

The Effect of Font Variation in the Accuracy of Image to Text Conversion

Vera Firmansyah and Amalia Rakhmawati

Academy of Metrology and Instrumentation, Ministry of Trade, Bandung, Indonesia

Keywords: Font, OCR, OpenCV, Accuracy.

Abstract: Image processing is using both hardware and software as tools to analyze and as an interface to process an image. These tools are able to improve the welfare and the quality of life of people with visual impairments to help them to read articles. The level of impaired vision can vary from person to person. Thus, this research develops initiate step in image to text conversion with font variation. Image to text conversion is done by extracting text from images obtained through the camera from the article. Previous research used microprocessor equipped with a camera module and the Tesseract OCR in the Python Pence. The Tesseract OCR program on Pence is an open source program used to extract text from images and save it in the form of a text. This research using 5 font variation chosen randomly which are Times New Roman, Arial, Calibri, Comic Sans and Courier New. Image Processing using Dilation, Crop, Canny and Median Blur. The result shows that Comic Sans Font has the highest accuracy and Times New Roman has the lowest accuracy. Comic Sans has the highest accuracy because the overall font does not have much curves than the other while Times New Roman Font has the lowest accuracy because it has more curve characteristics.

1 INTRODUCTION

Image processing is using both hardware and software as tools to analyze and as an interface to process an image. In 2015 it is estimated that of the 7.33 trillion world population, there are 253 million people (3.38%) who suffer from visual disturbances, consisting of 36 million people experiencing blindness, 217 million experiencing moderate to severe visual impairment. In addition, there are 188 million people with mild visual disturbances (M. Patil and R. Kagalkar, 2014). The classification of visual impairments used is in accordance with the WHO classification, which is based on visual acuity (Ministry of Health of the Republic of Indonesia, 2018).

Therefore, those tools are able to improve the welfare and the quality of life of people with visual impairments to help them to read articles. The level of impaired vision can vary from person to person. So this research develops initiate step in image to text conversion with font variation. Image to text conversion is done by extracting text from images obtained through the camera from the article. Previous research used microprocessor equipped with a camera module and the Tesseract OCR

(Optical Character Recognition) program in the Python OpenCV (Open Computer Vision) programming (Rithika, H., B. N. Santhoshi, 2016).

OpenCV is an API (Application Programming Interface) library used because it has familiarity with computer vision image processing. Computer vision is a branch of image processing field which allows computers to see like humans. With computer vision, the computer can make decisions, take action, and recognize objects. Some of the implementations of computer vision are face recognition, face detection, face / project tracking, road tracking, etc. (Widja. I. B. P., 2017). The Tesseract OCR program on OpenCV is an open source program used to extract text from images and save it in the form of a text. Fig. 1 shows the Tesseract OCR program on OpenCV to convert image to text.

This research aims to get the effect of font variation in the accuracy of image to text conversion. The fonts are chosen randomly which are Times New Roman, Arial, Calibri, Comic Sans and Courier New.

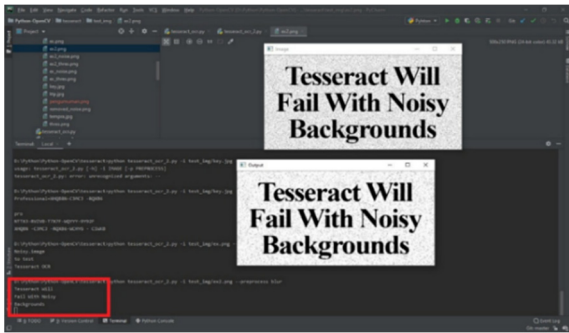


Figure 1: Tesseract OCR Example.

2 DIGITAL IMAGE

When the light source hits the object, the object will reflect some of the light back. The reflection of light will be captured by an optical sensing device i.e. digital camera, then the image of the object is received by the sensor according to the intensity of the reflected light and will be converted into a digital image. This study uses an active complementary metal-oxide-semiconductor (CMOS) sensor. With CMOS sensors the power consumption should be lower, integration capabilities should be high and should comes with lower price.

