

Image Text Enhancer

Algorithm Engineering 2026 Project Paper

Daniel Motz

Friedrich Schiller University Jena
Jena, Germany
daniel.motz@uni-jena.de

Leonard Teschner

Friedrich Schiller University Jena
Jena, Germany
leonard.teschner@uni-jena.de

Mher Mnatsakanyan

Friedrich Schiller University Jena
Jena, Germany
mher.mnatsakanyan@uni-jena.de

Abstract

The five-finger pattern:

- (1) **Topic and background:** What topic does the paper deal with? What is the point of departure for your research? Why are you studying this now?
- (2) **Focus:** What is your research question? What are you studying precisely?
- (3) **Method:** What did you do?
- (4) **Key findings:** What did you discover?
- (5) **Conclusions or implications:** What do these findings mean? What broader issues do they speak to?

Keywords

noise reduction, background removal, image filter, binarization

ACM Reference Format:

Daniel Motz, Leonard Teschner, and Mher Mnatsakanyan. 2026. Image Text Enhancer: Algorithm Engineering 2026 Project Paper. In *Proceedings of AEPROM 2026: Algorithm Engineering Projects (AEPROM 2026)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnn>

1 Introduction

1.1 Background

In der Zeit der Digitalisierung werden viele gedruckte, handschriftliche und historische Dokumente mittels Scanner oder Smartphone-Kamera digitalisiert. Dabei entstehen häufig digitale Bilder mit schlechter Qualität, die nicht für die digitalen Verfahren der image analysis, wie text detection, geeignet sind.

Bei der digitalisierung der Bildern können unterschiedliche challenges durch die Aufnahmearbeit, als auch durch den Dokumententyp entstehen: Durch Smartphone-Kameras gescannte Dokumente können nicht alle Details wie ein dedizierte Dokumenten-Scanner aufnehmen, wodurch blurry und distorted Bilder entstehen. Auch der Winkel und der Abstand der Kamere zum Dokument haben einfluss auf die Qualität. So können distorted Perspektiven und Probleme für das Text Alignment verursacht werden. Die Lichteigenschaften können Schatten werfen, glare und reflektionen hervorrufen. Zusätzlich können Smartphone-Kameras Text auffangen, der nicht zum gescannten Dokument selbst gehört [5][2].

Historische Dokumente haben durch ihr schieres Alter Probleme. So können sie fading und noise aufweisen, die den Text unlesbar machen. Die digitalisierung der Dokumente wird komplexer

durch handschriftliche Notizen, überlappende Texte und variationen im Style, sowie durch beschädigte Seiten und risse oder auch fungal spots [5].

1.2 Related Work

Der erste Schritt bei der image analysis, wie OCR Texterkennung und Recognition Systemen, ist die Segmentierung von Hintergrund und dem Fordergrund. [17]. Die Segmentierung wird druch Binarisierung durchgeführt. Es gibt Traditionelle Methoden, welche einen globalen [11], lokalen [4][6][12] oder auch gemixten Threshold berechnen[16][17]. Mit Image Feature Methoden wurden Edge Detection [9][17] und Fuzzy Logics [14][17] eingesetzt. In den letzten Jahren sind zusätzlich zu den Traditionellen Methoden noch Deep learning Binarisierungsmethoden hinzugekommen. Diese beruhen auf Convolutional Neural Networks, Generative Adversarial Networks oder Attention Mechanisms [17].

Neben den reinen Binarisierungsmethoden gibt es Ansätze, die mehrere Bildverarbeitungsmethoden kombinieren, um die Qualität von Textbildern zu verbessern. So wird in [2] eine Pipeline vorgestellt, die Entropy Filter und Morphologische Operationen mit Binarisierung kombiniert. In [10] wird Niblacks Binarisierung mit einem Laplace Filter und globaler Optimierung kombiniert. Ein weiterer Ansatz ist in [15] beschrieben, wo verschiedene Binarisierungsmethoden kombiniert werden und per Votingmechanismus entschieden wird, welcher Pixel Vorder- oder Hintergrund ist.

1.3 Our Contribution

In diesem Paper stellen wir eine Pipeline vor, die mehrere Bildverarbeitungsmethoden kombiniert, um die Qualität von gescannten oder fotografierten Textbildern zu verbessern. Die Pipeline umfasst Schritte wie Deskewing, Kontrastverbesserung, Rauschunterdrückung, Binarisierung, Despeckle, und morphologische Operationen. Das binarisierte Ergebnis der Pipeline kann außerdem zurück in ein Farbbild konvertiert werden.

Die Pipeline ist modular und userfriendly aufgebaut, sodass Nutzer auswählen können, welche Schritte sie anwenden möchten, um die Pipeline auf ihre spezifischen Anwendungsfälle anzupassen können. Die Methodenparameter können einzeln angepasst werden, sind aber für die userfriendlyness mit den state of the art Werten standardmäßig konfiguriert.

