https://doi.org/10.2352/issn.2168-3204.2022.19.1.01

Multispectral Scheimpflug: Imaging Degraded Books That Open Less Than 30 Degrees

Gregory Heyworth, Department of English, University of Rochester, Rochester, New York, USA
Keith T. Knox, Early Manuscripts Electronic Library, Rolling Hills Estates, California, USA
Kenneth Boydston, MegaVision, Goleta, California, USA
Yuhao Zhu, Department of Computer Science, University of Rochester, Rochester, New York, USA

Abstract

This paper presents an imaging system that reads texts from books that open less than 30 degrees (due to their fragile bindings) and whose paper quality is degraded. In particular, the system operates on the Scheimpflug principle to correct the geometric distortion necessarily introduced when imaging barely open books. We introduce the guiding principles behind such a system, discuss how it is calibrated and set up, present the results of imaging two dime novels from early twentieth century, both with fragile bindings, and discuss lessons we learned.

Motivation

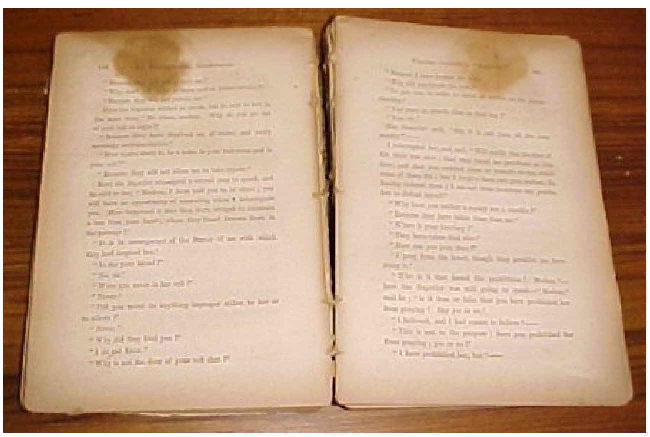


FIGURE 1. *Dime Novel, 1914. University of Rochester Collection.*

The purpose of the Multispectral Scheimpflug project was to devise a method for imaging fragile books and manuscripts that could not be opened more than thirty degrees and that remained illegible to scholars. For rare book and manuscript archives around the world, this problem of un-openability, for lack of a better term, affects between 3% and 8% of holdings on average, not only among medieval and Renaissance manuscripts but especially among modern books printed on, and bound in, pulp paper made between 1840 and 1930. In the United States alone, hundreds of thousands of books and manuscripts are currently inaccessible for this reason. In certain cases, the manuscript was from the Middle Ages, and the binding was from the 20th century, allowing one to send the book to a conservation lab to be rebound. However, if those same manuscripts had medieval or even 16th-century bindings, every effort would have been made to not only keep them intact but also preserve the binding.

Nor is the problem only that the texts do not open widely enough to permit reading, but that the poor-quality pulp paper has

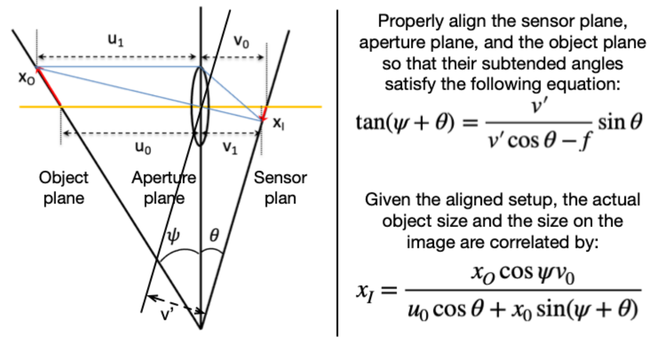


FIGURE 2. *Scheimpflug Principle.*

frequently degraded and become so severely that stained, even when open, the writing is often difficult to read. This problem affects handwritten books as well as marginalia, both of which are often central to scholarly research.

Figure 1 shows one such example. It is a dime novel published in 1914 and currently in the University of Rochester Collection. Notice the broken spine caused by forcefully opening book and dark brown stains that obscure the text.

