Browsing Through Closed Books: Fully Automatic Book Page Extraction from a 3-D X-ray CT Volume

Daniel Stromer, Vincent Christlein and Andreas Maier
Pattern Recognition Lab, Friedrich-Alexander-University
Erlangen-Nuremberg, Germany
Email: daniel.stromer@fau.de, vincent.christlein@fau.de,
andreas.maier@fau.de

Tobias Schoen and Wolfgang Holub
Fraunhofer Dev. Center for X-ray Technology (EZRT),
Fuerth, Germany
Email: tobias.schoen@iis.fraunhofer.de,
wolfgang.holub@iis.fraunhofer.de

***Abstract*—When digitizing or investigating historical documents, it is often the case that a document can not be opened, page-turned or touched anymore. Damages such as moisture or fire and aging processes disallow browsing through a book. To address these particular cases, our earlier work showed that Micro-CT X-ray scanners are able to image documents written with iron gall ink. A self-made book consisting of ten hand written pages was scanned and investigated without opening or page-turning. However, when analyzing the reconstruction results, we faced the problem of a proper automatic page segmentation and 2-D mapping within the volume in an acceptable time without losing information of the writings. The main problem is that the pages can be arbitrary deformed or squeezed together. In this paper, we present a fully automatic algorithm for the segmentation and extraction of book pages from the original 3-D volume. Our method delivers high quality results for our book model and can be easily adapted to other imaging modalities. We show that it performs well even for an extreme case with low resolution input data and wavy pages. To keep it simple for users, our algorithm works without any need of prior information or user interactions.**

***Index Terms*—Historical Document Analysis; 3-D X-ray CT Reconstruction; Automatic Page Extraction**

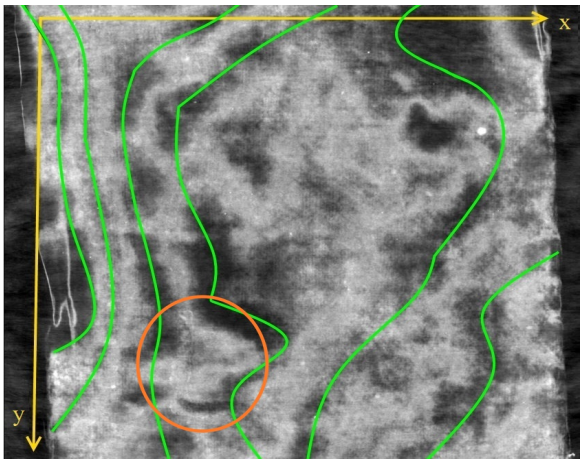


Fig. 1. Original 3-D volume center slice acquired by a X-ray CT scan. The green lines denote the air gaps between two neighboring pages whereas the orange circle denotes a part of a letter.

I. Introduction

The standard procedure of using a book scan robot [1] is not applicable for imaging of many historical documents. The documents may have a moisture, fire or aging damage that prohibits turning, opening or even touching them. However, some methods exist for imaging and reconstructing such documents. Our work concentrates on writings made of historical iron gall ink [2], an ink that has been used since the 5th century until present [3]. Not only Mozart used it for writing down his music but also the 'Declaration of Independence' as well as many other important documents were written with it.

X-ray phase contrast imaging (Synchrotron radiation) [4] is one approach for imaging such documents delivering a high quality output volume. However, this method is complex, expensive and immobile. Another approach is based on Terahertz radiation [5] which has the drawback of a limited depth resolution such that only a few pages can be imaged at once. A third method is using X-ray CT systems, such as it is used for non-destructive material testing or medical imaging [6], [7]. The scanners are mobile and cheap compared to the Synchrotron system as well as the field-of-view is only limited to the detector size. This approach is based on the material characteristics of the used ink. Since iron gall ink consists of metallic particles, its absorption is higher than papers' giving the ability to read the writings without opening the book itself. After acquiring the volume by a 3-D X-ray CT scan, a major problem lies in the visualization and separation of the pages for inspecting the writings or drawings on the paper. This is also mentioned in a related work of Albertin et al [8], [9]. Not only that the rather thin pages are stacked together in a very small area (sometimes more than 100 pages per cm) and squeezed together at many points, but they are also wavy and the writings have a higher absorption than the paper. Without a proper volume processing, it is hardly possible to investigate the delivered output. Fig. 1 shows the center slice of our generated volume where the green lines denote air gaps between two neighboring pages and the orange circle showing a part of a letter. Seals et al. showed that using the Structure Tensor performs well for scrolls [10] but it has not been tested on book structures yet where pages are squeezed together and not rolled up. Additionally, a drawback of this method is that the user needs to manually set initial seed points along the scroll and correct them if the calculation

