Data Collection and Image Processing System for Ancient Arabic Manuscripts

Somaya Al-Maadeed, Syed F. K. Peer, Nandhini Subramanian
Department of Computer Science & Engineering
Qatar University
Doha, Qatar
{s_alali, syed.peer, nandhini.s}@qu.edu.qa

***Abstract*—This paper presents a general-purpose data collection system that combines a DSLR camera with directional LED lamps in order to capture a large quantity of high-resolution manuscript images in such a way as to maximize the speed of data collection while minimizing time and the need for specialized equipment. By integrating custom image processing software, the captured document images are mapped to lie on a planar surface, thereby enabling the application of more sophisticated computer vision algorithms. For this purpose, we also introduce an optional binarization tool that allows researchers to perform basic image pre-processing to simplify later analysis. The hardware setup and software tools presented in this paper can be combined to yield a simple system capable of producing large image datasets for use in document analysis research projects.**

Keywords—document digitization, hardware design, image post-processing, image pre-processing, scanning software

I. INTRODUCTION

Making use of a reliable dataset of high-quality images is an essential part of the computer vision research workflow, particularly in the domain of historical document analysis. In addition to using pre-existing datasets, it may be necessary to build a new set of images for use in the research process by performing document digitization. However, current approaches to the digitization process require the use of highly specialized hardware and/or software which is unsuitable for use by non-experts. The net effect of these factors is that constructing new image datasets for use in computer vision research often becomes a complicated and tedious affair for the data collection team.

Our goal is to reduce both the time and effort associated with this process by presenting a relatively low-cost solution that combines conventional hardware and simple software tools to help researchers in producing large, high-resolution document datasets while reducing the effort and costs involved with data collection.

II. BACKGROUND

Reducing the overhead associated with the data collection process often requires a system that combines hardware- and software-based solutions in some way. Many of these systems present trade-offs between cost, ease-of-use, and output image quality.

Most hardware setups for document digitization require at least one piece of image capturing equipment (e.g. DSLR camera) and possibly additional equipment to facilitate document image capture. The complexity of these hardware solutions can range anywhere from simple, single-camera setups [1] to augmented camera systems [2] to advanced, specialized hardware designed specifically for document digitization purposes [3]. We will briefly go over some of the hardware solutions that have been proposed before.

A. Low-Cost Solutions

With regard to low-cost solutions, we could use a single, off-the-shelf digital camera for image capture. Although using only a single camera (with no additional modifications) helps to minimize the cost of data collection, using such simple equipment can give rise to optical distortion in the captured images. Most commonly, these distortions are caused by either the nature of the document being scanned (e.g. reflective ink, golden borders, curved pages, etc.) or due to the orientation of the camera, which can result in a skewed perspective of the digitized page [1]. One way to overcome these issues is to capture images from two different perspectives (using the same camera) and combine these to create a single 3D model of the page structure [1]. By doing so, we can digitize the document reliably without requiring any additional equipment. However, 3D document modeling often places certain restrictions on the original documents themselves, such as requiring them to be in a "naturally unfolded" state [1], which might be easy to satisfy with certain types of documents, like magazines and books with an intact spine, but may be impossible to satisfy for older documents (which have heavily damaged spines), such as those found in archives of ancient historical manuscripts.

B. Medium-Cost Solutions

In order to overcome this limitation, one proposed solution is to create a system capable of handling arbitrary deformations in the input document, independent of any assumptions regarding the underlying structure of the document pages [2]. This approach requires the use of a coupled camera-projector setup, such that the projector "sweeps" a moving line across the input page while a stationary camera observes the location and shape of the line as it moves across the page [2]. By observing the relative warping of the line as it moves across the deformed page, the system is capable of computing a depth map and creating a

978-1-5386-1459-4/18/$31.00 ©2018 IEEE

final 3D output model [2]. This approach allows us to capture and digitally "flatten" documents which have been subjected to a variety of deformations, as is often the case with older, more ancient historical manuscripts. Beyond the increased equipment costs for the additional projector, this solution suffers from reduced speed due to the "sweep line" approach which allows for the 3D reconstruction of page structure at the cost of additional capture and post-processing time compared to the previous single-camera, two-image solution [1][2].

