

11-16-2017

# A Book Reader Design for Persons with Visual Impairment and Blindness

Luis E. Galarza  
luis.galarza@fiu.edu

**DOI:** 10.25148/etd.FIDC004025

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

 Part of the [Signal Processing Commons](#)

---

## Recommended Citation

Galarza, Luis E., "A Book Reader Design for Persons with Visual Impairment and Blindness" (2017). *FIU Electronic Theses and Dissertations*. 3541.

<https://digitalcommons.fiu.edu/etd/3541>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact [dcc@fiu.edu](mailto:dcc@fiu.edu).

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

A BOOK READER DESIGN FOR PERSONS WITH VISUAL IMPAIRMENT AND  
BLINDNESS

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

Luis Galarza

2017

To: Dean John L. Volakis  
College of Engineering and Computing

This dissertation, written by Luis Galarza, and entitled A Book Reader Design for Persons with Visual Impairment and Blindness, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

---

Armando Barreto

---

Sakhrat Khizroev

---

Mercedes Cabrerizo

---

Chunlei (Peggy) Wang

---

Malek Adjouadi, Major Professor

Date of Defense: November 16, 2017

The dissertation of Luis Galarza is approved.

---

Dean John L. Volakis  
College of Engineering and Computing

---

Andrés G. Gil  
Vice President for Research and Economic Development  
and Dean of the University Graduate School

Florida International University, 2017

© Copyright 2017 by Luis Galarza

All rights reserved.

## DEDICATION

I dedicate this dissertation to my father, Rusty J. Galarza, who in life taught me with his words and actions to persevere in the face of adversity. As he used to tell me “Sí se puede”. I also dedicate this work to my mother, Maria and beloved sisters, Maria Grace, and Patricia. Thanks to their patience, understanding, support, and love, the completion of this endeavor has been made possible.

## ACKNOWLEDGMENTS

I am very grateful for all the guidance and support from Dr. Malek Adjouadi whose positive attitude, incredible patience, and deep insights have made this dissertation possible. When I faced challenges, it was always Dr. Adjouadi that inspired me to persevere. Furthermore, he provided me with a plan of action, which led to publications that have culminated in the work presented in this dissertation.

I also wish to express my gratitude to all my committee members, Dr. Armando Barreto, Dr. Sakhrat Khizroev, Dr. Mercedes Cabrerizo and Dr. Chunlei (Peggy) Wang, for their valuable insights which improved the quality of this dissertation.

Furthermore, I would like to acknowledge the research support provided from the Department of Electrical and Computer Engineering at Florida International University, the support of the National Science Foundation under grants: CNS-1532061, CNS-1042341, CNS 1551221, CNS-1429345 and IIP 1338922. The support of the Ware Foundation is also greatly appreciated.

I would also like to extend my gratitude to all the members of the Center for Advanced Technology and Education for creating an amazing research environment. In particular, I would like to acknowledge Zhenzhong Wang and Harold Martin, with whom I had the pleasure to collaborate with.

## ABSTRACT OF THE DISSERTATION

### A BOOK READER DESIGN FOR PERSONS WITH VISUAL IMPAIRMENT AND BLINDNESS

by

Luis Galarza

Florida International University, 2017

Miami, Florida

Professor Malek Adjouadi, Major Professor

The objective of this dissertation is to provide a new design approach to a fully automated book reader for individuals with visual impairment and blindness that is portable and cost effective. This approach relies on the geometry of the design setup and provides the mathematical foundation for integrating, in a unique way, a 3-D space surface map from a low-resolution time of flight (ToF) device with a high-resolution image as means to enhance the reading accuracy of warped images due to the page curvature of bound books and other magazines. The merits of this low cost, but effective automated book reader design include: (1) a seamless registration process of the two imaging modalities so that the low resolution (160 x 120 pixels) height map, acquired by an Argos3D-P100 camera, accurately covers the entire book spread as captured by the high resolution image (3072 x 2304 pixels) of a Canon G6 Camera; (2) a mathematical framework for overcoming the difficulties associated with the curvature of open bound books, a process referred to as the dewarping of the book spread images, and (3) image correction performance comparison between uniform and full height map to determine which map provides the highest Optical Character Recognition (OCR) reading accuracy possible. The design concept could also

be applied to address the challenging process of book digitization. This method is dependent on the geometry of the book reader setup for acquiring a 3-D map that yields high reading accuracy once appropriately fused with the high-resolution image. The experiments were performed on a dataset consisting of 200 pages with their corresponding computed and co-registered height maps, which are made available to the research community ([cate-book3dmaps.fiu.edu](http://cate-book3dmaps.fiu.edu)). Improvements to the characters reading accuracy, due to the correction steps, were quantified and measured by introducing the corrected images to an OCR engine and tabulating the number of miss-recognized characters. Furthermore, the resilience of the book reader was tested by introducing a rotational misalignment to the book spreads and comparing the OCR accuracy to those obtained with the standard alignment. The standard alignment yielded an average reading accuracy of 95.55% with the uniform height map (i.e., the height values of the central row of the 3-D map are replicated to approximate all other rows), and 96.11% with the full height maps (i.e., each row has its own height values as obtained from the 3D camera). When the rotational misalignments were taken into account, the results obtained produced average accuracies of 90.63% and 94.75% for the same respective height maps, proving added resilience of the full height map method to potential misalignments.



## TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION.....	1
1.1 Motivation.....	1
1.2 Current acquisition devices and assistive technology systems.....	2
1.3 Prior correction and dewarping studies.....	8
1.4 Problem definition and contributions of this dissertation.....	11
1.5 Dissertation organization .....	12
2. CONCEPT OF DESIGN AND THEORY .....	14
2.1 Book reader design .....	14
2.2 Concepts for the curvature correction method.....	16
3. IMPLEMENTATION STEPS AND INITIAL RESULTS .....	20
3.1 Height maps .....	20
3.2 Registration of imaging modalities.....	21
3.3 Flattening and extending correction algorithm.....	26
3.4 Initial testing .....	27
3.4.1 Flattening correction effect.....	30
3.4.2 Effects of specular lighting on the flattening process.....	34
3.4.3 Full correction: flattened and extended book spread.....	43
3.4.4 Resilience of the full correction on a rotated book spread.....	48
4. UNIFORM VS. FULL HEIGHT MAP.....	51
4.1 Exploring the effect of the height map .....	51
4.2 Standard placement.....	52
4.3 Misaligned placement .....	61
5. FURTHER EXPLORATION OF UNIFORM VS. FULL HEIGHT MAPS .....	70
5.1 Adjusting an alternative method to utilize height maps.....	70
5.2 Summary of adjustments.....	70
5.3 Standard placement.....	74
5.4 Misaligned placement .....	78
6. CONCLUSIONS AND FUTURE WORK .....	82
6.1 Summary.....	82
6.2 Future Work.....	84
LIST OF REFERENCES.....	86
APPENDICES .....	92
VITA.....	101

## LIST OF TABLES

TABLE	PAGE
Table 3.1 Effect of flatten excerpts .....	34
Table 3.2 Result of OCR engine from flatten excerpt .....	42
Table 3.3 Excerpt of results for old approach using ABBYY FineReader 8.0 .....	43
Table 3.4 Results with just flattening using ABBYY FineReader 12.0 .....	44
Table 3.5 Results of flattening with linear extension approach using ABBYY FineReader 12.0 .....	47
Table 3.6 Results of flattening with natural neighbor extension approach using ABBYY FineReader 12.0 .....	47
Table 3.7 Results of Flattening with nearest neighbor extension approach using ABBYY FineReader 12.0 .....	48
Table 3.8 Results of Flattening with linear extension approach using ABBYY FineReader 12.0 for rotated book spread 6.....	50
Table 4.1 Summary of the standard positioned book spreads.....	56
Table 4.2 Summary of the rotated positioned book spreads .....	65
Table 5.1 Summary of the standard positioned book spreads.....	78
Table 5.2 Summary of the rotated positioned book spreads .....	81
Table A.1 Summary of devices .....	100

## LIST OF FIGURES

FIGURE	PAGE
Figure 2.1 Experimental setup of the book reader design .....	15
Figure 2.2 Conceptual design and geometry of the book reader .....	17
Figure 3.1 Geometry for registering the two imaging modalities.....	22
Figure 3.2 The 2-D converted low-resolution book spread .....	24
Figure 3.3 The 3-D full height map with the low-resolution image .....	25
Figure 3.4 The 3-D higher-resolution full height map with the high-resolution image ..	26
Figure 3.5 Comparison of low-resolution raw height map and polynomial fitting by subtracting both values.....	28
Figure 3.6 Low-resolution height polynomial fitting with corresponding low- resolution pixel values. ....	29
Figure 3.7 Higher-resolution height polynomial fitting with corresponding high- resolution pixel values .....	30
Figure 3.8 Corrections via (a) raw height & (b) polynomial fitting. ....	31
Figure 3.9 Results of the OCR engine on the original warped image .....	32
Figure 3.10 Results of the OCR engine on the raw corrected image.....	33
Figure 3.11 Results of the OCR engine on the polynomial corrected image .....	33
Figure 3.12 Low-resolution raw height map with corresponding low-resolution pixel values .....	35
Figure 3.13 Comparison of low-resolution raw height map and polynomial fitting by subtracting both values.....	36
Figure 3.14 Low-resolution height polynomial fitting with low-resolution image .....	37
Figure 3.15 Higher-resolution image (warped) with inadequate lighting.....	37
Figure 3.16 Higher-resolution raw height map with high-resolution image .....	38
Figure 3.17 Higher-resolution height polynomial fitting with high-resolution image .....	38

Figure 3.18 Images under poor lighting corrected with (a) raw height and (b) polynomial fitting.....	39
Figure 3.19 Results of the OCR engine on the original warped image .....	40
Figure 3.20 Results of the OCR engine on the corrected image using the raw height map.....	41
Figure 3.21 Results of the OCR engine on the corrected image using polynomial fitting.....	42
Figure 3.22 Example of the original and fully corrected images.....	46
Figure 3.23 Example of the original (a) and rotated (b) book spread 6.....	49
Figure 4.1 Standard align book spread .....	52
Figure 4.2 Examples of text: (a) original warped, (b) corrected with uniform map, and (c) corrected with full map.....	54
Figure 4.3 Accuracy of the original and corrected book spreads .....	55
Figure 4.4 Page 79 for standard placement: a) original, b) uniform, c) full .....	57
Figure 4.5. Examples of standard placement results for: a) original, b) corrections with uniform map, c) corrections with full map .....	60
Figure 4.6 Rotated book spread .....	61
Figure 4.7 Examples of text: (a) original warped, (b) corrected with uniform map, and (c) corrected with full map.....	63
Figure 4.8 Accuracy of the original and corrected book spreads .....	64
Figure 4.9 Examples of rotated placement results for: a) original, b) corrections with uniform map, c) corrections with full map .....	68
Figure 5.1 Barrel Distorted Image and its Correction.....	71
Figure 5.2 Push Operation .....	72
Figure 5.3 Section View of the Page Curvature .....	73
Figure 5.4 Examples of text: (a) original warped, (b) corrected with uniform map, and (c) corrected with full map.....	76
Figure 5.5 Accuracy of the original and corrected book spreads .....	77

Figure 5.6 Examples of text: (a) original warped, (b) corrected with uniform map, and (c) corrected with full map.....	79
Figure 5.7 Accuracy of the rotated original and corrected book spreads .....	80
Figure A.1. Types of Qidenus V- Shaped Book Scanners.....	92
Figure A.2. Original FUJITSU Image Scanner ScanSnap SV600.....	93
Figure A.3. Modified FUJITSU Image Scanner ScanSnap SV600 .....	94
Figure A.4. Bookeye 4 with add-ons currently located at the FIU Library .....	95
Figure A.5. Xcanex bookscanner by piQx.....	96
Figure A.6. Transformer HD with Wi-Fi.....	97
Figure A.7. Smart Reader .....	98
Figure A.8. IRIScan Book 5 Wifi.....	99
Figure A.9. ReadingPen 2.....	100

## ABBREVIATIONS AND ACRONYMS

CCD	Charged Coupling Devices
IR	Infrared
JAWS	Job Access With Speech
KNFB	Kurzweil National Federation of the Blind
OCR	Optical Character Recognition
ToF	Time of Flight
TTS	Text To Speech
USB	Universal Serial Bus

# 1. INTRODUCTION

## 1.1 Motivation

Individuals with visual impairment and blindness have a difficult time reading bound books and magazines in the privacy of their own homes or offices. One of the reasons for this difficulty is due to the warping effect that occurs while a book or magazine is being digitized. A carefully designed book reader can address some of the difficulties and provide the freedom to check out any book from a public library or acquire a book from any other venue (old or new), without worrying if its audio version (abridged or unabridged) is available and then be required to dispense a monthly fee to Audiobooks.com or Audible.com to access their existing catalog of titles. This technology, although useful primarily as a reading device for visually impaired and blind individuals as part of the challenge for universal access, it can also serve as part of a practical solution to the remarkably challenging book digitization project. The digitization benefits the user as an audio version can now be created and saved in memory for later use or for re-reading certain sections of the book, but it will also serve the public at large by automating the audio recording. Currently it is estimated that about 40% of audiobook utilization is accomplished through public libraries, while the remaining 60% is through retail book stores.

The desirable features of this design include: portability, cost-effectiveness, and a user-friendly graphical interface. This design meets these requirements, as it can be connected

to any computer through USB ports and any platform could serve as the book holder as long as its surface is not black (as it absorbs most of the IR light). The acquisition devices included a CCD camera which was chosen for its relatively low cost, while the time of flight device, although it had a low resolution, was selected for its adequate although sparse depth map but at a relatively low cost and with a fast processing speed. All the operational steps of the book reader, including image capture from both acquisition devices, dewarping of the images to correct for the book curvature, and OCR for reading text have all been automated, and through the interface the user could initiate all these tasks through a single click on “START CAPTURE” button. Furthermore, any pages that have been read are automatically saved in memory as digitized text for future use.

## **1.2 Current acquisition devices and assistive technology systems**

There are different tools that have been proposed and used to digitize text. Flatbed scanners are among the most commonly used tools; however, these scanners require the user to physically press on each page of the bound book against the flatbed to reduce blurring. Portable scanners, although smaller in size, are only able to scan unbound pages and struggle with bound books and magazines. The emergence of digital cameras has made them ideal candidates for documents and book digitization, however there are some distortions which should be addressed to improve their performance [1]. Further developments in technology have made digital cameras available in portable devices, which have, in turn, expanded their use to scanning documents [2], however, they still are burdened by the inherit text distortion due to book curvature and barrel distortion. To address some of the shortcomings, laser scanner-based systems are used. Laser scanners



do provide very accurate information regarding the geometry of a book or document, but, due to their architecture, these devices can encounter digitization errors [3]. The cost of these devices can also be quite high, which increases depending on the accuracy sought [4].

Recent papers in assistive technologies for the visually impaired [5, 6] provide insightful application scope that show growth in the field of assistive technologies and the improvements they bring into the broad field of universal accessibility. Some of the technologies described [5] rely on components only available online. For instance, there are some devices, currently in the market, designed primarily for document digitization, such as the V-Shape Book Scanners by Qidenus Technologies with weights ranging from 190lb to 220lb. These devices require contact with the book by using a V shape base and flattening glass to reduce perspective distortions that occur during the capture process. They are bulky, industrial-like machines that require large volumes of space and are considerably heavy. They are used to digitize many pages automatically, but are expensive in terms of purchase and maintenance cost, thus not cost effective for users who will only use the device sparingly. The best model has a maximum resolution of 600 dpi and can process 2500 pages per hour, roughly 1.44 pages per second, which is quite impressive for the technology it offers in terms of digitalization. The FUJITSU Image Scanner ScanSnap SV600 is another scanning device that uses lens reduction optics with a CCD sensor to capture images. It is a small and lightweight product weighting in at only 6.6lb. The best scanning resolution it can offer is 600dpi for color scans and 1200dpi for monochrome, completing a capture in no more than 3 seconds. Once the scanning process completes, the deformation corrections, segmentation task, and limited OCR capabilities are available

through a graphical user interface. The ScanSnap SV600 costs around \$900. A magazine article [7] reviewed the operational aspects of this device and provides comparisons to other systems in the market. Bookeye 4 Professional is another scanning device available to users. It offers 600 dpi resolution with an adjustable scanning bed that changes its angle from flat to V-shape to reduce curvature distortions. It weighs in at 90lbs, not including accessories that require more space. It requires manual placement and scanning of the documents, and is very sensitive to document placement. It can achieve scanning speeds from 400 to 1000 pages per hour and additional OCR software capabilities can be purchased from the manufacturer at an additional cost, adding to its already high price tag.

All the previously mentioned scanners can be used in conjunction with OCR engines to implement a book reader system. An evaluation of the performance of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym (some of the most accurate and popular OCR software available) can be found in [8], where the authors analyze their accuracy and reliability using over 1200 images of different types, including hand- or machine-written characters, and with images that introduce blur, skewness, and different orientations, among other aspects. Results showed that there is still room for improvement as these OCR engines are not 100 percent accurate.

Many of the commercially available book readers combine a digital camera and some form of OCR software to ensure portability and accessibility to consumers. It is difficult to gauge the merits of such products as they are not cited in the literature, and the descriptions we can obtain come from the distributors' websites. Some examples of such products are piQx,

Transformer HD, and Smart Reader. The piQx system is very sensitive to document placement and alignment, as it utilizes a single inexpensive camera design to capture the image, and acquires the text via ABBYY. The complete product is fairly inexpensive, in contrast to the previously discussed scanners, coming in at only \$299. The Transformer HD is a portable device that relies on a manually operated high definition camera, with the capability to physically zoom in and stream video to capture the document image before post-processing and introducing it to the OCR software. However, due to its higher-resolution, the cost of this device is significantly higher, currently priced above \$2,500. The Smart Reader, just like the piQx, relies on a simple camera design but, unlike the piQx, it integrates a computing system, OCR, and TTS module making it a more complete and portable system that comes at a price of \$1,795. Another self-sufficient and extremely portable device is the ReadingPen2, which is priced at \$209; however, unlike the other devices, it requires significantly more interaction from the user, as he/she must follow every line of text that is intended to be read.

Alternatively, some users might prefer to use their own portable devices (such as smartphones or tablets) along with specialized software to read text. This text could be found on the screen of the devices, either previously extracted from images or from printed documents. Other assistive technologies which read text that has already been digitized to visually impaired individuals by using Digital Accessible Information Systems are described [9]. One of the most popular screen readers for persons with visual impairment and blindness is Job Access with Speech (JAWS), which is another effective software solution that reads out loud the text present in the screen of electronic devices to users.

Spatialized speech is also explored through auditory interfaces for computer users who are visually impaired [10], where JAWS as a screen reader equipped with an ALVA 544 Satellite braille display is compared with this custom-made auditory interface based on spatialized speech. However, such devices are designed for the spatial location of text to provide positioning context within a page or in some graphical computer environment. Another useful tool is Google Translate, a free software tool that can be used to read and translate text present in images obtained from the device's camera. A more inclusive and complete software is the KNFB (Kurzweil- National Federation of the Blind) award winning reader, which incorporates image capture, by utilizing the device's camera, processing, and text to synthesized speech conversion. These are some of the tools currently available and while they do have benefits not they do not have all the features required for a book reader.

In terms of assistive technology tools, a thorough assessment of the many technologies that are useful for persons with visual impairment and blindness is provided in [11]. Although its focus is on students with visual impairment and blindness, the main intent is to explore how assistive technologies could improve the learning and functional capabilities of all blind and visually impaired users. Some systems use two cameras and range sensors to gather data about a surrounding environment to enhance navigation. One example is the case of [12] where a Tyflos wearable system provides a 3D space representation made possible in part by utilizing stereo vision. The concepts behind the portable framework were also interestingly applied to document reading as was explored in [13], where the Tyflos reader responded to voice commands to perform many tasks, such as, but not limited

to, image capture, dewarping and segmentation. Other tools include spectacles and handheld video magnifiers that enable access to reading materials to individuals with low vision impairments. Whereas, audio books and digitized text, read by text to speech (TTS) software, are more general and inclusive tools. It is clear that the overarching theme for such technologies is universal accessibility [14, 15], but with the intent to exploit, in an optimal fashion, the information acquired by the camera(s) to yield useful descriptions of the viewed environment [16, 17].

The proposed design aims to help individuals with low and high visual impairments alike. As a testament to the operational success of our initial book reader design which was based on a dual digital camera system [18], four individuals with blindness from the Lighthouse of Broward, along with its director, have expressed full satisfaction with the reading accuracy of the initial prototype of our system after having evaluated and tested it. They have also indicated that they were not worried about the text recognition errors because they're quite adept at using context to overcome them and still capture the meaning as text is read to them. One of them expressed desire to have such a system incorporate handwritten text recognition so it could read personal notes, cards, and letters in the privacy of his own home. Although the current implementation of the proposed system only deals with printed text, future implementation will incorporate the ability to read hand written text as well as mathematical formulas and equations. In fact, our group has looked into the task of recognizing handwritten text using an extensive public database [19, 20]. The challenge remains in addressing the heavy computational requirements before hand-written text can be integrated into the proposed book reader design.

### 1.3 Prior correction and dewarping studies

Several book and document reader designs with different system set ups, architectures, and dewarping approaches have previously been proposed. A great deal of information regarding the difficulties encountered and how they are addressed by document readers can be found in [21]. The ultimate goal in these designs is the proper recognition of characters in documents. To this end the survey compares different characteristics of different document image processing methods, and provides different insights on the manner in which images are enhanced by adjusting capture environment, increased resolution, post processing to address distortions, among other imaging challenges.

With the same intent to help persons with visual impairment and blindness, related studies have considered alternative design options and methodologies. Remarkably, the work in [22] uses a single high-resolution camera, and the dewarping process is accomplished by taking into consideration cues present in the image. By supposing baseline fitting, assumed to be straight, horizontal and vertical vanishing point estimations on the perspective transformation on parallel lines are assumed for the text. In this study, although they report a high reading accuracy of nearly 96%, the rectified text seemed skewed, with some words being difficult to read to the naked eye, while others appear italicized next to some correctly rectified. In another study, the shape is obtained from shading rectification, where the shape of the document is inferred from the shadow it casts, has been explored [23]. Although this approach yields positive results, it is dependent on lighting conditions and is prone to variations. Whereas [24] aims to correct deformations of planar documents by using minimization functions that yield rectification transformations.

Studies in [25, 26] are rather unique as they propose to convert e-books or e-documents available in PDF into formats that are easily accessible by persons with visual impairments (i.e., Braille, audiotape, or large print format). Understandably, descriptions of figures and graphs must be added manually in a semiautomatic process [25]. In a sophisticated image analysis strategy and seeking a higher reading accuracy, the study reported in [27] integrates the so-called iconic model (character-image classifier) and a linguistic model (word occurrence probability) with mutual entropy being the measure that assesses the disagreement between the two models (a high mutual entropy would mean a disagreement or that the answer given out by the iconic model has no corresponding entry in the dictionary). The hamming-distance-based template matching is used for classification in the iconic model relying on the character classifier. Other research groups have considered automating reading and text tools for visually impaired persons [28] to be able to effectively interact and construct mathematical expressions. This case focused on linear algebra as a good example to facilitate access to mathematics.

The work in [29] focuses more on the removal of the so-called moiré-pattern noise and specular highlights are present in document images acquired through a smartphone in order to benefit the OCR reading outcome. The challenge of recognizing and then reading text also extends to other languages [30], where the interesting case of Arabic text, which is cursive in its writing nature and where characters assume different shapes as a function of their location in a word (start, within, and ending of a word), is explored. The authors in [30] proposed using different Arabic characters including standard characters (28 with no dots in them) and other more unique characters that have dots in them, which can be found

above the character (1 to 3) or below it (1 or 2). Then use a sliding window for feature extraction which are fed into a Hidden Markov Model for text recognition. Therefore, the recognition in this case is a probabilistic pattern recognition process that bypasses any segmentation process.

However, there is still room for improvement in these systems, especially in terms of recognition rate and reading accuracy. For instance, during image capture, warping effects occur due to the curvature of the book. These aberrations make the character recognition process challenging and error-prone. Methods using special de-warping or flattening mechanisms extract orientation information from cues implied directly from the printed text or from shading information to rebuild the 3-D surface [22, 23, 24]. A thorough and interesting survey [31] highlights the merits and challenges of the different existing systems and approaches used for reading documents.

Several studies have considered the challenge of de-warping documents in order to improve text recognition and ultimately the reading accuracy by means of existing OCR. Interesting studies have been reported on means to rectify distorted text by either estimating the curled text lines [32, 33], looking into the geometric properties of the captured images [34], or using two images from a single camera at different angles to estimate the document's surface from corresponding points akin to stereo vision. Several studies proposed different flattening mechanisms through either grid modeling [35], 3D shape modeling [36], by considering parallelism and equal line spacing [37], modeling lines as wave-fronts of planar waves [38], looking into slopes in words for realignment [39], or by



using curved surface projection modeling and baseline estimation as a goal oriented rectification [40].

The concepts of de-warping and flattening have also been explored in applications that include the restoring and enhancement of historical documents [41, 42], and for recognizing street signs from 3D scenes [43]. Cylindrical surface modeling has also been considered through the use of an isometric mesh [44] or through optical flow and structure from motion, by using the camera of a mobile phone, which is swept across the book spread for image capture [45]. In retrospect, is possible to agree with the assessment provided in [46] that in many surveys related to assistive technology, there is a lack of details in the technical and functional improvements that such devices provide.

#### **1.4 Problem definition and contributions of this dissertation**

This dissertation focuses on the development of a fully automated book reader design. First, I overcame distortions and alignment issues that arise throughout the image capturing process, whether it is from the ToF sensor for the 3D map or from the high-resolution camera for reading the text. Furthermore, one of the essential objectives of this dissertation is to solve the book curvature as a surface geometry problem, as obtained directly from depth extraction [47, 48, 49]. With notable advances made in Time of Flight (ToF) sensors, together with their increased range of applications [50, 51, 52, 53, 54], it is now possible to obtain depth maps that reflect the book's curvature with sufficient height accuracy. However, current technological limitations allow for only a small image resolution to be obtained from the ToF devices. Therefore, this dissertation also addresses this problem by

using a high-resolution digital camera and designing a co-registration process to match the high-resolution image with the low-resolution image output from the ToF devices. The matching process also yields a high-resolution height map which can in turn be used to obtain the height map of the book spread. This dissertation then explores the benefits of algorithms relying on height maps to improve character recognition through distortion correction. The focus is on using two different height maps: 1) a full height map as obtained from a full scan of the book's surface by a ToF sensor, and 2) a uniform height map where the values from the central row of the 3-D map are used for all other rows (as was originally used in [55]). Once these maps are determined, the curvature correction process takes place, and the corrected images are subsequently fed into the OCR engine. The resilience of our book reader's correction algorithm is further explored by rotating the book slightly ( $\pm 30^\circ$ ) to the left or to the right to introduce a type of misalignment that could happen when a user places the book on the reading platform of the book reader.

## **1.5 Dissertation organization**

This dissertation will detail the theoretical as well as the experimental set up of the approach considered in this dissertation for the design of a portable, fully automated, and highly effective book reader design, both in terms of its computational capability and high reading accuracy. To this end, this dissertation is structured as follows:

Chapter 2 defines the theory and concepts to correct curvature of books which are known to cause a warping effect in the book spread images that need to be read. Furthermore, this chapter elaborates on the experimental setup used to prove the validity of the approach,

and describes the integration aspects of the equipment/hardware elements and the software developed in support of the mathematical framework and system geometry.

Chapter 3 elaborates on the implementation steps that are necessary to perform the correction, as well as provide insight into the correction process based on the initial results obtained.

Chapter 4 explores other means to improve the reading accuracy—specifically, the effect the height map has on the curvature correction process.

