TEXT DETECTION USING TEXT FLOW IN CERTIFICATE

Indra Rianto[1], Ednawati Rainarli[2]

[1,2]Program Studi Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatikukur 112-114 Bandung
Email: indrarianto9@gmail.com[1], ednawati.rainarli@email.unikom.ac.id[2]

ABSTRACT

This study uses a scan of the certificate to detect the text contained in it, there are different font sizes and fonts on certificates are not the usual fonts used on ordinary documents, but fonts that are quite similar to handwriting. The results of the text detection are dependent by preprocessing and the test data used.

In this study, the text detection method used was Text Flow and letter classification using Tesseract. Before carrying out the classification process, the image goes through preprocessing method. Then the sample used for test data is three certificates with different events. Based on the results carried out using three combinations, the best results are obtained is 67% for Recall, 58% for Precision and 62% for F-Score with the combination used is the threshold value in 128 with min confidence score 50, horizontal distance is less than 2, vertical distance is less than 0.2 and the size of the letters is less than 2. The results of the Text Flow method on this certificate are less than 70% but the Text Flow method on the certificate is able to not select objects other than text such as signatures and logos.

Keyword: Text Detection, Certificate, Text Flow, Min-Cost Flow, Segmentation.

1. INTRODUCTION

The need to detect text on certificates is very important, one of the factors driving the need to detect text on certificates is technological progress. At this time the certificate is not only needed in print (hardcopy), but also a file (softcopy) of the certificate is needed to support several things such as speed of delivery, security and integrity of the information contained in the certificate. The certificate contains important information which generally consists of letterhead, letter number, participant's name, time and place, and endorsement [1]. Research for the text detection in the previous certificate was conducted by Reza Yogi Andria who concluded that there was a problem in the segmentation section. The use of the Connected Component method in this study caused the backgrounds in the certificate will be detected as objects / letters.

The research on Optical Character Recognition (OCR) by Raden Sofian Bahri and Irfan Maliki entitled *Perbandingan Algoritma Template Matching dan Feature Extraction pada Optical Character Recognition* concluded that based on accuracy, development, and time, the feature extraction algorithm is superior to the template matching algorithm on OCR [2].

Segmentation has a different approach process depending on the type of input received, each input has a different approach to the application of the segmentation process. Although they have differences, the goal to be achieved is the same, namely identification or recognition. In a certificate, object recognition or identification aims to recognize the text contained in a certificate. The prevalent scene text detection approach follows four sequential steps comprising character candidate detection, false character candidate removal, text line extraction, and text line verification. However, errors occur and accumulate throughout each of these sequential steps which often lead to low detection performance. To address these issues, a unified scene text detection system namely Text Flow used utilizing the minimum cost (min-cost) flow network model is used.

With character candidates detected by cascade boosting, the min-cost flow network model integrates the last three sequential steps into a single process which solves the error accumulation problem at both character level and text line level effectively. Text Flow technique has been tested on three public datasets, i.e, ICDAR2011 dataset, ICDAR2013 dataset and a multilingual dataset and it outperforms the state-of-the-art methods on all three datasets with much higher recall and F-score from Baseline, Huang et al, dan Neuman and Matas [3]. Therefore, this research will use the Text Flow method with the aim of measuring the success of the textflow method in character detection.

2. THE CONTENT OF RESEARCH

2.1. Problem Analysis

The problem is how to detect the text that is on the certificate because not all certificates use letters that are used to write in a document, but the typeface is quite similar to handwriting. The method used for the classification of the certificate letter is Text Flow. In previous studies Text Flow has been implemented for text recognition in image results [3]. Text Flow technique has been tested on three public datasets, i.e, ICDAR2011 dataset, ICDAR2013 dataset and a multilingual dataset and it outperforms the state-of-the-art methods on all three datasets with much higher recall and F-score from Baseline, Huang et al, dan Neuman and Matas [3]. Therefore, this research will use the Text Flow method with the aim of measuring the success of the textflow method in character detection.

2.2. System Analysis

The system that will be built have input, process, and output. The input is certificate that used as test data. The following is the process that occurs in the system seen in Fig 1.

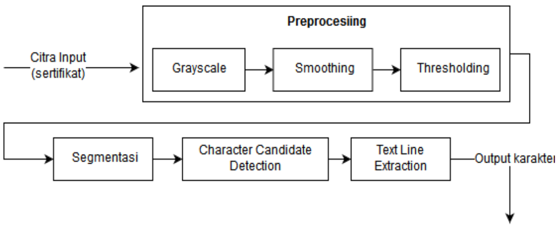

Fig 1. Block Diagram of Main Process System

2.3. Analysis of Data Input

In the system, the input data is an image. Input data is input image in the form of certificate scan results which are used as test data. The following is an example of input data in Fig 2.

