International Journal of Innovative Computing, Information and Control
Volume **14**, Number **3**, June **2018**

ICIC International ©2018 ISSN 1349-4198
pp. **797–816**

INTEGRATING LOW-RESOLUTION DEPTH MAPS TO HIGH-RESOLUTION IMAGES IN THE DEVELOPMENT OF A BOOK READER DESIGN FOR PERSONS WITH VISUAL IMPAIRMENT AND BLINDNESS

Luis Galarza, Harold Martin and Malek Adjouadi

Center for Advanced Technology and Education
Department of Electrical and Computer Engineering
Florida International University
10555 W. Flagler Street, Miami, FL 33174, USA
{ Luis.galarza; hmart027; adjouadi }@fiu.edu

Received September 2017; revised January 2018

Abstract. *The objective of this study 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 × 120 pixels) depth information, acquired by an Argos3D-P100 camera, accurately covers the entire book spread as captured by the high-resolution image (3072 × 2304 pixels), acquired by a Canon G6 Camera; (2) a mathematical framework for overcoming the curvature of open bound books, a process we refer 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 testing dataset consisting of 200 pages with their corresponding computed and co-registered height maps, which are made available to the research community for their own testing (cate-book3dmaps.fiu.edu). Improvements of the 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 missrecognized 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 depth information 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.*

Keywords: Book reader, Visual impairment and blindness, Book curvature correction, Time of flight (ToF) device, Height/depth map, Text digitization, Optical character recognition (OCR), Assistive technology

1. **Introduction.** In this study, we assert that there is great need and demand for an automated book reader design to be used as an assistive technology tool by individuals with visual impairment and blindness in the comfort and privacy of their own homes or offices, and be able to use it on any bound book or magazine they wish to read. This provides them 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. Although such technology can be useful primarily as a reading device for visually impaired and blind individuals as part of the challenge for universal access to education, it can also serve as part of a practical solution to the remarkably challenging book digitization project, which benefits primarily the user as this audio version can now be 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 needed of this design include: portability, cost-effectiveness, and a user-friendly graphical interface. The proposed design meets these requirements as it can be connected to any computer through USB ports, any platform could serve as the book holder as long as its surface is not of black color (as it absorbs most of the IR light). The two cameras were chosen for their relatively low costs, the upright camera holder costs around $20.00 and can be easily set up, and with all the operational steps of the user interface automated, including image acquisition, dewarping of the images, and OCR for reading text. Furthermore, any pages that have been read are also automatically saved in memory as digitized text for future use.

Recent papers in assistive technologies for the visually impaired [1,2] provide insightful applications 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 in [1] rely on components available only online. There are other reading devices, currently in the market, but they are designed primarily for document digitization [3,4], which are mostly bulky and rather expensive.

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 in [5]. JAWS is another software solution which 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 [6], 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 more tuned 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.

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 [7]. 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. Other prototypes use two cameras and range sensors to gather data about a surrounding environment to enhance navigation. One example is the case of [8] 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 [9] 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 [10,11], 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 [12,13].

As a testament to the operational success of our initial book reader design which was based on a dual digital camera system [55], four blind individuals from the Lighthouse of Broward and its director have expressed full satisfaction with the reading accuracy of this initial prototype after having evaluated and tested it. They have also indicated that they were not worried about the text recognition errors as they are quite adept at using context to overcome these errors 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 [14,15]. The challenge remains in addressing the heavy computational requirements before hand-written text it could to be integrated into the proposed book reader design.

Several book and document reader designs with different system setups, 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 [16]. The ultimate goal 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. Our group's early design attempts involved the use of a lateral camera to estimate the page curvature, and a top camera, to read the text once the dewarping was accomplished [17,18]. In this case, page flattening is accomplished mathematically by estimating the curvature captured by the lateral camera using polynomials, and determining the mathematical framework for stretching the polynomial and projecting the text characters into a flattened 2D plane are the contributions made by these early studies.

Remarkably, the work in [19] uses a single high-resolution camera, and the dewarping process is accomplished by taking consideration of 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 with the naked eye, while other words in a same paragraph appeared italicized next to others that looked correctly rectified. Shape from shading rectification approaches, where the shape of the document is inferred from the shadow it casts, has also been explored in [20]. Although this approach could yield positive results, it is highly dependent on lighting conditions and is prone to variations. Whereas, [21] aims to correct the deformations of planar documents by using minimization functions that yield rectification transformations. Studies in [22,23] 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 have to be added manually in a semiautomatic process [22]. In a sophisticated image analysis strategy seeking a higher reading accuracy, the study reported in [24] 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 (high mutual entropy to mean 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 looked into automating reading and text tools for visually impaired persons [25] 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 [26] focuses more on the removal of the so-called Moiré-pattern noise and specular highlight 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 [27], 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 [27] 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), and 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 which completely bypasses any segmentation process.

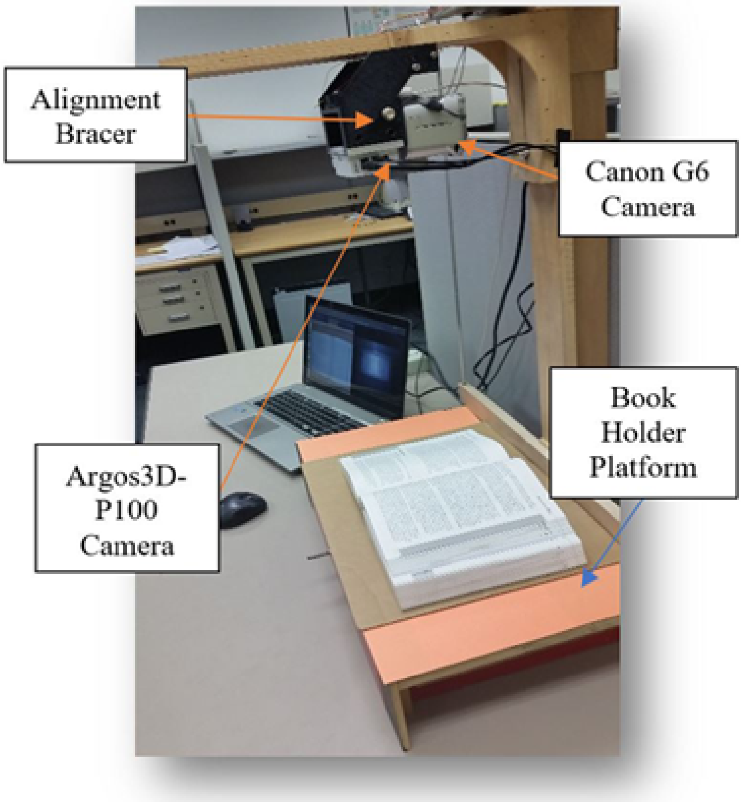
However, there is still room for improvement in these systems, especially in terms of recognition rate and reading accuracy. A thorough and interesting survey [28] highlights the merits and challenges of the different existing systems and approaches used for reading documents.