Uns ist die effiziente Verarbeitung in der Pipeline wichtig. Daher vergleichen wir uns mit traditionellen open-source Binarisierungsmethoden und Bildverbesserungsverfahren und lassen bewusst Deep Learning Ansätze, sowie zahlungspflichtige Software außen vor.

Wir stellen die Pipline als C++ Bibliothek, zusätzlich mit den einzelnen Methoden zur Verfügung. Die Implementierung nutzt

parallelisierung durch OpenMP, sowie weitere Optimierungsmethoden wie Loop-blocking um eine effiziente Verarbeitung großer Bilddatenmengen zu gewährleisten.

1.4 Outline

This paper is structured as follows: Section 2 provides a detailed description the developed pipeline and its methods. Section 3 demonstrates the performance of our pipeline using experiments. Finally, Section 4 summarises our results and provides an outlook on possible future work.

2 The Pipeline

Many approaches and best practices already exist for improving the quality of scanned images as seen in section 1.2. We have developed a pipeline that combines several of these methods in order to achieve potentially good results. Users can choose which steps to apply from the pipeline. The individual methods of the pipeline are shown in algorithm 1.

Algorithm 1 Image Text Enhance Pipeline

- (1) convert image to grayscale
 - (2) Deskew (if requested)
 - (3) Contrast enhancement
 - (4) Denoising
 - (5) Binarization
 - (6) Despeckle (if requested)
 - (7) Morphological operations (if requested)
 - (8) Color passthrough
-

The individual methods of the pipeline are explained in more detail below.

2.1 Convert image to graysacle

All pipeline methods work on grayscale images. Therefore, the first step is to convert the input image into a grayscale image. This is achieved by applying the weighted sum $Y = 0.299R + 0.587G + 0.114B$, as defined by the International Telecommunication Union [8], to each pixel. The result is a grayscale image in which the brightness value of each pixel, represented by Y , corresponds to that of the original RGB-pixel. All steps in the pipeline are performed on the converted grayscale image in-place after the conversion.

2.2 Deskew

Damit der Text im Bild horizontal ausgerichtet ist, wird ein Deskew Schritt auf den wunsch des Nutzers hin durchgeführt. Besonders text analysis Methoden wie OCR profitieren von horitonal ausgerichteten Texten [3]. Unser Deskew Algorithmus nutzt die Projection Profile Methode, die ähnlich zu [1] ist. Zuerst wird das Bild in Graustufen konvertiert und mittels Sauvols Methode Binarisiert. Im zweiten Schritt wird der Winkel gesucht, der die horizontale projections Varianz maximiert. Abschließend wird das Bild rotiert, um den Skew zu korrigieren.

siehe Abschnitt 2.5 oder [12]

Die Vorteile dieser Methode sind, dass sie Polarity-safe ist, eine coarse-to-fine Winkelsuche für Effizienz nutzt und Neumann Randbedingungen verwendet, um schwarze Ecken zu vermeiden.

2.3 Contrast enhancement

Um die unterschiedlichen Lichtverhältnisse und daraus entstandenen Kontrastprobleme bei der Digitalisierung von Dokumenten zu adressieren, wird ein Kontrasverbesserungsschritt durchgeführt. Dies hilft der Binarisierung im weiteren Verlauf [2]. Dabei wird ein robuster linearer Kontrast Streching Algorithmus angewendet. Dabei werden die unteren 1% und oberen 1% der Intensitäten abgeschnitten, um Ausreißer zu ignorieren. Der verbleibende Bereich wird dann auf den vollen Bereich von 0 bis 255 gestreckt.

2.4 Denoising

Digitale Bilder sind durch ihre Aufnahmeart, Kompression oder durch ihren Transmissionchannel mit Rauschen vershen, bei der Bildinformationen verloren gehen. Durch das Vorhandensein von Rauschen sind unteranderem Bild analysis Schritte in ihrer effektivität eingeschränkt [7]. Daher wird in der Pipeline ein Denoising Schritt durchgeführt, der das Rauschen reduziert. Es stehen verschiedene einfache und adaptive Filtermethoden zur Verfügung, die der Nutzer auswählen kann. Es werden zwei einfache Filtermethoden angeboten, die von CImg bereitgestellt werden: Ein Gaussian Blur Filter mit Neumann Randbedingungen und ein nicht-linearer Median Filter[13]. Zusätzlich werden zwei adaptive Filtermethoden angeboten: Ein adaptiver Gaussian Blur Filter, der eine variable Standardabweichungen nutzt, um an Kanten weniger zu verwischen und in flachen Regionen, mit hoher Varianz, stärker zu verwischen. Zusätzlich zum einfachen Median Filter wird ein adaptiver Median Filter bereitgestellt, der besonders gut für die Entfernung von Impulsrauschen (salt-and-pepper) geeignet ist, während Kanten und feine Details erhalten bleiben. Dabei wird mit einem 33 Fenster begonnen, das bei der Erkennung von Impulsrauschen bis zu einer definierten maximalen Fenstergröße erweitert wird, während nicht-Impuls-Pixel unverändert bleiben. Dies macht ihn ideal für die Entfernung von Scan-Speckle in Textbildern.

