Browsing Through Closed Books: Evaluation of Pre-processing Methods for Page Extraction of a 3-D CT Book Volume

D Stromer\textsuperscript{1a}, V Christlein\textsuperscript{1b}, T Schön\textsuperscript{2d}, W Holub\textsuperscript{2e} and A Maier\textsuperscript{1c}

\textsuperscript{1}Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Martensstraße 3, Germany
\textsuperscript{2}Fraunhofer Development Center for X-ray Technology, Flugplatzstraße 75, Fuerth, Germany

\textsuperscript{a}daniel.stromer@fau.de, \textsuperscript{b}vincent.christlein@fau.de, \textsuperscript{c}andreas.maier@fau.de, \textsuperscript{d}tobias.schoen@iis.fraunhofer.de, \textsuperscript{e}wolfgang.holub@iis.fraunhofer.de

Abstract. It is often the case that a document can not be opened, page-turned or touched anymore due to damages caused by aging processes, moisture or fire. To counter this, special imaging systems can be used. One of our earlier work revealed that a common 3-D X-ray micro-CT scanner is well suited for imaging and reconstructing historical documents written with iron gall ink – an ink consisting of metallic particles. We acquired a volume of a self-made book without opening or page-turning with a single 3-D scan. However, when investigating the reconstructed volume, we faced the problem of a proper automatic extraction of single pages within the volume in an acceptable time without losing information of the writings. Within this work, we evaluate different appropriate pre-processing methods with respect to computation time and accuracy which are decisive for a proper extraction of book pages from the reconstructed X-ray volume and the subsequent ink identification. The different methods were tested for an extreme case with low resolution, noisy input data and wavy pages. Finally, we present results of the page extraction after applying the evaluated methods.

1. Introduction
The standard procedure of using a book scanner is not applicable for imaging many historical documents. Although the page-turning process is done by an air flow, the documents may have a moisture, fire or aging damage that prohibits turning or opening. However, some methods exist for imaging and reconstructing such documents.

For our work, we assumed that writings are made with iron gall ink, an ink that has been widely used since the 5th century. Galilei, Mozart, da Vinci and even the Romans used this ink \cite{1} for writing or drawing their work. This kind of ink mainly consists of three materials – gum arabic, tannic acid and iron(II) sulfate. The latter element has a higher X-ray energy absorption allowing to differentiate it from the paper. Mocella et al used X-ray phase contrast imaging (Synchrotron) for recovering papyri rolls from Herculaneum \cite{2} having the drawback of being a highly complex, expensive, and not feasible approach for everyday applications due to the immobility of the scanning system. Another approach is based on Terahertz radiation \cite{3} which has the drawback of a limited depth resolution meaning that only a few pages can be imaged at once and has not been proofed for historical inks.
Our approach is using a common micro-CT X-ray system, used for medical imaging or material testing. We showed that such systems are also capable of imaging such documents [4,5]. We encountered the problem, that performing 3-D X-ray scans on books differ a lot from scrolls which is due to the shape of the objects. While the scroll has a uniform diameter, a book’s diameter differs with regard to the position of the gantry. This varying X-ray penetration length produces artifacts that do not occur with scrolls. Another major challenge when visualizing such documents by using 3-D imaging techniques is the extraction of single pages to inspect the writings on the paper. With higher resolutions and reduced X-ray energies, the acquisition and reconstruction times get higher. Also the extraction of rather thin pages out of the volume is a problem. They are stacked together in a very small area (sometimes more than 100 pages per cm) and squeezed at many points, but they are also wavy and the writings have a higher absorption than the paper. Without extraction, it is hardly possible to inspect single pages in the output volume of a reconstructed X-ray 3-D scan. Figure 1 shows the center slice of our generated volume. The green lines denote the transitions between two neighboring pages. To counter this problem, Seals et al. showed that using the Structure Tensor [6] performs good. However, the approach has been developed for scrolls and not for books, where many pages are squeezed together in the volume and the pages are not rolled up. A drawback of the structure tensor is that it has a poor corner response and edges are not a proper model for pages. The Vesselness filter defined [7] and normally used in medical imaging eliminates the drawbacks of the Structure Tensor and models the page as a curved surface.