does not provide a valid estimate of the local surface normal. In an earlier work, we already presented an approach where the user had to set two points on a single page resulting in an extraction of only this selected page [11]. The algorithm of this work is based on this algorithm.
Our contributions in the paper are:

1) We present a fully automatic algorithm with low computational time capable of extracting the entire pages of a book without the need of any user interaction or prior knowledge. The pages are mapped to 2-D such that the writings can be investigated by the naked eye.
2) We adopt an algorithm used in medical imaging to our case of the book pages. Taking a look at the sectional view of the volume's xz-layer, as shown for example in Fig. 4, reveals that the pages look like small vessels. This characteristic is utilized by replacing the Structure Tensor with the Vesselness filter defined by Frangi [12], [13]. The method is normally used in medical imaging to segment blood vessels by modeling the vessel as a curved surface.
3) We show that our approach is efficient by applying the developed algorithm to an extreme case. Our scanned book has small paper thicknesses down to $2-3\,\mathrm{pixel}$ per page, wavy pages, overlaps as well as fully and partially ink-saturated paper. As we used a pressboard holder to squeeze the pages, the noise level and artifacts were increased.

It should be further mentioned that the identification of the writings itself is not part of this work. However, we present a way of post-processing the 2-D mapped output to reduce noise and artifacts while simultaneously increasing the edge informations of the writings.

II. Materials and Methods

A. The Book Model

The used book consists of ten pages of handmade paper, each with a thickness of about $200\,\mu\mathrm{m}$. Letters of different font sizes, line thicknesses and paper saturations have been written on the pages with a quill. The height of the letters range from $1\,\mathrm{cm}$ to $2.5\,\mathrm{cm}$. To enable a scan, the pages were put in a pressboard holder. Fig. 7(a) and Fig. 7(b) show two pages on which we will concentrate for the following investigations.

B. CT Scan Trajectory

For this work, we used a micro-CT system consisting of a X-ray source and a detector. For the scan, an X-ray energy of $70\,\mathrm{kVp}$ and a detector pixel shift of $74\,\mu\mathrm{m}$ was configured. The object is placed centered on a turntable. A $360°$ circular scan is driven with 2400 projections and an exposure time of $1\,\mathrm{s}$ using cone-beam geometry. The book model was positioned upright as shown in Fig. 2 and the y-axis denotes the axis of rotation.

C. 3-D Reconstruction

For the reconstruction, the commonly used FDK algorithm [14] consisting of three major steps – cosine weighting, ramp filtering and back projection – was applied. With the large number of projections used, the algorithm should be accurate enough. The voxel size of the generated volume was $64\,\mu\mathrm{m}^3$ resulting in a mean page thickness of about $3-4\,\mathrm{pixel}$ which is sufficient for the presented extraction algorithm. With the used parameters, the scan and reconstruction time was kept relatively low (about 1.5 hours in total).

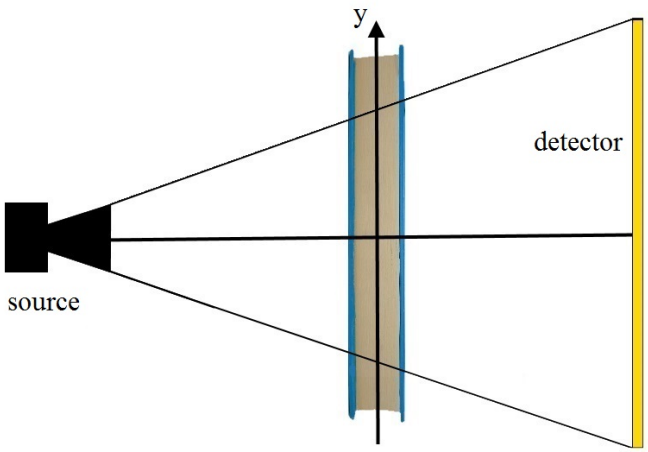


Fig. 2. Experimental setup for the 3-D CT scan. The rotation axis is the y-axis of the book.