Several studies have looked into the challenge of dewarping documents in order to improve text recognition and ultimately the reading accuracy by means of existing and much improved OCR systems. Interesting studies have been reported on means to rectify distorted text by either estimating the curled text lines [29,30], looking into the geometric properties of the captured images [31], 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 [32], 3D shape modeling [33], by considering parallelism and equal line spacing [34], modeling lines as wavefronts of planar waves [35], looking into slopes in words for realignment [36], or by using curved surface projection modeling and baseline estimation as a goal-oriented rectification [37]. The concepts of dewarping and flattening have also been explored for other applications that include the restoration and enhancement of historical documents [38,39], and for recognizing street signs from 3D scenes [40]. Cylindrical surface modeling has also been considered through the use of an isometric mesh [41] 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 [42]. In retrospect, we do agree with the assessment provided in [43] that in many surveys related to assistive technology, there is a lack of details in the technical and functional improvements that such devices provide.

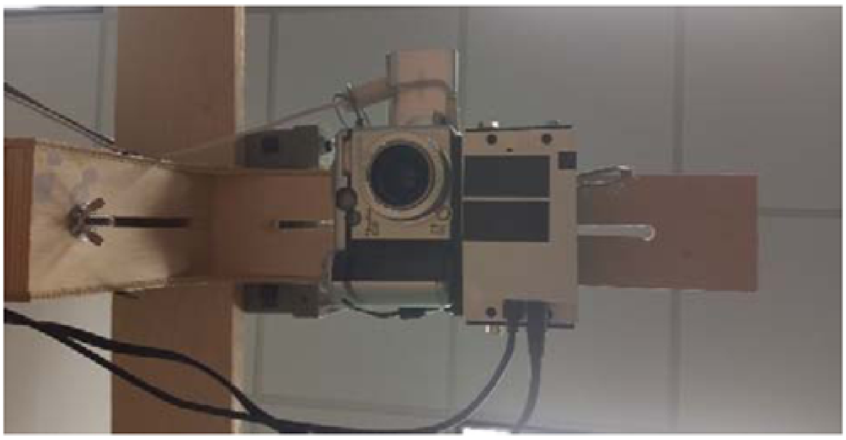
Our design focuses on the book curvature as a surface geometry problem, as obtained directly from height (depth) extraction [44-47], given the remarkable advances made in time of flight (ToF) sensors along with their increased range of applications [48-54]. With this kind of ToF devices, it became possible through the proposed design to determine height maps that reflect the book's curvature with sufficient accuracy. Therefore, algorithms relying on height maps to improve character recognition through distortion correction can now be further explored. We particularly focus 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 3D map are used for all other rows (as was originally used in [17]). Once these maps are determined, the curvature correction process takes place, and the corrected images are then fed into the OCR engine.

Our earlier efforts at designing automated book readers [45,46,47] have been extended to include improvements in the alignment of the Canon G6 Camera and Argos3D P100 sensor. An alignment bracer is used for the proper placement of heat sink and fan which can extend the operational time of the ToF device. The new design is also fully automated, requiring the pressing of single button we named *Start Capture*. Furthermore, the testing database has also been extended from 142 pages to 200 pages to include new examples that address different alignment issues that could lead to reading errors, thus providing more statistical meaningfulness to the reading accuracy of our new approach. More empirical and mathematical fine-tuning resulted from these additional experiments. Furthermore, examples of the matched high-resolution images with the corresponding high-resolution height maps are included in cate-book3dmaps.fiu.edu for the research community to use as an accessible database. Researchers could thus validate their work or create new algorithms based on 3D maps of over 200 processed pages, something that has never been made possible in the past. Research on this data website that we offer as part of this publication could provide other researchers the means to propose new mathematical ways to correct for the page curvature, perform comparative studies with methods that use stereo vision, grid modeling, or any other approaches they use for de-warping book spreads. This certainly will advance the research potential in the automated book reader design and related applications. The following sections will go into more details as to how these text distortion corrections are accomplished. The resilience of our book reader's correction algorithm is further explored by rotating the book slightly (30° clockwise and counterclockwise) to introduce a type of misalignment that could happen when a user places the book on the book holder platform.

2. **Experimental Setup and System Design.** In the design shown in Figure 1(a), a Bluetechnix Argos3D-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 depth map image of 160×120 pixels, which was converted to a height map (i.e., depth 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 set up as shown in Figure 1.