Chapter 5 provides a more thorough investigation of the so-called uniform vs full height maps, and reports on some of the adjustments that were made to understand how the reading accuracy could be optimized.

Finally, Chapter 6 concludes with a retrospective on the different design steps that led to the optimized version of the fully integrated and automated book reader design proposed here. It also provides an analysis of the techniques used and subsequent results obtained, as well as offers insights for further research and subsequent improvements to character recognition as part of the book digitalization effort as well as being able to improve the quality of life of blind individuals.

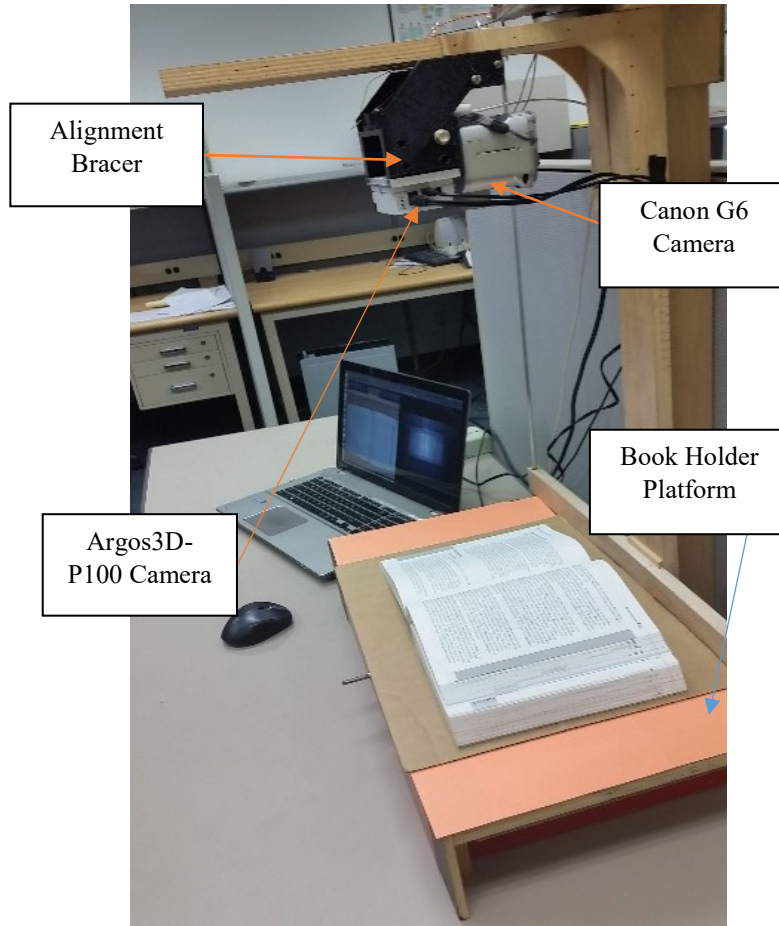
## 2. CONCEPT OF DESIGN AND THEORY

### 2.1 Book reader design

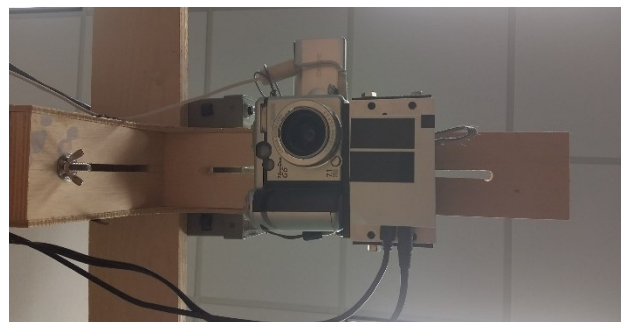
The book reader design is shown in Figure 2.1, a Bluetechnix Argos 3D-P100 depth sensor is used to capture the book's depth data. The information provided by this sensor is captured based on the principle of time of flight (ToF). This sensor can produce a grayscale low resolution height map image of 160x120 pixels, which was converted to a height map, and fused with the high-resolution image using homogeneous transformations. The high-resolution camera used in our setup is the Canon G6, an inexpensive camera which offers a resolution of 7.1 megapixels. The high-resolution image of this camera is needed as input to the OCR engine, once dewarped to overcome the curvature effect, to attain the highest reading accuracy possible. The two camera devices of the proposed design were setup as shown in Figure 2.1

With this setup in place, the full height map of the book's surface was obtained by subtracting the average depth map of 60 frames of the book stand from the average depth map of 60 frames of the book's surface. This averaging is performed in order to obtain reliable measurements of depth maps for use as height maps for estimating curvature of the book spread. The uniform height was created by using only the center row of the full height map and replicating it for all the other rows. Both these types of maps were obtained for each of the book spreads. In this design, the capture planes are evidently different for both devices and so are the resolutions of the two images. A challenge faced when fusing the

two images is the difficulty of matching the sparse height map determined from the low resolution ToF device to the high-resolution image that is used for reading text.



(a) Design setup



(b) Placement of the cameras for fusing the 3D height map to the high-resolution book spread image

Figure 2.1 Experimental setup of the book reader design

## 2.2 Concepts for the curvature correction method

The location of a given point or object (in this case text) relative to the camera position may result in the so-called warping effect due to page curvature. Ideally, a high-resolution depth camera would be sufficient, but the unit price for such 3D camera is prohibitive. To address this challenge, the low-resolution plane is fused through affine transformations to the main high-resolution plane. The distance from the high-resolution camera to the object remains fixed, then the factor affecting the display can be attributed to the height of the object, which in this setup reflects the curvature of the book spread. Therefore, the desired dewarping process, which includes the extension of the curved pages, was derived based on the conceptual design shown in Figure 2.2, introduced in earlier studies of our research group [48, 49, 56].

In Figure 2.2,  $V$  is defined as the distance of the image sensor from the lens,  $U$  is the distance to the base of the platform where the book is placed from the lens, and point,  $B_1$  given by  $(X, Y, -U')$  is a point along the book spread with a height of  $H(X, Y)$ . When this location is captured by the camera, it yields the warped location  $(x^*, y^*, V)$  as noted by point  $C_1$ . However, the desired flattened location on the image sensor plane should be instead located at point  $C_2$ , position shown since the point  $(X, Y, -U')$  should be at point  $B_2$ , location when it is flattened in reference to the book holder platform plane.

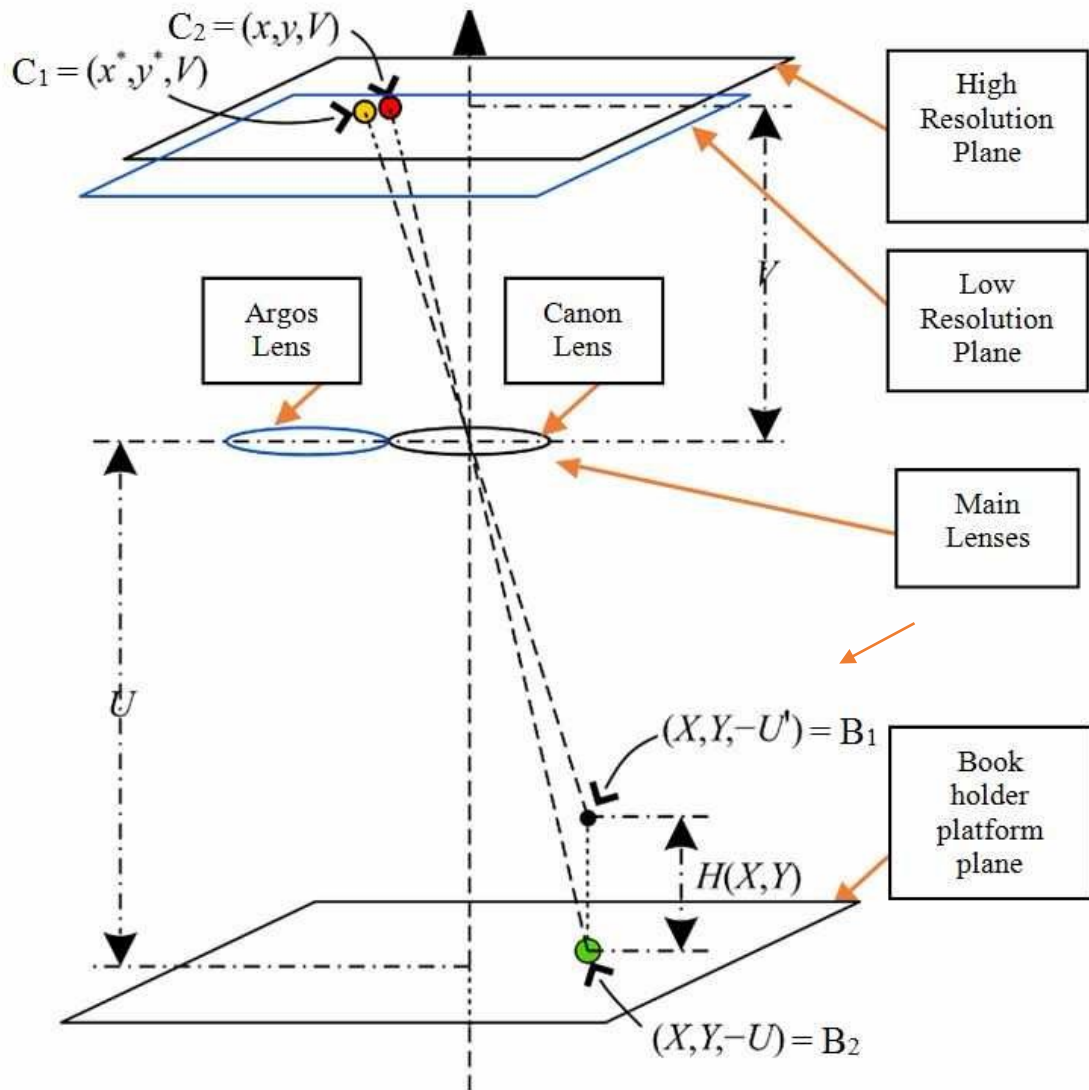


Figure 2.2 Conceptual design and geometry of the book reader

To make the necessary correction, a modified lens equation is used. Using the properties of the lens and the geometrical layout of this system design, the following relationship is established:

$$y/Y = x/X = V/U \quad (2.1)$$

Thus, by taking into account the height measure  $H(X, Y)$ , the 2-D warped coordinates could be derived using equation (2.2) as follows:

$$x^*/X = V/(U - H(X, Y)), \text{ and } y^*/Y = V/(U - H(X, Y)) \quad (2.2)$$

Since location  $(X, Y)$  on the book spread is unknown, and because the desired location is the corrected location  $(x, y)$  in the captured image, the height in terms of  $x$  and  $y$  can be expressed as follows:

$$H(X, Y) = H(x \times (U/V), y \times (U/V)) = h(x, y) \quad (2.3)$$

Measure  $h(x, y)$  is the height to be recovered from the unwarped image. Since the warped location  $(x^*, y^*)$  is known, and in order to obtain  $h(x, y)$ , our assumption is that the changes along  $x^*$  and  $y^*$  are minimal, allowing for  $h(x, y)$  to be estimated as:

$$h(x, y) = h(x^* + \Delta x^*, y^* + \Delta y^*) \approx h(x^*, y^*) \quad (2.4)$$

Therefore, the corrected location  $(x, y)$  on the captured image plane can be expressed as follows:

$$x = x^* \times (U - h(x^*, y^*)) / U, \text{ and } y = y^* \times (U - h(x^*, y^*)) / U \quad (2.5)$$

Applying this transformation (2.5) along all points will allow for the image to be flattened. To further improve these results to yield characters that look more natural, the book spread needs to be extended, as if one is pressing on the curved part near the spine of the book and see the pages extend over to the left and right sides of the book. This can also be observed when a book is pressed on a scanner or copier when making a copy to overcome the effect of the book curvature. Given the differential change in height defined as:

$$dh_i = h_{i+1}(x^*, y^*) - h_i(x^*, y^*) \quad (2.6)$$

With the differential change along the x direction being 1 pixel as in (2.7)



$$dx_i = x_{i+1} - x_i = 1 \text{ pixel} \quad (2.7)$$

The extended lengths ( $L_i$ ) can then be calculated as follows:

$$L_i = (dh_i - dx_i)^{\frac{1}{2}} \quad (2.8)$$

Using equation (2.8), as a function of  $x_i$ , it is then possible to express the new extended locations as:

$$x_{e_{i+1}} = \sum_{i=0}^n (x_{e_i} + L_i) \quad (2.9)$$

Where  $x_{e_0} = x_0 = 0$ , representing the starting point at leftmost beginning of the book, and  $n$  is the number of rows in the high-resolution image. This extension adjustment is applied after the flattening process has been performed. This extension process is akin to pressing near the spine of the book in an effort to flatten out the pages. But this process is challenging in that it extends the book spread to the left and to the right of the spine of the book and hence effects the dimension of the text as reflected in the resulting images in the following sections and chapters. The results would confirm later, the flattening process followed by the extension adjustment allowed for the image to be corrected, attenuating significantly the warping effect consequently.

There are certain considerations and modalities of operation that need to be followed in order to optimize the effectiveness of the design. The implementation steps taken for appropriate calibration, overcoming the different distortions, and registering the two imaging modalities of the ToF sensor and high-resolution camera to together with the initial results obtained are presented in Chapter 3.

### 3. IMPLEMENTATION STEPS AND INITIAL RESULTS

This chapter will go over the implementation steps required for establishing a multi-step mathematical foundation that ensures a fully-automated and highly accurate book reader design. The first step is a calibration process needed to ensure the accuracy at extracting the depth map. Specifically, for this application, the distance from the sensor to the surface of the book holder platform was measured using external tools and was determined to be 80 cm. Therefore, the depth acquisition distance on the ARGOS3D - P100 was set to 100 cm to ensure the required depth was obtained. Information about adjustments and calibration that are possible can be found in the camera manufacturer's website [56]. Also, as it has been demonstrated in [52], to further ensure better depth resolution, that a native bilateral filter can be used. The focus is thus placed on the book spread area. The ARGOS3D - P100 can be used to capture multiple frames per second, yet setting this too high could cause the device to overheat, which negatively impacts the depth information. In the case of the book reader, this setting was not changed from the default setting since the book is static. However, to reduce inaccuracies due to depth sampled variances, an average of 60 depth map frames was obtained for both the book holder with and without the book.

#### 3.1 Height maps

The previously derived equations in Chapter 2 rely on using a height map rather than a depth map directly. Therefore, the height of the book is then calculated by taking the averaged depth map with respect to the book holder platform and subtracting it from the

averaged depth map of the book spread. This height can be used directly as a raw height and is expressed as:

$$h(x^*, y^*) = \text{Raw Height} \quad (3.1)$$

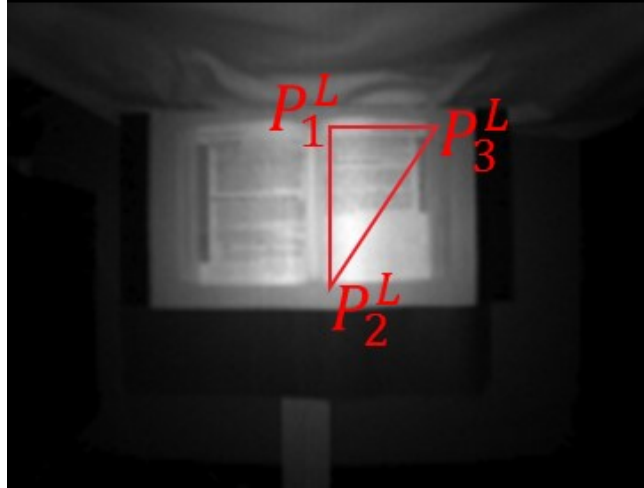
However, another approach would be to use polynomial approximations of given order  $n$  with  $p$  coefficients of the height points for each of the rows as shown in equation (3.2):

$$h(x^*, y^*) = \begin{cases} p_{11}x^{*n} + p_{12}x^{*n-1} + \dots + p_{1n}x^* + p_{1n+1} & , y^* = 1 \\ \vdots & \vdots \\ p_{m1}x^{*n} + p_{m2}x^{*n-1} + \dots + p_{mn}x^* + p_{mn+1} & , y^* = m \end{cases} \quad (3.2)$$

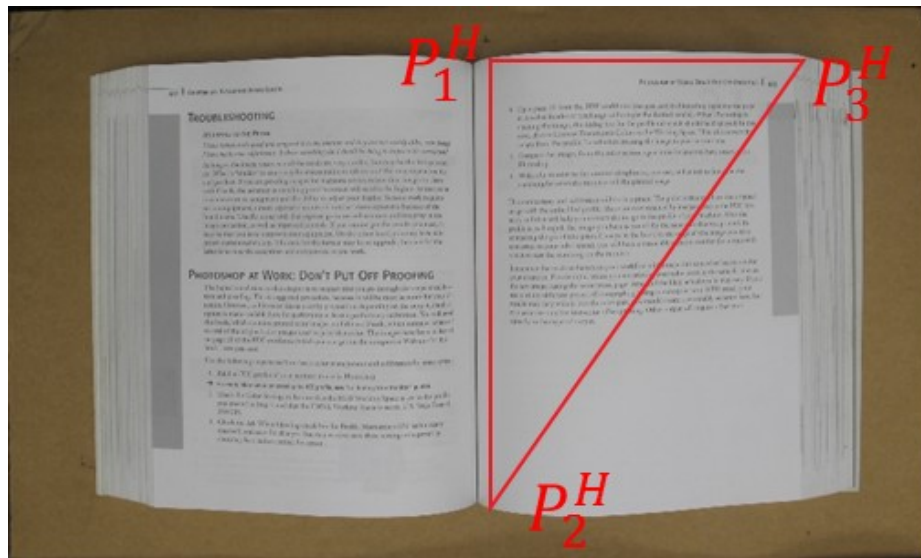
### 3.2 Registration of imaging modalities

The idea of combining low and high-resolution images to overcome the current resolution deficiency of time-of-flight device has been proposed in [51] and [54]. In the setting of this design, the conversion of the low-resolution image was then accomplished using homogeneous (affine) transformations of rotation and scaling about the origin (top left most point) and translation until the center of the two triangles are matched. These affine transformations were derived on the basis of the geometry of the virtual triangles as defined by the three points, denoted as  $(P_1^L, P_2^L, \text{ and } P_3^L)$  in the low-resolution image as shown in Figure 3.1(a) and as  $(P_1^H, P_2^H, \text{ and } P_3^H)$  in the high-resolution image as shown in Figure 3.2(b). The new pixel values were then estimated using bilinear interpolation.

The conversion of the low-resolution image is then accomplished by using the homogeneous (affine) transformations of scaling, rotation (about the center of the image), and translation until the center of the two triangles are matched.



(a) The 2-D low-resolution book spread



(b) The 2-D higher-resolution book spread as that found in Figure 2.2.

Figure 3.1 Geometry for registering the two imaging modalities

The steps to perform such corrections are as follows:

(1) Determine the measurements of the vectors:

$\vec{v}_{H1,2}$  defines the vector from  $P_1^H$  to  $P_2^H$  and

$\vec{v}_{H1,3}$  defines the vector from  $P_1^H$  to  $P_3^H$

$\vec{v}_{L1,2}$  defines the vector from  $P_1^L$  to  $P_2^L$  and

$\vec{v}_{L1,3}$  defines the vector from  $P_1^L$  to  $P_3^L$

(2) Compute the scaling factor of the low-resolution image to the high-resolution one as defined in equation (3.3)

$$S = \frac{\|\vec{v}_{H1,2}\|}{\|\vec{v}_{L1,2}\|} = \frac{\|\vec{v}_{H1,3}\|}{\|\vec{v}_{L1,3}\|} \quad (3.3)$$

(3) Determine the angle of rotation  $\theta$  between the low and high-resolution images using equation (3.4)

$$\theta = \cos^{-1} \frac{\|\vec{v}_{H1,2}\| \|\vec{v}_{L1,2}^*\|}{\|\vec{v}_{H1,2}\| \|\vec{v}_{L1,2}\|} \quad (3.4)$$

Once the scaling and rotation operations are performed based on the scale factor and angle of rotation as determined in equations (3.3) and (3.4), respectively, new reference points are computed for the new low-resolution image, where  $P_1^L$  becomes  $P_1'^L$ ,  $P_2^L$  becomes  $P_2'^L$ , and  $P_3^L$  becomes  $P_3'^L$ .

At this point, the reference triangles in both images are of the same size and orientation, but their centers are offset from one another. Therefore, the translation required along the x and y axes is determined through the  $\Delta x$  and  $\Delta y$  measurements found using equation (3.5).

$$\Delta x = P_{1x}^H - P_{1x}'^L \text{ and } \Delta y = P_{1y}^H - P_{1y}'^L \quad (3.5)$$

When these affine transformations are applied to the original low-resolution image shown earlier in Figure 3.1 (a), the result is the image shown in Figure 3.2, which is still a low-resolution image but now has the same dimension as the high-resolution image of the Canon G6 camera.



Figure 3.2 The 2-D converted low-resolution book spread

Bilinear interpolation, as the next step, allows me to generate a complete (full) height map to fill in for all the missing pixels in between when converting the low resolution from its original dimensions of 160 x 120 pixels to the 3072 x 2304 pixels of the Canon G6 camera. The result of combining the low-resolution image and low-resolution map is shown in Figure 3.3.

Once the registration and matching of the images are accomplished, using the same transformation on the low-resolution height map, a new high-resolution height map can be obtained, yielding the results shown in Figure 3.4.

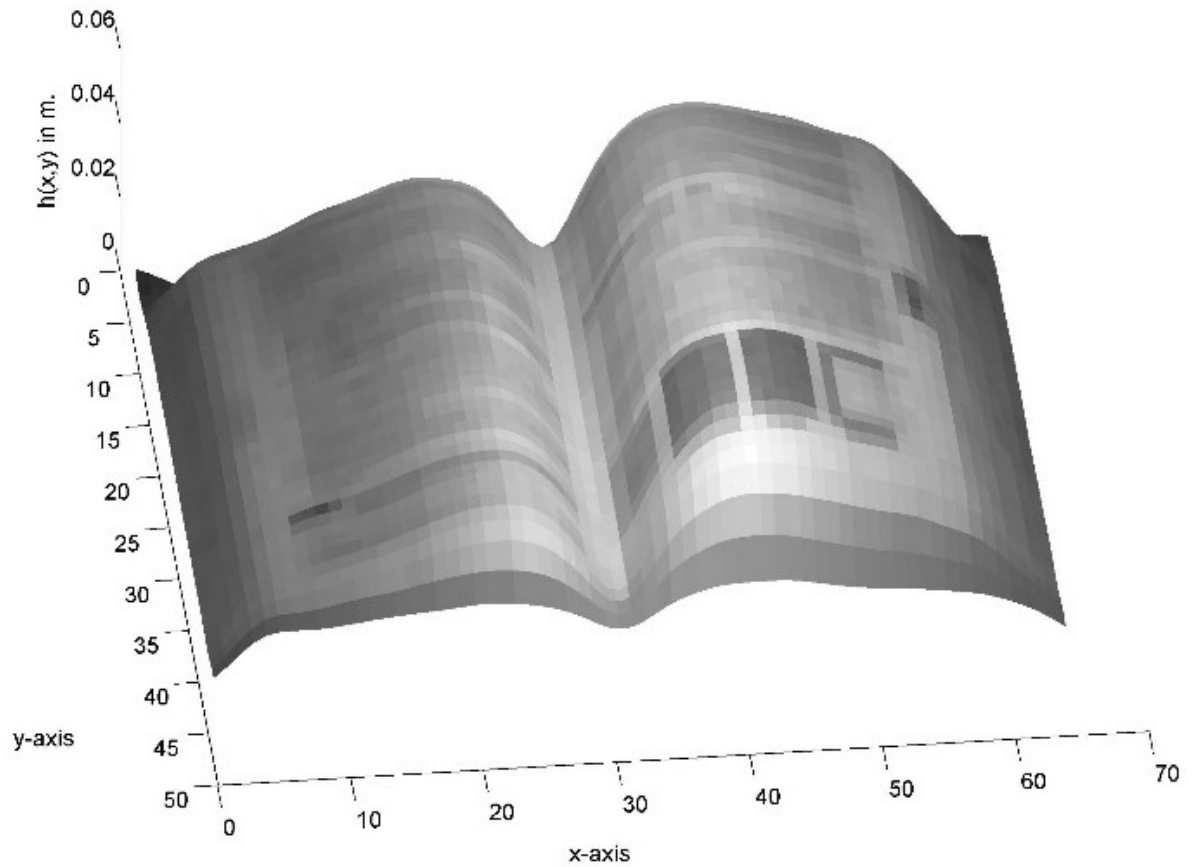


Figure 3.3 The 3-D full height map with the low-resolution image

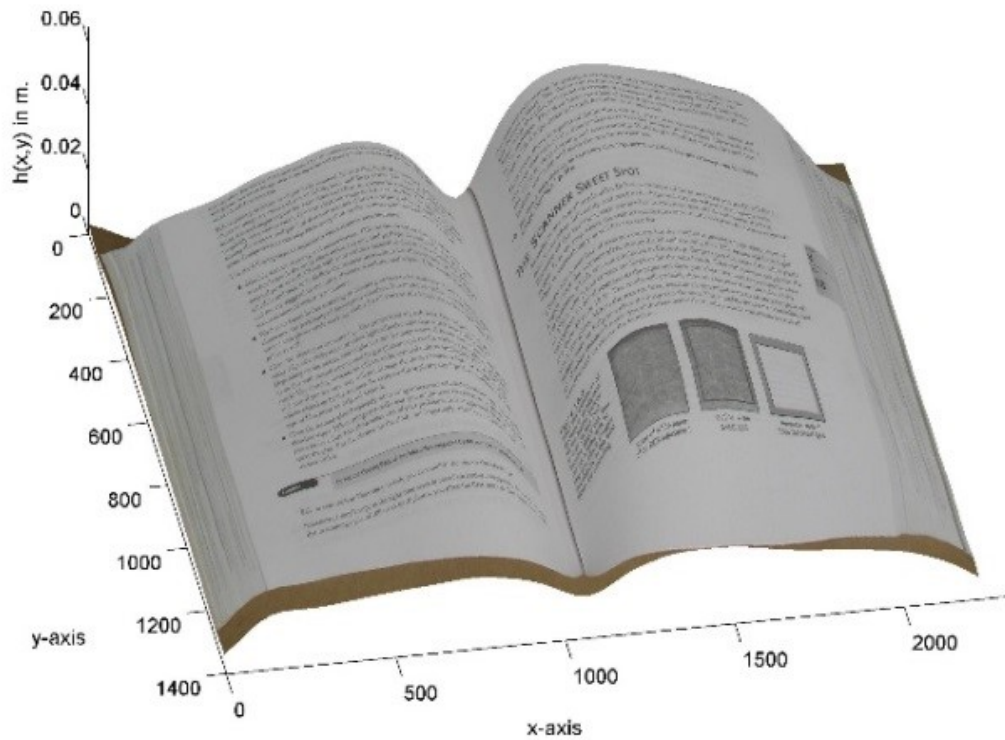


Figure 3.4 The 3-D higher-resolution full height map with the high-resolution image

### 3.3 Flattening and extending correction algorithm

The height data of the pixels can be used raw as expressed in equation (3.1) or approximated via polynomials as expressed by equation (3.2) to then initiate the page flattening process using equations (2.5). However, when using these equations, the value of  $U$  needs to be specified, which can be measured or determined experimentally. In this implementation,  $U$  was calculated by placing a single flat page with label points on the book holder platform and then changing the height to a known elevation. For this setup,  $U$  was determined to be 58.05 cm.