D. Volume Processing Pipeline

All calculations were performed with the CONRAD framework for cone beam geometries [15]. The volume processing pipeline exists of three major steps – binarization, page segmentation and texturing – which will be discussed in the following sections. Fig. 3 demonstrates the entire process steps after the 3-D X-ray volume has been acquired. The binarization process is denoted by blue boxes, the page segmentation by green and the texturing and final output by orange boxes.

1) Volume Binarization: As a first step, a 2-D guided filter [16] resulting in $\boldsymbol{V_{gf}}$ (guidance equals input image, $r = 0.5$, $\epsilon = 0.01$), is applied on each slice of the acquired X-ray 3-D volume $\boldsymbol{V}$ to enhance the gaps between pages such that a better separation between air and page is possible while simultaneously denoising flat patches. Afterwards, a 2-D Vesselness filter is applied on the xz-layer (Fig. 4) with i denoting a specific slice with K number of slices. The original Vesselness approach performs three subsequent steps to calculate a resulting probability map of the vessels. The first step is to calculate multiple Hessian matrices of each slice with different scales of the Gaussian filter $G(x, s)$. We can simplify this approach with the prior knowledge that the diameters of the pages are approximately constant such that the initial guided filtering is sufficient. This reduces the original multi-scale approach to a single-scale approach. For this case, the 2-D Hessian matrix is calculated by Eq. (1), where I denotes the guided filtered xz-slice image. To suppress noise, we apply a 2-D anisotropic Gaussian filter (Eq. (2)) on every matrix element with $\sigma_x = n \cdot \sigma_y$ with $(n \geq 2)$ since we know

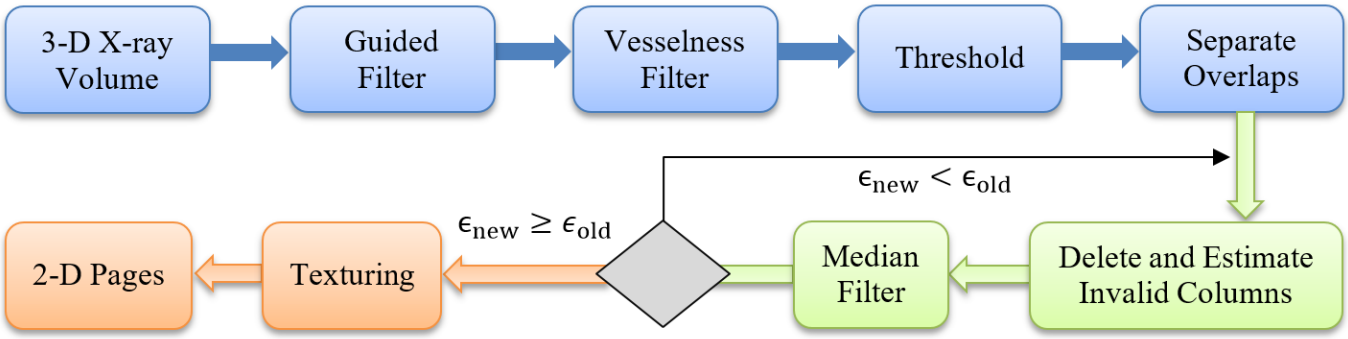


Fig. 3. Complete volume processing pipeline for automatic page extraction.

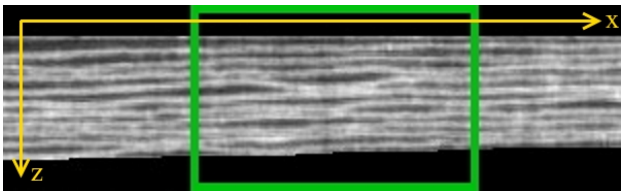


Fig. 4. Sectional view of the volume's xz-layer showing the ten pages.