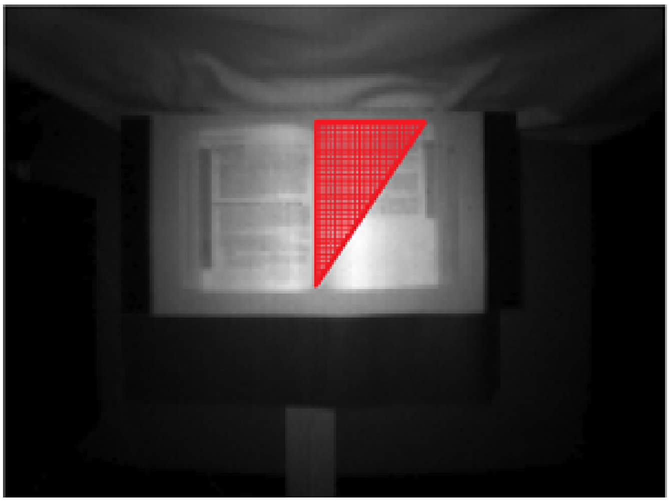
(a) Design setup



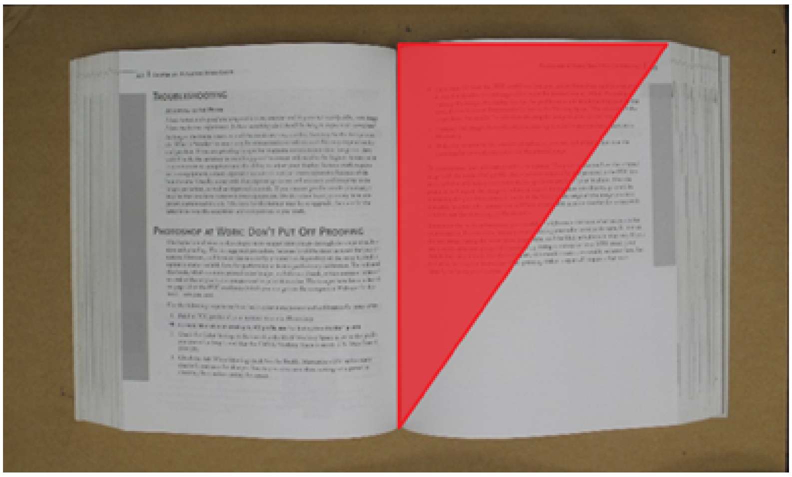
(b) Placement of the cameras for fusing the 3D depth map (or height map) to the high-resolution book spread image

Figure 1. Experimental setup of the book reader design

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



(a) The 2-D low-resolution book spread of the ToF sensor



(b) The 2-D higher resolution book spread of the Canon G6 Camera

FIGURE 2. Visual appreciation of the two imaging modalities from (a) the low-resolution ToF Sensor and (b) the high-resolution Canon G6 Camera

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.

For visual appreciation of this challenge, an example of the low-resolution image can be seen in Figure 2(a), and the corresponding high-resolution capture with the Cannon G6 can be seen in Figure 2(b). The virtual triangles shown in these two imaging modalities reflect the 3 points (2 along the spine of the book and a third point on one of the corners of the book) to be matched using affine transformations for registering the two images. Details of this registration method along with the mathematics used to derive the required affine transformations are presented in Section 3.

3. **Methods.** The main objectives sought in this book reader design are (1) to address the effect of the curvature of the pages of any bound book or magazine, and (2) to incorporate a height map that can best describe the curvature of the book spread. Within these two objectives, the main goal is to determine the best way to mathematically flatten the book spread so as to yield the highest reading accuracy. The main challenge in seeking this outcome is in the ability to create a complete high-resolution 3D image by fusing a

low-resolution (sparse) 3D map for evaluating the page curvature to a high-resolution 2D image used primarily for reading text.

3.1. **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 3, introduced in earlier studies of our research group [45-47].

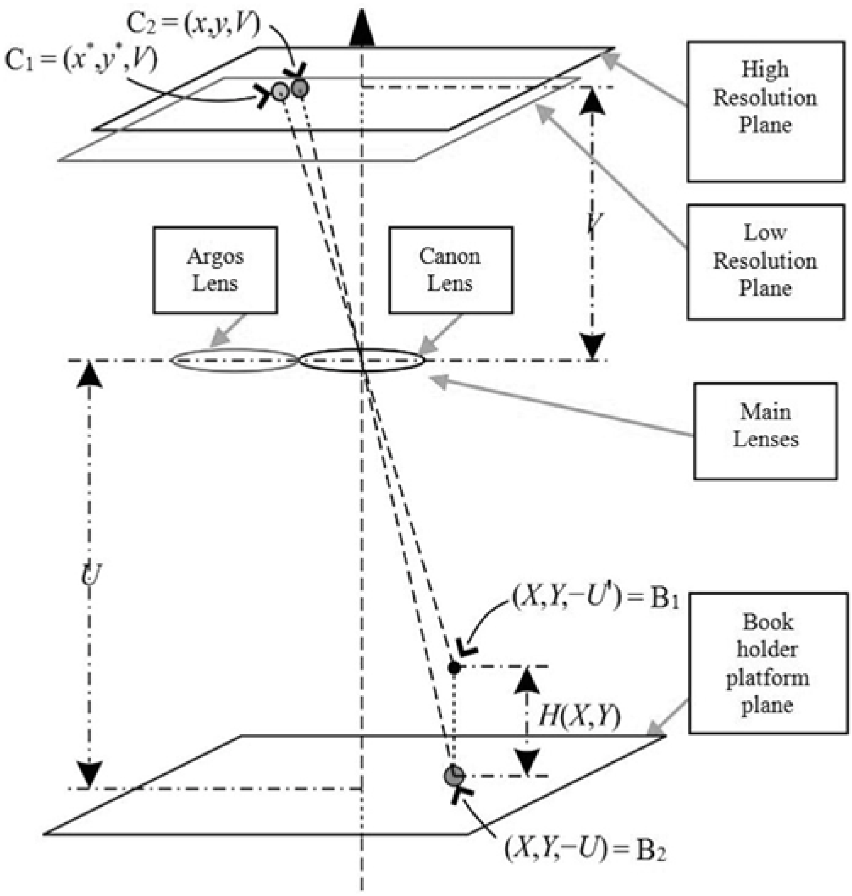


FIGURE 3. Conceptual design and geometry of the book reader

In this figure, 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. 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:

$$y/Y = x/X = V/U \tag{1}$$

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

$$x^*/X = V/(U - H(X,Y)), \text{ and } y^*/Y = V/(U - H(X,Y)) \tag{2}$$

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:

$$H(X,Y) = H(x \times (U/V), y \times (U/V)) = h(x,y) \tag{3}$$

Measure $h(x,y)$ is the height to be recovered from the dewarped 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:

$$h(x,y) = h(x^* + \Delta x^*, y^* + \Delta y^*) \approx h(x^*,y^*) \tag{4}$$

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

$$x = x^* \times (U - h(x^*,y^*))/U, \text{ and } y = y^* \times (U - h(x^*,y^*))/U \tag{5}$$

