Capture and Dewarping of Page Spreads with a Handheld Compact 3D Camera

Michael P. Cutter
University of California at Santa Cruz
Baskin School of Engineering (Computer Engineering department)
Santa Cruz, CA, USA
mcutter@soe.ucsc.edu

Patrick Chiu
FX Palo Alto Laboratory
Palo Alto, CA, USA
chiu@fxpal.com

***Abstract*—This paper describes a system for capturing images of books with a handheld 3D stereo camera, which performs dewarping to produce images that are flattened. A Fujifilm consumer grade 3D camera provides a highly mobile and low-cost 3D capture device. Applying standard computer vision algorithms, camera calibration is performed, the captured images are stereo rectified, and the depth information is computed by block matching. Due to technical limitations, the resulting point cloud has defects such as splotches and noise, which make it hard to recover the precise 3D locations of the points on the book pages. We address this problem by computing curve profiles of the depth map and using them to build a cylinder model of the pages. We then employ meshes to facilitate the flattening and rendering of the cylinder model in virtual space. We have implemented a prototype of the system and report on a preliminary evaluation based on measuring the straightness of resulting text lines.**

***Keywords*-document capture, document analysis, dewarping, 3D camera, stereo camera**

I. Introduction

It is desirable to be able to quickly and easily produce a digital copy of a page spread of a book, whether for note taking and collaboration or for archival purposes. The traditional method for doing this involves physically pressing the page spread against a flatbed scanner. Camera based document capture offers advantages over conventional flatbed scanning. A non-exhaustive list includes ease of capture, reduction in damage to the document, and the ubiquitousness of digital cameras. However, with these benefits there are also challenges. Documents that are captured by scanners are flattened almost to the point where geometric and perspective distortions become negligible. For camera captured documents, it is necessary to post-process the images to digitally correct for geometric and perspective distortions.

Our work addresses creating a system that uses a handheld stereo camera and processes the page spread images to compensate for perspective and geometric distortions. The capture device is a low-cost consumer grade compact 3D camera which takes two images simultaneously from the left and right lenses. Since these images are captured from a fixed pair of lenses, dense depth information can be calculated by standard block matching algorithms. However, the depth information obtained is not entirely accurate. To address this, we have developed algorithms and techniques based on constructing a cylinder model with depth profile curves which are robust to noisy depth information.

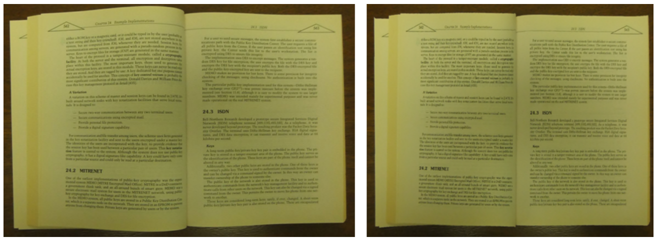


Figure 1. Before and after our dewarping technique. Both of these images were cropped to fit the column width.

Camera based document dewarping techniques can be split into monocular, structured light, and stereo solutions. Unlike monucular systems, by using dense block matching, our system is able to dewarp any type of content independent of reading order, layout, and language. It can dewarp any content as long as there is sufficient texture on the page to fit a page surface model to. Unlike other stereo systems ours is designed for dewarping images that are captured without extra equipment such as a tripod or a structured light apparatus.

II. Related work

For non-3D capture and dewarping of book page images, the 3D information is typically computed by detecting curved text lines or other content features, and the dewarping is based on using various models including the cylinder model; see [1], [2]. In the Document Image Dewarping Contest at CBDAR 2007 [3], the cylinder model by [4] performed the best (though the difference was not statistically significant).

For 3D capture, there are two primary approaches that have been studied: structured light and multiple cameras. Structured light can sense highly accurate 3D information; the drawback is that it requires more complicated apparatus. An example of a structured light system is [5].