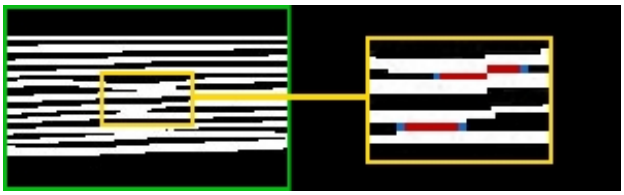


Fig. 5. The left part shows the snippet of the xz-slice in Fig. 4 after applying the Guided and Vesselness filter. The right part shows the page separation after applying the proposed algorithm. The red line is the new air gap and the blue dots are the touching points of the two pages verifying the gap.

that the page is supposed to move along the x-axis and is rather thin in y-direction.

$$H_i(I_i) = \begin{bmatrix} \frac{\delta^2}{\delta x^2} I_i & \frac{\delta}{\delta x \delta y} I_i \\ \frac{\delta}{\delta x \delta y} I_i & \frac{\delta^2}{\delta y^2} I_i \end{bmatrix} \text{ for } i = 1, ..., K. \quad (1)$$

$$G_{\sigma_x,\sigma_y} = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)\right) \quad (2)$$

The 2-D Vesselness measures are defined by Eq. (3) and Eq. (4), where λ_1, and λ_2 denote the descending eigenvalues of the Hessian matrix. These eigenvalues are used to describe the structure of the image at a certain area as they indicate the principal directions of the image's second order structure leading to the smallest curvature along the page. The structure can be plate-like, blob-like or tubular-like. Eq. (3) is the blobness measure and is close to 0 in case of a page. S will become larger since at least one of the eigenvalues will be large.

$$R_B = \frac{\lambda_2}{\lambda_1} \quad (3)$$

$$S = \sqrt{\lambda_1^2 + \lambda_2^2} \quad (4)$$

A good page (or vessel) in 2-D is defined by evaluating Eq. (5).

$$\boldsymbol{V_{ves}} = \begin{cases} 0, & \text{if } \lambda_1 \approx 0 \\ \exp\left(-\frac{R_B^2}{2\beta^2}\right)\left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right), & \text{if } \lambda_1 > 0 \end{cases} \quad (5)$$

$\boldsymbol{V_{ves}}$ is the final probability map for the pages where β and c are control parameters ($\beta = 0.5$ and $c = 4$). A small median filter ($radius = 1$) is applied to remove outliers. Next, the binarization is performed. Therefore Eq. (6) is evaluated for each pixel p in I. The pixel in the binarized slice I_{bin} is set to 1 if the Vesselness measure I_{ves} is greater than 0.1 and the value in the original slice I is greater than 30 % of the maximal gray value.

$$\boldsymbol{I_{bin}(p)} = \begin{cases} 1, & \text{if } I_{ves}(p) > 0.1 \wedge I(p) > (0.3 \cdot 255) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The following step is used to separate the pages automatically. The mean page thickness t_m is estimated by iterating through all xz-slices while simultaneously estimating the mean number of pages N. Afterwards, the pages are separated column-wise. If the thickness of a page is greater than 125 % of the mean thickness, the page is split in the center. If the page thickness is greater than $2 \cdot t_m$, the page is split twice at 33 % and 66 %. To improve this method, the split is only applied if two pages intersect in the neighborhood of the new line of intersection. The approach is visualized in the right part of Fig. 5 showing the page after applying the algorithm. The page is separated in the center because the pages intersect in the neighborhood (blue points) of the newly generated line of intersection (red line).

2) Page Segmentation: The page segmentation consists of two major steps that are repeated until the stop criterion is reached. First, for each xz-slice of the binarized volume, every column where the counted pages are different to N is saved in a list. This results in K lists of corrupted columns. The total number of corrupted columns ϵ is saved. Next, for each entry in a list, the nearest slice is searched where the specific column was identified correctly. The column in the corrupted volume is replaced by the found column. This step is repeated

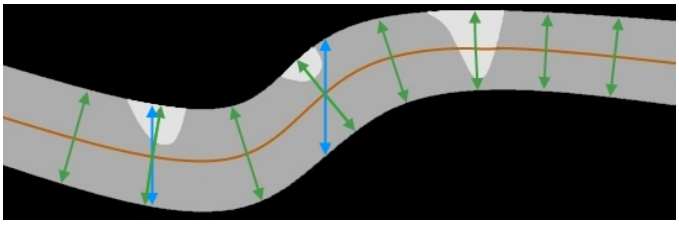


Fig. 6. Texturing along the centerline of the page. The normal (green) on the centerline (orange) is calculated for each point p. The maximal value along the normal of the centerline is considered as resulting intensity value at p. The black area is air, the gray area is the page and white is ink. The standard procedure of straight filtering along the z-axis could lead to error-prone mapping which is denoted by the blue arrows.