The flattened image can then be extended by calculating the change in height and the change along the x direction using equations (2.6) and (2.3), respectively. It should be noted that either one of these equations should be converted so that the units match (i.e. in pixels or in centimeters). Given the geometry of the experimental setup, the conversion value was 2668/58 (Pixel per centimeters). Hence, once converted, equation (2.8) can be used to calculate the extended lengths as expressed in (2.9). The image can then be reconstructed via scattered interpolation using any number of methods, such as linear interpolation, natural neighbor interpolation, and nearest-neighbor interpolation. The resulting images can then be introduced to the OCR engine for character recognition and for the text to be read to a person with visual impairment, or for the text to be digitized and saved in memory for future use. For testing purposes, the ABBYY FineReader 12 OCR engine was used for character identification, which in turn can then be used to quantify the performance of the correction.

### **3.4 Initial testing**

In retrospect, once the ARGOS3D - P100 was calibrated and adjusted, a typical book spread image obtained was shown in Figure 3.1, where the resolution of the book spread image is limited to only 65x41 out of the 120x160 of the full resolution of the ToF camera. Combining the pixel values with the raw height values as express by equation (3.1) yielded the results shown in Figure 3.3. However, instead of using raw height values, a polynomial approximation such as equation (3.2) could be used as an alternative approach to provide continuous height maps. Different orders of equation (3.2) were attempted; however, the

polynomials either did not match the shape of the book or they suffered from the lack of smoothness. Therefore, for the raw height data of Figure 3.3, we settled for using a 25<sup>th</sup> order polynomial. Comparing the raw height points with the polynomial points which gave the smallest deviations as shown in Figure 3.5.

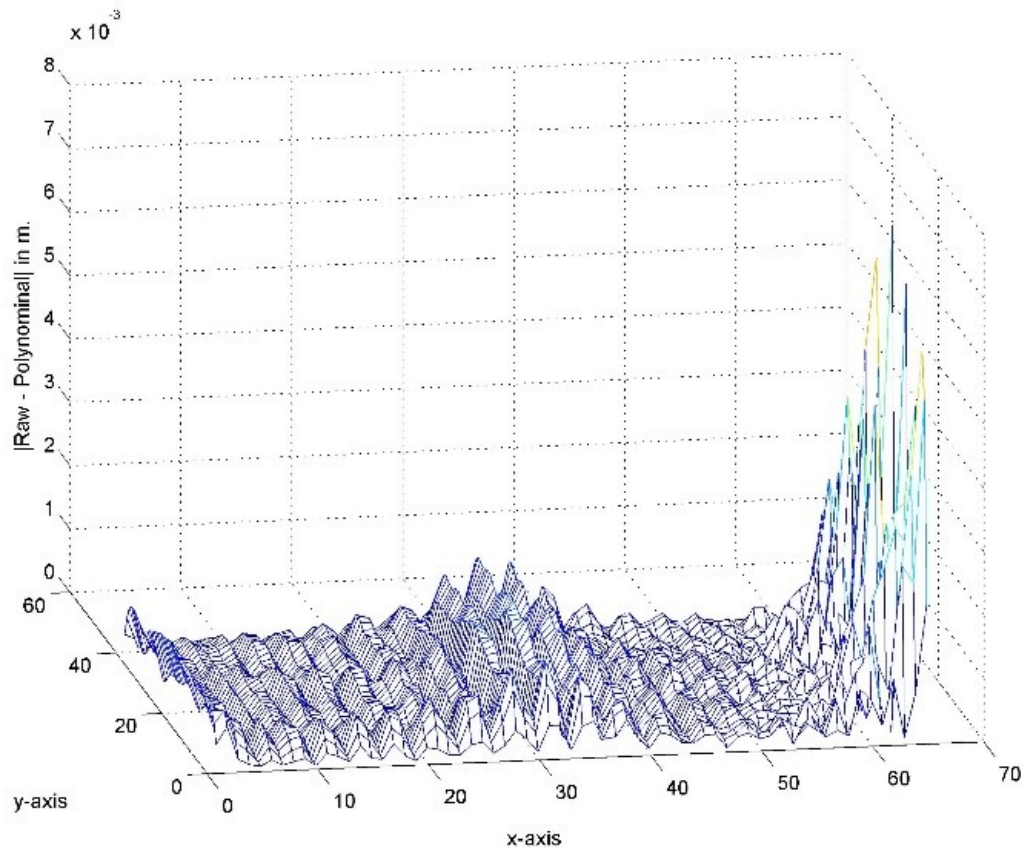


Figure 3.5 Comparison of low-resolution raw height map and polynomial fitting by subtracting both values

It can also be seen that the most noticeable differences occur along the spine of the book spread as well on its edges. Using this polynomial height fitting along with the low-resolution image in Figure 3.1, yields the results shown in Figure 3.6.

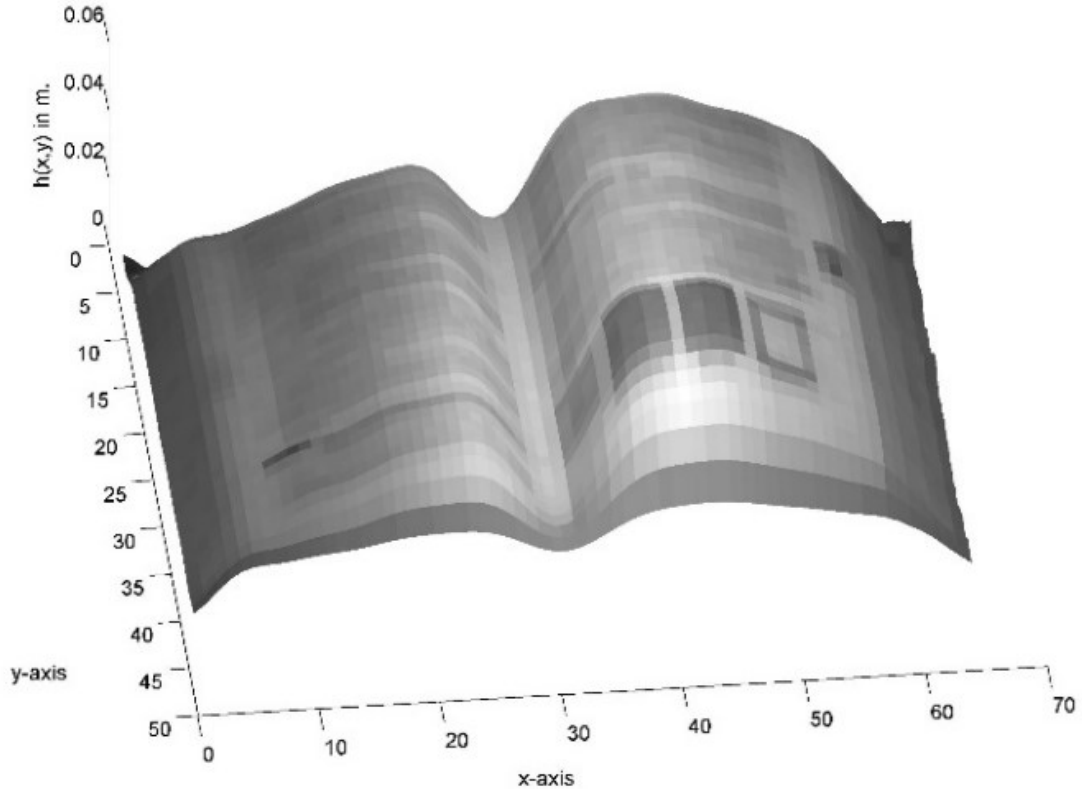


Figure 3.6 Low-resolution height polynomial fitting with corresponding low-resolution pixel values.

The flattening and length extension corrections would have been made at this point of the process; however, due to the low-resolution of the ARGOS3D - P100 camera, the text is illegible. Because of this drawback, the use of the higher-resolution 7.1-megapixel Canon G6 camera became a necessity. This higher-resolution is evidently more suited for OCR applications. The high-resolution image of the book spread to be used is given in Figure 3.2 which has a resolution of (2226 x 1374) out of the 7.1 megapixels available, as taken by the Canon G6 camera. This high-resolution image was then matched to the raw data from Figure 3.3 and separately to the polynomial fitting data from 3.6. Then the higher-resolution height map (raw and polynomial fitted, respectively) was calculated and combined with the higher-resolution image to yield the results shown in 3.4 (using raw

data) and Figure 3.7 (using polynomial curve fitting) were the text is now legible as confirmed by the reading accuracy of the OCR results.

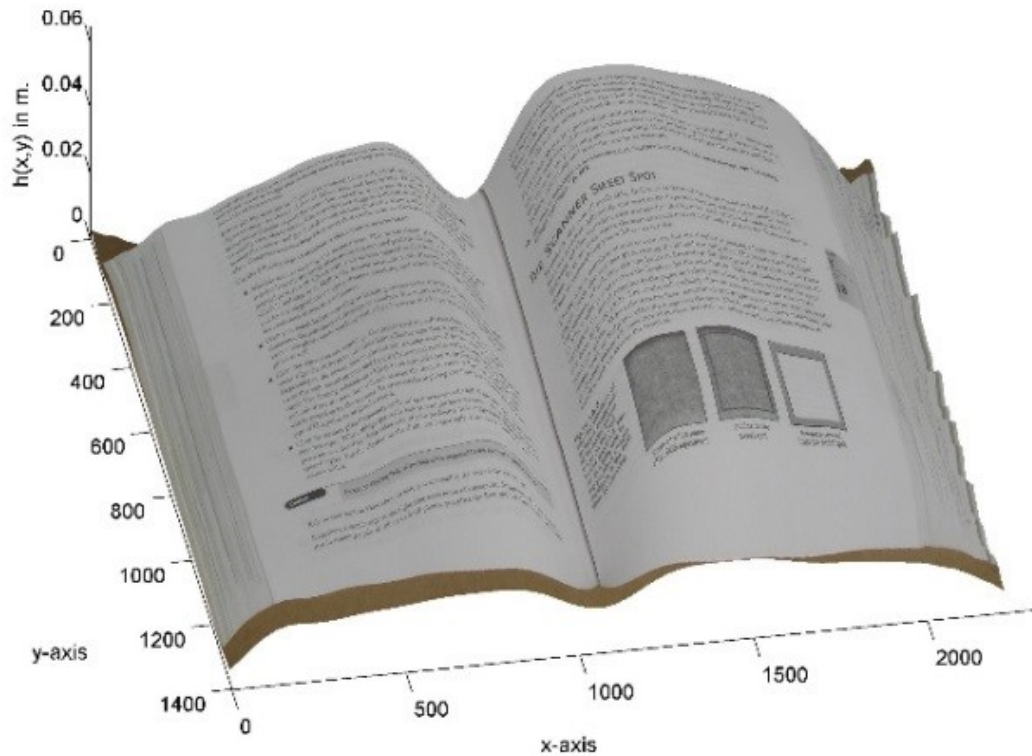
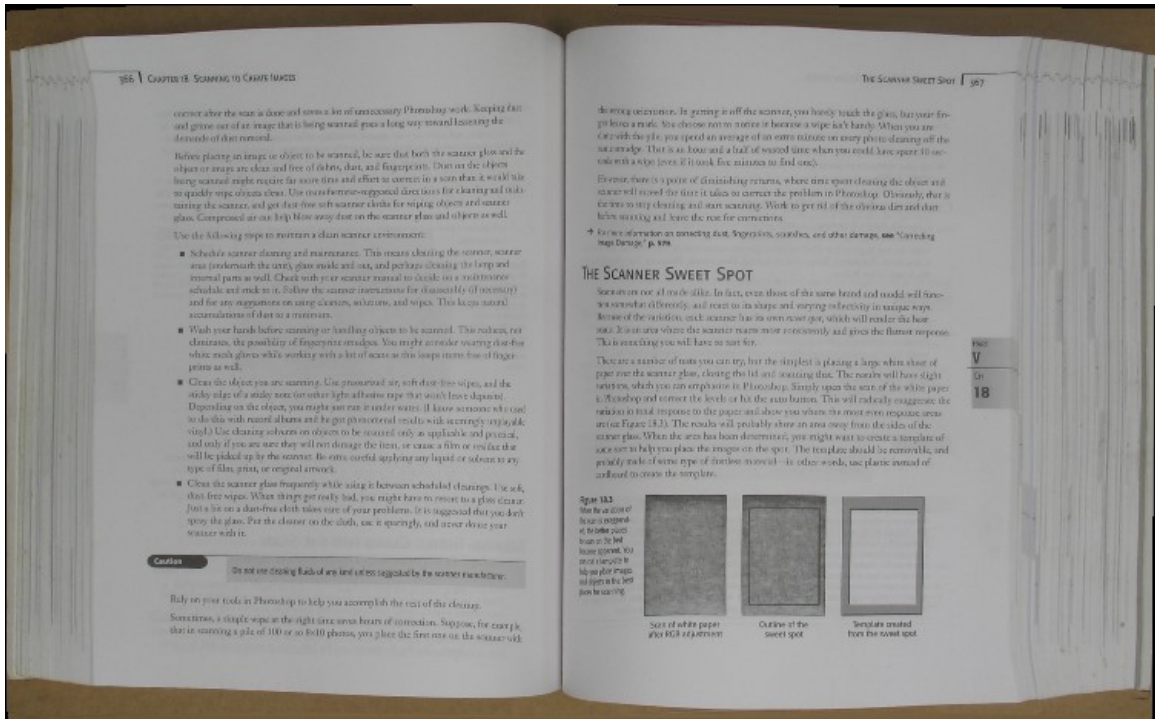


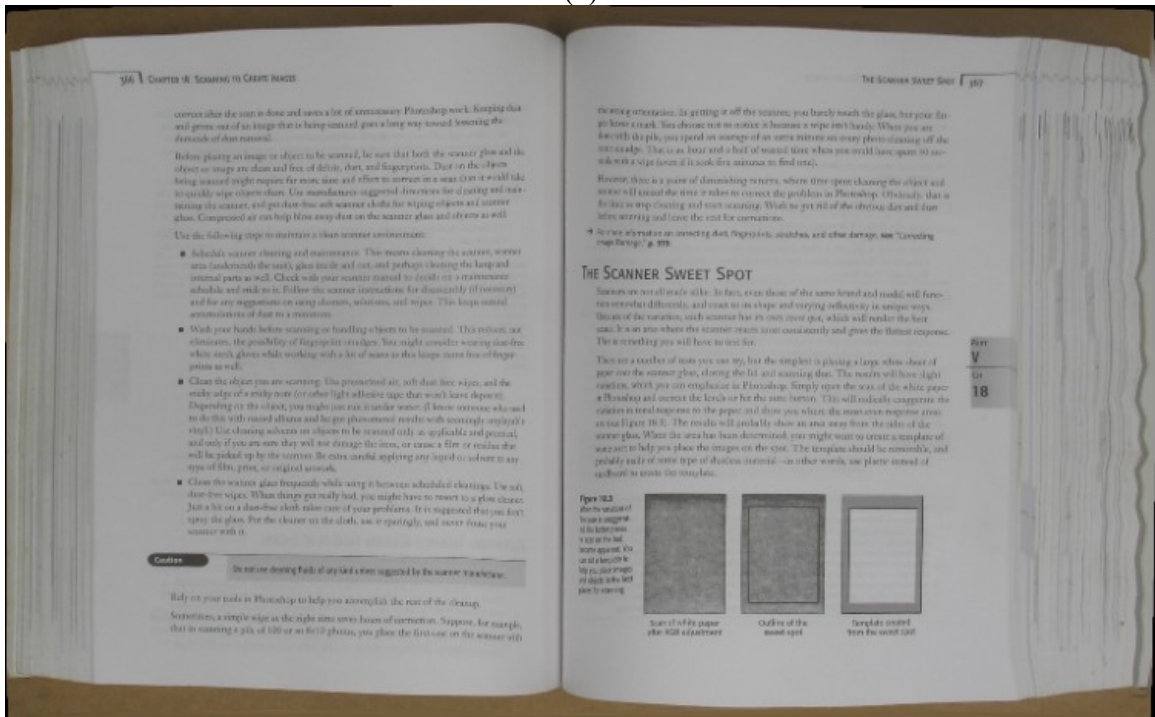
Figure 3.7 Higher-resolution height polynomial fitting with corresponding high-resolution pixel values

### 3.4.1 Flattening correction effect

At this stage, the flattening transformation process is applied using both the raw height data and the polynomial approximation of the book curvature. In Figure 3.8, the book spread correction using the raw height values is shown in (a), while the height polynomial fitting correction is shown in (b).



(a)



(b)

Figure 3.8 Corrections via (a) raw height & (b) polynomial fitting.

A zoom-in on the text provides better visual appreciation of the effect of the correction. Also, to further illustrate the effect on the characters, excerpts of the optical character recognition (OCR) engine results will be included in the following figures. Where Figure 3.9 shows partial results of the OCR engine on the top corner of the original warp book spread previously seen in Figure 3.2. In Figure 3.10 shows the same top corner flattened with the raw height map as well as the corresponding OCR results. Lastly, Figure 3.11 shows the results of the correction with polynomial approximated height map. The flattening effect is visually noticeable in contrast with the original image which has this wave like undulation, while the difference between the two flattened excerpts are not as obvious to the naked eye as the impact on the OCR results.

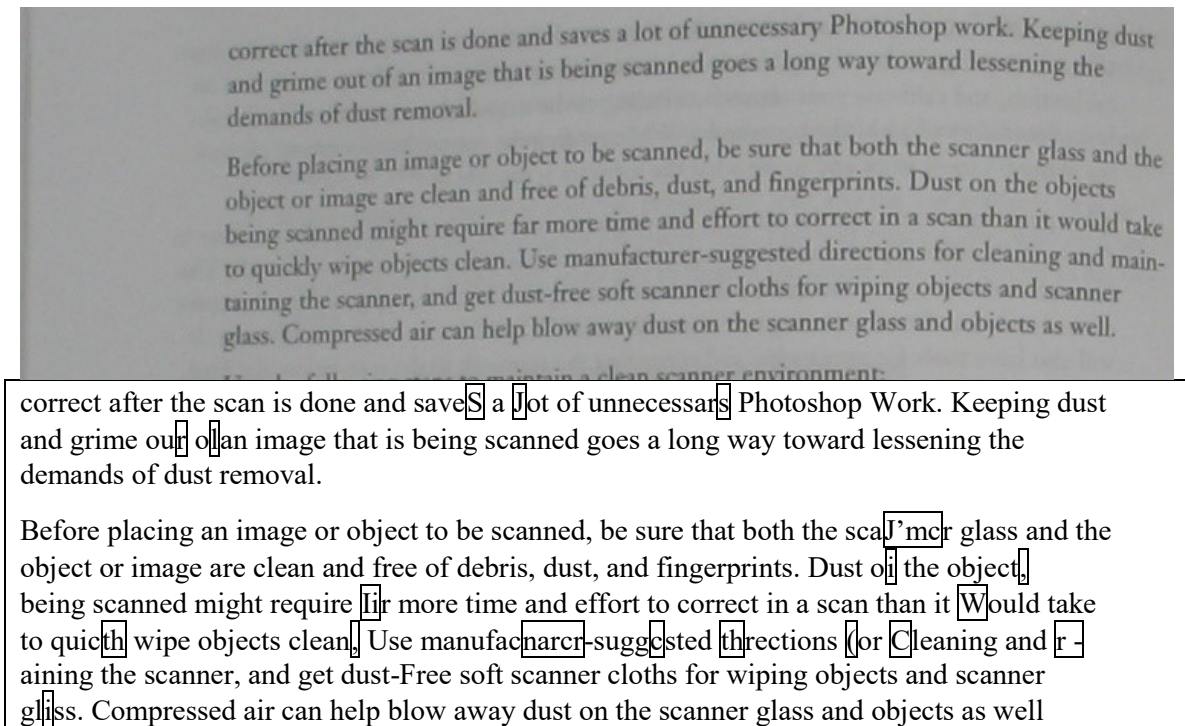
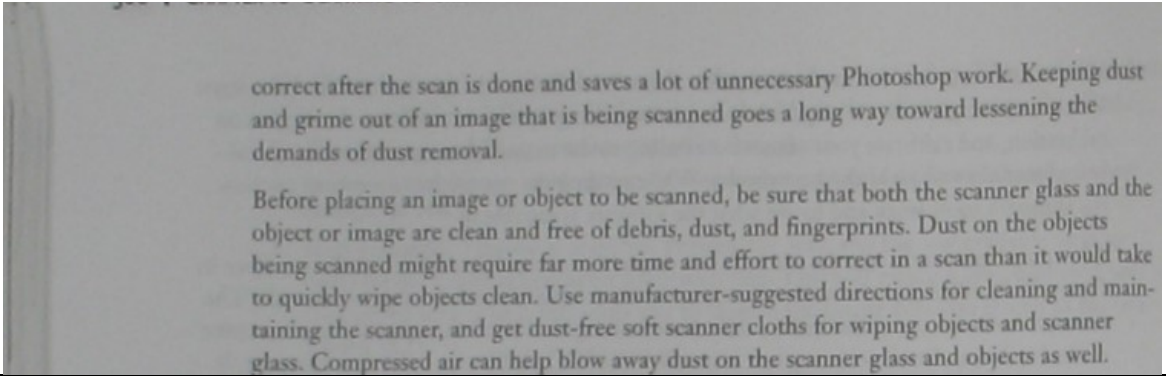


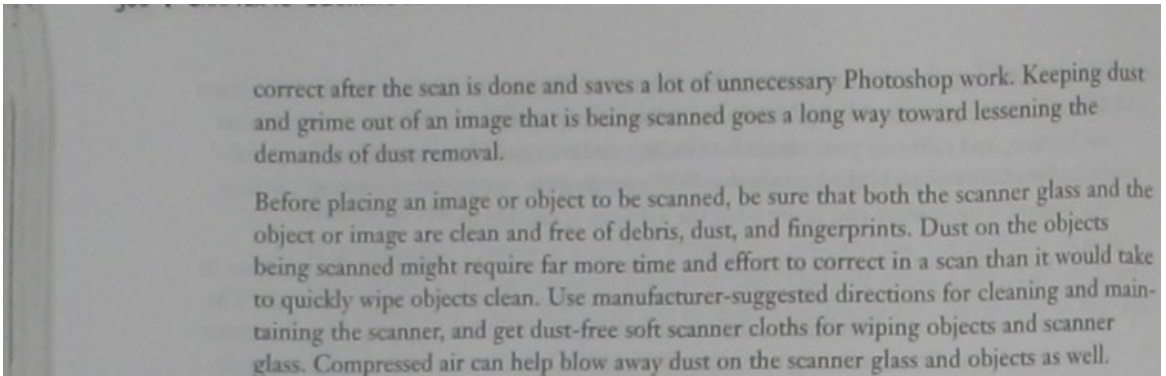
Figure 3.9 Results of the OCR engine on the original warped image



correct after the scan is done and saves a lot of unnecessary Photoshop work. Keeping dust and grime out of an image that is being scanned goes a long way toward lessening the demands of dust removal.

Before placing an image or object to be scanned, be sure that both the scanner glass and the object or image are clean and free of debris, dust, and fingerprints. Dust on the objects being scanned might require far more time and effort to correct in a scan than it would take to quickly wipe objects clean. Use manufacturer-suggested directions for cleaning and maintaining the scanner, and get dust-free soft scanner cloths for wiping objects and scanner glass. Compressed air can help blow away dust on the scanner glass and objects as well.

Figure 3.10 Results of the OCR engine on the raw corrected image



correct after the scan is done and saves a lot of unnecessary Photoshop work. Keeping dust and grime out of an image that is being scanned goes a long way toward lessening the demands of dust removal.

Before placing an image or object to be scanned, be sure that both the scanner glass and the object or image are clean and free of debris, dust, and fingerprints. Dust on the objects being scanned might require far more time and effort to correct in a scan than it would take to quickly wipe objects clean. Use manufacturer-suggested directions for cleaning and maintaining the scanner, and get dust-free soft scanner cloths for wiping objects and scanner glass. Compressed air can help blow away dust on the scanner glass and objects as well.

Figure 3.11 Results of the OCR engine on the polynomial corrected image

Therefore, to quantify the results, Table 3.1 summarizes the improvements obtained from the OCR engine from these different image captures, which for this particular example contained 129 words.

Table 3.1 Effect of flatten excerpts

	Warped image	Flattened image using raw height map	Flattened image using polynomial fitted height map
# Word errors	19	5	7
% Correct	85.27	96.12	94.57

### 3.4.2 Effects of specular lighting on the flattening process

Undesired specular effects occur due to factors such as poor lighting conditions and glossiness in the pages of the book. Such effects have a negative impact on character legibility and on depth map extraction. Despite these effects however, the ARGOS3D - P100 was able to retrieve an adequate depth profile. This particular deficiency of the ARGOS3D - P100 camera was expected because of its use of active IR illumination. The combined low-resolution (47x62) of the raw height profile of the book spread can be seen below.



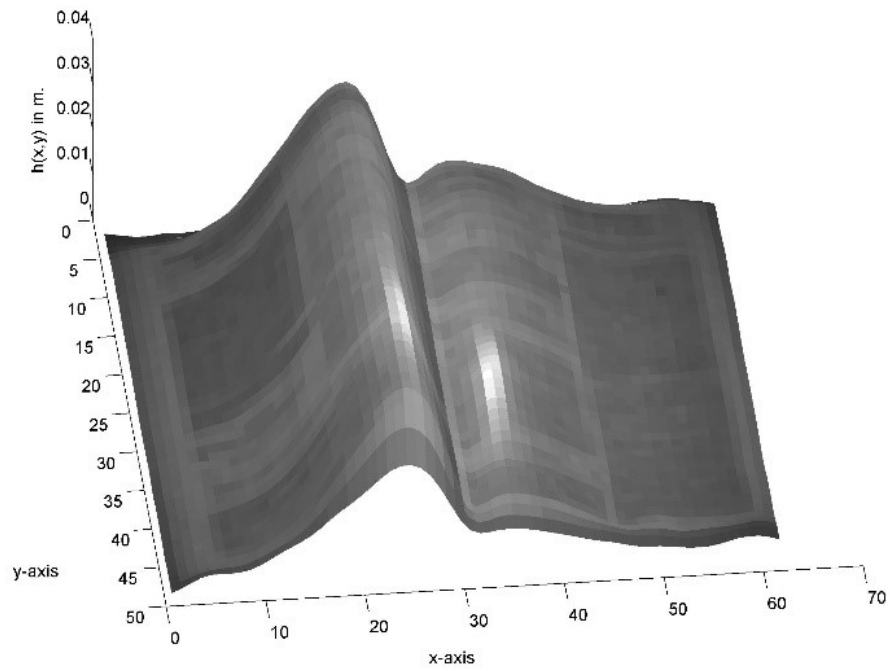


Figure 3.12 Low-resolution raw height map with corresponding low-resolution pixel values

The specular effects due to the glossiness of the pages as illustrated in Figure III-12 had the most impact on the sensor. A comparison of the polynomial fitting versus the raw height values in this case is given in Figure 3.13, where such effects impacted the height map in the range 20 to 40 in the x axis.

