Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 227 (2023) 922–931

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Computer Science and Computational Intelligence (ICCSCI 2023)

Image Pre-Processing Effect on OCR's Performance for Image Conversion to Braille Unicode

Andrean Ignasius[a], Joerio Christo Chandra[a], Roberto Oscadinata[a], Derwin Suhartono[a]*

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University Jakarta 11480, Indonesia*

Abstract

The number of cases of visual impairment has been rising, particularly in Indonesia. Additionally, most disabilities in Indonesia are visual impairments, according to data from the Population Census. Blindness is one aspect of this condition that accounts for a sizable percentage of it. Considering this, we are making an effort to assist them by translating literature, which entails converting text included in images into braille Unicode. A braille embosser, a device that can print braille sheets for the blind to read, will then be used to print this translation into braille. In this study, we discuss the consequences of pre-processing while converting a text-containing image to braille Unicode. Here, we employ the pre-processing techniques of binarization, denoising, and skew correction. These techniques are used before using the OCR, and the results are documented in this study. Additionally, the outcomes of pre-processed and raw photos are contrasted to determine the efficacy of applying pre-processing to an image.

Keywords: Visual Impairment; Optical Character Recognition (OCR); Binarization; Denoising; Skew Correction' Text to Braille conversion

1. Introduction

According to research, cases of visual impairment has increased dramatically since the last century and is happening across the world. The statistical estimation stated that the number of blind people in 2020 had reached 49.1 million, a staggering increase of 42.8% (34.4 million) from 1990 [1]. In Indonesia, visual disabilities alone made up around 29.6 % of all disabilities based on 2012 Population Census Data, making it the most common disabilities among Indonesian (ignoring the percentage of people with more than one disability). Visual Impairments in Indonesia translates to 3.05% of Indonesian total citizen which is roughly about 5,820,000 lives.

* Corresponding Author.
E-mail address: dsuhartono@binus.edu

10.1016/j.procs.2023.10.599

Table 1. Percentage of population above age of 10 experiencing functional difficulties based on the 2010 population census data [2]

	Number of people (in thousands)		Percentage
	Low	High	
Blind	5313	507	3.05
Deaf	5268	456	1.58
Walk	2432	656	1.62
Concentrate	2126	616	1.44
Be Independent	9046	9046	4.74

Back then, it was impossible for people with blindness to read books, that was until the year 1824 with the invention of Braille language. The inventor of this writing system, Louis Braille, was one of the visually impaired, as well. He was blinded at the age of three, and he came up with this idea when he was a student at the National Institute for Blind Children in Paris [3].

Since then, possibilities of blind people reading has been brought back as many have used the writing system for their daily lives. Unfortunately, not all books are translated to Braille and as a fact there are not many authors and translator out there capable of writing in Braille. There are only a few individuals, and most of them doesn't seem to be willing to contribute their time and energy into translating the lengthy pages of a book into Braille. Therefore, with this idea at hand, we are trying to broaden the opportunity of blind people to read books.

In almost every area, text-based information systems have become increasingly relevant. The source of the input text in many cases (such as physical newspapers or old printed books) is not editable documents, but documents in their original paper form. This is a very serious issue for visually impaired people as getting information from this text-based information can be nigh impossible due to the few to non-Braille versions of this type of information. But, in this modern time, we can get the text content by using a technology called OCR (Optical Character Recognition).

In this paper, we will be using OCR to convert physical text into machine-edible text then convert it into Braille Unicode. We will implement OCR by using an architecture called Tesseract. The accuracy of OCR varies on what quality of image is given to the OCR tool and in our case the word accuracy is really important because we need to convert a document or a book into Braille Unicode because of that we will explore image pre-processing effect on OCR accuracy.

