger time step to be taken during the integration phase and hence speeds up the simulation significantly.

However, by satisfying the plane collision constraint, we might end up invalidating

the stick constraint and vice versa. This can be resolved by solving both constraints simultaneously using a local iteration scheme known as point relaxation [Jak03]. By repeatedly satisfying various local constraints, the whole system will eventually converge to a global configuration that satisfies all the constraints. The system terminates when all the particles have reached the $z = 0$ plane or when the change of the surface area between time t and $t + \Delta t$ is small enough so that further significant improvement is unlikely. In general, the number of iterations usually depends on the particle system being simulated and the amount of motion needed to reach the planar state.

Gravitational Force. In order to drive all the particles down to a plane, we have proposed to exert a downward gravitational force on all the particles to force them to move toward the $z = 0$ plane. This force is a constant since $\mathbf{f} = \mathbf{ma}$, where $\mathbf{m} = 1$ and $\mathbf{a} = \mathbf{g} = 9.81$.

Drag Force. Sometimes using gravitational force alone is unable to correctly simulate the deformation process. This is especially true when the document warping is too severe like the deep spine in some thick bound books. In such cases, the mesh points near the spine region often collide with each other during the flattening phase and results in a self-collapsed spine as shown in Figure 5.13 (a). Therefore, actions need to be taken to avoid such self-collisions and inter-penetrations. Many studies on the self-collision handling problem have been targeted to cloth simulation and garment design [HMB01, VT94]. However, despite the many optimization techniques such as based on the curvature of the surface [VT00] and bounding volume hierarchies [Pro97], self-collision handling is still hard to achieve real-time performance. Nevertheless, our objective here is mainly to obtain a final flattened state of the 3D model. We do not care about the exact deformation at each time frame. This is different from cloth simulation and garment design in which

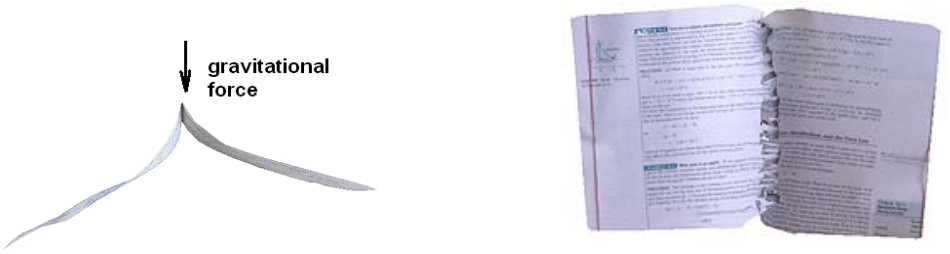
every single detail of the material motion needs to be taken care of delicately at each time step. Therefore, instead of detecting the self-collisions, we try to prevent them from happening. Volino *et al.* [VT94] have pointed out that self-collisions only can happen at areas with large curvatures. As being observed in our experiments, they only happen near the steep spines or folds on the mesh. To avoid particle self-collisions at these regions, we try to add an additional drag force on each particle in both the horizontal and the vertical directions to force the particles spread out during the simulation process. This drag force is similar to the viscous drag used in many particle systems [Wit01] except that it is not to resist the motion but to slightly accelerate the motion as defined by:

$$\begin{aligned} f^h_{drag} &= k_{drag} x^n_i \\ f^v_{drag} &= k_{drag} y^n_i \end{aligned} \tag{5.13}$$

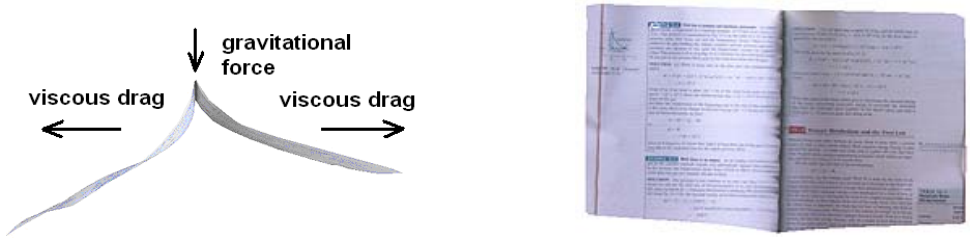
where k_{drag} is a constant known as the coefficient of drag and (x^n_i, y^n_i) is the current 2D location of the particle i at the time step n. An empirical value of k_{drag} is chosen to be 0.05 in all our experiments and does not require special tuning. The origin of the particle system is defined as the center of the 3D shape and is therefore consistent across different inputs. Figure 5.13 (b) shows the geometrically corrected image with both the bending resistance and the drag force added.

5.5.5 Numerical Simulation

As mentioned earlier, each particle is associated with two main variables, position $\mathbf{x}$ and velocity $\mathbf{v}$. Brown and Seales proposed the use of the simple explicit Euler method to handle the particle system [BS01b, BS04]. In each time step, the new position x_{n+1} and



(a)



(b)

Figure 5.13: Geometric correction results with or without drag forces: (*a*) Restored image with gravitational force alone; (*b*) Restored image with additional drag force.

velocity v_{n+1} are calculated as follows:

$$\begin{aligned} v_{n+1} &= v_n + a\Delta t \\ x_{n+1} &= x_n + v_{n+1}\Delta t \end{aligned} \tag{5.14}$$

where x_n and v_n are the current position and velocity, Δt is the time step and a is the acceleration computed using $\mathbf{f} = \mathbf{ma}$. However, as being discussed by Witkin [Wit01], Euler's method can be unstable, inaccurate and inefficient. Due to errors in the formulation, x and v often become asynchronous as the time accumulates. This causes inaccuracy and instability problems and need to be avoided by taking small time steps. Though Euler's method requires only one evaluation per step, the small step size is a bottleneck that restrains its overall performance. Therefore, the use of Euler's method results in a low

computational efficiency. On the other hand, the error introduced by Euler integration is given as $-\frac{1}{2}at^2 + O(\Delta t^3)$, which is much larger than other integration schemes such as Verlet integration and Runge-Kutta integration.

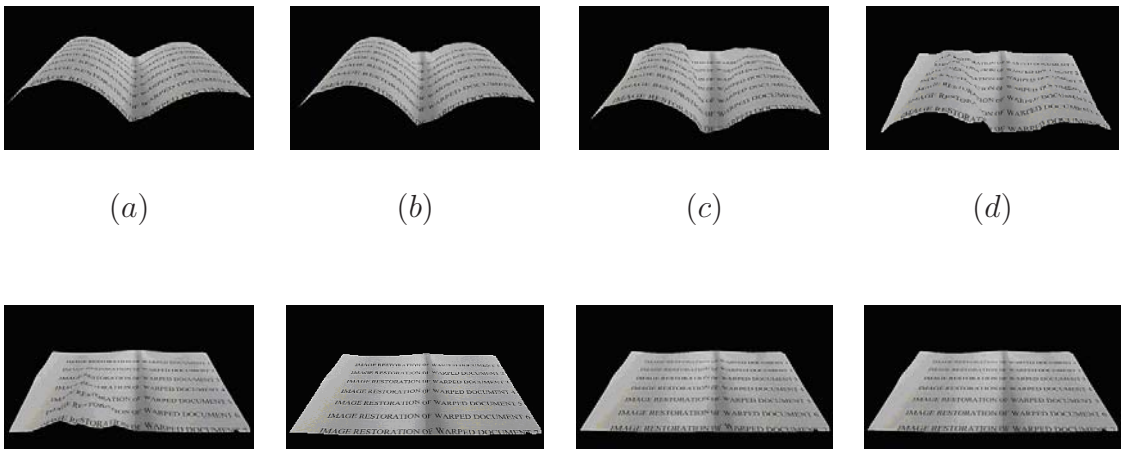
To increase numerical stability and to speed up the simulation, we use the Verlet integration [PFTV86] scheme here. Instead of storing both position and velocity, we only store the particle's current position x_n and its previous position x_{n-1}. The new position x_{n+1} can be calculated as:

$$x_{n+1} = 2x_n - x_{n-1} + a\Delta t^2 \tag{5.15}$$

Verlet integration has been used extensively to simulate molecular dynamics. One example is to create water ripple effect in computer graphics simulation. This scheme is stable because the velocity is implicitly given. Now only the position of the particle need to be taken care of. A large time step can thus be used to speed up the simulation process. In addition, compared with Euler integration, the error produced by Verlet integration is only $O(\Delta t^4)$, which implies a more accurate result. Figure 5.14 shows eight frames of a typical flattening process.

5.6 Experimental Results III

We have conducted experiments on a set of 52 images captured from books and brochures with arbitrary warpings to test the efficiency and stability of the proposed method. The character and word precisions on the restored images were computed and compared with those on the original images. Furthermore, we have also compared our method with Brown and Seales' method using a multi-folded page image. In addition, examples on



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 5.14: Simulation of the flattening process showing eight frames of the textured mesh.

extremely crumpled pages are shown here to discuss the breakdown points of the current method. In all the experiments, the 3D models were acquired using Minolta Vivid 900 by using a mid lens with a measuring distance of about 0.9m. The simulation process was run on a Pentium III 996 MHz PC (512 MB RAM).

5.6.1 Results on Real Images of Books and Brochures

In this experiment, we evaluated the restoration results based on their OCR performance. The images used here were captured from various sources such as thick bound books, advertising brochures and posters, etc. Therefore, they span a wide range of warping distortions such as folding, rolling curl, binding curl, rumples, etc. A total of 52 images were examined. Figure 5.15 shows eight selected images with different warping curvatures and their restored images. For better visualization and OCR performance, shadings in the restored images are removed using the photometric correction method discussed in Chapter 3. Graphical contents as shown in Figure 5.15 (b),(e) and (h) are also restored back to

their normal size and layout. This demonstrates that our method is content independent and can be applied to a large set of documents including those fragile art materials or newspapers to be digitized in many digital libraries. Besides simple warpings such as rolling curls, binding curls and multi-folds, we also tested on randomly crumpled pages with different degrees of rumples and creases. Figure 5.15 (f_1) shows a crumpled document image with large creases but less fine rumples. In order to represent the fine rumples more accurately, we used a relatively dense mesh with about 5,000 points. Figure 5.15 (f_2) shows the restored image.

Next, we fed both the original images and the restored images into the OmniPage OCR for recognition. The word and character precisions are computed and compared in Table 5.3. In general, for a set of 52 images with a total of 12,000 words, the average improvements on word precision and character precision are 18.8% and 16.5%, respectively. Due to the low resolution of the input images captured by the laser range scanner (640×480), the OCR precision tends to be low for some document images with smaller font. This can be further enhanced through a super-resolution algorithm such as [FREM03]. It is also possible that the 3D scanners could incorporate a higher resolution 2D imaging function in the future.