for every list entry and all slices K. The second step lies in the application of a 2-D median filter on the volume in $(x, y)-$direction with a radius of 5. This smooths the pages and removes outliers. The corrupted columns correction and median filtering steps are repeated until the newly calculated ϵ_m is greater or equal than the previously calculated ϵ_{m-1} where m denotes the number of iterations:

$$\epsilon_m - \epsilon_{m-1} \geq 0 \longrightarrow \text{stop} \tag{7}$$

3) Texturing: The texturing step takes the approximated pages from the page segmentation as input. With a standard straight maximum value filter along the z-dimension of the volume, wavy pages are not sampled properly. Our approach takes the separated pages, extracts them and calculates their specific centerlines for each slice K. Next, the normal at each point p of the centerline is calculated. We sample along the normal, save the maximal value since ink appears brighter than paper in the X-ray volume and set it as intensity at p of the 2-D page. Fig. 6 shows an example for the improvement with the proposed method for a small part of the page. The blue arrows denote the insufficient standard procedure, where the maximal value is searched along the z-direction. The left ink sport is found (left blue arrow), because the page is flat at that image part but as the page gets curved the ink spot is missed (right blue arrow). Sampling along the normals (green arrows) of the centerline (orange line) hits the ink spots even in the wavy page part. This process is repeated for all pages leading to the 2-D mapped result of the pages, where the writings can now be investigated.

E. Page Post-processing

The 2-D page output tends to be noisy and has streak artifacts which is caused by the relatively low used X-ray energy. To counter this we applied a Guided filter. This enhances the writings' edges and smooths the flat patches. Additionally, unsharp masking [17] is applied to sharpen the image and further increase the contrast at the intensity transitions.

III. Results

The acquired X-ray volume has a size of $1050 \times 980 \times 90$ voxel. The calculation times for all steps of the processing pipeline is shown in Table I. Thresholding and the overlap segmentation are the fastest steps with computation times lower than $1\,\text{second}$. Also the texturing is quite fast with $0.92\,\text{seconds}$ per page. Guided filtering (pre- and post-processing) and the Vesselness filter are still below $1\,\text{minute}$. The page segmentation process took the longest computation time of approximately $4\,\text{minutes}$. This is caused by the number of performed iterations. In total, 24 iterations were performed until the new number of corrupted columns ϵ_m was greater than the old number ϵ_{m-1}. This results in approximately $10\,\text{seconds}$ per iteration. When looking at Table II one can obtain that ϵ does not change a lot after iteration 10. To further reduce the computation time, we set a threshold α and defined that the algorithm should stop when the difference between the new and old number of corrupted columns is below that threshold:

$$\epsilon_m - \epsilon_{m-1} + \alpha \geq 0 \longrightarrow \text{stop} \tag{8}$$

After setting a threshold of 50 columns, the page segmentation stopped after 10 iterations with a computation time of $107.05\,\text{seconds}$. The results of the two segmentations were compared to each other having a $f1$-score of 0.996 meaning that the segmentations differ only slightly. Finally, the 2-D mapped Page 1 and Page 7 are shown in Fig. 7(c) and Fig. 7(d) and the post-processed versions in Fig. 7(e) and Fig. 7(e). Within this work we only showed two exemplary pages out of ten, however, all ten pages where extracted properly.

TABLE I

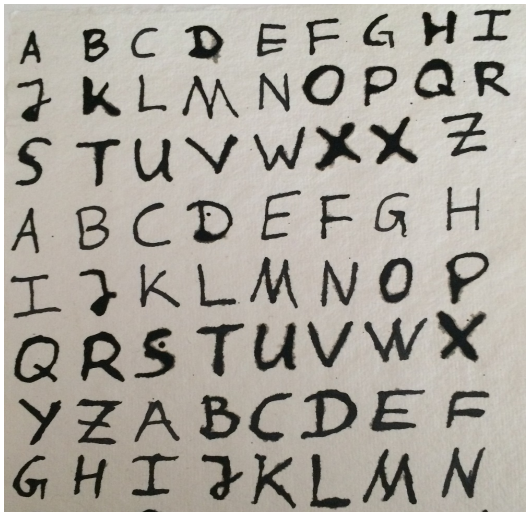
Calculation times for the pipeline processing steps. The page segmentation with 24 iterations has the longest computation time. The complete process of the automatic page extraction took approximately 6 minutes.