C. High-Cost Solutions

In order to reduce the time required for image capture, while taking into account possible deformation of the input document, it is possible to make use of hardware solutions that have been specifically designed for historical document digitization. In particular, the thinkMOTION system is flexible enough to account for a wide variety of document types (books, journals, manuscripts, etc.) while providing high-resolution images [3]. By combining the use of a fixed V-shaped cradle with the presence of two DSLR cameras, each oriented to face an individual page, operators are able to quickly capture images from the input document, while minimizing damage to more fragile manuscripts [3]. Although this system provides a good balance between high-quality imaging and very fast data collection speed, it too suffers from some significant limitations. In particular, this specialized machine presents significant costs compared to other, less hardware-intensive approaches. Furthermore, the software interface for the thinkMOTION system is only capable of providing image capture control and relies on other third-party software tools to handle image processing [3].

III. Methodology

A. Hardware Design

In order to overcome the limitations of previous solutions, we have built a relatively easy-to-use hardware setup that is intended to speed up the image capture process while preserving the physical integrity of scanned documents.

The high-level design of the image capture assembly is illustrated in Fig. 1 and Fig. 2. A single DSLR camera is mounted to the boom of a fixed tripod whose height can be adjusted to change the FOV on a per-document basis. Furthermore, the mounted camera can be angled atop the tripod to manually adjust for image skew at the time of capture. By adjusting the working distance between the document and the camera and fine-tuning the angle of the camera itself, we are able to modify the setup as needed to capture a wide variety of document types.

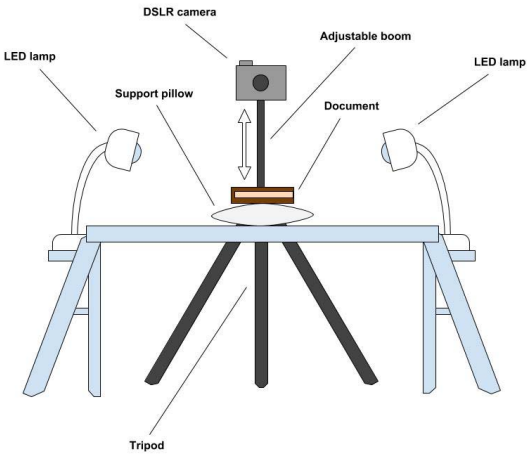

Fig 1. Frontal View of Hardware Setup

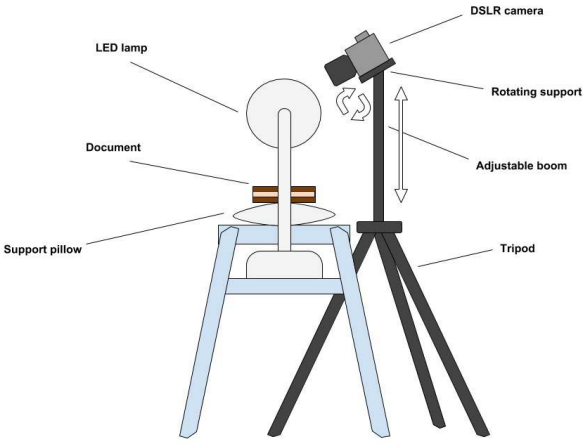

Fig 2. Side View of Hardware Setup

Instead of using the traditional V-shaped cradle, we opted to use a "support pillow" whose shape can be dynamically adjusted on a page-by-page basis. By doing so, we are able to further modify the setup to account for physical variation in the underlying document. Furthermore, with regards to the lighting equipment, we chose to use two LED lamps whose direction can be modified at the time of image capture, allowing for further fine-tuning of lighting conditions. In fact, the use of these LED lamps allows researchers to conduct the image capture process in any environment, irrespective of the presence (or absence) of external light sources.

Finally, the document itself is placed on top of the "support pillow", which is then placed atop a horizontal surface that is anchored at either end to fixed support structures, providing a stable working environment for the duration of the data collection process.