5.6.2 Comparisons with Brown and Seales' Method

In comparison with the geometric restoration framework proposed by Brown and Seales [BS01b, BS04], the current method has made two key improvements in terms of stability and efficiency. Firstly, the current method is able to handle irregular triangular meshes, which better represent some complex shapes compared to quad meshes when they are both

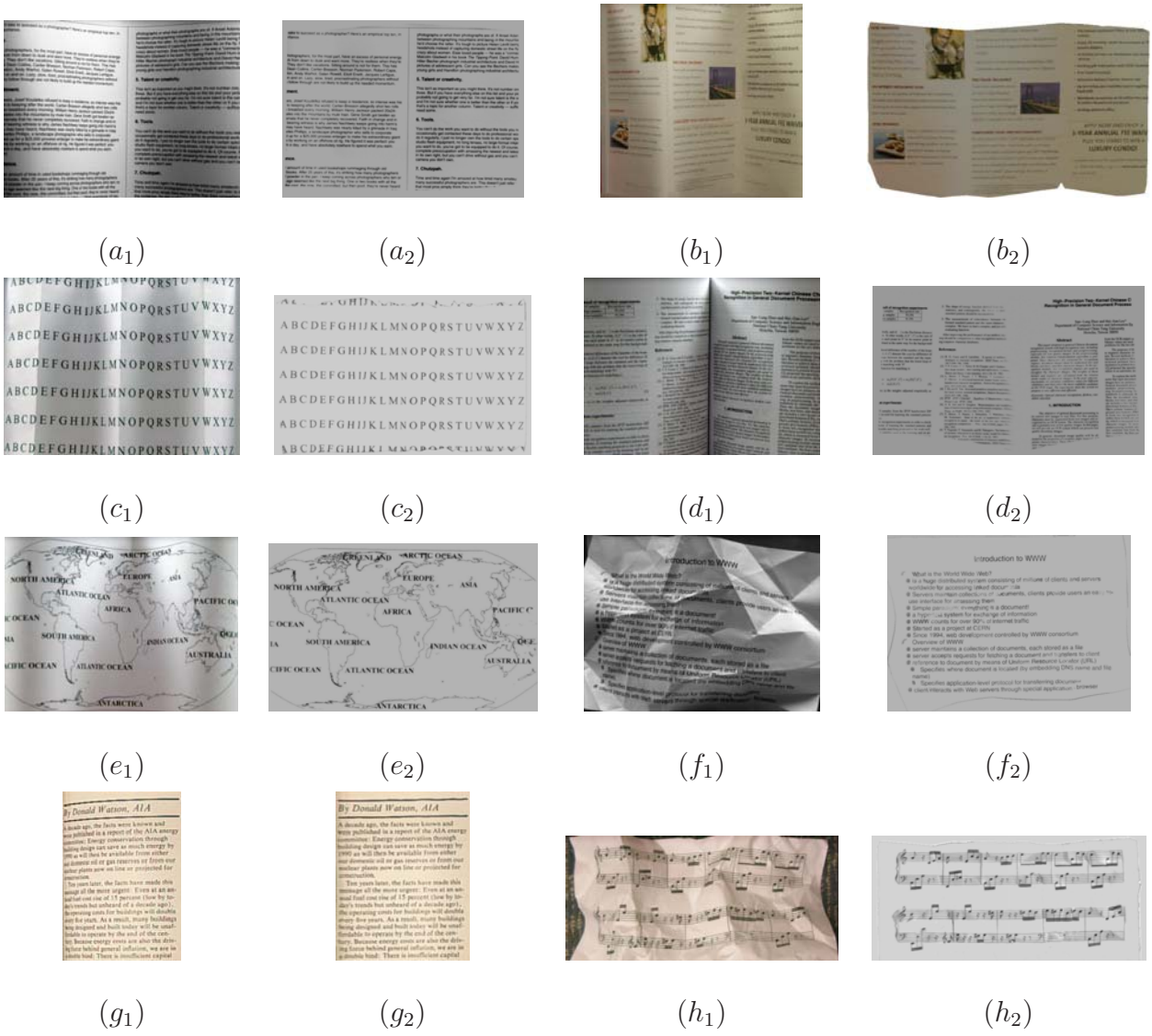


Figure 5.15: Geometric correction results on real images of books and brochures: (a_1-h_1) Original distorted 2D image; (a_2-h_2) Geometrically restored images with shadings removed.

sparse. The use of sparse meshes also helps to reduce the simulation time tremendously. Secondly, by using a stick constraint model, we eliminate the potential instability problems caused by the manually adjusted parameters in the mass-spring model. Figure 5.16 compares the results of a multi-folded paper restored using our stick model and the existing mass-spring model. Figure 5.16 (c) shows the restored mesh using the mass-spring

Table 5.3: Comparisons of OCR results on images shown in Figure 5.15.

Images	OCR Word Precision		OCR Character Precision	
	Original Image	Restored Image	Original Image	Restored Image
Figure 5.15(a)	78.6%	92.3%	83.4%	95.6%
Figure 5.15(b)	64.1%	82.6%	80.5%	91.8%
Figure 5.15(c)	88.7%	100%	88.7%	100%
Figure 5.15(d)	64.9%	85.4%	81.2%	92.4%
Figure 5.15(f)	51.4%	94.3%	57.8%	97.7%
Figure 5.15(g)	83.3%	93.8%	86.4%	96.8%

model with the author's original implementation, which is overall satisfactory except some minor errors at the right boundary where some mesh points overlap with each other during flattening and thus cause some distortions in the restored image. More overlappings are observed at the folds when the mesh is denser and the spring constants are less carefully tuned. It also takes a longer time to reach a stable state during flattening. The current method is able to deal with denser meshes without self-folding problems attributing to the bending resistance and the drag force as described earlier. Figure 5.16 (d) and (f) show the restored mesh and image, respectively. In terms of speed, Brown and Seales' method takes about 50 seconds to converge while our method only needs 3.13 seconds with two meshes containing both 999 points. To show the efficiency of the current method, we have evaluated the simulation time on the eight images shown in Figure 5.15 and compared with that derived from Brown and Seales' method. All the above experiments were run on an Intel Pentium III 996MHz PC (512 MB RAM). Time step used for Verlet integration was 0.1 seconds. The comparison shows that the current method enhances the efficiency significantly (see Table 5.4).

For numerical simulation, stability means the simulation is convergent under the condition of the adopted time step. Therefore, it can be measured quantitatively with the maximal time step a method can take without being divergent. For all explicit numerical simulation approaches such as Euler and Runge-Kutta, they are conditionally stable, i.e., there is a maximum time step above which these approaches will become divergent. Normally Euler and Runge-Kutta approaches need small time steps to maintain stability. The use of Verlet integration in our method allows us to use a large time step without divergence. Meanwhile, the simulation needs very few iterations to reach the equilibrium state and thus achieves computational efficiency as well. On average, the maximum time step our method can adopt is more than 10 times larger than that of Brown and Seales' method. In particular, for the example shown in Figure 5.16, the maximum time step our method can take is 0.32 seconds in contrast to the 0.02 seconds in the existing method under our simulation.

Table 5.4: Evaluation of the efficiency on images shown in Figure 5.15.

Images	Our Method				Brown and Seales'
	Mesh Points	Add BR (s)	Flatten (s)	Total (s)	Total (s)
Figure 5.15(a)	1087	0.25	5.37	5.62	52.35
Figure 5.15(b)	1264	0.27	9.25	9.52	64.36
Figure 5.15(c)	1287	0.28	9.68	9.96	66.42
Figure 5.15(d)	1154	0.26	5.56	5.82	56.67
Figure 5.15(e)	1181	0.26	7.25	7.51	63.78
Figure 5.15(f)	4986	0.34	10.35	10.69	72.43
Figure 5.15(g)	1446	0.29	7.87	8.16	62.54
Figure 5.15(h)	4732	0.33	10.13	10.46	71.07

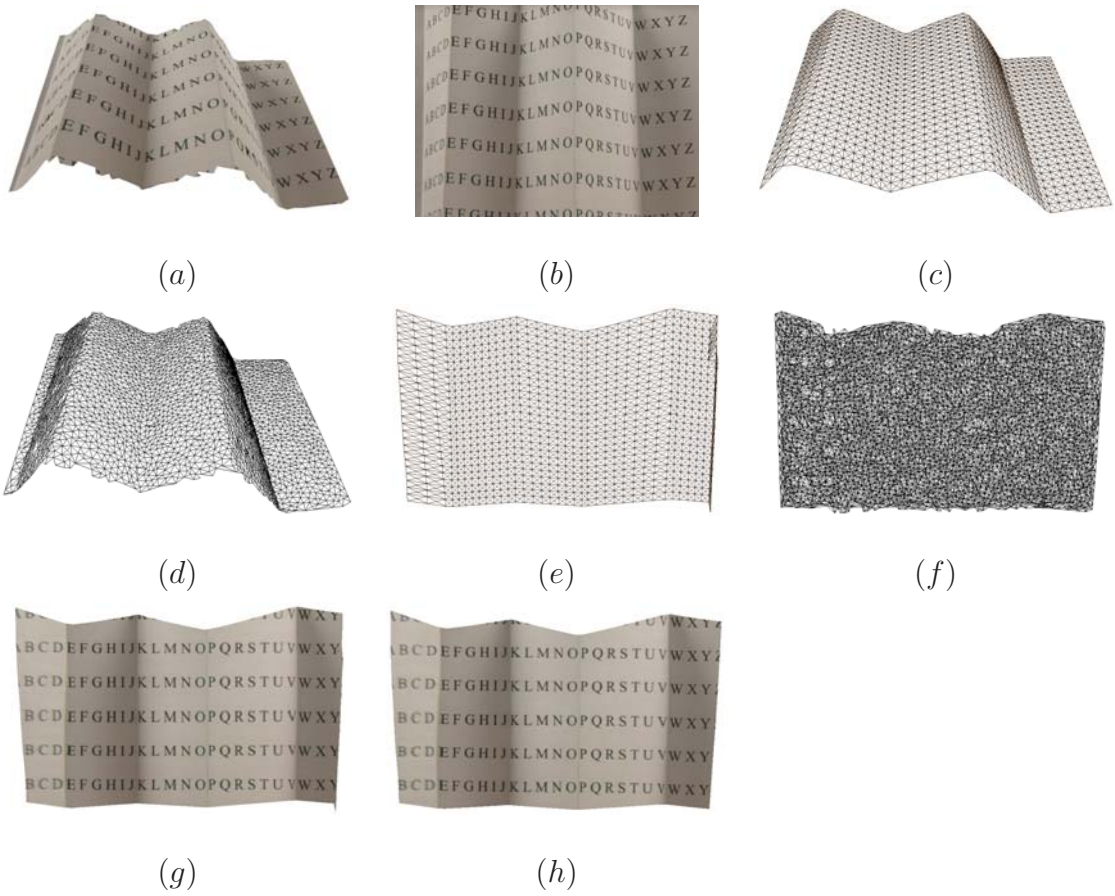
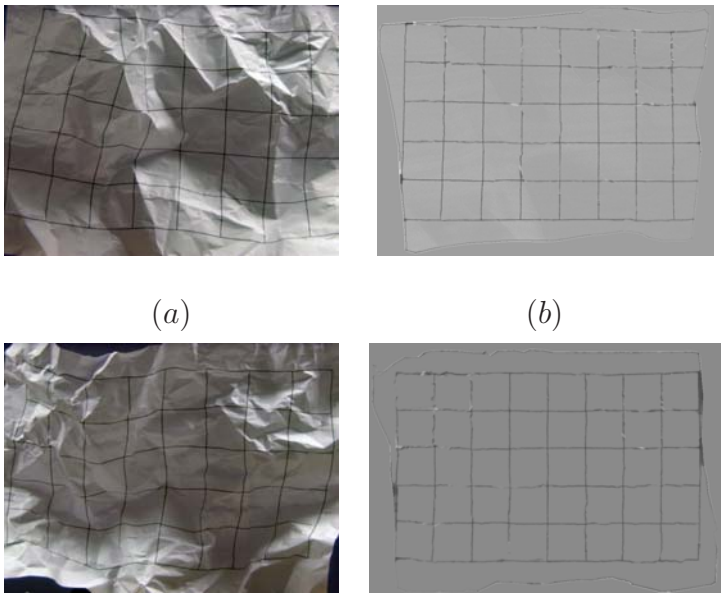


Figure 5.16: Comparison between our method and Brown and Seales' method on the restoration of a folded page: (a) A multi-folded page; (b) Original 2D warped image; (c) Sparse quad mesh with shear springs (999 points); (d) Sub-sampled sparse triangular mesh (2,909 points); (e) Restored mesh using Brown and Seales' method; (f) Restored mesh using our method; (g) Restored image using Brown and Seales' method; (h) Restored image using our triangular mesh with the stick model.

5.6.3 Results on Crumpled Pages

Furthermore, Figure 5.17 shows the results of a crumpled page with more severe rumples. Although there are many fine wrinkles across the page, no interiors are occluded by other parts of the surface. Therefore, the image can still be restored nicely without loss of content as shown in Figure 5.17 (b). However, if we further increase the level of distortion

as shown in Figure 5.17 (c), some parts of the surface will be occluded from the scanner so that the structure and the image are not correctly captured. Therefore, the restored image will lose information at the occluded parts such as the disconnected square edges as shown in Figure 5.17 (d).