Step	**Time** $[s]$
Guided Filter	26.83
Vesselness Filter	55.16
Threshold	0.50
Overlap Separation	0.37
Page Segmentation	243.71
Texturing	9.22
Post-processing	3.82
Total	339.61

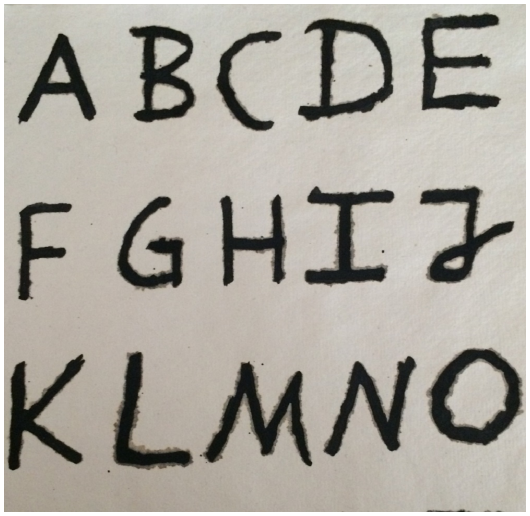
TABLE II

Number of corrupted columns ϵ after m iterations and percentage of the total number. After 10 iterations the output differs only slightly.

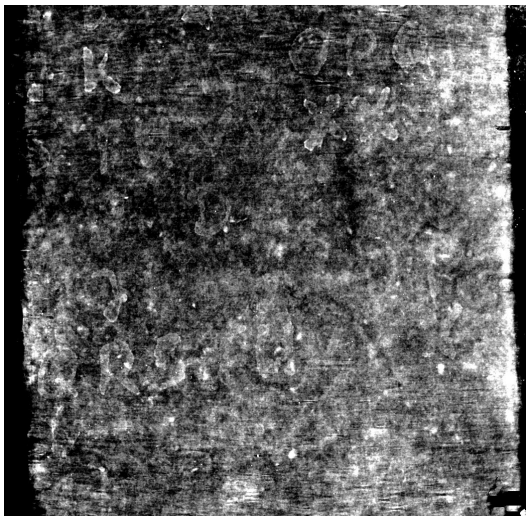
Iteration m	**Number of Corrupted Columns** ϵ	ϵ [%]
1	289011	28.30
2	9692	0.94
10	3081	0.30
20	2791	0.27
23	2752	0.27
24	2756	0.27