2.5 Binarization

Sauvola.

Bataineh.

Adaptive binarization using local threshold
Adaptive window size

2.6 Despeckle

2.7 Morphological operations

2.8 Color passthrough

3 Experiments

4 Conclusions

References

- [1] Teruo Akiyama and Norihiro Hagita. 1990. Automated entry system for printed documents. *Pattern Recognition* 23, 11 (1990), 1141–1154. doi:10.1016/0031-3203(90)90112-X

erkennt hellen vs. dunklen Hintergrund

- [2] Musab Kasim Alqudah, Mohammad F. Bin Nasrudin, Bilal Bataineh, Mashal Alqudah, and Arwa Alkhatafneh. 2015. Investigation of binarization techniques for unevenly illuminated document images acquired via handheld cameras. In *2015 International Conference on Computer, Communications, and Control Technology (I4CT)*. 524–529. doi:10.1109/I4CT.2015.7219634
- [3] Wuzhida Bao, Cihui Yang, Shiping Wen, Mengjie Zeng, Jianyong Guo, Jingting Zhong, and Xingmiao Xu. 2022. A Novel Adaptive Deskewing Algorithm for Document Images. *Sensors* 22, 20 (2022). doi:10.3390/s22207944
- [4] Bilal Bataineh, Siti Abdullah, Khairuddin Omar, and Mohammad Faidzul Nasrudin. 2011. Adaptive Thresholding Methods for Documents Image Binarization, Vol. 6718. 230–239. doi:10.1007/978-3-642-21587-2_25
- [5] Bilal Bataineh, Mohamed Tounsi, Nuha Zamzami, Jehan Janbi, Waleed Abdel Karim Abu-ain, Tarik AbuAin, and Shaima Elnazer. 2025. A Comprehensive Review on Document Image Binarization. *Journal of Imaging* 11, 5 (2025). doi:10.3390/jimaging11050133
- [6] Derek Bradley and Gerhard Roth. 2007. Adaptive Thresholding using the Integral Image. *J. Graphics Tools* 12 (01 2007), 13–21. doi:10.1080/2151237X.2007.10129236
- [7] Linwei Fan, Fan Zhang, Hui Fan, and Caiming Zhang. 2019. Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art* 2, 1 (2019), 7. doi:10.1186/s42492-019-0016-7
- [8] International Telecommunication Union. 2011. *Recommendation ITU-R BT.601-7: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*. Recommendation BT.601-7. ITU Radiocommunication Sector (ITU-R), Geneva, Switzerland. https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-7-201103-I!!PDF-E.pdf
- [9] Shijian Lu, Bolan Su, and Chew Lim Tan. 2010. Document image binarization using background estimation and stroke edges. *International Journal on Document Analysis and Recognition (IJDAR)* 13, 4 (2010), 303–314. doi:10.1007/s10032-010-0130-8
- [10] Sergey Milyaev, Olga Barinova, Tatiana Novikova, Pushmeet Kohli, and Victor Lempitsky. 2013. Image Binarization for End-to-End Text Understanding in Natural Images. In *2013 12th International Conference on Document Analysis and Recognition*. 128–132. doi:10.1109/ICDAR.2013.33
- [11] Nobuyuki Otsu. 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9, 1 (1979), 62–66. doi:10.1109/TSMC.1979.4310076
- [12] J. Sauvola and M. Pietikäinen. 2000. Adaptive document image binarization. *Pattern Recognition* 33, 2 (2000), 225–236. doi:10.1016/S0031-3203(99)00055-2
- [13] Jean-Philippe Tarel. 2024. ClImg: A Simple C++ Toolkit for Image Processing. <https://cimg.eu>. Version 3.x.
- [14] Li-Jing Tong, Kan Chen, Yan Zhang, Xiao-Ling Fu, and Jian-Yong Duan. 2009. Document Image Binarization Based on NFCM. In *2009 2nd International Congress on Image and Signal Processing*. 1–5. doi:10.1109/CISP.2009.5305330
- [15] Giorgiana-Violeta Vlăceanu, Caraman Ghenadie, Răzvan Nitu, and Costin-Anton Boianiu. 2022. A voting method for image binarization of text-based documents. In *2022 21st RoEduNet Conference: Networking in Education and Research (RoEduNet)*. 1–5. doi:10.1109/RoEduNet57163.2022.9921086
- [16] You Yang. 2008. OCR Oriented Binarization Method of Document Image. In *2008 Congress on Image and Signal Processing*, Vol. 4. 622–625. doi:10.1109/CISP.2008.262
- [17] Zhengxian Yang, Shikai Zuo, Yanxi Zhou, and Jianwen He, Jinlong and Shi. 2024. A Review of Document Binarization: Main Techniques, New Challenges, and Trends. *Electronics* 13, 7 (2024). doi:10.3390/electronics13071394