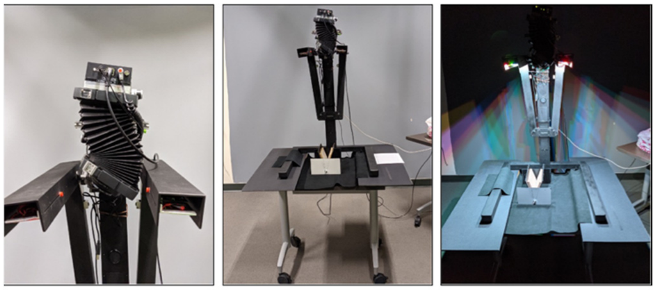
Our objective, therefore, was twofold: to image the book's contents safely without causing damage to the spine, while ensuring material legibility in the resulting images.

Approach

Our approach is to build a multispectral Scheimpflug imaging system. The system operates on the Scheimpflug principle [1] to reconstruct texts from barely open books. We also use multispectral imaging to read texts from stained and degraded papers. While the Scheimpflug principle was used in other fields such as climate science [2] [3] and extended depth-of-field photography [4], to our best knowledge, this is the first application of Scheimpflug principle for manuscript imaging. Our contribution is to demonstrate the feasibility and to describe the system design that can serve as the basis for similar imaging and digitization efforts in the future.

The challenge in imaging barely open books is that the object plane, where a to-be-imaged page lies on, is not parallel to the aperture plane (where the lens is); therefore, the object would not be in-focus if the sensor plane and the aperture plane were parallel, which is the case in normal photography.

We propose to address this issue using the Scheimpflug principle [1], as illustrated in Figure 2, which shows that the object plane, the aperture plane, and the sensor plane can be carefully arranged in such a way that an object on the object plane is imaged in-focus on the sensor plane. Critically, note that the sensor plane is now not parallel to the aperture plane.



The imaging camera whose lens can be tilted with respect to the sensor plane. | The copy stand. | The lighting system: a white light (from three LEDs) and 18 monochrome lights from UV (365nm) to IR (940nm).

FIGURE 3. *Imaging setup.*

This in turn requires us to have a camera where we can flexibly tilt the sensor plane. Figure 3 shows the imaging setup, where the left panel shows our imaging camera whose sensor plane can be tilted with respect to the lens. The exact angle between the sensor and the lens (θ) and the angle between the object plane and the lens (ψ) are governed by the first equation in Figure 2, where v' is the distance along the line of sight from the sensor plane to the center of the lens.

Given the aligned setup, we could also derive the relationship between the object size in the world space (x_O) and on the image (x_I), which allows us to correct the geometric distortion. The relationship is given by the second equation in Figure 2.

Figure 3 shows the imaging system with a conservation book cradle and a 50-megapixel monochrome tilt camera. The lighting system is a white light composed of three LED bands. To support multispectral imaging, we use, in addition to the white light, also 18 monochromatic lights from UV (365 nm) to IR (940 nm).

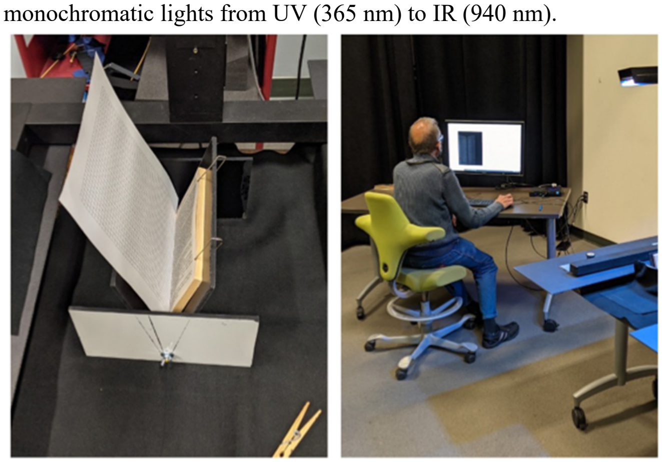


FIGURE 4. *Imaging graph paper, in place of the page, allows measurement of the calibration parameters.*

We carefully measured the books dimensions, which along with the camera inherent parameters (e.g., focal length) to determine the tilt of the sensor. We then imaged a grid paper to measure the geometric parameters, as shown in Figure 4, which allow us to correct the geometric distortion. Finally, we placed a Macbeth