(a) (b) (c) (d)

Figure 5.17: Geometric correction results on crumpled pages: (a) A crumpled page with more severe rumples; (b) Geometrically restored image with shadings removed; (c) A severely crumpled image with occlusions; (d) Geometrically restored image with shadings removed.

5.7 Discussion

In this Chapter, we have discussed three geometric correction methods which deal with different types of document images in terms of their content types such as text line styles or other available input information such as the surface shape. The first method is a purely 2D-based method which only requires a single input 2D image with dominant text

line features. The main idea is to estimate the surface warping based on the extracted text lines and use a 2D interpolation technique to correct the warping curvature. This 2D estimation may not be accurate enough especially when the warpings are complicated such as rumples and creases in some crumpled pages. However, this does provide a fast and reliable way of correcting simple geometric distortions in normal textual documents with single-column text blocks. On the other hand, the second method works on the same set of images but provides a more accurate estimation of the surface warping by incorporating the 3D information into the previous 2D text lines. This helps to improve the restoration results especially in the warping regions. The main limitation of these two methods is the requirement on the document content, which confines their application to only a limited set of document images.

Compared to the first two methods, the last method is more general and also content independent. This physically-based method is a 3D-based method that models the 3D geometry of the document surface as a particle system and flattens it to a plane to obtain the final restored image. The advantage of this method is the content independent property which makes this method applicable to a wide range of document images regardless whether they are textual documents or graphical documents. Moreover, this method is fast and robust as long as the given surface geometry is reliable. In particular, some special 3D devices such as the laser range scanner is perfect for this task. Such 3D devices can easily capture the 3D geometry accurately as long as there are no shape occlusions from the view angle. For laser range scanners, there might be limitations in terms of the low image resolution and the inability to capture pure black, blue or green surfaces. These can be avoided by combining structured lighting with normal high resolution digital cameras as discussed in [BS04]. Alternatively, the surface reconstruction component described in

Chapter 4 can also provide a close approximation of the surface shape for the restoration purpose and it can work on a single 2D image so long as the imaging environment is given. On the other hand, the requirement of a reliable surface geometry also becomes the main disadvantage because sometimes the shape of the document may not be easily available especially when the imaging environment is uncontrolled with complex lighting and unknown device parameters. In general, the shape of the document is the key to the success of this restoration method.

CHAPTER 6

OVERALL FRAMEWORK EVALUATION

So far, we have discussed the three main components of the restoration framework in terms of their functionality and the various alternative methods that can be employed to accomplish the specific tasks in each component. For each method presented in the previous Chapters, we have illustrated the detailed methodology used to solve the problem and also evaluated its performance for each specific task using both synthetic and real data set acquired under different experimental setups. In this Chapter, we look at the entire framework as a whole with a selected set of procedures/methods from the available method pool and evaluate its performance based on several benchmarks over the intermediate and final restoration results. First, Section 6.1 describes the set of methods we select from each component to assemble the whole framework. Next, Section 6.2 il-

lustrates the experimental setup and demonstrates how the final restoration results can be affected by each intermediate step using several examples of real document images. Finally, Section 6.3 analyzes and discusses the pros and cons of the whole framework and its potential applications in real-world situations.

6.1 A Typical Assembly of the Framework

Having discussed several alternative methods in each component of the framework, we now select a subset of methods to form a typical assembly of the framework for restoring the three most common distortions in document images namely shading, perspective and geometric distortions. This typical assembly is illustrated in Figure 6.1. In particular, we first use the generic photometric correction method discussed in Chapter 3 to remove the shading distortions in the input distorted image. Meanwhile, a shading image is constructed by smoothing the inpainted background layer image to a RBF fitted surface. Next, suppose a 3D range scanner is not available which is often the case due to its high cost, we then try to reconstruct the document surface using the Shape-from-Shading method based on Lax-Friedrichs update and the Fast Sweeping scheme with a known close point light source such as the on-camera flash as discussed in Section 4.4. Depending on the quality of the extracted shading image and the accuracy of the known light source location, we can then decide whether a shape refinement step is necessary. Finally, once the document's surface shape is obtained, we can map the shading corrected image to the reconstructed surface shape based on the relationship between the real-world coordinates and the image coordinates defined as:

$$\frac{x}{u} = \frac{y}{v} = \frac{z}{f} \tag{6.1}$$

where (x, y, z) is the real-world coordinates of the a surface point and (u, v) is the corresponding image coordinates of its projection in the image plane. Finally, we can apply the physically-based modeling approach to model the reconstructed document surface as a particle system and flatten it to a plane through a numerical simulation method as discussed in Section 5.5. Different from the original experimental setup where a 3D scanner is used to capture the 2D image and the 3D geometry simultaneously with a known image-to-shape mapping, we now have to map the image to our reconstructed surface explicitly before the flattening step. In addition, instead of using an irregular triangular mesh given by the 3D scanner, we construct a quad mesh to approximate the reconstructed surface. This does not affect much of the restoration accuracy since we are mainly focusing on documents with large warping curvatures instead of those with rumples and creases as discussed in Section 5.6. This last step will give us the final geometrically restored image by flattening the shading corrected image attached to the warped surface to its planar state.

6.2 Experimental Results

The typical assembly of the restoration framework that we discuss here is mainly for document images captured in a daily environment using normal digital cameras. The method used in each component does not require any additional input information except a single input image captured using a calibrated digital camera with a known light source location which can be the location of the on-camera flash with respect to the optical center. Based on these assumptions, we have conducted extensive experiments to evaluate the overall framework using real document images with various perspective and geometric

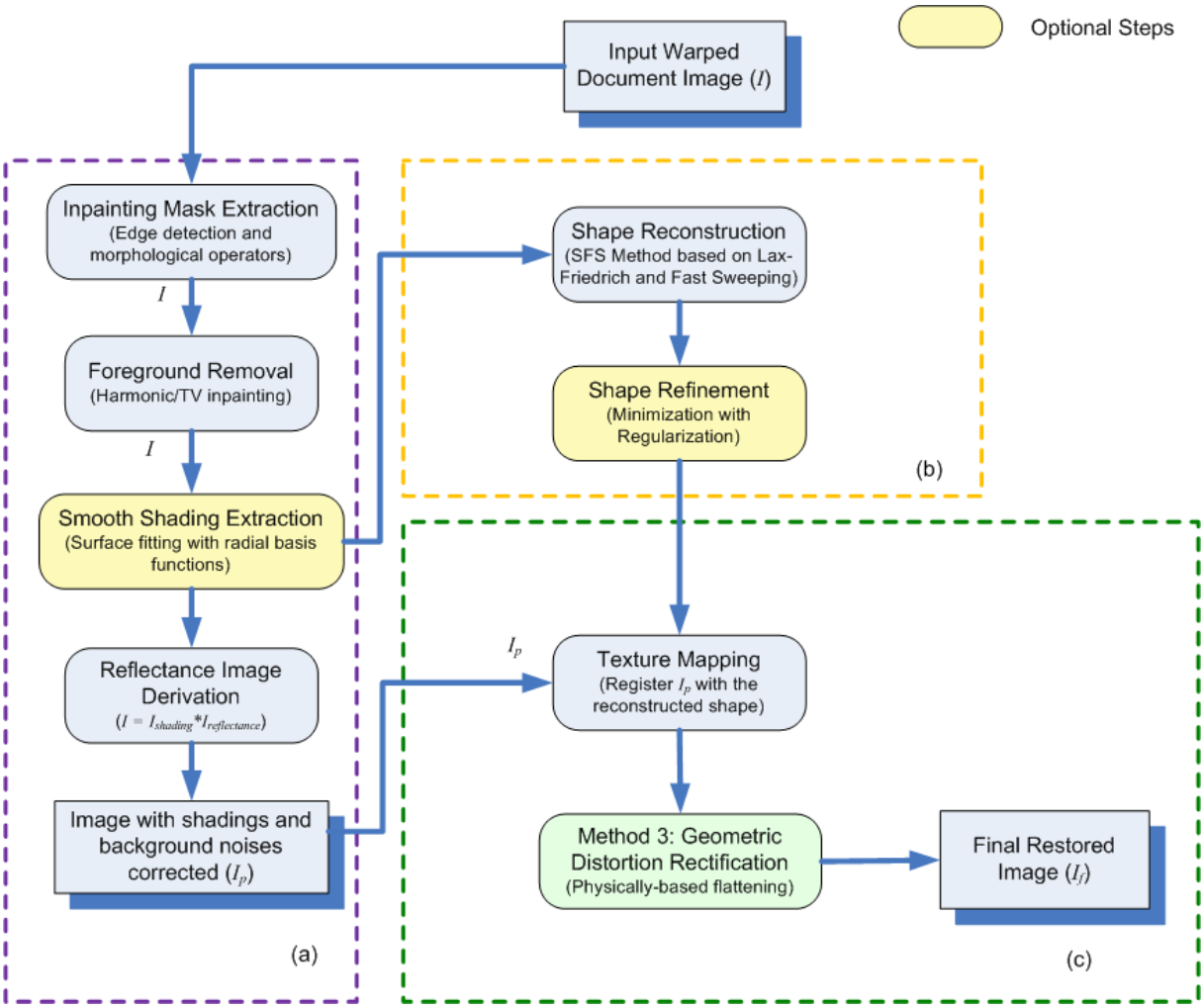


Figure 6.1: A typical assembly of the restoration framework: (a) Shading correction; (b) Surface shape reconstruction; (c) Geometric and perspective correction.

distortions under different orientations.

All the images in our experiments were taken in a relatively dark environment. The light source used was either the on-camera flash or a small halogen light both simulating a close point light source. The location of the light source was measured with respect to the camera's optical center. The camera's focal length f and the principle component (u_0, v_0) were obtained through a camera calibration process. Typically, for an image of

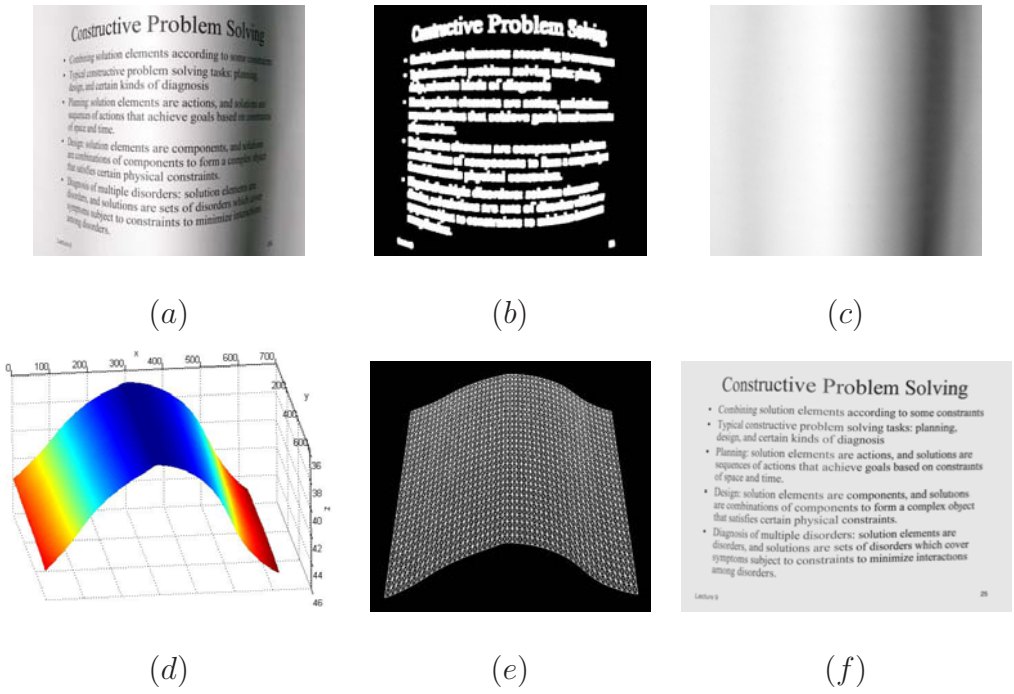
size 1600×1200, we got $f = 1286$ and $(u_0, v_0) = (820, 585)$ in pixel size. The images were cropped to avoid lens distortions near the corners. The shading image was often normalized to a certain dynamic range for effective SFS processing.