Applying this transformation (5) along all points will flatten the image. To further improve these results so that the characters 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 has 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 so as to undo the book curvature. Given the differential change in height defined as:

$$dh_i = h_{i+1}(x^*,y^*) - h_i(x^*,y^*) \tag{6}$$

with the differential change along the x direction being 1 pixel as in (7)

$$dx_i = x_{i+1} - x_i = 1 \text{ pixel} \tag{7}$$

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

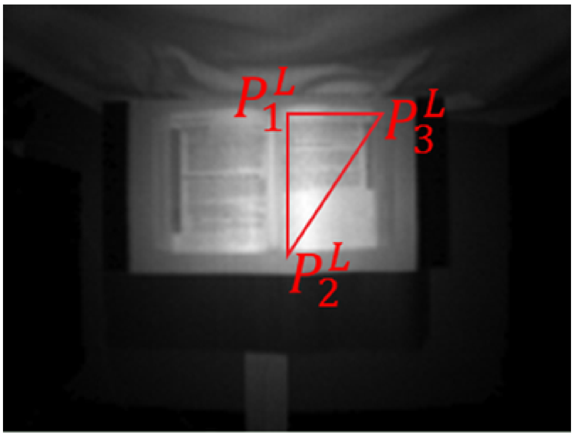
$$L_i = (dh_i - dx_i)^{\frac{1}{2}} \tag{8}$$

Using Equation (8), as a function of x_i, it is then possible to express the new extended locations as:

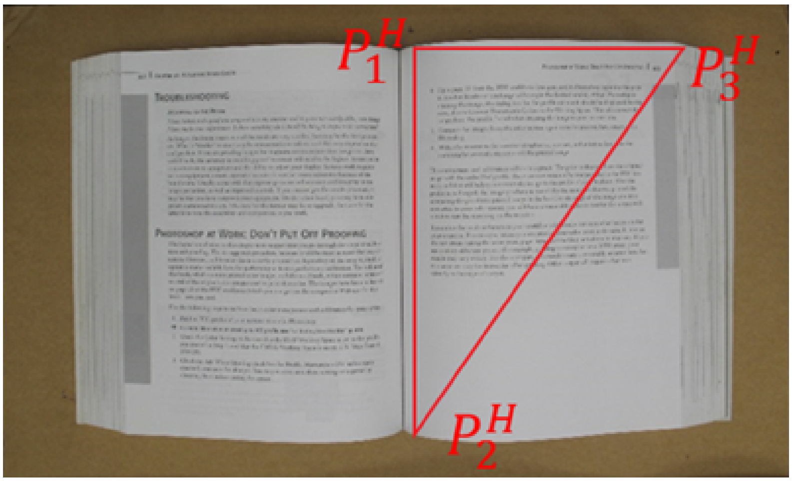
$$x_{e_{i+1}} = \sum_{i=0}^{n} (x_{e_i} + L_i) \tag{9}$$

where $x_{e_0} = x_0 = 0$, representing the starting point at the leftmost beginning of the book (opposite side of P_3^H in Figure 4(b)) with n being 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. However, 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 affects the dimension of the text as reflected in Figures 9 and 10 for example. Thus, flattening of the book spread followed by the extension process allowed for the image to be corrected appropriately. These corrections have significantly attenuated the warping effect as the results in Section 4 of this new design construct would confirm later.

3.2. **Registration of imaging modalities.** 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, and P_3^L) in the low-resolution image as shown in Figure 4(a) and as (P_1^H, P_2^H, and P_3^H) in the high-resolution image as shown in Figure 4(b). The new pixel values were then estimated using bilinear interpolation.



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



(b) The 2-D higher resolution book spread as that found in Figure 3

FIGURE 4. Geometry for registering the two imaging modalities

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 is matched.

The steps that perform such corrections are as follows:

(1) Determine the measurements of the vectors:
$\overrightarrow{v}_{H1,2}$ defines the vector from P_1^H to P_2^H and
$\overrightarrow{v}_{H1,3}$ defines the vector from P_1^H to P_3^H
$\overrightarrow{v}_{L1,2}$ defines the vector from P_1^L to P_2^L and
$\overrightarrow{v}_{L1,3}$ 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 (10)

$$S = \frac{\|\overrightarrow{v}_{H1,2}\|}{\|\overrightarrow{v}_{L1,2}\|} = \frac{\|\overrightarrow{v}_{H1,3}\|}{\|\overrightarrow{v}_{L1,3}\|} \tag{10}$$

(3) Determine the angle of rotation θ between the low- and high-resolution images using Equation (11)

$$\theta = \cos^{-1} \frac{\|\overrightarrow{v}_{H1,2} * \overrightarrow{v}_{L1,2}\|}{\|\overrightarrow{v}_{H1,2}\| * \|\overrightarrow{v}_{L1,2}\|} \tag{11}$$

Once the scaling and rotation operations are performed on the basis of the scale factor and angle of rotation as determined in Equations (10) and (11), respectively, new reference points are computed for the new low-resolution image, where P_1^L, P_2^L and P_3^L become $P_1'^L$, $P_2'^L$ and $P_3'^L$, respectively.

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 Δx and Δy measurements found using Equation (12).

$$\Delta x = P_{1_x}^H - P_{1_x}'^L \text{ and } \Delta y = P_{1_y}^H - P_{1_y}'^L \tag{12}$$

When these affine transformations are applied to the original low-resolution image as shown earlier in Figure 4(a), the result is the image shown in Figure 5, which is still a low-resolution image but now has the same dimension as the high-resolution image of the Canon G6 Camera.

Figure 5. The 2-D converted low-resolution book spread

Bilinear interpolation, as the next step, allows 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 160×120 pixels to the 3072×2304 pixels of the Canon G6 Camera. The result of combining the low-resolution image and low-resolution map is shown in Figure 6.

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 7.