The multiple camera (including stereo camera) approach can be done with a simpler system, but it is not as robust because it depends on the page areas containing visible text content or texture. An example of a system employing two separate cameras is the Decapod [6]; dewarping functionality is not yet available in this system.

Some dewarping methods that have been used with structured light are energy minimization [5] and conformal mapping [7]. These methods require highly accurate 3D information and are difficult to use with low-cost consumer grade stereo cameras.

Our approach is to use a cylinder model, which has worked well in non-3D capture (as discussed above), but to generate the model using stereo captured depth information without requiring the use of text line or other content analysis. This also enables the method to be applied to pages that have mostly figures and pictures.

A related cylinder model approach for using two camera images is explored in a system developed by [8]. The camera positions are not fixed, so it is more complicated than a stereo system with fixed cameras, which can be calibrated. We point out these differences:

1) Their system uses feature points and bundle adjustment for computing the 3D data, whereas we use camera calibration and block matching.
2) Their system performs curve fitting as an optimization problem on all the points, whereas we compute a depth map with imperfections and compute two profiles to model the cylinder. Both of our systems use polynomials to model the cylinder shaped curves.
3) Their system uses hi-end DSLR cameras with large sensors, and we use compact consumer grade stereo camera with small sensors.

III. System and Methods

A. System description

The 3D camera used in our work is the Fujifilm FinePix W3 (see Figure 2). Our system intends to create a reproducable dewarping framework that is high-speed and robust. Therefore, we utilize as much well tested and optimized open-source computer vision code that is available. OpenCV [9] provides many routines that are suitable for camera calibration and stereo block matching and foreground-background segmentation. For modeling the depth profile curves and performing dewarping, we developed several modules in Java. Figure 3 shows an overview of the system.

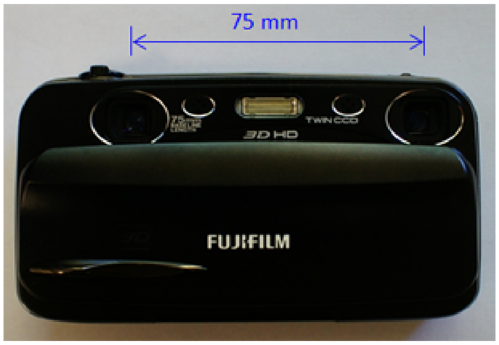


Figure 2. Fujifilm FinePix W3 3D camera

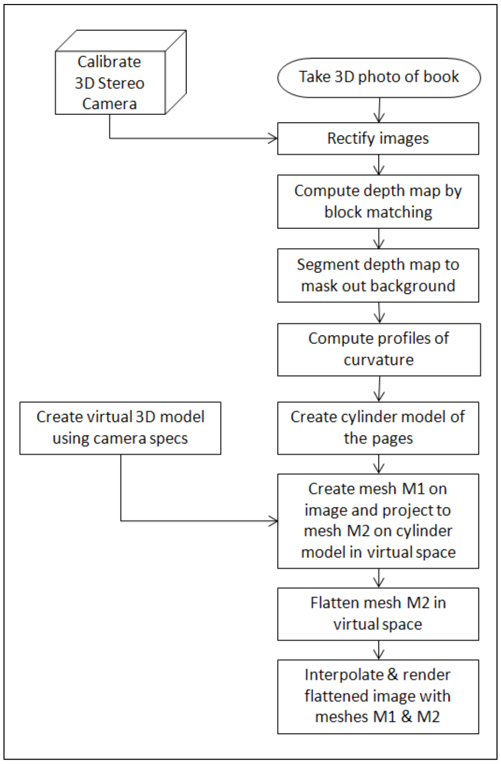

Calibrate 3D Stereo Camera
Take 3D photo of book
Rectify images
Compute depth map by block matching
Segment depth map to mask out background
Compute profiles of curvature
Create virtual 3D model using camera specs
Create cylinder model of the pages
Create mesh M1 on image and project to mesh M2 on cylinder model in virtual space
Flatten mesh M2 in virtual space
Interpolate & render flattened image with meshes M1 & M2

Figure 3. Method pipeline

B. Calibration

The first step is to compute each camera's intrinsic properties and then the entire stereo systems' extrinsic properties. The intrinsic properties allow for correction of tangential and radial distortion for each individual camera. The extrinsic properties are used to compute the homography that rectifies the images so the epipolar geometry of the image pair is aligned. Once aligned and undistorted the image pair is ready to be block matched.