In this work, we concentrate on the pre-processing of the acquired X-ray images. We show different approaches to create a binarized volume that sets the pages to ‘1’ and the air to ‘0’. Therefore, we use global and adaptive thresholding methods as well as the Structure Tensor and the Vesselness filter. Afterwards, a selected page is extracted by a straight-forward segmentation algorithm. After applying all methods on a small snippet of the volume and comparing it to a manually segmented version with respect to computation time and exactness, we apply the methods on a larger part of the page.

2. Materials and Methods

2.1. The Book Model
We built a book model with ten pages. The paper is handmade and has a thickness of about 200 μm and the writings were made with iron gall ink and a quill. Letters of different font sizes, line thicknesses and paper saturations have been written on the pages. The height of the letters range from 1 cm to 2.5 cm. The pages were put in a pressboard holder during the X-ray scan to allow an upright book scan. The book consists of three different materials. Iron gall ink is assumed to be iron(II) sulfate heptahydrate (FeSO₄ ⋅ 7H₂O) with a density 1.89 g ⋅ cm⁻³ having the highest intensities in the output. Pages were expected to consist of cellulose with the chemical formula C₆ H₁₀ O₅ and a specific density 1.5 g ⋅ cm⁻³ resulting in medium range intensities. Gaps between pages are assumed to be air and appear dark in the images.

2.2. CT Scan Trajectory and 3-D Reconstruction
For the scan, we used a micro-CT system built for non-destructive material testing and performed a 360° circular scan with 2400 projections using cone-beam geometry. This geometry is commonly used for 3-D X-ray CT scans and derives its name from the cone-shaped field-of-view as shown in Figure 2. We used a quadratic 2864 × 2864 px flat panel detector with a pixel shift of 80 μm and set up an energy of 70 kVp which is sufficient to see larger objects written with the ink. The book model was positioned upright on the turntable and y denotes the axis of rotation.

For the 3-D reconstruction the commonly used FDK algorithm [8] was used. The algorithm consists of three steps – cosine weighting, ramp filtering and back projection. The chosen voxel size was 68 μm² such that approximately 4 px cover one page. In total, the scan time was kept relatively
low (about 1.5 hours). All calculations were performed with the CONRAD framework for cone beam geometries [9].

Figure 1. Center slice of the generated 3-D volume with the green lines denoting the transition between two pages.

2.3. Page Extraction

As we want to investigate single pages, we need to extract single pages out of the volume. We use an Figure 3 shows a xz-layer slice of the 3-D X-ray volume showing the ten book pages. We select two start points \( p_1 \) and \( p_2 \) on the page-of-interest (green points), one close to the left and one close to the right. Afterwards, the algorithm walks along the page, segments the page (orange line) and crops the rest of the book. The key for the distinction between page and air is the use of two thresholds since one intensity threshold is not sufficient enough due to the squeezed pages. The red box shows such an area where the lower page is clearly differentiable from the other two but the upper two pages’ overlap such that a simple threshold method would run into problems. Other problems are: ink that is brighter than the paper, as well as X-ray artifacts and noise disrupting the image.

Figure 3. Center xz-layer slice of the 3-D X-ray volume showing the ten book pages that were stacked together in the wooden holder.

Figure 4 shows the complete processing pipeline. In the following, we will show and evaluate different approaches to perform the binarization process where the page is set to ‘1’ and air is set to ‘0’ such that the presented algorithm extracts (Page Extraction) only those parts of the page that really belong to it. After the binarization, we multiply the binary image with the original volume and receive the final page.

Figure 4. Processing pipeline for the extraction and 2-D mapping of a single page out of the volume.
2.4. 3-D Volume Pre-processing

Different approaches for binarization exist, where we picked out six common approaches to compare them to each other. Before applying those algorithms to the volume, we initially smoothed the single slices with a Gaussian Blur filter (σ = 2). This will not affect the output, because we afterwards will multiply the generated binary volume with the original volume, but we get rid of the noise and smooth the writings. Changes in the kernel size or σ had only small effects on the end result.