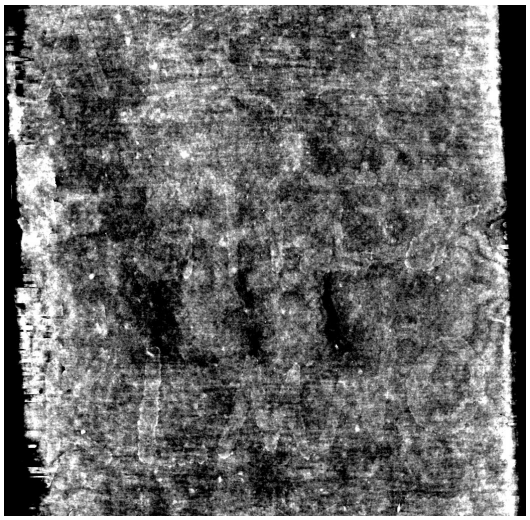
(a) Original Page 1



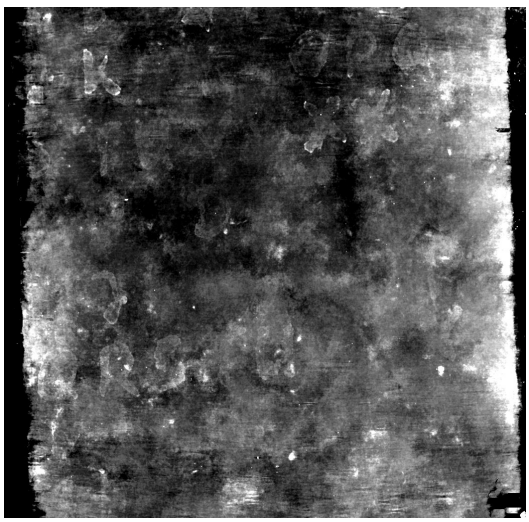
(b) Original Page 7



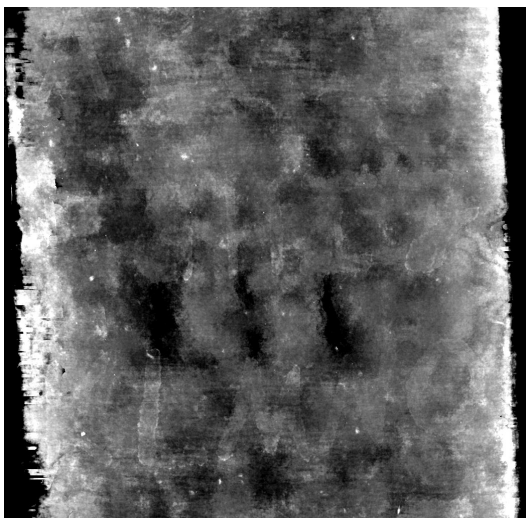
(c) 2-D Mapped Page 1



(d) 2-D Mapped Page 7



(e) Post-processed Page 1



(f) Post-processed Page 7

Fig. 7. Top row: Original pages of the book model notated with letters written with iron gall ink and a quill. Center row: Extracted and 2-D mapped pages of the 3-D reconstructed X-ray volume after applying the presented algorithm. Bottom row: Post-processed extracted and 2-D mapped pages.

While within the original volume shown in Fig. 1, where only the 'L' could be partly seen (orange circle), the 2-D mapped result shown in Fig. 7(f) of this page shows all letters when taking a closer look. It should be mentioned, that the aim of this work was to extract and 2-D map all pages. The identification of the writings is not part of this work. Therefore, more work on the scan parameters has to be done. However, letters where the ink deeply penetrated the paper can be easily identified with the naked eye. Writings with insufficient ink saturation appear ghost-like like in Fig. 7(d) and Fig. 7(f) caused by the well-know partial volume effect in CT. Applying the post-processing filters removes artifacts and noise significantly providing a good basis for further letter identification algorithms.

IV. Conclusion and Outlook

We think that using micro-CT scanners is a highly promising approach for the digitization of historical documents that can not be touched or opened any more. We developed a fully automatic algorithm to extract and 2-D map all book pages from a 3-D X-ray CT volume ($size = 1050 \times 980 \times 90$ voxel), acquired with a micro-CT system, used for material testing. We were able to identify writings on each without opening the book. The algorithm works robustly and efficiently without the need of any user interaction or prior knowledge. All pages of our book model could be extracted properly within approximately $6\,\mathrm{minutes}$. Setting a threshold within the segmentation reduced the total computation time to $4\,\mathrm{minutes}$. We showed that our algorithm is robust to work even on wavy pages, having a lot of spots where the pages were squeezed together. The pages were rather thin having a thickness of about $200\,\mu\mathrm{m}$ resulting in a diameter of approximately $3 - 4$ pixel. The algorithm can easily be adapted to other imaging modalities and also to other problems where the segmentation of thin lines has to be done. Although we were able to identify letters on the pages of the closed book, we were not able to visualize all writings of the pages. The reason for this is that the scan was not optimal and the used pressboard holder also influenced the measurement. Please note that the identification of letters was out of the scope of this paper. Instead we proposed an algorithm that can extract and visualize the pages without any user interaction.

In future, we want to find an optimal trajectory and parameters for the 3-D micro-CT scan. For example, we need to figure out which slice thicknesses are appropriate for given ink saturations to counter the mentioned partial volume effect. This could be achieved by creating a phantom with increasing ink saturations. We also want to evaluate the best fitting 3-D reconstruction algorithm for the book scan. Due to the rather low used X-ray energy, the volume is affected by many artifacts. Using iterative reconstructions such as SART [18] or TV-based methods [19] could be improve the results. Afterwards, we want to test our algorithm on real historical documents and extend it so that it works with scrolls. Given a correctly extracted page, we want to focus now our work on the identification of the writings and evaluate different modalities such as different paper, or ink.