As for the imaging equipment itself, we chose to use a Nikon D5000 [4] equipped with a Nikkor AF-S DX [5] objective lens. The hardware specifications for both camera and objective lens can be found in Table I and Table II.

TABLE I. DSLR CAMERA SPECIFICATIONS

Image Sensor	
Type	23.6 mm x 15.8 mm CMOS
Effective Pixels	12.3 megapixels
Aspect Ratio	3:2
Focusing	
AF System / Points	11 AF Points
Selected AF Point Display	Superimposed on viewfinder and on LCD display
Shutter	
Type	Electronically-controlled vertical-travel focal plane
Speed	30 sec. - 1/4000 sec.
File Type	
Image Size	(L) 4288 x 2848, (M) 3216 x 2136, (S) 2144 x 1424
File Format	Compressed 12-bit NEF (RAW), JPEG (Baseline Compliant)
Storage Media	SD, SDHC
Interface	
Computer	Hi-speed USB
Other	HDMI, NTSC

TABLE II. OBJECTIVE LENS SPECIFICATIONS

Parameter	**Value**
Focal Length Range	18 mm - 55 mm
Minimum aperture	f/22
Maximum aperture	f/3.5 - 5.6
Optical construction	11 elements in 8 groups
Angle of view (minimum)	20° 50'
Angle of view (maximum)	76°
Minimum focus distance	0.28 m
Lens length	79 mm

B. *Software Tools*

The raw images captured by the system are transferred via SD card to the user's laptop, which is loaded with software to manipulate and post-process these files. In particular, we developed two Python programs which were used to automatically process the output files: one for renaming the raw files to follow a more convenient naming scheme and the other for performing basic image cropping, deskewing, and rotation operations. Given that these programs output full-color images, we also we developed a C++ program that provides researchers with the option of applying simple image processing operations, like binarization, on the resulting document images.

In order to speed up the process of image capture, we opted for a workflow where all even pages were captured in a single run, followed by all odd pages in a single run. After doing so, we obtained two folders, one containing images of even pages and the other containing images of odd pages.

In order to merge these two folders into a single folder containing the entire book, we developed a script that automatically converts the filenames, located in each "even" and "odd" sub-folder, from the camera's internal format to a custom filename format that follows an easy-to-read naming scheme.

Following this renaming stage, we processed the full set of document images using our Python-based image processing script, which allows the user to manually select the four corners of each page. After the user has made their selection, our program uses the 4-point perspective transform function given in [6] to compute (and apply) a homography matrix that maps the original four points to the four corners of a "flattened" rectangle, while also cropping and deskewing the input image. After this stage is complete, our program then allows the user to rotate the image into an upright orientation, using the "inbound rotation" function included in the `imutils` library [7]. The user interface for this program can be seen in Fig. 3 and Fig. 4.

After using this program, we obtain a complete set of full-color, "flattened" page images, like the one in Fig. 5, which can be used as input to more sophisticated computer vision algorithms.