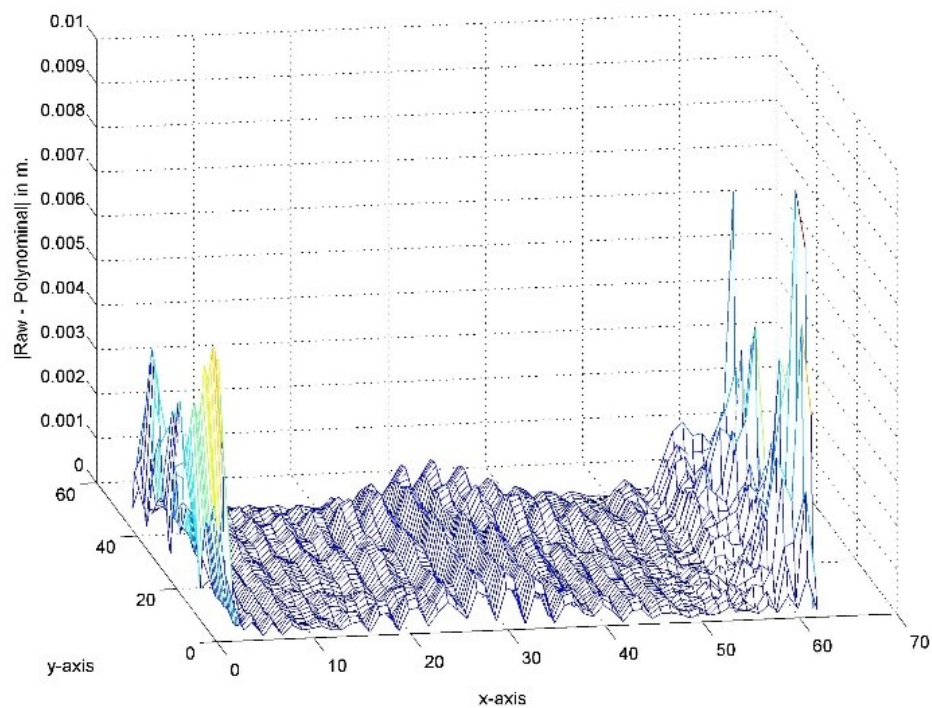


Figure 3.13 Comparison of low-resolution raw height map and polynomial fitting by subtracting both values

As was previously observed in Figure 3.5, discrepancies between the raw height and polynomial fitting appear to still be on the edges and near the spine of the book spread. The combination of the polynomial fitting on the low-resolution image can then be gauged from Figure 3.14, where the discrepancy of the edges is more pronounced. However, the target for correction is the higher-resolution (2113x1603) image of the book spread. Figure 3.15 shows the high-resolution image. Note the lighting is not adequate and some of the text is affected because of it. The next step was to interpolate the low-resolution height map and applying it to the higher-resolution image. The combination of the high-resolution image with the raw height map can be seen in Figure 3.16, while the counterpart

by applying polynomial fitting can be seen in Figure 3.17. Lastly, the flattened book spreads are shown in Figure 3.18.

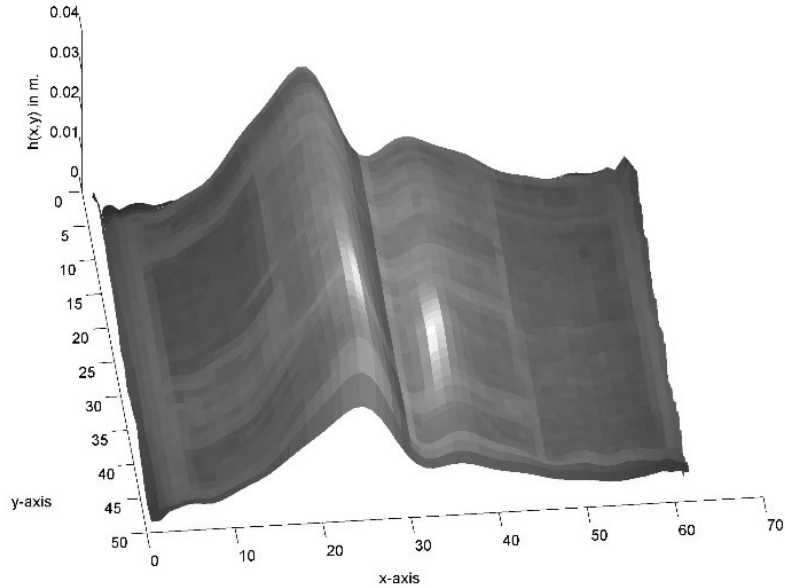


Figure 3.14 Low-resolution height polynomial fitting with low-resolution image

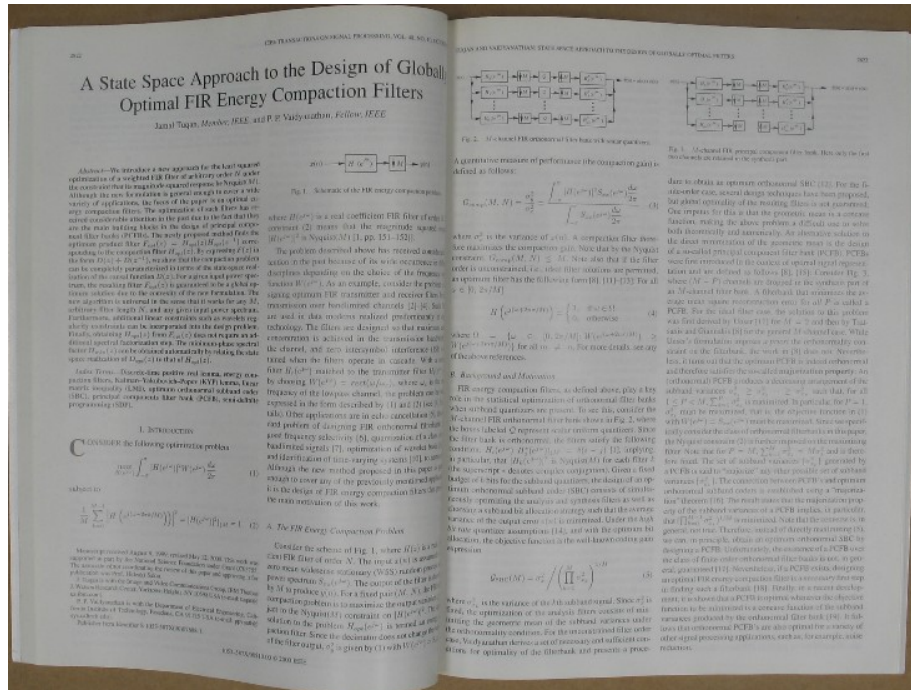


Figure 3.15 Higher-resolution image (warped) with inadequate lighting

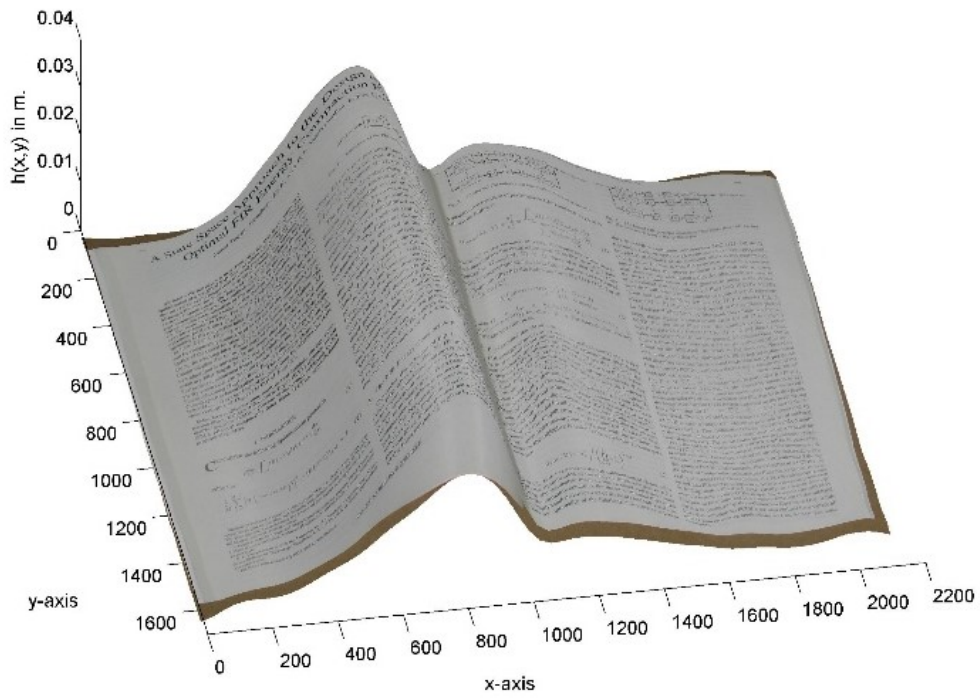


Figure 3.16 Higher-resolution raw height map with high-resolution image

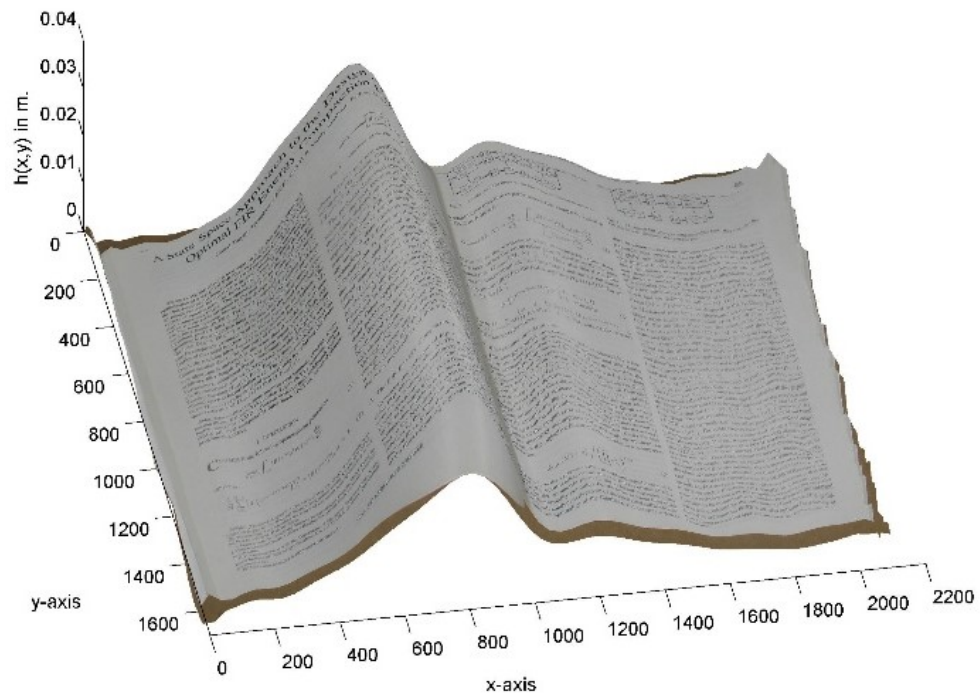


Figure 3.17 Higher-resolution height polynomial fitting with high-resolution image

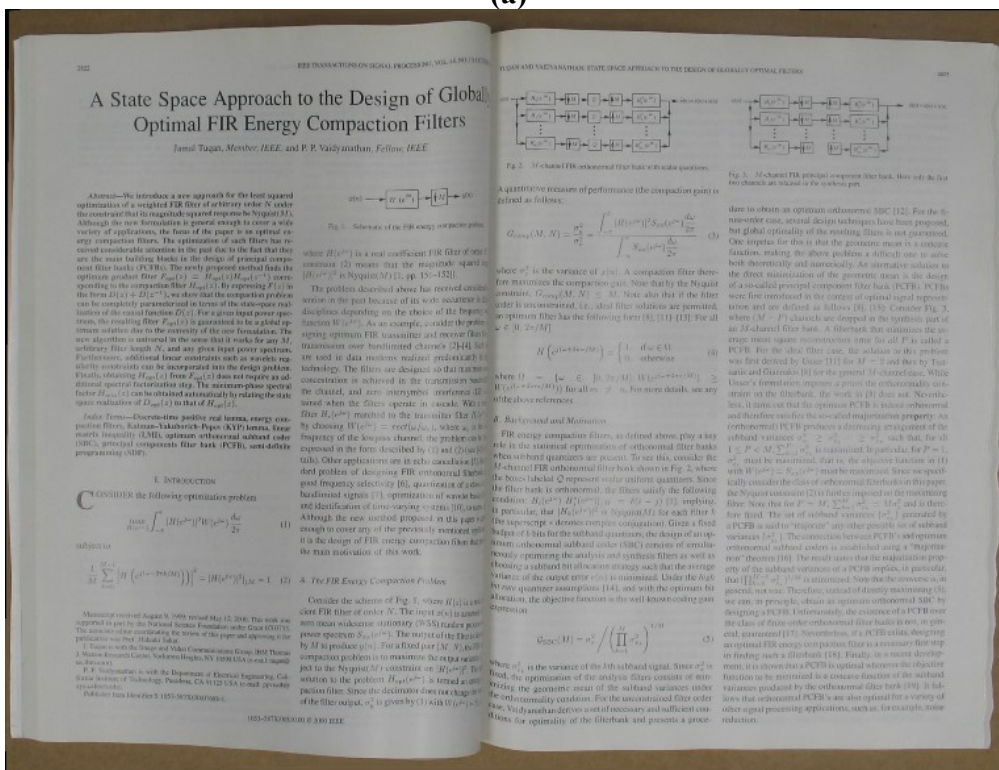
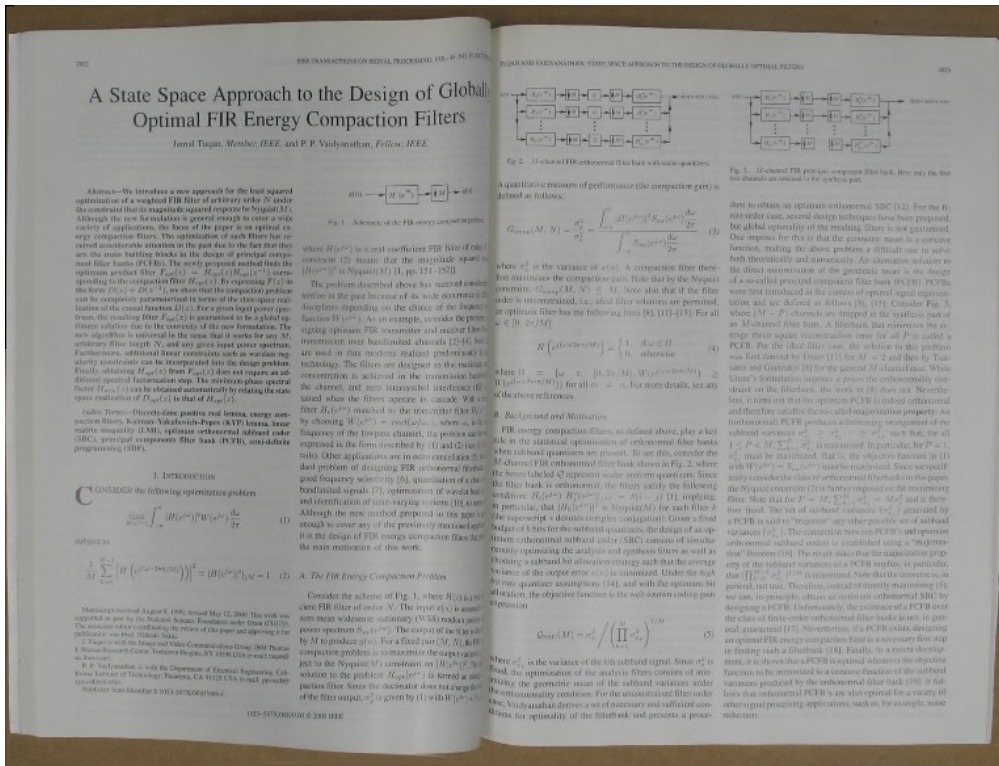
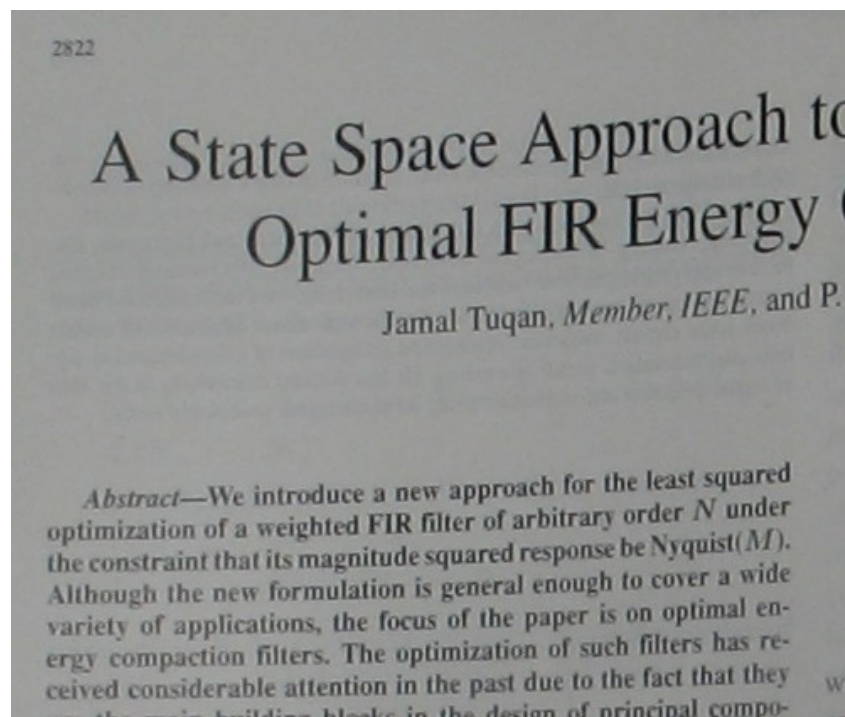


Figure 3.18 Images under poor lighting corrected with (a) raw height and (b) polynomial fitting

To gauge the effect of specular reflection on the book spread reading accuracy, a small section of the text was introduced into the OCR engine in which the visual results of the warped and its two corrected images through raw height maps and polynomial fitting are presented in Figure 3.19, Figure 3.20 and Figure 3.21, respectively. Table 2.2 summarizes the improvements obtained from the OCR engine from these different images, which in this example contained a total of 89 words.



A State Space Approach to  
 Optimal FIR Energy  
 Jamal Tuqan, Member, IEEE, and P.  
 Abstract—We introduce a new approach for the least squared optimization of a weighted FIR filter of arbitrary order  $N$  under the constraint that its magnitude squared response be Nyquist ( $M$ ). Although the new formulation is general enough to cover a wide variety of applications, the focus of the paper is on optimal energy compaction filters. The optimization of such filters has received considerable attention in the past due to the fact that they

Figure 3.19 Results of the OCR engine on the original warped image

# A State Space Approach to Optimal FIR Energy

Jamal Tuqan, *Member, IEEE*, and

*Abstract*—We introduce a new approach for the least squared optimization of a weighted FIR filter of arbitrary order  $N$  under the constraint that its magnitude squared response be Nyquist ( $M$ ). Although the new formulation is general enough to cover a wide variety of applications, the focus of the paper is on optimal energy compaction filters. The optimization of such filters has received considerable attention in the past due to the fact that they

A State Space Approach to

Optimal FIR Energy

Jamal Tuqan, Member, IEEE, and

*Abstract*—We introduce a new approach for the least squared optimization of a weighted FIR filter of arbitrary order  $N$  under the constraint that its magnitude squared response be Nyquist ( $M$ ). Although the new formulation is general enough to cover a wide variety of applications, the focus of the paper is on optimal energy compaction filters. The optimization of such filters has received considerable attention in the past due to the fact that they

Figure 3.20 Results of the OCR engine on the corrected image using the raw height map

# A State Space Approach Optimal FIR Energy

Jamal Tuqan, *Member, IEEE*, and

*Abstract*—We introduce a new approach for the least squared optimization of a weighted FIR filter of arbitrary order  $N$  under the constraint that its magnitude squared response be Nyquist ( $M$ ). Although the new formulation is general enough to cover a wide variety of applications, the focus of the paper is on optimal energy compaction filters. The optimization of such filters has received considerable attention in the past due to the fact that they

A State Space Approach  
Optimal FIR Energy  
Jamal Tuqan, Member, IEEE, and  
We introduce a new approach for the least squared  
optimization of a weighted FIR filter of arbitrary order  $N$  under  
the constraint that its magnitude squared response be Nyquist ( $M$ ).  
Although the new formulation is general enough to cover a wide  
variety of applications, the focus of the paper is on optimal en-  
ergy compaction filters. The optimization of such filters has re-  
ceived considerable attention in the past due to the fact that they

Figure 3.21 Results of the OCR engine on the corrected image using polynomial fitting

Table 3.2 Result of OCR engine from flatten excerpt

	Warped image	Flattened image using raw height map	Flattened image using polynomial fitted height map
# Word errors	23	20	19
% Correct	74.16	77.53	78.65



### 3.4.3 Full correction: flattened and extended book spread

Earlier work by a research group with the Center for Advanced Technology and Education (CATE) has made use of the ABBYY engine. The approach relied on using two cameras: a lateral camera used to extract a single representative curvature of the book assumed to be uniform throughout the page spread and a top camera for a high-resolution capture of the text [18]. The following is an excerpt of their findings using the older OCR engine of ABBYY FineReader 8.0.

Table 3.3 Excerpt of results for old approach using ABBYY FineReader 8.0

Book spread	Word count	Errors original image	Reading accuracy	Errors corrected image	Reading accuracy
1	464	267	42.46%	8	98.28%
2	383	85	77.81%	15	96.08%
3	944	400	57.63%	15	98.41%
4	963	300	68.85%	24	97.51%
5	496	192	61.29%	9	98.19%
6	456	180	60.53%	21	95.39%
7	513	243	52.63%	17	96.69%
8	898	560	37.64%	29	96.77%
9	567	320	43.56%	11	98.06%

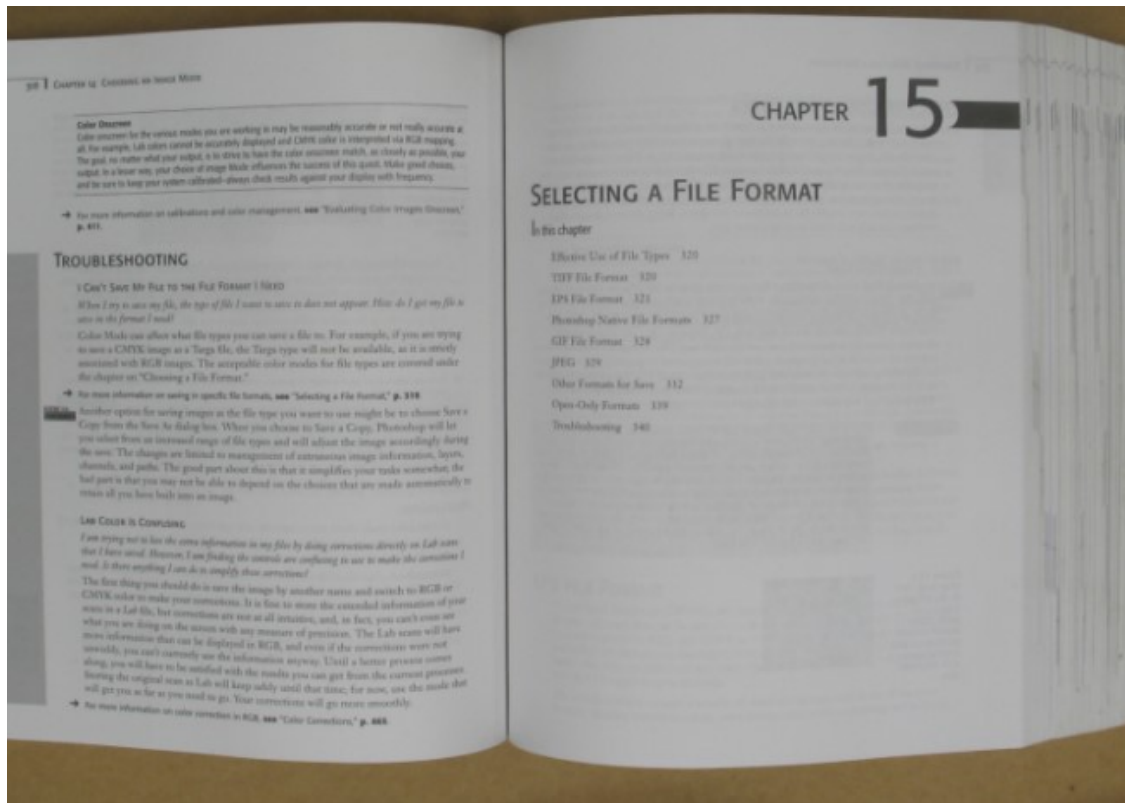
Since then, it is important to note that the ABBYY OCR engine appears to have been greatly improved, and it seems to include some form of geometrical distortion correction, which we assumed to be part of the newer version ABBYY FineReader 12.0 as can be observed by comparing Table 3.3 with Table 3.4. Furthermore, the approach taken to obtain the result in Table 3.3 relied on having 2 cameras set at different locations but were well aligned.

Excerpts of the tests showing improvements in reading accuracy can be seen in the tables that follow, which were obtained by relying only on the raw heights, as previously seen in Table 3.1 and Table 3.2, provided better results. The newer ABBYY version performed rather well as can be seen in Table 3.4, however the proposed method, which corrected the page curvature using the heights as extracted by the ARGOS3D - P100 camera performed better in majority of the cases.

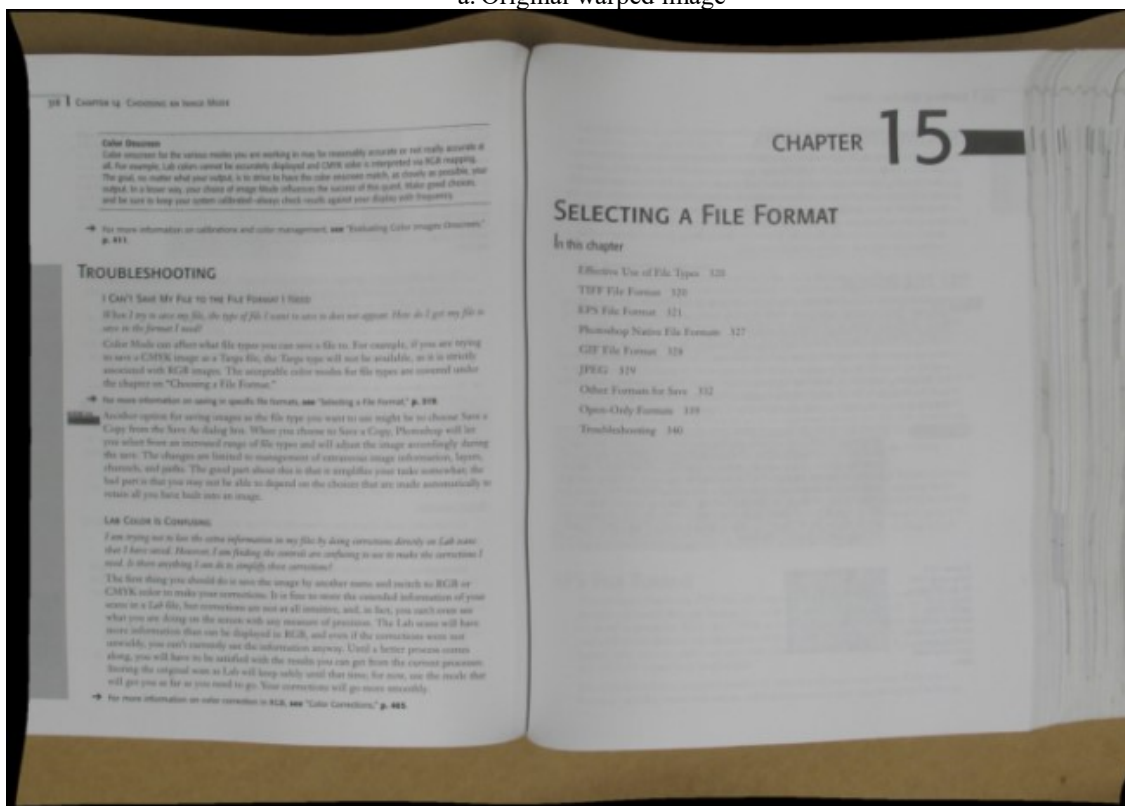
Table 3.4 Results with just flattening using ABBYY FineReader 12.0

Book spread	Word count	Errors original image	Reading accuracy	Errors corrected image	Reading accuracy
1	368	57	84.51%	19	94.84%
2	601	48	92.01%	27	95.51%
3	480	70	85.42%	30	93.75%
4	614	76	87.62%	16	97.39%
5	286	15	94.76%	4	98.60%
6	610	83	86.39%	10	98.36%
7	305	3	99.02%	4	98.69%
8	472	20	95.76%	11	97.67%
9	506	32	93.68%	7	98.62%

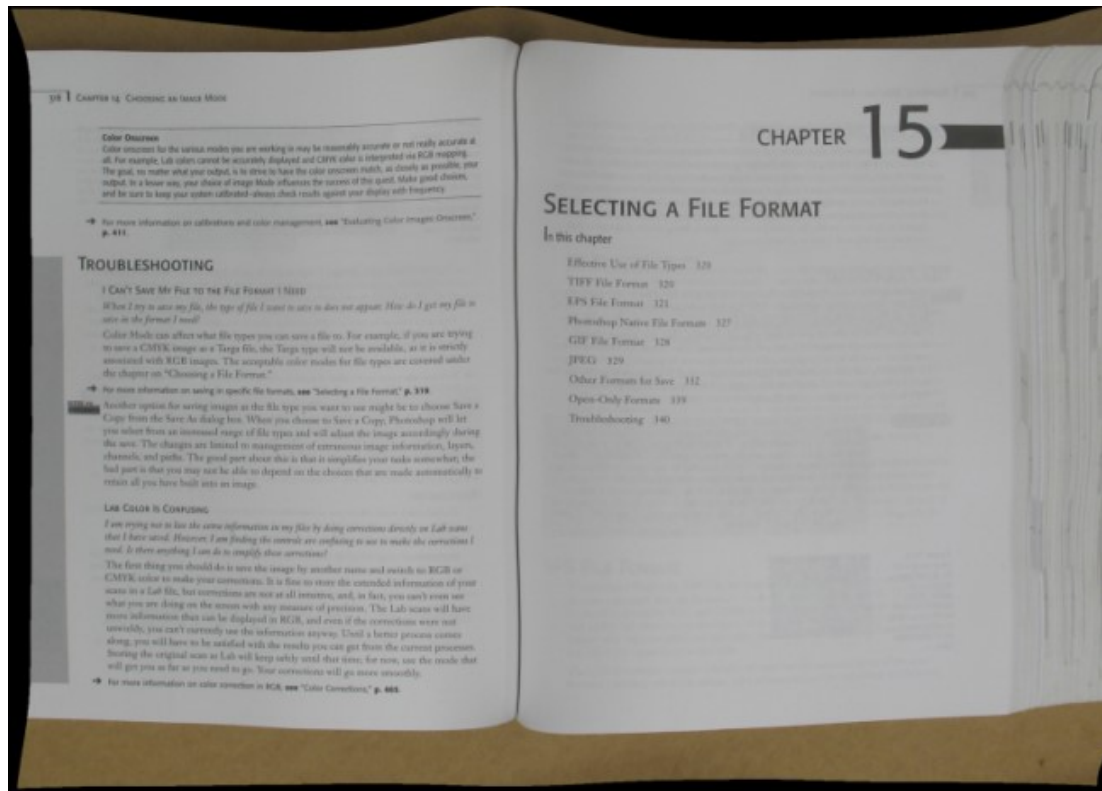
Figure 3.22 shows examples of the original and corrected versions via flattening with extension. The OCR results for these can be seen in Book spread 3 of Table 3.5, Table 3.6, and Table 3.7, respectively. To the naked eye, all 3 types of corrections of flattening with extension (linear, nearest neighbor, and natural) as shown in previous figures, look the same, but, as we can see in the following tables, they do have different impacts on the OCR performance.