C. Block matching

For block matching we chose Semi Global Block Matching routine implemented in OpenCV. The result of block matching is a disparity image (see Figure 4) which captures the depth or z information of the scene. This disparity image can be reprojected into a 3D point cloud (see Figure 6) and can be triangulated by Delaunay triangulation.

D. Segmentation

To perform the segmentation necessary to isolate the location of document pages within the depth map image, we apply the GrabCut algorithm [10]. This routine is available through OpenCV since version 2.1. The input to GrabCut is

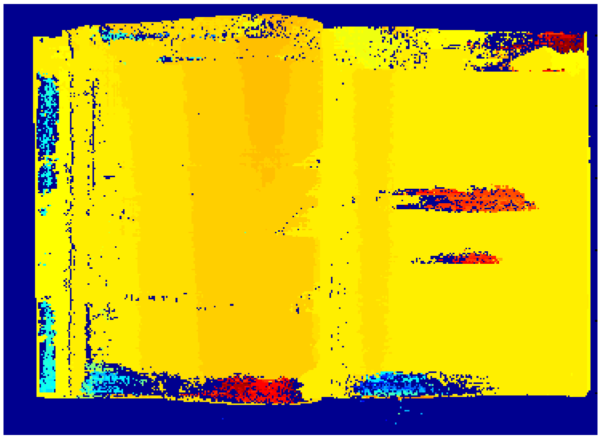


Figure 4. The disparity (depth) image calculated from left and right rectified stereo pair prior to any smoothing. It can be deduced from this image that this page spread is curved more on the left side.

a matrix of equal size as the image with one of four values in each entry: background, likely-background, foreground, and likely-foreground.

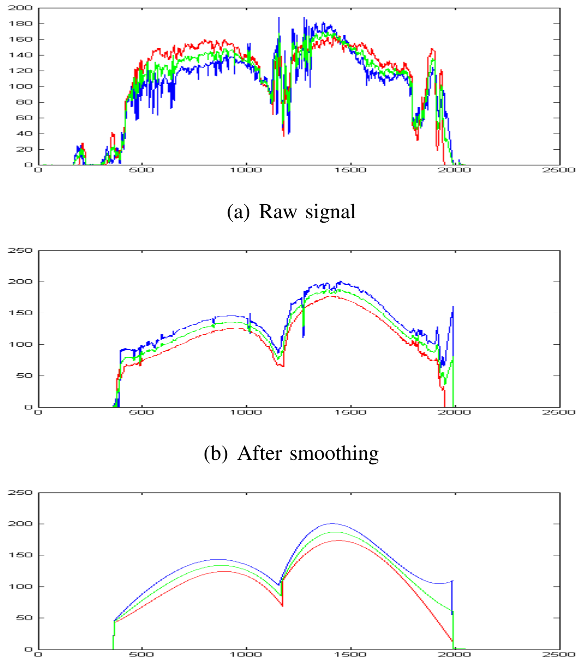
The following capture assumptions are used to estimate these values: The user is taking an image of a book within the capture frame, and the background forms at least a small border in both images. Therefore, all the pixels around the edge of the image can be labeled as likely background and all the pixels near the center of the image can be labeled as foreground.

E. Depth profiles

Depth profiles can be extracted once a segmented disparity image similar to Figure 4 has been computed. To compute the curve profiles of the depth map, the average depth is computed for each column of the depth map. The profiles are then modeled by detecting the location of the book spine and fitting cubic polynomials to the profiles on each side of the spine. The cylinder model is constructed by computing two profiles at the top and bottom halves of the depth map (see Figure 7). These are assigned y positions at $\frac{1}{4}$ an $\frac{3}{4}$ of the image height, respectively.