Fig 2. Test Data

2.4. Method Analysis

In this section that explains the method analysis that occurs in the implementation of the Text Flow method on the certificate.

2.4.1. Pre-processing

The stages of preprocessing consist of grayscale, smoothing, thresholding and segmentation.

a. Grayscale

Grayscale process is carried out to get grayscale images or images with gray colors. The formula used is "luma" or "luminance". Following is the grayscale formula in equation 1 below.

$$Gray = 0{,}299 * red + 0{,}587 * green + 0{,}114 * blue \qquad (1)$$

b. Smoothing

Smoothing is using integral sum image with the aim of improving image quality and eliminating noise. At this stage a mask is used as a calculation range with neighbor values [4]. Made of integran sum image $g_{x,y}$ form grayscale image $G_{i,j}$ with size as $M \times N$ used in equation 2 below.

$$g_{x,y} = \sum_{i=0}^{x}\sum_{j=0}^{y} G_{i,j}\,, x = 0,1,2,..,M-1, y = 0,1,2,..,N-1 \qquad (2)$$

Simplification will be done in equation 2 for more effective computing. Following is the simplification of equation 2.

$$g_{0,y} = G_{0,y} + g_{0,y-1} \qquad y=1,2,..,N-1 \qquad (3)$$

$$g_{x,y} = G_{x,0} + g_{x-1,0} \qquad x=1,2,..,M-1 \qquad (4)$$

$$g_{x,y} = G_{x,y} + g_{x,y-1} + g_{x-1,y} - g_{x-1,y-1}\,, x=1,2,..,M-1, \quad y=1,2,..,N-1 \qquad (5)$$

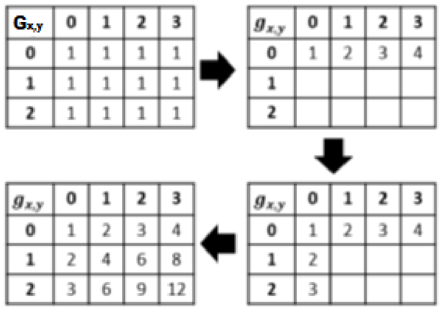

Fig 3. Step of integral sum calculation

Determine the dimensions of the mask that will affect the amount of intensity obtained from the mask area. The size of the mask is 3 x 3, with this size obtained by a neighbor of 8 pixels.

Calculation of the average value of integral images in the range of mask dimensions with equation 6 below.

$$mean(x,y) = \frac{s(x,y)}{wx \times wy}, \quad x = 0,1,2..M-1, \quad y = 0,1,2..N-1 \qquad (6)$$

where s(x,y) can be calculated with equation 7 below.

$$s(x,y) = (g_{x+dx-1,y+dy-1} + g_{x-dx,y-dy}) - (g_{x-dx,y+dy-1} + g_{x+dx-1,y-dy}) \quad \text{Dengan } dx = round\left(\frac{wx}{2}\right), \quad dy = round\left(\frac{wy}{2}\right), \qquad (7)$$

In this study, a technique for finding cells or pixels in a mask that has no value is proposed by adding an artificial function (x,y) to Equation 6 so that it becomes like Equation 8 below.

$$mean(x,y) = \frac{s(x,y)}{wx \times wy} + a(x,y), \quad x = 0,1,2..M-1, y = 0,1,2..N-1 \tag{8}$$

value of $a(x,y)$ can be calculated with equation 9 below.

$$a(x,y) = \begin{cases} \frac{nw(x,y)}{wx \times wy}, & if \begin{pmatrix} (x-nr<0 \; or \; x+nr \geq M) \\ or \; (y-nc<0 \; or \; y+nc \geq N) \end{pmatrix} \\ 0, & sebaliknya \end{cases}$$

$$with \; nr = trunc(\frac{wx}{2}) \; and \; nc = trunc(\frac{wy}{2}) \tag{9}$$

where $nw(x,y)$ can be calculated with equation 10 below.