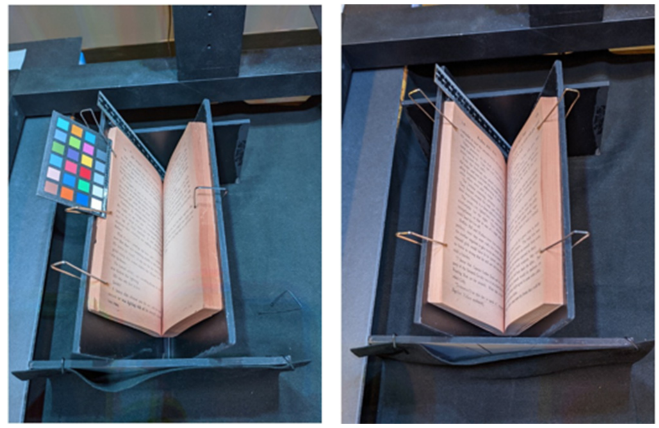


FIGURE 5. *The Macbeth ColorChecker chart is placed with the books for color calibration.*

ColorChecker chart in order to calibrate colors after capture, as shown in Figure 5.

Since the books are barely opened, we made extra effort to send lights to the books. Pointing the camera at the verso, we illuminated the recto so that light would reflect onto the opposite side diffusely all the way down to the gutter.

Results

FIGURE 6. *The two dime books and an imaging session.*

We used our imaging system to test two dime novels, *Where Peril Beckons* and *Buffalo Bill in the Land of Dread*, from the early twentieth century borrowed from the Department of Rare Books, Special Collections, and Preservation (RBSCP) at the University of Rochester. Dime novels are a form of middlebrow American literature that emerged in the 1860s. Produced weekly, they offer a large window into what topics were popular among readers, particularly young boys, at the time. The problem is much of RBSCP's collection is extremely fragile and difficult to use. And this is not unique to Rochester. The low-quality paper has made the copies that have survived the last hundred-plus years so brittle that to open them would be to pull them apart. That fragility made them perfect test subjects for an unprecedented imaging system.

Figure 6 shows a photo of an imaging session. Our initial tests were successful. The system and accompanying processing software we developed demonstrated their effectiveness in generating easily legible images.

Figure 7 shows an ortho-rectified black-and-white image that of the original that is in focus and legible into the gutter, a quality sufficient for optical character recognition to successfully render a digitized and searchable text from our images. On pages with stains

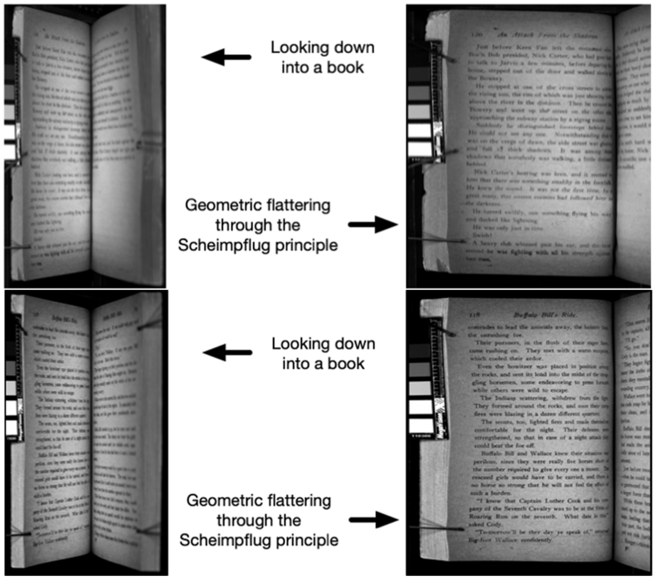


FIGURE 7. *Two dime books under imaging. Our system with the processing software can generate legible texts.*

and discolorations, we found that multispectral bands, especially the infrared, produced a clear black and white image of the obscured text without the need of image processing.

The color images we produced by combining 10 visible wavelengths initially generated images that were surprisingly colorimetrically inaccurate. Further tests revealed that early pulp paper, when imaged in wavelengths between 450 and 530 nanometers produces noticeable fluorescence in the red, causing our initial color images to have a pink tinge. Although more material research needs to be done on early pulp paper and the chemical used to produce it, we were able to measure that fluorescence with a spectrophotometer and compensate, generating highly color accurate images, adjustments we then wrote into the image processing software.

Our system is fast to set up. Adjusting the view camera to the correct angle was a process that became easier and faster with practice. To begin with, we spent 15 minutes setting the original angle, and another five adjusting it in the second shot. Once set, however, no more adjustments are necessary.

Our system also allows for fast imaging: the time it takes to position the book and complete a capture is short. Book positioning in cradle with Macbeth takes about 2 minutes, and image capture of sequence of 18 images in reflection from 365nm to 940nm takes about 3 minutes to finish. This shows that our system is feasible for batch imaging in the future.