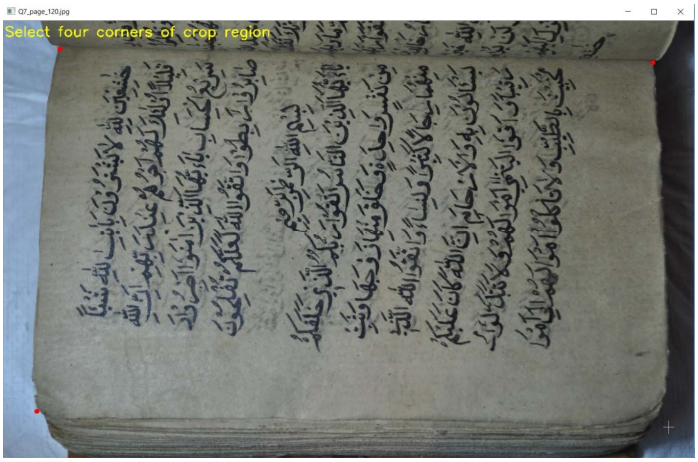

Fig 3. User Interface for Corner Selection

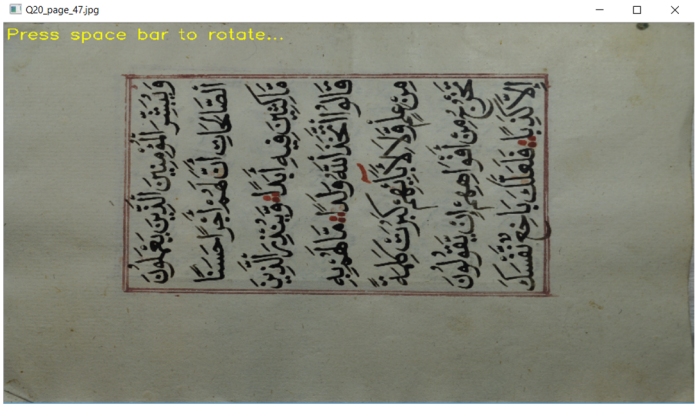

Fig 4. User Interface for Image Rotation

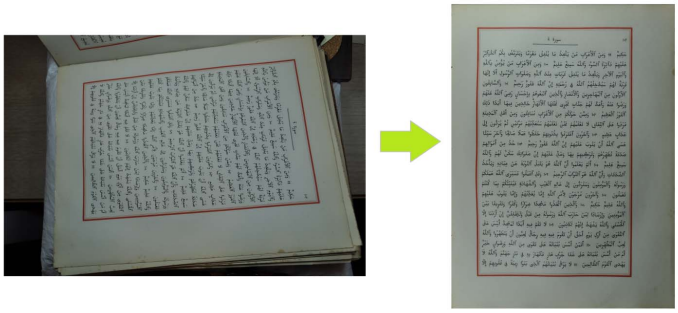

Fig 5. Results of Post-Processing

C. *Image Pre-Processing*

In order to facilitate the later application of computer vision algorithms, researchers can choose to make use of our C++-based program, which is capable of applying a variety of binarization algorithms. It is important to note that, before applying these algorithms, our program first converts each image to the grayscale colorspace to satisfy the input requirements for valid binarization. The user interface for accessing this functionality is given in Fig. 6.

The simplest binarization algorithm allows the user to specify a threshold value (t) which will be used as a cut-off to determine which input pixels (x_{ij}) should be rendered as black pixels and which should be rendered as white in the output image, as per the formula in Equation (1).

$$f(x_{ij}, t) = \begin{cases} 255 & \text{if } x_{ij} \geq t \\ 0 & \text{otherwise} \end{cases}$$

Equation 1. Basic Thresholding

The Otsu binarization algorithm [8] works by assuming that the pixels in the input grayscale image can be divided into background and foreground classes. Based on this assumption, the algorithm proceeds by iterating through (and applying) various threshold values, which range from 1 to 255, while keeping track of whichever threshold value yields the maximum variance between foreground and background classes. By doing so, the algorithm is able to automatically identify the optimal value for the threshold parameter (t) in Equation (1).

The K-means binarization algorithm is implemented such that it remaps each grayscale pixel from its current value to a new value based on the computed distance between each pixel value (x_{ij}) and two fixed reference values (v_1, v_2), as shown in Equation (2).

$$f(x_{ij}, v_1, v_2) = \begin{cases} v_1 & \text{if } |x_{ij} - v_1| \leq |x_{ij} - v_2| \\ v_2 & \text{otherwise} \end{cases}$$

Equation 2. K-Means Thresholding

The "combined" binarization algorithm, adapted from previous work in [9], combines the Otsu method, K-means thresholding, and various visual metrics to compute and apply the optimal threshold for binarization. An example of applying the Otsu binarization function is illustrated in Fig. 7.