2.4.1. Otsu’s Thresholding. The first method of choice is Otsu’s thresholding [10], a global thresholding method trying to minimize the intra-class variance between two classes (in our case: air and page) by

\[ \sigma_w^2(t) = w_1(t) \sigma_1^2(t) + w_2(t) \sigma_2^2(t). \]  

where \( w_1,2 \) are the class probabilities, \( \sigma_{1,2}^2 \) are the class variances and \( t \) is the threshold. \( w_{1,2}(t) \) is calculated by iterating through the histogram where the goal is to minimize the intra-class variance while maximizing the inter-class variance.

2.4.2. Adaptive Thresholding Approaches. In comparison to the global thresholding methods, local thresholding methods work on small patches that are slid over the whole image [11]. We picked out two commonly used approaches for our tests. The first method is the Adaptive Mean Threshold putting the same weight on every pixel in the given neighborhood \( N_n \) with \( n \) denoting the kernel size. This leads to the filter kernel \( k_m(x,y) \) of Equation (2) where \( i(x,y) \) denotes the specific intensity value.

\[ k_{m,n \times n}(x,y) = \frac{1}{n \cdot n} \sum_{k,l \in N_n} i(k,l) \]  

Secondly, the Adaptive Median Threshold function is tested. It uses the same filter kernel as the Mean filter, but sorts the pixel intensities in ascending order and considers the middle value as threshold. The third approach is the Adaptive Gaussian Threshold with the filter kernel \( k_g(x,y) \) shown in Equation (3) where the standard deviation \( \sigma \) defines the range of the filter function.

\[ k_g(x,y) = \frac{1}{2\pi \sigma} \exp \left( -\frac{x^2 + y^2}{2\sigma^2} \right) \]  

The entire volume is filtered with the specific filter kernel function and afterwards, the original pixel value is compared to the filter output value. If the value is smaller than the output, it is set to ‘0’, else it is set to ‘1’. As all filters are separable in each direction, the filtered image can be computed within a very short time.

2.4.3. Structure Tensor. Another approach that is part of this evaluation is the Structure Tensor. Seales et al. used this method for unwrapping the En-Gedi scroll successfully [6]. However, it has to be investigated that this method also fits on a book whose structure differs from the scroll in the way that the pages are compressed. The Structure Tensor is defined by the spatial averaged tensor product of the gradients in \( x \) - and \( y \)-direction as shown in Equation (4), with \( \nabla f_{\sigma_2} = (\nabla k_{\sigma_2}) \ast f \).

\[ J_{\sigma_1,\sigma_2}(x,y) = k_{\sigma_1}(x,y) \ast (\nabla f_{\sigma_2} \otimes \nabla f_{\sigma_2}). \]  

Afterwards, the eigenvalues \( \lambda_1 \) and \( \lambda_2 \) (with \( \lambda_1 \geq \lambda_2 \)) are computed which describe the structure of the image by the following three relations with \( \varepsilon \) denoting a small threshold value. \( \lambda_1 \approx \lambda_2 \leq \varepsilon \) describes a flat area, \( \lambda_1 \gg \varepsilon > \lambda_2 \) describes an edge and \( \lambda_1 \geq \lambda_2 > \varepsilon \) describes a corner. The edge image is set as the binarization image, but it has to be mentioned that an edge is not a good model for a page and the corner response of the Structure Tensor is poor such that is has to be proofed if this method is appropriate for the given task.
2.4.4. Vesselness filter. The Vesselness filter eliminates the drawbacks of the Structure Tensor by modeling the page as a curved surface and highlighting tube-similar structures. It is widely used in medical imaging for segmenting vessels with differing diameters. Instead of the tensor product, the image’s hessian matrix is computed by Equation (5) and (6). The second derivative of the Gaussian kernel at scale $\sigma$ is used to measure the contrast between the regions inside and outside the range $(-\sigma, +\sigma)$ in the derivatives’ directions.