6.2.1 Images with Shading and Geometric Distortions

Figure 6.2, Figure 6.3 and Figure 6.4 show the restoration of three real document images with both shading and geometric distortions under different orientations. The light source used here is a small halogen light source located at position $L = (-25, 0, -34)$ with respect to the camera's optical center. For each image, we show the intermediate result of each restoration step. First, we show the inpainting mask and the extracted shading image. From image (c) of Figure 6.2, 6.3 and 6.4, we can see that the extracted shading image reflects the illumination change nicely and is well separated from the reflectance image. A good shading image definitely accounts for a more accurate reconstruction since shading is the sole information used as the SFS input. Then we show the reconstructed surface shape based on the extracted shading image which is also given in Chapter 4. Next, we present the uniformly sampled mesh. Finally, we give the final restored image obtained by flattening the mesh with the shading corrected image as the texture.

6.2.2 Images with Mixed Distortions

Figure 6.5, Figure 6.6 and Figure 6.7 show three examples of real camera-captured document images with pure perspective and shading distortions or mixed perspective, geometric and shading distortions. The light source used here is the on-camera flash located at position $(-1.8, -2.4, 0)$ with respect to the camera's coordinate system. For each image,



(a) (b) (c)

(d) (e) (f)

Figure 6.2: Overall framework evaluation on real document images - Example 1: (a) Distorted real document images; (b) Inpainting mask; (c) Extracted shading image; (e) 3D mesh of the reconstructed surface shape; (f) Final restored image with all the distortions removed.

we also demonstrate the intermediate result from each restoration procedure. In particular, Figure 6.5 (a) is a planar page image with pure perspective distortions. Figure 6.5 (b) and (c) are the inpainting mask and the extracted shading image, respectively. The reconstructed shape shown in Figure 6.5 (d) is obtained by initializing the bottom of the page to be at a distance of $b_0 = 10$ to the camera. The black background is computed as faraway from the camera. Next, Figure 6.5 (e) shows the uniformly sampled quad mesh. Finally, Figure 6.5 (f) gives the restored image with all the distortions removed. In addition, Figure 6.6 and 6.7 show that the proposed framework is able to handle images with all three types of distortions namely shading, perspective and geometric distortions effectively.

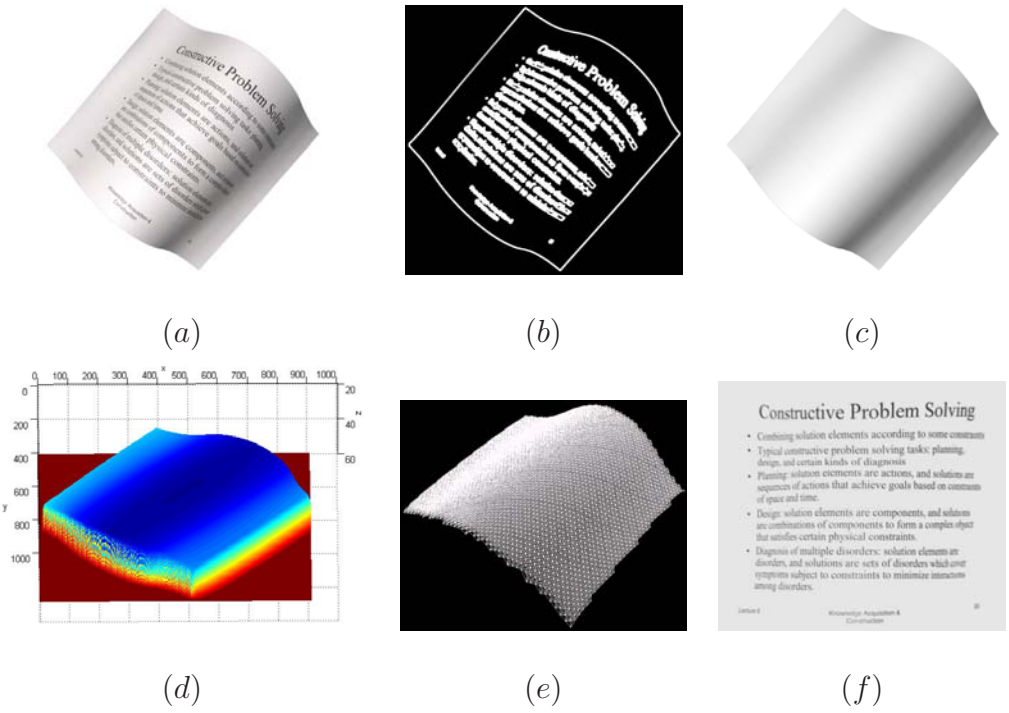
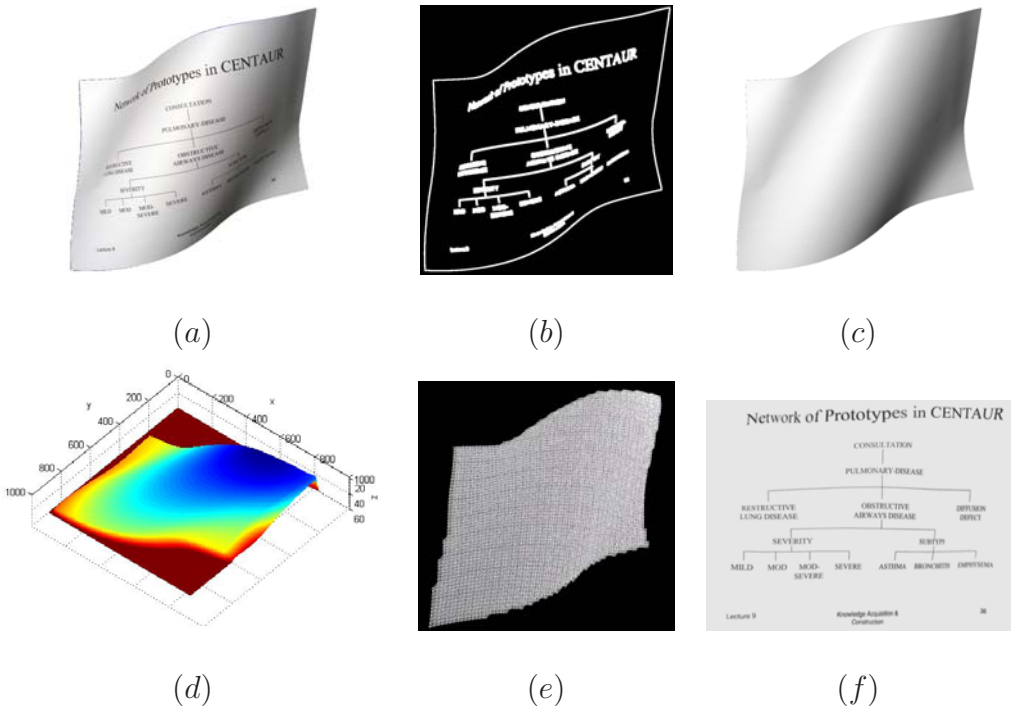


Figure 6.3: Overall framework evaluation on real document images - Example 2: (a) Distorted real document images; (b) Inpainting mask; (c) Extracted shading image; (e) 3D mesh of the reconstructed surface shape; (f) Final restored image with all the distortions removed.

6.2.3 OCR Results

All the above examples show that the restored images are much better improved compared to the original images although some distortions still remain due to the immaculate surface reconstruction based on irregular real-world shadings. Nevertheless, the OCR performance is greatly improved on the restored images with an average increase of 48.6% in terms of word precision over a set of 30 textual document images with approximately 3,500 words. Table 6.1 gives the word precision P_w and character precision P_c on the six examples shown in Figure 6.2 to 6.7. One thing worth mentioning is that Figure 6.7 does not achieve a very good OCR performance even after the restoration. This is because the geometric and perspective distortions have caused a servere resolution loss at the warping



(a) (b) (c)

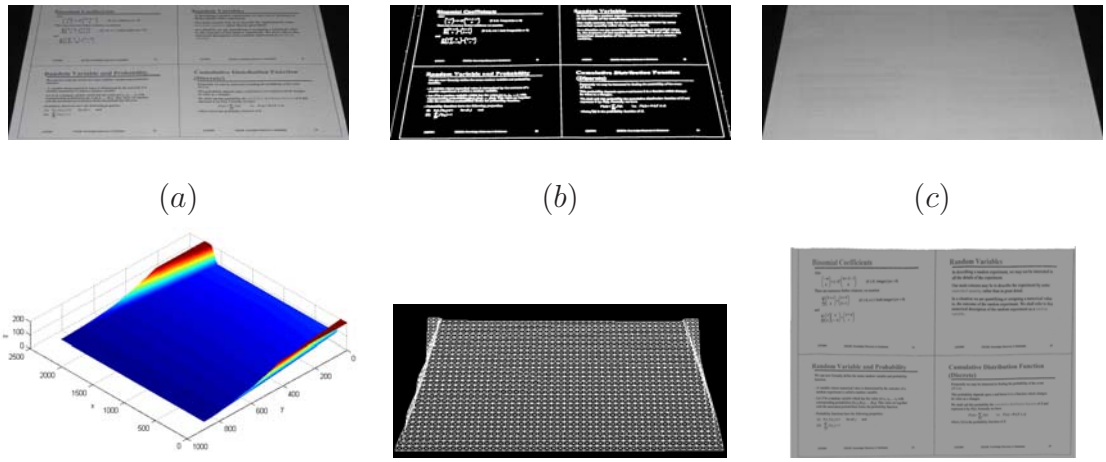
(d) (e) (f)

Figure 6.4: Overall framework evaluation on real document images - Example 3: (a) Distorted real document images; (b) Inpainting mask; (c) Extracted shading image; (e) 3D mesh of the reconstructed surface shape; (f) Final restored image with all the distortions removed.

regions and this will result in a blurring effect after the rectification due to pixel interpolation. One way to remedy this problem is to use a super resolution technique or an edge enhancement technique to further improve the quality of the image for better OCR performance.

6.2.4 Shape Refinement Step Evaluation

In the examples shown above, a one-step SFS method is used to derive the surface shape of the document without the additional refinement step. This is because the surface reconstruction does not improve much through the refinement step when the extracted shading image is a good approximation of the real shading and the light source location is



(a) (b) (c)

(d) (e) (f)

Figure 6.5: Overall framework evaluation on real document images - Example 4: (*a*) A planar page image with pure perspective and shading distortions; (*b*) Inpainting mask; (*c*) Extracted shading image; (*d*) Reconstructed surface shape; (*e*) 3D mesh of the surface shape; (*f*) Final restored image with all the distortions removed.

Table 6.1: OCR results on both the original images and the restored images shown in Figure 6.2-6.7.