2. Literature Review

A. OCR Tesseract

The mechanical or electronic conversion of images of handwritten or typed text into machine-editable text is known as Optical Character Recognition, also known as OCR [5]. The consistency of the input documents has a direct impact on OCR output and accuracy [5]. The OCR process can be divided into many other parts: image acquisition, image pre-processing, image segmentation, and matching & recognition [6]. OCR ability to scan physical images to then be converted into machine-editable text has allowed scientists to develop many new technologies such as. handwriting recognition, receipt imaging, captcha, number plate recognition, and many more[7]. Moreover, OCR has the ability to learn from its past misspellings and errors by simply comparing raw scan to the OCR-processed result side-by-side in order to increase its accuracy for the future conversions [8, 9]. This error- correcting part is undoubtedly very critical for document safekeeping, since e-books with digitized texts are meant to be accessed by a wide range of people on the internet for a variety of purposes [10]. Besides alphabetical characters, punctuations and symbols can confuse the machine, as well. In this case, the accuracy of OCR can be increased by using the method called convolutional preprocessing [11, 12].

B. Image preprocessing

Image preprocessing is done to create a clearer form of image. It is done to increase the accuracy of the image being scanned to be recognized by computers and algorithms. Techniques used in preprocessing an image are:

B.1. Binarization

Binarization technique involves changing a colored image into an image with only black and white pixels. Binarization can be done implementing a threshold (normally threshold=127, half the pixel range of the color white 0 and black 255). Binarization will then use that threshold to determine if the pixel value is greater than the threshold, it is then made into a white pixel. The techniques used to find the threshold are [15]:

- Otsu: A technique that divides the pixels into two classes foreground and background. Weight is counted by using Eq.1.

$$\sigma^2 p(t) = p1(t)\sigma_1^2 + p2(t)\sigma_1^2(t) \quad (1)$$

- Niblack's: A local thresholding method, meaning it will define different threshold values for every pixel by using Eq.2.

$$T(i, jj) = \mu + k \times \sigma \tag{2}$$

- Sauvola: The improvement of the Niblack method by using a variance method that utilizes standard deviation. Local thresholds are found by using Eq.3.

$$T(i, jj) = \mu \times [1 + k(\frac{\sigma\sigma}{R} - 1) \tag{3}$$

- Bernsen: The threshold is estimated as the average of highest and lowest intensity values in the window by using Eq. 4.

$$C(i. jj) = lmax - Imin \tag{4}$$

Algorithms in binarization do not work well for all types of images, but combining different algorithms for specific documents can result in a better performance [15].

B.2. Skew Correction

Skewed paper images obstruct subsequent image processing and can also result in incorrect outcomes, so skew identification and correction are critical measures in image preprocessing [17]. To fix this problem we need to apply an image preprocessing method called skew correction. This process will be done by correcting the skewness of the image being taken. While extracting the information from the scanned image, detecting & correcting the skew is crucial. The most widely used technique for this is the projection profile method. One such method is done by using the axes-parallel bounding box method proposed by Mahnaz Shafii and Maher Sid-Ahmed [16]. The method will firstly, calculate the area of an axes-parallel rectangle bounding box (Fig. 1), and then rotate only the outermost pixels of a text using a simple numerical method to fix the skewness of the image.

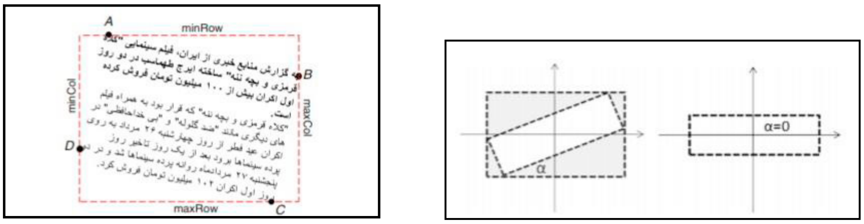
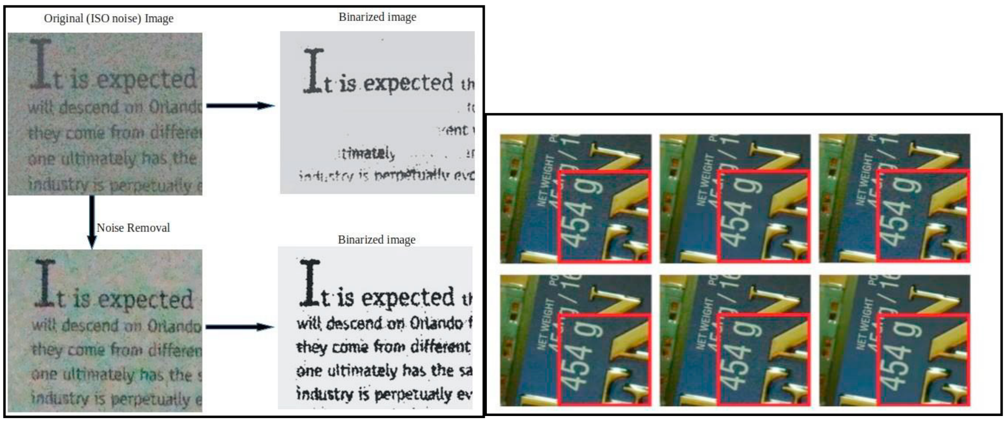