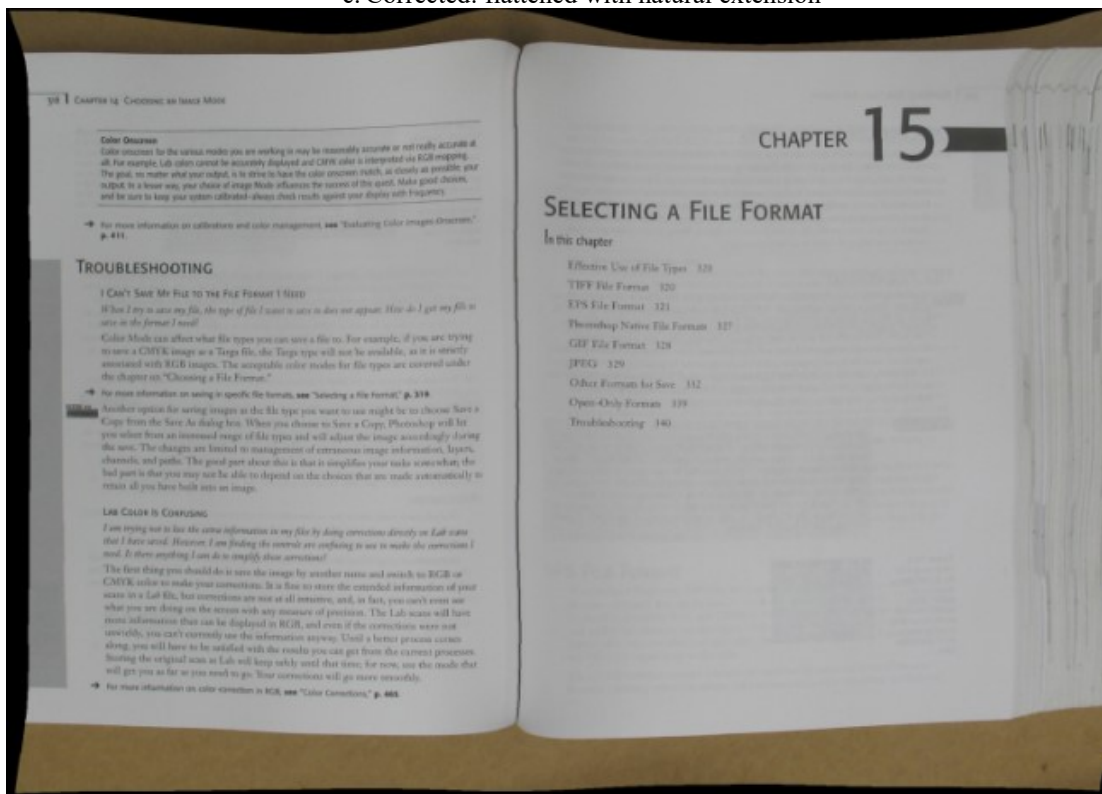
a. Original warped image



b. Corrected: flattened with linear extension



c. Corrected: flattened with natural extension



d. Corrected: flattened with nearest neighbor extension

Figure 3.22 Example of the original and fully corrected images

From the results shown in Tables 3.5, 3.6 and 3.7, it can be observed that the type of image correction process used does affect the results. Both nearest neighbor interpolation and natural neighbor interpolation appear to have the best results on these types of correction processes.

Table 3.5 Results of flattening with linear extension approach using ABBYY FineReader 12.0

Book spread	Word count	Errors original image	Reading accuracy	Errors corrected image	Reading accuracy
1	368	57	84.51%	22	94.02%
2	601	48	92.01%	42	93.01%
3	480	70	85.42%	32	93.33%
4	614	76	87.62%	24	96.09%
5	286	15	94.76%	4	98.60%
6	610	83	86.39%	10	98.36%
7	305	3	99.02%	2	99.34%
8	472	20	95.76%	12	97.46%
9	506	32	93.68%	9	98.22%

Table 3.6 Results of flattening with natural neighbor extension approach using ABBYY FineReader 12.0

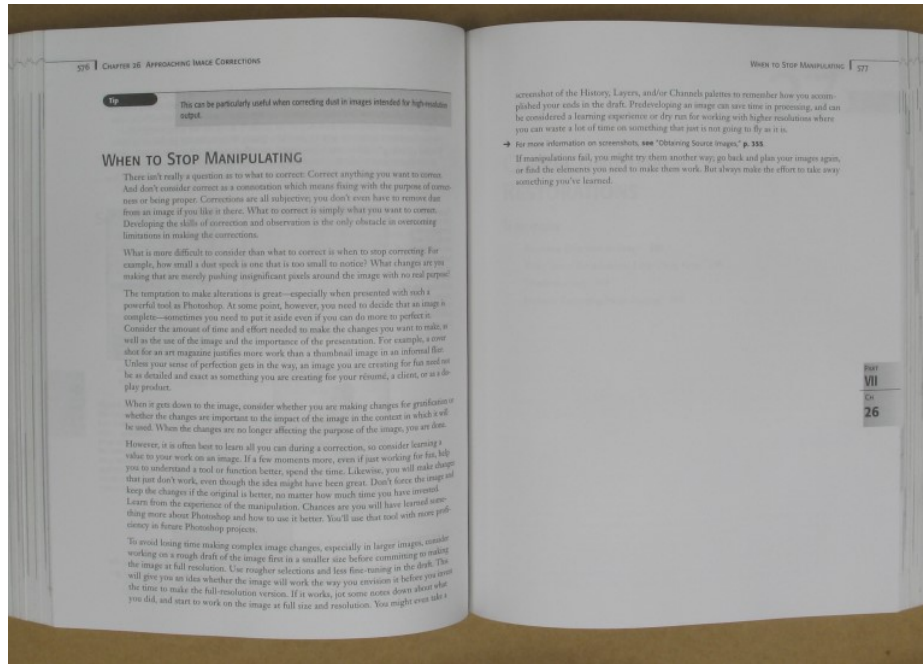
Book spread	Word count	Errors original image	Reading accuracy	Errors corrected image	Reading accuracy
1	368	57	84.51%	16	95.65%
2	601	48	92.01%	33	94.51%
3	480	70	85.42%	39	91.88%
4	614	76	87.62%	12	98.05%
5	286	15	94.76%	2	99.30%
6	610	83	86.39%	9	98.52%
7	305	3	99.02%	4	98.69%
8	472	20	95.76%	14	97.03%
9	506	32	93.68%	7	98.62%

Table 3.7 Results of Flattening with nearest neighbor extension approach using ABBYY FineReader 12.0

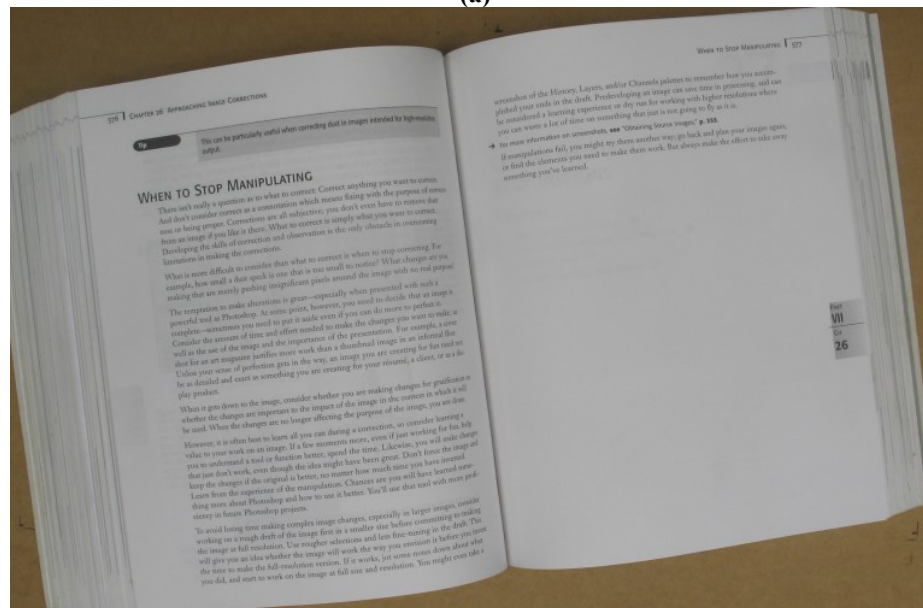
Book spread	Word count	Errors original image	Reading accuracy	Errors corrected image	Reading accuracy
1	368	57	84.51%	16	95.65%
2	601	48	92.01%	29	95.17%
3	480	70	85.42%	23	95.21%
4	614	76	87.62%	14	97.72%
5	286	15	94.76%	1	99.65%
6	610	83	86.39%	10	98.36%
7	305	3	99.02%	8	97.38%
8	472	20	95.76%	7	98.52%
9	506	32	93.68%	7	98.62%

#### 3.4.4 Resilience of the full correction on a rotated book spread

The resilience of the proposed curvature correction approach was further explored by testing the effect of having the book misaligned on the book reader platform as portrayed in Figure III-23. The following example shows that having good alignment helps in character recognition; however, this also proves the resilience of the proposed approach to such misalignments as evidenced in the results provided in Table 3.8.



(a)



(b)

Figure 3.23 Example of the original (a) and rotated (b) book spread 6

Book spread 6 as in Table 3.5 was purposely misaligned by physically rotating the book left and right by about 30 degrees. The original and corrected images were then introduced to the OCR engine yielding the results shown in Table 3.8. It is clear that such

misalignments do significantly affect the reading accuracy if they are not taken into consideration. Moreover, the results obtained showing a high reading accuracy, confirming that the proposed method is resilient to such misalignments.

Table 3.8 Results of Flattening with linear extension approach using ABBYY FineReader 12.0 for rotated book spread 6

Book spread	Word count	Errors original image	Reading accuracy	Errors corrected image	Reading accuracy
Rotated left	610	221	63.77%	20	96.72%
Rotated right	610	220	63.93%	21	96.56%



## 4. UNIFORM VS. FULL HEIGHT MAP

### 4.1 Exploring the effect of the height map

At this juncture, the approach taken to correct the warping effect on the book spread and improve the OCR results has been successful. This can be attributed in part to having an accurate height map. However, having part of the height map seems to be just as good as using the full height map after all, the book spread appears to be uniform throughout. This chapter will particularly focus on comparing these two different height maps, namely 1) a full height map as obtained from a full scan of the book's surface by the ToF sensor, and 2) a uniform height map that replicates a singular height profile from or near the center line of the page throughout the entire page. To perform the necessary tests for comparison, it was necessary to create a standardized database composed of a match low-resolution height map and pixel intensity with its corresponding high-resolution book spread data. The data set should contain 142 pages at an average of about 3245 characters per page.

Once the images were corrected, they were fed into the latest version of the OCR engine ABBYY FineReader v12.0. The resulting digital files were then compared on a character by character basis with the actual digital version of the book's chapter. The character by character comparison was performed by an implementation of the Myers' Difference Algorithm which finds the longest common chain of matching characters in two sequences (control and test sample). Utilizing this algorithm allowed us to both facilitate and expedite the comparison process, and was especially useful with the many pages that were used.

## 4.2 Standard placement

The following results demonstrate the corrections made possible by the proposed algorithm using both uniform and full height maps on de-warping the pages of bound documents. The character recognition accuracy of the uncorrected and corrected book spreads for all the 71 book spreads. The standard placement is achieved by positioning the book spread at the center of the book reader base and placed parallel to the horizontal axis which yields the image seen in Figure 4.1.

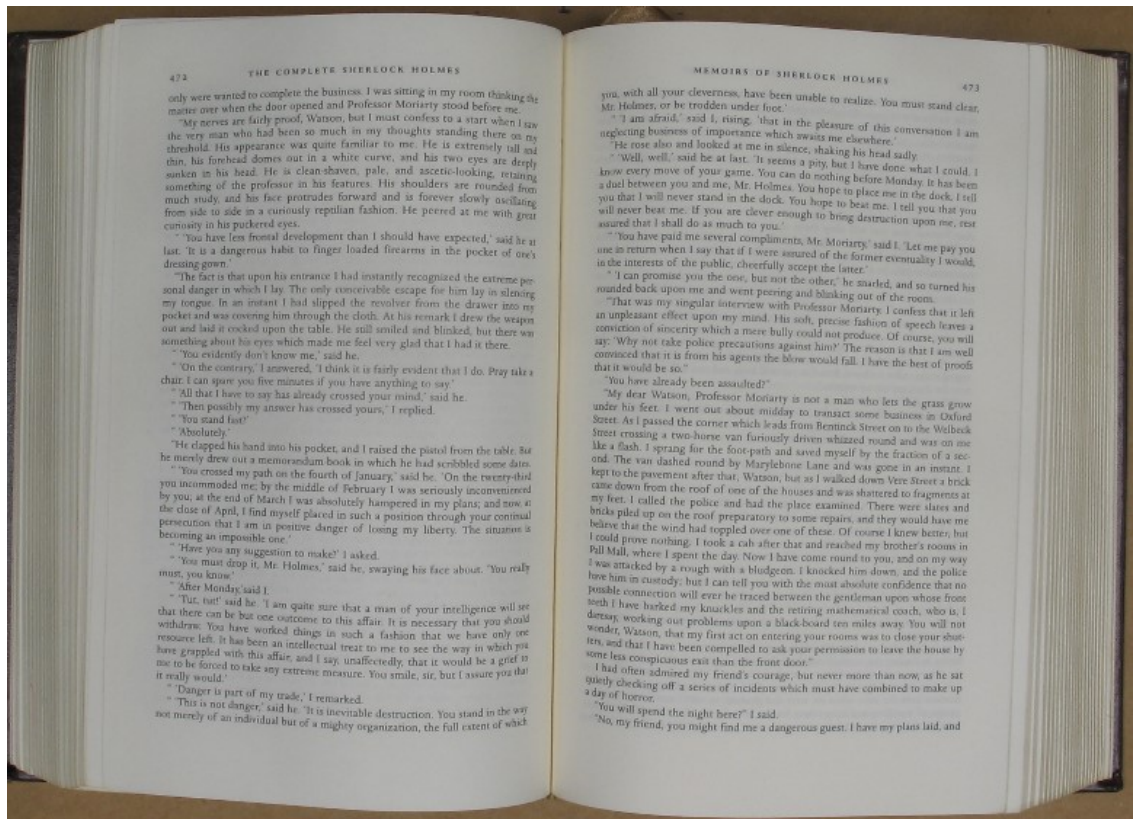


Figure 4.1 Standard align book spread

If we were to focus on a section of the book spread, it is possible to see the warping effect that occurs as demonstrated in Figure 4.2(a). This arching effect occurs in the majority of

the tested book spreads. Figure 4.2(b) shows the result of the usual correction of the warped text using the uniform height map, while Figure 4.2(c) illustrates the correction with the full height map. The overall accuracy results are given in Figure 4.3. illustrates the correction with the full height map. It is worth noting that the results of the proposed correction algorithm using either map were better than the original, although the newly improved OCR engine of the ABBYY FineReader v12.0. performed quite well even on the curved pages. Table 4.1 shows a summary of the results for all the methods tested. It is evident that both curvature correction procedures led to a much-enhanced reading accuracy. Furthermore, it can be observed that using the full height map yielded better results than when using the uniform one. This outcome was anticipated as the full height map better reflects the true nature of the page curvature, which obviously could not be replicated by simply assuming a uniformly distributed curvature using either a single row of text or when it is obtained from a lateral view of the book as in [18]. Correction errors most certainly occur due the complex nature of book curvatures as we leaf through the book. A smarter expansion process (i.e. recognizing key points as reference from which to start the expansion) could improve these results. Nonetheless, we still have to contend with the imperfections of the ToF technology in extracting depth information and the nonlinear nature of page curvatures.

daresay, working out problems upon  
wonder, Watson, that my first act on  
ters, and that I have been compelled  
some less conspicuous exit than the  
I had often admired my friend's  
quietly checking off a series of incidents  
a day of horror.  
"You will spend the night here?" I  
"No, my friend, you might find me

(a)

teeth I have barked my knuckles and  
daresay, working out problems upon  
wonder, Watson, that my first act on  
ters, and that I have been compelled to  
some less conspicuous exit than the friend  
I had often admired my friend's composure  
quietly checking off a series of incidents  
a day of horror.  
"You will spend the night here?" I said  
"No, my friend, you might find me a

(b)

teeth I have barked my knuckles and  
daresay, working out problems upon  
wonder, Watson, that my first act on  
ters, and that I have been compelled to  
some less conspicuous exit than the friend  
I had often admired my friend's composure  
quietly checking off a series of incidents  
a day of horror.  
"You will spend the night here?" I said  
"No, my friend, you might find me a

(c)

Figure 4.2 Examples of text: (a) original warped, (b) corrected with uniform map, and (c) corrected with full map

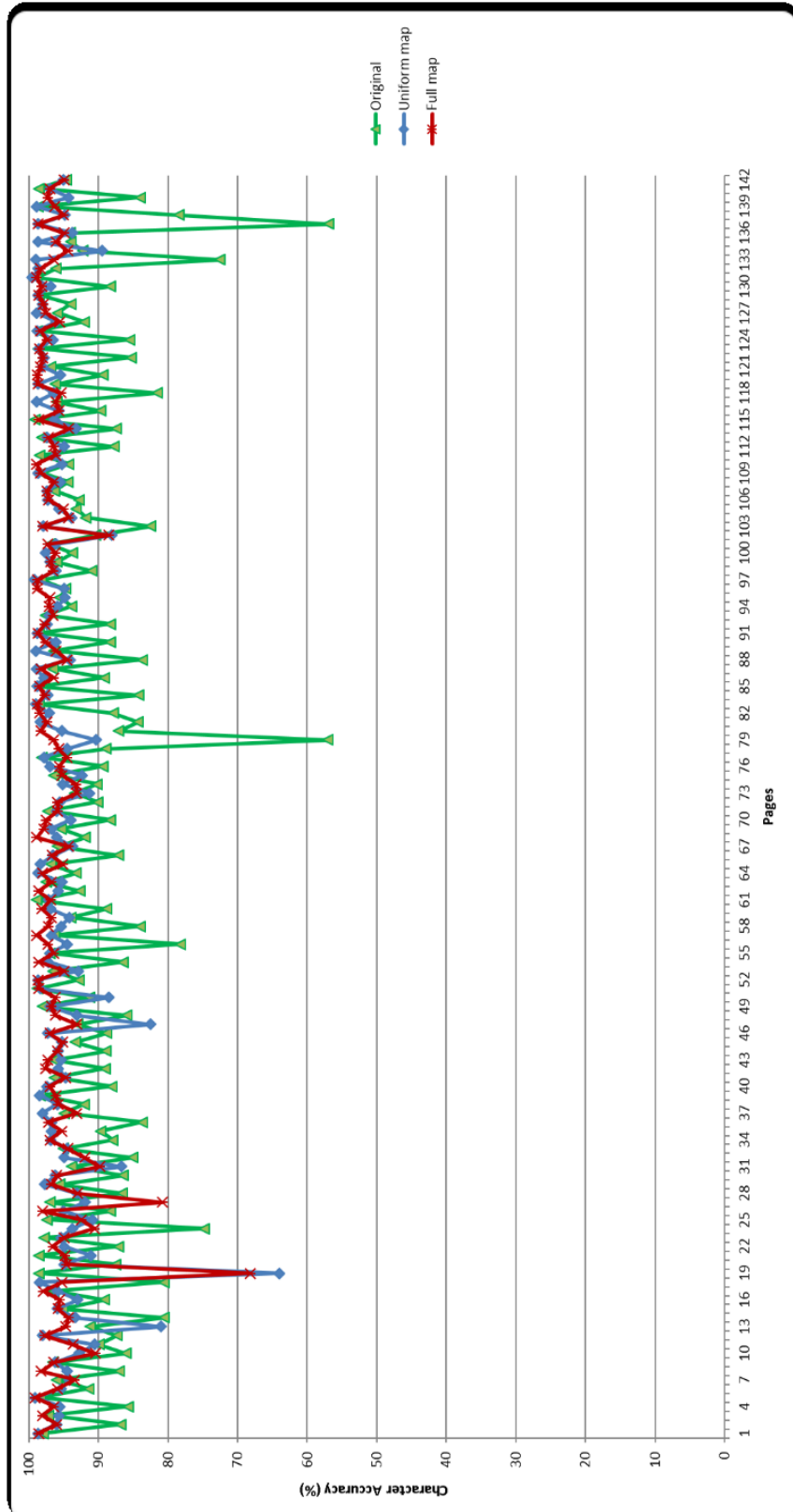
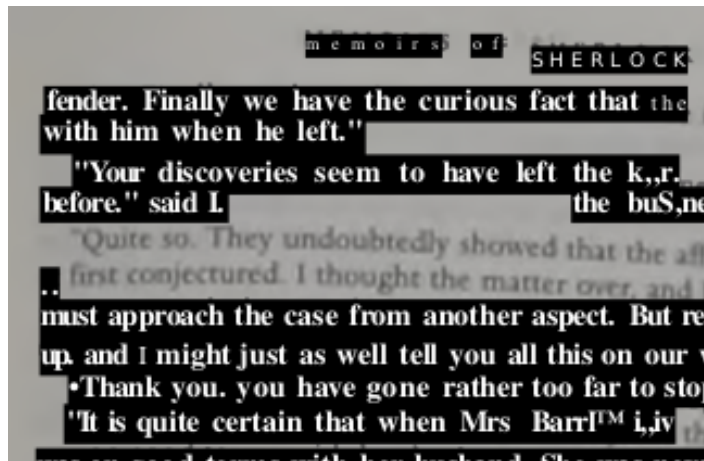


Figure 4.3 Accuracy of the original and corrected book spreads

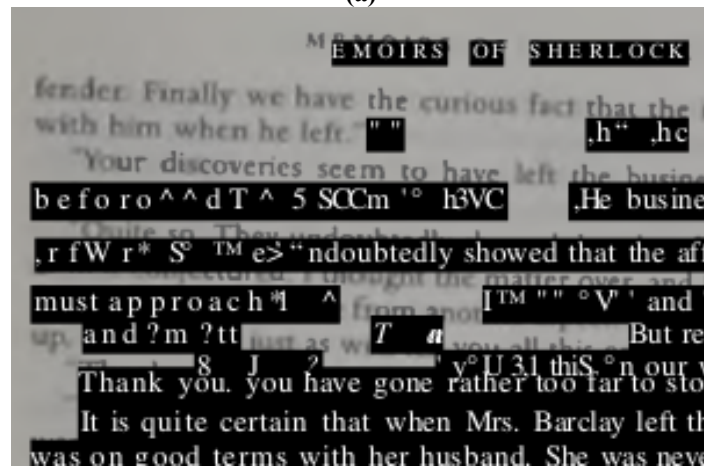
Table 4.1 Summary of the standard positioned book spreads

Procedure	Number of samples	Minimum accuracy	Maximum accuracy	Mean accuracy	Variance	Standard deviation	Median accuracy (after sort)
Original – warped	142	56.99%	99.18%	91.57%	48.17	6.94	93.25%
Uniform map	142	63.97%	99.55%	95.55%	15.74	3.97	96.10%
Full map	142	68.15%	99.12%	96.11%	11.15	3.34	96.56%

An example of the results of the OCR with bad accuracy can be observed in Figure 4.4. (a) In these cases, some portions of text are not recognized at all, and the parts that were readable contained misspellings and/or misrecognized characters. The complete correction with full height map shows much clear character as can be seen in Figure 4.4 (c).



(a)



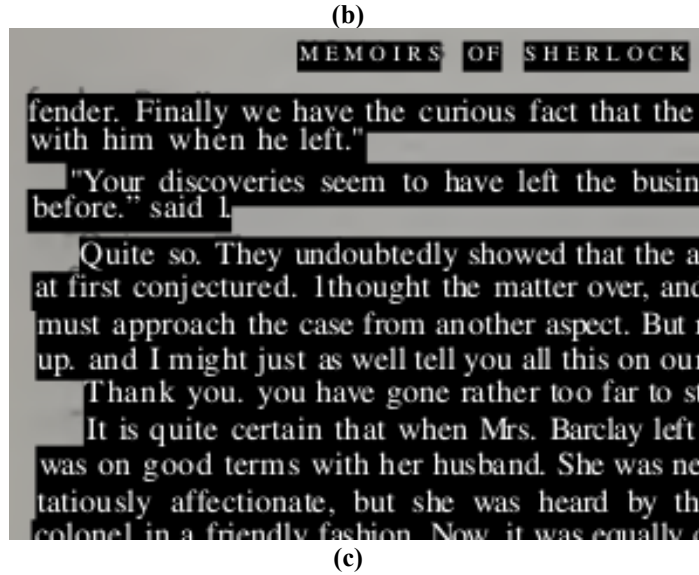


Figure 4.4 Page 79 for standard placement: a) original, b) uniform, c) full

Due to the warping present in the book's page image, much of the characters were not recognized or were improperly recognized as can be appreciated in Figure 4.5(a), with sections highlighted in white as being the text that was recognized. After performing the correction with the uniform map, many more characters are recognized as shown in Figure 4.5(b). Figure 4.6(c) shows how when the correction is made using the full height map, more characters are recognized in contrast to the results shown in Figure 4.5(b).

fingers were trembling as I [redacted] the [redacted] of her mantle [redacted] I never remember having done such a thing in my life before. The fact is that I felt as though I were choking and had a perfect longing for a breath of fresh air. I really think that I should have fainted if I had not gone out. I stood at the door for a few minutes, and now I am quite myself again.