OCR Results	Original image		Restored image	
	P_w	P_c	P_w	P_c
Figure 6.2	65.9%	68.6%	95.7%	96.3%
Figure 6.3	58.3%	62.4%	91.7%	93.1%
Figure 6.4	27.3%	21.6%	84.8%	83.4%
Figure 6.5	69.8%	70.3%	96.5%	97.3%
Figure 6.6	73.0%	76.8%	94.6%	95.5%
Figure 6.7	37.6%	38.2%	70.2%	74.8%

a close estimation of its actual position. However, when such good conditions are not met, the one-step surface reconstruction result may deviate from the actual shape significantly. For example, if we assume the on-camera flash is aligned with the optical center for

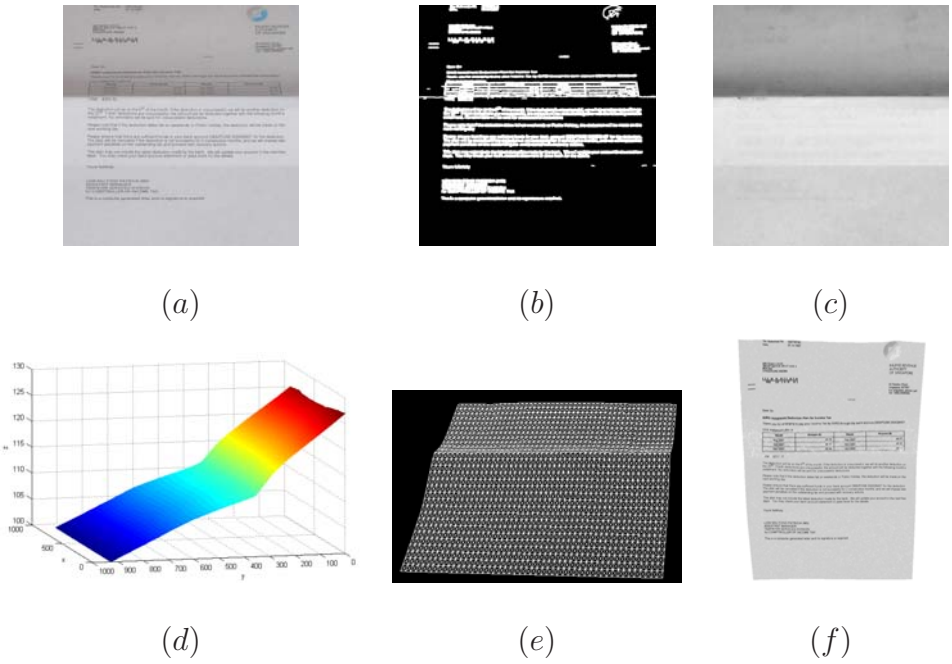
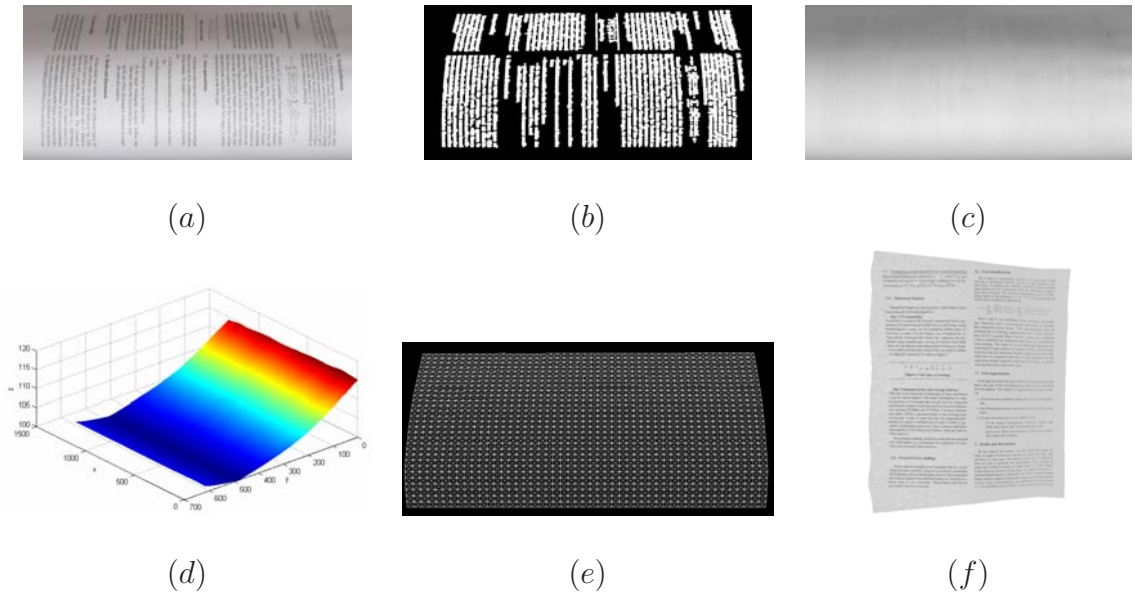


Figure 6.6: Overall framework evaluation on real document images - Example 5: (a) A folded page image with shading, geometric and perspective distortions; (b) Inpainting mask; (c) Extracted shading image; (d) Reconstructed surface shape; (e) 3D mesh of the surface shape; (f) Final restored image with all the distortions removed.

simplicity reasons, i.e. let $L = (0, 0, 0)$, then the one-step surface reconstruction based on the SFS method discussed in Section 4.4 may not produce a satisfactory result. Therefore, an additional shape refinement step is needed to further improve the reconstructed surface in order to ensure a good restored image at the final stage.

To illustrate how an inaccurate light source location may affect the surface reconstruction result, we try to produce two surfaces based on a synthetic shading image using the SFS method discussed in Section 4.4, one with an accurate light source location and the other with an inaccurate light source location. Using a cylindrical shape shown in Figure 6.8 (a) as an example, we generate a synthetic shading image as shown in Figure 6.8 (b) by specifying the on-camera flash's location at $L = (1.4, -5.0, 0)$. Next, we try



(a) (b) (c)

(d) (e) (f)

Figure 6.7: Overall framework evaluation on real document images - Example 6: (*a*) A thick bound book page with shading, geometric and perspective distortions; (*b*) Inpainting mask; (*c*) Extracted shading image; (*d*) Reconstructed surface shape; (*e*) 3D mesh of the surface shape; (*f*) Final restored image with all the distortions removed.

to reconstruct the surface shape by assuming that the on-camera flash is at the optical center and get the result as shown in Figure 6.8 (*c*). Similarly, we reconstruct another shape with the accurate light source and get the result shown in Figure 6.8 (*d*). Clearly, the first reconstruction deviates from the true shape significantly while the second reconstruction is close to the actual shape. In addition, we also conduct a similar experiment on a real document image. Figure 6.9 (*c*) and (*d*) show the reconstructed shapes using an inaccurate light source location and an accurate light source location, respectively. The results indicate that the light source location is an important factor that influences the surface reconstruction performance greatly.

On the other hand, when the light source locations are not readily available, we can still make a rough estimation of its location and try to improve the reconstructed surfaces

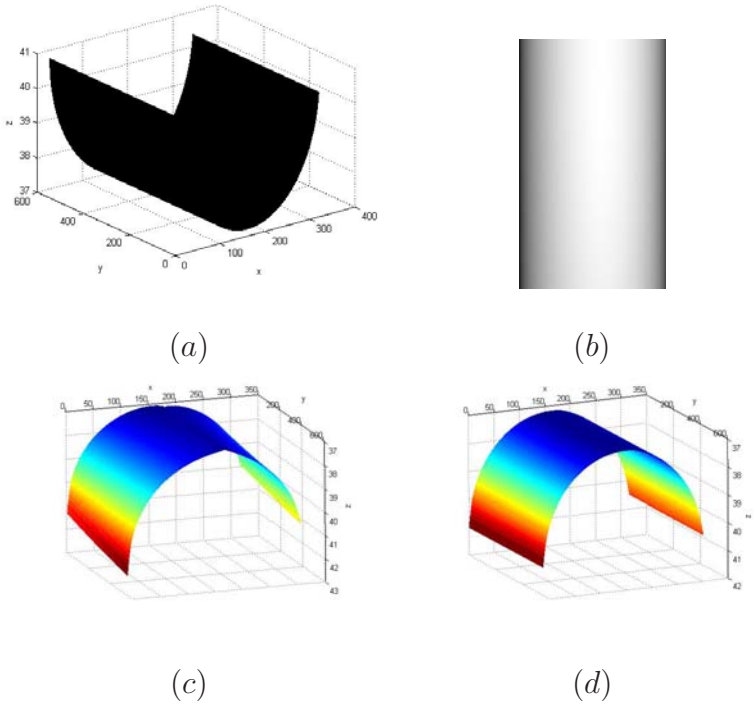


Figure 6.8: Effect of the light source locations on surface reconstruction - a synthetic example: (a) A cylindrical shape; (b) Synthetic shading image of (a) with $L = (1.4, -5.0, 0)$; (c) Reconstructed shape using $L = (0, 0, 0)$; (d) Reconstructed shape using $L = (1.4, -5.0, 0)$.

using the refinement step described in Section 4.4.3. Figure 6.10 shows an example of how the refinement step can improve the shape reconstruction result and subsequently improve the quality of the restored image when the known light source location is not so accurate. Figure 6.10 (a) is a real document image taken using a normal digital camera with its on-camera flash. By assuming the on-camera flash is located at the optical center, we obtain the one-step reconstructed surface as shown in Figure 6.10 (e). This step is initialized by setting the height at the two vertical boundaries to 0. This could be any arbitrary value because the reconstructed shape is translation invariant. As we can see, due to the inaccurate light source information, the reconstructed surface has significant errors near the central part. If we further apply the refinement step by using this shape

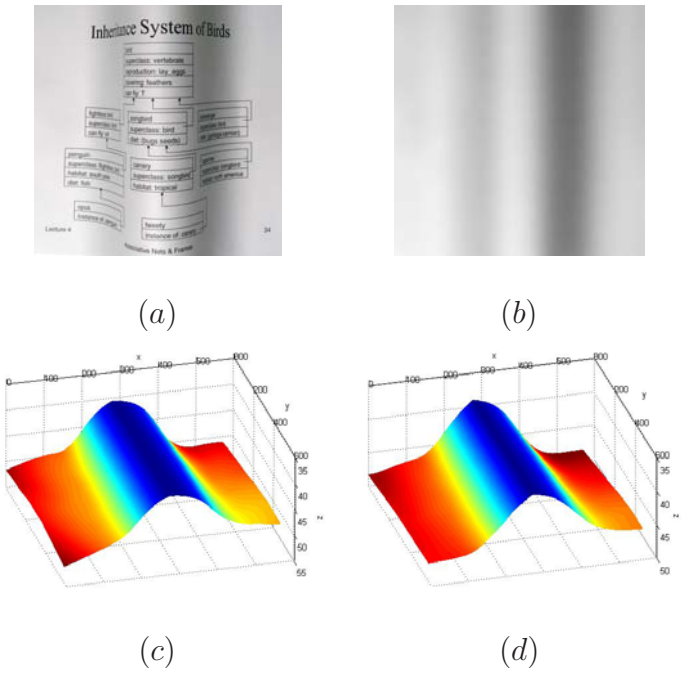


Figure 6.9: Effect of the light source locations on surface reconstruction - a real document image example: (a) A real image under a light source close to the optical center with $L = (-1.5, 4.5, -1.0)$; (b) Extracted shading image; (c) Reconstructed shape using $L = (0, 0, 0)$; (d) Reconstructed shape using $L = (-1.5, 4.5, -1.0)$.

as the initial value, we obtain the refined shape as shown in Figure 6.10 (f). Clearly, this is much better than the initial result with less abrupt errors. Finally, using this refined shape and the shading corrected image as the input to the flattening procedure, we obtain the final restored image as shown in Figure 6.10 (g). It is noticed that the restored image is much better improved compared to the original image although there are still some distortions due to imperfections of the estimated shape.

Figure 6.11 shows another example of a diagonally curved document restored based on a refined surface shape. This randomly curved document does not have obvious boundaries lying on the same plane as opposed to the two vertical boundaries in Figure 6.10. Nevertheless, we use the singular points as the initialization condition in the first step.