Fig. 1. Detecting the corner pixel of a text [16] (left), Transforming a skewed image by using an axes-parallel bounding box [16] (right)

B.3. Noise Removal

Noise removal is used to create a smooth-like image. This process is done by simply deleting patches that have higher intensity compared to the rest of the image. Noise is made to create a more refined and smoother image so that text in images can be viewed better.



(a) (b)

Fig. 2. (a) Example of denoising an image. From left to right and top to bottom: the noise image, the ground truth, denoising by BM3D, GCBD, Noise2Noise and the proposed method of Nam [18]. (b) Comparing ISO noise image and pre-processed image.

B.4. ASCII Braille

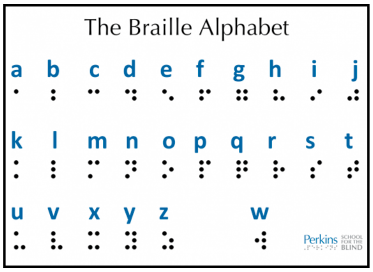


Fig. 3. Braille Alphabet [24]

The 64 ASCII characters between 32 and 95 inclusive are used for Braille ASCII. In ASCII, all capital letters correspond to their uncontracted English Braille [18]. Unlike regular print, each letter of the alphabet has only one Braille symbol. As a result, unless followed by a capitalization symbol, letters in Braille are lowercase by default [18]. When viewing Braille ASCII in a word processor, it appears as a jumbled mess of letters, numbers, and punctuation. Because of that we need to use special fonts that are available to see and print Braille ASCII [18]. Braille ASCII was originally used to store and transmit six dot Braille in a digital format. Because of that, we have decided to convert machine-edible text into Braille ASCII mappings for Unicode Braille characters.

C. Comparison to Previous Studies

Various research studies of OCR have been conducted in the past. These studies mainly focus on integrating the OCR algorithm to discern a character from images, or in other words, read text from images and videos with different languages. A study proposed by a team in a Kolkata college, RCC Institute of Information Technology, is one of the distinct research studies conducted on this subject [20]. They discuss on how to read text using OCR, that is by first converting the scanned image into binary so that the machine will be able to start processing the characters. From there, they extract and separate the words in each line, which is similar to our method. They were able to achieve, on average, 80-90% accuracy in five cases, not to mention the different fonts they have to process.

Another important OCR research study was proposed by Xiang Tong and David A. Evans of the Carnegie Mellon University's Computational Linguistics Laboratory [21]. Their study focuses on fixing OCR word errors based on the context of the previous words because sometimes "found" could be falsely interpreted as "fornd". At the end of their study, the overall word error correction or reduction the OCR achieved was 60.2%.

There are also some studies conducted on the same subject but with preprocessing beforehand. The preprocessing is meant to make the image much clearer for the machine to discern each letter in a word. One of the preprocessing methods is denoising, which means making the text sharper and reducing the blurs. Before the denoising, however, the three researchers compressed the images into JPEG format to significantly increase the denoising effect [17]. Researchers Nikolaos Ntogas and Dimitrios Ventzas came up with the same conclusion with their study on denoising as well. Only difference is that they used the background surface thresholding (BST) to help produce fewer OCR errors than Niblack's method [15].

Another group of researchers, Yang and Yan (2000) achieved binarization by using a logical adaptive thresholding method [22]. Their binarization study shows that gray-scale document images in poor condition can be greatly enhanced with complex signal-dependent noise. A group of students in Bangladesh University of Engineering and Technology proposed a method to help visually impaired people by creating a low-cost text-to- Braille scanner. They achieved this by using Google's open-source OCR engine with a smartphone app that will capture the image and turn the image into machine-edible text and then send that raw text to a Braille display by using bluetooth which then will interpret the text into Braille [25].