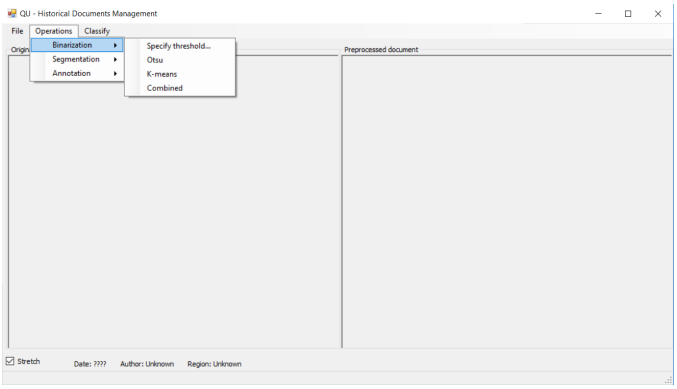

Fig 6. Binarization options in User Interface

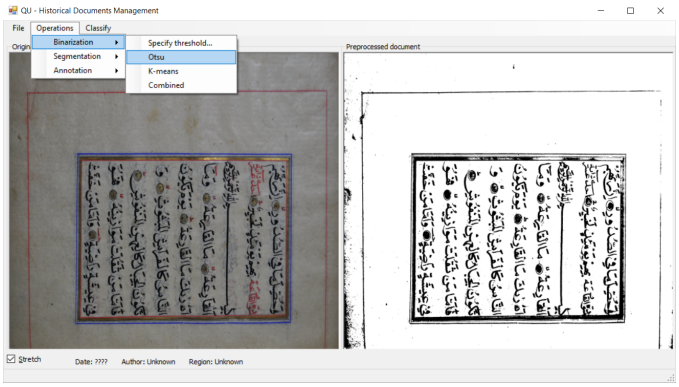

Fig 7. Applying Otsu Binarization to Image

IV. Comparison with Existing Systems

Although the proposed hardware setup shares some similarities with the single-camera solution presented in [1], our system is able to generate high-quality results while taking advantage of lower computational complexity. In particular, with regards to the "page flattening" process, the system proposed by Kim, Koo, and Cho relies upon the assumption that pages are in a "naturally unfolded" state [1], whereas our system makes no such assumption regarding the physical state of the document pages being captured. As such, our system is able to utilize a simplified deskewing algorithm as opposed to the more complicated and computationally intensive 3D page reconstruction method presented in [1].

Given that our system makes extensive use of off-the-shelf, conventional hardware components, we are also able to avoid the financial costs associated with the highly specialized solution suggested by Brown and Seales [2], which requires the use of an expensive coupled camera-projector setup. Although the setup in [2] may result in higher quality output images, the time and effort associated with calibrating and using such a system can cause significant delays in the image capture workflow, as opposed to a simpler single-camera setup.

Furthermore, unlike the solution presented by Ciupe, Lovasz, and Gruescu [3], our proposed system provides greater flexibility while granting the benefit of reduced equipment costs. In particular, although the system in [3] enables researchers to capture high-quality images across a wide variety of document types, the lack of built-in image processing functionality and the limited number of customization options ultimately restricts the kinds of research environments in which the thinkMOTION system can be used. By utilizing modifiable hardware components and providing essential image processing functionality, our proposed system is able to provide a more comprehensive data collection platform than that presented in [3].

V. Discussion & Future Work

The design of this system proves that it is possible to develop a low-cost solution that combines general-purpose hardware with custom software to deliver high-quality document digitization results. In addition to providing a concrete implementation of this system, our work provides a

template for the future construction of modular and highly customizable setups that can provide greater flexibility and ease-of-use over existing solutions, which often suffer from high implementation costs and are generally unsuitable for use by non-experts.

Although the overall system is currently functional, there are several aspects which can be improved upon or modified.

With respect to the physical hardware setup, the use of a "support pillow" allows for fine-tuned adjustments, but often places a strain on the user who needs to re-adjust the pillow every dozen pages or so. One potential solution would be to use an adjustable V-shaped cradle with glass platen, as mentioned in [3], which would maintain the benefits of an adjustable "support pillow" while providing a more stable support structure than is currently available.