$$nw(x,y) = \begin{cases} \left(rp(x) \times wy + cp(y) \times (wx - rp(x))\right) \times 255, & \begin{array}{l} if (x-nr<0 \; and \; y-nc<0) \\ or \; (x-nr<0 \; and \; y+nc \geq N) \\ or \; (x+nr \geq M \; and \; y-nc<0) \\ or \; (x+nr \geq M \; and \; y+nc \geq N) \end{array} \\ rp(x) \times wy \times 255, & if \begin{pmatrix} (x-nr<0 \; or \; x+nr \geq M) \\ and \; (y>0 \; and \; y<N) \end{pmatrix} \\ cp(y) \times wx \times 255, & if \begin{pmatrix} (y-nc<0 \; or \; y+nc \geq N) \\ and \; (x>0 \; and \; x<M) \end{pmatrix} \end{cases} \tag{10}$$

(x) = Row padding, the number of lines in the mask that are outside the image dimensions are calculated based on the outer distance.

(y) = Column padding, the number of columns in the mask outside the dimensions of the image is calculated based on the outer distance.

$rp(x)$, and $cp(y)$) can be calculated with equation 11 bellow

$$rp(x) = \begin{cases} |x-nr|, & if(x-nr<0) \\ x+nr-M+1, & if(x+nr \geq M) \end{cases}$$

$$cp(y) = \begin{cases} |y-nc|, & if(y-nc<0) \\ y+nc-N+1, & if(y+nc \geq N) \end{cases} \tag{11}$$

c. Thresholding

Thresholding process is processing grayscale images into binary images or black and white images. The threshold value used in this study was 128 [5]. Berikut algoritma yang digunakan pada persamaan 12 dibawah ini.

```
if (gray > 128) {
        val = 0
} else {
        val = 1
}
```
(12)

First of all, take pixels starting from pixels (0,0). After taking the pixels, do the calculation using equation 2 above. Do it until the last pixel in the image.

d. Segmentation

The segmentation process is used to divide the image into base elements according to specified criteria. The method used in the segmentation process is *Connected Component Labeling* (CCL) [6]. These are the steps of the method.

1) Define variables.
   Matrix for storing object coordinates = $pt_{i,j}$
   Matrix for storing object coordinate boundaries = $br_{i,j}$
   One dimension array to temporarily store x and y coordinate values = c
2) The search starts from the coordinates (0,0) until finding a matrix with value is one (1).
3) Save the coordinates in the matrix $pt_{i,j}$, then change the value of the matrix one (1) with a zero (0).
4) Save the coordinate boundary of the object into the matrix $br_{0,j}$. Because the variable does not currently have a value, the upper and lower limits will be filled with coordinate *x* values, Left border and right border will be filled with coordinates *y*.
5) Next find the pixel value of the neighboring matrix with a center coordinate (3,3) with a dimension of mask is 5 X 5.
6) Save the pixel coordinates with one (1) value into the matrix $pt_{i,j}$ then initialize the values in matrix coordinate one (1) dewith a zero (0).
7) Use coordinates with the value one in the neighbor matrix, initialize the x coordinate to the variable C_0 and the y coordinate on the variable C_1. Next do a comparison of values on the matrix $br_{0,j}$ from the previous data with coordinates of value one in the neighbor matrix. This process obtains the outer boundary value for an object.
8) The process in steps number 5 to 7 will be repeated using the next coordinate in the matrix $pt_{0,j}$ until there are no more coordinates that can be processed on the matrix.

2.4.2. Min-Cost Flow Network

At this stage will discuss the design, determination of the cost of each prospective candidate character and implementation of the Min-Cost Flow Network. Candidate characters detected by the segmentation process will be processed by the tesseract before entering the Min-Cost Fow Network stage, this is used to check each prospective character, whether it is a character or not. If the result of the tesseract states that a candidate character is a true character, then the character will be marked with a green colored box, if considered not a character it will be marked with a red boxes. Then the green box will be processed for the Min-Cost Flow Network process.

1. Before the formation of the text will go through the checking process, to find out the input image has more than one candidate text line or not.

Checking is done based on the Y axis on the character results of segmentation in the previous process. Following is an example to check rows with examples in table 1 below.

Table 1 Candidate Character I as A (red) and Candidate Character N as B (yellow)

x/y	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
3	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
4	0	0	1	1	0	0	0	1	1	1	0	0	1	1	0
5	0	0	1	1	0	0	0	1	1	1	0	0	1	1	0
6	0	0	1	1	0	0	0	1	1	1	1	1	1	1	0
7	0	0	1	1	0	0	0	1	1	1	1	1	1	1	0
8	0	0	1	1	0	0	0	1	1	1	1	1	1	1	0
9	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0
10	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0
11	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0

From the table, let's say that b1 is marked with a red box and b2 is marked with a yellow.

a. Information of b1
   b1.y = 3
   b1.h = 9
b. Information of b2
   b2.y = 3
   b2.h = 9

Following are the functions used to do line checking.