Future Work

Our initial experiments have shown that the proposed system can reconstruct texts from barely open, degraded books. The results enabled us to better understand refinements to the system necessary to produce a more robust solution.

Hardware Setup

First, we will need to design a cradle that can accommodate larger manuscripts while being able to hold pages open without causing cockling

Second, the system must be able to be moved along the x-axis relative to the camera to accommodate changes in focal length that come with changes in the thickness of the book as one turns pages to image. For instance, having positioned the view cameras at the appropriate angle for Scheimpflug capture, we will mount the cradle on an adjustable translation arm on the x-axis. As we progress through the book, flipping through the pages, each side of the book will experience equivalent and opposite changes in focus that can be compensated for by moving the cradle in the appropriate direction along the x-axis.

Finally, to increase the efficiency of the workflow, an ideal system would involve two cameras capable of capturing recto and verso simultaneously. For instance, we could have two cameras with more modest sensors mounted on the copystand arm, each capturing one page.

Multispectral Imaging

Multispectral imaging, has on occasion, been used as a crude spectrophotometer to distinguish in a rough fashion among various types of ink. In general, however, we find that the use of MSI for material analysis is best for detection, not diagnosis or analysis.

MSI is, however, very good at producing accurate color. Although our experimental set-up pictured above employs 9-wavelengths to produce images with a sub-1 ΔE, an ideal workflow would reduce the number of bands to 5 – UV, IR, 630, 535, 450 – producing normal 3-band color. Note that depending on the object, the number of useful wavelengths will differ. A damaged medieval manuscript may require up to 28 bands in both reflectance and fluorescence. For printed books that are otherwise undamaged, if somewhat discolored or faded with age, we find that five bands usually suffice. The five bands we plan to use are UV and IR (the latter is ideal for removing stains and the former for increasing the contrast of faint or faded text) along with a red, green, and blue band to create a color image.

What we have noticed, however, is that early pulp paper is manufactured with chemicals that cause the paper to fluoresce in green light, producing a long red fluorescence at c. 680nm that alters the color of images. No doubt, this is a subtle problem that affects an entire corpus of books produced with the same methods and materials. Having noticed the shift, we corrected the color image by image. A future system will provide a batch-color correction option.

Conclusion

We demonstrated an imaging system, including the hardware setup and the post-processing software, that can reconstruct texts from barely open and degraded books. The idea is to use the Scheimpflug principle that is commonly used in conventional photography. We demonstrate a specific use-case of applying the Scheimpflug principle to cultural heritage.

References

[1] Scheimpflug, Theodor. “Improved Method and Apparatus for the Systematic Alteration or Distortion of Plane Pictures and Images by Means of Lenses and Mirrors for Photography and for other purposes.” GB Patent No. 1196 (1904).

[2] Mei, Liang, and Mikkel Brydegaard. “Atmospheric aerosol monitoring by an elastic Scheimpflug lidar system.” Optics Express 23.24 (2015).

[3] Brydegaard, Mikkel, and Sune Svanberg. “Photonic monitoring of atmospheric and aquatic fauna.” Laser & Photonics Reviews (2018).
[4] Nagahara, Hajime, et al. “Flexible depth of field photography.” European Conference on Computer Vision (2008).

Author Biography

Gregory Heyworth is Associate Professor of English, History, and Computer Science at the University of Rochester. With degrees from Columbia (BA, Comparative Literature), Cambridge (English BA, MA), and Princeton (Ph.D., Comparative Literature) Heyworth is a medievalist with expertise in Latin and vernacular philology, codicology, paleography, and textual editing.

Keith Knox is an imaging scientist affiliated with the Early Manuscripts Electronic Library. Knox received a B.S. in Electrical Engineering in 1970 and a Ph.D. in Optics in 1975, both from the University of Rochester in Rochester, NY. He has over 40 years of research experience in multispectral imaging of manuscripts.

Ken Boydston is president of MegaVision, which developed the first integrated commercial system for the multispectral imaging of cultural heritage treasures. MegaVision’s multispectral imaging systems have imaged many of the world’s most iconic cultural heritage treasures including the Dead Sea Scrolls, Magna Carta, and Declaration of Independence.

Yuhao Zhu (BS in Computer Science from Beihang University, 2010, and Ph.D. in Electrical and Computer Engineering from The University of Texas at Austin, 2017), is an Assistant Professor of Computer Science at University of Rochester. His work focuses on applications, algorithms, and systems for imaging and visual computing.