All the time that she was telling me this story she never once looked in my direction, and her voice was quite unlike her usual tones. It was evident to me that she was saying what was false. I said nothing in reply, but turned my face to [redacted] sick at heart, with my mind filled with a thousand venomous doubts and suspicions. What was it that my wife was concealing from me? Where had she been during that strange expedition? I felt that I should have no peace until I knew, and yet I shrank from asking her again after once she had told me what was false. All the rest of the night I tossed and tumbled, framing theory after theory, each more unlikely than the last.

I should have gone to the City that day, but I was too disturbed in my mind to be able to pay attention to business matters. My wife seemed to be as upset as myself, and I could see from the little questioning glances which she kept shooting at me that she understood that I disbelieved her statement, and that she was at her wit's end what to do. We hardly exchanged a word during breakfast, and immediately afterwards I went out for a walk that I might think the matter out in the fresh morning air.

I went as far as the Crystal Palace, spent an hour in the grounds, and was back in Norbury by one o'clock. It happened that my way took me past the cottage, and I stopped for an instant to look at the windows and to see if I could catch a glimpse of the strange face which had looked out at me on the day before. As I stood there, imagine my surprise, Mr. Holmes, when the door suddenly opened and my wife walked out.

I was struck dumb with astonishment at the sight of her, but my emotions were nothing to those which showed themselves upon her face when our eyes met. She seemed for an instant to wish to shrink back inside the house again; and then, seeing how useless all concealment must be, she came forward, with a very white face and frightened eyes which belied the smile upon her lips.

Ah, Jack, she said, I have just been in to see if I can be of any assistance to our new neighbours. Why do you look at me like that, Jack? You are not angry with me?

"No," said I, "where did you go during the night?"

"What do you mean?" she cried.

"You came here. I am sure of it. Who are these people that you should visit them at such an hour?"

"I have not been here before."

"How can you tell me what you know is false?" I cried. "Your very voice changes as you speak. When have I ever had a secret from you? I shall enter that cottage, and I shall probe the matter to the bottom."

"No, no, Jack, for God's sake!" she gasped in uncontrollable emotion. Then, as I approached the door, she seized my sleeve and pulled me back with convulsive strength.

"I implore you not to do this, Jack," she cried. "I swear that I will tell you everything some day, but nothing but misery can come of it if you enter that cottage." Then, as I tried to shake her off, she clung to me in a frenzy of entreaty.

(a)



fingers were trembling as she undid the fastenings of her mantle. 'Why, I never remember having done such a thing in my life before. The fact is that I felt as though I were choking and had a perfect longing for a breath of fresh air. I really think that I should have fainted if I had not gone out. I stood at the door for a few minutes, and now I am quite myself again.'

'All the time that she was telling me this story she never once looked in my direction, and her voice was quite unlike her usual tones. It was evident to me that she was saying what was false. I said nothing in reply, but turned my face to the wall, sick at heart, with my mind filled with a thousand venomous doubts and suspicions. What was it that my wife was concealing from me? Where had she been during that strange expedition? I felt that I should have no peace until I knew, and yet I shrank from asking her again after once she had told me what was false. All the rest of the night I tossed and tumbled, framing theory after theory, each more unlikely than the last.'

'I should have gone to the City that day, but I was too disturbed in my mind to be able to pay attention to business matters. My wife seemed to be as upset as myself, and I could see from the little questioning glances which she kept shooting at me that she understood that I disbelieved her statement, and that she was at her wit's end what to do. We hardly exchanged a word during breakfast, and immediately afterwards I went out for a walk that I might think the matter out in the fresh morning air.'

'I went as far as the Crystal Palace, spent an hour in the grounds, and was back in Norbury by one o'clock. It happened that my way took me past the cottage, and I stopped for an instant to look at the windows and to see if I could catch a glimpse of the strange face which had looked out at me on the day before. As I stood there, imagine my surprise, Mr. Holmes, when the door suddenly opened and my wife walked out.'

'I was struck dumb with astonishment at the sight of her, but my emotions were nothing to those which showed themselves upon her face when our eyes met. She seemed for an instant to wish to shrink back inside the house again; and then, seeing how useless all concealment must be, she came forward, with a very white face and frightened eyes which belied the smile upon her lips.'

'Ah, Jack,' she said, 'I have just been in to see if I can be of any assistance to our new neighbours. Why do you look at me like that, Jack? You are not angry with me?'

'So,' said I, 'this is where you went during the night.'

'What do you mean?' she cried.

'You came here. I am sure of it. Who are these people that you should visit them at such an hour?'

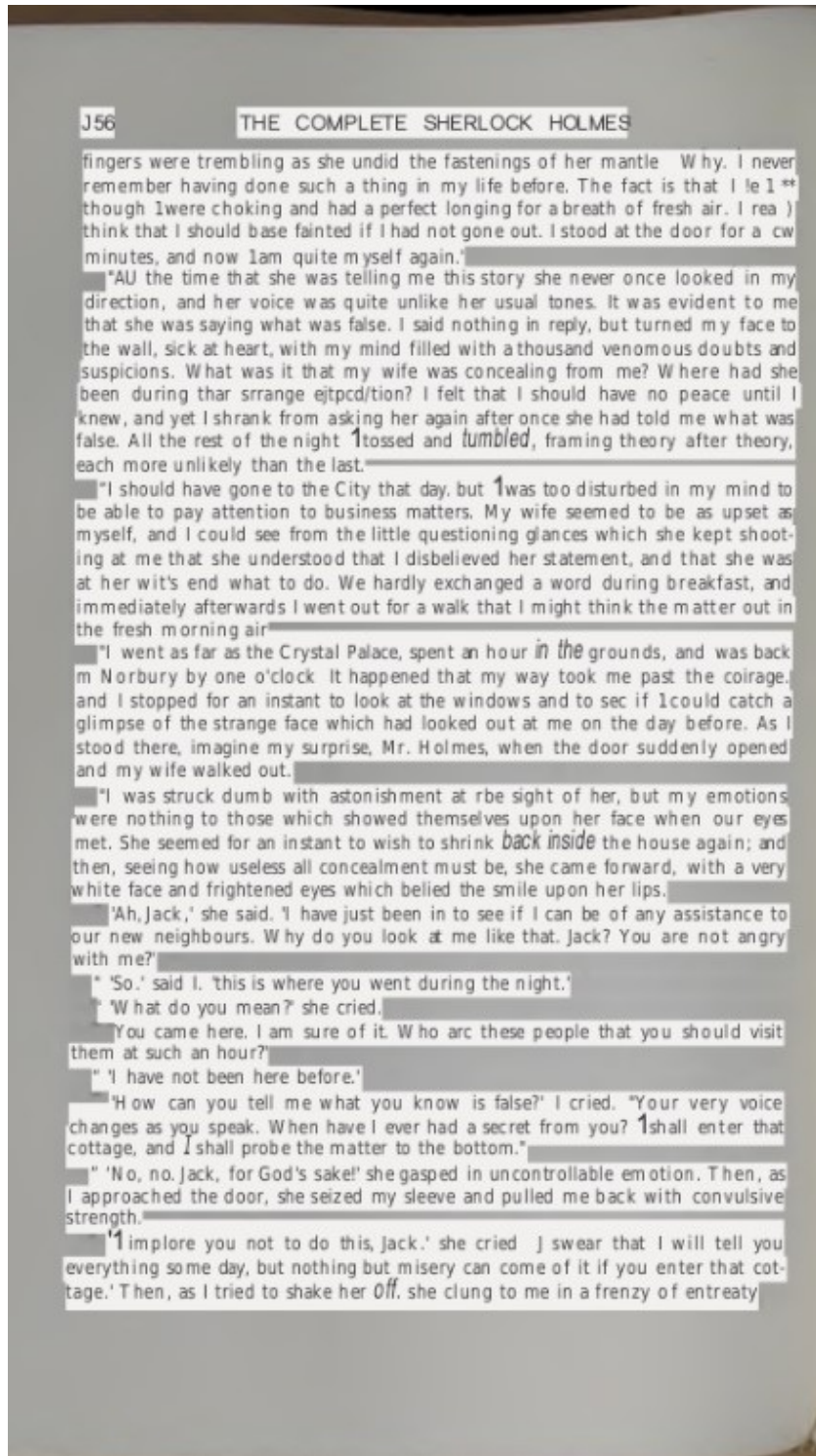
'I have not been here before.'

'How can you tell me what you know is false?' I cried. 'Your very voice changes as you speak. When have I ever had a secret from you? I shall enter that cottage, and I shall probe the matter to the bottom.'

'No, no, Jack, for God's sake!' she gasped in uncontrollable emotion. 'Then, as I approached the door, she seized my sleeve and pulled me back with convulsive strength.'

'I implore you not to do this, Jack,' she cried. 'I swear that I will tell you everything some day, but nothing but misery can come of it if you enter that cottage. Then, as I tried to shake her off, she clung to me in a frenzy of entreaty.'

(b)



(c)

Figure 4.5. Examples of standard placement results for: a) original, b) corrections with uniform map, c) corrections with full map

### 4.3 Misaligned placement

Many approaches rely on having a standard placement of the book spread due to their sensitivity to placement/alignment disturbances. Therefore, to further test the resilience of the proposed method, the book spreads were physically made to be rotated by 30 degrees. In this experiment, half of the 142 pages were rotated 30 degrees clockwise and the other half 30 degrees counter clockwise; an example can be seen in Figure 4.6.

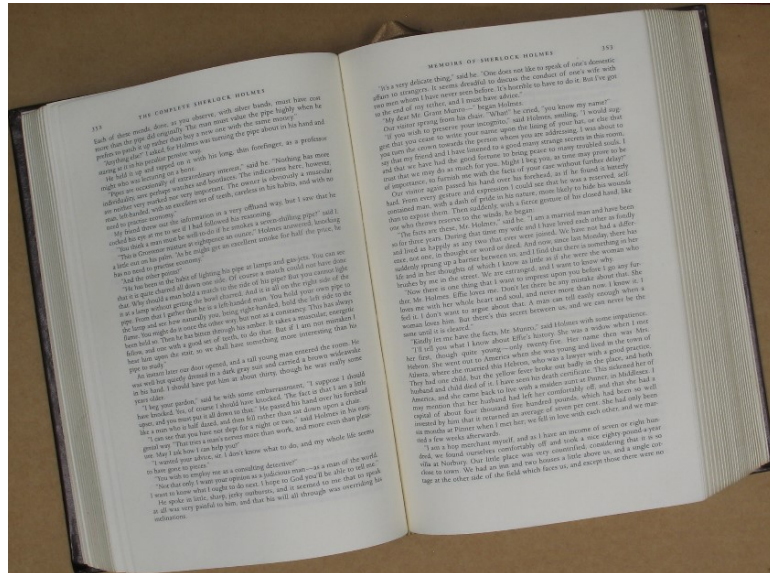
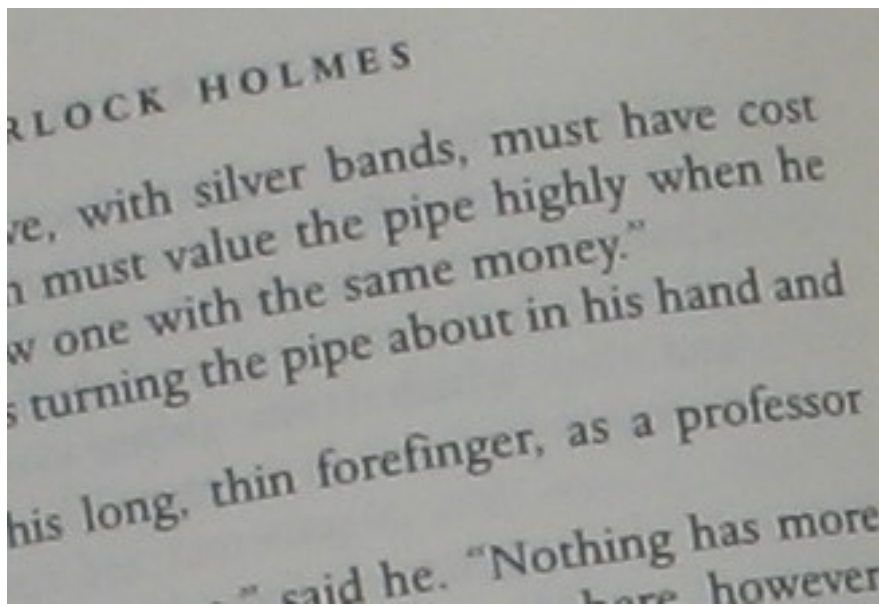


Figure 4.6 Rotated book spread

The accuracy of the results for the rotated book spread images remained quite high despite the rotation. However, the overall accuracy was still lower than the standard placement. Figure 4.7(a) illustrates a zoomed-in section of the warping effect where the text lines appear curved instead of straight. Figure 4.7(b) shows how the correction algorithm does straighten the lines so they are no longer warped, whereas Figure 4.7(c) depicts the correction with a full height map. The overall accuracy results are given in Table 4.2 shows

a statistical summary of the results of the proposed algorithm with both height maps. According to our expectations, the highest overall accuracy and lowest variance was obtained when using the full height map. On the other hand, the uniform maps yielded a less accurate performance, indicating that replicating a single row height profile throughout the page was not resilient to misalignment, represented here as a rotation. Nonetheless the variance in the reading accuracy through all the pages remained lower than for the uncorrected version.



(a)

LOCK HOLMES  
e, with silver bands, must have cost  
must value the pipe highly when he  
one with the same money."  
turning the pipe about in his hand and  
is long, thin forefinger, as a professor  
interest," said he. "Nothing has mor  
indications here, howeve  
muscul

(b)

LOCK HOLMES  
e, with silver bands, must have cost  
must value the pipe highly when he  
one with the same money."  
turning the pipe about in his hand and  
is long, thin forefinger, as a professor  
interest," said he. "Nothing has mor  
indications here, howeve  
muscul

(c)

Figure 4.7 Examples of text: (a) original warped, (b) corrected with uniform map, and (c) corrected with full map

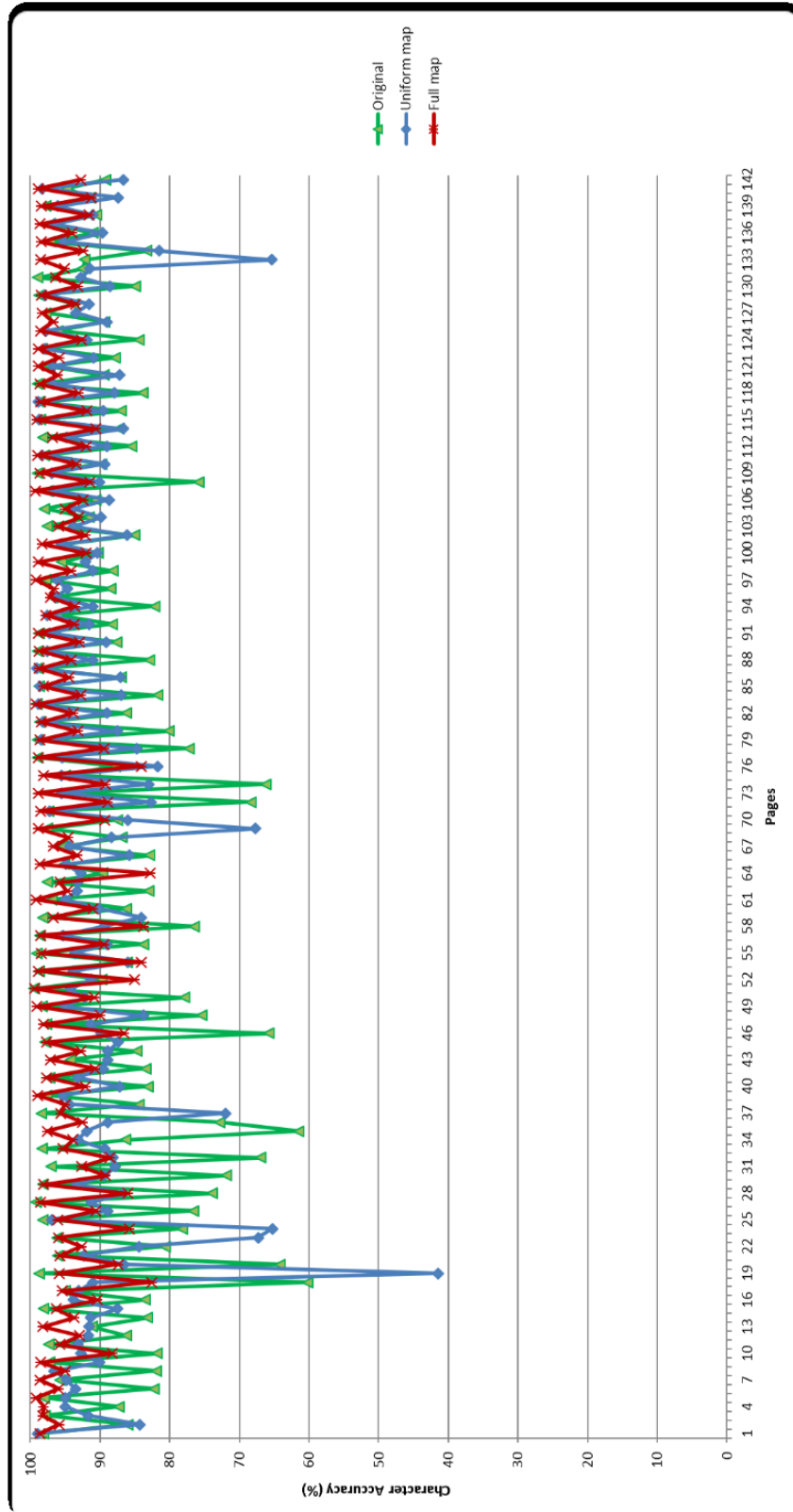


Figure 4.8 Accuracy of the original and corrected book spreads

Table 4.2 Summary of the rotated positioned book spreads

Procedure	Number of samples	Minimum accuracy	Maximum accuracy	Mean accuracy	Variance	Standard deviation	Median accuracy (after sort)
Original – warped	142	60.26%	99.51%	90.00%	84.82	9.21	95.74%
Uniform map	142	41.51%	99.03%	90.63%	54.88	7.41	91.56%
Full map	142	82.60%	99.37%	94.75%	16.88	4.11	95.74%

An example of the effect of the disturbance caused by the rotation can be appreciated in Figure 4.9, where the rotation caused the OCR to recognize more character on the warp image as can be seen in Figure 4.9(a) than those in Figure 4.5 (a). However, this does not always occur and yet there still many character which were not accurately recognize. The correction with the uniform map led to an improvement on the recognition on the upper portion of the text in the page. Unfortunately, the bottom part of the page did not fit that particular uniform map therefore was not properly corrected and thus many of the characters were improperly or not recognized at all as shown in Figure 4.9 (b). The full correction did improve the overall recognition of the characters as shown in Figure 4.9 (c) however this was not as high as the corrections on the standard placement image shown earlier in Figure 4.5 (c). Notable improvements are made over previous similar studies reported by our research group, mainly because of a unique registration process of the two imaging modalities. The low resolution (160 x 120 pixels) height information, as acquired by an Argos3D-P100 camera, accurately covers the entire book spread as captured by the high-resolution image (3072 x 2304 pixels). This has led to the seamless fusion of the two imaging modalities. Additionally, this process has helped in the development of a highly accurate dewarping procedure, which is supported through a strong mathematical framework and extensive empirical evaluations involving 200 pages.

lindid the fastenings of her mantle. Why, I never  
 fingers were trembling as she remembered having done  
 though I were choking? I had not gone out. I stood at the door for a few  
 minutes, and now I am quiet.

The tone of her voice was quite unlike her usual tones. It was evident to me  
 direction, and her voice was filled with a thousand venomous doubts and  
 that she was saying what she was concealing from me? Where had she  
 the wall, sick at heart, with her suspicions. What was it that I felt that I should have no peace until I  
 been during that strange episode? I had not seen her again after once she had told me what was  
 knew, and yet I shrink from asking her again after once she had told me what was  
 false. All the rest of the night I tossed and turned, each more unlikely than the last.  
 each more unlikely than the last. I should have gone to the doctor, but I was too disturbed in my mind

to be able to pay attention to business matters. My wife seemed to be as upset as  
 myself, and I could see from the little questioning glances which she kept shoot-  
 ing at me that she understood that I disbelieved her statement, and that she was  
 at her wits' end what to do. We hardly exchanged a word during breakfast, and  
 immediately afterwards I went out for a walk that I might think the matter out in  
 the fresh morning air.

I went as far as the Crystal Palace, spent an hour in the grounds, and was back  
 in Norbury by one o'clock. It happened that my way took me past the cottage,  
 and I stopped for an instant to look at the windows and to see if I could catch a  
 glimpse of the strange face which had looked out at me on the day before. As I  
 stood there, imagine my surprise, Mr. Holmes, when the door suddenly opened  
 and my wife walked out.

I was struck dumb with astonishment at the sight of her, but my emotions  
 were nothing to those which showed themselves upon her face when our eyes  
 met. She seemed for an instant to wish to shrink back inside the house again; and  
 then, seeing how useless all concealment must be, she came forward, with a very  
 white face and frightened eyes which belied the smile upon her lips.

"Ah, Jack," she said, "I have just been in to see if I can be of any assistance to  
 our new neighbours. Why do you look at me like that. Jack? You are not angry  
 with me?"

"So," said I, "this is where you went during the night."

"What do you mean?" she cried.

"You came here. I am sure of it. Who are these people that you should visit  
 them at such an hour?"

"I have not been here before."

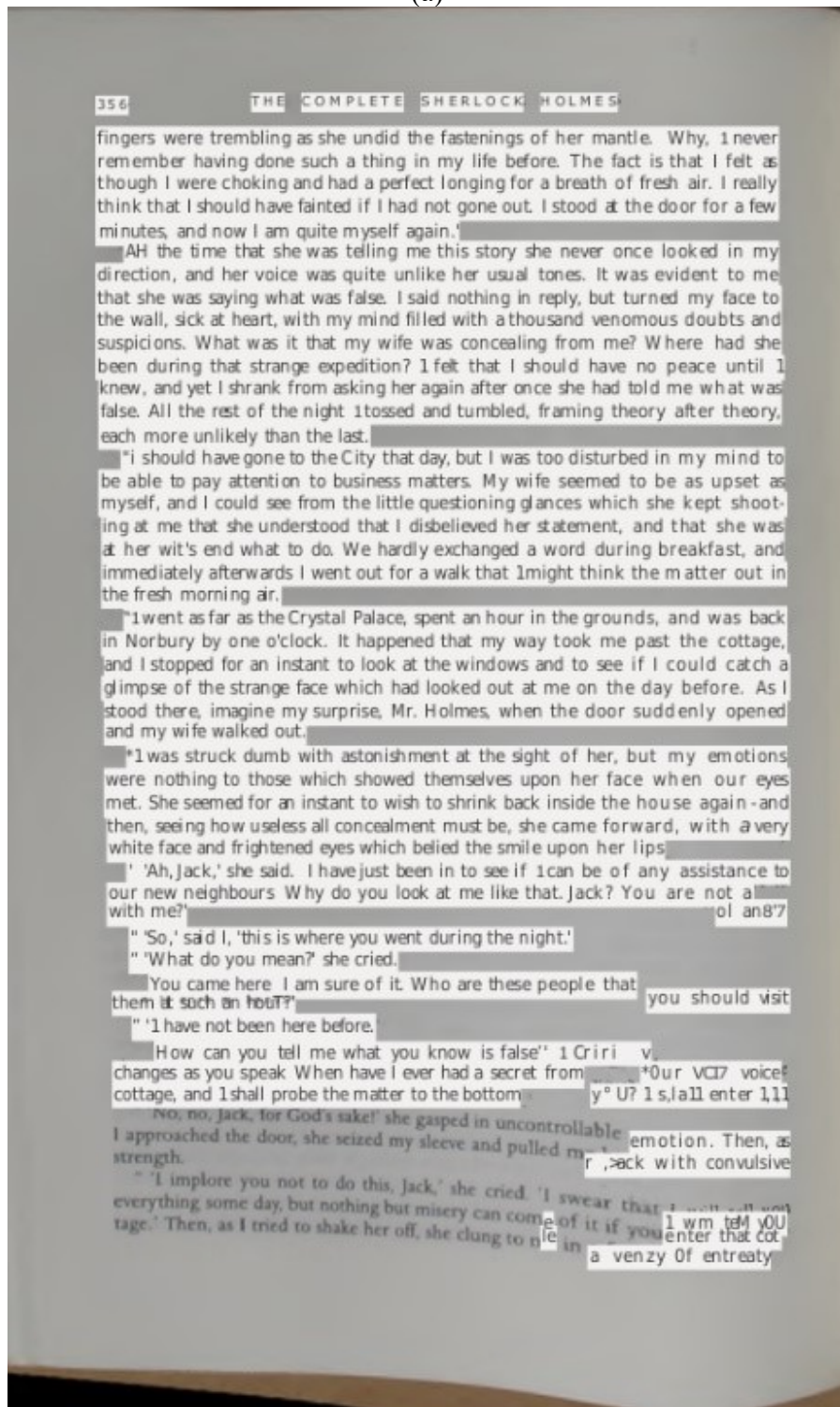
"How can you tell me what you know is false?" I cried. "Your very voice  
 changes as you speak. When have I ever had a secret from you? I shall enter that  
 cottage, and I shall probe the matter to the bottom."

"No, no, Jack, for God's sake!" she gasped in uncontrollable emotion. Then, as  
 I approached the door, she seized my sleeve and pulled me back with convulsive  
 strength.