Box boxA; Box boxB;
if (b1.y > b2.y) {
boxA = b2; boxB = b1;
} else {
boxA = b1; boxB = b2;
}
return (boxB.y >= boxA.y && boxB.y <= (boxA.y + boxA.h)) || (((boxB.y + boxB.h) >= boxA.y && (boxB.y + boxB.h) <= (boxA.y + boxA.h)));

if this function result is true, then the two objects that are entered into the function will be considered as one line, otherwise it will be considered as a different line.

2. Based on the assumption that all lines of text start from left to right, all candidate characters will be sorted first to get their horizontal coordinates. Candidates for characters are obtained from the results of previous segmentation process. Here is a picture for designing a min-cost flow network in Fig 4 .

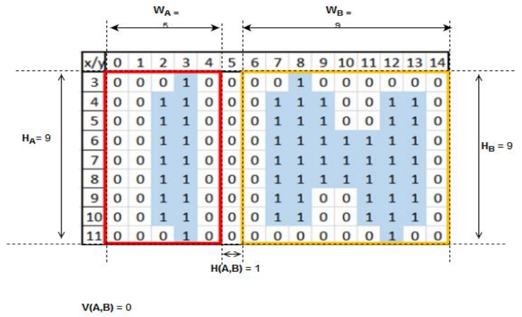

Fig 4 Illustration of Distance Beteween Candidate Character

For each candidate character A, the next candidate character B, which can be connected by A, must be limited by certain conditions to reduce errors in calculating the distance between characters A and B. The following conditions must be checked and followed in table 2 below.

Table 2 Condition of Distance Between Candidate Character

No	Kondisi	Syarat
1.	Horizontal distance between A and B	$\frac{H(A,B)}{\min(W_{A,}W_b)} < T_H$
2.	Vertical distance between A and B	$\frac{V(A,B)}{\min(W_{A,}W_b)} < T_V$
3.	Size of A and B	$\frac{\lvert W_A - W_b \rvert}{\min(W_{A,}W_b)} < T_S$

After all checks have been made, and all conditions are met, it can be interpreted that the two characters A and B are neighbors and can be interpreted as a text.

2.5. Accuracy Testing

In accuracy testing, the accuracy or compatibility value of the text recognition application on the certificate will be calculated.

In this study the accuracy testing used several combinations of parameters in the test data. The table 3 below is a combination of parameters used.

Table 3 Combination of Testing Parameters

Pengujian Ke-	**Min *Confidence Score***	**Nilai T_H, T_V dan T_S**
1	65	(2 , 0.6 , 0.2)
2	60	(2 , 0.6 , 1)
3	50	(2, 0.6 , 2)

2.5.1 First Combination Accuracy Test Result

Testing the accuracy of the first combination is done by using a combination of parameters as follows.

Min Confidence Score value : 65

TH, TV dan TS value : 2 , 0.6 , 0.2

Based on testing on the application using the parameters above, then table 4 below is the test result.

Table 4. the First Test Result

No	Image Name	Recall	Precision	F-score
1	dataUji1	0	0	0
2	dataUji2	0.09	0.11	0.099
3	dataUji3	0.09	0.11	0.099
average		**0.06**	**0.07**	**0.066**

2.5.2 Second Combination Accuracy Test Result

Testing the accuracy of the second combination is done by using a combination of parameters as follows.

Min Confidence Score value : 60

TH, TV dan TS value : 2 , 0.6 , 1

Based on testing on the application using the parameters above, then table 5 below is the test result.

Table 5. The Second Test Result

No	Image Name	Recall	Precision	F-score
1	dataUji1	0.125	0.14	0.12
2	dataUji2	0.047	0.11	0.06
3	dataUji3	0.5	0.66	0.56
average		**0.224**	**0.3**	**0.25**

2.5.3 Third Combination Accuracy Test Result

Testing the accuracy of the third combination is done by using a combination of parameters as follows.

Min Confidence Score value : 50

TH, TV dan TS value : 2 , 0.6 , 2

Based on testing on the application using the parameters above, then table 6 below is the test result.