$$H_\sigma = \begin{bmatrix} \frac{\delta^2}{\delta x^2} I(x, \sigma) & \frac{\delta}{\delta x \delta y} I(x, \sigma) \\ \frac{\delta}{\delta x \delta y} I(x, \sigma) & \frac{\delta^2}{\delta y^2} I(x, \sigma) \end{bmatrix}, \quad I(x, \sigma) = I(x) * k_g(x, \sigma) = I(x) * \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{|x|^2}{2\sigma^2}\right)$$

As we know that the paper has a constant diameter of about 4 $px$ over the entire length and all slices, we can simplify the method and use only a single $\sigma$, which also reduces the computation time. Afterwards, the vessel parameters are calculated by Equation (7) and (8) and the total ‘Vesselness’ is computed by Equation (9) where $\beta$ and $c$ are control parameters ($\beta = 0.5$, $c$ depends on the scale). $\lambda_1$ and $\lambda_2$ are the eigenvalues of $H_\sigma$ (with $\lambda_1 \geq \lambda_2$). The output $V(x, \sigma)$ is the finally generated probability for the vessel, or in our case the page. To binarize this output, we set the threshold value to 95 % meaning that only the most-likely part is considered to be part of the page.

$$R_B = \frac{\lambda_2}{\lambda_1}, \quad S = \sqrt{\lambda_1^2 + \lambda_2^2}, \quad V(x, \sigma) = \begin{cases} 0 & \lambda_1 \approx 0 \\ \exp\left(-\frac{R_B^2}{2\beta^2}\right) \left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right) & \lambda_1 > 0 \end{cases} \quad (7,8,9)$$

2.5. Measurements
The abovementioned methods are compared with regard to the total computation time (in seconds) and common segmentation measures (RMSE, Accuracy (acc) and f1-score (f1)) [12]. Finally, the resulting 2-D images will also be inspected with the naked eye. For the RMSE, a smaller value denotes a better match. For the remaining four accuracy measurements, the result gets more exact the closer the value is to ‘1’.

3. Results
First, a small snippet with a total size of $380 \times 180 \times 85$ voxel was cropped and manually segmented. For the Vesselness filter, the Structure Tensor, and the Adaptive Gaussian Thresholding, $\sigma = 2$ was the best solution (tested: $\sigma = [0.25, 0.5, 1, 2, 3, 5, 10]$). For Mean and Median filtering, a kernel size of (5,2) was configured. Table 1 shows that the Vesselness approach is the most exact approach while the Structure Tensor performs worst. For the adaptive thresholding methods, the Gaussian performs best with respect to the RMSE, accuracy and f1-score. The Vesselness filter had the longest computation time while the Mean and Gaussian thresholding methods performing about eight times faster.

| Table 1. Results for the comparison of the manually segmented volume with the proposed methods with regard to accuracy and computation time. |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                | Otsu           | Mean          | Median         | Gaussian       | Structure      | Vesselness     |
| RMSE           | 0.509          | 0.397         | 0.390          | 0.387          | 0.606          | 0.374          |
| acc            | 0.740          | 0.843         | 0.848          | 0.850          | 0.633          | 0.860          |
| f1             | 0.723          | 0.785         | 0.773          | 0.790          | 0.522          | 0.797          |
| time [s]       | **0.081**      | 0.415         | 1.101          | 0.502          | 3.554          | 4.129          |
Next, we performed the calculations on the complete volume. Table 2 shows the results. The computation time for the Vesselness filter was 38.542 s for processing a volume of 1430 × 500 × 85 voxel. Ignoring Otsu’s thresholding, the Gaussian filter performed best again with being approximately 17 times faster than the Vesselness filter.

|                | Otsu | Mean | Median | Gaussian | Structure | Vesselness |
|----------------|------|------|--------|----------|-----------|------------|
| *time [s]*     | 0.300| 3.891| 11.484 | 2.292    | 33.626    | 38.542     |