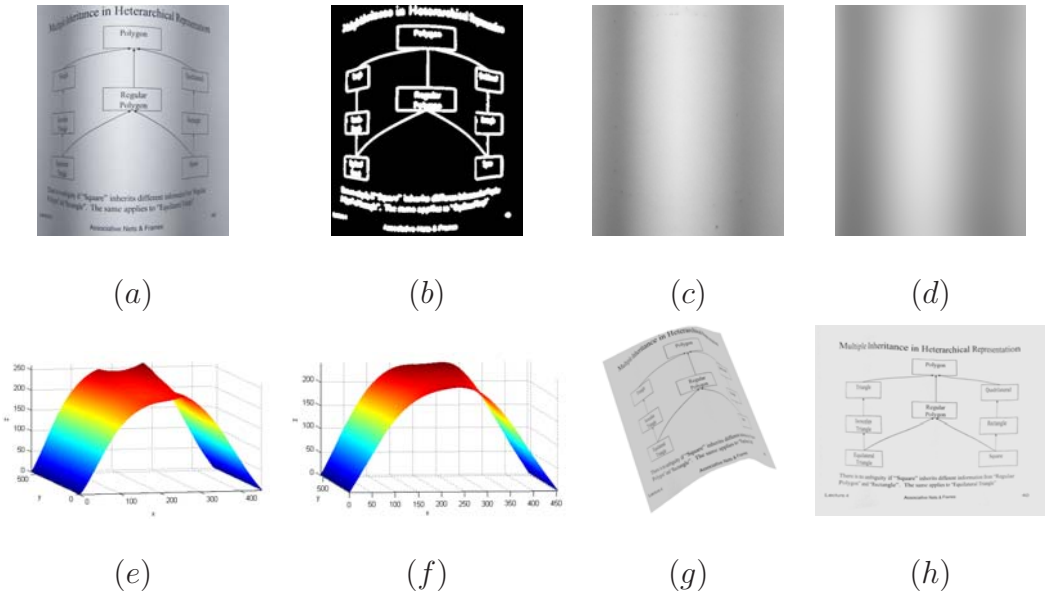
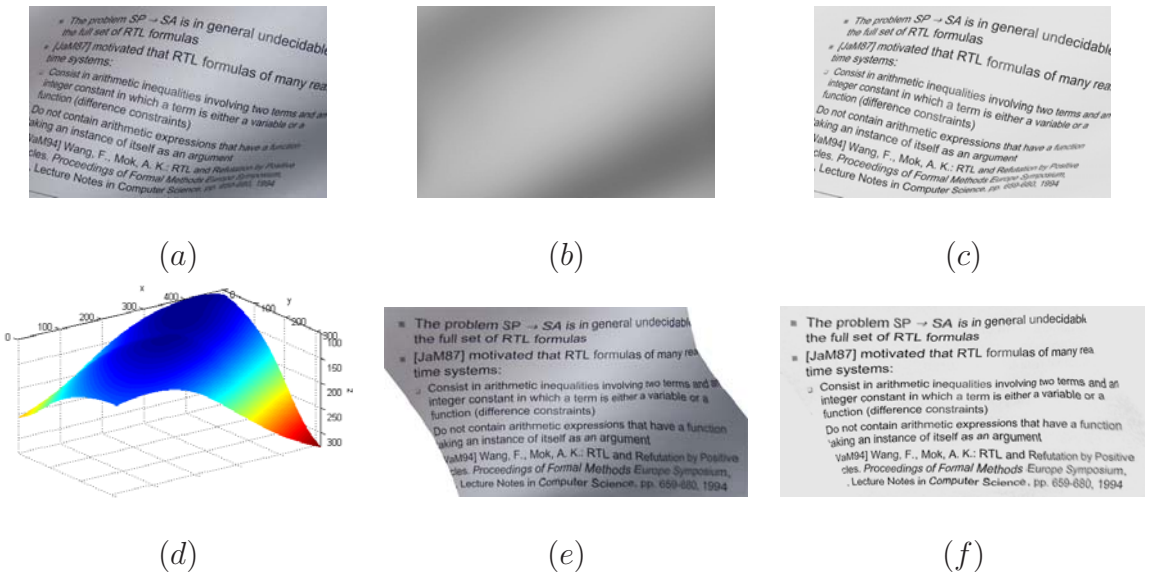


Figure 6.10: Restoration result based on a refined surface shape - Example 1: (a) Original warped image (cropped from an image of size 1600×1200); (b) Inpainting mask; (c) Inpainted image; (d) Extracted shading image; (e) One-step reconstructed shape; (f) Shape after refinement; (g) Surface mesh textured with the shading corrected image; (h) Final restored image.

Experiments show that even if the singular points are slightly off, the result is not affected much. Figure 6.11 (d) and (e) show the reconstructed shape after the refinement step and the final restored image, respectively. We can see that the shape does emulate the original curvature though it is not a perfect reconstruction. The restored image also shows a better visual appearance despite some remaining distortions. In terms of OCR performance, the restored image gives a word precision of 95.6% comparing to 38.9% on the original image.



(a) (b) (c)

(d) (e) (f)

Figure 6.11: Restoration result based on a refined surface shape - Example 2: (*a*) Original warped image; (*b*) Extracted shading image; (*c*) Photometrically restored image with $k = 0.9$; (*d*) Reconstructed shape by initializing singular points; (*e*) Geometrically restored image; (*f*) Final restored image.

6.3 Discussion

As shown in all the examples above, the photometric correction part is a critical component of the whole restoration process. On the one hand, it removes the shading or shadow distortions that often appear in daily snapshot images. On the other hand, it also produces a shading image for the subsequent shape reconstruction process if a SFS-based shape recovery method is used as in our case. The quality of the extracted shading image is one of the factors that influence the quality of the reconstructed surface shape which subsequently affects the quality of the final restored image. The generic photometric correction method we proposed in Chapter 3 proves to be an effective and efficient method that handles different degrees of shadings, shadows and background noise in both modern and historical document images. The experiments in this Chapter have further demonstrated its robust performance and illustrated how it integrates into the whole restoration

framework by providing a qualitative shading image to the shape reconstruction process.

The experiments in this Chapter have further demonstrated that the SFS method discussed in Section 4.4 is able to effectively recover smoothly warped surfaces in aid of the geometric and perspective restoration process. Due to the unconstrained imaging environment and the difficulty in extracting a perfect shading image, it is hard to obtain a very accurate reconstruction from real images as compared to the synthetic case. However, our objective is to be able to use a single input image to produce a rough estimation of the physical warping so that we can use this information to restore the geometric and perspective distortions to a certain extent. As our experiments have demonstrated, the restored images based on this approach have been greatly improved compared to the original images in terms of the OCR performance. In addition, the SFS approach integrates well into the full restoration framework by providing a good surface estimation to the geometric correction process.

Last but not least, we have demonstrated in this Chapter that the geometric correction method using the physically-based modeling technique as discussed in Section 5.5 can be well adapted to self-reconstructed shapes instead of only 3D scanned surfaces as shown in the earlier examples in Chapter 5. One additional step needed is to construct a uniformly sampled mesh from the reconstructed surface and register it with the shading corrected image. As for the 3D scanned surfaces, this is automatically provided by the 3D scanner since the shape and the image are essentially captured together. The experiments in this Chapter have also proved the robustness and effectiveness of the flattening procedure on self-reconstructed surface meshes. This makes the flattening procedure a flexible restoration routine that can be combined with different types of shape recovery method effectively.

CHAPTER 7

CONCLUSION AND FUTURE DIRECTIONS

7.1 Summary of Contributions

Document imaging is a fundamental application of computer vision and image processing. The ability to image printed documents has contributed greatly to the creation of vast digital collections now available from libraries and publishers. While traditional document imaging has been performed using flatbed scanning devices, a trend towards more flexible camera-based imaging is also emerging. The goal of document imaging, either via flatbed scanners or digital cameras, is to capture an image that is a reasonable substitute for the original printed content. However, with both approaches unavoidable distortions can

be present in the resulting image due to the printed materials' construction such as non-planar surfaces, the imaging setup, or the environmental conditions such as non-uniform illumination. Such distortions can make the document difficult to read as well as adversely affect the performance of subsequent processing, namely Optical Character Recognition (OCR) and Document Layout Analysis (DLA).

In this thesis, we presented several restoration techniques that try to rectify various document image distortions as a pre-processing step in order to produce restored images that can facilitate subsequent document image analysis tasks. Each of these techniques can be used by themselves to address the specific distortions effectively. Moreover, combined together they form a reasonably generic framework that is suitable for correcting mixed distortions in both traditional flatbed scanned images and current less-restrictive camera-based images. The three types of distortions that we dealt with are: 1) shading distortions including both shadings and shadows; 2) background noise including ink bleed-through, water stains, show-through, etc; 3) geometric distortions arising in non-planar documents including perspective distortions for camera-imaged materials.

More specifically, the whole framework consists of three components: photometric correction component, surface shape reconstruction component, and geometric correction component as shown in Figure 1.3 in Section 1.2. In the photometric correction component, we proposed an inpainting-based method to deal with shading distortions and background noise. This method by itself can be used to address photometric-related distortions in both scanned or camera-captured document images ranging from modern printed materials to historical manuscripts. On the other hand, many shading distortions are often mixed with geometric distortions due to the change of the surface normal with respect to the illumination direction. Such shading variations have direct relationships

with the surface geometry and therefore can be used to derive the surface shape of the document. This naturally led to the surface shape reconstruction component which utilizes the shading image obtained from the first component to recover the 3D structure of the document. In this component, we proposed two shape recovery methods based on the Shape-from-Shading methodology. The first method is based on an iterative Fast Marching strategy that takes an initialization of the surface shape and gradually improves it based on an upwind finite difference scheme. To improve the dependence on the large set of initialization values, the second method tries to model the image irradiance equation into a Hamilton-Jacobi equation and solves it using a Lax-Friedrichs-based Fast Sweeping strategy. This method only requires an initialization on the boundary values and has a relatively faster convergence rate. Once the surface shape is available, we can then move to the geometric correction component. The three methods proposed in this component can stand alone by themselves without relying on the surface shape reconstruction component. In particular, the first method is a purely 2D method that uses 2D text lines to estimate the warping curvature and rectify it through an interpolation technique. To better represent the physical warpings, 3D information is necessary. Therefore, the second method tries to incorporate the surface shape information into the text line representation, which effectively improves the restoration result. Finally, the third method presents a content independent approach that models the document's surface shape as a particle system and flattens it to obtain the restored shape and image. Although the last two methods both rely on the surface shape of the document, the shape can be obtained through various shape recovery techniques or captured using special devices. This demonstrates that the whole framework is very flexible and different methods can be substituted into each of the three components with ease.

7.2 Future Directions

In reference to our earlier discussions in Section 3.4, Section 4.6 and Section 5.7 on the limitations of the methods described in each component, further improvements and extensions can be explored in all three directions in terms of photometric correction, surface reconstruction and geometric correction.

7.2.1 Future Work on Photometric Correction

As discussed in Section 3.4, when the ink bleed-through in historical documents is too severe in a sense that the bleed-through pixels may appear even darker than the foreground pixels, then our inpainting-based technique will fail to produce satisfactory results. This is mainly due to the edge detection routine which is unable to distinguish foreground pixels and background pixels based on purely edge features. In this case, we may not be able to do much about it using a single image. However, if we can have both the front-side image and the back-side image, we may use the additional information to dim the back-side strokes and intensify the front-side strokes to increase the distinction between the foreground and the background pixels. On the other hand, to make the edge detection process more automatic, techniques can be devised to select the detector's parameters automatically based on features in the input image. Furthermore, more complex inpainting techniques such as texture inpainting can be applied to fill in the masked regions more accurately especially when the masked regions are large as in the case of a large embedded picture. There will certainly be efficiency tradeoffs when using complex algorithms in this case. Therefore, depending on the type of the application, one may choose to use different techniques according to the specific needs. Last but not least, if a color restoration is

needed, one may substitute the restored luminance component back to the original YUV model and reconstruct the corresponding color image accordingly.

7.2.2 Future Work on Surface Reconstruction

The Shape-from-Shading methods we proposed in Chapter 4 require a priori knowledge on the illumination direction of the imaging environment. Although this can be easily measured for the on-camera flash, it may be hard to obtain in general situations. One future direction is therefore to explore possible ways to estimate the illumination direction based on the input image. Moreover, because of the numerical schemes used to solve the image irradiance equation, the proposed Shape-from-Shading methods only deal with smoothly warped surfaces such as those shown in most of our examples. Although some ridges can still be recovered to a certain extent in some examples, more robust methods need to be developed to handle general discontinuous surfaces such as folds, creases, etc. Besides Shape-from-Shading techniques that derive the shape based on a single input image, other types of shape recovery methods using multiple images can also be explored such as those Shape-from-Stereo and Structure-from-Motion techniques mentioned in Section 2.3.3. Using a single 2D image makes the methods more applicable to real-world imaging applications, but it may also suffer the problem of resolution loss when the warpings are severe. This is because the pixel values of the restored regions are essentially interpolated from the existing values. For example, when the warped book spine is restored, the image is typically stretched out with new pixel values interpolated from the existing values. Multi-view imaging can be applied to address this problem. Meanwhile, the surface reconstructed from multi-view images are also more accurate compared to that from a single image.