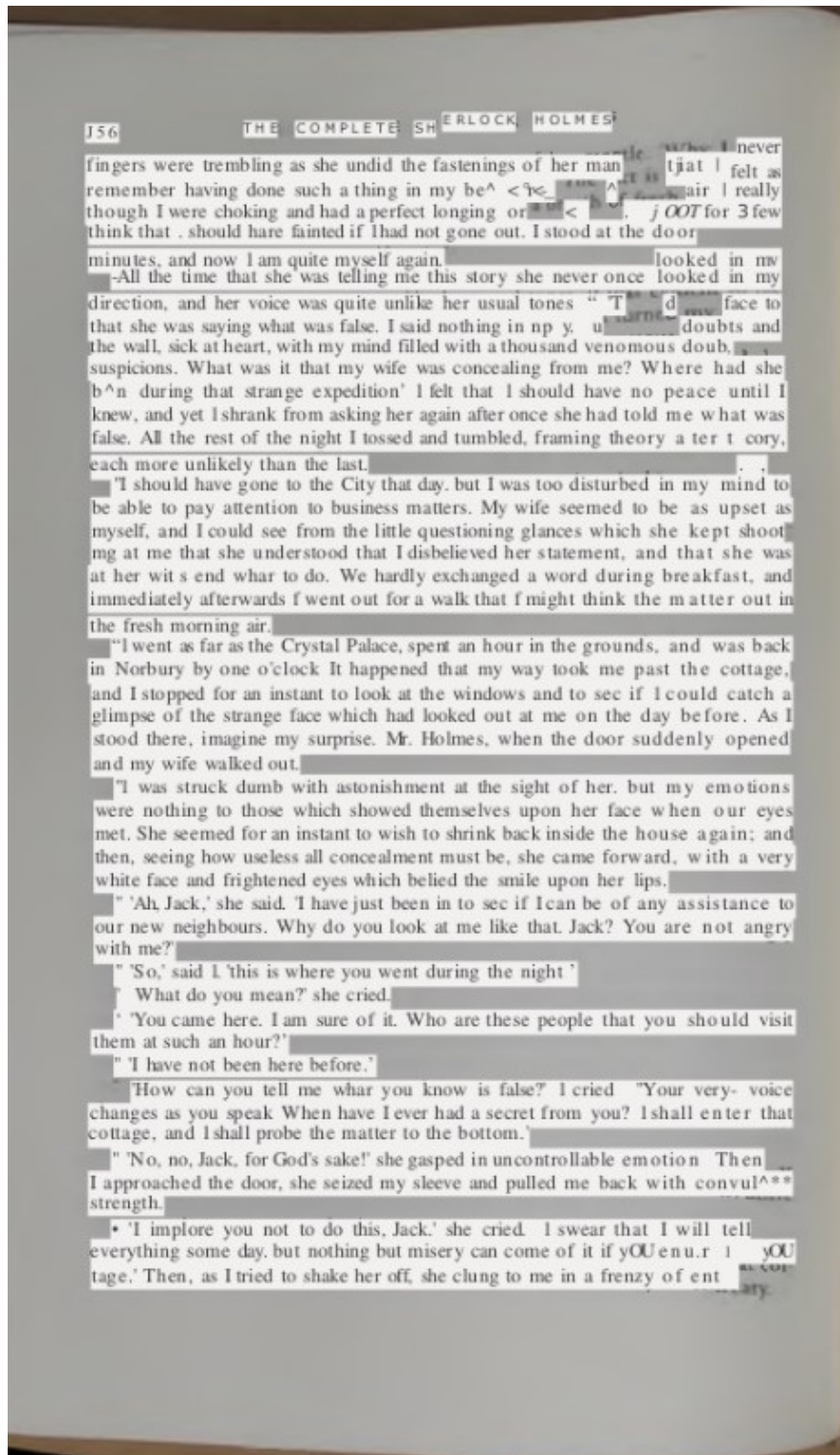
"I implore you not to do this, Jack," she cried. "I swear that I will tell you  
 everything some day, but nothing but misery can come of it if you enter that cot-  
 tage. I implore you, as I tried to shake her off, she clung to me in a frenzy of entreaty



(a)



(b)



(c)

Figure 4.9 Examples of rotated placement results for: a) original, b) corrections with uniform map, c) corrections with full map

Consequently, with the proposed book reader design, each pixel is now defined by its spatial coordinates  $x$  and  $y$  and its height ( $z$ ) which expresses the curvature of the book at that particular point. Certainly, additional improvements may relate to the inherent technological advances of the OCR engine itself. The results seen so far illustrate the effects of the uniform and height map. In the next chapter, it will be possible to observe if that behavior is exclusive to this correction algorithm.

## 5. FURTHER EXPLORATION OF UNIFORM VS. FULL HEIGHT MAPS

### 5.1 Adjusting an alternative method to utilize height maps

The approach which is updated for this chapter was chosen for its initial dewarping characteristics via lens correction from a surface map. This approach [18, 55, 57], which is based on a uniform map, was initially developed by obtaining the book's curvature information (height map) from a single side view (lateral) image, and assumed uniformity through the rest of the book's surface. However, ToF devices can be used instead to reduce the size of the book reader. The uniform map is then made possible by simply replicating the center curvature. Whereas, the full height map can be obtained from the raw height data provided by the ToF device. This height information was then used to perform the desired corrections.

### 5.2 Summary of adjustments

The location of a given object relative to the camera can cause the so-called warping effect. The first step of this method involves correcting the barrel distortion. This distortion is created because the magnification of the lens is inversely proportional to the distance from the optical axis. This makes a straight line in the real world seem as a curved line in the image plane as it moves away from the optical axis.

provides a visual representation of how a distorted straight line is corrected, whereas equation (5.1) provides the mathematical expression describing the distorted radius ( $r_{dist}$ ).

In equation (5.1)  $P^C$  is a vector containing the inherent parameters of the camera lens and

$R_{cor}^T$  is the transpose of a vector containing the radius of each pixel in the image. The radius of the distorted pixel's location in terms of its x and y coordinates ( $x_{dist}$  and  $y_{dist}$ , respectively) is described by equation (5.2).

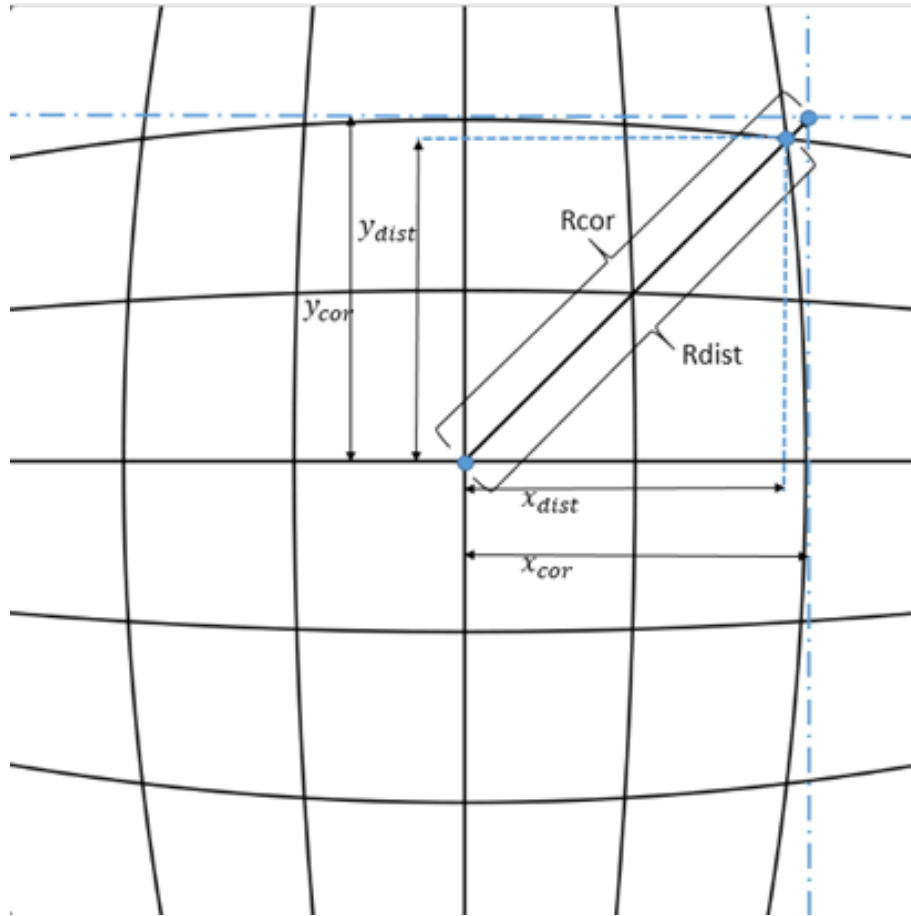


Figure 5.1 Barrel Distorted Image and its Correction

$$r_{dist} = P^C \cdot R_{cor}^T \quad (5.1)$$

$$r_{dist} = (x_{dist}^2 + y_{dist}^2)^{\frac{1}{2}} \quad (5.2)$$

In the implementation phase, an expansion of equation (5.1) as given by equation (5.3) is used. In this expression, term  $P_1^C$  represents the first coefficient of the  $P^C$  vector,  $P_2^C$  the second, and so on for the other terms. The  $r_{cor}$  terms are the elements of the  $R_{cor}^T$  vector.

$$r_{dist} = P_1^C * r_{cor}^4 + P_2^C * r_{cor}^3 + P_3^C * r_{cor}^2 + P_4^C * r_{cor}^1 \quad (5.3)$$

Thus, after the barrel correction is completed, a perspective transformation is applied to the image, referred to in this study as a “push operation”. This operation corrects the perspective transformation, which causes straight lines to appear curved due to the camera’s point of view as seen in **Error! Reference source not found.**.

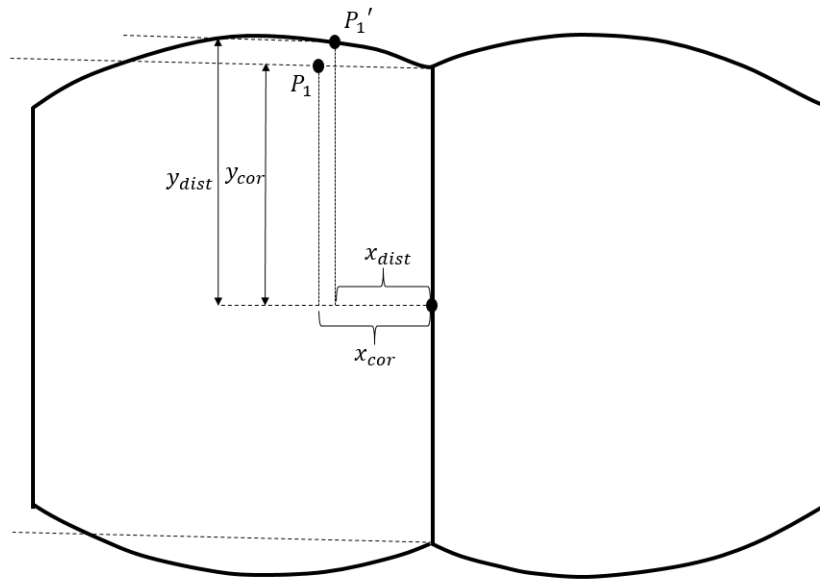


Figure 5.2 Push Operation

The push operation as described by equation (5.4) is used to obtain a pixel’s undistorted vertical distance from the camera’s center axis ( $y_{cor}$ ). In this equation,  $f$  is the camera focal length,  $h(x_{dist}, y_{dist})$  is the height of the distorted pixel (*i. e.*,  $P_1'$  in **Error! Reference**

source not found.) from the book stand surface in the real world, and  $y_{dist}$  is the distorted vertical distance of a given pixel from the camera's optical axis.

$$y_{cor} = (f \cdot y_{dist}) / (h(x_{dist}, y_{dist}) - f) \quad (5.4)$$

The last step for this method is an extension operation along the x-axis to eliminate the horizontal distortion that the book's curvature introduces in the image. This operation, originally introduced in [30] can be derived from Error! Reference source not found. to yield equation (5.5).

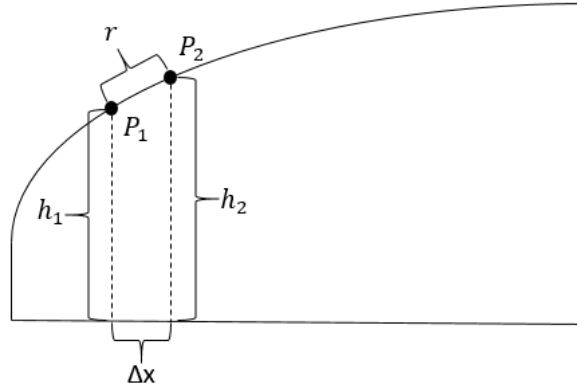


Figure 5.3 Section View of the Page Curvature

$$r = (\Delta x^2 + (h_2 - h_1)^2)^{\frac{1}{2}} \quad (V-5)$$

From equation (V-5), the flattened distance between two pixels, as represented by  $r$ , can be approximated by a triangle when the horizontal distance between said points ( $\Delta x$ ) is sufficiently small, and where  $h_1$  and  $h_2$  represent the respective heights of the two points with respect to the book holder's surface.

This method originally obtained the pixels' height values by extracting the book curvature from a side picture of the book spread, which is then uniformly distributed throughout the entire surface thereby generating what we refer to as a uniform height map. For this study, the height maps will be instead obtained using the experimental setup described in Chapter 2 using the data based described in Chapter 4. This setup will generate the uniform and full height maps which will then be used as the new values for  $h$  and  $h_i$  respectively.

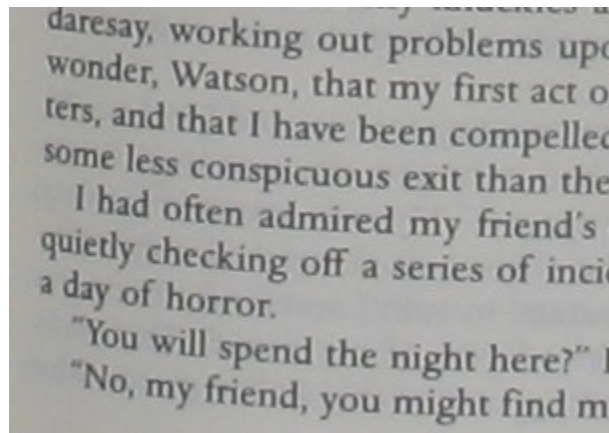
The following sections demonstrate the effect of the height maps on the de-warping corrections.

### 5.3 Standard placement

The standard placement is achieved by positioning the book spread so that it is aligned with the base of the book reader, which yields book spreads as previously seen in Figure 5.1. If we focus on a section of the book spread, it is possible to see the warping effect that occurs as demonstrated in Figure 5.4(a). This arching effect occurs in the majority of the tested book spreads. Figure 5.4(b) shows the result of the usual correction of the warped text using the uniform height map, while Figure 5.4(c) illustrates the correction with the full height map. The OCR results of the warped images can be seen in Figure 5.5. It should be noted from the results that for this type of correction, the OCR engine performs quite well. The results of the corrections using both the uniform map and full map are also depicted in Figure 5.5, both of which performed better than on the original image. Table 5.1 shows a statistical summary of the results for all the methods.



It is observed that all the different curvature correction procedures led to enhanced reading accuracy. Furthermore, using the full height map yielded better results than the uniformly distributed height map. This outcome was anticipated as the full height map reflects better the true nature of the page curvature, which may not necessarily be replicated by simply assuming a uniformly distributed curvature from single row of text or from a lateral view of the book. Correction errors most certainly occur due the complex nature of book curvatures as we flip through the pages. A smarter expansion process (i.e. recognizing key points as references from where to start the expansion) could improve these results. Nonetheless, we are still bound by limited range of current ToF technology which produces imperfections in extracting depth information. These perturbations become more noticeable when the height map is scaled to the higher resolution.



(a)

teeth I have barked my knuckles at  
daresay, working out problems upon  
wonder, Watson, that my first act of  
ters, and that I have been compelled  
some less conspicuous exit than the  
I had often admired my friend's  
quietly checking off a series of inci  
a day of horror.  
"You will spend the night here?" I  
"No, my friend, you might find m

(b)

teeth I have barked my knuckles and th  
daresay, working out problems upon a l  
wonder, Watson, that my first act on ente  
ters, and that I have been compelled to a  
some less conspicuous exit than the fron  
I had often admired my friend's coura  
quietly checking off a series of incidents  
a day of horror.  
"You will spend the night here?" I said.  
"No, my friend, you might find me a da

(c)

Figure 5.4 Examples of text: (a) original warped, (b) corrected with uniform map, and (c) corrected with full map

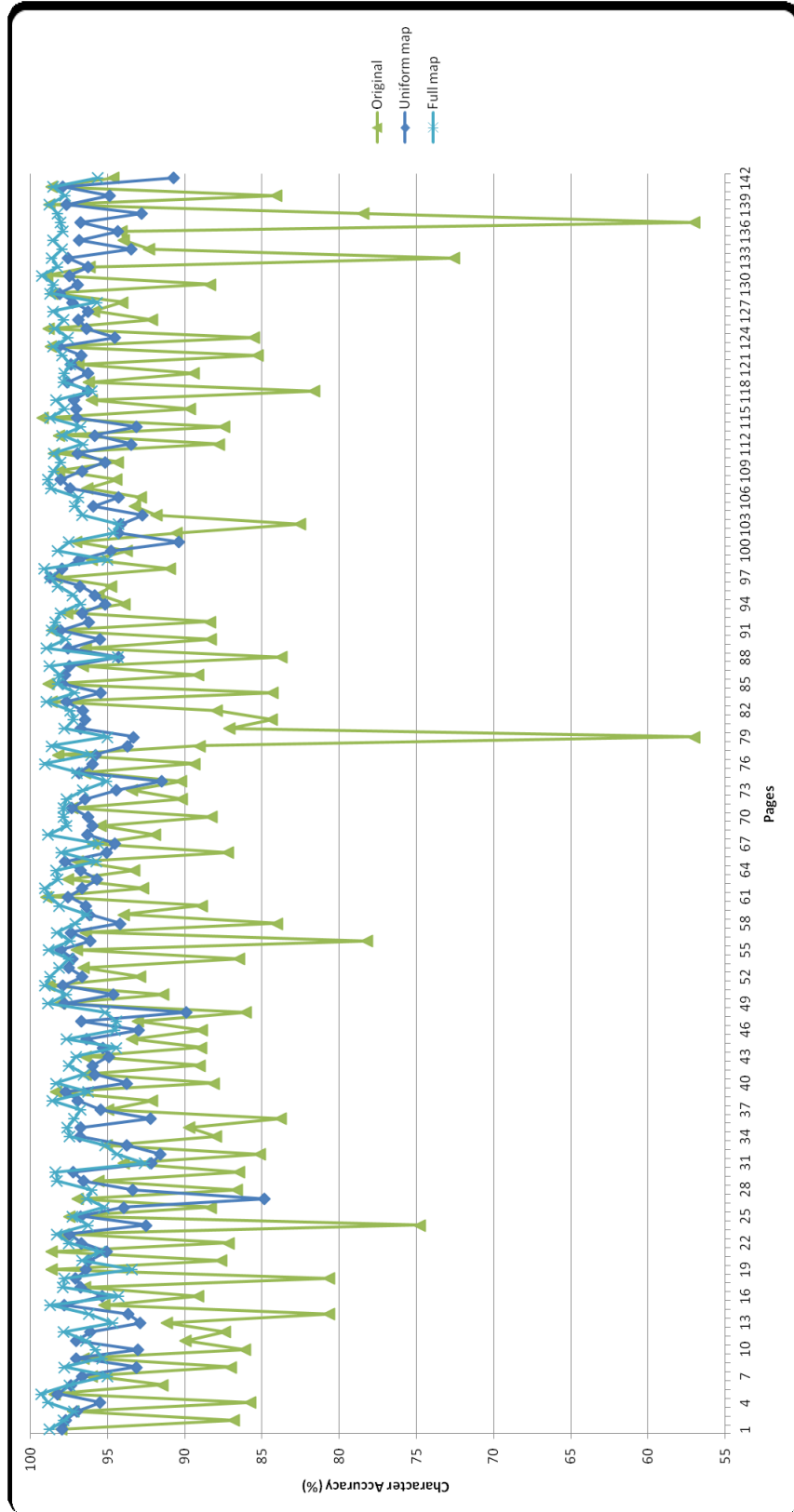


Figure 5.5 Accuracy of the original and corrected book spreads

Table 5.1 Summary of the standard positioned book spreads

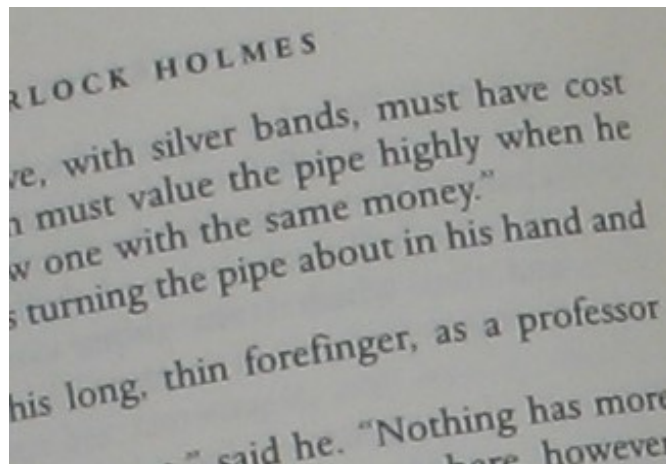
Procedure	Number of samples	Minimum accuracy	Maximum accuracy	Mean accuracy	Variance	Standard deviation	Median accuracy (after sort)
Original – warped	142	56.99%	99.18%	91.57%	48.17	6.94	93.25%
Uniform map	142	84.82%	98.70%	95.80%	4.14	2.03	96.40%
Full map	142	92.59%	99.28%	97.33%	1.99	1.41	97.78%

#### 5.4 Misaligned placement

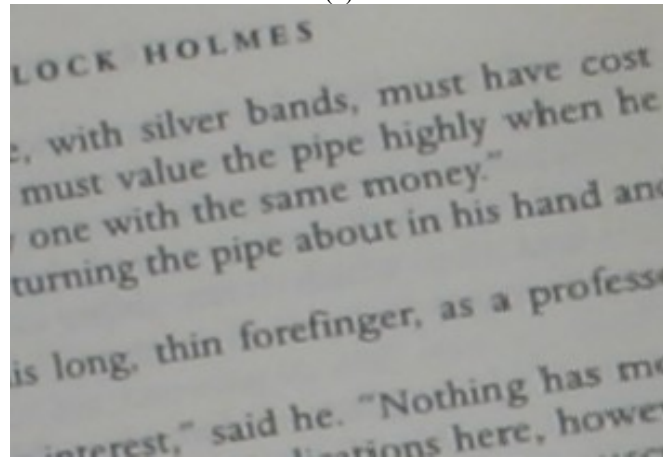
Many approaches rely on having a standard placement of the book spread, because of their sensitivity to misalignment in their standard placement. Therefore, to further test the resilience of the methods tested, as well as which type of height map has the best performance, the book spreads were physically rotated by 30 degrees. Then half of 142 pages were rotated 30 degrees left and the other half 30 degrees right, an example of which was previously seen in Figure 4.5.

The accuracy of the results for the rotated book spread images were quite high despite the rotation of the book spread. However, the overall accuracy was still lower than their standard placement counterparts. Figure 5.6(a) illustrates a zoomed-in section of the warping effect where the text lines appear curved instead of straight. Figure 5.6(b) shows how the correction algorithm does straighten the lines so that they are no longer warped while using uniform height map; while Figure 5.6(c) depicts the correction with a full height map. Table 5.2 shows a statistical summary of all methods with their corresponding maps. As expected, the method which had the highest overall accuracy and lowest variance was the full height map. On the other hand, the uniform maps had a very poor performance,

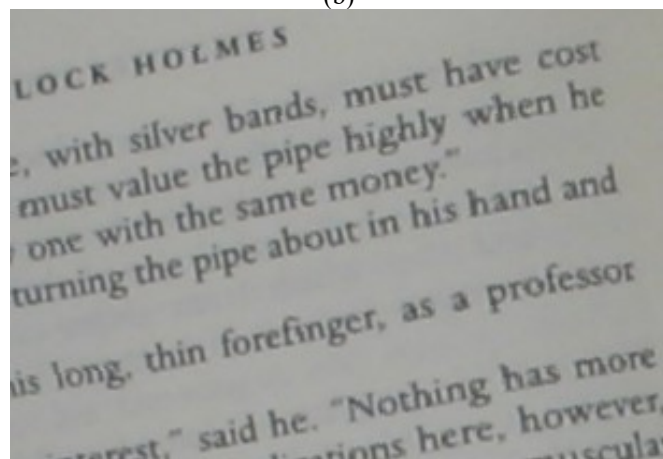
as they created a twisting effect in some cases that greatly decreased its performance even below the uncorrected instances, as observed in Figure 5.7.



(a)



(b)



(c)

Figure 5.6 Examples of text: (a) original warped, (b) corrected with uniform map, and (c) corrected with full map

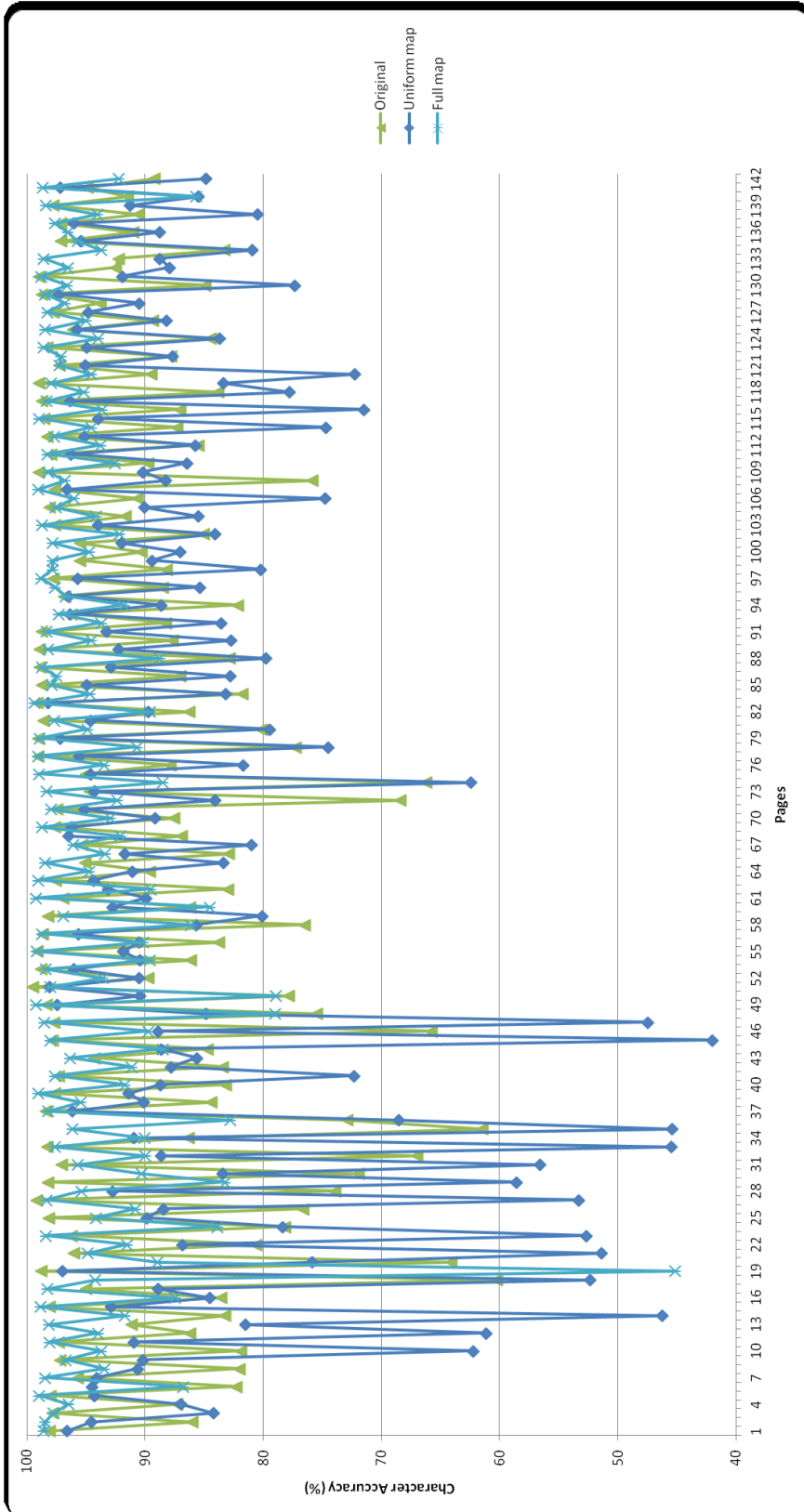


Figure 5.7 Accuracy of the rotated original and corrected book spreads

Table 5.2 Summary of the rotated positioned book spreads

Procedure	Number of samples	Minimum accuracy	Maximum accuracy	Mean accuracy	Variance	Standard deviation	Median accuracy (after sort)
Original – warped	142	60.26%	99.51%	90.00%	84.82	9.21	95.74%
Uniform map	142	42.00%	98.20%	85.28%	160.95	12.69	88.90%
Full map	142	45.16%	99.38%	94.62%	36.06	6.01	96.55%

## 6. CONCLUSIONS AND FUTURE WORK

In this chapter, the main results of this dissertation are summarized. Then other potential directions for future work are discussed.