Figure 5(a) shows the original part of the page that should be extracted out of the volume. The extraction algorithm of the binarized volume mentioned in Section 2.3. took about 2.5 minutes with identical starting points. Figures 5(b-f) show the outputs starting with the Vesselness filter as object of comparison at the top left position, and the outputs sorted by decreasing f1-score starting at the top left cell (Vesselness) ending at the bottom right position (Otsu’s Thresholding). The result of the Structure Tensor was not considered. The examples show that the result gets slightly worse, visible through the increasing black holes, which confirms the measurements. The RMSE of the Vesselness 2-D image (Figure 5(b)) and the Gaussian 2-D image (Figure 5(c)) is 0.075. Additionally, it should be mentioned, that applying the 3-D versions of the Structure Tensor and the Vesselness filter did not yield to improved results but led to higher computation times.

4. Discussion and Outlook
In this work, we evaluated different approaches for the binarization of a book volume acquired with a micro-CT X-ray system which is an essential part for the investigation of single page’s writings. Our pipeline consists of an initial volume smoothing and denoising followed by one of the binarization approaches. Finally, the page-of-interest is getting extracted and 2-D mapped. The total computation time takes about 3 – 4 minutes (volume size = 1430 × 500 × 85 voxel). We showed, that with the straight-forward segmentation method, the Vesselness filter delivered the best results with respect to accuracy. However, this method had the longest computation time (approximately 26 s). The Adaptive Thresholding Methods performed quite well, with the Gaussian filter having the edge over the Mean and Median filter. Otsu’s threshold and the Structure Tensor turned out to be unsuited for the
implemented extraction algorithm. In future, we want to evaluate the methods also on scrolls. We also want to find the best scanning system parameters, such as X-ray energy, collimation and detector pixel shift. Furthermore, we want to try reconstructing different inks and papers, such as parchment or modern paper, with other reconstruction approaches.

References
[1] Kolar J, et al. "Historical iron gall ink containing documents - properties affecting their condition." Analytica chimica acta 555.1: 167-174, 2006.
[2] Mocella V, Brun E, Ferrero C, and Delattre D. “Revealing letters in rolled herculaneum papyri by X-ray phase-contrast imaging.” Nature communications, vol. 6, 2015.
[3] Redo-Sanchez A, Heshmat B, Aghasi A, Naqvi S, Zhang M, Romberg J and Raskar R. “Terahertz time-gated spectral imaging for content extraction through layered structures.” Nature Communications, vol. 7, p. 12665, 2016.
[4] Stromer D, Christlein V, Anton G, Kugler P and Maier A. “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.
[5] Stromer D, Schoen T, Holub W and Maier A. “3-D Reconstruction of Iron Gall Ink Writings.” In Proceedings of the 7th Conference on Industrial Computed Tomography, Leuven, 2017.
[6] Seales W B, Parker C S, Segal M, Tov E, Shor P and Porath Y. “From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi.” Science Advances, 2(9), e1601247, 2016.
[7] Budai A, Bock R, Maier A, Hornegger J and Michelson G. “Robust vessel segmentation in fundus images.” International journal of biomedical imaging, 2013.
[8] Feldkamp L, Davis L and Kress J. “Practical cone-beam algorithm.” JOSA A, vol. 1, no. 6, pp. 612–619, 1984.
[9] Maier A, Hofmann H, Berger M, Fischer P, Schwemmer C, Wu H, Müller K, Hornegger J, Choi J H, Riess C, Keil A and Fahrig R. “CONRAD - A software framework for cone-beam imaging in radiology.” Medical Physics, vol. 40, no. 11, 2013.
[10] Otsu N. “A threshold selection method from gray-level histograms.” Automatica, vol. 11, no. 285-296, pp. 23–27, 1975.
[11] González R C and Woods R E, Digital Image Processing (2nd Ed). Prentice Hall, 2002.
[12] Powers, D M. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." Journal of Machine Learning Technologies 2 (1), pp. 37–63, 2011.