Digital image is a discrete data set in the form of a two-dimensional matrix where the numbers from the matrix indicate the brightness level of the points. Pixel is the representation of the smallest point in the digital image and the value in the coordinates (x, y) as shown in Fig. 2 which indicates the intensity value. For digital images with red, green, blue (RGB) coordinates, the image has an intensity value in the RGB color coordinates of each pixel.

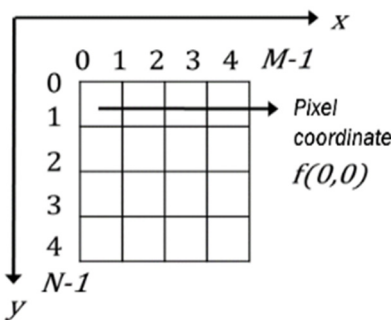


Figure 2: Pixel Coordinate (Rithika, H., B. N. Santhoshi, 2016).

3 IMAGE PROCESSING

This research using image processing steps shown in

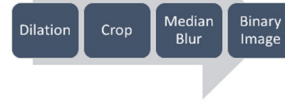


Figure 3: Image Processing.

Sample is taken in a constant lighting of 5 lumen, the font size is 12 and the font variation is Times New Roman, Arial, Calibri, Comic Sans and Courier New. Sample taken 5 times to get repeatability. The image taken from camera is shown in Fig. 4.

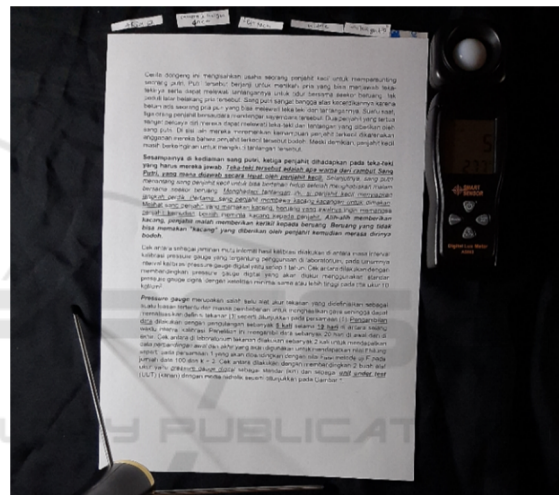


Figure 4: Digital Sample.

Sample is using words and sentences in Fig. 5.

Cerita dongeng ini mengisahkan usaha seorang penjahit kecil untuk mempersunting seorang putri. Putri tersebut berjanji untuk menikahi pria yang bisa menjawab teka-tekiya serta dapat melewati tantangannya untuk tidur bersama seorang beruang, tak peduli latar belakang pria tersebut. Sang putri sangat bangga atas kecerdikannya karena belum ada seorang pria pun yang bisa melewati teka-teki dan tantangannya. Suatu saat, tiga orang penjahit bersaudara mendengar sayembara tersebut. Dua penjahit yang tertua sangat percaya diri mereka dapat melewati teka-teki dan tantangan yang diberikan oleh sang putri. Di sisi lain mereka meremehkan kemampuan penjahit terkecil dikarenakan anggapan mereka bahwa penjahit terkecil tersebut bodoh. Meski demikian, penjahit kecil masih berkeinginan untuk mengikuti tantangan tersebut.