7.2.3 Future Work on Geometric Correction

There has been many existing work that tackles geometric distortions in document images as we have reviewed in Section 2.3. Compared to most 2D-based approaches, 3D-based shape dependent approaches generally produce better results since the warpings are more accurately represented. The main concern for most 3D-based approaches is the availability of the surface shape. This also re-emphasizes the importance of the surface reconstruction component in our framework. Once the shape is available, either surface interpolation or numerical methods can be applied to restore the warping. Problems concerning the surface interpolation method might be how to deal with irregular surfaces such as crumpled pages with all the rumples and creases. For the numerical methods, the concern might be the stability and efficiency. More research can thus be focused on these directions to produce better restoration methods that address all these concerns.

BIBLIOGRAPHY

[AS95] D. Adalsteinsson and J.A. Sethian. A fast level set method for propagating interfaces. *Journal of Computational Physics*, 118(2):269–277, 1995.

[Bai93] H.S. Baird. Document image defect models and their uses. *In proceedings of Second International Conference on Document Analysis and Recognition (ICDAR'03)*, 1:62–67, 1993.

[Bai00] H.S. Baird. The state of the art of document image degradation modeling. *In Proceedings of IAPR International Workshop on Document Analysis System (DAS'00)*, 1:1–16, 2000.

[Bai03] H.S. Baird. Digital libraries and document image analysis. *In proceedings of Seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, 1:2–14, 2003.

[Bar97] D. Baraff. An introduction to physically based modeling: Rigid body simulation i - unconstrained rigid body dynamics. *ACM SIGGRAPH Tutorial Notes*, 1997.

[BBY92] H.S. Baird, H. Bunke, and K. Yamamoto. *Structured Document Image Analysis*. Springer-Verlag, 1992.

[Ber86] J. Bernsen. Dynamic thresholding of gray level images. *International Conference of Pattern Recognition*, pages 1251–1255, 1986.

[BP05] M.S. Brown and C.J. Pisula. Conformal deskewing of non-planar documents. *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:998–1004, 2005.

[BS00] M.S. Brown and W.B. Seales. Beyond 2d images: Effective 3d imaging for library materials. *In proceedings of ACM Conference on Digital Library (ACMDL'00)*, 1:27–36, 2000.

[BS01a] M.S. Brown and W.B. Seales. Digital atheneum: New approach for preserving, restoring and analyzing damaged manuscripts. *In proceedings of IEEE/ACM Joint Conference on Digital Library (JCDL'01)*, 1:437–443, 2001.

[BS01b] M.S. Brown and W.B. Seales. Document restoration using 3d shape: a general deskewing algorithm for arbitrarily warped documents. *In proceedings of International Conference on Computer Vision (ICCV'01)*, 2:367–374, 2001.

[BS04] M.S. Brown and W.B. Seales. Image restoration of arbitrarily warped documents. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(10):1295–1306, 2004.

[BSBC00] M. Bertalmio, G. Sapiro, C. Ballester, and V. Caselles. Image inpainting. *SIGGRAPH 2000*, 1:417C424, 2000.

[BT78] H.G. Barrow and J.M. Tenenbaum. *Recovering Intrinsic Scene Characteristics from Images*. Academic Press, 1978.

[BT06] M.S. Brown and Y.C. Tsoi. Geometric and shading correction for images of printed materials using boundary. *In proceedings of IEEE Transactions on Image Processing*, 15(6):1544–1554, 2006.

[BYBKD07] I. Bar-Yosef, I. Beckman, K. Kedem, and I. Dinstein. Binarization, character extraction, and writer identification of historical hebrew calligraphy documents. *International Journal on Document Analysis and Recognition*, pages 89–99, 2007.

[BZA07] W. Boussellaa, A. Zahour, and A. Alimi. A methodology for the separation of foreground/background in arabic historical manuscripts using hybrid methods. *ACM symposium on Applied computing*, pages 605–609, 2007.

[CBM+03] J.C. Carr, R.K. Beatson, B.C. McCallum, W.R. Fright, T. McLennan, and T.J. Mitchell. Smooth surface reconstruction from noisy range data. *Graphite 2003*, pages 119 –297, 2003.

[CCDG04] F. Courteille, A. Crouzil, J.D. Durou, and P. Gurdjos. Towards shape from shading under realistic photographic conditions. *International Conference on Pattern Recognition*, 2:277–280, 2004.

[CDD03] A. Crouzil, X. Descombes, and J.D. Durou. A multiresolution approach for shape from shading coupling deterministic and stochastic optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1416–1421, 2003.

[CDL03a] H. Cao, X. Ding, and C. Liu. A cylindrical model to rectify the bound document image. *In Proceedings of the IEEE International Conference on Computer Vision (ICCV'03)*, 2:228–233, 2003.

[CDL03b] H. Cao, X. Ding, and C. Liu. Rectifying the bound document image captured by the camera: A model based approach. *In proceedings of Seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, 1:71–75, 2003.

[CK03] K.J. Choi and H.S. Ko. Extending the immediate buckling model to triangular meshes for simulating complex clothes. *In Proceedings of EUROGRAPHICS*, pages 187–191, 2003.

[CM02] P. Clark and M. Mirmehdi. Recognizing text in real scenes. *Int'l Journal on Document Analysis and Recognition*, 4:243–257, 2002.

[CM03] P. Clark and M. Mirmehdi. Rectifying perspective views of text in 3d scenes using vanishing points. *Pattern Recognition*, 36:2673–2686, 2003.

[CS02] T.F. Chan and J.H. Shen. Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics*, 62(3):1019–1043, 2002.

[DBQ97] A. Doncescu, A. Bouju, and V. Quillet. Former books digital processing: Image warping. *In proceedings of International Workshop of Document Image Analysis (DIA'97)*, 1:5–9, 1997.

[DFS04] J. D. Durou, M. Falcone, and M. Sagona. A survey of numerical methods for shape from shading, 2004. Research Report from IRIT.

[DLL03] D. Doermann, J. Liang, and H.P. Li. Progress in camera-based document image analysis. *7th Int'l Conf. on Document Analysis and Recognition*, pages 606–616, 2003.

[Dri06] F. Drira. Towards restoring historic documents degraded over time. *Second International Conference on Document Image Analysis for Libraries*, pages 350–357, 2006.

[DWPL04] X.Q. Ding, D. Wen, L.R. Peng, and C.S. Liu. Document digitization technology and its application for digital library in china. *In proceedings of First International Workshop on Document Image Analysis for Libraries (DIAL'04)*, 1:46–53, 2004.

[ED01] A. Pathak E. Dubois. Reduction of bleed-through in scanned manuscripts documents. *In Proceedings of the IS&T conference on image processing, image quality, image capture systems*, pages 177–180, 2001.

[Far90] G. Farin. *Curves and surfaces for computer aided geometric design : a practical guide*. Boston: Academic Press, 1990.

[FC88] R.T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:439–451, 1988.

[FDB92] B. V. Funt, M. S. Drew, and M. Brockington. Recovering shading from color images. *2nd European Conference on Computer Vision*, pages 124–132, May 1992.

[Fle87] R. Fletcher. *Practical Methods of Optimization*. John Wiley and Sons, New York, 2nd edition, 1987.

[FREM03] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Fast and robust super-resolution. *IEEE Int. Conf. on Image Processing*, pages 291–294, 2003.

[GK06] P. A. Gremaud and C. M. Kuster. Computational study of fast methods for the eikonal equation. *SIAM Journal on Scientific Computing*, 27(6):1803–1816, 2006.

[HB86] B.K.P. Horn and M.J. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, 1986.

[HMB01] S. Huh, D.N. Metaxas, and N.I. Badler. Collision resolutions in cloth simulation. *In Proceedings of Computer Animation*, pages 122–127, 2001.

[Hor75] B.K.P. Horn. *Obtaining shape from shading information*. The Psychology of Computer Vision (P.H. Winston), McGraw-Hill, New York, 1975.

[IB06] R. Kimmel I. Blayvas, A. Bruckstein. Efficient computation of adaptive threshold surfaces for image binarization. *Pattern Recognition*, 39(1):89–101, 2006.

[IH81] K. Ikeuchi and B.K.P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17(1):141–184, 1981.

[ISI06] A. Iketani, T. Sato, and S. Ikeda. Video mosaicing for curved documents based on structure from motion. *In proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, 1:391–396, 2006.

[Jak03] T. Jakobsen. Advanced character physics. *Game Developer*, 2003.

[KB95] R. Kimmel and A.M. Bruckstein. Tracking level sets by level sets: A method for solving the shape from shading problem. *Computer Vision, Graphics, Image Processing: Image Understanding*, 62(2):47–58, 1995.

[KHB95] T. Kanungo, R.M. Haralick, and H. Baird. Validation and estimation of document degradation models. *In Proceedings of Annual Symposium on Document Analysis and Information Retrieval (SDAIR'95)*, 1:217–225, 1995.

[KHB+00] 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'00)*, 22:1209–1223, 2000.

[KHP93] T. Kanungo, R.M. Haralick, and I. Phillips. Global and local document degradation models. *In proceedings of Second International Conference on Document Analysis and Recognition (ICDAR'93)*, 1:730–734, 1993.

[KOQ04] C.Y. Kao, S. Osher, and J.L. Qian. Lax-friedrichs sweeping scheme for static hamilton-jacobi equations. *Journal of Computational Physics*, 196(1):367–391, 2004.

[KS01] R. Kimmel and J.A. Sethian. Optimal algorithm for shape from shading and path planning. *Journal of Mathematical Imaging and Vision*, 14(2):237–244, 2001.

[LDD05] J. Liang, D. DeMenthon, and D. Doermann. Flattening curved documents in images. *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2:20–25, 2005.

[LDK00] H. Li, D. Doermann, and O. Kia. Automatic text detection and tracking in digital video. *IEEE Trans. on Image Processing*, 9(1):147–156, 2000.

[LK93] K. M. Lee and C. C. J. Kuo. Shape from shading with a linear triangular element surface model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(8):815–822, 1993.

[LMAB01] O. Lavaille, X. Molines, F. Angella, and P. Baylou. Active contours network to straighten distorted text lines. *In Proceedings of International Conference on Image Processing (ICIP'01)*, 1:1074–1077, 2001.

[LS05] Y. Lin and W.B. Seales. Opaque document imaging: Building images of inaccessible texts. *In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)*, 1:662–669, 2005.

[LT07] S. Lu and C. L. Tan. Thresholding of badly illuminated document images through photometric correction. *ACM Symposium on Document Engineering*, pages 3–8, 2007.

[LVPG02] G. Leedham, S. Varma, A. Patankar, and V. Govindaraju. Separating text and background in degraded document images - a comparison of global threshholding techniques for multi-stage threshholding. *8th International Workshop on Frontiers in Handwriting Recognition*, pages 244–249, 2002.

[MBL+05] G. K. Myers, R. C. Bolles, Q. T. Luong, J. A. Herson, and H. B. Aradhye. Rectification and recognition of text in 3-d scenes. *International Journal on Document Analysis and Recognition*, pages 147–158, 2005.

[Nib86] W. Niblack. *An Introduction to Digital Image Processing*. Prentice-Hall, Englewood Cliffs, 1986.

[NS02] H. Nishida and T. Suzuki. Correcting show-through effects on document images by multiscale analysis. *16th International Conference on Pattern Recognition*, pages 65–68, 2002.

[OD91] J. Oliensis and P. Dupuis. Direct method for reconstructing shape from shading. *SPIE*, pages 116–128, 1991.

[OK04] A. Olmos and F. A. A. Kingdom. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception*, 33(12):1463–1473, 2004.

[Oli01] G. Oliveira. Exploring spring models. *Game Developer*, 2001.