### 6.1 Summary

This dissertation introduced a new fully-integrated and automated book reader design as an assistive technology tool for persons with visual impairment and blindness. Although such a system could be used for digitization purposes, its intent here is for reading books and other bound reading materials where text is distorted due to nonlinear and dynamic page curvatures as one leaves through the pages of the book.

The mathematical foundation that explores the geometry of the book reading setup, along with the height maps that were calculated, showed that performing dewarping corrections by integrating the full height maps with high resolution images provides significant text reading accuracy improvements. Furthermore, the proposed book reader proved to be more resilient to book spread misalignments, as reflected through the homogeneous rotation transformation examples. Such misalignments are shown to affect the reading accuracy when curvature correction is achieved using the uniform map (i.e., a map obtained through a single row height profile, which is then replicated throughout the other rows). However, when using the full height map (i.e., each pixel of the high-resolution image has a height associated with it, which means a full 3D map of the book spread) the results were as good as for those images acquired from a book that was properly placed.



Although the extension component of the algorithm introduces some noise, it still provided improved results in most cases. The extension operation of the algorithm (correcting for the text as the book spread is mathematically flattened) is shown to yield better results by selecting better reference points and/or performing smarter expansions; and by taking into account the orientation or any other misalignment of the book spread. Other factors which could improve the performance of the book reader could be a better registration process for matching or fusing the low-resolution height map to the high-resolution image (e.g. through trilinear image registration) as used with the renown FMRIB Software Library (FSL) software for the registration of the low-resolution PET to the high-resolution MRI brain scans in medical imaging.

It should be noted that if we were to seek high resolution height maps, which would help resolve or facilitate the registration process, the price of the ToF device that provides such high-resolution height information will increase manifold, but the intent here was to build a cost-effective book reader. Furthermore, although significant improvements have been made by current OCR engines such as the ABBYY FineReader 12, there is still room for improvement in order to achieve enhanced text recognition in distorted images of bound books and magazines due to curvature. Such curvature is dynamic in nature and a function of the number of pages that have been flipped to one side in relation to the other, with respect to the spine of the book, which makes the use of a ToF the more meaningful. This last challenge, of dynamic nonlinear curvatures of pages, is what led us to use a ToF device to extract the depth information, which in turn are converted into height maps.

In seeking cost effectiveness of the design, we managed to set up a mathematical framework to match the low-resolution height maps obtained using the cheapest ToF device available in the market to the high-resolution camera used for capturing the text to be fed into the OCR. It is worth noting that of all the studies that relate to this work and are referenced here, this research endeavor included the largest number of processed pages (200) that were acquired. Therefore, a website is made available to our research community to use such a database not only for developing new ways of constructing smoother height maps from the raw results of the ToF device, as they are matched to high-resolution images (also provided in the named website), but also in developing algorithms for new dewarping and flattening mechanisms that will further enhance OCR and text to speech accuracy, perhaps to be utilized in future book reader designs or other applications involving the use of raw height maps. This database is accessible through the following website: [cate-book3dmaps.fiu.edu](http://cate-book3dmaps.fiu.edu).

## **6.2 Future Work**

For future work, we intend to perform adjustments for improving the matching of the low-resolution height map with the high-resolution image through new different image registration processes that could account for the normalized mutual information between the two imaging modalities. Other improvement could be focused on the correction algorithm, for instance in the selection on smarter points to perform the expansion such as by considering the orientation of the book spread, and by determining the best fit for these nonlinear page curvatures given the raw height maps.

As observed through our experimental evaluations, current state-of-the-art OCR engines such as ABBYY FineReader 12 have experienced great design improvements that are more tolerant of misalignments, perspective and barrel effects, and rotation. However, additional improvements are still needed to contend with book curvature, which are inherently nonlinear with respect to both the spine of the book as well as the page number that is being read as one leaves through the book. In addressing this contentious issue, this study presents an alternative design for a portable book reader which can yield enhanced text identification, and hence higher reading accuracy, but with the added potential of using such a design construct for fully automating the process for both book readers and document digitization in line with the initiative of the Open Content Alliance (OCA) for the book digitization project.

## LIST OF REFERENCES

- [1] S. Pollard and M. Pilu, "Building cameras for capturing documents," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 7, no. 2-3, pp. 123-137, 2005.
- [2] M. M. Luqman, P. Gomez-Krämer and J.-M. Ogier, "Mobile phone camera-based video scanning of paper documents," *International Workshop on Camera-Based Document Analysis and Recognition*, vol. 8357, pp. 164-178, 2013.
- [3] H.-Y. Feng, Y. Liu and F. Xi, "Analysis of digitizing errors of a laser scanning system," *Precision Engineering*, vol. 25, no. 3, pp. 185-191, 2001.
- [4] V. Campanelli, S. M. Howell and M. L. Hull, "Accuracy evaluation of a lower-cost and four higher-cost laser scanners," *Journal of biomechanics*, vol. 49, no. 1, pp. 127-131, 2016.
- [5] A. Bhowmick and S. M. Hazarika, "An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends," *Journal on Multimodal User Interfaces*, vol. 11, no. 2, pp. 149-172, 2017.
- [6] Y. Tian, X. Yang, C. Yi and A. Ardit, "Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments," *Machine Vision and Applications*, vol. 24, no. 3, pp. 521-535, 2013.
- [7] B. Holton, "ScanSnap SV600 Overhead Scanner from Fujitsu: Turning the Page on Text Recognition," *AccessWorld Magazine*, vol. 15, no. 7, July 2014.
- [8] A. P. Tafti, A. Baghaie, M. Assefi, H. R. Arabnia, Z. Yu and P. Peissig, "OCR as a Service: An Experimental Evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym," in *International Symposium on Visual Computing: Advances in Visual Computing*, 2016.
- [9] S. Spooner, "'What page, Miss?' Enhancing Text Accessibility with DAISY (Digital Accessible Information SYstem)," *Journal of Visual Impairment & Blindness (Online)*, vol. 108, no. 3, pp. 201-211, 2014.
- [10] J. Sodnik, G. Jakus and S. Tomazic, "The use of spatialized speech in auditory interfaces for computer users who are visually impaired," *Journal of Visual Impairment & Blindness*, vol. 106, no. 10, pp. 634-645, 2012.

- [11] A. M. Mulloy, C. Gevarter, M. Hopkins, K. S. Sutherland and S. T. Ramdoss, "Assistive technology for students with visual impairments and blindness," *Assistive technologies for people with diverse abilities*, pp. 113-156, 2014.
- [12] N. Bourbakis, S. K. Makrogiannis and D. Dakopoulos, "A system-prototype representing 3d space via alternative-sensing for visually impaired navigation," *IEEE Sensors Journal*, vol. 13, no. 7, pp. 2535-2547, 2013.
- [13] R. Keefer, P. Kakumanu and N. Bourbakis, "A wearable document reader for the visually impaired: Dewarping and segmentation," *International Journal on Artificial Intelligence Tools*, vol. 18, no. 03, pp. 467-486., June 2009.
- [14] G. Annable, G. Goggin and D. Stienstra, "Accessibility, disability, and inclusion in information Technologies: Introduction," *The Information Society*, vol. 23, no. 3, pp. 145-147, 2007.
- [15] J. Abascal, S. D. Barbosa, C. Nicolle and P. Zaphiris, "Rethinking universal accessibility: a broader approach considering the digital gap," *Universal Access in the Information Society*, vol. 15, no. 2, pp. 179-182, 2016.
- [16] M. Adjouadi, J. Riley, F. Candocia, J. Andrian and H. Sumargo, "An augmented computer vision approach for enhanced image understanding," *Journal of Rehabilitation Research and Development*, vol. 32, no. 3, pp. 264-279, 1995.
- [17] M. Adjouadi, "A Man-Machine Vision Interface for Sensing the Environment," *Journal of rehabilitation research and development*, vol. 29, no. 2, pp. 57-76, 1992.
- [18] L. Wang, "An Automated Book Reader Design as an Assistive Technology Tool for Persons with Blindness," Florida International University, Miami, 2007.
- [19] H. Changan, "Neural Network Based Off-line Handwritten Text Recognition System," Florida International University, Miami, 2011.
- [20] C. Han, M. Adjouadi, A. Barreto, N. Rishe and J. Andrian, "Improved Pyramidal Neural Network for Segmented Handwritten Characters Recognition," in *International Conference on Image Processing, Computer Vision, and Pattern Recognition*, 2009.
- [21] R. Keefer and N. Bourbakis, "A survey on document image processing methods useful for assistive technology for the blind," *International Journal of Image and Graphics*, vol. 15, no. 01, p. 1550005, January 2015.

- [22] P. Kakumanu, N. Bourbakis, J. Black and S. Panchanathan, "Document Image Dewarping Based on Line Estimation for Visually Impaired," in *18th IEEE International Conference on Tools with Artificial Intelligence*, 2006.
- [23] F. Courteille, A. Crouzil, J.-D. Durou and P. Gurdjos, "Shape from shading for the digitization of curved documents," *Machine Vision and Applications*, vol. 18, no. 5, pp. 301-316, 2007.
- [24] J. An, H. I. Koo and N. I. Cho, "Rectification of planar targets using line segments," *Machine Vision and Applications*, vol. 28, no. 1, pp. 91-100, 2017.
- [25] E. Contini, B. Leporini and F. Paternò, "A Semi-automatic Support to Adapt E-Documents in an Accessible and Usable Format for Vision Impaired Users," in *International Conference on Computers for Handicapped Persons*, 2008.
- [26] A. Calabrò, E. Contini and B. Leporini, "Book4All: A tool to make an e-book more accessible to students with vision/visual-impairments," in *Symposium of the Austrian HCI and Usability Engineering Group*, 2009.
- [27] P. Xiu and H. S. Baird, "Whole-Book Recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 12, pp. 2467-2480, 2012.
- [28] B. Almasri, I. Elkabani and R. Zantout, "An Interactive Workspace for Helping the Visually Impaired Learn Linear Algebra," in *International Conference on Computers for Handicapped Persons*, 2014.
- [29] C. Simon and I. K. Park, "Correcting geometric and photometric distortion of document images on a smartphone," *Journal of Electronic Imaging*, vol. 24, no. 1, pp. 013038-013038, 2015.
- [30] I. Ahmad, S. A. Mahmoud and G. A. Fink, "Open-vocabulary recognition of machine-printed Arabic text using hidden Markov models," *Pattern Recognition*, vol. 51, pp. 97-111, 2016.
- [31] J. Liang, D. Doermann and H. Li, "Camera-based analysis of text and documents: a survey," *International Journal of Document Analysis and Recognition (IJ DAR)*, vol. 7, no. 2-3, pp. 84-104, 2005.
- [32] A. Ulges, C. H. Lampert and T. M. Breuel, "Document image dewarping using robust estimation of curled text lines," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, 2005.

- [33] V. Kluzner and A. Tzadok, "Information, Page Curling Correction for Scanned Books Using Local Distortion," in *2011 International Conference on Document Analysis and Recognition*, 2011.
- [34] J. Liang, D. DeMenthon and D. Doermann, "Geometric rectification of camera-captured document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 591-605, 2008.
- [35] S. Lu and C. L. Tan, "Document flattening through grid modeling and regularization," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006.
- [36] C. L. Tan, L. Zhang, Z. Zhang and T. Xia, "Restoring warped document images through 3D shape modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 195-208, 2006.
- [37] J. Liang, D. DeMenthon and D. Doermann, "Flattening curved documents in images," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
- [38] H. K. Aghajan, B. H. Khalaj and T. Kailath, "Estimation of skew angle in text-image analysis bySLIDE: Subspace-based line detection," *Machine Vision and Applications*, vol. 7, no. 4, pp. 267-276, 1994.
- [39] L. Song, Y. Wu and B. Sun, "A Robust and Fast De-warping Method of Document Images," in *2010 International Conference on E-Product E-Service and E-Entertainment*, 2010.
- [40] N. Stamatopoulos, B. Gatos, I. Pratikakis and S. J. Perantonis, "Goal-Oriented Rectification of Camera-Based Document Images," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 910-920, 2011.
- [41] L. Likforman-Sulem, J. Darbon and E. H. B. Smith, "Enhancement of historical printed document images by combining total variation regularization and non-local means filtering," *Image and vision computing*, vol. 29, no. 5, pp. 351-363, 2011.
- [42] K. Pal, M. Terras and T. Weyrich, "Interactive exploration and flattening of deformed historical documents," in *Computer Graphics Forum*, 2013.
- [43] 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 of Document Analysis and Recognition (IJDAR)*, vol. 7, no. 2-3, pp. 147-158, 2005.

- [44] G. Meng, C. Pan, S. Xiang and J. Duan, "Metric Rectification of Curved Document Images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 4, pp. 707-722, 2012.
- [45] C. Kim, P. Chiu and S. Chandra, "De-warping Book Page Spreads Captured with a Mobile Phone Camera," in *International Workshop on Camera-Based Document Analysis and Recognition*, 2013.
- [46] M. Leo, G. Medioni, M. Trivedi, T. Kanade and G. Farinella, "Computer vision for assistive technologies," *Computer Vision and Image Understanding*, vol. 154, pp. 1-15, 2017.
- [47] M. S. Brown, M. Sun, R. Yang, L. Yun and W. B. Seales, "Restoring 2D content from distorted documents," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, no. 11, pp. 1904-1916, 2007.
- [48] L. Galarza, Z. Wang and M. Adjouadi, "Book spread correction using a time of flight imaging sensor," in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, 2014.
- [49] L. Galarza, Z. Wang and M. Adjouadi, "Book reader optimization using a time of flight imaging sensor," in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, 2015.
- [50] L. Zhang, Y. Zhang and C. Tan, "An improved physically-based method for geometric restoration of distorted document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 728-734, 2008.
- [51] S. Foix, G. Alenya and C. Torras, "Exploitation of time-of-flight (ToF) cameras," Institut de Robòtica i Informàtica Industrial, Technical Report IRI-TR-10-07, 2010.
- [52] C. Kim, H. Yu and G. Yang, "Depth super resolution using bilateral filter," in *4th International Congress on Image and Signal Processing (CISP)*, 2011.
- [53] S. Schwarz, M. Sjöström and R. Olsson, "A weighted optimization approach to time-of-flight sensor fusion," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 214-225, 2014.
- [54] F. Garcia, B. Mirbach, B. Ottersten, F. Grandidier and A. Cuesta, "Pixel weighted average strategy for depth sensor data fusion," in *17th IEEE International Conference on Image Processing*, 2010.



- [55] M. Adjouadi, E. Ruiz and L. Wang, "Automated book reader for persons with blindness," in *International Conference on Computers for Handicapped Persons*, 2006.
- [56] L. Galarza, M. Harold and A. Malek, "Uniform vs Full Height Maps Using a Time of Flight Device for Dewarping Book Spread Images in the Design of an Automated Book Reader," in *International Conference on Computational Intelligence*, Las Vegas, 2016.
- [57] "Argos 3D - P100," Bluetechnix, 7 March 2016. [Online]. Available: [https://support.bluetechnix.at/wiki/Argos\\_3D\\_-\\_P100](https://support.bluetechnix.at/wiki/Argos_3D_-_P100).
- [58] L. Wang and M. Adjouadi, "Automated Book Reader Design for Persons with Blindness," in *International Conference on Computers for Handicapped Persons*, 2008.
- [59] K. Coyle, "Mass digitization of books," *The Journal of Academic Librarianship*, vol. 32, no. 6, pp. 641-645, 2006.

## APPENDICES

### Appendix A. Types of text digitization and reading devices

#### A.1 V-Shaped Book Scanners

## V-Shaped Book Scanners

### Robotic, Semi-automatic and Manual Book Scanners

Get a Quote on Qidenus Book Scanners




 <p><b>QIDENUS SMART</b></p> <ul style="list-style-type: none"><li>• Manual operating mode</li><li>• Available in A1 and A2 sizes</li><li>• 100-degree V-shaped book cradle</li><li>• Offers affordability and ease-of-use</li><li>• QIScan Software Suite integrated</li></ul> <p>Brochure</p>	 <p><b>QIDENUS MASTERED</b></p> <ul style="list-style-type: none"><li>• Manual and semi-automatic operating modes</li><li>• Available in A1, A2, A2+ and A3+ sizes</li><li>• 80-degree V-shaped book cradle</li><li>• Scan speeds up to 1,500 pages/hour</li></ul> <p>Brochure</p>	 <p><b>QIDENUS ROBOTIC</b></p> <ul style="list-style-type: none"><li>• Manual, semi- and fully-automatic operating modes</li><li>• Available in A2 and A3+ sizes</li><li>• 80-degree V-shaped book cradle</li><li>• Scan speeds up to 2,500 pages/hour</li></ul> <p>Brochure</p>
--	---	---

Figure A.1. Types of Qidenus V- Shaped Book Scanners

#### Properties

- Bulky, large storage space is required
- Manual, semi, and fully automatic operation
- Can handle large volume of pages
- Scanner only, corrections are made via hardware and software
- Very expensive (price is not posted and quotes are required)

For additional information on this system refer to

<http://www.thecrowleycompany.com/scanners-software/product-brands/qidenus/>

## A.2 FUJITSU Image Scanner ScanSnap SV600 and modification

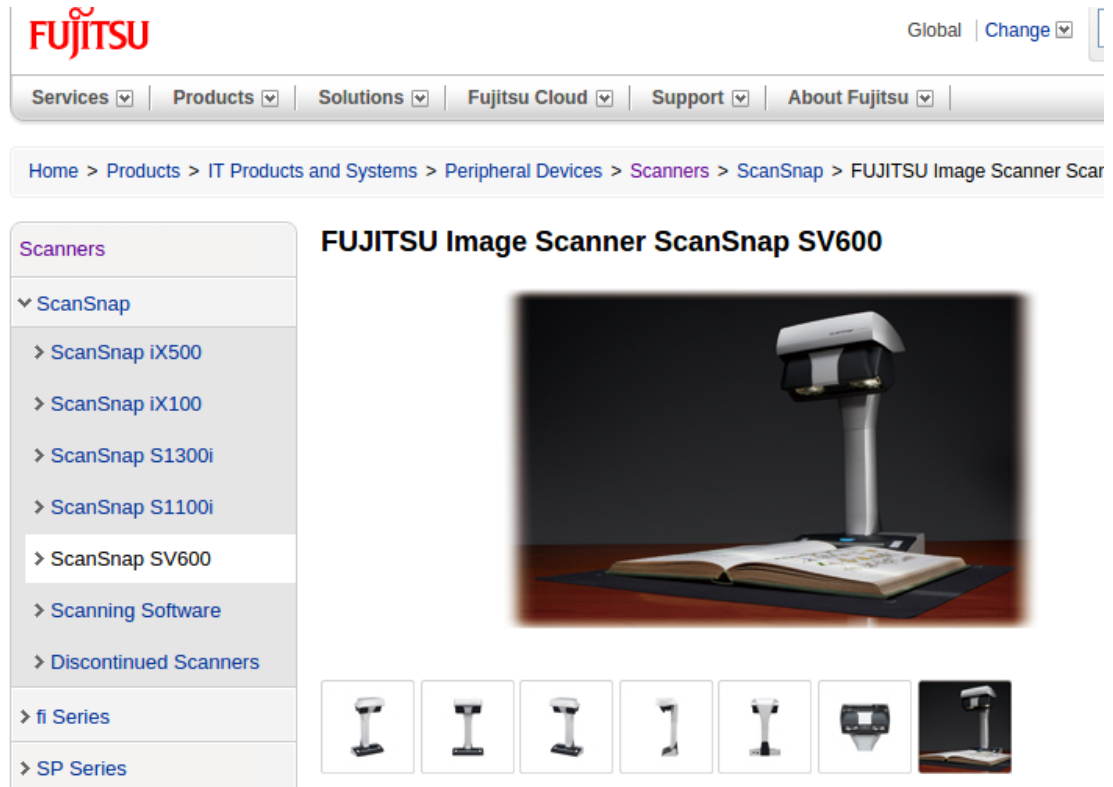


Figure A.2. Original FUJITSU Image Scanner ScanSnap SV600

### Properties

- Medium size scanner
- Portable
- Manual operation only
- Scanner only which can make corrections via software (based on text)

For additional information on this system refer to

<http://www.fujitsu.com/global/products/computing/peripheral/scanners/scansnap/sv600/index.html>



Figure A.3. Modified FUJITSU Image Scanner ScanSnap SV600

#### Properties

- This version includes an embedded OCR module
- Manual operation only
- Medium cost
- Requires specific drivers for operation

For additional information on this system refer to  
<http://www.recogniform.net/eng/recogniform-book-scanning.html>

### A.3 Bookeye 4



Figure A.4. Bookeye 4 with add-ons currently located at the FIU Library

#### Properties

- Medium to bulky depending on add-ons
- Very sensitive to placement of the book
- User must make manual adjustments to improve scan
- High resolution
- Expensive to very expensive depending on add-ons

For additional information on this system refer to  
<http://www.imageaccess.de/?page=ScannersBE4-SGS-V2Professional#>

#### A.4 Xcanex bookscanner



Figure A.5. Xcanex bookscanner by piQx

#### Properties

- Medium to small portable size depending on add-ons
- Manual operation only
- OCR software add-on is available
- Low cost

For additional information on this system refer to <http://www.piqximaging.com/>

## A.5 Transformer HD



Figure A.6. Transformer HD with Wi-Fi

### Properties

- Medium portable size depending on add-ons
- Manual operation only
- OCR software add-on is available
- Text to speech is available
- Very expensive

For additional information on this system refer to

<https://www.enhancedvision.com/low-vision-product-line/transformer-hd-portable-electronic-magnifier.html>

## A.6 Smart Reader



Figure A.7. Smart Reader

### Properties

- Medium portable size depending on add-ons
- Manual operation only
- Includes OCR module for conversion
- Text to speech can be performed without using add-ons
- Very expensive

For additional information on this system refer to <https://www.enhancedvision.com/low-vision-product-line/smart-reader.html>



## A.7 IRIScan Book 5



Figure A.8. IRIScan Book 5 Wifi

### Properties

- Small portable device
- Needs a great deal of manual intervention
- Prone to mistakes by user
- OCR module included
- Low cost

For additional information on this device refer to

<http://www.irislink.com/EN-US/c1485/IRIScan-Book-5---IRIScan-Book-5-Wifi---Book-scanner.aspx>

## A.8 ReadingPen 2

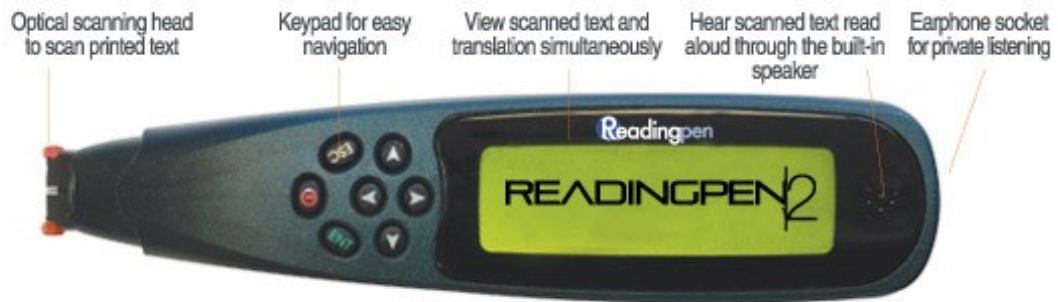


Figure A.9. ReadingPen 2

### Properties

- Small portable device
- Limited number of characters can be capture at a time
- Needs a great deal of manual intervention
- Prone to mistakes by user
- Text to speech can be performed without using add-ons

For additional information on this device refer to

<http://www.wizcomtech.com/a-product?id=97&lang=en>

A summary on the properties of these devices is provided in Table A.1

Table A.1 Summary of devices

Device	Size	Mode of Operation	OCR	TTS	Price Range (\$)
V-Shape Book Scanners	Large	Manual, Semi or Automatic	Yes	No	28,000-110,000
ScanSnap SV600	Medium	Manual	Add-on	No	400-1,300
Bookeye 4	Medium - Large	Manual	Add-on	No	expensive - very expensive
Xcanex bookscanner	Medium	Manual	Add-on	No	300-600
Transformer HD	Medium	Manual	Yes	Add-on	2,600
Smart Reader	Medium	Manual	Yes	Yes	1,800
IRIScan Book 5	Small	Extensive Manual	Yes	No	100-150
ReadingPen 2	Small	Extensive Manual	No	Yes	170-300

## VITA

### LUIS GALARZA

- 2002      B.S., Electrical Engineering  
Florida International University  
Miami, Florida
- 2005      M.S., Electrical Engineering  
Florida International University  
Miami, Florida
- 2017      Ph.D., candidate, Electrical Engineering  
Florida International University  
Miami, Florida

### PUBLICATION AND PRESENTATIONS

1. Galarza, Luis, Harold Martin, and Malek Adjouadi. 2018. "Time of Flight Sensor in a Book Reader System Design for Persons with Visual Impairment and Blindness." *IEEE Sensons Journal* – Second revision.
2. Galarza, Luis, Harold Martin, and Malek Adjouadi. 2018. "Integrating Low Resolution Depth Maps to High Resolution Images in the Development of a Book Reader System for persons with Visual Impairment and Blindness." *Journal of Innovative Computing, Information and Control (IJICIC)* ISSN 1349-4198. 14 pages.
3. Galarza, Luis, Harold Martin, and Malek Adjouadi. 2016. "Uniform vs Full Height Maps Using a Time of Flight Device for Dewarping Book Spread Images in the Design of an Automated Book Reader." *International Conference on Computational Intelligence (CSCI)*. Las Vegas: IEEE. 692-697.
4. Galarza, Luis, Harold Martin, and Malek Adjouadi. 2016. "Improving Word Recognition with a Time of Flight 3D Camera." *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*. Las Vegas: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). 56-62.
5. Galarza, Luis, Zhenzhong Wang, and Malek Adjouadi. 2015. "Book reader optimization using time of flight imaging." *CAHSI Summit 2015*. San Juan. 96-99.
6. Martin, Harold, Luis Galarza, Lu Wang, and Malek Adjouadi. 2015. "Uniform vs full height map approaches for dewarping of book spread images." *CAHSI Summit 2015*. San Juan. 100-103.

7. Galarza, Luis, Zhenzhong Wang, and Malek Adjouadi. 2015. "Book reader optimization using a time of flight imaging sensor." *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*. Las Vegas: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). 324-330.
8. Galarza, Luis, Zhenzhong Wang, and Malek Adjouadi. 2014. "Book spread correction using a time of flight imaging sensor." *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*. Las Vegas: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). 1-7.
9. Wen-Cheng, Lai, Sun An-Cheng, Chen Nan-Wei, Ching-Wen Hsue, and Luis Galarza. 2013. "RFID Reader Range Extension using Magnetodielectric Material An Extended Version." *i-manager's Journal on Electronics Engineering* 3 (2): 1-6.
10. Guo, Lilin, Luis Galarza, Jeffrey Fan, and Chiu Choi. 2013. "High accuracy three-dimensional radar sensor design based on fuzzy logic control approach." *45th Southeastern Symposium on System Theory*. IEEE. 22 - 26. doi:10.1109/SSST.2013.6524941.
11. Niu, Xinwei, Luis Galarza, Ying Gao, and Jeffrey Fan. 2012. "Software-hardware co-design for video coding acceleration." *Proceedings of the 2012 44th Southeastern Symposium on System Theory (SSST)*. IEEE. 57 - 60. doi:10.1109/SSST.2012.6195122 .
12. Galarza, Luis, and Frank Candocia. 2006. "Optimal and Dense Small Baseline Stereo Image Correspondence." *2006 International Conference on Image Processing*. IEEE. 1037-1040. doi:10.1109/ICIP.2006.312732.