Table 6. The Third Test Result

No	Image Name	Recall	Precision	F-score
1	dataUji1	0.57	0.57	0.57
2	dataUji2	0.66	0.54	0.59
3	dataUji3	0.77	0.63	0.69
average		**0.67**	**0.58**	**0.62**

Testing is done one by one on each test data, so that the average accuracy obtained from each test. The following is a table of test results obtained in table 7.

Table 7. Accuracy Results for Each Combination

Test Result	***Confidence Score*, T_H, T_V dan T_S**	Recall	Precision	F-score
1	(65, 2 , 0.6 , 0.2)	0.06	0.07	0.066
2	(60, 2 , 0.6 , 1)	0.224	0.3	0.25
3	(50, 2 , 0.6 , 2)	0.67	0.58	0.62

Based on the table above, the best parameter combination is the third parameter combination with an average accuracy of Recall of 67% Precision of 58% and F-Score of 62%, The parameter combination used is the threshold value 128 with min confidence score 50, horizontal distance (T_H) 2, vertical distance (Tv) 0.2 and size between letters(T_S) is 2. Size values between letters (T_S) in this study is different from previous studies, i.e 0.2 [3], this is most likely due to the different types of letters used and the different sizes in a certificate.

The results obtained in testing is that there are several weaknesses of the method of character recognition or letters. Following are the weaknesses of the method implemented in this system:

1. In this study, the recognized object should only be letters and numbers, but after testing, the noise in the certificate is recognized as an object as well.
2. This method cannot handle attached text. Because the certificate, the type of letter that there is sometimes a letter attached to another letter. The result is that if there are several letters attached, they will be read as one object / letter only.
3. This method cannot handle a letter that has two separate objects, for example like the letter i (lowercase I) and the letter j (lowercase J). Therefore, the letter is not recognized as one object but two objects.

3. CLOSURE

3.1. Conclusion

Based on the results of testing the system of Text Recognition Using the Text Flow Method on the Certificate, the best accuracy is obtained for Recall by 67% Precision by 58% and F-Score by 62%. This accuracy is influenced by preprocessing, character recognition and test data used. Because each certificate has a different letter size for each text, making it difficult to determine the exact distance to form a text. Detection of characters before entering the process of detecting text is also very influential because, if there is one letter that is not detected in the middle of a text then the text will be considered to be two different parts of the text.

3.2. Suggestion

So that further research on Text Recognition Using the Text Flow Method on Certificates can have higher accuracy, the following are suggestions that can be taken into consideration, namely improving the method of segmentation and recognition of letters that are attached, and recognizing a letter that has two separate objects

BIBLIOGRAPHY

[1] F. A. Prabowo and M. Syani, "Sistem Informasi Pengolahan Sertifikat Berbasis Web di Divisi Training Seamolec," *Jurnal Masyarakat Informatika Indonesia*, vol. 2, no. 1, pp. 73-81, 2017.

[2] R. S. Bahri and I. Maliki, "Perbandingan Algoritma Template Matching dan Feature Extraction pada Optical Character Recognition.," *KOMPUTA: Jurnal Komputer dan Informatika,* vol. 1, no. I, pp. 29-35, 2012.

[3] S. Tian, Y. Pan, C. Huang, S. Lu, K. Yu and a. C. L. Tan, "Text Flow: A Unified Text Detection System in Natural Scene Images," in *IEEE International Conference on Computer Vision*, 2015.

[4] T. Singh, S. Roy, O. Singh, T. Sinam and K. Singh, "A New Local Adaptive Thresholding Technique in Binarization," *International Journal of Computer Science,* vol. 8, no. 6, No 2, p. 271, 2011.

[5] P. Hidayatullah, Pengolahan Citra dan Aplikasinya, Bandung: Informatika, 2017.

[6] K. S. FU and J. K. MUI, "A Survey on Image Segmentation," *Pattern Recognition,* vol. 13, no. 1, pp. 3-16, 1981.

[7] A. Bissacco, M. Cummins, Y. Netxer and H. Neven, "Photo to OCR: Reading text in uncontrolled condition.," *International Conference on Computer Vision (ICCV),* pp. 785-792, 2003.

[8] R. K. Ahuja, T. L. Magananti and J. B. Orlin, Network flows: Theory, algorithm, and application, 1993.

[9] X. Chen and A. L. Yuille, "Detecting and Reading Text in Natural Scenes," *Computer Vision and Pattern Recognition (CVPR),* no. 2004, pp. 366-373, 2010.

[10] Y. Liu, C. Luo and L. Jin, "Tightness-aware Evaluation Protocol for Scene Text Detection," 2019.