[OS88] S. Osher and J.A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on the hamilton-jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[PCF06] E. Prados, F. Camilli, and O. Faugeras. A unifying and rigorous shape from shading method adapted to realistic data and applications. *Journal of Mathematical Imaging and Vision*, 25(3):307–328, 2006.

[Pen84] A.P. Pentland. Local shading analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:170–187, 1984.

[PF03] E. Prados and O. Faugeras. Perspective shape from shading and viscosity solutions. *In Proceedings of IEEE International Conference on Computer Vision (ICCV'03)*, 2:826–831, 2003.

[PFR02] E. Prados, O. Faugeras, and E. Rouy. Shape from shading and viscosity solutions. *European Conf. on Computer Vision*, 1:709–804, 2002.

[PFTV86] W.H. Press, B.P. Fannery, S.A. Teukolsky, and W.T. Verrerling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, 1986.

[Pil01a] M. Pilu. Extraction of illusory linear clues in perspectively skewed documents. *Computer Vision and Pattern Recognition*, pages 363–368, 2001.

[Pil01b] M. Pilu. Undoing page curl distortion using applicable surfaces. *In IEEE Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 1:67–72, 2001.

[PO01] M. Pietikainen and O. Okun. Edge-based method for text detection from complex document images. *6th Int'l Conf. on Document Analysis and Recognition*, pages 286–291, 2001.

[Pro95] X. Provot. Deformation constraints in a mass-spring model to describe rigid cloth behaviour. *Graphics Interface*, pages 155–174, 1995.

[Pro97] X. Provot. Collision and self-collision detection handling in cloth model dedicated to design garments. *Graphics Interface*, pages 177–189, 1997.

[PS05] E. Prados and S. Soatto. Fast marching method for generic shape from shading. *3rd International Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, 1:320–331, 2005.

[RNN99] S.V. Rice, G. Nagy, and T. A. Nartker. *Optical character recognition: an illustrated guide to the frontier*. Kluwer Academic Publishers, 1999.

[RT92] E. Rouy and A. Tourin. A viscosity soluition approach to shape from shading. *SIAM Journal on Numerical Analysis*, 3(29):867–884, 1992.

[Set99] J.A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, New York, 1999.

[Sha01] G. Sharma. Show-through cancellation in scans of duplex printed documents. *IEEE Transactions on Image Processing*, 10(5):736–754, 2001.

[SL04] W.B. Seales and Y. Lin. Digital restoration using volumetric scanning. *In proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries (JCDL'04)*, 1:117–124, 2004.

[SP00] J. Sauvola and M. Pietikainen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.

[SYY+05] M. Sun, R. Yang, L. Yun, G. Landon, B. Seales, and M.S. Brown. Geometric and photometric restoration of distorted documents. *In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)*, 2:17–21, 2005.

[TB04] Y.C. Tsoi and M.S. Brown. Geometric and shading correction for images of printed materials - a unified approach using boundary. *In IEEE Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR'04)*, 1:240–246, 2004.

[TBS04] A. Tonazzini, L. Bedini, and E. Salerno. Independent component analysis for document restoration. *International Journal on Document Analysis and Recognition*, 7:17–27, 2004.

[TCOZ03] Y.H.R. Tsai, L.T. Cheng, S. Osher, and H.K. Zhao. Fast sweeping algorithms for a class of hamilton-jacobi equations. *SIAM J. on Numerical Analysis*, 41(2):659–672, 2003.

[TCS02a] C. L. Tan, R. Cao, and P. Shen. Restoration of archival documents using a wavelet technique. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24:1399–1404, 2002.

[TCS+02b] C. L. Tan, R. Cao, P. Shen, J. Chee, and J. Chang. Text extraction from historical handwritten documents by edge detection. *6th International Conference on Control, Automation, Robotics and Vision*, 24:1399–1404, 2002.

[TFA05] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 27(9):1459–1472, 2005.

[TPB87] D. Terzopoulos, J.C. Platt, and A.H. Barr. Elastically deformable models. *In Proceedings of ACM SIGGRAPH*, 21:205–214, 1987.

[TS93] Y.Y. Tang and C.Y. Suen. Image transformation approach to nonlinear shape restoration. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(1):155–171, 1993.

[TS94] P.S. Tsai and M. Shah. Shape from shading using linear approximation. *Journal of Image and Vision Computing*, 12(8):487–498, 1994.

[TSY03] A. Tankus, N. Sochen, and Y. Yeshurun. A new perspective on shape-from-shading. *IEEE International Conference on Computer Vision*, 2:862–869, 2003.

[TSY04] A. Tankus, N. Sochen, and Y. Yeshurun. Perspective shape-from-shading by fast marching. *IEEE Conf. on Computer Vision and Pattern Recognition*, 1:43–49, 2004.

[TT95] O. D. Trier and T. Taxt. Evaluation of binarization methods for document images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(3):312–315, 1995.

[TZAF04] J. Toro, D. Ziou, and M. F. Auclair-Fortier. Recovering the shading image under known illumination. *1st Canadian Conference on Computer and Robot Vision*, pages 92–96, 2004.

[TZZX06] C.L. Tan, L. Zhang, Z. Zhang, and T. Xia. Restoring warped document images through 3d shape modeling. *In proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'06)*, 28(2):195–208, 2006.

[Viv] Vivid. 3d digitizers - non-contact laser range scanner.

[VT94] P. Volino and N.M. Thalmann. Efficient self-collision detection on smoothly discretized surface animations using geometrical shape regularity. *In Proceedings of EUROGRAPHICS*, 13:155–166, 1994.

[VT00] P. Volino and N.M. Thalmann. Implementing fast cloth simulation with collision response. *In Proceedings of Computer Graphics Internaltional (CGI'00)*, pages 257–268, 2000.

[Wei01] Y. Weiss. Deriving intrinsic images from image sequences. *IEEE International Conference on Computer Vision*, 2:68–75, 2001.

[Wit97] A. Witkin. An introduction to physically based modeling: Particle system dynamics. *ACM SIGGRAPH*, 1997.

[Wit01] A. Witkin. Physically based modeling - particle system dynamics. *ACM SIGGRAPH Course Notes*, 2001.

[Wol90] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, 1990.

[WR83] J. M. White and G. D. Rohrer. Image thresholding for optical character recognition and other applications requiring character image extraction. *IBM Journal of Research and Development*, 27(4):400–411, 1983.

[WUM97] T. Wada, H. Ukida, and T. Matsuyama. Shape from shading with interreflections under a proximal light source: Distortion-free copying of an unfolded book. *In proceedings of International Journal of Computer Vision (IJCV'97)*, 24(2):125–135, 1997.

[WZ96] Y. Weng and Q. Zhu. Nonlinear shape restoration for document images. *In IEEE Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR'96)*, 1:568–573, 1996.

[YB89] S. D. Yanowitz and A. M. Bruckstein. A new method for image segmentation. *Computer Vision, Graphics and Image Processing*, 46(1):82–95, 1989.

[YKKM04] 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. *In proceedings of 17th International Conference on Pattern Recognition (ICPR'04)*, 1:482–485, 2004.

[YTLC02] S.Y. Yuen, Y.Y. Tsui, Y.W. Leung, and R.M.M. Chen. Fast marching method for shape from shading under perspective projection. *Visualization, Imaging, and Image Processing*, pages 584–589, 2002.

[Zha99] Z. Y. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. *In Proceedings of IEEE International Conference on Computer Vision (ICCV'99)*, 1:666–673, 1999.

[ZT01] Z. Zhang and C.L. Tan. Recovery of distorted document image from bound volumes. *In proceedings of Sixth International Conference on Document Analysis and Recognition (ICDAR'01)*, 1:429–433, 2001.

[ZT02] Z. Zhang and C.L. Tan. Straightening warped text lines using polynomial regression. *In proceedings of International Conference on Image Processing (ICIP'02)*, 1:977–980, 2002.

[ZT03] Z. Zhang and C.L. Tan. Correcting document image warping based on regression of curved text lines. *In proceedings of Seventh International Conference on Document Analysis and Recognition (ICDAR'03)*, 1:589–593, 2003.

[ZTCS99] R. Zhang, P.S. Tsai, J.E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.

[ZTF04] Z. Zhang, C.L. Tan, and L.Y. Fan. Restoration of curved document images through 3d shape modeling. *In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, 1:10–16, 2004.

[ZZJ00] Y. Zhong, H. Zhang, and A.K. Jain. Automatic caption localization in compressed video. *IEEE Pattern Analysis and Machine Intelligence*, 22(4):385–392, 2000.

[ZZQ04] Y.T. Zhang, H.K. Zhao, and J.L. Qian. High order fast sweeping methods for eikonal equations. *In Proceedings of the 74th SEG Annual International Meeting*, pages 1901–1904, 2004.

[ZZTX05] L. Zhang, Z. Zhang, C.L. Tan, and T. Xia. 3d geometric and optical modeling of warped document images. *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1:337–342, 2005.

AUTHOR BIOGRAPHY

Li Zhang is a PhD candidate in the Department of Computer Science, School of Computing, National University of Singapore. Her research interests include Document Image Processing, Analysis and Understanding, Information Extraction and Retrieval, Computer Vision and Pattern Recognition.

During her PhD candidature, her publications include:

- Li Zhang, Andy M. Yip, Michael S. Brown, Chew Lim Tan. A Unified Framework for Document Restoration Using Inpainting and Shape-from-Shading. Submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'08)*.
- Li Zhang, Andy M. Yip, Chew Lim Tan. An Improved Physically-Based Method for Geometric Restoration of Distorted Document Images. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI'08)*. To appear.

- Li Zhang, Andy M. Yip, Chew Lim Tan. A Restoration Framework for Correcting Photometric and Geometric Distortions in Camera-Based Document Images. In *Proceedings of the 11th International Conference on Computer Vision (ICCV'07)*. Rio de Janeiro, Brazil, October 14-20, 2007.

- Li Zhang, Andy M. Yip, Chew Lim Tan. Photometric and Geometric Restoration of Document Images Using Inpainting and Shape-from-Shading. In *Proceedings of the 22nd Conference on Artificial Intelligence (AAAI'07)*. Vancouver, Canada, July 22-26, 2007.

- Li Zhang, Andy M. Yip, Chew Lim Tan. Shape from Shading Based on Lax-Friedrichs Fast Sweeping and Regularization Techniques with Applications to Document Image Restoration. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*. Minneapolis, Minnesota, June 18-23, 2007.

- Li Zhang, Andy M. Yip, Chew Lim Tan. Removing Shading Distortions in Camera-Based Document Images Using Inpainting and Surface Fitting With Radial Basis Functions. In *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR'07)*. Brazil, September 23-26, 2007.

- Li Zhang and Chew Lim Tan. Warped Document Image Restoration Using Shape-from-Shading and Physically-Based Modeling. In *Proceedings of IEEE Workshop on Applications of Computer Vision (WACV'07)*. Austin, Texas, February 21-22, 2007.

- Li Zhang and Chew Lim Tan. Restoring Warped Document Images Using Shape-from-Shading and Surface Interpolation. In *Proceedings of 18th International Conference on Pattern Recognition (ICPR'06)*. Hong Kong, August 20-24, 2006.

- Li Zhang, Zheng Zhang, Chew Lim Tan and Tao Xia. 3D Geometric and Optical Modeling of Warped Document Images from Scanners. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*. San Diego, CA, June 20-25, 2005.

- Li Zhang and Chew Lim Tan. Warped Image Restoration with Applications to Digital Libraries. In *Proceedings of 8th International Conference on Document Analysis and Recognition (ICDAR'05)*. Seoul, Korea, August 29 - September 1, 2005.

- Kok Beng Chua, Li Zhang and Chew Lim Tan. A Fast and Stable Approach for Restoration of Warped Document Images. In *Proceedings of 8th International Conference on Document Analysis and Recognition (ICDAR'05)*. Seoul, Korea, August 29 - September 1, 2005.

- Xingyu Qi, Li Zhang and Chew Lim Tan. Motion Deblurring for Optical Character Recognition. In *Proceedings of 8th International Conference on Document Analysis and Recognition (ICDAR'05)*. Seoul, Korea, August 29 - September 1, 2005.