On the software side, one major limitation is the lack of sophisticated user interface controls in the Python-based post-processing script. Although this program was originally designed for ease-of-use over rich functionality, future versions of this software will focus on providing additional features and interactivity, such as the ability to move or delete selected control points on the page, which would greatly benefit the quality and usability of the overall system.

In addition to integrating these improvements, further work is currently under way to augment the proposed system with the ability to capture documents in the infrared spectrum, thereby providing researchers with the opportunity to identify and explore salient features which may be difficult to discern using traditional DSLR imaging. Finally, future work will focus on enhancing the proposed system by including more sophisticated document pre-processing algorithms, which will help in increasing the legibility of ancient handwritten manuscripts, as well as developing a framework for evaluating the overall quality of the proposed document digitization system.

At this point in time, the document imaging results obtained from the proposed system are actively being used as the basis of a computer vision platform for automatic document analysis. The full dataset and system will be made available to researchers in the field of Arabic handwriting recognition beside our other available datasets [9]-[11].

VI. Conclusion

The proposed hardware- and software-based system is capable of handling the full data collection workflow, from image capture and post-processing to (optional) output binarization. By combining readily available, cost-effective hardware components, we were able to create a highly configurable system that is capable of digitizing large sets of ancient documents, while presenting lower costs than other, more specialized solutions. Furthermore, we developed various user-friendly software tools to streamline the conversion of raw images into high-quality datasets that can be used as input to more sophisticated computer vision algorithms.

Acknowledgment

This work was conducted as part of the NPRP-7-442-1-082 project. The authors are particularly grateful to the staff at the Museum of Islamic Art (MIA) in providing certain components of the hardware setup and for allowing access to their ancient historical manuscript archives.

References

[1] J. Kim, H. Il Koo and N. Ik Cho, "Camera-based document digitization using multiple images", in *International Conference on Image Processing* (*ICIP*), San Diego, California, 2008, pp. 1025-1028.

[2] M. Brown and W. Seales, "Document restoration using 3D shape: a general deskewing algorithm for arbitrarily warped documents", in *International Conference on Computer Vision* (*ICCV*), Vancouver, Canada, 2002, pp. 367-374.

[3] V. Ciupe, E. Lovasz and C. Gruescu, "High quality document digitization equipment", *Applied Mechanics and Materials*, vol. 162, pp. 589-596, 2012.

[4] "D5000 from Nikon", *Nikon Americas | USA*. [Online]. Available: https://www.nikonusa.com/en/nikon-products/product-archive/dslr-cameras/d5000.html#tab-ProductDetail-ProductTabs-TechSpecs.

[5] "AF-S DX Zoom-Nikkor ED 18-55mm F3.5-5.6G from Nikon", *Nikon Americas | USA*. [Online]. Available: https://www.nikonusa.com/en/nikon-products/product/camera-lenses/af-s-dx-zoom-nikkor-ed-18-55mm-f3.5-5.6g.html#tab-ProductDetail-ProductTabs-TechSpecs.

[6] A. Rosebrock, "4 point OpenCV getPerspective transform example", *PyImageSearch*, 2014. [Online]. Available: http://www.pyimagesearch.com/2014/08/25/4-point-opencv-getperspective-transform-example/.

[7] A. Rosebrock, *imutils*. Github, 2018. [Online]. Available: https://github.com/jrosebr1/imutils/

[8] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, 1979.

[9] S. Al Maadeed, W. Ayouby, A. Hassaine, and J. Aljaam, "QUWI: an Arabic and English handwriting dataset for offline writer identification", in *International Conference on Frontiers in Handwriting Recognition* (*ICFHR*), Bari, Italy, 2012, pp. 746-751.

[10] S. Al-Ma'adeed, D. Elliman, and C. Higgins, "A data base for Arabic handwritten text recognition research", *The International Arab Journal of Information Technology*, vol. 1, no. 1, pp. 117-221, 2004.

[11] A. Hassaine and S. Al Maadeed, "ICFHR2012 competition on writer identification - challenge 2: Arabic scripts", in *International Conference on Frontiers in Handwriting Recognition* (*ICFHR*), Bari, Italy, 2012, pp. 835-840.