Sesampainya di kediaman sang putri, ketiga penjahit dihadapkan pada teka-teki yang harus mereka jawab. **Teka-teki tersebut adalah apa warna dari rambut Sang Putri, yang mana dijawab secara tepat oleh penjahit kecil.** Selanjutnya, sang putri menantang sang penjahit kecil untuk bisa bertahan hidup setelah menghabiskan malam bersama seekor beruang. **Menghadapi tantangan ini, si penjahit kecil menyiapkan langkah cerdik. Pertama, sang penjahit membawa kacang-kacangan untuk dimakan.** Melihat sang penjahit yang memakan kacang, beruang yang awalnya ingin memangsa penjahit kemudian berubah meminta kacang kepada penjahit. **Alih-alih memberikan kacang, penjahit malah memberikan kerikil kepada beruang. Beruang yang tidak bisa memakan "kacang" yang diberikan oleh penjahit kemudian merasa dirinya bodoh.**

Cek antara sebagai jaminan mutu internal hasil kalibrasi dilakukan di antara masa interval kalibrasi pressure gauge yang tergantung penggunaan di laboratorium, pada umumnya interval kalibrasi pressure gauge digital yaitu setiap 1 tahun. Cek antara dilakukan dengan membandingkan pressure gauge digital yang akan diukur menggunakan standar pressure gauge digital dengan ketelitian minimal sama atau lebih tinggi pada titik ukur 10 kg/cm².

Pressure gauge merupakan salah satu alat ukur tekanan yang didefinisikan sebagai suatu luasan tertentu dan massa pembebanan untuk menghasilkan gaya sehingga dapat merealisasikan definisi tekanan [3] seperti ditunjukkan pada persamaan (1). **Pengambilan data dilakukan dengan pengulangan sebanyak 5 kali selama 10 hari** di antara selang waktu interval kalibrasi. Penelitian ini mengambil data sebanyak 20 hari di awal dan di akhir. Cek antara di laboratorium tekanan dilakukan sebanyak 2 kali untuk mendapatkan data **perbandingan awal dan akhir** yang akan digunakan untuk mendapatkan nilai F hitung seperti pada persamaan 1 yang akan dibandingkan dengan nilai F_{tabel} metode uji F pada jumlah data 100 dan k = 2. Cek antara dilakukan dengan membandingkan 2 buah alat ukur yaitu **pressure gauge digital** sebagai standar (kiri) dan sebagai **unit under test** (UUT) (kanan) dengan media hidrolik seperti ditunjukkan pada Gambar 1.

Figure 5: Words and Sentences in Sample.

The image from the last step of image processing in this reasearch is binary image shown in Fig. 6.

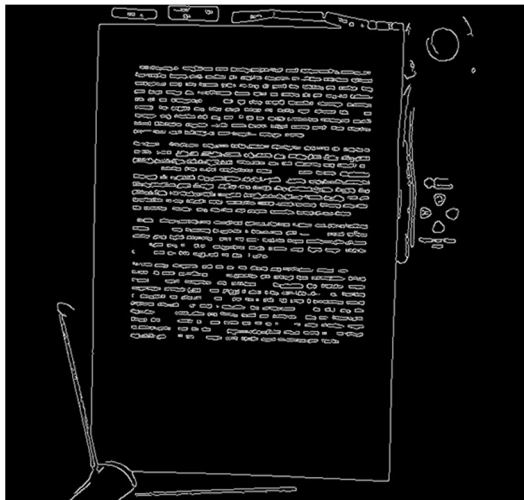


Figure 6: Binary Image of Sample.

4 TESSERACT OCR

The binary image taken is the input of Tesseract OCR, then on the Page Layout Analysis with analysis of connected component to find where the component outline is stored. The outline is gathered together to form a blob. The blob is the area of the image that overlaps together. Then the blobs organized into a text line, and the lines and regions are analyzed to find fixed pitch and proportional writing (Smith. R., 2007). Posts with fixed pitch are broken down into character cells. Proportional writing is divided into words using defined spaces and fuzzy spaces. Furthermore, image word recognition is carried out in two stages called pass-two (Smith. R., 2007).

The first pass is made to recognize each word. The words that pass the first pass are words that match the dictionary and are passed on to the adaptive classifier to be used as training data. After sufficient samples, this adaptive classifier can also provide classification results even on the first pass. Words that may not be recognized or missed on the first pass will be continued in the pass two process. In this condition, the adaptive classifier that has received more information on the first pass will be more able to recognize words that were missed or less recognized before (Smith. R., 2007).