3. Methodology

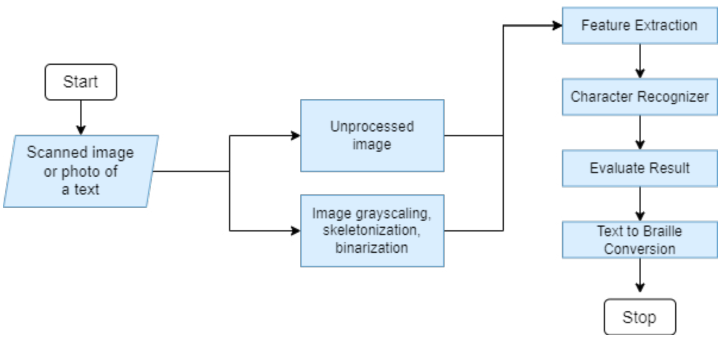


Fig. 4. Workflow Model

A. Datasets

The dataset we use for this research mainly is from images that contain text. Since this research paper's focus is on aiding the visually impaired to access books that do not have braille translations, the dataset we use is extracted from a GitHub Repository that provides scanned pages of books [23]. The books are mainly physical ones, which have texts printed on the paper with ink. With this in mind, we also have to account for the quality of the paper, that includes background color, the typeface, and the size of the font. The transparency of the pages also matters. Not so seldom that we get see-through pages in books (as we can see in Fig. 5), making it hard to read due to the back page's text below the current page's text.

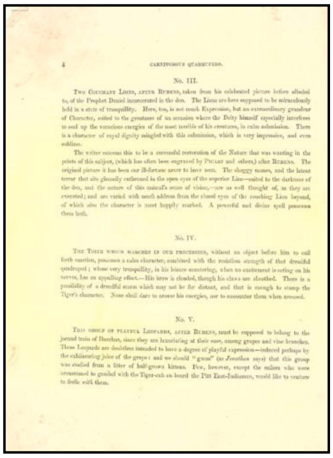
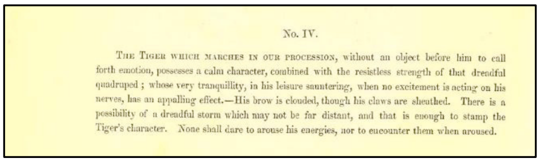


Fig. 5. An old scan of a book page.

A. Preprocessing (Grayscaling, Skeletonization and binarization of image & Raw unprocessed image)

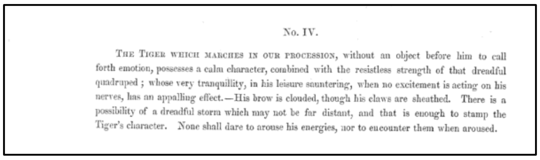
To test how high the accuracy of OCR is in different conditions, we make several modifications to the scanned images first. The purpose of this is to anticipate pages with mistyped words that made it into the final print or physical defect of the pages, such as getting wet, decolorized due to age, degraded, et cetera. Some of the modifications that we apply to the images are gray scaling, skeletonization, and binarization. The preprocessing steps are done by selecting random images from the dataset and creating a new image which is gray scaled, skeletonized and binarized while also leaving the raw original image (unprocessed).



No. IV.

THE TIGER WHICH MARCHES IN OUR PROCESSION, without an object before him to call forth emotion, possesses a calm character, combined with the resistless strength of that dreadful quadruped ; whose very tranquillity, in his leisure sauntering, when no excitement is acting on his nerves, has an appalling effect.—His brow is clouded, though his claws are sheathed. There is a possibility of a dreadful storm which may not be far distant, and that is enough to stamp the Tiger's character. None shall dare to arouse his energies, nor to encounter them when aroused.

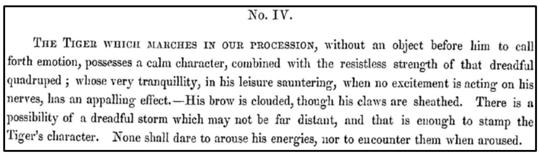
(a)



No. IV.