A triangulated point cloud of a depth map is visualized in Figure 6.

Due to the large amount of noise in the disparity image,[1] it is necessary to smooth the column averages prior to polynomial fitting. The first step is to perform basic smoothing by removing values that fall two standard deviations away from the mean. Then a low degree polynomial will suffice to accurately model the curvature. Visually these steps can be seen in Figure 5.



(a) Raw signal

(b) After smoothing

(c) After polynomial fitting

Figure 5. Depth profiles before and after smoothing

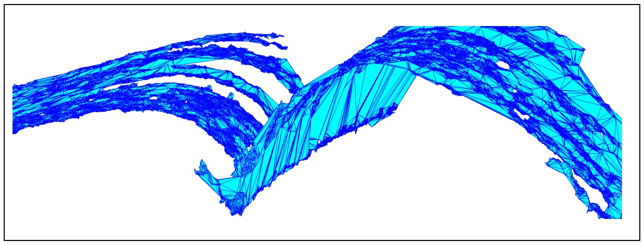


Figure 6. Triangulated point cloud

F. Dewarping using a cylinder model with curve profiles

As shown in the point cloud in Figure 6, the 3D information obtained from the pipeline using the standard algorithms in `OpenCV` described above does not provide sufficiently clean or accurate data to reconstruct the page directly. We deal with this by using a cylinder model to parameterize the paper pages, which can be considered as rigid ruled surfaces. The high level idea is to extract two curve profiles from the depth map (described above), and the depth value of a point $P'_{ic}(x'_{ic}, y'_{ic})$ on the rectified photo can be interpolated by the depth values at $(x'_{ic}, \frac{1}{4} \cdot \text{height})$ and $(x'_{ic}, \frac{3}{4} \cdot \text{height})$ on the two curve profiles. An additional refinement is to estimate the slight angle correction caused by the tilt of the camera. With the interpolated depth value z, we can determine the point $P(x, y, z)$ on the cylindrical surface in world coordinates. This is achieved by using a camera model based on the hardware specifications of the camera.

The camera model is illustrated in Figure 8(a). From the

[1] By inspecting the triangulated point cloud in Figure 6, holes still exist even after the points have been fused by triangulation, which motivates a robust solution based on smoothing and polynomial fitting.

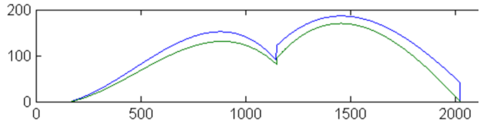
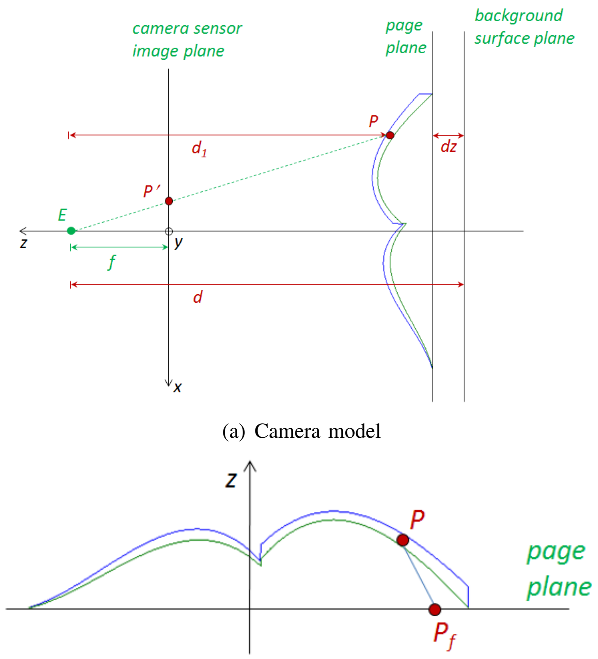


Figure 7. The height profiles of a book

FinePix W3 camera specifications [11], the sensor width and height, and the focal length f can be obtained. The focal length can also be extracted from the EXIF data of the photo. There are two parameters that depend on the context when taking a photo of a book. One is the distance d from the camera to the background surface plane (e.g. a table) and the other is the thickness dz of the book between the background and the pages being captured. Since it may be inconvenient for the user to measure these during the capture step, these parameters can be estimated and adjusted during processing.