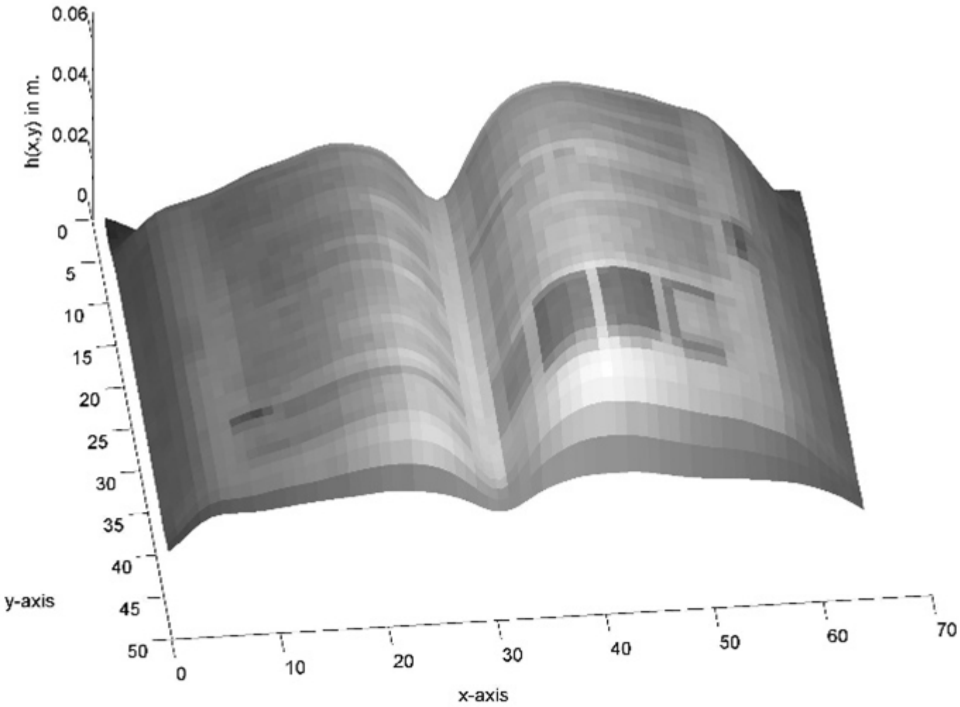


FIGURE 6. The 3-D full height map with the low-resolution image

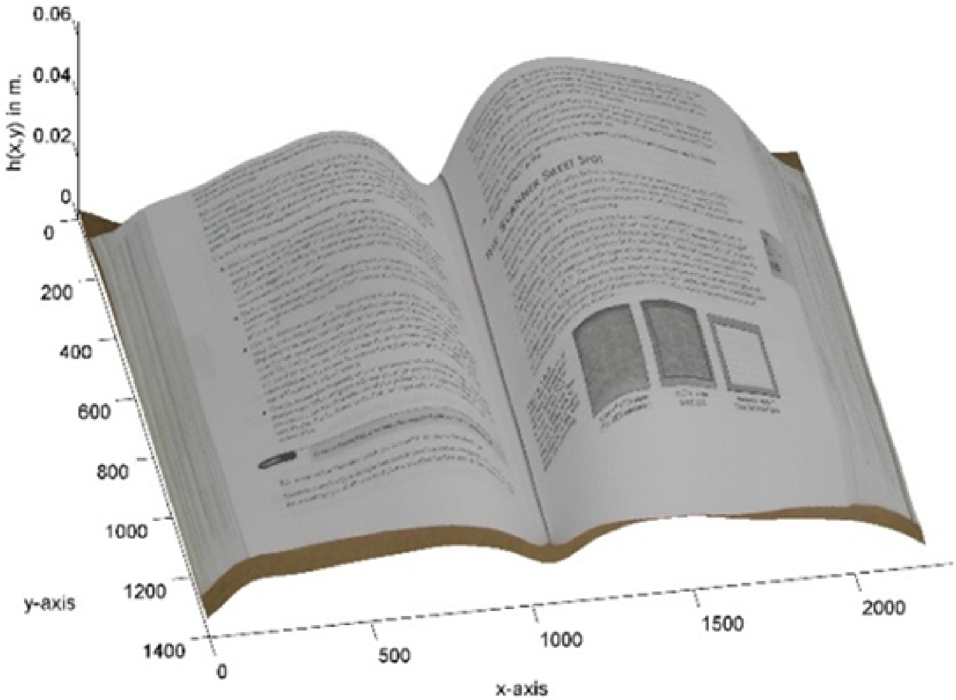


FIGURE 7. The 3-D higher resolution full height map with the high-resolution image. Note the high resolution of the text in comparison to full height map with the low-resolution image shown in Figure 6.

4. **Results.** This approach was tested using 200 pages of a book with their corresponding full height maps. The uniform height maps were then obtained, by replicating the center row of the height map (h) on all the other rows, as an alternative to the full height map to see if such an approximation is sufficient for the curvature correction process. Lastly, the book was physically rotated by 30 degrees clockwise for half of the 200 pages and then 30 degrees counterclockwise for the other half to test the book reader's resilience to such types of misalignments.

Once the images were corrected, they were fed into the latest and more robust 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 unprecedented large number of pages (200) that were used in this study.

The following results demonstrate the corrections made possible by the proposed algorithm using both uniform and full height maps on dewarping the pages of books or other bound documents. The character recognition accuracy of the uncorrected and corrected book spreads for all of the 200 pages at an average of about 3245 characters per page were then compared.

4.1. **Standard placement.** 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 8.

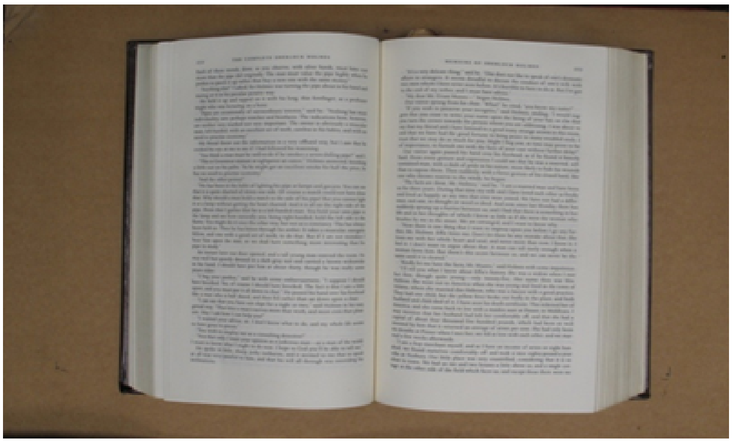
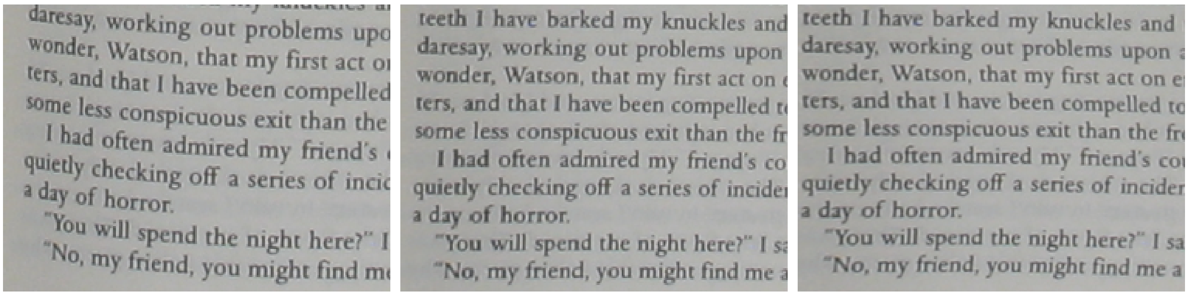