THE TIGER WHICH MARCHES IN OUR PROCESSION, without an object before him to call forth emotion, possesses a calm character, combined with the resistless strength of that dreadful quadruped ; whose very tranquillity, in his leisure sauntering, when no excitement is acting on his nerves, has an appalling effect.—His brow is clouded, though his claws are sheathed. There is a possibility of a dreadful storm which may not be far distant, and that is enough to stamp the Tiger's character. None shall dare to arouse his energies, nor to encounter them when aroused.

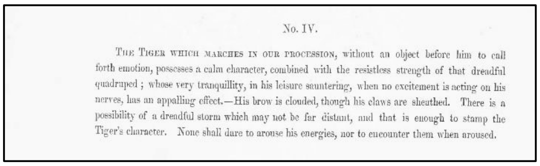
(b)



No. IV.

THE TIGER WHICH MARCHES IN OUR PROCESSION, without an object before him to call forth emotion, possesses a calm character, combined with the resistless strength of that dreadful quadruped ; whose very tranquillity, in his leisure sauntering, when no excitement is acting on his nerves, has an appalling effect.—His brow is clouded, though his claws are sheathed. There is a possibility of a dreadful storm which may not be far distant, and that is enough to stamp the Tiger's character. None shall dare to arouse his energies, nor to encounter them when aroused.

(c)



No. IV.

THE TIGER WHICH MARCHES IN OUR PROCESSION, without an object before him to call forth emotion, possesses a calm character, combined with the resistless strength of that dreadful quadruped ; whose very tranquillity, in his leisure sauntering, when no excitement is acting on his nerves, has an appalling effect.—His brow is clouded, though his claws are sheathed. There is a possibility of a dreadful storm which may not be far distant, and that is enough to stamp the Tiger's character. None shall dare to arouse his energies, nor to encounter them when aroused.

(d)

Fig. 6 (a). An old scan of a book page. (b). Gray Scaled (c). Skeletonization. (d). Binarization with Otsu method

B. Feature Extraction and Character Recognition

This step is done by using Tesseract OCR in Python programming language. Comparing between the two Unprocessed and processed images. All the Python program does is basically convert all images into .txt file format which contains the translated text.

C. Result Evaluation

The result of the .txt file is then evaluated by simply calculating the rate of error and mistranslation within each .txt files. Further information about this step is listed in this paper at the result and evaluation section.

D. Converting text to Braille

To make it less time-consuming, we use an automated text-to-braille conversion with the machine. Fig. 7 below is the Braille transcript for the text we can see in Fig. 6. Some symbols are still not converted and left as they are since the Braille version that we used here does not cover all symbols in its Unicode.

Fig. 7. Text-to-Braille transcript.

E. Tesseract OCR

In this research, we decided to use Tesseract OCR because it is one of the most accurate open-source engines to pre-process and use OCR on. Tesseract is also very versatile in adapting to different kinds of images and contains many built-in functions which can help to further enhance the results of OCR text.

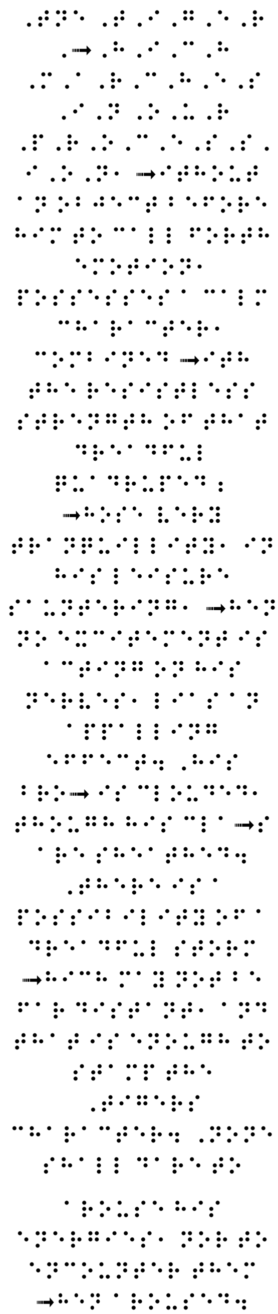
4. Results and Discussion

A. Raw Scan vs. Pre-Processed Text Comparison

A.1. Raw Scan

OCR	Braille conversion
Tne TIGER WHICH MARCHES IN OUR PROCESSION, without an object before him to call forth emotion, possesses a calm character, combined with the resistless strength of that dreadful quadruped ; whose very tranquillity, in his leisure sauntering, when no excitement is acting on his nerves, lias an appalling effect.— His brow is clouded, though his claws are sheathed. There is a possibility of a dreadful storm which may not be far distant, and that is enough to stamp the Tiger's character. None shall dare to arouse his energies, nor to eucounter them when aroused.	[illegible]