According to (Smith, 2007) several steps taken by Tesseract for character recognition are as follows:

1. Line and Word Search
2. Character and Word Introduction

5 RESULT AND ANALYSIS

The result of image to text conversion with font variation is shown in Table 1. The articles consist of 4 paragraphs in Latin text with bold, italic and underline text combined.

Accuracy of the image to text conversion depends on how much the Tesseract OCR recognize the font with bold, italic and underline text combined. Also the image processing to get binary image which affects the input of Tesseract OCR The result shows that Comic Sans Font has the highest accuracy and Times New Roman has the lowest accuracy. Comic Sans has the highest accuracy because the overall font does not have much curves and details than the other while Times New Roman Font has the lowest accuracy because it has more curve characteristics. The differences of Comic Sans and Times New Roman font shape shown in Table 2.

Table 1: Image to Conversion Result.

	Font	Accuracy (%)	Mean (%)
Para 1	<u>Times New Roman</u>	98.2	99.1
	Arial	99.1	
	Calibri	99.2	
	Comic Sans	99.6	
	Courier New	99.4	
Para 2	<u>Times New Roman</u>	97.4	98.9
	Arial	99.2	
	Calibri	99.2	
	Comic Sans	99.6	
	Courier New	99.1	
Para 3	<u>Times New Roman</u>	97.3	98.8
	Arial	99.1	
	Calibri	99.3	
	Comic Sans	99.6	
	Courier New	98.9	
Para 4	<u>Times New Roman</u>	97.1	98.8
	Arial	99.2	
	Calibri	99.1	
	Comic Sans	99.4	
	Courier New	99.2	
Total Mean			98.9

Table 2: Font Shape.

<p>Comic Sans the quick brown fox jumps over the lazy dog. THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.</p>	<p>Times New Roman the quick brown fox jumps over the lazy dog. THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.</p>
---	--

From Table 2, Times New Roman font has more details shape in each letter than Comic Sans. The details contain curve in the letter especially for the letter 'g' which the Tesseract OCR has not recognized the letter very well. Even though the performance represented by total mean of the system is 98.9 %, the system needs to be improved by applying image

processing technique specified in extracting curve (Z. Martinez) letter for each font styles.

REFERENCES

- M. Patil and R. Kagalkar. (2014). A Review on Conversion of Image to Text As Well As Speech Using Edge Detection and Image Segmentation. *International Journal of Science and Research (IJSR)*.
- I. Isewon, J. Oyelade, O. Oladipupo. (2014). Design and Implementation of Text to Speech Conversion for Visually Impaired People. *International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA*.
- S. Edward, A. Jothimani, V. Jayaprakash, J.B. Xavier. (2018). Text-To-Speech Device for Visually Impaired People. *International Journal of Pure and Applied Mathematics Volume 119 No. 15 2018, 1061-1067*.
- Ministry of Health of the Republic of Indonesia. (2018) . Situation of Vision Disorders. Infodatin of the Ministry of Health of the Republic of Indonesia.
- Rithika, H., B. N. Santhoshi. (2016). Image Text to Speech Conversion In The Desired Language By Translating With Raspberry Pi, *IEEE International Conference on Computational Intelligence and Computing Research (ICIC)*.
- Widja. I. B. P. (2017). Image Binarization Design and Text Character Recognition with Raspberry Pi, *National Conference on Systems & Information Technology*.
- Rakhmawati, A., Juliastuti. E., Umma. A. K., Sari. B. P. (2019). Prototype of Orifice Plate Diameter Measurement Based on OpenCV Image Analysis, *Instrumentation and Control Seminar*.
- Smith. R. (2007). An Overview of the Tesseract OCR Engine, *Ninth International Conference on Document Analysis and Recognition*.
- Z. Martinez. Curves Extraction in Images. *Revista Colombiana de Estadística*, January 2015, Volume 38, Issue 1, pp. 295 to 320.