FIGURE 8. Standard alignment of the book spread

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 9(a). This arching effect occurs in the majority of the tested book spreads. Figure 9(b) shows when the result of the usual correction of the warped text using the uniform height map, while Figure 9(c) 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 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



(a) (b) (c)

FIGURE 9. Examples of the text correction process: (a) original warped text, (b) corrected text with uniform height map and (c) corrected text with full height map

TABLE 1. Summary of the standard positioned book spreads

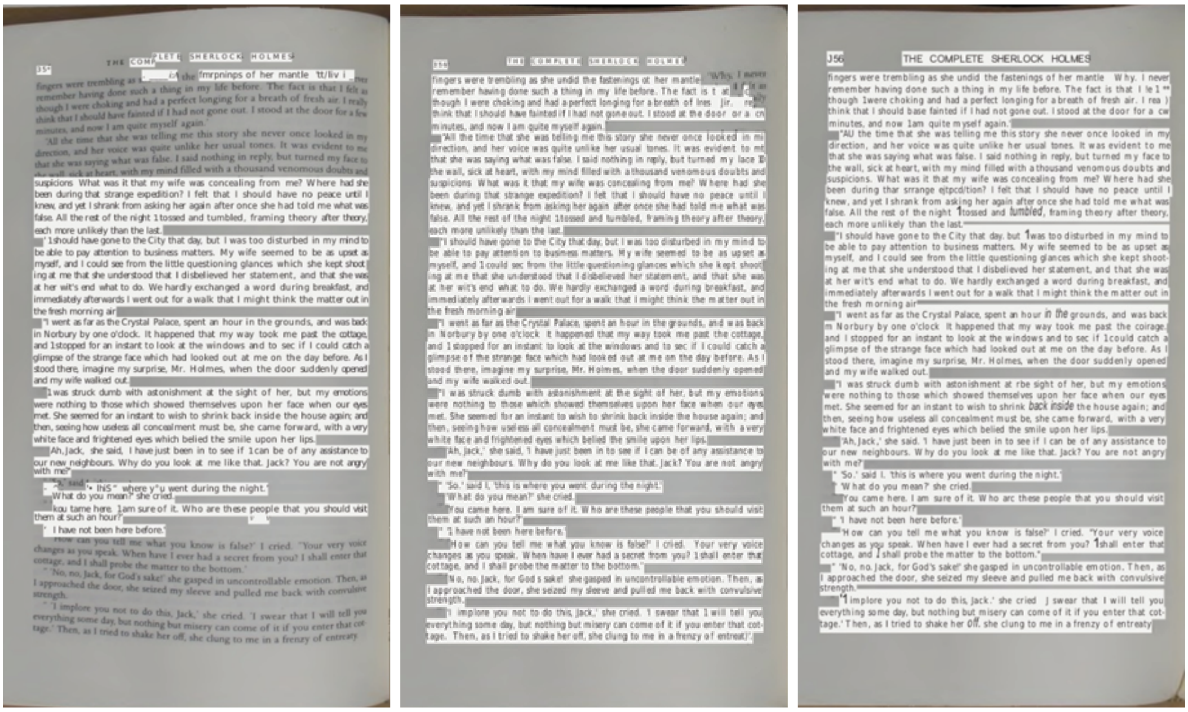
Procedure	Min accuracy	Max accuracy	Mean accuracy	VAR σ^2	STD σ	Median accuracy (after sort)
Original	56.99%	99.18%	91.57%	48.17	6.94	93.25%
Uniform height map	63.97%	99.55%	95.55%	15.74	3.97	96.10%
Full height map	68.15%	99.12%	96.11%	11.15	3.34	96.56%

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 [55]. Correction errors most certainly occur due to 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.

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 10(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 10(b). On the other hand, Figure 10(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 10(b).

4.2. **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 rotated by 30 degrees. In this experiment, half of the 200 pages were rotated 30 degrees clockwise and the other half 30 degrees counter clockwise; an example can be seen in Figure 11.

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 with the standard placement. The overall accuracy results are given in Table 2 which 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 were obtained when



(a) (b) (c)

FIGURE 10. Examples of standard placement results for: (a) original, (b) corrections with uniform map, and (c) corrections with full height map