(a) Camera model

(b) Flattening the surface based on arclength

Figure 8.

From the camera model, given a 2D point $P'_{ic}(x'_{ic}, y'_{ic})$ in the rectified photo image coordinates and the computed depth value z, we can compute the corresponding 3D point $P(x, y, z)$ in the virtual world space of the camera model as follows. The photo image coordinates are converted to image sensor world coordinates $P'_{wc}(x'_{wc}, y'_{wc})$, and the depth value is also converted to world coordinates z_{wc}. Using trigonometry on the illustration in Figure 8(a), we determine the point $P(x, y, z)$ on the cylindrical surface in world coordinates.

Next, the cylindrical surface can be flattened based on the arclength to obtain a point $P_f(x_f, y_f, z_f)$ on the page plane. See Figure 8(b). More precisely, we compute the arclength of the curve profile from $x = 0$ to P, and this arclength will be the magnitude of x_f . Finally, P_f can be converted to a point P_{fic} in the output image coordinates by scaling the page plane coordinates appropriately.

Once we know how to convert and transform one point P'_{ic} to P_{fic}, we can set up a rectangular mesh to facilitate the transformation of the whole image. Meshes are also used in [4] for non-stereo dewarping. Each mesh point from the rectified input photo image is mapped to a point on the flattened output image, and the points inside the sub-rectangles can be interpolated from the corresponding mesh rectangles. A close up example can be seen in Figure 9.

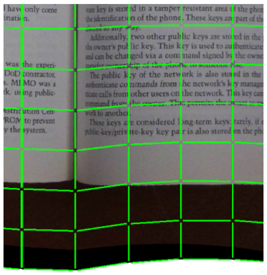


Figure 9. Close up of flattened image with mesh overlay

IV. Preliminary Evaluation

Despite that our dewarping technique has no dependence on any specific text information, printed text-lines have useful consistent properties which make it informative to study. Additionally, text will inevitably be encountered while dewarping document images. Before a line of text is warped by page curl, in theory it is completely straight. Therefore our method of evaluation is to measure how straight the before and after dewarped text lines are.

A. Metric

The straightness of a text-line is determined by first fitting a regression to all the y coordinates of every centroid of each connected components in a text-line. See Figure 10. Evaluating the straightness is more effective with a linear regression than a horizontal mean fit because stereo rectification introduced a slight rotation to the dewarped images. The slope of the regression compensates for this rotation.

The straightness standard error is measured by the standard deviation of the absolute difference of each centroid from the linear regression.

$$\text{Straightness}_{\text{SE}} = \frac{\text{std}(\vec{y} - \vec{r})}{\sqrt{n}}.$$

Where n is the number of centroids, $\vec{y}$ is vector of the centroid of the connected components y-coordinate, $\vec{r}$ is the vector of the regression fit and std() stands for the standard deviation function.

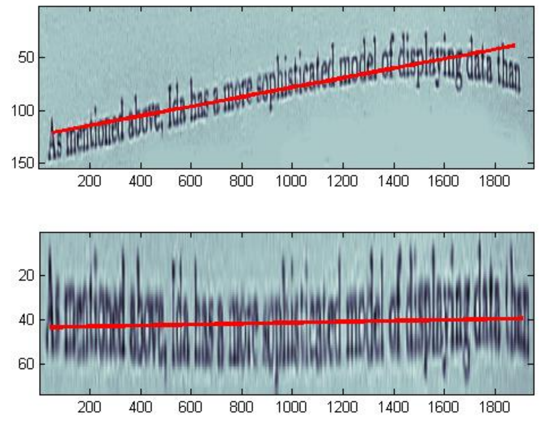


Figure 10. Above: text-line before dewarping. Below: text-line after dewarping. The line through each image represents the linear fit.