A.2. Pre-Processed (Gray-scaled, Skeletonized, and Binarized)

OCR	Braille conversion
Tne TIGER WHICH MARCHES IN OUR PROCESSION, without an object before him to call forth emotion, possesses a calm character, combined with the resistless strength of that dreadful quadruped ; whose very tranquillity, in his leisure sauntering, when no excitement is acting on his nerves, lias an appalling effect.— His brow is clouded, though his claws are sheathed. There is a possibility of a dreadful storm which may not be far distant, and that is enough to stamp the Tiger's character. None shall dare to arouse his energies, nor to encounter them when aroused.	

The results above are obtained through the sample dataset extracted from the GitHub Repository [23]. Translations of the text to braille is done by an automated text-to-Braille conversion. Some symbols are not converted since our model does not recognize certain symbols and characters. There are also some words that are not properly translated as they're special characters and words that are in bold or in italic format.

B. Table of Comparison

Table 3. Comparison

Techniques	Precision	Recall	F1 Score
Raw	97.62%	97.34%	0.97
Preprocessing	98,38%	98.06%	0.98

The result above is tested on samples based on different pages from the GitHub Repository [23]. We have found out that errors found when establishing the OCR are due to the fact that the model does not recognize bold and italic letters, causing some words to be mashed into a single word at certain times. Furthermore, special characters such as quotation marks and brackets cannot be recognized by our model. This resulted in some inconceivable words in the text format.

5. Conclusion And Future works

Pre-processing is an important part when converting text contained in an image to braille as shown in the results in this paper. By implementing various pre-processing techniques to an image, it results in a better input data, which subsequently improve the number of letters correctly transformed to text format. This is further reinstated with the results we have gotten in this paper as improvements can be found with an increase in the percentage of precision when applying pre-processing before translating the image to text. As a conclusion, pre-processing an image is one of the factors that can increase the accuracy of OCR-related words, especially in this case using OCR to translate document images into braille Unicode.

References

[1] Bourne, R. R. A., Adelson, J., & Flaxman, S. (2020). Global Prevalence of Blindness and Distance and Near Vision Impairment in 2020: Progress Towards the Vision 2020 Targets and What the Future Holds. *Investigative Ophthalmology & Visual Science Volume 61*. https://iovs.arvojournals.org/article.aspx?articleid=2767477. Accessed on 5:45 PM, March 1 2021.

[2] Kementerian Kesehatan RI. (2014). Buletin Jendela Data dan Informasi Kesehatan: Situasi Penyandang Disabilitas. Jendela.

[3] Editors of Encyclopaedia Britannica. (2019). *Braille*. Encyclopedia Britannica. https://www.britannica.com/topic/Braille-writing-system. Accessed on March 7 2021.

[4] Singla, S., & Yadav, R. (2014). Optical character recognition based speech synthesis system using labview. *Journal of Applied Research and Technology, 12*(5), 919-926. doi:10.1016/s1665-6423(14)70598-x

[5] Hamad, K. & Kaya, M. (2016). *A Detailed Analysis of Optical Character Recognition Technology.* International Journal of Applied Mathematics, Electronics and Computers. 4. 244-244. 10.18100/ijamec.270374.

[6] Rice, S.V., Jenkins, F.R. & Nartker, T.A. (1995). *The Fourth Annual Test of OCR Accuracy, Technical Report 95-03*. Information Science Research Institute. University of Nevada, Las Vegas.

[7] Mithe, R., Indalkar, S. & Divekar, N. (2013). Optical Character Recognition. *International Journal of Recent Technology and Engineering (IJRTE) Volume 2 Issue 1*. https://www.ijrte.org/wp-content/uploads/papers/v2i1/A0504032113.pdf. Accessed on 5:21 PM, March 9 2021.