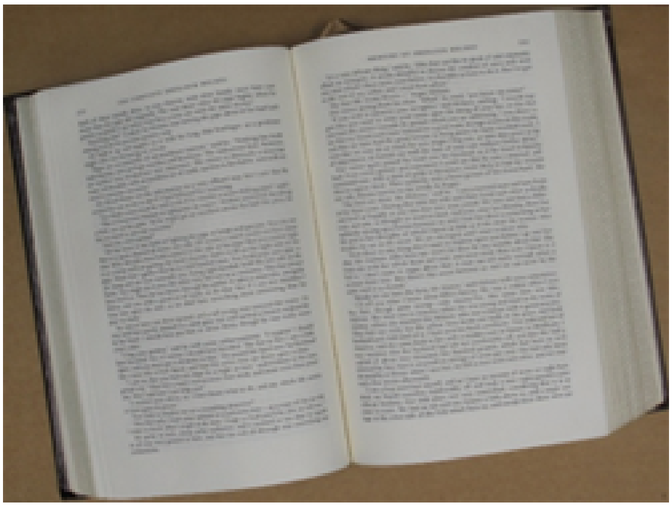


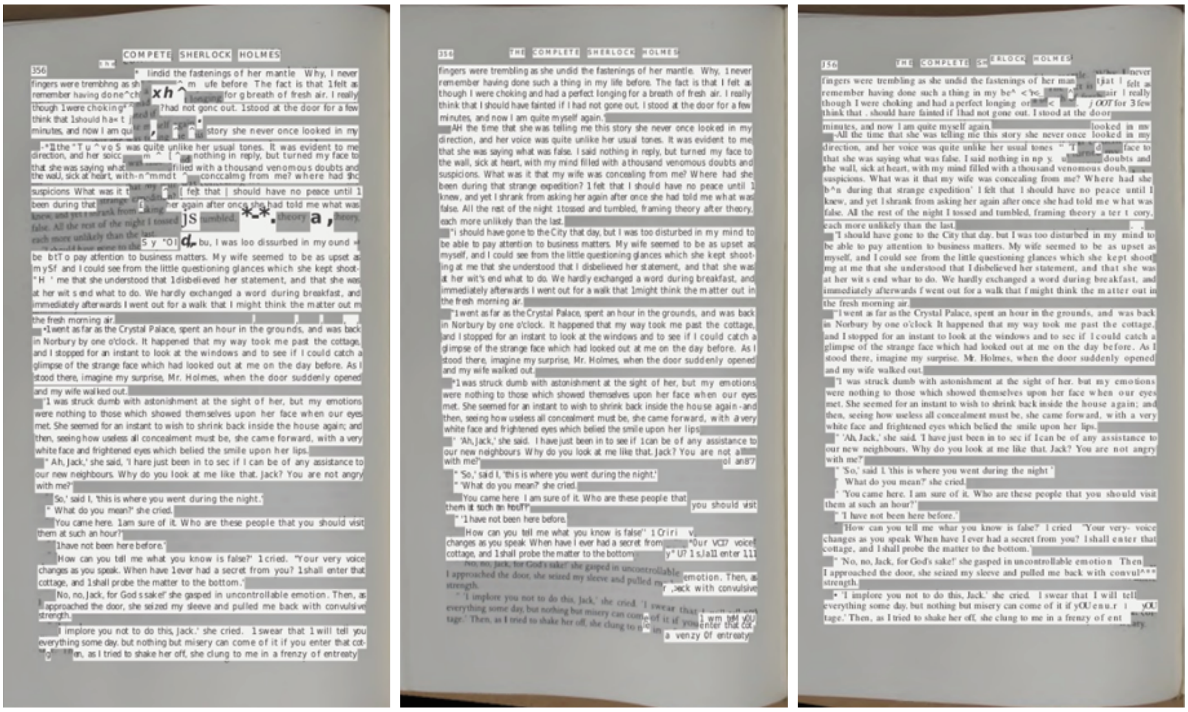
FIGURE 11. Rotated book spread

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.

An example of the effect of the disturbance cause by the rotation can be appreciated in Figure 12 where the rotation caused the OCR to recognize more character on the warp image as can be seen in Figure 12(a) than those in Figure 10(a), however, this does not always occur and yet there are still many characters which were not accurately recognized.

Table 2. Summary of the rotated positioned book spreads

Procedure	Min accuracy	Max accuracy	Mean accuracy	VAR σ^2	STD σ	Median accuracy (after sort)
Original	60.26%	99.51%	90.00%	84.82	9.21	95.74%
Uniform map	41.51%	99.03%	90.63%	54.88	7.41	91.56%
Full map	82.60%	99.37%	94.75%	16.88	4.11	95.74%



(a) (b) (c)

Figure 12. Examples of rotated placement results for: (a) original, (b) corrections with uniform map, and (c) corrections with full map

The correction with the uniform map led to an improvement on the recognition on the upper portion of the text in the page, however, the bottom part of the page did not fit that particular uniform map therefore was not properly correct and thus many of the characters were improperly or not recognized at all as shown in Figure 12(b). The full correction did improve the overall recognition of the characters as shown in Figure 12(c); however, this was not as high as the corrections on the standard placement image shown earlier in Figure 10(c).

The results shown in this section have certainly improved significantly from the group initial work as reported in [55]. 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 so that the low-resolution (160 $\times$ 120 pixels) depth information, as acquired by an Argos3D-P100 camera, accurately covers the entire book spread as captured by the high-resolution image (3072 $\times$ 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 in the testing phase. 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.

5. **Conclusion.** This study 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 account of 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 depth 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 depth maps. This database is accessible through the following website: cate-book3dmaps.fiu.edu.

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 depth 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.

Acknowledgments. This research is made possible thanks to the support of the National Science Foundation through grants CNS-1532061, CNS-1551221, CNS-1042341, and IIP 1338922. The support of the Ware Foundation is also greatly appreciated. This work is also supported through the graduate research fellowship (NSF-GRFP) granted to Mr. Harold Martin.

REFERENCES

[1] 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.
[2] Y. Tian, X. Yang and C. Yi, 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.
[3] B. Holton, ScanSnap SV600 overhead scanner from Fujitsu: Turning the page on text recognition, *AFB AccessWorld® Magazine*, vol.15, no.7, 2014.
[4] 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, *International Symposium on Visual Computing*, pp.735-746, 2016.
[5] S. Spooner, "What page, Miss?" Enhancing text accessibility with DAISY (digital accessible information system), *Journal of Visual Impairment & Blindness*, vol.108, no.3, pp.201-211, 2014.
[6] 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.
[7] A. M. Mulloy, C. Gevarter, M. Hopkins, K. S. Sutherland and S. T. Ramdoss, Assistive technology for students with visual impairments and blindnes, *Assistive Technologies for People with Diverse Abilities*, pp.113-156, 2014.
[8] 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.