References

[1] S. Noguchi, M. Yamada, Y. Watanabe, and M. Ishikawa, "Real-time 3d page tracking and book status recognition for high-speed book digitization based on adaptive capturing," in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*. IEEE, 2014, pp. 137–144.

[2] A. Stijnman, "Historical iron-gall ink recipes: art technological source research for inkcor," *Papier Restaurierung*, vol. 5, no. 3, pp. 14–17, 2004.

[3] O. Hahn, W. Malzer, B. Kanngiesser, and B. Beckhoff, "Characterization of iron-gall inks in historical manuscripts and music compositions using x-ray fluorescence spectrometry," *X-Ray Spectrometry*, vol. 33, no. 4, pp. 234–239, 2004.

[4] V. Mocella, E. Brun, C. Ferrero, and D. Delattre, "Revealing letters in rolled herculaneum papyri by x-ray phase-contrast imaging," *Nature communications*, vol. 6, 2015.

[5] A. Redo-Sanchez, B. Heshmat, A. Aghasi, S. Naqvi, M. Zhang, J. Romberg, and R. Raskar, "Terahertz time-gated spectral imaging for content extraction through layered structures," *Nature Communications*, vol. 7, p. 12665, 2016.

[6] D. Stromer, T. Schön, W. Holub, and A. Maier, "3-D Reconstruction of Iron Gall Ink Writings," in *Proceedings of 7th Conference on Industrial Computed Tomography (iCT 2017)*, K. Leuven, Ed., 2017.

[7] D. Stromer, V. Christlein, G. Anton, P. Kugler, and A. Maier, "3-D Reconstruction of Historical Documents using an X-Ray C-Arm CT System," in *Proceedings of the 31th conference on Image and Vision Computing New Zealand 2016*, M. University, Ed., 2016.

[8] F. Albertin, A. Patera, I. Jerjen, S. Hartmann, E. Peccenini, F. Kaplan, M. Stampanoni, R. Kaufmann, and G. Margaritondo, "Virtual reading of a large ancient handwritten science book," *Microchemical Journal*, vol. 125, pp. 185–189, 2016.

[9] F. Albertin, M. Romito, E. Peccenini, M. Bettuzzi, R. Brancaccio, M. Morigi, M. del Rio, D. Raines, G. Margaritondo, D. Psaltis *et al.*, "From closed testaments to books: Virtual x-ray reading as an alternate digitization technology for fragile documents."

[10] W. B. Seales, C. S. Parker, M. Segal, E. Tov, P. Shor, and Y. Porath, "From damage to discovery via virtual unwrapping: Reading the scroll from en-gedi," *Science Advances*, vol. 2, no. 9, 2016.

[11] D. Stromer, V. Christlein, T. Schön, W. Holub, and A. Maier, "Fast, Robust and Efficient Extraction of Book Pages from a 3-D X-ray CT Volume," in *Proceedings of the 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, G. Wang and X. Mou, Eds., 2017, pp. 401–404. [Online]. Available: https://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2017/Stromer17-FRA.pdf

[12] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1998, pp. 130–137.

[13] A. Budai, R. Bock, A. Maier, J. Hornegger, and G. Michelson, "Robust vessel segmentation in fundus images," *International journal of biomedical imaging*, 2013.

[14] L. Feldkamp, L. Davis, and J. Kress, "Practical cone-beam algorithm," *JOSA A*, vol. 1, no. 6, pp. 612–619, 1984.

[15] A. Maier, H. Hofmann, M. Berger, P. Fischer, C. Schwemmer, H. Wu, K. Müller, J. Hornegger, J.-H. Choi, C. Riess, A. Keil, and R. Fahrig, "CONRAD - A software framework for cone-beam imaging in radiology," *Medical Physics*, vol. 40, no. 11, 2013.

[16] A. Maier and R. Fahrig, *GPU Denoising for Computed Tomography*, 1st ed. Boca Raton, Florida, USA: CRC Press, 2015, vol. 1.

[17] A. Polesel, G. Ramponi, and V. J. Mathews, "Image enhancement via adaptive unsharp masking," *IEEE transactions on image processing*, vol. 9, no. 3, pp. 505–510, 2000.

[18] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm," *Ultrasonic imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[19] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE transactions on image processing*, vol. 7, no. 6, pp. 813–824, 1998.