[8] Nguyen, T., Jatowt, A., Coustaty, M., Nguyen, N., & Doucet, A. (2019). *Deep Statistical Analysis of OCR Errors for Effective Post-OCR Processing*.

[9] Hämäläinen, M., & Hengchen, S. (2019). *From the Paft to the Fiiture: A Fully Automatic NMT and Word-Embeddings Method for OCR Post-Correction*. Helsinki: University of Helsinki.

[10] Chiron, G., Doucet, A., Coustaty, M., Visani, M., & Moreux, J. (2017). *Impact of OCR Errors on the Use of Digital Libraries*. New York: IEEE. https://doi.org/10.1109/jcdl.2017.7991582

[11] Breuel, T. M. (2017). *High Performance Text Recognition Using a Hybrid Convolutional-LSTM Implementation*. Santa Clara: NVIDIA. https://doi.org/10.1109/icdar.2017.12

[12] Sporici, D., Cușnir, E., & Boiangiu, C. (2020). *Improving the Accuracy of Tesseract 4.0 OCR Engine Using Convolution-Based Preprocessing.* https://doi.org/10.3390/sym12050715

[13] Patel, C., Patel, A. & Patel, D. (2012). Optical Character Recognition by Open-Source OCR Tool Tesseract: A Case Study. *International Journal of Computer Applications*.

[14] Harraj, A. & Rassouni, N. (2015). OCR Accuracy Improvement on Document Images Through a Novel Pre-Processing Approach. *Signal & Image Processing : An International Journal*. 6. 10.5121/sipij.2015.6401.

[15] Nikolaos, N., & Dimitrios, V. (2009). Binarization of Pre-Filtered Historical Manuscripts Images. *International Journal of Intelligent Computing and Cybernetics*, 2(1), 148–174. doi:10.1108/17563780910939282

[16] Shafii, M., & Sid-Ahmed, M. (2014). Skew Detection and Correction Based on an Axes-Parallel Bounding Box. *International Journal on Document Analysis and Recognition (IJDAR)*, 18(1), 59–71. doi:10.1007/s10032-014-0230-y

[17] Yahya, A. A., Tan, J., Su, B. et al. (2020). BM3D Image Denoising Algorithm based on an Adaptive Filtering. *Multimed Tools and Applications*. https://doi.org/10.1007/s11042-020-08815-8

[18] Nam, S., Hwang, Y., Matsushita, Y. & Kim, S. J. (2016). A Holistic Approach to Cross-channel Image Noise Modeling and its Application to Image Denoising. *In Computer Vision and Pattern Recognition*, pg. 1683– 1691.

[19] Stella, J. & Valsan, K. S. (2018). Text to Braille Conversion: A Survey. *International Journal of Management and Applied Science (IJMAS)*, pp. 15-18, Volume 4, Issue 1

[20] Jana, R., Chowdhury, A. R. & Islam, S. M. (2014). Optical Character Recognition from Text Image. *International Journal of Computer Applications Technology and Research*, *3(4):240-244, Volume 3, Issue 4*. doi: 10.7753/IJCATR0304.1009

[21] Tong, X. & Evans, D. A. (1996). *A Statistical Approach to Automatic OCR Error Correction in Context*, p. 88-89, 96-97. Pittsburgh: Carnegie Mellon University.

[22] Yang, Y. & Yan, H. (2000). Pattern Recognition. *An Adaptive Logical Method for Binarization of Degraded Document Images*. https://doi.org/10.1016/S0031-3203(99)00094-1

[23] Barcha, P. (2017). Old Scanned Book. GitHub Repository. https://github.com/PedroBarcha/old-books-dataset

[24] *How Braille Alphabet Works*. Perkins School of the Blind. https://www.perkins.org/stories/how-the-braille-alphabet-works. Accessed on 12:32 AM, 28 March 2021.

[25] Hossain, S., Raied, A. A., Rahman, A., Abdullah, Z. R., Adhikary, D., Khan, A. R., ... & Fattah, S. A. (2018, October). *Text to Braille Scanner with Ultra Low Cost Refreshable Braille Display*. In 2018 IEEE Global Humanitarian Technology Conference (GHTC) (pp. 1-6). IEEE