[9] 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.3, pp.467-486, 2009.
[10] G. Annable, G. Goggin and D. Stienstra, Accessibility, disability, and inclusion in information technologies: Introduction, *Information Society*, vol.23, no.3, pp.145-147, 2007.
[11] J. Abascal, S. D. J. 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.
[12] 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.
[13] 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.
[14] C. Han, *Neural Network Based Off-line Handwritten Text Recognition System*, Ph.D. Thesis, Department of Electrical and Computer Engineering, Florida International University, 2011.
[15] C. Han, M. Adjouadi, A. Barreto, N. Rishe and J. Andrian, Improved pyramidal neural network for segmented handwritten characters recognition, *The 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV 09)*, vol.2, pp.695-699, 2009.
[16] 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.1, 2015.
[17] M. Adjouadi, E. Ruiz and L. Wang, Automated book reader for persons with blindness, *Computers Helping People with Special Needs*, vol.4061, pp.1094-1101, 2006.
[18] L. Wang and M. Adjouadi, Automated book reader design for persons with blindness, *Lecture Notes in Computer Science*, vol.5105, pp.318-325, 2008.
[19] P. Kakumanu, N. Bourbakis, J. Black and S. Panchanathan, Document image dewarping based on line estimation for visually impaired, *The 18th IEEE International Conference on Tools with Artificial Intelligence*, pp.625-631, 2006.
[20] F. Courteille, A. Crouzil, J. D. Durou et al., Shape from shading for the digitization of curved documents, *Machine Vision and Applications*, vol.18, pp.301-316, 2007.
[21] J. An, H. I. Koo and N. I. Cho, Rectification of planar targets using line segments, *Machine Vision and Applications*, vol.28, nos.1-2, pp.91-100, 2017.
[22] 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, *Computers Helping People with Special Needs*, vol.5105, pp.242-249, 2008.
[23] A. Calabrò, E. Contini and B. Leporini, Book4All: A tool to make an e-book more accessible to students with vision/visual-impairments, *HCI and Usability for e-Inclusion*, vol.5889, pp.236-248, 2009.
[24] P. P. Xiu and H. S. Baird, Whole-book recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.34, pp.2467-2480, 2012.
[25] B. Almasri, I. Elkabani and R. Zantout, An interactive workspace for helping the visually impaired learn linear algebra, *Computers Helping People with Special Needs*, vol.8547, pp.572-579, 2014.
[26] 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, 2015.
[27] 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.
[28] J. Liang, D. Doermann and H. Li, Camera-based analysis of text and documents: A survey, *International Journal of Document Analysis and Recognition (IJDAR)*, vol.7, nos.2-3, pp.84-104, 2005.
[29] A. Ulges, C. H. Lampert and T. M. Breuel, Document image dewarping using robust estimation of curled text lines, *The 8th International Conference on Document Analysis and Recognition*, vol.2, pp.1001-1005, 2005.
[30] V. Kluzner and A. Tzadok, Page curling correction for scanned books using local distortion information, *International Conference on Document Analysis and Recognition (ICDAR)*, vol.18, no.21, pp.890-894, 2011.
[31] J. Liang, D. DeMenthon and D. Doermann, Geometric rectification of camera-captured document images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.30, pp.591-605, 2008.
[32] S. Lu and C. L. Tan, Document flattening through grid modeling and regularization, *The 18th International Conference on Pattern Recognition*, vol.1, pp.971-974, 2006.

[33] C. L Tan, L. Zhang, Z. Zhang and T. Xia, Restoring warped document images through 3D shape modeling, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.28, pp.195-208, 2006.
[34] J. Liang, D. DeMenthon and D. Doermann, Flattening curved documents in images, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.2, nos.20-25, pp.338-345, 2005.
[35] H. K. Aghajan, B. H. Khalaj and T. Kailath, Estimation of skew angle in text-image analysis by SLIDE: Subspace-based line detection, *Machine Vis. Apps.*, vol.7, pp.267-276, 1994.
[36] L. Song, Y. Wu and B. Sun, A robust and fast dewarping method of document images, *International Conference on E-Product E-Service and E-Entertainment*, pp.1-4, 2010.
[37] N. Stamatopoulos, B. Gatos, I. Pratikakis and S. J. Perantonis, Goal-oriented rectification of camera-based document images, *IEEE Trans. Image Processing*, vol.20, pp.910-920, 2011.
[38] 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.
[39] K. Pal, M. Terras and T. Weyrich, Interactive exploration and flattening of deformed historical documents, *Computer Graphics Forum*, vol.32, pp.327-334, 2013.
[40] 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, nos.2-3, pp.147-158, 2005.
[41] G. F. Meng, C. H. Pan, S. M. Xiang, J. Y. Duan and N. N. Zheng, Metric rectification of curved document images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.34, pp.707-722, 2012.
[42] C. Kim, P. Chiu and S. Chandra, Dewarping book page spreads captured with a mobile phone camera, *Camera-Based Document Analysis and Recognition*, vol.8357, pp.101-112, 2014.
[43] M. Leo, G. Medioni, M. Trivedi, T. Kanade and G. M. Farinella, Computer vision for assistive technologies, *Computer Vision and Image Understanding*, vol.154, pp.1-15, 2017.
[44] M. S. Brown, M. X. Sun, R. G. Yang, L. Yun and W. B. Seales, Restoring 2D content from distorted documents, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.29, pp.1904-1916, 2007.
[45] L. Galarza, Z. Wang and M. Adjouadi, Book spread correction using a time of flight imaging sensor, *Proc. of the International Conference on Image Processing, Computer Vision, and Pattern Recognition*, 2014.
[46] L. Galarza, Z. Wang and M. Adjouadi, Book reader optimization using a time of flight imaging sensor, *Proc. of the International Conference on Image Processing, Computer Vision, and Pattern Recognition*, pp.324-330, 2015.
[47] L. Galarza, H. Martin and M. Adjouadi, 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*, pp.692-697, 2016.
[48] Y. Zhang and C. L. Tan, An improved physically-based method for geometric restoration of distorted document images, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.30, pp.728-734, 2008.
[49] S. Foix, G. Alenyà and C. Torras, *Exploitation of Time-of-Flight (ToF) Cameras*, Technical Report IRI-TR-10-07, Institut de Robòtica i Informàtica Industrial, 2010.
[50] C. Kim, H. Yu and G. Yang, Depth super resolution using bilateral filter, *The 4th International Congress on Image and Signal Processing*, pp.1067-1071, 2011.
[51] S. Schwarz, M. Sjostrom and R. Olsson, A weighted optimization approach to time-of-flight sensor fusion, *IEEE Trans. Image Processing*, vol.23, no.1, pp.214-225, 2014.
[52] D. Jiménez, D. Pizarro, M. Mazo and S. Palazuelos, Modeling and correction of multipath interference in time of flight cameras, *Image and Vision Computing*, vol.32, no.1, pp.1-13, 2014.
[53] T. Fersch, R. Weigel and A. Koelpin, A CDMA modulation technique for automotive time-of-flight LiDAR systems, *IEEE Sensors Journal*, vol.17, no.11, pp.3507-3516, 2017.
[54] R. Lussana, F. Villa, A. D. Mora, D. Contini, A. Farina, L. D. Sieno and F. Zappa, Non-contact inclusion detection in food through a single-photon time-of-flight imager, *IEEE Sensors Journal*, vol.17, no.1, pp.78-83, 2017.
[55] L. Wang, *An Automated Book Reader Design as an Assistive Technology Tool for Persons with Blindness*, Ph.D. Thesis, Department of Electrical and Computer Engineering, Florida International University, Miami, 2007.