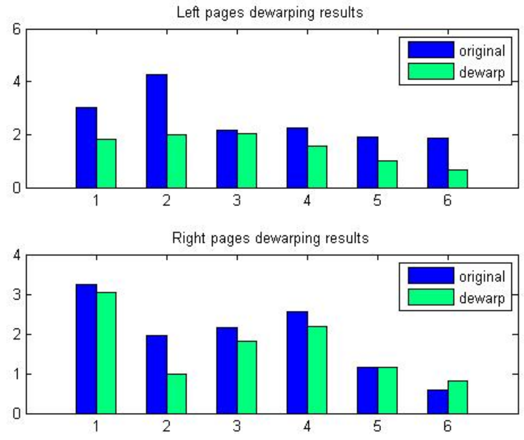


Figure 11. The graph above reports the aggregated straightness error prior and post dewarping for the left and right sides of the book images.

B. *Sample page spreads*

The data for this preliminary evaluation are six images of book page spreads captured with a consumer grade handheld 3D camera (see Figure 2). The six images are of several books with varying types of paper, in various backgrounds and lighting.

Straightness is measured on four representative lines of text on each page spread. These four text-lines are chosen in a systematic and simple way: top-left, bottom-left, top-right, and bottom-right. Since the images are captured with a handheld camera, which is almost never perpendicular to the page spread, there is a difference between the top and bottom of the page due to perspective distortion. Testing on these text-lines provides an indication of the effectiveness of our dewarping technique.

C. *Results*

Initial evaluation of these images shows that our dewarping method reduces straightness standard error in almost all cases (see Figure 11). Documents with greater page curl on the left or right side can be dewarped more producing a greater reduction in straightness error.

V. Conclusion and future work

We presented a system for using a handheld compact 3D camera to capture and dewarp page spreads. Our preliminary evaluation shows that this system is able to successfully dewarp these page spread images. Next steps for future work include evaluation of the system based on OCR and on visual appearance.

Acknowledgment

This work was done at FX Palo Alto Laboratory. We thank Don Kimber for helpful discussions.

References

[1] H. Cao, X. Ding, and C. Liu, "Rectifying the bound document image captured by the camera: A model based approach," in *Proc. ICDAR 2003*, pp. 71–75.

[2] J. Liang, D. DeMenthon, and D. Doermann, "Geometric rectification of camera-captured document images," *IEEE TPAMI*, vol. 30, pp. 591–605, April 2008.

[3] F. Shafait and T. Breuel, "Document image dewarping contest," in *Proc. CBDAR 2007*, pp. 181–188.

[4] B. Fu, M. Wu, R. Li, W. Li, and X. Xu, "A model-based book dewarping method using text line detection," in *Proc. CBDAR 2007*, pp. 63–70.

[5] M. S. Brown and W. B. Seales, "Image restoration of arbitrarily warped documents," *IEEE TPAMI*, vol. 26, pp. 1295–1306, October 2004.

[6] F. Shafait, M. Cutter, J. van Beusekom, S. Bukhari, and T. Breuel, "Decapod: A flexible, low cost digitization solution for small and medium archives," in *Proc. CBDAR 2011*, pp. 41–46.

[7] M. S. Brown and C. J. Pisula, "Conformal deskewing of non-planar documents," in *Proc. CVPR 2005*, pp. 998–1004.

[8] H. I. Koo, J. Kim, and N. I. Cho, "Composition of a dewarped and enhanced document image from two view images," *IEEE Trans. Image Processing*, vol. 18, pp. 1551–1562, July 2009.

[9] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[10] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, pp. 309–314, August 2004.

[11] Fujifilm finepix w3 specifications. [Online]. Available: http://www.fujifilm.com/products/3d/camera/finepix_real3dw3/specifications/